Thèse de doctorat de l'**Université de Lyon**

N° d'ordre NNT : **2022LYSEC015**
Spécialité : **Mathématiques et applications**

Présentée par **Antoine Piguet**

Dirigée par **Grégory Vial**, codirigée par **Céline Helbert**
et encadrée par **Astrig Benefice** et **Guillaume Bontron**

Préparée au sein de l'**École Centrale de Lyon** et de **CNR (Compagnie Nationale du Rhône)**, dans l'école doctorale **InfoMaths (ED 512)**

Optimisation sous incertitudes par métamodèle pour la planification à court terme de la production conjointe d'une chaîne hydroélectrique et d'actifs de production variable

Thèse soutenue publiquement le **28 avril 2022** devant le jury composé de :

Marianne Clausel, Professeure des universités, Université de Lorraine IECL, Rapporteure

Mathilde Mougeot, Professeure des universités, ENSIIE, Rapporteure

Olivier Daxhelet, Chief expert Sustainability Solutions EMEAI, Engie Impact, Examinateur

Fabrice Gamboa*, Professeur des universités, Institut Mathématiques de Toulouse, Examinateur

Yannig Goude, Ingénieur-chercheur, EDF R&D, Examinateur

Mélodie Mouffe, Senior analyst Sustainability Solutions EMEAI, Engie Impact, Invitée

Astrig Benefice, Cheffe de projets, CNR, Encadrante de thèse

Guillaume Bontron, Responsable de projets, CNR, Encadrant de thèse

Céline Helbert, Maîtresse de conférences, École Centrale de Lyon, Co-directrice de thèse

Grégory Vial, Professeur des universités, École Centrale de Lyon, Directeur de thèse

* : *Président du jury*

Optimisation sous incertitudes par métamodèle pour la planification à court terme de la production conjointe d'une chaîne hydroélectrique et d'actifs de production variable

Antoine Piguet

28 avril 2022



Résumé

Optimisation sous incertitudes par métamodèle pour la planification à court terme de la production conjointe d'une chaîne hydroélectrique et d'actifs de production variable.

La planification à court terme de la production conjointe d'une chaîne hydroélectrique et d'actifs de production variable est un problème d'optimisation. Le programme de production hydroélectrique est construit à un horizon de quelques jours pour respecter les contraintes d'exploitation et pour maximiser le chiffre d'affaires obtenu par la vente sur le marché *day-ahead* et par la pénalisation des écarts, en considérant la production totale de tous les actifs.

Le travail de la thèse consiste à prendre en compte les incertitudes qui entachent les prévisions des apports en eau, des prix de l'électricité et des productions variables dans le problème d'optimisation. À partir d'une modélisation des incertitudes par un ensemble fini de scénarios multivariés, le problème d'optimisation en univers probabiliste proposé consiste en un problème de programmation dynamique stochastique à deux niveaux linéaires mixtes.

La problématique des temps de calcul étant cruciale pour un usage opérationnel, le problème d'optimisation probabiliste est résolu numériquement en remplaçant la valeur optimale du second niveau par un métamodèle calibré par apprentissage supervisé durant une phase de prétraitement. Pour calibrer ce métamodèle, les données d'entrée sont simplifiées par des approches spécifiques. Un échantillonnage par analogie permet d'abord d'approcher le domaine de la donnée d'entrée associée aux variables de décision du premier niveau. Les données d'entrée fonctionnelles sont ensuite réduites par analyse en composantes principales. Un plan d'expérience peut ainsi être construit par échantillonnage par hypercube latin pour former le jeu de données nécessaire à l'apprentissage du métamodèle. Plusieurs métamodèles linéaires sont ensuite proposés.

La méthodologie proposée est testée sur un cas d'étude réel simplifié. Le métamodèle obtenu par régression linéaire sur les données d'entrée réduites donne des performances acceptables et il permet d'obtenir une solution du problème d'optimisation en univers probabiliste. Néanmoins, en l'absence de productions variables dans le cas d'étude, le problème d'optimisation en univers probabiliste ne permet pas d'apporter des gains significatifs par rapport à celui en univers déterministe. En outre, les temps de calcul ne permettent pas une utilisation en opérationnel sans calcul distribué. Plusieurs pistes de recherche sont toutefois proposées pour améliorer la méthodologie. Une validation sur plusieurs cas d'étude, voire sur un horizon roulant, permettrait d'estimer les performances de la méthodologie de manière plus robuste.

Mots-clés : programmation linéaire mixte, programmation stochastique à deux niveaux, programmation dynamique stochastique, prévision d'ensemble, prévision par analogie, réduction de données fonctionnelles, réduction de scénarios, plan d'expérience, régression linéaire.

Abstract

Optimization under uncertainty with a surrogate model for the short-term production planning of a hydropower cascade combined with variable energy sources.

Short-term production planning of a hydropower cascade combined with variable energy sources is an optimization problem. The hydropower production plan is constructed within a time horizon of a few days to comply with operational requirements and to maximize revenues obtained by selling the total production from all energy sources on the day-ahead market and by penalizing imbalances.

This PhD thesis focuses on the optimization problem considering data uncertainty on water inflows, electricity prices and variable generation, modeled by a finite set of multivariate scenarios. The proposed optimization problem in the stochastic framework is then written with a two-stage stochastic dynamic mixed-integer linear programming formulation.

Since computation time is a crucial issue for an operational use, the optimization problem is solved by replacing the value function of the second stage with a surrogate model fitted by supervised learning during a pre-processing step. The surrogate model is fitted on input data that are simplified by specific approaches. The domain of the input data related to the decision variables of the first stage is first estimated by analogue-based sampling, and the functional inputs are reduced by principal components analysis. The learning data set can then be obtained from a design of experiments using latin hypercube sampling. Several linear models are finally proposed for the surrogate model.

The proposed methodology is tested on a real but simplified case study. The surrogate model obtained by linear regression on the reduced inputs provides acceptable performance, and it can be used to get a solution to the optimization problem in the stochastic framework. Since variable generation is not considered in the case study, the stochastic framework provides, however, minor profit compared to the deterministic framework. Besides, computation time is not compatible with an operational use without distributed computing. Several research tracks are then proposed to improve the methodology. A validation procedure on several case studies or on a rolling horizon could be used to assess the performance of the methodology in a more robust way.

Keywords: mixed-integer linear programming, two-stage stochastic programming, stochastic dynamic programming, ensemble forecasting, analog forecasting, functional reduction, scenario reduction, design of experiments, linear regression.

Table des matières

Introduction générale **1**

I Contexte et formulation mathématique **5**

1 Contexte et problématique **7**
1.1 Généralités sur le mix électrique français 7
1.1.1 Le mix électrique français 7
1.1.2 Les énergies variables 9
1.1.3 L'énergie hydraulique 10
1.2 Généralités sur le marché de l'énergie 13
1.2.1 Équilibre entre production et consommation 13
1.2.2 Marchés de l'électricité 15
1.2.3 Services système et mécanisme d'ajustement 18
1.2.4 Écarts entre production réelle et engagements 19
1.2.5 Agrégation 20
1.2.6 Mécanisme de capacité 20
1.3 Gestion opérationnelle d'actifs de production 21
1.3.1 Planification de la production 21
1.3.2 Vente de la production 22
1.3.3 Problème d'optimisation 26
1.4 Gestion sous incertitudes 27
1.4.1 Impact des incertitudes sur la planification de la production 27
1.4.2 Vers un problème d'optimisation en univers probabiliste 28
1.4.3 Corrélations des paramètres incertains 28
1.4.4 Identification des paramètres incertains d'intérêt 29
1.4.5 Quantification et modélisation des incertitudes 29
1.4.6 Adaptation et résolution du problème d'optimisation en univers probabiliste 35
1.5 Objectifs de la thèse 36
1.6 Conclusion du chapitre 38

2 Données et cas d'étude **39**
2.1 La chaîne hydroélectrique du Rhône 39
2.1.1 Le Rhône et CNR 39
2.1.2 Structure des aménagements hydroélectriques du Rhône 42
2.2 Modélisation du Rhône 42
2.2.1 Modèle hydraulique 43
2.2.2 Modulation de la production 48
2.2.3 Contraintes d'exploitation 50
2.2.4 Modèle de simulation 50

2.3 Planification à court terme de la production hydroélectrique du Rhône 52
2.3.1 Programme de production 52
2.3.2 Placement de l'énergie 54
2.3.3 Calcul du chiffre d'affaires 55
2.3.4 Prévisions 56
2.3.5 Outil d'optimisation 59
2.4 Vers une gestion du Rhône en univers probabiliste 63
2.4.1 Impact des erreurs de prévision des apports en eau et des prix 63
2.4.2 Gestion conjointe du Rhône et des actifs de production variable 64
2.4.3 Prise en compte des incertitudes dans l'optimisation 67
2.4.4 Modélisation des incertitudes 68
2.5 Cas test étudié durant la thèse 77
2.5.1 Situation étudiée 77
2.5.2 Simplifications 78
2.5.3 Prévisions 79
2.6 Conclusion du chapitre 84

3 Problème de planification à court terme de la production 87
3.1 Généralités sur l'optimisation 87
3.1.1 Problème d'optimisation 87
3.1.2 Programmation linéaire 89
3.1.3 Programmation linéaire mixte 90
3.1.4 Optimisation sous incertitudes 91
3.2 Formulation mathématique du problème d'optimisation 99
3.2.1 Description 99
3.2.2 Scénario 100
3.2.3 Solutions admissibles 101
3.2.4 Chiffre d'affaires 106
3.2.5 Optimiseur 107
3.2.6 Solution déterministe 108
3.3 Prise en compte des incertitudes sur le scénario 110
3.3.1 Modélisation des incertitudes sur le scénario 110
3.3.2 Comportement de l'optimiseur déterministe face aux incertitudes 112
3.3.3 Quelle méthode d'optimisation sous incertitudes ? 115
3.3.4 Problème d'optimisation sous incertitudes 121
3.3.5 Optimisation en univers déterministe ou probabiliste ? 125
3.4 Conclusion de la première partie 128

II Résolution numérique par métamodèle 129

4 Méthodologie générale et transformation des données d'entrée 131
4.1 Description méthodologique 131
4.1.1 Rappel du problème d'optimisation probabiliste 131
4.1.2 Résolution numérique du problème d'optimisation probabiliste 132
4.1.3 Méthodologie générale proposée 134
4.2 Réduction de la dimension des données fonctionnelles 136
4.2.1 Problématique 136
4.2.2 Approches de réduction de la dimension 137
4.2.3 Comparaison des données fonctionnelles 140
4.2.4 Application aux scénarios du cas d'étude 141
4.3 Réduction des scénarios 147

4.3.1 Problématique . . . 147
4.3.2 Algorithmes testés . . . 148
4.3.3 Critères de qualité de la réduction . . . 149
4.3.4 Application au cas d'étude . . . 152
4.3.5 Comparaison sur un historique de vérification . . . 156
4.3.6 Algorithme *tubing* . . . 159
4.4 Conclusion du chapitre . . . 161

5 Plan d'expérience et métamodèle 163
5.1 Approximation par métamodèle . . . 163
5.1.1 Rappel de la méthodologie . . . 163
5.1.2 Construction du métamodèle . . . 165
5.1.3 Étapes de la phase de pré-traitement . . . 166
5.2 Discrétisation des prochaines ventes *day-ahead* . . . 167
5.2.1 Problématique et méthodologie . . . 167
5.2.2 Hyperparamètres pour la prévision par analogie . . . 170
5.2.3 Qualité de discrétisation . . . 173
5.2.4 Application au cas d'étude . . . 178
5.3 Plan d'expérience . . . 181
5.3.1 Problématique . . . 181
5.3.2 Échantillonnage par hypercube latin . . . 182
5.3.3 Jeu de données . . . 185
5.4 Régression . . . 188
5.4.1 Choix de la méthode d'apprentissage . . . 188
5.4.2 Choix des entrées du métamodèle linéaire . . . 191
5.4.3 Critères de qualité intrinsèque du métamodèle . . . 191
5.4.4 Apprentissage du métamodèle . . . 193
5.5 Conclusion du chapitre . . . 201

6 Conclusions méthodologiques et perspectives 203
6.1 Résumé de la méthodologie proposée . . . 203
6.1.1 Problématique . . . 203
6.1.2 Formulation mathématique du problème d'optimisation probabiliste . . . 204
6.1.3 Résolution numérique à l'aide d'un métamodèle . . . 205
6.2 Validation de la méthodologie proposée . . . 205
6.2.1 Résolution du problème d'optimisation probabiliste approché . . . 205
6.2.2 Erreurs d'approximation de la fonction objectif . . . 210
6.2.3 Temps de calcul . . . 213
6.2.4 La méthodologie proposée est-elle satisfaisante ? . . . 218
6.3 Perspectives générales . . . 219
6.3.1 Validation robuste de la méthodologie . . . 219
6.3.2 Corrélations entre les données d'entrée . . . 220
6.3.3 Choix du métamodèle . . . 221
6.3.4 Généralisation de la méthodologie . . . 223

Conclusion générale 225

Bibliographie 227

A Compléments sur les scénarios 241
A.1 Étude des corrélations . . . 241
A.1.1 Théorie sur les corrélations croisées . . . 241

A.1.2 Application 242
A.2 Qualité de la prévision probabiliste des prix 245
A.3 Réduction des scénarios par distance de Kantorovich 247
A.3.1 Notations générales 247
A.3.2 Distance de Kantorovich entre deux probabilités finies 248
A.3.3 Redistribution optimale 257
A.3.4 Lien avec la quantification et la classification de Voronoï 260
A.3.5 Choix optimal des scénarios à supprimer à nombre fixé 263
A.3.6 Algorithmes numériques 265
A.4 Génération de nouveaux scénarios 266

Introduction générale

La gestion d'actifs de production électrique repose sur la planification de leur production, principalement pour la vendre sur le marché de l'énergie plusieurs heures à plusieurs années en avance. Lorsque les actifs permettent beaucoup de latitude pour décider de leur production (par exemple les centrales hydroélectriques dotées d'une grande retenue, les centrales à gaz ou charbon, ...), la production est peu contrainte par des éléments extérieurs et peut être programmée en avance. Dans ce cas, le programme de production est généralement construit pour optimiser la valorisation économique en maximisant le chiffre d'affaires obtenu par la vente de la production sur le marché et en minimisant les coûts de production. On forme alors un problème d'optimisation, appelé *problème d'attribution d'unités de production*, pour lequel la littérature est abondante.

En revanche, le problème se complexifie dès lors que les productions des actifs sont fortement dépendantes et que la latitude pour décider de leur production est très limitée. Dans ce cas, la production est en partie contrainte par des éléments extérieurs. C'est par exemple le cas pour une chaîne hydroélectrique au fil de l'eau le long d'un fleuve, comme celle constituée par les 18 aménagements hydroélectriques exploités par CNR (Compagnie Nationale du Rhône) le long du Rhône. L'optimisation doit alors être effectuée sur l'ensemble de la chaîne hydroélectrique. Comme les capacités de stockage des aménagements de la chaîne hydroélectrique sont très faibles, l'horizon d'intérêt est celui à court terme, permettant de planifier la production quelques jours en avance. Le problème d'optimisation repose donc sur une modélisation détaillée de la chaîne hydroélectrique et des contraintes d'exploitation qui assurent la sûreté hydraulique (par exemple les cotes minimales et maximales de l'eau dans les retenues à ne pas dépasser). Il est défini à partir d'une prévision des apports en eau (issus notamment des affluents) pour déterminer les débits entrants aux aménagements hydroélectriques, ainsi qu'une prévision des prix ou des coûts pour déterminer la valeur économique d'un programme de production. Une telle planification à court terme s'inscrit dans le processus opérationnel mené quotidiennement pour gérer la chaîne hydroélectrique. En particulier, le problème d'optimisation doit être résolu dans un temps très limité. Un tel optimiseur est utilisé quotidiennement à CNR comme outil d'aide à la décision pour appuyer l'équipe qui planifie la production du Rhône.

Le problème se complexifie encore davantage lorsque l'on souhaite optimiser la production de la chaîne hydroélectrique conjointement avec celle d'actifs de production variable (par exemple des parcs éoliens, des centrales photovoltaïques). L'intérêt d'une telle gestion conjointe est de pouvoir utiliser la faible flexibilité de production de la chaîne hydroélectrique comme moyen pour compenser, le cas échéant, les écarts dus aux erreurs de prévision des productions variables.

En univers déterministe, les prévisions donnent, pour chaque échéance, une seule valeur pour les apports en eau, les prix/coûts et les productions variables. Elles ignorent donc les incertitudes sur leurs réalisations effectives, i.e. les écarts possibles entre les valeurs prévues et les réalisations effectives connues en temps réel. L'optimalité des programmes de production renvoyés par l'optimiseur en univers déterministe peut alors être fortement remise en question du fait des incertitudes réelles entachant les prévisions. Ces programmes peuvent non seulement s'avérer peu

optimaux en termes de valeur économique avec les réalisations effectives, mais aussi nécessiter des corrections qui viennent grever la valeur économique afin d'assurer à tout instant le respect des contraintes d'exploitation. Or, ces incertitudes de prévision peuvent être estimées en avance, par exemple grâce à des prévisions d'ensemble permettant de former un ensemble fini de scénarios multivariés des paramètres incertains munis de leurs probabilités. On peut donc envisager l'optimisation en univers probabiliste pour tenir compte des incertitudes de prévision. Même si la littérature est abondante pour la planification d'actifs de production en univers probabiliste de manière générale, elle est assez rare pour la planification à court terme, dans un contexte opérationnel, de la production conjointe d'une chaîne hydroélectrique et d'actifs de production variable, en considérant les incertitudes modélisées à partir de prévisions d'ensemble.

Le travail de la thèse consiste à proposer une méthodologie pour faire passer l'optimiseur d'un univers déterministe à un univers probabiliste, en nous appuyant sur la gestion de la production à l'échelle de CNR. Nous cherchons à répondre à deux problématiques principales.

La première problématique abordée dans la thèse concerne l'adaptation de la formulation mathématique du problème d'optimisation de l'univers déterministe à l'univers probabiliste. La difficulté est en effet de mettre en place une stratégie de valorisation sous incertitudes qui n'existe pas encore à CNR dans le processus opérationnel. Compte tenu de l'enchaînement des planifications au cours du temps dans le processus opérationnel, nous proposons d'utiliser une formulation par programmation dynamique stochastique à deux niveaux linéaires mixtes. L'originalité de cette formulation réside dans le choix de chercher les solutions dans le même ensemble admissible que celui de la formulation en univers déterministe. Les incertitudes ne sont donc utilisées dans la fonction objectif que pour évaluer la valeur optimale du second niveau du problème d'optimisation probabiliste, laquelle représente une estimation du chiffre d'affaires des re-planifications ultérieures éventuellement nécessaires en fonction de l'évolution effective du scénario. Un autre point original dans la formulation mathématique proposée est le choix de porter la dynamique entre les deux niveaux sur les prochaines ventes *day-ahead*, et non sur les pas de temps comme c'est généralement le cas dans la littérature pour la gestion sous incertitudes de centrales hydroélectriques.

La seconde problématique abordée dans la thèse concerne la résolution numérique du problème d'optimisation en univers probabiliste. Le passage de l'univers déterministe à l'univers probabiliste complexifie en effet le problème d'optimisation, rendant sa résolution numérique inaccessible en pratique avec un temps de calcul adapté à une utilisation en opérationnel. Nous proposons de remplacer la valeur optimale du second niveau du problème d'optimisation probabiliste, coûteuse à évaluer numériquement, par un métamodèle rapide à évaluer. Ce métamodèle est obtenu par apprentissage supervisé dans une phase de pré-traitement à l'optimisation, où le jeu de données nécessaire à l'apprentissage est obtenu grâce à un plan d'expérience. Bien que l'utilisation d'un métamodèle pour accélérer la résolution d'un problème d'optimisation ait déjà été abordée dans la littérature, une telle démarche est, à notre connaissance, peu commune pour l'optimisation probabiliste à deux niveaux. Une des difficultés rencontrées est que le domaine de définition d'une des données d'entrée du métamodèle n'est pas connu sous une forme explicite. Nous proposons d'approcher ce domaine grâce à un échantillonnage par analogie. Une autre difficulté est que les données d'entrée sont fonctionnelles, si bien que nous proposons de réduire leur dimension par analyse en composantes principales. Une réduction du nombre des scénarios est également étudiée dans le but de simplifier les données d'entrée du métamodèle.

Ce manuscrit de thèse est divisé en deux parties.

La première partie est consacrée à la présentation du contexte dans lequel s'inscrit le travail de la thèse et à la formulation mathématique du problème d'optimisation en univers probabiliste. Le chapitre 1 donne une vision générale de la gestion opérationnelle d'actifs de production en

lien avec le marché de l'énergie. Nous y présentons le rôle de l'optimisation de la production et de la maîtrise des incertitudes. Ces éléments de contexte sont utilisés en fin de chapitre pour définir plus précisément le cadre de travail et les objectifs de la thèse. Le chapitre 2 présente le contexte plus spécifique de la planification de la production à CNR, ainsi que les données utilisées pour tester et illustrer la méthodologie proposée tout le long du manuscrit. Ces deux premiers chapitres apportent peu d'éléments mathématiques mais ils dressent un panorama assez exhaustif du contexte général dans lequel s'inscrit la thèse et qui intéressera principalement les lecteurs et lectrices du domaine de l'énergie. Le chapitre 3 concerne la formulation mathématique du problème d'optimisation en univers probabiliste. Nous nous appuyons pour cela sur le modèle de simulation et la stratégie de valorisation présentés dans le chapitre 2 pour le contexte de CNR, mais en essayant de nous placer dans un cadre légèrement plus général.

La seconde partie est consacrée à la résolution numérique du problème d'optimisation en univers probabiliste. Le chapitre 4 présente la méthodologie générale de résolution à l'aide d'un métamodèle dans une phase de pré-traitement à l'optimisation. Ce chapitre est également consacré à deux problématiques qui interviennent à plusieurs étapes de la méthodologie proposée : la réduction des données fonctionnelles et la réduction des scénarios. Le chapitre 5 décrit les étapes successives de la construction du métamodèle dans une phase de pré-traitement à l'optimisation. La méthodologie est ensuite testée dans le chapitre 6. Ce dernier chapitre permet également de faire une synthèse de la méthodologie proposée et de conclure sur l'intérêt de la méthodologie avant de donner quelques pistes d'amélioration.

Par ailleurs, l'annexe A en fin de manuscrit présente quelques compléments sur les scénarios.

Ce manuscrit de thèse est volontairement rédigé pour qu'il puisse être lu à plusieurs niveaux, selon l'intérêt des lecteurs et lectrices pour le domaine de l'énergie ou pour les aspects mathématiques. Nous rappelons en particulier certains éléments importants tout le long du manuscrit, quitte à être parfois redondants. Les idées principales développées dans chaque chapitre peuvent donc être comprises sans avoir lu les chapitres précédents, même si les chapitres suivent un ordre logique avec quelques renvois.

Première partie

Contexte et formulation mathématique

Chapitre 1

Contexte et problématique

La gestion opérationnelle d'un ensemble d'actifs de production électrique repose sur de nombreuses problématiques, dont nous présentons les plus importantes dans ce chapitre introductif. Ce premier chapitre intéressera principalement les lecteurs et lectrices qui souhaitent comprendre le cadre de travail de la thèse de manière exhaustive. Dans une première section, nous présentons un bref aperçu du mix électrique français et de la diversité des moyens de production. Dans une deuxième section, nous présentons les éléments clefs du fonctionnement du marché de l'énergie, en lien avec l'équilibre à tout instant entre la consommation et la production. Dans une troisième section, nous présentons le rôle de l'optimisation de la production et de la maîtrise des incertitudes dans la gestion opérationnelle, en lien avec le marché de l'énergie. À la fin de ce chapitre, nous nous appuyons sur ces éléments de contexte pour définir plus précisément les objectifs de la thèse, le point essentiel de ce chapitre.

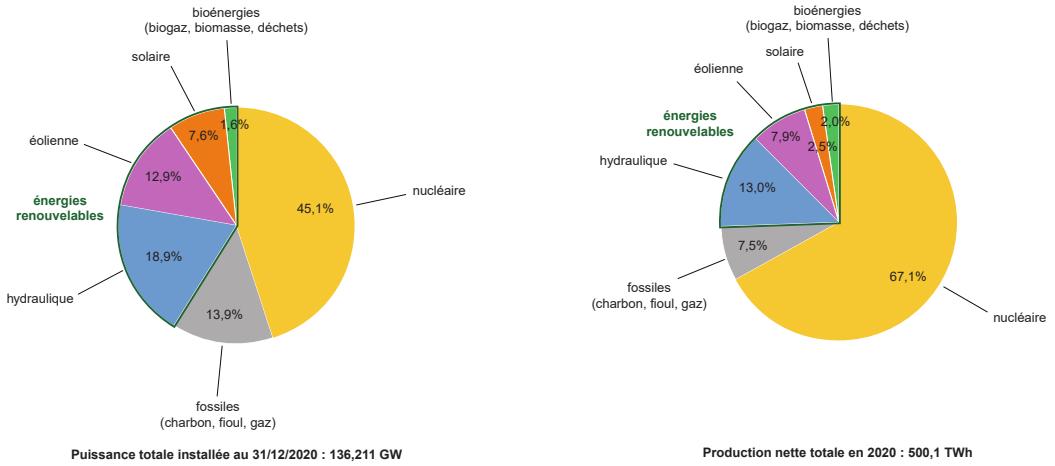
1.1 Généralités sur le mix électrique français

1.1.1 Le mix électrique français

L'électricité est une source d'énergie secondaire. Elle est produite à partir de sources d'énergie primaire (éolienne, solaire, nucléaire, charbon, gaz naturel, pétrole etc.) par des centrales électriques puis elle est convertie sous d'autres formes d'énergie finale (énergie mécanique, lumière, chaleur etc.). Comme l'électricité est transportable sur de longues distances à l'aide de câbles conducteurs, la conversion en énergie finale peut s'effectuer loin des lieux de production. On dit que l'électricité est un vecteur d'énergie.

L'ensemble des énergies primaires utilisées pour la production d'électricité constitue le *mix électrique*. La figure 1.1 illustre, par exemple, la composition du mix électrique français en 2020. Une source primaire d'électricité est dite *renouvelable* si son exploitation n'entraîne pas l'épuisement de ressources naturelles à l'échelle de temps humain (les ressources naturelles ainsi utilisées sont disponibles en grande quantité et elles sont renouvelées assez rapidement par rapport à l'échelle de temps humain). En France, les énergies renouvelables principalement utilisées pour la production d'électricité sont l'énergie hydraulique, l'énergie éolienne, l'énergie solaire et les bioénergies (biogaz, biomasse, déchets). La part des énergies renouvelables dans la production totale d'électricité en 2020 en France était de 25,4%. Une source primaire d'électricité est dite *décarbonnée* si elle ne rejette pas directement de dioxyde de carbone dans l'atmosphère durant sa phase d'exploitation [1]. L'énergie nucléaire et les énergies renouvelables en France sont décarbonnées (on considère que les émissions de dioxyde de carbone pour les bioénergies sont compensées

1. Les phases en amont et en aval de la phase d'exploitation, comme la fabrication ou le démantèlement des centrales électriques par exemple, peuvent néanmoins être émettrices de dioxyde de carbone.



(a) Puissance totale installée au 31/12/2020 (b) Énergie totale produite en 2020

FIGURE 1.1 – Mix électrique français en 2020. Source : `https://bilan-electrique-2020.rte-france.com/production-production-totale/`.

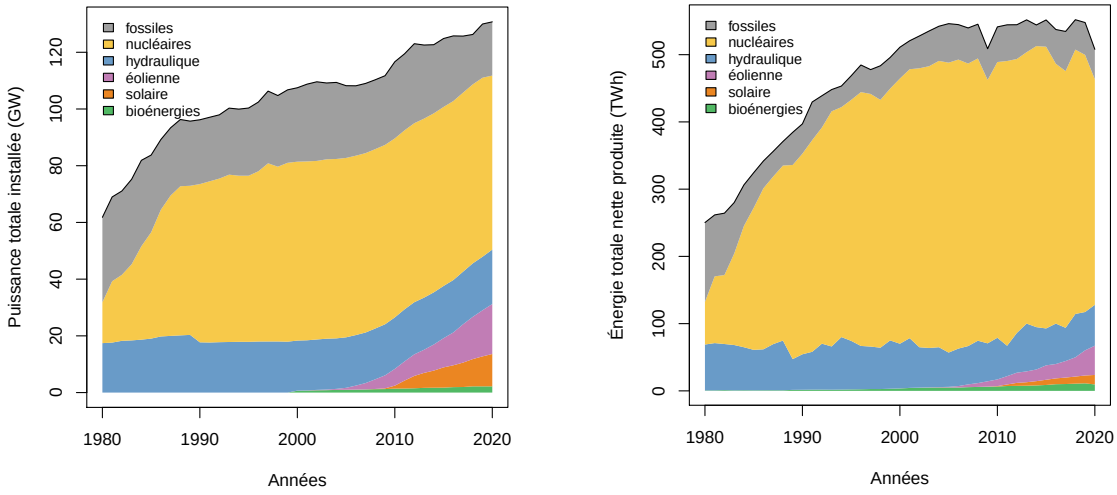
par les absorptions de dioxyde de carbone par les matières organiques qui se renouvellent naturellement). Les énergies fossiles (charbon, fioul, gaz) sont en revanche carbonnées. La part des énergies décarbonnées dans la production totale d'électricité en 2020 en France était de 92,5% (voir Figure 1.1b).

Le rejet massif de dioxyde de carbone par les activités humaines étant une des principales causes du dérèglement climatique [2], des accords internationaux (par exemple à la COP21 [3]) et des lois européennes et nationales relatives à la transition énergétique ont vu le jour ces dernières années pour davantage développer les énergies décarbonnées en substitution des énergies fossiles. En particulier, les Programmations Pluriannuelles de l'Énergie (PPE [4]), prévues par la loi relative à la transition énergétique pour la croissance verte (LTECV) adoptée en 2015, fixent des objectifs pour atteindre la neutralité carbone en 2050 en France.

Pour atteindre ces objectifs, des dispositifs de soutien public, mis en place par l'État et gérés par la Commission de Régulation de l'Énergie (CRE), permettent de soutenir le développement des énergies renouvelables et des solutions de flexibilité en France (stockage, pilotage de la consommation etc., voir Sous-section 1.2.1). Ces dispositifs permettent notamment de garantir une rentabilité aux investisseurs selon deux modalités de rémunération : l'obligation d'achat ou le complément de rémunération.

L'obligation d'achat [5] est le mécanisme historique qui proposait aux producteurs un prix de vente fixe (défini par l'État) intéressant par rapport au marché pour permettre de rentabiliser les investissements. La période de l'obligation d'achat est de 15 ans pour les parcs éoliens et 20 ans pour les centrales photovoltaïques. Après cette période, la production de ces actifs doit être vendue sur le marché (voir Sous-section 1.2.2), comme n'importe quelle production d'autres actifs. Aujourd'hui, le dispositif d'obligation d'achat n'est proposé que pour l'installation ou la modernisation de centrales de production renouvelable de petite taille (certaines Petites Centrales Hydroélectriques (PCH), certaines centrales photovoltaïques sur toiture, ou certains actifs de production innovants par exemple).

2. `https://www.ipcc.ch/`
3. `https://www.gouvernement.fr/action/la-conference-de-paris-sur-le-climat`
4. `https://www.ecologie.gouv.fr/programmations-pluriannuelles-lenergie-ppe`
5. `https://www.centrales-next.fr/glossaire-energies-renouvelables/obligation-achat/`



(a) Puissance installée par année

(b) Énergie totale produite par année

FIGURE 1.2 – Évolution du mix électrique français dans le temps (les sources d'énergie sont empilées par coût marginal croissant moyen en 2020). Source : `https://db.nomics.world/`.

Le complément de rémunération [6] est le mécanisme, introduit par la LTECV, qui remplace l'obligation d'achat pour l'installation de nouvelles grandes centrales de production renouvelable. Avec ce nouveau dispositif, les productions renouvelables entrent directement dans le marché, mais les producteurs perçoivent une compensation financière si les prix de vente au marché sont inférieurs à un tarif seuil fixé par l'État (ancien tarif de l'obligation d'achat). À cette compensation s'ajoute une prime pour prendre en compte les surcoûts liés à la participation au marché (coût des prévisions pour vendre l'énergie à l'avance, pénalité des écarts en cas de déséquilibre entre production et vente etc.). Le mécanisme de complément de rémunération est proposé par l'intermédiaire d'appels d'offres gérés par la CRE.

Le mix électrique français évolue donc fortement depuis les années 2000 (voir Figure 1.2), au rythme du développement des moyens de production d'énergie renouvelable, des fins de vie des centrales électriques historiques (nucléaires et fossiles) et des évolutions de la consommation électrique (le parc électrique français est globalement dimensionné par rapport à la consommation électrique française).

1.1.2 Les énergies variables

Une source primaire d'électricité est dite *variable* (ou intermittente, fluctuante, fatale) si la production n'est pas entièrement contrôlable [7] et qu'elle varie fortement dans le temps entre un état minimal et maximal. Il s'agit principalement des énergies éolienne et solaire.

L'énergie éolienne pour la production d'électricité est issue des parcs éoliens terrestres (*onshore*) ou en mer (*offshore*) qui exploitent l'énergie cinétique du vent. La production d'une éolienne dépend des caractéristiques des pales (formes, hauteur, longueur), de la vitesse du vent [8] (intensité et direction) et de la densité de l'air (liée, notamment, à la température). Pour plus de détails sur l'énergie éolienne, voir par exemple [Rapin et Noël, 2010].

6. `https://www.centrales-next.fr/glossaire-energies-renouvelables/complement-de-remuneration/`

7. Les énergies variables peuvent être très partiellement contrôlées à la baisse (bridage des éoliennes par exemple) ou à la hausse (si on choisit un point de fonctionnement qui n'est pas nominal). Ce léger contrôle peut être notamment utilisé pour les services système (voir Sous-section 1.2.3).

8. La production produite par une éolienne est proportionnelle au cube de la vitesse du vent (avant le bridage ou l'arrêt automatique de l'éolienne en cas de forts vents).

L'énergie solaire pour la production d'électricité est issue des centrales photovoltaïques implantées principalement sur les sols et les toitures. Des centrales photovoltaïques sont aussi développées sur des plans d'eau (photovoltaïque flottant) et au-dessus de cultures agricoles (agrivoltaïsme). Ces centrales exploitent l'énergie du rayonnement solaire qui arrive sur la surface des panneaux photovoltaïques qui, selon le trajet suivi, est direct, diffus (par les nuages et l'atmosphère) ou réfléchi (par le sol par exemple). Une grande partie du spectre des rayonnements solaires peut être exploitée par les matériaux semi-conducteurs (principalement le silicium) qui composent les panneaux photovoltaïques pour produire de l'électricité. La production d'un panneau photovoltaïque dépend de l'intensité du rayonnement reçu (liée à la nébulosité, à la position du Soleil dans le ciel, à la présence d'aérosols etc.), de l'angle d'incidence du rayonnement reçu et de la température du panneau. La puissance installée d'une centrale photovoltaïque est souvent exprimée en puissance crête, i.e. la puissance qu'elle peut produire dans des conditions standards d'éclairement et d'orientation. Pour plus de détails sur l'énergie solaire, voir par exemple [Labouret et Villoz, 2006].

Le caractère variable des énergies éolienne et solaire provient donc de leur dépendance aux conditions météorologiques. Il convient de noter que ces deux énergies ne sont pas totalement indépendantes car le vent a pour origine les différences de température et de pression dans l'atmosphère qui dépendent en partie de l'absorption du rayonnement solaire par l'atmosphère. En outre, la forte variabilité de ces énergies constatée lorsque l'on considère un unique actif de production se réduit considérablement si on les considère à l'échelle de la France car les actifs de production sont localisés dans des zones où les conditions météorologiques sont différentes, voire indépendantes (couloirs de vents différents, ensoleillement différents etc.). On dit qu'il y a *foisonnement* des énergies variables.

Les parcs éoliens et les centrales photovoltaïques se comparent généralement à partir de leur *facteur de charge*. Le facteur de charge d'un actif de production est par définition le rapport entre l'énergie produite par cet actif sur une période donnée et l'énergie qu'il aurait produit sur cette même période s'il avait fonctionné à sa puissance maximale (puissance installée).

1.1.3 L'énergie hydraulique

L'énergie hydraulique pour la production d'électricité est issue des aménagements hydroélectriques qui exploitent l'énergie mécanique de masses d'eau en mouvement (due à la gravité et l'écoulement) pour la convertir en énergie électrique à l'aide d'ensembles turbines-alternateurs appelés *groupes de production*. Les centrales hydroélectriques tirent profit d'un dénivelé pour créer une retenue d'eau en amont et une hauteur de chute au niveau des groupes de production (différence entre le niveau d'eau en amont et celui en aval).

1.1.3.1 Catégories des centrales hydroélectriques

Les centrales hydroélectriques sont classées en trois grandes catégories, en fonction de leur hauteur de chute (et donc aussi du type de turbine).

— Les centrales de haute chute (ou de lac) ont une hauteur de chute supérieure à 200 m. Elles sont souvent construites dans des zones de haute montagne, pour tirer profit d'un dénivelé important et d'un grand lac de retenue créé par un barrage et alimenté par les torrents et la fonte nivale et glacière. L'eau est évacuée du barrage dans des conduites forcées jusqu'à la centrale située en contre-bas du barrage où elle est turbinée par des turbines de type Pelton. Le débit en amont du barrage est assez faible mais la hauteur de chute est assez importante.

— Les centrales de moyenne chute (ou d'éclusée) ont une hauteur de chute entre 30 et 200 m. Elles sont souvent construites dans des zones de moyenne montagne sur le cours de fleuves

ou rivières. La retenue est créée par un barrage, et la centrale, généralement située au pied du barrage, turbine l'eau par des turbines de type Francis. Le débit en amont du barrage et la hauteur de chute sont moyens.

— Les centrales de basse chute (ou au fil de l'eau) ont une hauteur de chute inférieure à 30 m. Elles sont souvent construites sur le cours de fleuves ou rivières. Si le dénivelé naturel le permet, la centrale hydroélectrique est directement construite sur le cours d'eau. Sinon, la grande partie du cours d'eau naturel est déviée par un barrage vers un canal de dérivation qui est creusé pour cumuler tout le dénivelé en un point où est construite la centrale. L'eau est turbinée à la centrale par des turbines de type Kaplan ou bulbe. Le débit en amont de la centrale est assez fort mais la hauteur de chute est assez faible. Une centrale de basse chute est dite *parfaitement au fil de l'eau* s'il n'est pas possible de moduler le débit sortant de la centrale par rapport au débit entrant. C'est généralement le cas des Petites Centrales Hydroélectriques (PCH) installées sur le cours de rivières dans des zones de faibles dénivelés (sur les barrages des aménagements basse chute par exemple).

Pour plus de détails sur l'énergie hydraulique, notamment sur les types de turbines, voir par exemple [Ginocchio et Viollet, 2012].

Il existe aussi, à plus petite échelle, des aménagements hydroélectriques qui tirent profit du mouvement des masses d'eau marines [CETIM, 2009] :

— les hydroliennes exploitent les courants marins en profondeur,

— les centrales marémotrices [9] exploitent le courant et le marnage des marées pour remplir un bassin et créer une hauteur de chute,

— les centrales houlomotrices exploitent l'énergie transportée par la houle (il existe plusieurs procédés).

1.1.3.2 Production hydroélectrique

L'eau qui traverse une centrale hydroélectrique est turbinée par les groupes de production, i.e. les ensembles constitués d'une turbine et d'un alternateur. La puissance active brute instantanée produite par un groupe de production s'écrit (voir [Ginocchio et Viollet, 2012])

$$P = \eta(Q)\rho g H(Z,Q)Q, \tag{1.1}$$

où :

— $\rho = 1\,000\,\mathrm{kg/m^3}$ est la masse volumique de l'eau (supposée constante) et $g = 9{,}81\,\mathrm{m/s^2}$ l'accélération de la pesanteur (supposée constante),

— $Q \in \mathbb{R}_+$ le débit qui traverse le groupe de production,

— $Z \in \mathbb{R}_+$ le niveau de la retenue (hauteur de l'eau),

— $\eta(Q) \in [0,1]$ le rendement global du groupe de production (rendement turbine, alternateur, transformateur),

— $H(Z,Q) \in \mathbb{R}_+$ la hauteur de chute nette, i.e. la hauteur de chute en tenant compte des pertes de charges dues aux frottements de l'eau en mouvement dans les conduites [10].

Le rendement global dépend du débit tandis que la hauteur de chute nette dépend du niveau de la retenue et du débit. La puissance active brute instantanée produite par un groupe de production n'est donc pas linéaire en fonction du débit. En pratique (par exemple, en effectuant des mesures empiriques), elle a une allure en forme de cloche, voir Figure 1.3.

9. La centrale marémotrice de la Rance près de Saint-Malo en France, exploitée par EDF (Électricité De France), fut la première centrale marémotrice mondiale à l'échelle industrielle, voir `https://www.youtube.com/watch?v=jNXiwcZYMpU`. Elle tire profit des fortes marées du littoral de la Manche.

10. $H(Z,Q) = H_0(Z,Q) - \frac{\Delta p(Z,Q)}{\rho g}$, où $H_0(Z,Q) \in \mathbb{R}_+$ est la hauteur de chute brute (différence entre le niveau amont et le niveau aval) et $\Delta p(Z,Q) \in \mathbb{R}_+$ les pertes de charges.

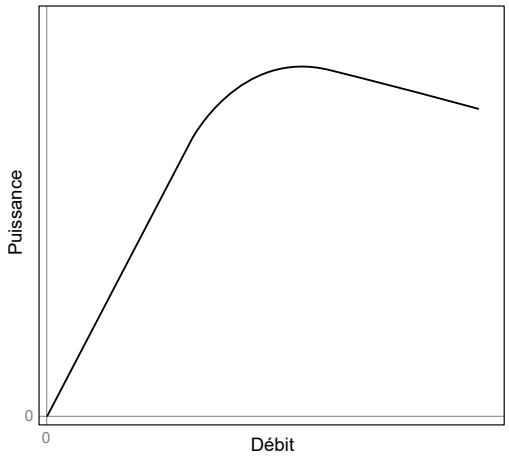


FIGURE 1.3 – Allure de la puissance instantanée produite par un groupe de production hydraulique à un niveau de la retenue fixé.

La formule de l'équation (1.1) est valable pour tous les types de turbines, donc pour les 3 catégories de centrales hydroélectriques. Pour les centrales de haute chute, la puissance est surtout issue de la hauteur de chute, alors que pour les centrales de basse chute, elle est surtout issue du débit.

1.1.3.3 Flexibilité de la production hydroélectrique

Selon la capacité de marnage (i.e. la différence entre le volume maximal et le volume minimal de la retenue), le débit sortant de la centrale hydroélectrique peut différer du débit entrant en faisant varier le niveau de la retenue. La retenue sert alors de stockage d'eau qui augmente lorsque le débit sortant est plus faible que le débit entrant et qui diminue dans le cas contraire. Comme ce stock d'eau est situé en amont de la hauteur de chute, son énergie potentielle représente un stock d'énergie pour la centrale hydroélectrique, qui permet de moduler (contrôler) le débit turbiné et donc la production. Contrairement aux actifs de production d'énergie variable, la production hydroélectrique est donc généralement *flexible*[11]. Plus la capacité de marnage de la retenue est importante, plus il est possible de moduler le débit turbiné par rapport au débit entrant de la centrale qui alimente la retenue.

Comme le débit entrant dépend généralement des conditions hydro-météorologiques (précipitation, fonte nivale et glacière etc.), cette flexibilité peut permettre d'atténuer la dépendance du débit turbiné aux conditions hydro-métérologiques, et cela d'autant plus que la capacité de marnage de la retenue est importante.

L'utilisation de la flexibilité offerte par la capacité de marnage de la retenue n'est possible qu'avec une gestion temporelle de la production. L'eau de la retenue utilisée pour augmenter le débit sortant par rapport au débit entrant doit en effet avoir été stockée antérieurement. L'horizon temporel de gestion de la production de la centrale hydroélectrique dépend alors du temps de remplissage de la retenue, donc de la capacité de marnage et de l'ordre de grandeur du débit entrant. Plus la capacité de marnage est faible ou plus les débits entrants sont élevés, plus l'échelle de temps pour la gestion de la production de la centrale hydroélectrique est réduite. En général, les centrales hydroélectriques de haute chute sont gérées à l'année, celles de moyenne chute à la semaine ou saison et celles de basse chute à la journée. Cette gestion de la production est présentée plus en détails dans la section 1.3.

11. La flexibilité d'un actif de production est par définition sa capacité à avoir une partie de sa production qui peut être modulée à la guise du producteur.

Les centrales hydroélectriques parfaitement au fil de l'eau sont le cas extrême où la flexibilité est inexistante. Dans ce cas, l'énergie produite est variable car elle dépend entièrement du débit entrant, donc des conditions hydro-météorologiques.

1.1.3.4 L'hydroélectricité au fil de l'eau

L'hydroélectricité produite à l'échelle industrielle en France est réglementée par un régime de concession [12]. Les aménagements hydroélectriques appartiennent à l'État mais leur exploitation (et leur construction) est confiée à un concessionnaire. Ce dernier se rémunère grâce à la vente de la production, mais en contrepartie, il verse une redevance à l'État. Les contrats de concession sont de durée limitée et ils portent sur un domaine hydraulique donné pouvant regrouper plusieurs aménagements hydroélectriques. Le contrôle des contrats de concession est effectué par les Directions Régionales de l'Environnement, de l'Aménagement et du Logement (DREAL).

La grande majorité des centrales hydroélectriques en France sont de moyenne ou basse chute et elles sont installées en chaîne (i.e. en série) sur les cours d'eau qui sillonnent la France. Sur une chaîne hydroélectrique, une partie du débit entrant à un aménagement provient du débit sortant de l'aménagement amont. Ainsi, comme la production d'une centrale de moyenne ou basse chute dépend des débits entrants (voir §1.1.3.3), la production d'un aménagement d'une chaîne hydroélectrique dépend des productions des autres aménagements de cette chaîne.

En outre, des contraintes d'exploitation fortes sont définies sur les aménagements d'une chaîne hydroélectrique afin de pouvoir garantir la sûreté hydraulique pour les personnes et les activités liées à l'usage du cours d'eau (navigation, irrigation, tourisme, gestion des crues etc.). La sûreté hydraulique fait l'objet d'un cadre réglementaire très strict imposé et contrôlé par les DREAL. Ces contraintes d'exploitation se caractérisent notamment par des bornes minimales et maximales sur les niveaux d'eau à ne pas dépasser.

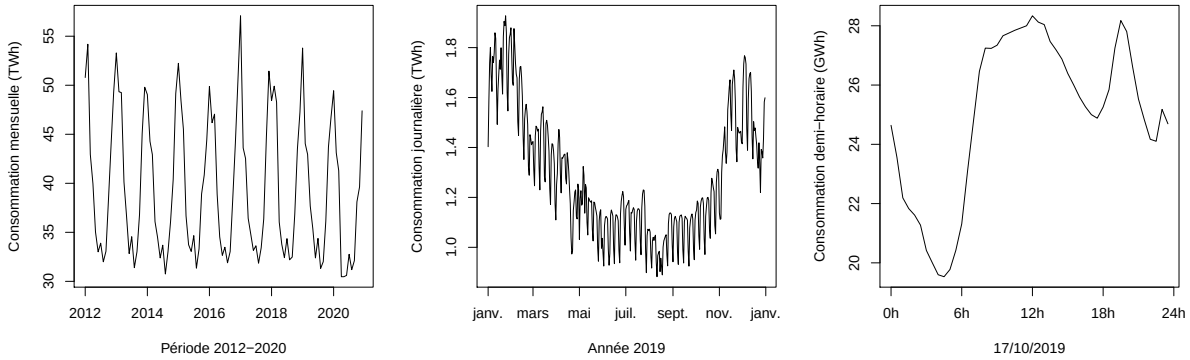
1.2 Généralités sur le marché de l'énergie

1.2.1 Équilibre entre production et consommation

La consommation électrique en France n'est pas constante dans le temps. Depuis les années 2010, la consommation annuelle est stable autour de 475 TWh, mais elle fluctue de manière cyclique selon le mois (saison), le jour et l'heure, comme l'illustre la figure 1.4. La consommation est plus importante en hiver qu'en été (voir Figure 1.4a), plus importante en semaine que les weekends (voir Figure 1.4b) et les pics de consommation au sein d'une journée sont situés autour de 8 h et 20 h (voir Figure 1.4c).

Or, la puissance totale injectée sur le réseau électrique doit être strictement égale à la puissance totale soutirée à tout instant, pour ne pas que le réseau s'effondre (*black-out*). Le gestionnaire du Réseau de Transport d'Électricité (RTE) est chargé de maintenir cet équilibre en France. La variation de la fréquence du courant alternatif du réseau, autour de la valeur de référence de 50 Hz pour le réseau européen, est une mesure de l'équilibre entre la production et la consommation. Une augmentation (respectivement diminution) de la fréquence indique une sur-production globale (respectivement une sous-production globale) par rapport à la consommation. Comme la consommation est moins pilotable que la production, l'équilibre est atteint en adaptant en permanence la puissance injectée sur le réseau à la consommation.

12. `https://www.ecologie.gouv.fr/hydroelectricite`



(a) Consommation mensuelle (b) Consommation journalière (c) Consommation demi-horaire

FIGURE 1.4 – Exemple de l'évolution de la consommation électrique en France (source RTE).

L'équilibre entre la production et la consommation peut être facilité par le stockage de l'énergie. Une partie du surplus d'énergie en période de sur-injection globale peut être stockée puis éventuellement réinjectée en période de sous-injection globale. En France, la technologie de stockage d'énergie la plus utilisée est celle des Stations de Tranfert d'Énergie par Pompage (STEP) [Rehman *et al.*, 2015]. Il s'agit d'ouvrages hydroélectriques composés de plusieurs bassins (souvent 2) à des altitudes différentes reliés entre eux par des conduites équipées de dispositifs hydroélectriques réversibles. Ces aménagements servent de stockage d'énergie en pompant l'eau d'un bassin inférieur vers un bassin supérieur puis en turbinant l'eau d'un bassin supérieur vers un bassin inférieur. Il s'agit aujourd'hui du principal moyen de stockage de l'électricité car le rendement (rapport entre l'énergie produite par turbinage et l'énergie consommée pour le pompage) est assez bon (entre 70% et 85%). La flexibilité de production des centrales hydroélectriques (voir §1.1.3.3) permet également un stockage limité de l'énergie. De même, le stockage par batterie se développe à l'échelle industrielle et le pilotage de la consommation (effacement, recharge intelligente de voiture électrique, stockage intelligent en hydrogène etc.) permet, à petite échelle, de faciliter l'équilibre en jouant sur la consommation. Néanmoins, le stockage et la flexibilité de ces technologies sont aujourd'hui très limitées à l'échelle du réseau électrique [13]. La production totale doit donc s'adapter en permanence à la consommation.

Les différents actifs de production sont alors plus ou moins sollicités en fonction de la consommation. Chaque actif de production est associé à un coût marginal qui définit le seuil du prix de vente de l'énergie à partir duquel il est rentable de le solliciter. Le coût marginal d'un actif de production est le coût de production d'une unité supplémentaire (coût opérationnel, combustible etc.) sans tenir compte des coûts fixes (entretien, exploitation et investissement initial). Il dépend donc essentiellement des coûts des combustibles (pouvant être volatils) et de l'impact sur l'environnement (empreinte carbone via la taxe carbone). Les énergies variables, dont la production serait perdue si elle n'est pas utilisée, ont des coûts marginaux très faibles. Au contraire, les énergies fossiles ont des coûts marginaux plus élevés. Il convient de noter que le coût marginal des centrales hydroélectriques de haute chute est souvent plus élevé que celui des centrales nucléaires car il correspond à la valeur de l'eau stockée dans les retenues qui reflète la limitation de la ressource en eau. En revanche, les coûts marginaux des centrales hydroélectriques de moyenne ou basse chute sont généralement inférieurs à ceux des centrales nucléaires.

13. Dans un scénario visant à atteindre un mix électrique 100% renouvelable en 2050 en France, l'ADEME (Agence De l'Environnement et de la Maîtrise de l'Énergie) suggère notamment le développement des solutions de flexibilité (stockage de l'électricité et pilotage de la consommation) [ADEME, 2015].

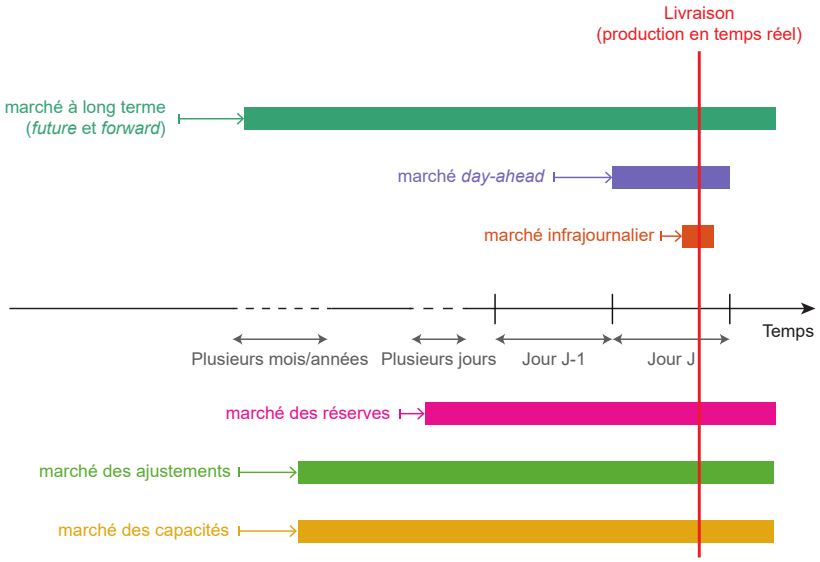


FIGURE 1.5 – Schéma des périodes des marchés de l'électricité.

1.2.2 Marchés de l'électricité

Les marchés de l'électricité font le lien entre les producteurs et les consommateurs (fournisseurs [14] et gros industriels [15]). Les produits échangés sur ces marchés correspondent à un volume d'électricité (en énergie, souvent exprimée en MWh) injecté par le vendeur et soutiré par l'acheteur sur le réseau sur une période donnée à venir. Les échanges se font de gré à gré (contrats bilatéraux ou par l'intermédiaire d'un courtier) ou en bourse. Les transactions sont toujours effectuées en avance par rapport à la livraison de la production. On distingue les marchés à long terme et ceux à court terme. À long terme, l'échéance de livraison peut porter de quelques jours jusqu'à 6 ans et les échanges sont généralement effectués de gré à gré. À court terme, l'échéance de livraison porte sur le jour suivant (marché *day-ahead*) ou sur le jour même (marché infrajournalier) et les échanges sont effectués en bourse. Les prix des produits sont souvent exprimés en EUR/MWh. La figure 1.5 illustre les périodes des différents marchés qui sont présentés dans cette sous-section.

1.2.2.1 Produits à terme

Les produits à terme ont une échéance de livraison allant de quelques jours jusqu'à 6 ans. Ils couvrent des périodes qui durent plusieurs jours, semaines, mois, semestres, voire saisons. On distingue deux types de contrats de produits à terme : les contrats *forward* et les contrats *futures*. Les contrats *forward* sont conclus de gré à gré avec des conditions qui peuvent être spécifiques aux deux parties, alors que les contrats *futures* sont échangés sur des plateformes centralisées et portent sur des produits standardisés (heure de pointe ou de base, volumes multiples de 1 MWh). Dans les deux cas, les prix sont négociés à la date de création des contrats.

Les produits à terme permettent aux acteurs de sécuriser leurs revenus ou leurs coûts. Ils couvrent la variation globale de la consommation à l'échelle des semaines et des saisons. Les deux parties s'engagent à respecter les contrats de vente ou d'achat. En particulier, cela donne aux producteurs des engagements de production à long terme qu'ils doivent respecter du mieux possible à l'échéance de livraison.

14. Les fournisseurs sont les intermédiaires entre le marché de l'électricité et les clients finaux (particuliers et entreprises).

15. Certains gros industriels s'approvisionnent eux-mêmes sur les marchés sans passer par des fournisseurs.

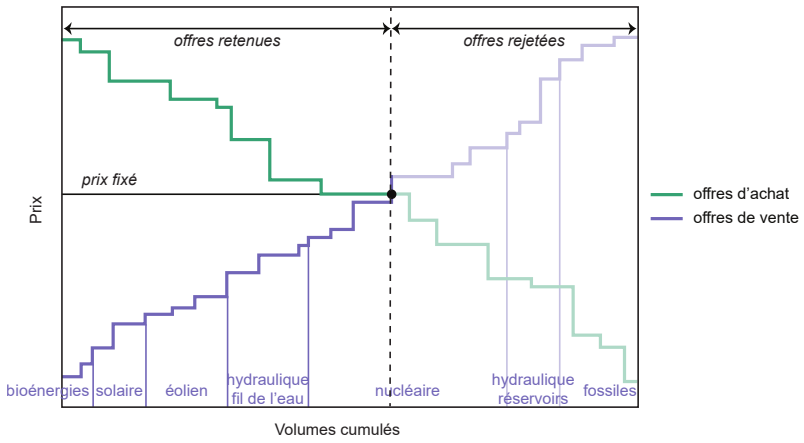


FIGURE 1.6 – Schéma du croisement des courbes des offres de vente et d'achat au marché *day-ahead* pour une heure donnée. Les volumes des offres sont cumulés dans l'ordre de leur prix, croissant pour les offres de vente et décroissant pour les offres d'achat.

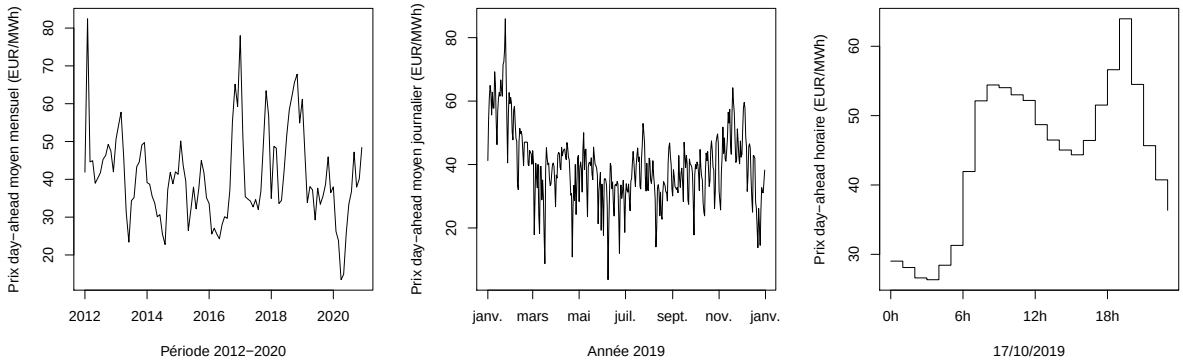
Le fait de recourir à des prix fixés longtemps en avance permet également de réduire les risques de financement pour le développement de nouveaux actifs de production, notamment lorsque les dispositifs d'aide de l'État pour le développement des énergies renouvelables (voir Sous-section 1.1.1) ne s'appliquent pas ou plus. Les contrats directs d'électricité (PPA pour *Power Purchase Agreement* en anglais) entrent dans ce cadre. Il s'agit de contrats de gré à gré à très long terme entre un producteur d'énergie renouvelable et un acheteur d'électricité (un consommateur par exemple) qui portent sur une longue période (généralement plus de 10 ans), avec des conditions généralement très spécifiques aux deux parties.

1.2.2.2 Produits *day-ahead*

Les produits échangés quotidiennement en bourse au marché *day-ahead* portent sur la production à livrer sur une ou plusieurs heures du lendemain. Une session du marché *day-ahead* est ouverte chaque jour. Avant l'heure de fermeture d'une session (à 12 h en France), les acteurs soumettent leurs offres de vente et d'achat pour chaque heure du lendemain (ou sur des blocs de plusieurs heures). Une offre de vente (respectivement d'achat) correspond à un volume d'énergie à injecter (respectivement à soutirer) sur le réseau durant une période donnée et à un prix minimum de vente (respectivement à un prix maximum d'achat).

Les prix *day-ahead* horaires du lendemain sont alors fixés juste après la fermeture du marché *day-ahead* par le point de croisement des courbes des offres de vente et des offres d'achat, comme l'illustre la figure 1.6. Seules sont retenues les offres de vente à un prix inférieur au prix *day-ahead* fixé et les offres d'achat à un prix supérieur, mais toutes les transactions s'effectuent au prix *day-ahead* fixé (principe du *pay-as-cleared*). Cela permet d'atteindre l'équilibre entre les offres de vente et d'achat d'énergie, reflétant le besoin d'équilibre entre la production et la consommation (durant le lendemain). Ainsi, le marché *day-ahead* couvre la variation globale de la consommation horaire à l'échelle d'une journée. Il s'agit d'enchères à l'aveugle : les offres de vente et d'achat sont définies avant de connaître les prix de vente ou d'achat. En général, les prix des offres de vente des producteurs sont très légèrement supérieurs aux coûts marginaux de leurs actifs de production [16]. Ainsi, la définition des prix *day-ahead* suit le mécanisme du *merit order* qui consiste à faire appel aux actifs de production dans l'ordre croissant de leur coût marginal, en lien avec la consommation. Les énergies variables sont sollicitées en premier, puis en période de

16. Dans certains cas, les prix des offres de vente peuvent être significativement plus élevés que les coûts marginaux. Cela s'explique notamment par la présence des quelques gros producteurs [Pikk et Viiding, 2013].



(a) Moyennes mensuelles (b) Moyennes journalières (c) Valeurs horaires

FIGURE 1.7 – Exemple de l'évolution des prix *day-ahead* en France (EPEX SPOT).

forte consommation, elles sont complétées par les énergies à coût marginal plus élevé. Le prix *day-ahead* correspond alors approximativement au coût marginal du système, i.e. celui du dernier actif de production appelé (voir Figure 1.6). Ainsi, les prix *day-ahead* faibles correspondent généralement à des périodes de faible consommation et les prix *day-ahead* forts correspondent généralement à des périodes de forte consommation (voir Figure 1.7). Les prix horaires en période de forte consommation (en hiver notamment) dépendent beaucoup de la disponibilité du parcs nucléaire (car la production française est principalement d'origine nucléaire, voir Sous-section 1.1.1), des coûts marginaux des centrales hydroélectriques de haute chute (liés à la quantité d'eau stockée dans leurs retenues) et ceux des centrales fossiles de pointe (liés aux prix des combustibles, comme le gaz par exemple). Le mécanisme du *merit order* rend les prix *day-ahead* assez volatils. En outre, les prix *day-ahead* dépendent en partie des conditions météorologiques car il en est de même pour les variations de la consommation et la production des énergies variables.

Une caractéristique importante du marché *day-ahead* est que sa liquidité est très forte, faisant des prix *day-ahead* les prix de référence du marché de l'énergie.

Les acteurs du marché *day-ahead* s'engagent à respecter leurs offres qui ont été retenues. En particulier, cela donne aux producteurs des engagements de production à court terme, qui s'ajoutent aux éventuels engagements à long terme déjà existants pour le lendemain et qu'ils doivent respecter du mieux possible à l'échéance de livraison.

1.2.2.3 Produits infrajournaliers

Les produits échangés en bourse au marché infrajournalier portent sur la production à livrer le jour même, jusqu'à quelques heures ou minutes avant la livraison. Les produits infrajournaliers sont définis sur les périodes d'une demi-heure ou d'une heure (ou par blocs de plusieurs périodes). Ils permettent d'effectuer les ajustements de dernières minutes, après les offres retenues au marché *day-ahead*, pour assurer l'équilibre presque en temps réel entre la consommation et la production. En France, le marché infrajournalier s'ouvre à 15 h la veille du jour de livraison.

Les transactions au marché infrajournalier sont effectuées en continu et elles sont exécutées dès qu'il y a, pour la même période, deux offres compatibles (il faut donc un vendeur et un acheteur). La liquidité du marché infrajournalier est limitée. Les prix sur ce marché sont définis selon le prix des offres et non le prix de la dernière offre retenue comme au marché *day-ahead*. En outre, il est difficile de compenser un déséquilibre soudain entre la consommation et la production quelques heures ou minutes avant livraison, donc les prix au marché infrajournalier sont assez volatils et incertains.

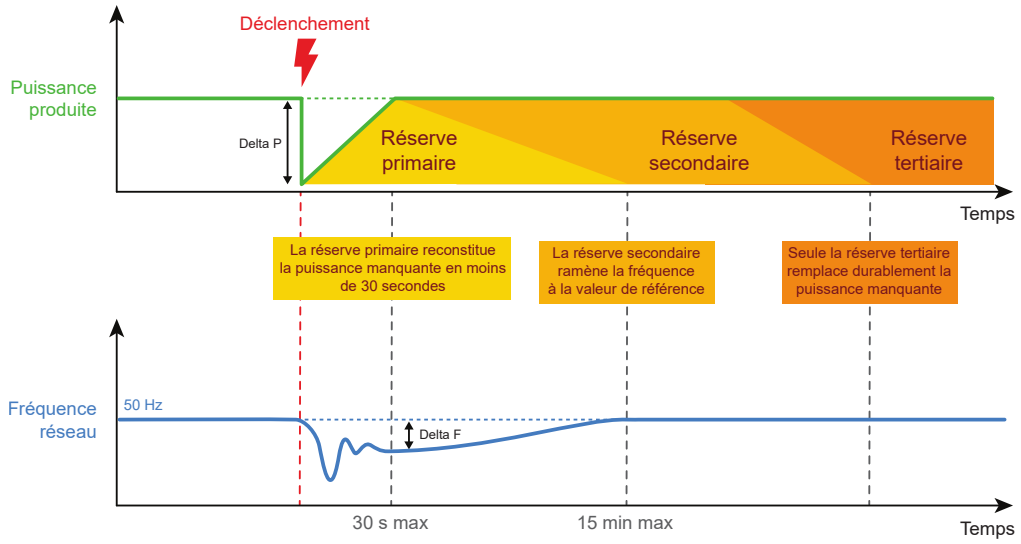


FIGURE 1.8 – Schéma du fonctionnement des services système et du mécanisme d'ajustement (largement inspiré de `https://www.cre.fr/Electricite/Reseaux-d-electricite/services-systeme-et-mecanisme-d-ajustement`).

1.2.3 Services système et mécanisme d'ajustement

Les services système et le mécanisme d'ajustement sont des réserves d'équilibrage constituées par RTE qui permettent d'assurer l'équilibre en temps réel entre la production et la consommation sur le réseau. Les réserves d'équilibrage sont des parts de production ou consommation disponibles, contractualisés par RTE auprès de producteurs et consommateurs, qui peuvent être mobilisées à tout moment en cas de déséquilibre. Il s'agit donc d'une des dernières actions possibles pour assurer l'équilibrage du réseau après le marché infrajournalier [17].

Il existe trois types de réserves selon leurs délais d'activation (voir Figure 1.8) : la réserve primaire, la réserve secondaire et la réserve tertiaire. Les réserves primaire et secondaire constituent les services système et la réserve tertiaire constitue le mécanisme d'ajustement.

La réserve primaire se déclenche en premier, de manière automatique et décentralisée, pour contenir en quelques secondes les déviations de fréquence. La participation à la réserve primaire est définie selon un processus d'appel d'offres hebdomadaire géré par RTE. La réserve primaire est constituée par des actifs de production (les groupes de production équipés de régulateurs de vitesse par exemple) et, à une moindre échelle, par des consommateurs flexibles.

La réserve secondaire se déclenche automatiquement par RTE, après la réserve primaire, pour ramener la fréquence du réseau à sa valeur de référence de 50 Hz. Le délai d'action de la réserve secondaire est de quelques minutes. Tous les actifs de production qui le peuvent et qui ont une puissance installée supérieure à 120 MW participent à la réserve secondaire (prescription).

La réserve tertiaire (mécanisme d'ajustement) se déclenche manuellement par RTE, après la réserve secondaire, pour régler les déséquilibres à plus long terme (dizaines de minutes à plusieurs heures) et reconstituer les réserves primaire et secondaire. La participation au mécanisme d'ajustement, gérée par RTE, est définie selon un processus d'appel d'offres annuel, complété par des appels d'offres journaliers (du jour pour le lendemain). Tous les actifs de production qui peuvent moduler leur production participent à la réserve tertiaire. Les consommateurs flexibles peuvent également faire des offres sur le marché d'ajustement (capacité d'effacement ou ajustement à la

17. En tout dernier recours, il y a les délestages de consommation ou de production.

hausse). La rémunération de la participation à la réserve tertiaire est complétée par une prime fixe selon la puissance mobilisée et le temps de réponse.

Les participants aux services système et au mécanisme d'ajustement encourent des pénalités par RTE s'il s'avère *a posteriori* qu'ils n'ont pas été capables de garantir leurs réserves.

1.2.4 Écarts entre production réelle et engagements

L'équilibrage entre la production et la consommation est donné par les engagements de vente et d'achat après toutes les transactions sur les marchés. Si un producteur ou un consommateur ne peut pas garantir exactement ses engagements, alors il peut créer un déséquilibre sur le réseau.

Des producteurs et des consommateurs ont un engagement contractuel auprès de RTE pour financer les coûts de remise à l'équilibre du réseau (voir Sous-section 1.2.3). Ils forment ce que l'on appelle des *responsables d'équilibre* (responsabilité financière). Tous les producteurs et consommateurs sont associés à un responsable d'équilibre. Le *périmètre d'équilibre* d'un responsable d'équilibre correspond alors à tous les sites d'injection et soutirage dont il est responsable.

Pour un responsable d'équilibre, on appelle *écarts* la différence, à l'échelle de son périmètre d'équilibre, entre la puissance totale algébrique[18] réellement injectée sur le réseau et celle des engagements. En cas d'écarts positifs, le responsable d'équilibre est rémunéré (éventuellement à prix négatif) par RTE pour le surplus d'énergie injectée. En cas d'écarts négatifs, le responsable d'équilibre est facturé par RTE pour la carence d'énergie injectée.

Les prix de règlement des écarts utilisés par RTE sont définis au pas demi-horaire à partir de la tendance du réseau (à la hausse ou à la baisse), du sens des écarts (positif ou négatif) et du prix moyen pondéré des offres d'ajustement activées par RTE [RTE, 2020]. Si le prix moyen pondéré π_{PMP} est positif ou nul, alors le prix de règlement des écarts positifs est défini par $\pi^+ = (1-\kappa)\pi_{\text{PMP}}$ et celui des écarts négatifs par $\pi^- = (1+\kappa)\pi_{\text{PMP}}$, avec $\kappa = 0{,}05$ en 2020. Si le prix moyen pondéré est négatif, alors les formules sont les mêmes mais en remplaçant κ par son opposé $-\kappa$. Les prix moyens pondérés sont *a priori* différents des prix horaires de l'électricité (voir §1.2.2.2). Le coefficient κ est défini chaque année par RTE pour que les prix des écarts soient, en moyenne sur l'année, moins intéressants que les prix *day-ahead*, afin d'inciter les responsables d'équilibre à réduire le plus possibles les écarts (i.e. être à l'équilibre du mieux possible). En revanche, les prix de règlement des écarts peuvent être plus attractifs que les prix *day-ahead* sur certaines heures, auquel cas il peut s'avérer plus intéressant financièrement pour un responsable d'équilibre de rester en écarts. En outre, de par leur définition, les prix de règlement des écarts sont peu prévisibles, même à très courte échéance. La faible liquidité du marché infrajournalier (peu d'offres disponibles) est une conséquence du fait que les prix de règlement des écarts sont peu prévisibles et qu'ils ne sont pas nécessairement pénalisant par rapport au prix *day-ahead*.

Pour anticiper les possibles écarts et les besoins d'ajustement subséquents, les responsables d'équilibre envoient à RTE leur prévision de production et/ou consommation du lendemain chaque jour à 16h30 (donc après les transactions au marché *day-ahead*) au pas demi-horaire. Les déclarations peuvent être corrigées le jour même si besoin (apparition d'une avarie par exemple).

Les productions variables sont sources d'écarts car elles ne peuvent pas être parfaitement prévues la veille. Les erreurs (erreurs absolues moyennes par rapport à la puissance installée) de prévision à court terme (horizon d'un ou plusieurs jours) sont de l'ordre de 15% pour un parc éolien et de l'ordre de 25% pour une centrale photovoltaïque [Burtin et Silva, 2015]. Néanmoins, les erreurs de prévision se compensent en partie par foisonnement à l'échelle de plusieurs actifs de production. Par exemple, à l'échelle de la France, les erreurs de prévision tombent à 3% pour

18. La puissance totale algébrique est la différence entre la production totale et la consommation totale.

la production éolienne et 5% pour la production photovoltaïque [Burtin et Silva, 2015]. Avec le développement des énergies variables, ces chiffres pourraient encore diminuer. Le foisonnement des énergies variables permet alors un lissage passif des erreurs de prévision, et donc des écarts.

1.2.5 Agrégation

L'agrégation est le mécanisme qui consiste à effectuer des transactions sur le marché de l'énergie pour le compte d'un tiers. L'agrégation peut porter sur les productions variables (hors obligation d'achat) ou sur les flexibilités de consommation. Les acteurs agrégés entrent dans le périmètre d'équilibre de l'agrégateur.

Un enjeu de l'agrégation des productions variables est de tirer profit du foisonnement des écarts des productions agrégées pour obtenir une meilleure valorisation économique (réduction des pénalités dues aux écarts). Cela devient un enjeu économique de plus en plus important en raison du développement des énergies variables et du nombre de producteurs. En outre, l'agrégateur dispose souvent de compétences en prévision et en transactions sur le marché de l'énergie qui permettent aussi d'obtenir une meilleure valorisation économique.

De manière analogue, l'agrégation sur les flexibilités de consommation (effacement [19], modulation ou stockage) permet de disposer d'une capacité suffisante de modulation de la consommation pour en tirer une valorisation économique sur le marché des réserves (services système et mécanisme d'ajustement, voir Sous-section 1.2.3).

1.2.6 Mécanisme de capacité

Le mécanisme de capacité est un dispositif mis en place en 2017 pour assurer la sécurité d'approvisionnement en électricité en France.

Les fournisseurs d'électricité ont l'obligation de couvrir la consommation agrégée (puissance et énergie) de leur clients en périodes de pointe (i.e. de forte consommation, principalement en hiver). Ils acquièrent pour cela des certificats de capacité (de production et/ou d'effacement) auprès des producteurs et/ou exploitants de capacité d'effacement. Les échanges de certificats se font de gré à gré ou sur un marché organisé (enchères réalisées plusieurs fois pas an). Les fournisseurs sont pénalisés ou indemnisés *a posteriori* par RTE si la différence entre les capacités des certificats achetés sont inférieures ou supérieures à la consommation agrégée effective. Les producteurs ont l'obligation de mettre à disposition et de faire certifier par RTE leurs moyens de production. Ils sont pénalisés *a posteriori* par RTE s'il s'avère qu'ils n'ont pas été capables de mettre à disposition les capacités des certificats vendus.

Les prix des certificats de capacité sont issus des négociations de gré à gré et des enchères. En revanche, les prix des pénalités ou indemnités sont proportionnels aux prix de règlement des écarts (voir Sous-section 1.2.4).

Le mécanisme de capacité permet d'inciter l'investissement dans les moyens de production utilisables en période de pointe (qui seraient peu ou pas rentables sans ce dispositif).

19. L'effacement de consommation consiste à modérer volontairement la consommation sur une courte période. Ce mécanisme s'appuie sur des applications connectées et sur des *smart grids* [Siano, 2014].

1.3 Gestion opérationnelle d'actifs de production

1.3.1 Planification de la production

Les producteurs peuvent exploiter plusieurs actifs de production et ils sont également associés à des responsables d'équilibre et/ou des agrégateurs. Les actifs de production sont donc rarement exploités individuellement mais de manière conjointe et coordonnée avec d'autres actifs. Cela permet de travailler sur la production totale cumulée plutôt que sur les productions de chaque actif prises séparément, afin de tirer profit du foisonnement de la production, qui peut être renforcé par des actifs de stockage [20]. Cette vision conjointe est par exemple indispensable pour l'exploitation d'une chaîne hydroélectrique (voir §1.1.3.4).

La *gestion opérationnelle* d'un ensemble d'actifs de production est par définition l'ensemble des opérations qui sont régulièrement répétées pour leur exploitation (maintenance, contrôle, conduite etc.). Elle est généralement effectuée en lien avec les marchés de manière à optimiser la valorisation économique conjointe des actifs de production, en minimisant les coûts de production et/ou en maximisant le chiffre d'affaires obtenu par la vente de la production totale. Par exemple, les opérations de maintenance sont réalisées dans la mesure du possible sur des périodes à bas prix.

Comme les transactions sur les marchés se font en avance, la production totale doit aussi être prévue en avance. Ainsi, la gestion opérationnelle comprend une phase de *planification* de la production de chaque actif à un horizon souhaité. Cette planification consiste en la construction d'un programme de production pour les actifs flexibles et/ou en l'utilisation de modèles de prévision pour les actifs peu ou pas flexibles. Par exemple, les productions variables peuvent être prévues en appliquant un modèle de simulation à partir de prévisions météorologiques, et un programme de production peut être construit pour les centrales hydroélectriques qui ont une flexibilité suffisante.

Dans la suite du manuscrit, on appelle *horizon caractéristique de planification* l'horizon pour lequel il est pertinent de planifier la production d'un actif de production. Si l'actif de production est peu ou pas flexible, alors l'horizon caractéristique de planification correspond à l'horizon maximal pour lequel les modèles de prévision de la production sont performants. Par exemple, l'horizon caractéristique de planification d'un parc éolien est de l'ordre de quelques jours (les prévisions du vent, et donc de la production éolienne, étant peu performantes au-delà de 4 ou 5 jours). Si l'actif de production est flexible, alors l'horizon caractéristique de planification correspond à l'horizon pour lequel il est pertinent de construire le programme de production. Un programme de production construit avec un horizon en-deçà de l'horizon caractéristique de planification n'offrirait pas une vision suffisamment étendue pour tirer profit de toute la modulation possible. Au contraire, une programmation au-delà de l'horizon caractéristique de planification serait superflue car globalement indépendante de la programmation jusqu'à l'horizon caractéristique de planification. L'horizon caractéristique de planification dépend du type des actifs de production. Par exemple, l'horizon caractéristique de planification des centrales hydroélectriques est celui qui correspond au temps de remplissage de leur retenue et de la modulation de leur production. Il est de plusieurs mois pour les centrales hydroélectriques de haute chute et de quelques jours pour les centrales hydroélectriques de moyenne et basse chute (voir Sous-section 1.1.3). La modulation de la production est donc décidée à long terme pour les centrales hydroélectriques de haute chute et à court terme pour celles de moyenne ou basse chute.

Bien que les productions des actifs de production se planifient à des horizons différents, elles peuvent être vendues à tous les horizons des marchés (voir Sous-section 1.2.2), même pour ceux

20. Plus généralement, on forme ainsi ce que l'on appelle une *centrale électrique virtuelle* [Pudjianto *et al.*, 2007 ; Saboori *et al.*, 2011] mais ce concept est volontairement ignoré dans ce manuscrit.

qui sont au-delà des horizons caractéristiques de planification. Pour une échéance de livraison donnée, les ventes à terme sont d'abord décidées plusieurs mois ou années en avance (par exemple à partir de statistiques de productions passées). Les ventes à terme sont progressivement complétées à mesure que l'échéance de livraison approche et que la planification de la production se précise. Les transactions au marché *day-ahead* (le marché de référence du fait de sa très forte liquidité) sont effectuées la veille du jour de livraison pour atteindre l'équilibre en fonction des engagements des ventes à terme pour l'échéance de livraison. Le marché infrajournalier peut être utilisé quelques heures ou minutes avant l'échéance de livraison, si la liquidité le permet, pour compenser une partie des écarts qui n'ont pas été prévus la veille. Les écarts effectivement réalisés en temps réel sont pénalisés par RTE. Ce cumul des engagements de production est illustré à la figure 1.9 pour une livraison à une heure donnée. La figure 1.10 illustre un exemple de planification de la production à court terme pour décider des prochaines ventes *day-ahead*. La sous-section 1.3.2 suivante présente plus en détail le lien entre les horizons caractéristiques de planification et ceux des marchés.

1.3.2 Vente de la production

Le producteur répartit sa production entre les différents marchés de la sous-section 1.2.2. Il se positionne également sur les marchés des services système, d'ajustements et de capacité (voir Sous-sections 1.2.3 et 1.2.6). Nous présentons dans cette sous-section les étapes générales qui permettent au producteur de se positionner sur les différents marchés. Ces étapes peuvent néanmoins varier d'un producteur à un autre.

Les ventes à terme sont effectuées progressivement à partir de la production potentielle prévue à long terme, en engageant de plus en plus de production à mesure que l'échéance de livraison approche et que la prévision de la production s'affine (voir Figure 1.9). Les ventes à terme peuvent commencer jusqu'à 6 ans avant la livraison et être complétées progressivement jusqu'à quelques jours ou semaines avant la livraison. Comme leurs prix sont généralement fixés longtemps en avance sur de longues périodes (voir §1.2.2.1), les ventes à terme permettent au producteur de se protéger du risque associé au prix (notamment une chute des prix *day-ahead*). En revanche, les ventes à terme engagent le producteur sur des volumes de production à produire sur des périodes parfois lointaines avec peu de visibilité. Le placement de la production à long terme résulte donc d'une gestion des risques : le producteur vend à terme une proportion de la production potentielle prévue à long terme afin de diminuer le risque associé au prix, sans pour autant s'exposer au risque associé au volume de production à livrer. Pour prévoir la production potentielle à long terme, deux cas se présentent selon que les dates de livraison sont au-delà ou en-deçà de l'horizon caractéristique de planification (voir Figure 1.11).

— Si les dates de livraison des ventes à terme sont au-delà de l'horizon caractéristique de planification, alors il n'est pas pertinent de prévoir la production potentielle en planifiant la production (voir Sous-section 1.3.1). Dans ce cas, la production potentielle prévue est obtenue à partir de statistiques de productions passées pour des périodes analogues. Les ventes à terme ne permettent alors pas de valoriser économiquement la production par rapport aux variations hebdomadaires ou mensuelles des prix mais elles permettent de sécuriser une partie des revenus. Si, en outre, la production dépend des conditions météorologiques (actif de production variable ou chaîne hydroélectrique au fil de l'eau par exemple), alors la production potentielle prévue à long terme peut aussi être obtenue à partir de tendances météorologiques (pour des livraisons dans les 15 jours), de tendances saisonnières (pour des livraisons dans les 3 à 4 mois) ou de la climatologie [21] (pour des livraisons au-delà de 4 mois). En revanche, plus l'horizon est lointain, moins ces prévisions sont performantes, d'où le risque associé au volume de production à livrer.

21. La climatologie correspond à la répartition statistique des observations passées.

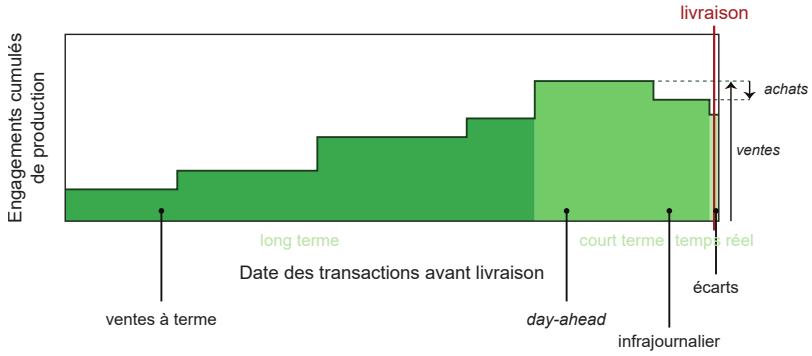


FIGURE 1.9 – Schéma du cumul des engagements de production pour une livraison à une heure donnée à mesure que l'échéance approche.

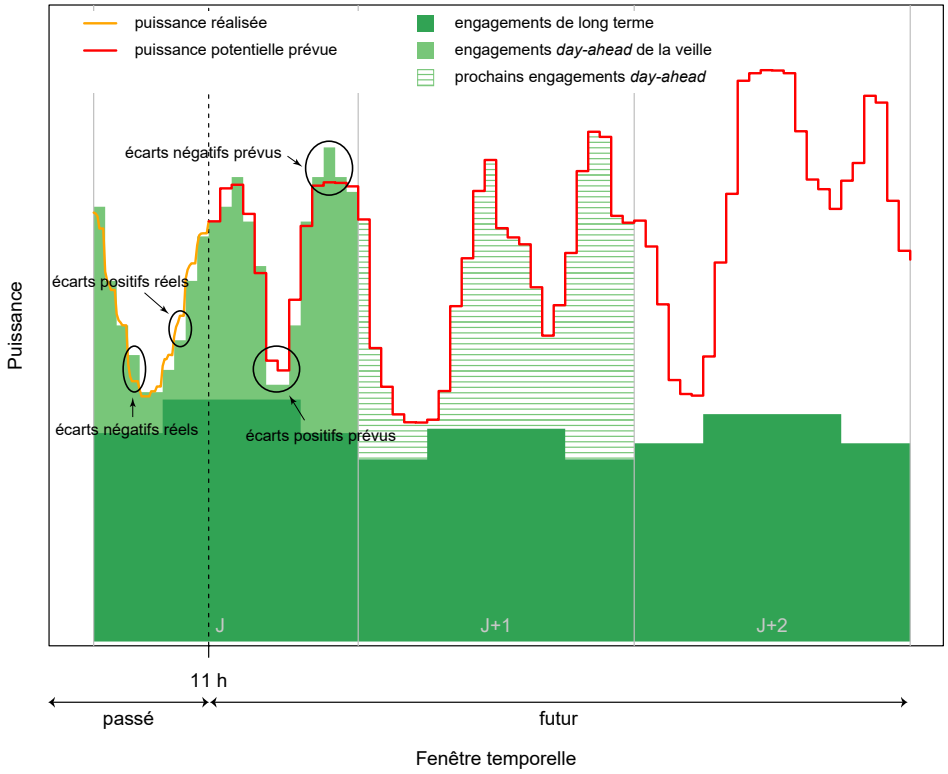


FIGURE 1.10 – Exemple de la planification de la production à court terme effectuée à 11 h juste avant les prochaines ventes *day-ahead* à un horizon correspond à la fin du sur-lendemain (+61 h).

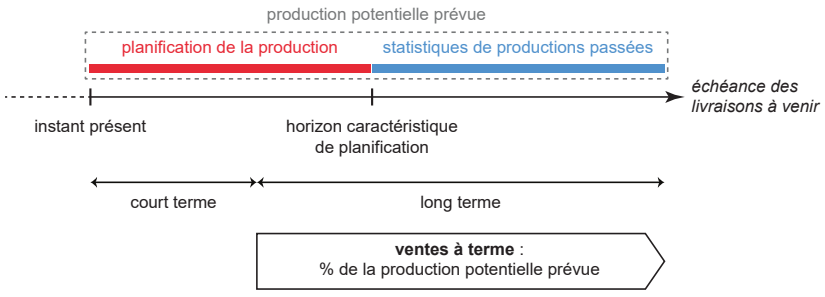


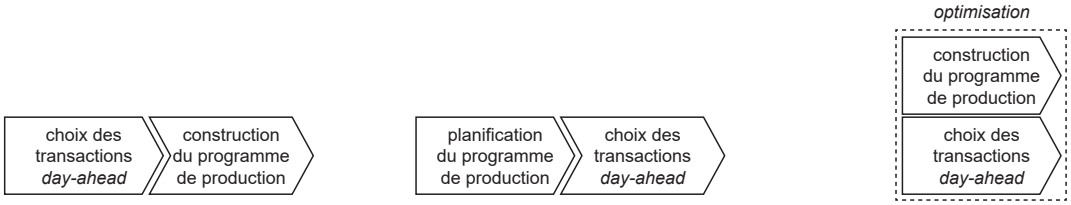
FIGURE 1.11 – Schéma de la prévision de la production potentielle à long terme.

— Si les dates de livraison des ventes à terme sont en-deçà de l'horizon caractéristique de planification, alors la production potentielle est prévue en planifiant la production à long terme, i.e. en utilisant des modèles de prévision de production pour les actifs de production peu ou pas flexibles et en construisant le programme de production des actifs de production flexibles. Dans le cas où un programme de production est construit à long terme, la production potentielle prévue tient compte des variations hebdomadaires ou mensuelles des prix pour placer, dans la mesure du possible, la production aux périodes où les prix prévus sont les plus élevés en fonction de la flexibilité de production à disposition. En plus de sécuriser une partie des revenus, les ventes à terme peuvent alors servir à valoriser économiquement la production. La construction du programme de production à long terme des actifs de production flexibles nécessite une prévision des prix à long terme (avec, au moins, les variations hebdomadaires ou mensuelles). Si la production dépend des conditions météorologiques (centrales hydroélectriques de haute chute par exemple), alors la planification de la production à long terme nécessite des tendances météorologiques, des tendances saisonnières ou de la climatologie selon la date de livraison.

À moyen terme (à l'échelle de quelques semaines), les problématiques sont similaires à celles à long terme sauf que la production potentielle prévue est plus précise et avec une résolution temporelle plus fine. La production potentielle prévue à long et moyen terme permet également de se positionner sur le marché des ajustements (voir Sous-section 1.2.3) pour les périodes en-deçà de l'horizon caractéristique de planification et sur le marché de capacité (voir Sous-section 1.2.6) quel que soit l'horizon caractéristique de planification [22].

La production potentielle prévue est affinée et régulièrement mise à jour à mesure que l'échéance de livraison se rapproche et que la connaissance de l'état futur du système se précise. Les offres de vente ou d'achat au marché *day-ahead* s'obtiennent à partir de la production potentielle prévue à court terme (quelques jours) pour compléter les engagements de long terme déjà existants pour le lendemain, dans le but d'atteindre l'équilibre (voir Figure 1.10). Le marché *day-ahead* est en effet le dernier guichet de vente où la liquidité est totale, le marché infrajournalier n'ayant pas une liquidité suffisante pour y échanger autre chose que des déséquilibres prévues seulement à très court terme (quelques heures ou minutes) après avoir effectué les transactions au marché *day-ahead*. En général, la production potentielle prévue à court terme est supérieure aux engagements de long terme, donc la différence entre les deux sont des offres de vente au marché *day-ahead*. Néanmoins, s'il n'est pas possible de produire au moins les engagements de long terme (avarie récente, conditions météorologiques défavorables etc.), alors la production manquante est une offre d'achat au marché *day-ahead* (normalement, ces cas sont gérés par la gestion des risques des ventes à terme). Pour les actifs de production flexibles, les prochaines offres au marché *day-ahead* et le programme de production à court terme peuvent être définis par trois approches différentes selon la flexibilité disponible à court terme (voir Figure 1.12).

22. Au-delà de l'horizon caractéristique de planification, des règles de calcul statistique existent pour déterminer les garanties de capacité. C'est ainsi que les parcs éoliens et les centrales photovoltaïques peuvent participer au marché de capacité.



(a) Flexibilité élevée (b) Flexibilité inexistante (c) Flexibilité faible

FIGURE 1.12 – Schéma de l'enchaînement de la décision des prochaines offres au marché *day-ahead* et de la planification du programme de production à court terme en fonction de la flexibilité de production à disposition.

1. Les offres de vente ou d'achat au marché *day-ahead* sont d'abord décidées dans la limite de ce qui est possible de produire à court terme compte tenu de la vision globale de l'optimisation à long terme, des puissances maximales et des contraintes d'exploitation, mais sans limitations sur les ressources disponibles pour la production à court terme (eau, vent, soleil par exemple) en fonction des prix attendus sur le marché. Le programme de production à court terme est ensuite construit ou mis à jour de manière à respecter les engagements pour la période concernée (long terme et *day-ahead*) ainsi que les contraintes d'exploitation (voir Figure 1.12a). La valorisation économique consiste alors à minimiser les écarts : le programme de production doit atteindre une puissance cible (celle des engagements) du mieux possible. Cette approche est surtout pertinente lorsque la flexibilité à court terme est suffisante. C'est le cas par exemple pour une ou plusieurs centrales hydroélectriques de haute ou moyenne chute en cascade, où la production totale horaire est définie dans le temps grâce aux ventes *day-ahead* puis répartie entre les centrales par la construction du programme de production à court terme (les valeurs de l'eau des retenues à la fin de l'horizon court terme sont données par la vision globale de l'optimisation à long terme).
2. Le programme de production à court terme est d'abord planifié en fonction des ressources disponibles pour la production (eau, vent, soleil par exemple) et des contraintes d'exploitation. Les offres de vente ou d'achat au marché *day-ahead* sont ensuite obtenues en considérant la différence entre la production potentielle prévue à court terme et les engagements de long terme (voir Figure 1.12b). Cette approche ne permet pas d'effectuer une valorisation économique sur le marché *day-ahead*. Elle est surtout pertinente lorsque la flexibilité à court terme est inexistante, i.e. lorsque le programme de production est entièrement imposé par les ressources disponibles et les contraintes d'exploitation. C'est le cas par exemple pour les actifs de production variable et les centrales hydroélectriques parfaitement au fil de l'eau.
3. Le programme de production à court terme et les offres de vente ou d'achat au marché *day-ahead* sont décidés conjointement (voir Figure 1.12c). La valorisation économique sur le marché *day-ahead* consiste ici en la maximisation des gains (différence entre les revenus obtenus par la vente de la production et les coûts de production), notamment en plaçant la production sur les heures où les prix prévus sont les plus élevés. Cette approche est surtout pertinente lorsque la flexibilité à court terme est située entre les deux cas extrêmes précédents, i.e. lorsque la flexibilité à court terme est faible mais suffisante pour être valorisée économiquement sur le marché *day-ahead*. Elle s'applique par exemple pour une chaîne hydroélectrique où la production moyenne est imposée par les apports en eau (principalement issus des affluents) mais dont le marnage des retenues est suffisant pour moduler légèrement la production autour de la production moyenne à l'échelle de quelques heures. Dans ce cas, la construction du programme de production à court terme permet de définir la production à la fois dans le temps et entre les centrales.

Pour la première et la dernière des approches précédentes, une prévision des prix horaires *day-ahead* à court terme est nécessaire. Si les actifs de production dépendent des conditions météorologiques, alors les trois approches précédentes nécessitent des prévisions météorologiques à court terme. Il convient de noter que la production potentielle prévue pour le lendemain sert également pour la déclaration quotidienne de production à RTE (voir Sous-section 1.2.4).

À très court terme (le jour même), la production potentielle prévue issue de la mise à jour du programme de production et/ou des prévisions météorologiques peut être légèrement différente des engagements à livrer (long terme et *day-ahead*). La différence entre les deux représente une prévision à très court terme des écarts (voir Figure 1.10). Une partie de ces écarts prévus peut éventuellement être rattrapée sur le marché infrajournalier, le reste risquant de se transformer au final en écart pénalisés financièrement par RTE (voir Sous-section 1.2.4). La proportion des écarts prévus à rattraper sur le marché infrajournalier dépend de la liquidité du marché et de la stratégie de valorisation [23]. Ensuite, les écarts effectivement réalisés en temps réel (qui peuvent être légèrement différents des écarts prévus à très court terme) sont effectivement pénalisés par RTE au prix de règlement des écarts.

1.3.3 Problème d'optimisation

Lorsque les actifs de production possèdent une flexibilité suffisante, la planification de la production consiste généralement en la résolution d'un problème d'optimisation pour valoriser économiquement la production, en la définissant dans le temps et entre les différents actifs, ainsi qu'entre les différents marchés. Le programme de production est construit, sur l'horizon caractéristique de planification, pour optimiser un critère de valorisation économique (minimisation des coûts de production, maximisation des revenus) tout en respectant les contraintes d'exploitation. Ce problème d'optimisation se base alors sur une prévision de la disponibilité des ressources pour la production (eau, vent, soleil par exemple) et sur une prévision des prix pour estimer la valeur économique. L'optimisation peut parfois être effectuée manuellement ou semi-manuellement par expertise et à l'aide d'outils de simulation. Dans ce cas, l'écriture mathématique du problème d'optimisation n'est pas nécessaire. L'optimisation peut également être effectuée numériquement à l'aide d'un optimiseur. Cela est principalement utile lorsque le fonctionnement conjoint des actifs de production est complexe et que l'on cherche à valoriser la production sur plusieurs marchés. Le problème d'optimisation est alors écrit mathématiquement (modélisation du fonctionnement des actifs de production et du critère de valorisation) puis résolu numériquement.

Pour la gestion opérationnelle à court terme, une des principales difficultés concerne le temps de calcul lié à la résolution numérique du problème d'optimisation, qui doit être compatible avec les délais opérationnels. Le problème d'optimisation à court terme est en effet généralement complexe car il est nécessaire de modéliser finement (résolution temporelle fine) le fonctionnement des actifs de production et les contraintes d'exploitation, le rendant *a priori* difficile à résoudre numériquement.

Ces problèmes mathématiques d'optimisation correspondent à la famille des problèmes d'attribution d'unités de production (*unit commitment problem*), pour laquelle la littérature est assez riche, voir par exemple [Bertsekas *et al.*, 1983 ; Abdi, 2020 ; Saravanan *et al.*, 2013 ; Huang *et al.*, 2017 ; Carrión et Arroyo, 2006]. Ils apparaissent également dans les systèmes de contrôle des centrales électriques virtuelles, voir [Pandžić *et al.*, 2013 ; Othman *et al.*, 2015 ; Nikonowicz et Milewski, 2012]. Une formulation mathématique du problème de la planification à court terme de la production à l'échelle d'EDF (Électricité De France) est présentée dans [Hechme-Doukopoulos *et al.*, 2010]. Pour le cas particulier d'une chaîne hydroélectrique qui fait

23. Même si les prix de règlement des écarts incitent les producteurs à rester, en moyenne, à l'équilibre du mieux possible, il peut s'avérer plus intéressant économiquement de rester en écarts sur certaines périodes, expliquant notamment la faible liquidité du marché infrajournalier (voir Sous-section 1.2.4).

l'objet de ce manuscrit, on peut citer les états de l'art généraux [Taktak et D'Ambrosio, 2017 ; Labadie, 2004]. Les systèmes hydroélectriques considérés sont souvent des multi-réservoirs (i.e. des centrales hydroélectriques en cascade) [Conejo *et al.*, 2002 ; Tufegdzic *et al.*, 1996 ; Diaz *et al.*, 2010 ; Piekutowski *et al.*, 1993]. Les principales difficultés du problème d'optimisation sont alors la non linéarité des courbes de puissance (voir Figure 1.3) et le caractère binaire des démarrages et arrêts des groupes de production. Des procédures de linéarisation ou de relaxation peuvent être utilisées mais il est crucial d'étudier les erreurs d'approximation qui en résultent. Certains auteurs tirent également profit du caractère dynamique (dans le temps et entre les centrales hydroélectriques) pour la résolution numérique du problème d'optimisation [Taktak et D'Ambrosio, 2017 ; Siu *et al.*, 2001]. À notre connaissance, peu d'articles traitent le cas concret d'une chaîne hydroélectrique au fil de l'eau le long d'un fleuve entier sujet à des contraintes d'exploitation fortes et pour une utilisation en opérationnel. Dans ce cas particulier, le problème d'optimisation est généralement difficile à résoudre en raison du faible degré de liberté dans le choix du programme de production (du fait des fortes contraintes d'exploitation et de la faible flexibilité de la production) et de la grande taille du problème d'optimisation (linéaire avec le nombre de centrales le long de la chaîne hydroélectrique). Pourtant, la résolution numérique doit être effectuée en temps très contraint pour une utilisation en opérationnel. Des exemples de chaînes hydroélectriques réelles sont néanmoins présentés dans, par exemple, [Chancelier et Renaud, 1994 ; Arce *et al.*, 2002 ; Li *et al.*, 2013].

1.4 Gestion sous incertitudes

En plus des difficultés dues à la taille et au nombre de contraintes du problème d'optimisation, la planification de la production peut être très sensible face aux incertitudes qui interviennent sur certains paramètres du problème d'optimisation. Nous présentons dans cette section les points clefs de la gestion opérationnelle sous incertitudes.

1.4.1 Impact des incertitudes sur la planification de la production

Certains paramètres utiles à la planification de la production sont estimés ou prévus en avance car ils dépendent d'événements futurs qui ne sont pas encore réalisés. Ces paramètres sont donc incertains, i.e. ils ne sont pas parfaitement connus à l'instant présent. Les sources d'incertitude peuvent être très variées, par exemple :

— les incertitudes qui entachent les prévisions des prix et des conditions météorologiques,

— les incertitudes liées à la modélisation du système (le cas échéant) qui simplifie le fonctionnement réel,

— les incertitudes induites par les incidents d'exploitation (avaries par exemple).

Les valeurs effectivement réalisées des paramètres incertains ne sont connues qu'en temps réel, une fois que les décisions concernant la planification de la production (décision des ventes par exemple) ont été prises. Ainsi, les incertitudes peuvent avoir plus ou moins d'impact sur le chiffre d'affaires. Par exemple, les incertitudes des prévisions météorologiques sont des sources d'écarts pour les productions variables qui engendrent des pénalités d'écarts (voir Sous-section 1.2.4) impactant le chiffre d'affaires (à la hausse ou à la baisse). De même, les incertitudes des prévisions hydrologiques impactent fortement les centrales hydroélectriques au fil de l'eau car leur production est fortement influencée par le débit entrant (voir Sous-section 1.1.3). Dans ce cas, le principal risque est de ne pas pouvoir respecter les contraintes d'exploitation pour les conditions hydrologiques qui se réalisent effectivement. Le programme de production doit donc être corrigé (mis à jour) régulièrement à très court terme en fonction des débits d'apports qui se réalisent effectivement en temps réel. Ces corrections sont des sources d'écarts qui engendrent des pénalités d'écarts impactant le chiffre d'affaires (à la hausse ou à la baisse). À l'inverse,

les incertitudes des prévisions hydrologiques impactent plus faiblement les centrales hydroélectriques de moyenne ou haute chute car leur production est faiblement influencée par le débit entrant (voir Sous-section 1.1.3).

1.4.2 Vers un problème d'optimisation en univers probabiliste

Pour faire face au risque d'écarts (les anticiper et les maîtriser) et adapter au mieux la production en fonction des incertitudes, il peut être pertinent de tenir compte des incertitudes dans la formulation mathématique du problème d'optimisation. Certains paramètres du problème d'optimisation sont alors considérés comme des variables aléatoires, i.e. des paramètres dont les valeurs qui seront effectivement observées sont issues d'une loi de probabilité. On dit que le problème d'optimisation est *sous incertitudes*, ou en *univers probabiliste* ou *stochastique* par opposition à un problème d'optimisation en univers déterministe lorsque les incertitudes ne sont pas considérées [Haneveld *et al.*, 2019].

Le passage de l'univers déterministe à l'univers probabiliste complexifie le problème d'optimisation, privilégiant la résolution à l'aide d'un optimiseur (résolution numérique) au détriment des méthodes (semi-)manuelles. Il consiste en les quatre étapes suivantes [King et Wallace, 2012] :

1. l'identification des paramètres incertains les plus importants,
2. la quantification et la modélisation des incertitudes,
3. l'adaptation de la formulation mathématique du problème d'optimisation,
4. la résolution numérique du problème d'optimisation sous incertitudes.

Les quatre étapes ci-dessus sont fortement dépendantes et elles sont idéalement étudiées en parallèle. Néanmoins, les deux premières étapes reposent souvent, en pratique, sur d'autres critères plus prioritaires que le problème d'optimisation (par exemple, les données disponibles, l'utilisation en opérationnel etc.). Elles sont alors traitées en amont et indépendamment des deux dernières étapes. Les méthodes possibles pour l'adaptation de la formulation mathématique et la résolution numérique dépendent donc des choix faits pour la modélisation de l'incertitude des paramètres considérés comme incertains.

1.4.3 Corrélations des paramètres incertains

Les paramètres incertains sont par définition les paramètres *scalaires* qui interviennent dans la planification de la production et qui sont sujets aux incertitudes (voir Sous-section 1.4.1). Ils sont donc modélisés par des variables aléatoires réelles, continues ou discrètes.

Néanmoins, les paramètres incertains ne sont généralement pas indépendants. Ils sont souvent associés à des grandeurs ayant une cohérence temporelle (voire spatio-temporelle) sur tout ou partie de la fenêtre temporelle de la planification (période entre l'instant présent et l'horizon caractéristique de planification). Autrement dit, les paramètres incertains sont généralement regroupés en différents processus stochastiques temporels (ou spatio-temporels). Par exemple, les prix (sujets aux incertitudes sur les conditions du marché) sont des paramètres incertains qui forment un processus stochastique temporel car il y a de fortes corrélations entre les prix d'instants voisins. De même, pour la planification d'une chaîne hydroélectrique, les débits d'apports en eau des affluents (qui interviennent dans les débits entrants des aménagements et qui sont sujets aux incertitudes sur les conditions hydrologiques) sont des paramètres incertains qui forment un processus stochastique spatio-temporel car les débits sont des grandeurs physiques continues dans le temps et il y a des corrélations entre les débits d'affluents de bassins versants voisins.

Au-delà des cohérences spatio-temporelles, des corrélations peuvent également exister entre les différentes grandeurs représentées par les paramètres incertains (corrélations entre les prix et les conditions météorologiques par exemple, voir Annexe A.1).

1.4.4 Identification des paramètres incertains d'intérêt

Comme nous venons de le voir, les paramètres qui interviennent dans le problème d'optimisation sont souvent fortement corrélés, rendant difficile la quantification des incertitudes de tous les paramètres sujets aux incertitudes. En outre, la complexité du problème d'optimisation en univers probabiliste, et donc le temps de calcul nécessaire pour sa résolution numérique, dépend généralement du nombre de paramètres pour lesquels on souhaite prendre en compte leurs incertitudes.

Une manière de simplifier la tâche est de ne considérer comme incertains qu'une partie des paramètres du problème d'optimisation, en identifiant ceux dont les incertitudes ont le plus d'impact sur la solution du problème d'optimisation. Cette étape peut être effectuée par une analyse de sensibilité à partir du problème d'optimisation en univers déterministe. On analyse alors le comportement de la solution optimale et de la valeur optimale face à des variations des paramètres. Cette analyse peut être effectuée par expertise, à partir de méthodes statistiques ou en utilisant les résultats théoriques des problèmes d'optimisation paramétriques [Gal et Greenberg, 2012 ; Pistikopoulos, 2007 ; Dupačová, 1990 ; Vardanyan et Amelin, 2012]. Les paramètres incertains les plus importants sont ceux ayant une influence suffisamment importante (critères objectifs ou subjectifs) pour que la prise en compte de leurs incertitudes dans l'optimisation soit pertinente.

Une méthode statistique connue d'analyse de sensibilité repose sur le calcul des indices de Sobol [Sobol, 2001 ; Saltelli *et al.*, 2008, 2010]. Étant données des variables aléatoires indépendantes X_1, ..., X_r sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ (facteurs explicatifs), et une variable aléatoire réelle Y sur $(\Omega, \mathcal{F}, \mathbb{P})$ qui dépend de X_1, ..., X_r (covariable) et admettant un moment d'ordre 2 avec une variance non nulle, les indices de Sobol du premier ordre pour chacun des facteurs explicatifs sont définis par :

$$\forall i \in \{1, \ldots, r\}, \quad \mathcal{S}_i = \frac{\mathrm{Var}[\mathbb{E}[Y|X_i]]}{\mathrm{Var}[Y]} \in [0, 1], \tag{1.2}$$

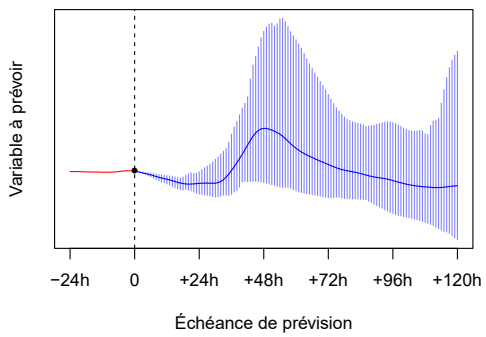
où $\mathbb{E}$ et Var désignent respectivement l'espérance et la variance pour des variables aléatoires réelles sur $(\Omega, \mathcal{F}, \mathbb{P})$. On a également la relation : $\mathcal{S}_1 + \cdots + \mathcal{S}_r \leq 1$. La valeur $1 - (\mathcal{S}_1 + \cdots + \mathcal{S}_r) \in [0, 1]$ correspond à l'influence sur la covariable des interactions entre les facteurs explicatifs. Plus un indice de Sobol $\mathcal{S}_i$ (où $i \in \{1, \ldots, r\}$) est proche de 1, plus le facteur explicatif X_i associé est influent sur la covariable Y (i.e. plus les variations de X_i se propagent sur les variations de Y).

1.4.5 Quantification et modélisation des incertitudes

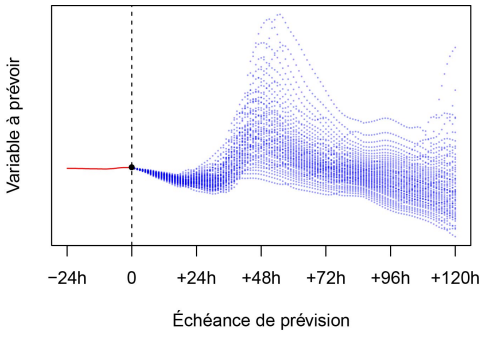
1.4.5.1 Méthodes de prévision probabiliste

La modélisation des incertitudes consiste à exprimer les lois de probabilité des paramètres incertains, selon une représentation choisie. On considère que les valeurs effectivement observées (par exemple, grâce à des mesures en temps réel) de ces paramètres sont des réalisations de lois de probabilité, appelée lois de probabilité *effectives* dans la suite du manuscrit. Il convient de noter qu'il s'agit d'une modélisation : rien ne garantit en effet l'existence de telles lois de probabilité en pratique.

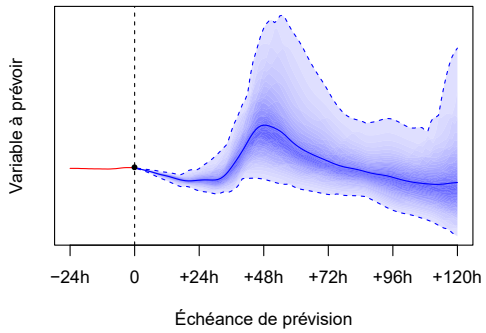
Les paramètres incertains pouvant être dépendants (processus stochastiques spatio-temporels pouvant être corrélés), on stipule l'existence d'une loi jointe, dite *effective* dans la suite du manuscrit, qui représente les informations les plus complètes que l'on peut avoir sur les incertitudes des paramètres. Les lois marginales de cette loi jointe effective sont les lois de probabilité univariées effectives correspondant à chacun des paramètres incertains. Les paramètres incertains sont alors représentés par une unique variable aléatoire multivariée qui suit la loi jointe effective.



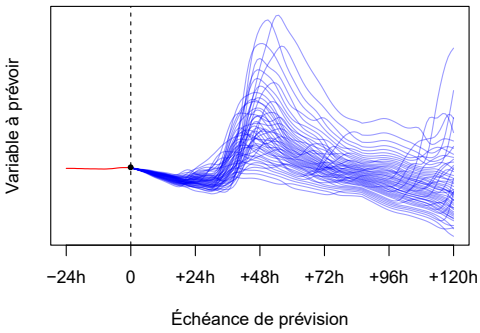
(a) Intervalles de confiance



(b) Lois discrètes univariées



(c) Lois continues univariées



(d) Scénarios multivariés (prévision d'ensemble)

FIGURE 1.13 – Exemples de modélisations des incertitudes pour des paramètres incertains regroupés en un processus temporel.

En pratique, la connaissance complète de la loi jointe effective (ni même des lois marginales effectives) est rarement accessible, mais il est possible de l'approcher par des méthodes de prévision probabiliste [Gneiting *et al.*, 2007]. Il en existe deux grandes catégories, chacune donnant lieu à des modélisations différentes des incertitudes :

1. les approches par habillement d'une prévision déterministe,
2. les approches ensemblistes.

Les approches par habillement consistent à entourer une prévision déterministe donnée (i.e. une prévision ayant une valeur prévue pour chaque paramètre incertain) à l'aide d'une information sur les dispersions univariées des paramètres incertains. Les incertitudes sont alors modélisées par des intervalles de confiance (voir Figure 1.13a), des valeurs de quantiles (voir Figure 1.13b) ou des lois continues univariées (voir Figure 1.13c), de manière à estimer les lois marginales effectives. Les dispersions univariées sont quantifiées, par exemple, à l'aide d'un modèle d'erreurs calibré sur un historique. Les approches par habillement ont l'avantage d'être assez faciles à mettre en place (il suffit, par exemple, d'avoir un historique de données passées) mais elles ne permettent pas de prendre en compte les corrélations entre les paramètres incertains (on perd les cohérences spatio-temporelles et les corrélations entre les différentes grandeurs). En outre, ces approches ne permettent généralement pas de modéliser les incertitudes propres à la situation considérée sachant les informations connues à l'instant présent, mais plutôt les erreurs de prévision estimées à grande échelle sur un historique (ayant souvent une plus grande dispersion). Les approches par habillement sont souvent utilisées pour modéliser les incertitudes de manière globale sans dissocier les différentes sources d'incertitude dont elles héritent. C'est par exemple le cas pour les paramètres dont les valeurs sont obtenues à l'aide de modèles numériques qui prennent eux-mêmes en entrée plusieurs autres paramètres dont les incertitudes sont délicates à quantifier.

Les approches ensemblistes consistent à considérer un ensemble fini de scénarios multivariés munis de leurs probabilités (voir Figure 1.13d). Les incertitudes sont alors modélisées par les scénarios, i.e. des réalisations possibles des paramètres incertains qui tiennent compte des corrélations (au moins celles liées aux cohérences spatio-temporelles), permettant d'estimer la loi jointe effective. Les approches ensemblistes ont l'avantage de tenir compte des corrélations entre les paramètres incertains (notamment pour représenter les cohérences spatio-temporelles) et d'être faciles à interpréter et manipuler. En revanche, leur mise en place est généralement plus difficile que pour les approches par habillement. Une autre limite des approches ensemblistes est que le nombre de scénarios considérés est fini : certains événements futurs ayant physiquement une probabilité d'occurrence non nulle, peuvent ne pas figurer dans les scénarios établis. Les scénarios ne modélisent donc pas les incertitudes pour des fréquences d'occurrence rare.

Une des approches ensemblistes est la *prévision météorologique d'ensemble* qui a été originellement mise en place pour tenir compte de la sensibilité des modèles de prévision météorologique à l'état initial de l'atmosphère [Bauer *et al.*, 2015 ; Gneiting et Raftery, 2005 ; Leutbecher et Palmer, 2008]. Elle consiste à lancer un ou plusieurs modèles météorologiques avec un état initial de l'atmosphère puis avec différentes perturbations autour de cet état initial. L'état initial est une estimation de l'état initial réel de l'atmosphère obtenu par un modèle météorologique à partir de l'ensemble des mesures atmosphériques disponibles. Cet état initial s'appelle *l'analyse* en météorologie. Les perturbations sont définies de manière à être compatibles avec l'ensemble des observations ayant servi à formuler l'analyse : elles forment d'autres estimations de l'état réel de l'atmosphère tout aussi probables que l'analyse. Dans certains systèmes de prévisions d'ensemble, les perturbations peuvent également porter sur les paramètres du modèle météorologique qui sert à établir l'analyse à partir des observations (à la place ou en complément des perturbations qui portent directement sur l'état initial). En appliquant les modèles météorologiques à chacun de ces estimations de l'état initial réel, on obtient un ensemble de prévisions, appelées *membres*, souvent considérés comme équiprobables. Le membre *contrôle* est par définition celui associé à l'état initial sans perturbations. Les membres *perturbés* sont par définition ceux associés aux états perturbés. Les prévisions météorologiques d'ensemble sont fournies par les centres météorologiques (Météo-France par exemple). Les membres d'une prévision météorologique d'ensemble peuvent ensuite être utilisés successivement en entrée d'un modèle numérique pour former une prévision d'ensemble d'une grandeur qui dépend des conditions météorologiques (on parle alors de *forçage*). On peut alors quantifier les incertitudes des paramètres qui dépendent des conditions météorologiques. Par exemple, une prévision d'ensemble des productions variables [Taylor *et al.*, 2009 ; Nielsen *et al.*, 2004 ; Sobri *et al.*, 2018] ou une prévision d'ensemble hydrologique [Bellier *et al.*, 2016] peut être obtenue à partir d'une prévision météorologique d'ensemble. Les prévisions d'ensemble hydrologiques peuvent ensuite servir à modéliser les incertitudes sur les apports en eau pour des actifs hydroélectriques [Faber et Stedinger, 2001 ; Raso *et al.*, 2014].

Une autre approche ensembliste est la *prévision d'ensemble par analogie*. Elle consiste à trouver, dans une archive de situations passées, un nombre fini de situations proches (analogues) à la situation courante et à utiliser les valeurs observées des paramètres de ces situations historiques comme scénarios. Elle nécessite donc une archive importante. Le principe d'analogie est applicable lorsque des situations proches produisent des effets similaires, comme c'est généralement le cas en météorologie [Lorenz, 1969]. En général, on considère que les scénarios obtenus par analogie sont équiprobables mais il est possible de définir des probabilités différentes selon la similitude des situations analogues. Cette approche par analogie repose sur le choix de la ou des variables d'analogie (prédicteurs de similitude) et du critère d'analogie (distance pour mesurer les similitudes entre les situations). Elle est également beaucoup utilisée pour les prévisions qui dépendent des conditions météorologiques car on considère que la dynamique de l'atmosphère est immuable (elle est similaire entre deux situations similaires) [Lorenz, 1969 ; Delle Monache *et al.*, 2013 ; Bontron, 2004].

Il est possible de passer des approches ensemblistes aux approches par habillement en considérant les lois de probabilité empiriques de chaque paramètre incertain, supprimant alors les corrélations. Il est aussi possible de passer des approches par habillement à une approche ensembliste en quantifiant, le cas échéant, la loi d'habillage de manière équiprobable puis en réarrangeant les valeurs quantifiées entre les lois univariées à condition de disposer d'un modèle de dépendance entre les paramètres [Clark *et al.*, 2004 ; Schefzik *et al.*, 2013 ; Bellier *et al.*, 2017, 2018]. Cette démarche transforme les lois de probabilité univariées en une loi de probabilité multivariée par reconstruction des corrélations entre les paramètres incertains. Cela permet donc de créer des scénarios multivariés, généralement considérés comme équiprobables.

1.4.5.2 Qualité des prévisions probabilistes

La qualité intrinsèque d'une prévision probabiliste correspond à sa capacité propre à représenter correctement la loi de probabilité effective sous-jacente. Il s'agit donc d'un critère important pour déterminer la confiance que l'on peut avoir de la modélisation des incertitudes [Jolliffe et Stephenson, 2012]. La qualité intrinsèque d'une prévision probabiliste se vérifie principalement à partir de deux attributs : la fiabilité et la finesse [Gneiting *et al.*, 2007 ; Bellier, 2018].

— La fiabilité est une mesure de similitude entre la loi de probabilité prévue par la prévision probabiliste et la loi de probabilité jointe effective. Les écarts en termes d'espérance sont des biais et les écarts en termes d'écart-type (ou variance) sont des défauts de dispersion, voir Figure 1.14. En pratique, la loi de probabilité effective de l'observation n'est pas accessible mais on dispose de sa réalisation. La fiabilité s'évalue donc en moyenne sur un grand nombre de situations différentes.

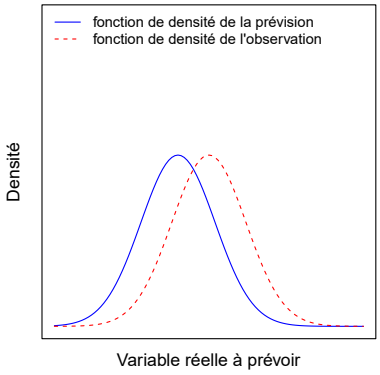
— La finesse est une mesure de la dispersion (i.e. l'écart-type ou variance) des prévisions probabilistes. Plus la finesse est élevée, plus la dispersion est faible (donc plus les incertitudes sont faibles), voir Figure 1.15.

Une prévision probabiliste est performante si sa fiabilité et sa finesse sont acceptables pour une utilisation en pratique. La performance de la modélisation des incertitudes est cruciale pour une utilisation dans un problème d'optimisation en univers probabiliste [King et Wallace, 2012 ; Ahmed *et al.*, 2002]. Si la modélisation des incertitudes n'est pas suffisamment fiable, alors les erreurs d'estimation de la loi de probabilité jointe effective peuvent se répercuter sur la qualité de la solution optimale du problème d'optimisation en univers probabiliste. Cela est renforcé par le fait que les paramètres considérés comme incertains sont généralement ceux dont les incertitudes impactent le plus le problème d'optimisation (voir Sous-section 1.4.4). De même, si la modélisation des incertitudes n'est pas suffisamment fine, alors cela risque de réduire inutilement l'optimalité ou l'admissibilité de la solution optimale (la solution optimale sera, par exemple, trop robuste par rapport aux événements futurs qui peuvent effectivement se réaliser).

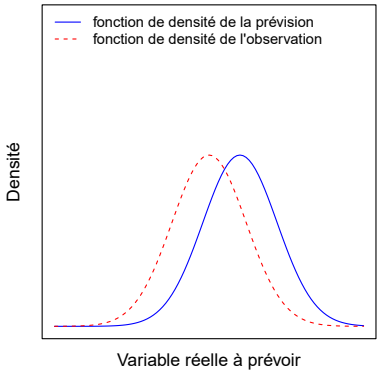
La fiabilité et la finesse peuvent être évaluées simultanément en calculant le CRPS (*Continuous Ranked Probability Score*) moyen sur un historique de vérification [Matheson et Winkler, 1976 ; Hersbach, 2000 ; Gneiting et Raftery, 2007]. Le CRPS individuel associé à une prévision probabiliste d'un paramètre incertain (réel), donnée par sa fonction de répartition $F_n \in \mathcal{F}(\mathbb{R}, \mathbb{R})$, et la réalisation $y_n \in \mathbb{R}$, connue (mesurée) en temps réel, est défini par (voir l'illustration de la figure 1.16) :

$$\mathrm{crps}(F_n, y_n) = \int_{-\infty}^{+\infty} \left(F_n(x) - H(x - y_n)\right)^2 \mathrm{d}x \in [0, +\infty], \tag{1.3}$$

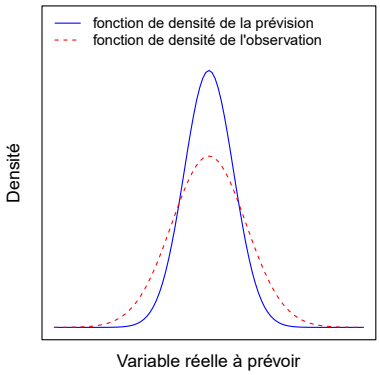
où $H : \mathbb{R} \to \{0, 1\}$ est la fonction d'Heaviside définie par : pour tout $u \in \mathbb{R}$, $H(u) = 1$ si $u \geq 0$ et $H(u) = 0$ si $u < 0$.



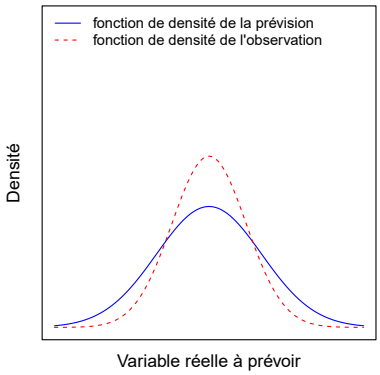
(a) Biais négatif



(b) Biais positif



(c) Sous-dispersion



(d) Sur-dispersion

FIGURE 1.14 – Exemple de défauts de fiabilité individuelle dans le cas idéal où la loi de probabilité effective de l'observation est connue (figure largement inspirée de [Bellier, 2018]).

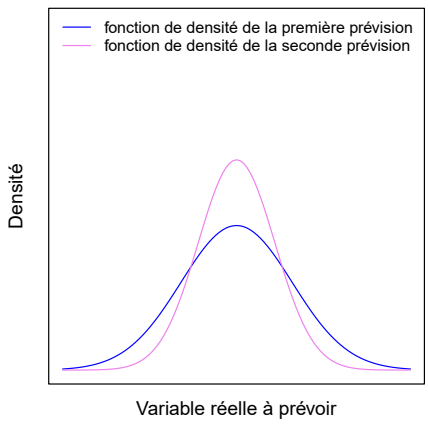


FIGURE 1.15 – Exemple des fonctions de densité de deux prévisions probabilistes d'une variable réelle ayant des finesses différentes. La seconde prévision probabiliste est plus fine que la première.

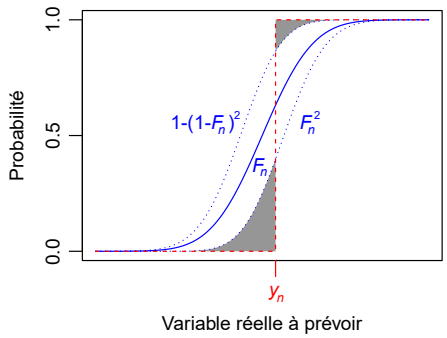


FIGURE 1.16 – Illustration du CRPS individuel qui correspond à l'aire en gris (figure largement inspirée de [Bellier, 2018]).

Dans le cas des valeurs univariées $(x_{n1}, \ldots, x_{nM}) \in \mathbb{R}^M$ d'une prévision d'ensemble à M membres ($M \in \mathbb{N}^*$) à un point donné (spatial ou temporel), la fonction de répartition F_n s'écrit

$$\forall x \in \mathbb{R}, \quad F_n(x) = \sum_{m=1}^{M} p_m H(x - x_{nm}),$$

où $p_m \in [0, 1]$, pour $m \in \{1, \ldots, M\}$, est la probabilité du membre d'indice m, $p_1 + \cdots + p_M = 1$. Dans ce cas, la formule de l'équation (1.3) s'écrit [Gneiting *et al.*, 2007] :

$$\mathrm{crps}(F_n, y_n) = \sum_{m=1}^{M} p_m |x_{nm} - y_n| - \frac{1}{2} \sum_{1 \leq m, m' \leq M} p_m p_{m'} |x_{nm} - x_{nm'}|,$$

ce qui facilite les calculs numériques. En particulier, pour une prévision déterministe, on a $M = 1$, donc le CRPS individuel correspond à l'erreur absolue entre la prévision et l'observation.

Le CRPS moyen est alors calculé sur un ensemble de couples (F_n, y_n), pour $n \in \{1, \ldots, N\}$ ($N \in \mathbb{N}^*$), pour une même situation (dans le cas où la prévision probabiliste est multivariée) ou sur un historique de situations différentes :

$$\mathrm{CRPS} = \frac{1}{N} \sum_{n=1}^{N} \mathrm{crps}(F_n, y_n) \in \mathbb{R}_+.$$

Comme le CRPS individuel d'une prévision déterministe correspond à l'erreur absolue entre la prévision et l'observation, le CRPS moyen est l'extension, en univers probabiliste, de l'erreur absolue moyenne (MAE, pour *Mean Absolute Error* en anglais) utilisée en univers déterministe.

Les fonctions $x \mapsto H(x - y_n) \in \mathcal{F}(\mathbb{R}, \mathbb{R})$, pour $n \in \{1, \ldots, N\}$, peuvent s'interpréter comme les fonctions de répartition des réalisations $y_n \in \mathbb{R}$ (les erreurs de mesure étant considérées comme nulles pour que les réalisations soient considérées comme des valeurs déterministes). Les valeurs du CRPS moyen s'interprètent grâce aux deux propriétés suivantes :

— un modèle de prévision probabiliste parfaitement juste a un CRPS moyen nul (dans ce cas, les prévisions probabilistes sont en fait déterministes (dispersion nulle) et confondues avec les réalisations),

— le CRPS moyen est un score orienté négativement : plus le CRPS moyen est proche de zéro, plus la qualité intrinsèque des prévisions issues du modèle de prévision probabiliste est bonne.

En pratique, la qualité intrinsèque des prévisions d'ensemble brutes (obtenues uniquement par forçage à partir de prévisions météorologiques d'ensemble) n'est en général pas satisfaisante pour une utilisation en opérationnel à cause de défauts de fiabilité [Bellier *et al.*, 2016]. Des corrections statistiques sont alors appliquées avant et après les forçages des modèles de calcul des paramètres. Les étapes de ces corrections sont généralement les suivantes [Bellier, 2018] :

1. l'estimation des lois de probabilité univariées et conditionnelles des paramètres incertains sachant la prévision d'ensemble brute multivariée, par exemple par une approche EMOS (*Ensemble Model Outputs Statistics*, [Gneiting *et al.*, 2005]) ou BMA (*Bayesian Model Averaging*, [Raftery *et al.*, 2005]),
2. l'échantillonnage des lois de probabilité univariées, par exemple en calculant des quantiles équiprobables,
3. le réarrangement des points échantillonnés pour reconstruire la structure multivariée, par exemple par *Schaake shuffle* [Clark *et al.*, 2004] ou ECC (*Ensemble Copula Coupling*, [Schefzik *et al.*, 2013 ; Bellier *et al.*, 2018]),
4. le lissage, si besoin, des membres de la prévision d'ensemble.

Dans notre contexte, les prévisions probabilistes sont utilisées pour la planification de la production. Elles servent donc à prendre des décisions qui cherchent une valorisation économique. En outre, elles modélisent les incertitudes des paramètres incertains qui ont le plus d'impact sur le problème d'optimisation, donc les légères erreurs d'approximation de la loi jointe effective peuvent se répercuter sur la solution et la valeur optimale du problème d'optimisation, même si la qualité intrinsèque est satisfaisante [King et Wallace, 2012 ; Ahmed *et al.*, 2002]. La qualité intrinsèque n'est alors pas un critère suffisant pour déterminer si la modélisation des incertitudes est pertinente pour l'optimisation. Il convient de considérer la valeur économique des prévisions probabilistes, i.e. leur capacité à fournir une solution optimale économiquement compte tenu des observations en temps réel des paramètres incertains [Cassagnole *et al.*, 2021]. Pour plus de détails sur la dualité entre la qualité et l'utilité d'une prévision probabiliste, voir [Bontron, 2004]. Ce critère permet d'évaluer la qualité de la modélisation des incertitudes, non pas de manière intrinsèque, mais avec l'ensemble des autres étapes du passage de l'univers déterministe à l'univers probabiliste (voir Sous-section 1.4.2).

1.4.6 Adaptation et résolution du problème d'optimisation en univers probabiliste

L'adaptation de la formulation mathématique du problème d'optimisation en univers probabiliste ne demande pas nécessairement beaucoup de changements. Il est possible, par exemple, d'utiliser la formulation déterministe avec un scénario « pire cas » ou avec des marges de sécurité définies de manière heuristique sur les contraintes les plus sensibles aux incertitudes, à partir de la quantification des incertitudes. Néanmoins, il est souvent difficile de maîtriser l'impact de ces modifications sur l'optimalité (voire sur l'admissibilité) des solutions. Il est donc généralement préférable d'utiliser les formulations mathématiques plus rigoureuses et les algorithmes de résolution spécifiques au cadre probabiliste. La littérature à ce sujet est très riche, voir par exemple les références générales [Haneveld et van der Vlerk, 2020 ; Haneveld *et al.*, 2019 ; Shapiro *et al.*, 2021 ; Ruszczyński et Shapiro, 2003 ; King et Wallace, 2012]. Le cadre probabiliste est également très présent dans le domaine de l'énergie, notamment pour la planification d'actifs de production électrique, voir par exemple [Conejo *et al.*, 2010 ; Wallace et Fleten, 2003 ; Huang *et al.*, 2017].

Pour le cas particulier d'une chaîne hydroélectrique qui fait l'objet de ce manuscrit, on peut citer les états de l'art généraux [van Ackooij *et al.*, 2018 ; Labadie, 2004]. Dans ce cas, l'intérêt principal de prendre en compte les incertitudes pour la planification de la production est de gérer les risques associés aux prix et à la disponibilité de la ressource en eau. Les incertitudes étant plus importantes pour des horizons caractéristiques de planification lointains, la littérature concerne principalement la planification de la production à moyen ou long terme [Cuvelier *et al.*, 2018 ; Zhao *et al.*, 2010 ; Ge *et al.*, 2019 ; Pereira *et al.*, 1998 ; Li *et al.*, 1990 ; Sherkat *et al.*, 1985]. Néanmoins, la prise en compte des incertitudes pour la planification à court terme peut donner des solutions plus stables et plus intéressantes économiquement, voir par exemple [Fleten et Kristoffersen, 2007 ; De Ladurantaye *et al.*, 2009] où les incertitudes sont considérées sur les prix, [Séguin *et al.*, 2017 ; Fan *et al.*, 2016 ; Boucher *et al.*, 2012] où les incertitudes sont considérées sur les débits entrants, et [Fleten et Kristoffersen, 2008] où les incertitudes sont considérées à la fois sur les prix et les débits entrants. La formulation mathématique du problème d'optimisation probabiliste peut également tenir compte d'une aversion au risque, comme par exemple dans [García-González *et al.*, 2007]. La prise en compte des incertitudes dans la formulation du problème d'optimisation complexifie le problème, déjà souvent difficile à résoudre en univers déterministe (voir Sous-section 1.3.3). Pour une utilisation en opérationnel, il est donc crucial de trouver une approche de résolution numérique efficace [van Ackooij, 2017 ; van Ackooij et Malick, 2016 ; Carpentier *et al.*, 2018]. Par ailleurs, lorsque l'on considère la production conjointe d'actifs hydroélectriques et de production variable (parcs éoliens, centrales photovoltaïques), les incertitudes sur les prévisions des productions variables peuvent être valorisées dans l'optimisation de la chaîne hydroélectrique, voir par exemple [Vardanyan *et al.*, 2013 ; Zhu *et al.*, 2020].

À notre connaissance, en revanche, peu d'articles traitent un exemple concret, pour une utilisation en opérationnel, de la planification à court terme et sous incertitudes de la production conjointe d'une chaîne hydroélectrique au fil de l'eau sous fortes contraintes d'exploitation et d'actifs de production variable. Les articles cités ci-dessus utilisent souvent des approches de résolution basées sur des modélisations trop simplifiées du fonctionnement des chaînes hydroélectriques pour une utilisation en opérationnel. Pour des raisons de temps de calcul, ces approches ne s'appliquent généralement pas sur une modélisation plus fine et détaillée.

1.5 Objectifs de la thèse

Dans le cadre de la thèse, nous nous intéressons à la planification conjointe d'une chaine hydroélectrique au fil de l'eau et d'actifs de production variable (parcs éoliens et centrales photovoltaïques), voir Figure 1.17. Nous nous intéressons particulièrement au cas concret des actifs de production exploités par CNR (Compagnie Nationale du Rhône [24]), où la chaîne hydroélectrique est celle le long du Rhône. Le travail de la thèse peut néanmoins être transposé à n'importe quel producteur ayant un accès aux marchés et opérant une chaîne hydroélectrique (constituée de centrales hydroélectriques de moyenne ou basse chute) et des actifs de production variable.

En raison de la flexibilité journalière des centrales hydroélectriques au fil de l'eau, l'horizon caractéristique de planification est de quelques jours. Parmi les phases de la gestion opérationnelle présentées à la sous-section 1.3.2, nous considérons donc uniquement celle à court terme qui comprend :

— la décision des ventes au marché *day-ahead*,

— la construction du programme de production de la chaîne hydroélectrique avec une résolution temporelle au moins horaire (adaptée aux pas de temps horaires du marché *day-ahead*) et une modélisation suffisamment détaillée pour décrire correctement les fortes contraintes d'exploitation (voir §1.1.3.4).

24. `https://www.cnr.tm.fr/`

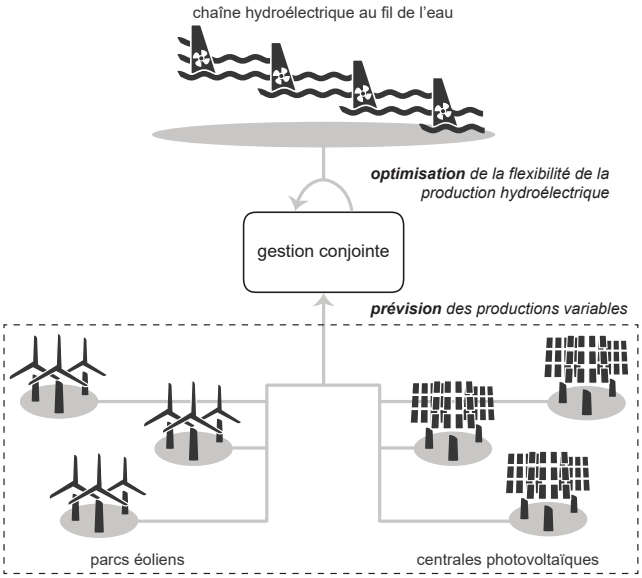


FIGURE 1.17 – Schéma de la gestion conjointe d'une chaîne hydroélectrique au fil de l'eau et d'actifs de production variable.

Comme les centrales hydroélectriques au fil de l'eau offrent une faible flexibilité, les ventes au marché *day-ahead* et le programme de construction sont définis conjointement (voir Sous-section 1.3.2). L'objectif est alors de construire le programme de production qui respecte les contraintes d'exploitation et qui maximise le chiffre d'affaires obtenu par les ventes de production au marché *day-ahead*, en agissant sur le placement de la production à la fois dans le temps et entre les centrales hydroélectriques. Nous nous intéressons plus précisément au problème mathématique d'optimisation qui en découle et à sa résolution numérique.

La problématique générale de la thèse concerne le problème de planification à court terme en univers probabiliste. Nous nous appuyons pour cela sur une formulation mathématique et un optimiseur déjà existants en univers déterministe pour tenter de les transposer en univers probabiliste.

Parmi les quatre étapes présentées dans la sous-section 1.4.2 pour passer l'optimiseur de l'univers déterministe à l'univers probabiliste, nous supposons que les deux premières, l'identification des paramètres incertains d'intérêt et la modélisation des incertitudes, sont des étapes réalisées en amont et indépendamment du travail de la thèse. Nous supposons également que les incertitudes sont modélisées par des scénarios multivariés (approche ensembliste).

L'objectif de la thèse est alors de construire les bases d'une méthodologie concernant les deux dernières étapes de la sous-section 1.4.2 :

— l'adaptation de la formulation mathématique du problème d'optimisation déterministe pour tenir compte des incertitudes de manière cohérente avec la gestion opérationnelle et la modélisation des incertitudes en scénarios multivariés,

— la résolution numérique du problème d'optimisation probabiliste de manière efficace pour une utilisation en opérationnel où les temps de calcul sont restreints.

Nous proposons également de tester et de discuter les développements méthodologiques sur un cas test réel à l'échelle de CNR.

1.6 Conclusion du chapitre

Le travail de la thèse concerne la planification à court terme de la production conjointe d'une chaîne hydroélectrique au fil de l'eau et d'actifs de production variable. Le problème consiste alors à effectuer, en même temps, la construction du programme de production hydroélectrique et la décision des prochaines ventes au marché *day-ahead*. L'objectif est de placer la production totale aux heures où les prix prévus sont les plus élevés afin de maximiser le chiffre d'affaires. Comme une telle gestion conjointe repose notamment sur des prévisions performantes, la prise en compte de leurs incertitudes dans la planification permet *a priori* d'améliorer la valorisation de la production. Nous nous intéressons donc au problème de planification en univers probabiliste.

Dans le chapitre suivant, nous nous appuyons sur le contexte général présenté dans ce chapitre pour situer le contexte plus spécifique de la planification de la production à CNR. Nous décrivons en particulier les points clefs du processus opérationnel et de l'optimiseur déterministe pour la planification de la production hydroélectrique du Rhône. Nous détaillons également les raisons qui nous poussent à faire passer l'optimiseur en univers probabiliste.

Chapitre 2

Données et cas d'étude

Dans ce chapitre, nous présentons le contexte plus spécifique de la gestion opérationnelle de la chaîne hydroélectrique du Rhône exploitée par CNR (Compagnie Nationale du Rhône) en le situant dans le contexte général présenté au chapitre précédent. Ce chapitre intéressera principalement les lecteurs et lectrices qui souhaitent comprendre la gestion opérationnelle menée par CNR de manière exhaustive. Nous mettons l'accent sur les points importants pour la compréhension du problème travaillé durant la thèse et présenté dans le reste du manuscrit. Nous présentons le modèle de simulation et les grandes lignes de la gestion opérationnelle du Rhône. Cela nous conduit à deux points essentiels de ce chapitre : la stratégie de placement de l'énergie et la présentation de l'optimiseur qui sert d'outil d'aide à la décision en opérationnel pour la planification de la production du Rhône. Au-delà de la gestion du Rhône seul, nous évoquons également l'intérêt d'une gestion conjointe du Rhône avec d'autres sources d'énergie renouvelable (éolien et solaire). Nous présentons également deux autres points essentiels de ce chapitre : les raisons qui nous poussent à prendre en compte les incertitudes de prévision dans l'optimiseur et les méthodes utilisées pour les modéliser. À la fin de ce chapitre, nous présentons également le cas d'étude utilisé tout le long du manuscrit pour les tests et les illustrations.

2.1 La chaîne hydroélectrique du Rhône

2.1.1 Le Rhône et CNR

Le Rhône est un des fleuves qui traversent la France (voir Figure 2.1). Il prend sa source au glacier du Rhône dans le massif du Saint-Gothard en Suisse, traverse le lac Léman jusqu'à Genève, longe et franchit la frontière franco-suisse puis se jette dans la mer Méditerranée (delta du Rhône qui forme la Camargue). Le Rhône s'écoule sur une longueur de 812 km, dont 531 km en France, drainant un bassin versant d'une superficie totale de 97 800 km^2, dont environ 90 000 km^2 en France.

L'aménagement de la partie française du Rhône a été confié dès 1934 à CNR (Compagnie Nationale du Rhône) autour de trois missions solidaires : la production d'électricité, la navigation fluviale et l'irrigation des terres agricoles. En tant que concessionnaire du Rhône, CNR exploite en particulier les 18 aménagements hydroélectriques en cascade le long de la partie française du Rhône (voir Figure 2.2 et Tableau 2.1), représentant une puissance installée d'environ 3 000 MW. En 2020, la production hydroélectrique CNR issue du Rhône s'est élevée à 13,5 TWh, ce qui représente un peu plus de 20% de toute la production hydroélectrique française [1] (voir Figure 1.1).

1. https://bilan-electrique-2020.rte-france.com/production-hydraulique/

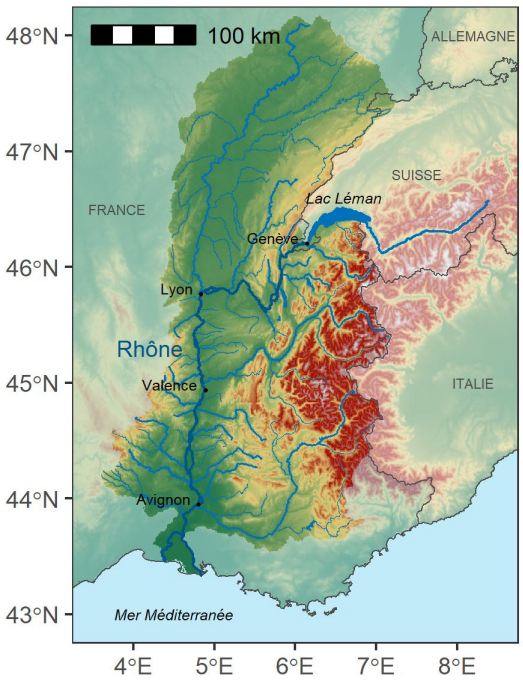


FIGURE 2.1 – Carte du bassin versant du Rhône (en aval du lac Léman) avec relief.

i	Abr.	Aménagements	Barrages	Centrales	Affluents principaux
1	GE	Génissiat	Génissiat*	Génissiat	Arve, Valserine
2	SY	Seyssel	Seyssel*	Seyssel	Usses
3	CE	Chautagne	Motz	Anglefort	Fier
4	BY	Belley	Savières**, Lavours	Brens-Virignin	-
5	BC	Brégnier-Cordon	Champagneux	Brégnier-Cordon	Séran
6	SB	Sault-Brénaz	Villebois	Porcieu-Amblagnieu	Guiers
7	PB	Pierre-Bénite	Pierre-Bénite	Pierre-Bénite	Ain, Saône
8	VS	Vaugris	Vaugris*	Vaugris	Gier
9	PR	Péage-de-Roussilon	St-Pierre-de-Boeuf	Sablons	-
10	SV	Saint-Vallier	Arras	Gervans	-
11	BV	Bourg-lès-Valence	La Roche-de-Glun	Bourg-lès-Valence	Doux, Isère
12	BE	Beauchastel	Charmes	Beauchastel	-
13	LN	Logis-Neuf	Le Pouzin	Logis-Neuf	Eyrieux, Drôme, Ouvèze RD
14	MO	Montélimar	Rochemaure	Châteauneuf	Roubion
15	DM	Donzère-Mondragon	Donzère	Bollène	-
16	CA	Caderousse	Caderousse	Caderousse	Ardèche
17	AV	Avignon	Villeneuve	Sauveterre, Avignon	Cèze, Aygues
18	VA	Vallabrègues	Vallabrègues	Beaucaire	Ouvèze RG, Durance, Gard†

* : aménagements constitués d'un barrage-centrale.
** : barrage secondaire servant de déversoir du lac du Bourget.
† : affluent qui conflue avec le Rhône en aval de l'aménagement (et non en amont), utilisé pour le calcul du niveau aval.

TABLEAU 2.1 – Aménagements hydroélectriques exploités par CNR le long du Rhône.

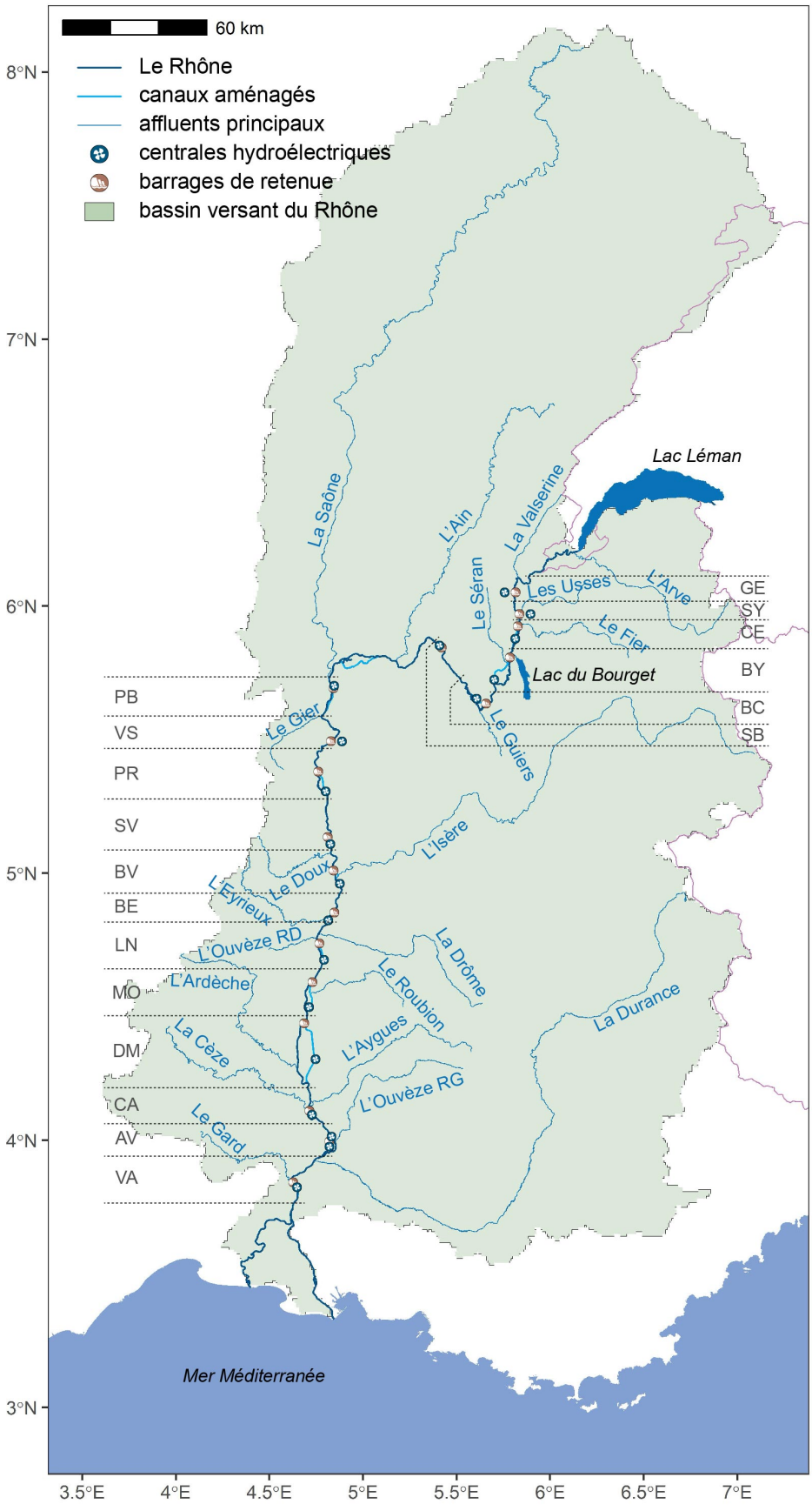


FIGURE 2.2 – Carte des aménagements hydroélectriques du Rhône exploités par CNR.

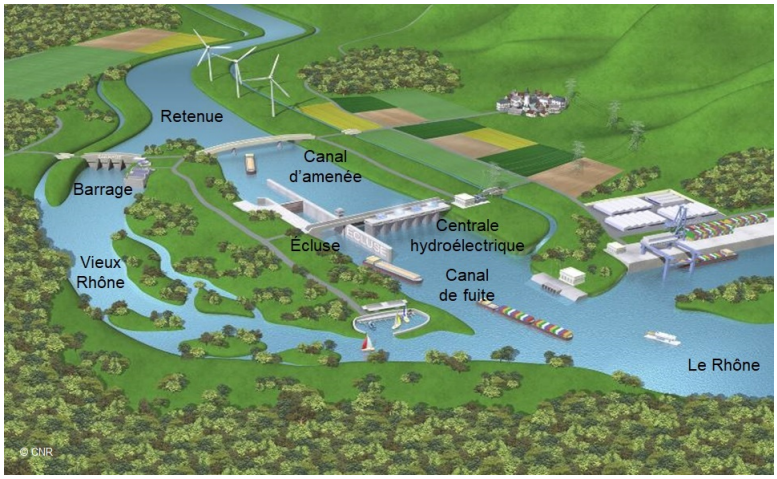


FIGURE 2.3 – Schéma d'un aménagement au fil de l'eau du Rhône.

2.1.2 Structure des aménagements hydroélectriques du Rhône

L'aménagement hydroélectrique de Génissiat, en tête de la chaîne hydroélectrique exploitée par CNR, est constitué d'un barrage-centrale de moyenne chute (hauteur de chute de 65 m). Tous les autres aménagements hydroélectriques en aval de Génissiat sont de type fil de l'eau, construits sur des tronçons du Rhône qui présentent naturellement un dénivelé suffisamment important (voir Figure 2.4). Ils sont composés d'un barrage de retenue, d'une centrale hydroélectrique et d'une écluse. Un canal de dérivation est creusé de manière à concentrer le dénivelé du tronçon en un point où sont implantées la centrale hydroélectrique et l'écluse. Le barrage de retenue, construit sur le cours naturel du Rhône, entraîne une grande partie de l'eau vers le canal de dérivation jusqu'à la centrale hydroélectrique (canal d'amenée). Le dénivelé concentré à la centrale hydroélectrique permet de créer une faible hauteur de chute. L'ensemble de l'aménagement forme une faible retenue en amont. Après être turbinée à la centrale, l'eau du canal de dérivation est restituée au Rhône à l'aval de la centrale (canal de fuite). La partie de l'eau qui n'est pas déviée vers la centrale transite par le barrage (où elle est généralement turbinée par des petites centrales hydroélectriques installées au barrage). Un débit minimal, appelé *débit réservé*, doit traverser en continu le barrage pour ne pas assécher le cours naturel du Rhône (Vieux Rhône), limitant l'impact de l'aménagement sur l'environnement et la biodiversité. En période de crue, le barrage sert d'évacuateur de crue en déversant le surplus d'eau une fois les capacités de turbinage de la centrale saturées. Un schéma de l'aménagement type est présenté à la figure 2.3. Quatre des 17 aménagements hydroélectriques au fil de l'eau font figures d'exception :

— pour les aménagements de Seyssel et Vaugris, le barrage et la centrale hydroélectrique sont accolés, formant un seul ouvrage, sans canal de dérivation,

— l'aménagement de Bourg-lès-Valence dispose d'un second barrage qui permet d'évacuer une partie du débit de l'Isère,

— l'aménagement d'Avignon comporte un barrage-centrale supplémentaire sur le bras mort de la Barthelasse qui contourne le canal de dérivation.

2.2 Modélisation du Rhône

La modélisation du Rhône consiste à mettre en équations le fonctionnement de la chaîne hydroélectrique (Sous-section 2.2.1) et les contraintes d'exploitation (Sous-section 2.2.3). Le modèle de simulation ainsi obtenu (Sous-section 2.2.4) est utilisé dans la gestion opérationnelle du Rhône (Section 2.3) [Grimaldi *et al.*, 2014 ; Gemaehling, 1962].

2.2.1 Modèle hydraulique

2.2.1.1 Système de réservoirs en cascade

La chaîne hydroélectrique du Rhône est modélisée par un système de réservoirs en cascade (voir Figure 2.4). On note $\mathcal{I} = \{1, \dots, I\}$, avec $I = 18$, les indices des aménagements hydroélectriques de l'amont vers l'aval. Pour chaque aménagement hydroélectrique, le niveau de la retenue utilisé pour la conduite de l'aménagement, appelé *cote*, est calculé et mesuré à un point virtuel de référence, appelé *point de réglage*, situé sur le cours du fleuve immédiatement en amont du canal d'amenée. La figure 2.5 illustre la modélisation en réservoir d'un aménagement hydroélectrique du Rhône.

2.2.1.2 Bilan d'eau

Les principales équations du système de réservoirs en cascade sont les bilans d'eau (i.e. les lois de conservation de l'eau) aux points de réglage des aménagements hydroélectriques :

$$\forall (i,t) \in \mathcal{I} \times \mathbb{R}, \quad Q_{\text{ent}}(i,t) = Q_{\text{sort}}(i,t) + \frac{\partial V}{\partial t}(i,t), \tag{2.1}$$

où $Q_{\text{ent}}(i,t) \in \mathbb{R}_+$, $Q_{\text{sort}}(i,t) \in \mathbb{R}_+$ et $V(i,t) \in \mathbb{R}_+$ désignent respectivement le débit entrant, le débit sortant et le volume de retenue de l'aménagement d'indice $i \in \mathcal{I}$ calculés au point de réglage à l'instant $t \in \mathbb{R}$. Le dernier terme représente le débit qui alimente la retenue. L'équation (2.1) exprime donc que l'eau qui entre à un aménagement en sort ou alimente la retenue.

Les cotes peuvent se calculer à partir des volumes, et inversement, grâce à des fonctions obtenues par calage expérimental :

$$\forall (i,t) \in \mathcal{I} \times \mathbb{R}, \quad \begin{cases} Z(i,t) = \varphi_{Z,i}(V(i,t), V(i,t-\delta t), Q_{\text{sort}}(i,t)), \\ V(i,t) = \varphi_{V,i}(Z(i,t), V(i,t-\delta t), Q_{\text{sort}}(i,t)), \end{cases}$$

où $Z(i,t) \in \mathbb{R}_+$ désigne la cote de l'aménagement d'indice $i \in \mathcal{I}$ à l'instant t (au point de réglage), $\delta t > 0$ la durée pour le calcul numérique des variations, et $\varphi_{Z,i}$ et $\varphi_{V,i}$ les fonctions obtenues par calage expérimental (lois de volume ou de surface).

2.2.1.3 Débit sortant

Le débit sortant à un aménagement est constitué du débit turbiné à la centrale hydroélectrique et du débit qui traverse le barrage de retenue, en tenant compte des temps de retard [2] (anticipation) par rapport au point de réglage situé légèrement en amont (voir figure 2.5) :

$$\forall (i,t) \in \mathcal{I} \times \mathbb{R}, \quad Q_{\text{sort}}(i,t) = Q_{\text{US}}(i, t - \tau_{\text{US}}(i,i)) + Q_{\text{BR}}(i, t - \tau_{\text{BR}}(i,i)), \tag{2.2}$$

où

- $Q_{\text{US}}(i,t) \in \mathbb{R}_+$ et $Q_{\text{BR}}(i,t) \in \mathbb{R}_+$ désignent respectivement le débit qui traverse la centrale hydroélectrique et celui qui traverse le barrage de retenue de l'aménagement d'indice $i \in \mathcal{I}$ à l'instant $t \in \mathbb{R}$,
- $\tau_{\text{US}}(i,i) \in \mathbb{R}_+$ et $\tau_{\text{BR}}(i,i) \in \mathbb{R}_+$ désignent respectivement les temps de retard (supposés indépendants des débits) des débits de la centrale hydroélectrique et du barrage de retenue par rapport au point de réglage de l'aménagement d'indice $i \in \mathcal{I}$.

La modélisation en réservoir d'un aménagement du Rhône de la figure 2.5 est aussi valable pour les aménagements de Génissiat, Seyssel et Vaugris, en considérant que le barrage de retenue et la centrale hydroélectrique sont accolés au sein d'un même ouvrage (les temps de retard τ_{US} et τ_{BR} de ces aménagements sont alors identiques).

2. Les temps de retard représentent le temps que mettrait une vague émise au barrage ou la centrale pour se propager jusqu'au point de réglage (en amont) en remontant le cours de l'eau.

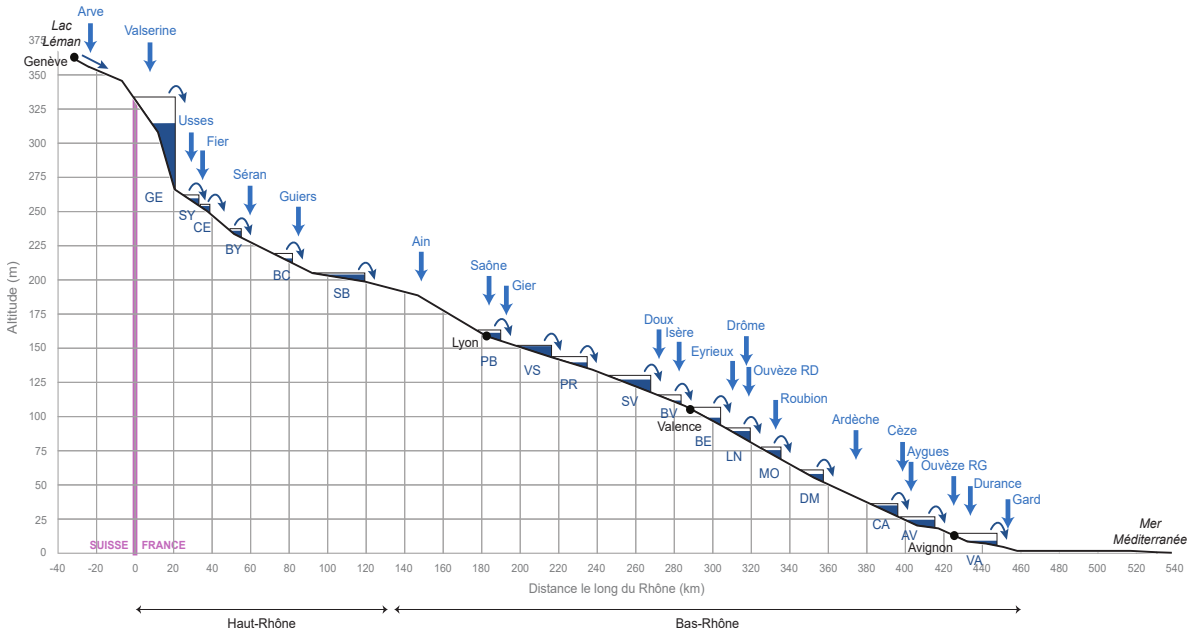


FIGURE 2.4 – Schéma du système de réservoirs en cascade (inspiré du schéma de la structure des pentes du Rhône de `https://fr.wikipedia.org/wiki/Rhône`). Les niveaux de remplissage des retenues n'ont pas de signification particulière dans ce schéma.

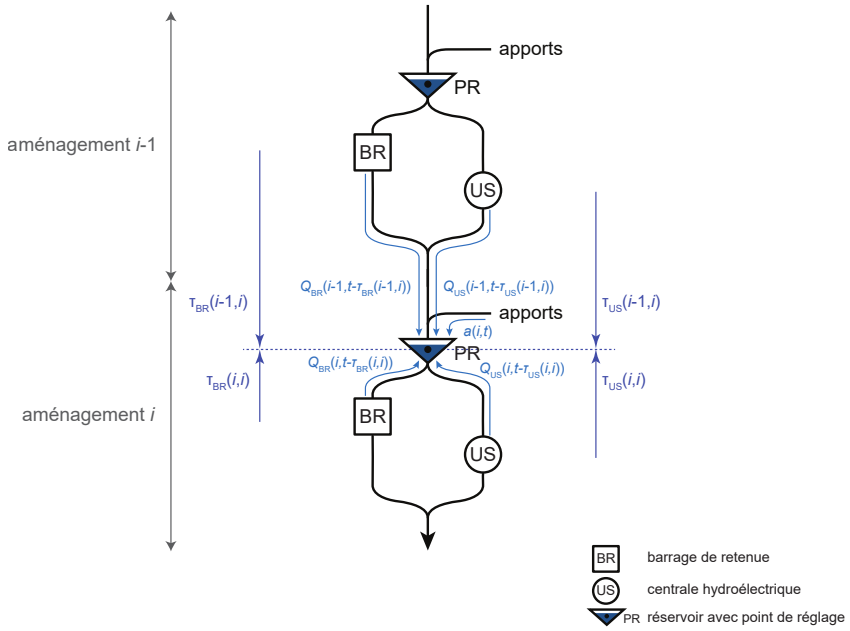


FIGURE 2.5 – Schéma de la modélisation en réservoir d'un aménagement du Rhône.

2.2.1.4 Débit entrant

Le débit entrant à un aménagement est constitué du débit sortant de l'aménagement amont qui s'est propagé et des débits d'apports en eau qui se déversent dans le Rhône entre l'aménagement amont et l'aménagement courant (voir Figure 2.5). Avec les notations précédentes :

$$\forall (i,t) \in \mathcal{I} \times \mathbb{R},\ \ Q_{\text{ent}}(i,t) = Q_{\text{US}}(i-1, t-\tau_{\text{US}}(i-1,i)) + Q_{\text{BR}}(i-1, t-\tau_{\text{BR}}(i-1,i)) + a(i,t), \quad (2.3)$$

où

- $a(i,t) \in \mathbb{R}$ désigne le débit d'apports en eau calculé au point de réglage de l'aménagement d'indice $i \in \mathcal{I}$ à l'instant t,
- $\tau_{\text{US}}(i-1,i) \in \mathbb{R}_+$ et $\tau_{\text{BR}}(i-1,i) \in \mathbb{R}_+$ désignent respectivement les temps de propagation (supposés indépendants des débits) du débit de la centrale hydroélectrique et du barrage de retenue de l'aménagement amont jusqu'au point de réglage de l'aménagement courant d'indice $i \in \mathcal{I}$.

Pour l'aménagement de tête d'indice $i = 1$ (Génissiat), on considère que l'indice $i - 1 = 0$ correspond à la centrale hydroélectrique de Verbois construite sans canal d'amenée sur le Rhône en Suisse, qui n'est pas exploitée par CNR mais par les Services Industriels de Genève. Ainsi, pour tout $t \in \mathbb{R}$, $Q_{\text{US}}(0,t)$ correspond au débit du Rhône à Verbois à l'instant t et $Q_{\text{BR}}(0,t) = 0$. Le débit entrant à l'aménagement de tête (Génissiat) est donc influencé par les opérations des Services Industriels de Genève.

2.2.1.5 Apports en eau

Les apports en eau sont issus des 21 affluents principaux pour lesquels les débits sont mesurés (voir Figures 2.2 et 2.4) et des autres petits affluents pour lesquels les débits ne sont pas mesurés (bassins versants intermédiaires). Avec ces apports en eau, le bilan d'eau de l'équation (2.1) n'est pas respecté avec les valeurs mesurées en pratique. Cela est dû aux simplifications de la modélisation hydraulique (par exemple, la chaîne hydroélectrique n'est pas parfaitement un système de réservoirs en cascade avec des temps de retard et de propagation constants) et aux erreurs de mesure ou d'estimation des débits. Des débits de correction (qui peuvent être négatifs) sont donc ajoutés aux apports en eau pour « forcer » le respect du bilan d'eau de l'équation (2.1). Plus de détails sur les débits de correction sont présentés dans le §2.3.4.2.

On suppose que les 21 affluents principaux (et mesurés) du Rhône sont indexés de l'amont vers l'aval de leur confluent. Les apports en eau s'écrivent ainsi :

$$\forall (i,t) \in \mathcal{I} \times \mathbb{R}, \quad a(i,t) = \sum_{k \in \mathcal{A}(i)} Q_{\text{aff}}(k, t - \tau_{\text{aff}}(k,i)) + Q_{\text{BVI}}(i,t) + Q_{\text{corr}}(i,t), \quad (2.4)$$

où

- $\mathcal{A}(i) \subset \{1, \ldots, 21\}$ contient les indices des affluents principaux qui se jettent dans le Rhône entre l'aménagement amont d'indice $i-1$ et l'aménagement courant d'indice i (si $i = 1$, alors il s'agit des indices des affluents principaux qui se jettent entre la centrale hydroélectrique de Verbois et l'aménagement de Génissiat),
- $Q_{\text{aff}}(k,t) \in \mathbb{R}_+$ désigne le débit de l'affluent d'indice $k \in \mathcal{A}(i)$ à une station de mesure située généralement en amont immédiat du confluent avec le Rhône,
- $\tau_{\text{aff}}(k,i)$ désigne le temps de propagation (supposé indépendant du débit) entre la station de mesure de l'affluent d'indice $k \in \mathcal{A}(i)$ et le point de réglage de l'aménagement d'indice $i \in \mathcal{I}$,
- $Q_{\text{BVI}}(i,t) \in \mathbb{R}_+$ et $Q_{\text{corr}}(i,t) \in \mathbb{R}$ désignent respectivement le débit des bassins versants intermédiaires et le débit de correction (présenté dans le §2.3.4.2) affectés de façon arbitraire au point de réglage de l'aménagement d'indice $i \in \mathcal{I}$ à l'instant t.

Les caractéristiques (régimes hydrologiques, climatologies, altitudes, superficies et morphologies des bassins versants, sensibilité aux crues etc.) des affluents sont très variées. Néanmoins, les débits d'apports des affluents dépendent des conditions hydro-météorologiques de chaque bassin versant qui peuvent être semblables du fait, notamment, de la proximité géographique des bassins versants. Les débits d'apports des affluents peuvent donc être corrélés entre eux. On dit qu'il y a concomitance des débits des affluents du Rhône.

Pour certains affluents du Rhône, notamment le Fier, l'Ain, l'Isère, l'Ardèche et la Durance, les débits d'apports ne correspondent pas à l'écoulement naturel de l'eau mais sont influencés par des aménagements hydroélectriques qui ne sont pas exploités par CNR mais par EDF (Électricité De France). En outre, le canal de Savières, l'émissaire du lac du Bourget, se jette dans le Rhône au niveau du barrage de Savières, en amont de l'aménagement de Belley. Le barrage de Savières permet de réguler le niveau du canal (pour la navigation) et celui du lac du Bourget. Le débit du canal de Savières est comptabilisé dans les apports en eau de l'aménagement de Belley. Ce débit peut être négatif car le cours d'eau du canal de Savières s'inverse naturellement en période de crue, refluant l'eau vers le lac du Bourget.

2.2.1.6 Lois de puissance

Les centrales hydroélectriques sont constituées de plusieurs groupes de production qui peuvent être démarrés ou arrêtés. L'eau qui arrive à la centrale se répartit dans les différents groupes de production démarrés. Le nombre de groupes de production démarrés et le débit qui arrive à la centrale sont fortement dépendants. Une partie du débit qui traverse la centrale, appelée *débit déchargé*[3], peut ne pas être turbinée lorsque les capacités de turbinage de la centrale sont saturées, lors des crues (débit entrant trop important) ou lors de certaines opérations de maintenance (diminution de la capacité de turbinage). Selon les opérations de maintenance ou les avaries, certains groupes de production peuvent être indisponibles et, dans ce cas, ils ne peuvent pas être démarrés pour la production d'électricité. Ainsi :

$$\forall (i,t) \in \mathcal{I} \times \mathbb{R}, \quad Q_{\mathrm{US}}(i,t) = Q_{\mathrm{dech}}(i,t) + \sum_{k \in \mathcal{G}(i)} Q_{\mathrm{Gr}}(i,k,t),$$

où

- $Q_{\mathrm{dech}}(i,t) \in \mathbb{R}_+$ désigne le débit déchargé à la centrale hydroélectrique de l'aménagement d'indice $i \in \mathcal{I}$ à l'instant t,
- $\mathcal{G}(i)$ désigne l'ensemble des indices des groupes de production de la centrale hydroélectrique de l'aménagement d'indice $i \in \mathcal{I}$,
- $Q_{\mathrm{Gr}}(i,k,t) \in \mathbb{R}_+$ désigne le débit qui traverse le groupe de production d'indice $k \in \mathcal{G}(i)$ de l'aménagement d'indice $i \in \mathcal{I}$ à l'instant t (nul si le groupe de production est arrêté).

Les groupes de production démarrés produisent de l'électricité en fonction du débit, de la hauteur de chute nette et du rendement (voir Équation (1.1)). Les hauteurs de chute nettes dépendent des cotes et des débits (à l'aménagement courant et l'aménagement aval) par des fonctions obtenues par calage expérimental :

$$\begin{aligned} \forall (i,t) \in \mathcal{I} \times \mathbb{R}, \quad H(i,t) = \varphi_{H,i}(&Z(i,t), Z(i+1,t), \\ &Q_{\mathrm{US}}(i,t), Q_{\mathrm{US}}(i+1,t), \\ &Q_{\mathrm{BR}}(i,t), Q_{\mathrm{BR}}(i+1,t), \\ &a(i+1,t)), \end{aligned}$$

3. Les centrales hydroélectriques disposent d'organes spécifiques de décharge dont l'un des buts principaux est de faire transiter le débit en cas d'arrêt brutal d'un groupe de production afin de limiter l'ampleur de la vague qui pourrait être générée dans le canal d'amenée.

où $H(i,t) \in \mathbb{R}_+$ désigne la hauteur de chute nette à la centrale hydroélectrique de l'aménagement d'indice $i \in \mathcal{I}$ à l'instant t. Si $i = I$ (aménagement de Vallabrègues en dernière position de la chaîne hydroélectrique), alors $a(i+1,t) \in \mathbb{R}$ désigne le débit d'apports en aval du dernier aménagement (principalement issu du Gard) à un point de référence situé au confluent du Gard. On utilise la convention : $\forall t \in \mathbb{R},\ Q_{\mathrm{US}}(I+1,t) = Q_{\mathrm{BR}}(I+1,t) = 0$.

Les rendements dépendent des débits qui traversent les groupes de production selon des fonctions obtenues par calage expérimental :

$$\forall (i,t) \in \mathcal{I} \times \mathbb{R},\ \forall k \in \mathcal{G}(i), \quad \eta(i,k,t) = \varphi_{\eta,i,k}(Q_{\mathrm{Gr}}(i,k,t)), \tag{2.5}$$

où $\eta(i,k,t) \in [0,1]$ désigne le rendement du groupe de production d'indice $k \in \mathcal{G}(i)$ de l'aménagement d'indice $i \in \mathcal{I}$ à l'instant t, et $\varphi_{\eta,i,k}$ la fonction obtenue par calage expérimental.

Les puissances brutes instantanées produites par les groupes de production sont alors données par (voir Équation (1.1)) :

$$\forall (i,t) \in \mathcal{I} \times \mathbb{R},\ \forall k \in \mathcal{G}(i), \quad P_{\mathrm{Gr}}(i,k,t) = \eta(i,k,t)\rho g H(i,t) Q_{\mathrm{Gr}}(i,k,t), \tag{2.6}$$

où $P_{\mathrm{Gr}}(i,k,t) \in \mathbb{R}_+$ désigne la puissance active brute instantanée produite par le groupe de production d'indice $k \in \mathcal{G}(i)$ de l'aménagement d'indice $i \in \mathcal{I}$ à l'instant t, $\rho = 1\,000\,\mathrm{kg/m^3}$ la masse volumique de l'eau (supposée constante), et $g = 9{,}81\,\mathrm{m/s^2}$ l'accélération de la pesanteur (supposée constante).

Les puissances nettes instantanées produites aux centrales hydroélectriques sont alors la somme des puissances produites par les groupes de production à laquelle on soustrait les éventuelles pertes de puissance :

$$\forall (i,t) \in \mathcal{I} \times \mathbb{R}, \quad P_{\mathrm{US}}(i,t) = \gamma_{\mathrm{US}}(i) \sum_{k \in \mathcal{G}(i)} P_{\mathrm{Gr}}(i,k,t) - \Delta P_{\mathrm{cor}}(i,t),$$

où $P_{\mathrm{US}}(i,t) \in \mathbb{R}_+$ et $\Delta P_{\mathrm{cor}}(i,t) \in \mathbb{R}_+$ désignent respectivement la puissance nette produite et les pertes de puissance à la centrale hydroélectrique de l'aménagement d'indice $i \in \mathcal{I}$ à l'instant t, et $\gamma_{\mathrm{US}}(i) \in [0,1]$ désigne un coefficient d'ajustement (presque toujours égal à 1) de la puissance pour la centrale hydroélectrique de l'aménagement d'indice $i \in \mathcal{I}$. Les coefficients d'ajustement et les pertes de puissance permettent de tenir compte des objets divers transportés par le Rhône (branchages, déchets etc.) qui s'accumulent sur les grilles d'entrée des groupes de production, entraînant des pertes de production par colmatage. Ces pertes par colmatage sont très difficiles à anticiper et à prévoir. Elles sont généralement ajustées manuellement via le coefficient d'ajustement γ_{US}. Lorsque les pertes par colmatage dépassent un certain seuil, l'exploitant effectue, sur un créneau souvent planifié en avance, une opération de dégrillage qui consiste à retirer les objets accumulés sur les grilles. Pour cela, il doit baisser manuellement la puissance des groupes de production (voire les arrêter selon le type de groupe et le système de dégrillage en place) pour limiter l'aspiration, entraînant des pertes dites par dégrillage. Une bonne partie de ces pertes par dégrillage sont prévues en avance et elles sont prises en compte dans les pertes de puissances ΔP_{cor}.

De manière analogue, on peut calculer les puissances nettes instantanées $P_{\mathrm{BR}}(i,t) \in \mathbb{R}_+$, pour $(i,t) \in \mathcal{I} \times \mathbb{R}$, produites par les petites centrales hydroélectriques (PCH) installées sur les barrages qui ne sont pas ou plus en obligation d'achat (voir Sous-section 1.1.1). Les productions des PCH en obligation d'achat n'étant pas échangées sur le marché de l'énergie, elles ne sont pas comptabilisées dans la puissance totale de la chaîne hydroélectrique.

La puissance totale nette instantanée produite par l'ensemble de la chaîne hydroélectrique est alors la somme des puissances nettes produites par tous les aménagements hydroélectriques :

$$\forall t \in \mathbb{R}, \quad P_{\mathrm{hydro}}(t) = \sum_{i \in \mathcal{I}} (P_{\mathrm{US}}(i,t) + P_{\mathrm{BR}}(i,t)), \tag{2.7}$$

où $P_{\text{hydro}}(t) \in \mathbb{R}_+$ désigne la puissance nette totale produite par la chaîne hydroélectrique à l'instant t.

2.2.1.7 Autres équations du modèle hydraulique

Les équations détaillées précédemment constituent le socle de base du modèle hydraulique, permettant notamment de décrire la dynamique en temps et en espace de la chaîne hydroélectrique (grâce au bilan d'eau de l'équation (2.1)). En réalité, le modèle hydraulique du Rhône est beaucoup plus riche afin de modéliser correctement le fonctionnement des aménagements hydroélectriques et les écoulements de l'eau (gestion du démarrage des groupes de production, de la répartition des débits entre barrages et centrales, de la saturation des centrales hydroélectriques etc.). Ces équations ne sont pas décrites dans ce manuscrit car nous n'avons pas besoin de ce niveau de détail pour décrire la méthodologie travaillée durant la thèse et certaines équations sont confidentielles. Néanmoins, la complexité du modèle hydraulique est une caractéristique importante de notre problème qui est, à notre connaissance, peu présente dans la littérature.

2.2.2 Modulation de la production

L'écriture du bilan d'eau de l'équation (2.1) permet de faire le lien entre la capacité de stockage de la retenue et la modulation du débit sortant par rapport au débit entrant au niveau des aménagements. La capacité de stockage de la retenue de l'aménagement de Génissiat (moyenne chute) est suffisamment grande pour moduler le débit sortant à l'échelle de quelques jours. En revanche, la capacité de stockage des autres aménagements (au fil de l'eau) ne permet une modulation du débit sortant qu'à l'échelle de quelques heures (voir §1.1.3.4). Plus précisément, les capacités de stockage des aménagements se caractérisent par :

$$\forall i \in \mathcal{I}, \quad |V(i,t_2) - V(i,t_1)| \ll \int_{t_1}^{t_2} Q_{\text{ent}}(i,t)\,\mathrm{d}t,$$

où $[t_1, t_2]$ est un intervalle de temps dont la durée est de l'ordre d'une semaine pour l'aménagement de Génissiat ($i = 1$) et de l'ordre d'une journée pour les autres aménagements ($i \in \{2, \ldots, 18\}$). Le terme de gauche représente la variation de volume durant l'intervalle de temps $[t_1, t_2]$. Le terme de droite représente le volume d'eau cumulé qui est entré à l'aménagement durant l'intervalle de temps $[t_1, t_2]$. Ainsi, à partir du bilan d'eau de l'équation (2.1), il vient :

$$\forall i \in \mathcal{I}, \quad \int_{t_1}^{t_2} Q_{\text{sort}}(i,t)\,\mathrm{d}t = \int_{t_1}^{t_2} Q_{\text{ent}}(i,t)\,\mathrm{d}t - (V(i,t_2) - V(i,t_1)) \simeq \int_{t_1}^{t_2} Q_{\text{ent}}(i,t)\,\mathrm{d}t.$$

La modulation du débit sortant par rapport au débit entrant durant l'intervalle de temps $[t_1, t_2]$ est donc presque nulle :

$$\forall i \in \mathcal{I}, \quad \int_{t_1}^{t_2} (Q_{\text{sort}}(i,t) - Q_{\text{ent}}(i,t))\,\mathrm{d}t \simeq 0.$$

Autrement dit, l'augmentation ou la diminution du débit sortant par rapport au débit entrant (vidange ou rétention de la retenue) ne peut s'effectuer que sur des périodes de temps de quelques jours pour l'aménagement de Génissiat et quelques heures pour les autres aménagements. En outre, une période d'augmentation du débit sortant (vidange) doit être compensée par une période de diminution (rétention) dans les heures qui suivent, et inversement. La figure 2.6 illustre sur un exemple la modulation du débit sortant : les périodes d'augmentation (respectivement de diminution) du débit sortant par rapport au débit entrant entraînent une diminution (respectivement une augmentation) du volume de la retenue.

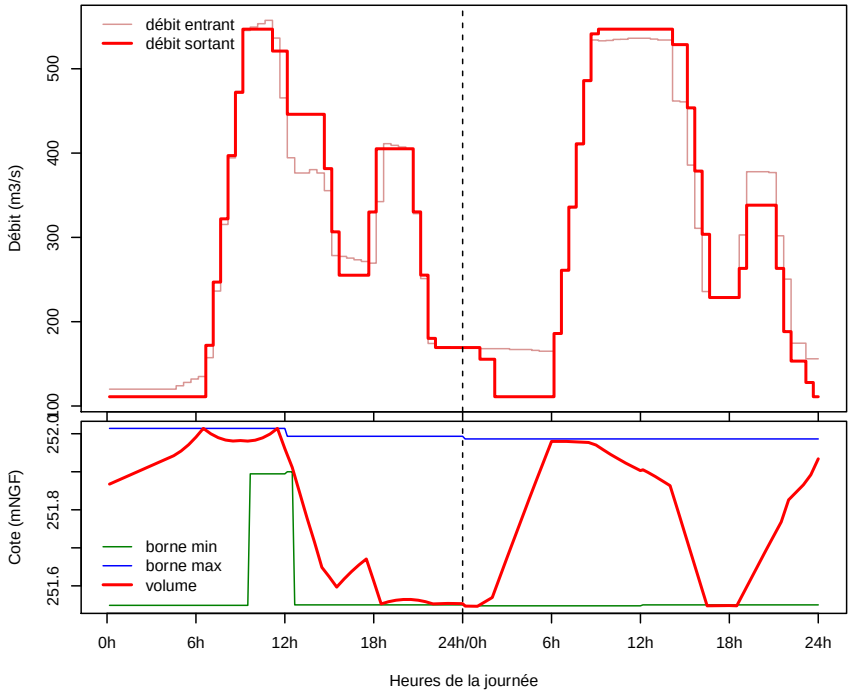


FIGURE 2.6 – Exemple du lien entre la cote de la retenue et la modulation du débit sortant par rapport au débit entrant à un aménagement du Rhône sur une fenêtre temporelle de 2 jours. Le graphique du haut représente l'évolution des débits entrant et sortant au cours du temps tandis que celui du bas représente l'évolution de la cote au cours du temps.

Le débit qui traverse le barrage est presque toujours uniquement constitué du débit réservé. La modification du débit à un barrage ne doit se faire que s'il n'y a pas le choix (en crue, par exemple, où il est nécessaire de déverser le surplus d'eau qui saturerait la centrale) et doit être justifiée aux autorités de tutelle (DREAL [4]). Ainsi, dans les conditions habituelles (zone énergétique), la modulation du débit sortant (permise par la capacité de stockage de la retenue) est entièrement déterminée par la modulation du débit qui traverse la centrale hydroélectrique.

Comme le rendement et la hauteur de chute nette dépendent du débit turbiné, la puissance produite (voir Équation (2.6)) est en pratique une fonction (non linéaire) du débit turbiné. La modulation du débit qui traverse la centrale (et donc la modulation du débit sortant) est donc entièrement déterminée par la modulation de la puissance produite, en agissant sur le démarrage et l'arrêt des groupes de production ou sur la position du système de vannage des groupes à bulbe réglant pour les aménagements qui en sont équipés [5].

Par conséquent, la faible capacité de stockage des aménagements au fil de l'eau ne permet que peu de modulation du débit sortant par rapport au débit entrant au niveau de chaque aménagement : tout au plus est-il possible de sur-produire ou de sous-produire durant quelques heures d'une même journée sur chaque aménagement. En particulier, la production d'un aménagement au fil de l'eau est fortement dépendante du débit entrant et donc du débit des apports en eau (voir Équation (2.3)). La production de la chaîne hydroélectrique du Rhône est ainsi fortement dépendante des conditions hydro-météorologiques.

4. Direction Régionale de l'Environnement, de l'Aménagement et du Logement.

5. Hormis les aménagements de Seyssel, de Bourg-lès-Valence, de Beauchastel, de Logis-Neuf, de Montélimar et de Donzère-Mondragon qui sont équipés de turbines Kaplan, les 12 autres aménagements de basse chute du Rhône sont équipés de groupes à bulbe réglant (et aussi des groupes à bulbe fixe pour les aménagements de Vaugris et Caderousse). L'aménagement de moyenne chute de Génissiat est équipé de turbines Francis.

2.2.3 Contraintes d'exploitation

Les équations de la Sous-section 2.2.1 permettent de définir un modèle hydraulique qui décrit le fonctionnement des aménagements hydroélectriques du Rhône et les écoulements de l'eau. À ce modèle hydraulique s'ajoutent les contraintes d'exploitation qui doivent être respectées pour assurer le bon fonctionnement des aménagements, garantir la sûreté hydraulique et limiter l'impact des aménagements sur l'environnement (voir §1.1.3.4). La gestion de la production du Rhône (i.e. la modulation des débits sortants par rapport aux débits entrants aux aménagements hydroélectriques) est effectuée en associant le modèle hydraulique et les contraintes d'exploitation.

Les contraintes d'exploitation se traduisent généralement par des bornes minimales et maximales sur des variables qui dépendent du modèle hydraulique. Elles peuvent porter sur des variables qui interviennent directement dans le modèle hydraulique (par exemple, les cotes, les débits entrants et sortants) ou des variables qui n'interviennent pas directement dans le modèle hydraulique mais qui s'en déduisent (par exemple, les variations de débits, les débits moyens journaliers). En pratique, les bornes minimales et maximales des contraintes d'exploitation considérées pour la gestion prévisionnelle de la production tiennent souvent compte de marges supplémentaires afin de faire face à des aléas en temps réel.

Par exemple, les niveaux d'eau (cotes) des aménagements doivent toujours rester entre des bornes minimales et maximales, limitant le marnage possible. Pour les aménagements du Rhône, ces bornes sur les niveaux dépendent du débit de pilotage (débit calculé au point de réglage) selon une courbe d'exploitation qui n'est pas nécessairement linéaire, voir Figure 2.7 [Piron *et al.*, 2015]. La conduite de l'aménagement est opérée en fonction de l'inclinaison du plan d'eau autour du point de réglage. En période d'étiage (faibles débits), la cote au point de réglage est aussi élevée que possible pour compenser une partie de l'abaissement du plan d'eau en amont du point de réglage et maintenir ainsi un niveau d'eau suffisant pour la navigation (passage des tirants d'eau des bateaux). En début de période de crue, le niveau maximal est diminué pour compenser l'élévation du plan d'eau en amont du point de réglage et éviter ainsi des inondations en amont. Si les débits augmentent et qu'il n'est plus possible de diminuer le niveau au point de réglage, alors l'aménagement devient neutre face à la crue, gardant le plus possible les niveaux du plan d'eau à leurs valeurs naturelles avant la construction de l'aménagement. En dehors de ces deux périodes extrêmes, en période dite *énergétique*, la production de l'aménagement peut être modulée en jouant sur le marnage de la retenue tout en respectant les contraintes d'exploitation.

2.2.4 Modèle de simulation

La gestion de la production de la chaîne hydroélectrique consiste à définir les valeurs des variables qui interviennent dans le modèle hydraulique et les contraintes d'exploitation.

Certaines variables du modèle hydraulique et des contraintes d'exploitation ont des valeurs fixées car elles ne dépendent pas de l'utilisation de la modulation de la production. C'est le cas, par exemple, des apports en eau (qui interviennent dans les débits entrants), des pertes par dégrillage, du nombre de groupes de production disponibles et de la plupart des bornes minimales et maximales des contraintes d'exploitation. Dans la suite du manuscrit, on dit que ces variables sont des *paramètres*. Au contraire, les autres variables sont dites de *décision* (ou simplement les *variables* s'il n'y a pas d'ambiguïté) car leurs valeurs dépendent directement de l'utilisation de la modulation de la production de la chaîne hydroélectrique. C'est le cas, par exemple, des débits sortants des aménagements, des volumes des retenues, des cotes, du nombre de groupes de production démarrés, des puissances produites. C'est aussi le cas des bornes minimales et maximales sur les cotes car elles se calculent à partir des débits (courbes d'exploitation du §1.1.3.4).

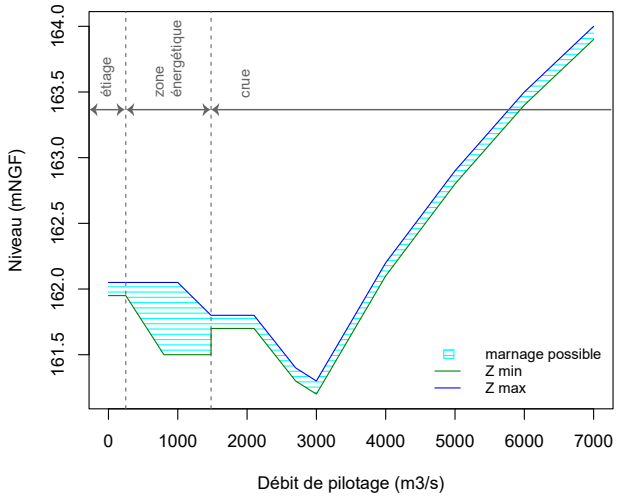


FIGURE 2.7 – Exemple de la courbe d'exploitation d'un aménagement du Rhône.

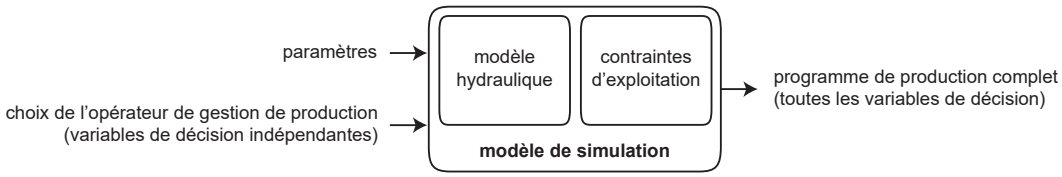


FIGURE 2.8 – Schéma du modèle de simulation du Rhône.

Les variables de décision sont redondantes car elles sont fortement dépendantes, réduisant les libertés de la gestion de la production. Souvent, le choix des valeurs d'un seul type de variables permet de déduire les valeurs de toutes les autres variables en appliquant les équations du modèle hydraulique et des contraintes d'exploitation.

Un modèle de simulation (numérique) est alors construit en combinant le modèle hydraulique et les contraintes d'exploitation (voir Figure 2.8). À partir des valeurs des paramètres et de variables indépendantes données en entrée, le modèle de simulation calcule numériquement les valeurs des autres variables dépendantes et il indique si les contraintes d'exploitation sont toutes respectées. En pratique, le modèle hydraulique et les contraintes d'exploitation (et donc le modèle de simulation) peuvent changer selon le régime hydraulique (étiage, crue ou période énergétique, voir Figure 2.7).

Le choix des variables indépendantes en entrée du modèle de simulation dépend du *mode de modulation*. Pour l'aménagement de Génissiat, on utilise généralement un mode de modulation en puissance ou nombre de groupes démarrés. Pour les autres aménagements, on utilise généralement un mode de modulation en débit ou en cote. Il existe également un mode de modulation, appelé *mode de modulation support*, qui impose une variation de volume nulle (fonctionnement parfaitement au fil de l'eau, i.e. sans stockage de l'eau). Des algorithmes de régulation, intégrés au modèle de simulation, sont également utilisés pour définir automatiquement les valeurs des variables de décision sur une plage de temps donnée pour atteindre une valeur cible (en cote ou débit) en fin de plage. Le mode de modulation peut changer d'un aménagement à un autre et d'une plage de temps à une autre. Le choix et le paramétrage des modes de modulation sont effectués par les opérateurs de gestion de production (selon une procédure bien définie).

Le modèle (numérique) de simulation est défini en pratique sur des pas de temps réguliers d'une durée de $\delta t = 10\,\text{min}$, une résolution temporelle suffisamment fine pour la planification à court terme de la production (voir Section 2.3). Le modèle n'est donc pas continu. En particulier, les valeurs des variables et des paramètres sont moyennées sur les pas de temps (granularité de 10 min) ou des périodes composées de plusieurs pas de temps (par exemple, granularité de 30 min ou 1 h). D'autres modèles de simulation, avec une modélisation plus fine et précise, sont également utilisés à CNR (notamment pour les consignes de conduite en temps réel) mais ils ne font pas l'objet de ce manuscrit.

2.3 Planification à court terme de la production hydroélectrique du Rhône

Dans le cadre de notre étude, nous nous intéressons principalement à la planification à court terme de la production hydroélectrique du Rhône (voir Sous-section 1.5). Il s'agit d'une des étapes de la gestion opérationnelle effectuée quotidiennement depuis le siège social de CNR à Lyon. Cette étape consiste à définir, en avance, l'utilisation de la modulation de la production du Rhône (i.e. les variables de décision du modèle de simulation, voir Sous-section 2.2.4) au niveau de chaque aménagement hydroélectrique jusqu'à un horizon futur de quelques jours. Les objectifs sont de respecter les contraintes d'exploitation et de maximiser le chiffre d'affaires obtenu par la vente de la production sur les marchés à court terme (marché *day-ahead* principalement, voir Sous-section 1.2.2). Le programme de production ainsi construit sert de référence pour les étapes ultérieures de la gestion opérationnelle, notamment les ventes de la production à court terme sur les marchés et la conduite en temps réel des aménagements effectuée à distance (téléconduite) depuis le siège social de CNR à Lyon [6]. La construction du programme de production à court terme est effectuée à partir du modèle de simulation de la Sous-section 2.2.4.

On suppose que les valeurs des paramètres (en particulier les prévisions des apports en eau et des prix) qui interviennent pour la planification de la production sont à disposition.

2.3.1 Programme de production

2.3.1.1 Définitions

Un *programme de production* regroupe par définition les valeurs des variables (de décision) du modèle de simulation du Rhône (voir Section 2.2) sur une fenêtre temporelle future, discrétisée en T pas de temps de même durée δt correspondant à la résolution temporelle du modèle de simulation ($T \in \mathbb{N}^*$). Le pas de temps final d'indice T est l'*horizon* de la planification, qui correspond à l'horizon caractéristique de planification de la sous-section 1.3.1. La *fenêtre temporelle* (future) est la période entre l'instant présent et l'horizon, représentée par les pas de temps d'indices $t \in \mathcal{T} = \{1, \dots, T\}$.

On dit qu'un programme de production est *applicable* (ou *réalisable*) si, pour le modèle de simulation, les équations du modèle hydraulique et les contraintes d'exploitation sont respectées sur toute la fenêtre temporelle (pour une résolution temporelle δt).

Une équipe d'opérateurs de gestion de production travaille en opérationnel à CNR pour, notamment, la construction des programmes de production, en considérant le Rhône dans son ensemble.

6. Les exploitants des aménagements peuvent néanmoins prendre localement la main sur la conduite d'un aménagement en cas d'incident. Les aménagements disposent également d'automates qui permettent de réguler les consignes de la téléconduite et de prendre la main sur les consignes si celles-ci ne permettent pas de respecter les contraintes d'exploitation.

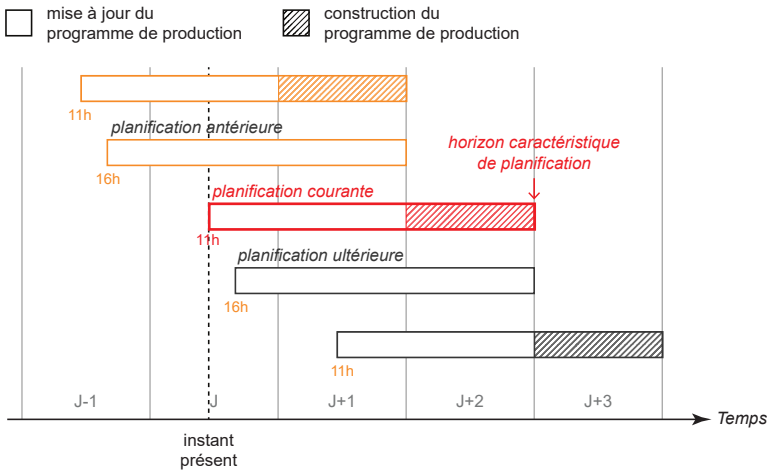


FIGURE 2.9 – Schéma de l'enchaînement des fenêtres temporelles des planifications au cours du temps dans le cas où deux planifications sont réalisées quotidiennement, une à 11 h et l'autre à 16 h, avec un horizon caractéristique de planification correspondant à la fin du sur-lendemain.

2.3.1.2 Horizons

Un premier programme de production de l'aménagement hydroélectrique de tête de Génissiat seul (ou avec l'aménagement aval de Seyssel) est construit, au moins chaque début de semaine, à un horizon hebdomadaire (moyen terme) car le volume de retenue de l'aménagement de Génissiat permet des modulations à l'échelle de plusieurs jours (aménagement de moyenne chute). Cela permet notamment de définir une cote cible à Génissiat en fin de chaque journée de la semaine.

Le programme de production de l'ensemble de la chaîne hydroélectrique (Génissiat compris) est construit à un horizon de 2 à 3 jours (court terme) et mis à jour 4 à 5 fois par jour à des instants réguliers. Ces instants sont définis pour que le processus opérationnel soit en phase avec les horaires des marchés de l'électricité et des déclarations de production à envoyer au gestionnaire de Réseau Transport d'Électricité (RTE), voir Section 1.2. Pour Génissiat, le programme de production à court terme permet d'affiner le programme de production à moyen terme, en gardant les cotes cibles en fin de journées comme objectif de gestion.

Dans tout le manuscrit, on considère uniquement l'horizon court terme (voir Sous-section 1.5) : l'indice T du pas de temps final correspond à une échéance de 2 à 3 jours.

2.3.1.3 Situation

L'ensemble des informations dont on dispose à un instant présent donné permet de définir la *situation*. Autrement dit, une situation regroupe :

— l'état du fleuve mesuré (observé) jusqu'à l'instant présent considéré,

— les valeurs de tous les paramètres du modèle de simulation nécessaires à la définition du programme de production à l'horizon considéré (jeux de contraintes d'exploitation, débits d'apports prévus des affluents, prévision des prix de l'électricité, ...).

À chaque planification de la production (i.e. mise à jour ou construction du programme de production) correspond une situation unique. La durée qui sépare deux planifications différentes est de quelques heures (typiquement 4 h en journée), donc petite par rapport à l'horizon. Une partie du programme de production est donc retravaillée (mise à jour) à chaque planification, comme l'illustre les parties non hachurées de la figure 2.9. Il y a donc une dépendance entre les planifications successives.

2.3.2 Placement de l'énergie

La grande majorité de la production CNR (environ 80%) est vendue à long terme de gré à gré (voir Sous-section 1.2.2). Les produits de ces ventes sont définis sur des périodes de plusieurs heures, jours ou mois. Avec cette granularité, les contrats à long terme ne tiennent pas compte de la modulation de la production du Rhône (qui s'effectue à l'échelle de quelques heures). Le volume de production engagé par ces contrats repose sur des tendances (scénarios) hydro-météorologiques à long terme en lien avec l'expérience passée et une gestion des risques par CNR (voir Sous-section 1.3.2).

L'autre partie de la production CNR (environ 20%) est vendue à court terme, de gré à gré ou aux marchés *day-ahead* et infrajournalier (voir Sous-section 1.2.2). Comme la modulation de la production du Rhône est adaptée à l'horizon court terme, les ventes à court terme et le programme de production doivent être définis conjointement (troisième approche de la gestion à court terme présentée à la sous-section 1.3.2). Nous nous intéressons alors principalement à la planification de la production pour respecter les contraintes d'exploitation et maximiser le chiffre d'affaires issu de la vente de la production à court terme (voir Sous-section 1.5).

Dans la suite du manuscrit, on considère que le programme de production à court terme n'est construit qu'en anticipant les ventes au marché *day-ahead* et les pénalités des écarts (voir Section 1.2). On ne considère donc pas les éventuelles ventes de gré à gré à court terme, que l'on suppose effectuées à la place de certaines ventes *day-ahead* si les contrats sont à des prix plus intéressants *a priori* que ceux du marché *day-ahead*. On ne considère pas non plus les transactions au marché infrajournalier, que l'on suppose être utilisées uniquement pour rattraper, si possible, une partie des écarts qui peuvent être anticipés à très court terme (gestion presque en temps réel), voir Sous-section 1.3.2.

La planification de la production à court terme est par définition la démarche qui consiste, pour une situation donnée (voir Figure 2.9), à :

— mettre à jour le programme de production sur la partie de la fenêtre temporelle située avant l'horizon caractéristique de la planification antérieure,

— construire le programme de production sur la partie de la fenêtre temporelle ultérieure à l'horizon caractéristique de la planification antérieure.

Les enjeux du programme de production à court terme sont doubles (voir Sous-section 1.3.2) :

— respecter les contraintes d'exploitation en priorité (le programme de production doit toujours être applicable),

— maximiser le chiffre d'affaires obtenu par la vente de la production sur le marché *day-ahead*, en déplaçant la production des heures à bas prix vers les heures à forts prix au sein d'une même journée et entre les différents aménagements.

En plus de l'intérêt financier pour CNR, la maximisation du chiffre d'affaires à court terme a également deux atouts (voir §1.2.2.2) :

— elle permet de mettre en adéquation le programme de production avec la demande du réseau car les faibles prix correspondent généralement aux heures de faible consommation et les forts prix aux heures de forte consommation,

— elle permet de substituer une énergie renouvelable à une énergie plus couteuse et souvent fortement carbonée (production de pointe au gaz, charbon, fioul) car les prix de l'électricité sont définis à partir des coûts marginaux des actifs de production mis en œuvre pour répondre à la demande du réseau.

2.3.3 Calcul du chiffre d'affaires

Le chiffre d'affaires associé à un programme de production est un indicateur d'optimalité pour la planification de la production. On ne considère ici que le chiffre d'affaires qui est directement lié à la modulation de la production de la chaîne hydroélectrique. Le chiffre d'affaire d'un programme de production est alors calculé en considérant les ventes au marché *day-ahead* et les pénalités des écarts (voir Sous-section 2.3.2) sur sa fenêtre temporelle $\mathcal{T} = \{1, \dots, T\}$:

$$\mathrm{CA}(\mathcal{T}) = \mathrm{CA}_{\mathrm{DA}}(\mathcal{T}) + \mathrm{CA}_{\text{écarts}}(\mathcal{T}), \tag{2.8}$$

où $\mathrm{CA}(\mathcal{T}) \in \mathbb{R}$ désigne le chiffre d'affaires du programme de production à court terme sur la fenêtre temporelle $\mathcal{T}$, $\mathrm{CA}_{\mathrm{DA}}(\mathcal{T}) \in \mathbb{R}$ la partie du chiffre d'affaires obtenue par les ventes au marché *day-ahead* et $\mathrm{CA}_{\text{écarts}}(\mathcal{T}) \in \mathbb{R}$ la partie du chiffre d'affaires due aux pénalités des écarts.

Pour obtenir le chiffre d'affaires total, il faudrait ajouter le chiffre d'affaires issu de toutes les autres transactions, notamment celles liées aux engagements connus à l'instant présent. Ce terme additionnel n'a cependant pas d'importance pour la planification de la production car il ne dépend pas du programme de production à court terme.

2.3.3.1 Marché *day-ahead*

Le chiffre d'affaires obtenu par les ventes au marché *day-ahead* sur la fenêtre temporelle $\mathcal{T} = \{1, \dots, T\}$ d'un programme de production est défini par :

$$\mathrm{CA}_{\mathrm{DA}}(\mathcal{T}) = \sum_{t \in \mathcal{T}} \pi(t) \delta t P_{\mathrm{DA}}(t) = \delta t \sum_{t \in \mathcal{T}} \pi(t) P_{\mathrm{DA}}(t), \tag{2.9}$$

où $\pi(t) \in \mathbb{R}$ désigne le prix horaire *day-ahead* à l'instant $t \in \mathcal{T}$ et $P_{\mathrm{DA}}(t) \in \mathbb{R}$ désigne la part de la production (en puissance moyenne sur un pas de temps) du programme de production restant à vendre au marché *day-ahead* (après prise en compte des engagements déjà pris sur les marchés moyens et longs termes) pour livraison au pas de temps $t \in \mathcal{T}$. Ainsi, $P_{\mathrm{DA}}(t) = 0$ si les ventes *day-ahead* à livrer au pas de temps t ont déjà été effectuées. Par exemple, pour une planification effectuée à 11 h, la production de la fin du jour courant (de 11 h à 24 h) a déjà été valorisée la veille au marché *day-ahead* mais pas celle au-delà du jour courant (voir Figure 1.10).

La multiplication par δt permet d'obtenir des énergies : $\delta t P_{\mathrm{DA}}(t)$ correspond à l'énergie, sur la durée du pas de temps $t \in \mathcal{T}$, à vendre au marché *day-ahead* (les produits échangés sur les marchés sont des énergies). Comme le marché *day-ahead* repose sur des contrats horaires, les fonctions π et P_{DA} sont constantes sur chaque heure des journées de la fenêtre temporelle (granularité horaire).

2.3.3.2 Pénalisation des écarts

Les écarts (prévus) du programme de production sont définis par :

$$\forall t \in \mathcal{T}, \quad e(t) = P(t) - (c(t) + P_{\mathrm{DA}}(t)), \tag{2.10}$$

où $e(t) \in \mathbb{R}$ désigne les écarts au pas de temps $t \in \mathcal{T}$, $P(t) \in \mathbb{R}_+$ la puissance totale produite dans le périmètre d'équilibre au pas de temps $t \in \mathcal{T}$, et $c(t) \in \mathbb{R}$ les engagements[7] sur le périmètre d'équilibre à livrer au pas de temps $t \in \mathcal{T}$ et qui sont connus à l'instant présent (issus des contrats à long terme et des dernières sessions du marché *day-ahead*). Si seule la chaîne hydroélectrique entre dans le périmètre d'équilibre, alors :

$$\forall t \in \mathcal{T},\ P(t) = P_{\mathrm{hydro}}(t) \quad \text{et} \quad \forall t \in \mathcal{T},\ c(t) = c_{\mathrm{hydro}}(t),$$

7. La notation c fait référence au terme anglais *commitments* pour désigner les engagements.

où $P_{\text{hydro}}(t) \in \mathbb{R}_+$ désigne la puissance produite par le Rhône (voir Équation (2.7)) et $c_{\text{hydro}}(t) \in \mathbb{R}$ les engagements associés à la production du Rhône au pas de temps $t \in \mathcal{T}$.

La fonction P représente la production potentielle prévue à partir du programme de production, tandis que la fonction $c + P_{\text{DA}}$ représente les engagements potentiels prévus à partir des engagements déjà pris et des ventes *day-ahead* qu'il reste à effectuer à partir du programme de production (voir Sous-section 1.3.2). La fonction $e = P - (c + P_{\text{DA}})$ représente donc les écarts potentiels prévus associés au programme de production (voir Sous-section 1.2.4). En général, la fonction c est constante sur chaque heure des journées de la fenêtre temporelle (granularité horaire), tout comme la fonction P_{DA} (voir §2.3.3.1), mais pas la fonction P *a priori*. Si les fonctions P et c sont constantes sur chaque heure des journées de la fenêtre temporelle, alors, avec la gestion opérationnelle décrite à la section 1.3, la fonction e est nulle sur la partie de la fenêtre temporelle pour laquelle les transactions *day-ahead* n'ont pas encore été effectuées (à partir du lendemain avec l'exemple de la figure 1.10).

Les écarts positifs (respectivement négatifs) du programme de production sont alors obtenus en prenant la partie positive (respectivement négative) des écarts à chaque demi-heure de la fenêtre temporelle :

$$\forall t \in \mathcal{T}, \quad \begin{cases} e^+(t) = \max(0, e_m(t)), \\ e^-(t) = \max(0, -e_m(t)), \end{cases} \tag{2.11}$$

où $e_m(t) \in \mathbb{R}$ désigne les écarts à la granularité demi-horaire au pas de temps $t \in \mathcal{T}$. La fonction e_m s'obtient à partir de la fonction e en calculant les écarts moyens sur chaque demi-heure. Par exemple, si $\delta t = 10\,\text{min}$ et si l'instant présent $t = 0$ correspond à une heure ronde, alors

$$\forall t \in \mathcal{T}, \quad e_m(t) = \begin{cases} (e(t) + e(t+1) + e(t+2))/3 & \text{si } t-1 \in 3\mathbb{Z}, \\ (e(t-1) + e(t) + e(t+1))/3 & \text{si } t-2 \in 3\mathbb{Z}, \\ (e(t-2) + e(t-1) + e(t))/3 & \text{si } t-3 \in 3\mathbb{Z}. \end{cases}$$

Le chiffre d'affaires dû à la pénalisation des écarts du programme de production est défini par :

$$\text{CA}_{\text{écarts}}(\mathcal{T}) = \delta t \sum_{t \in \mathcal{T}} \pi^+(t) e^+(t) - \delta t \sum_{t \in \mathcal{T}} \pi^-(t) e^-(t), \tag{2.12}$$

où $\pi^+(t) \in \mathbb{R}$ et $\pi^-(t) \in \mathbb{R}$ désignent respectivement les prix de règlement des écarts positifs et négatifs au pas de temps $t \in \mathcal{T}$. Les fonctions π^+, π^-, e^+ et e^- sont constantes sur chaque demi-heure des journées de la fenêtre temporelle. La multiplication par δt permet d'exprimer les écarts en termes d'énergie.

2.3.4 Prévisions

Un programme de production porte sur une fenêtre temporelle future. En conséquence, les valeurs de certains paramètres du modèle de simulation ne sont pas parfaitement connues à l'instant présent où l'on effectue la planification de la production. Il est donc nécessaire de prévoir ces valeurs avant d'utiliser le modèle de simulation (voir Sous-section 1.3.1).

2.3.4.1 Apports en eau des affluents et des bassins versants intermédiaires

Les apports en eau des affluents jouent un rôle majeur dans le modèle hydraulique. Ils ne sont cependant pas parfaitement connus à l'instant présent car ils dépendent des conditions hydrométéorologiques et, pour certains affluents, des opérations sur les ouvrages hydroélectriques exploités par EDF. C'est pourquoi une équipe de prévisionnistes (météorologues, hydrologues et hydrauliciens) travaille à CNR pour produire des prévisions expertisées des apports en eau

qui alimentent le Rhône. Cela permet de définir les paramètres Q_{aff} et Q_{BVI} de l'équation (2.4). Le travail des prévisionnistes, effectué plusieurs fois pas jour en opérationnel, est décomposé en deux étapes :

1. La première étape consiste à créer un scénario de précipitations liquides et de fonte nivale sur l'ensemble du bassin versant du Rhône à un horizon de plusieurs jours (typiquement 4 jours). Le scénario de précipitations est construit par expertise des prévisionnistes à partir d'observations (réseau pluviométrique de Météo-France, outil interne basé sur les données des radars météorologiques, ...) et de prévisions (modèles des grands centres météorologiques, modèles internes de fonte nivale et de prévision par analogie).
2. La seconde étape consiste à créer un scénario hydrologique (débits des affluents naturels et des bassins versants intermédiaires) à un horizon de plusieurs jours (typiquement 4 jours) avec une granularité horaire, voire demi-horaire. Le scénario hydrologique est construit par expertise des prévisionnistes à partir de modèles hydrologiques (modèles pluie-débit forcés par les scénarios de précipitation) et hydrauliques (propagation des débits) développés par CNR [Bompart *et al.*, 2009 ; Celie *et al.*, 2019]. Ces modèles s'appuient sur un large réseau de stations de mesure, réparti sur l'ensemble du bassin versant du Rhône, qui permet la mesure en temps réel de l'état hydro-météorologique du Rhône.

Une prévision des débits des affluents influencés par des aménagements hydroélectriques d'EDF jusqu'à la fin du lendemain est reçue quotidiennement à 12h30 par CNR. Ces prévisions successives ne couvrent qu'une partie de la fenêtre temporelle, surtout si l'instant présent est avant 12h30. C'est pourquoi CNR dispose de modèles de prévision spécifiques développés en interne, complétés par expertise des prévisionnistes, pour créer les prévisions des débits des affluents influencés.

2.3.4.2 Débits de correction

Le modèle hydraulique du Rhône n'est pas parfait. La chaîne hydroélectrique du Rhône n'est pas strictement un système de réservoirs en cascade, certaines équations reposent sur des hypothèses simplifiées (par exemple, pour une situation donnée, les temps de propagation et de retard sont supposés constants) et certains comportements du Rhône ne sont pas mis en équations ou ne peuvent pas l'être. En outre, les mesures des débits ne sont pas parfaites donc elles ne donnent pas les véritables valeurs des débits (les erreurs de mesure sont de l'ordre de 5% à 10%). De même, les prévisions des apports des bassins versants intermédiaires ne sont pas parfaitement justes et il est difficile de les confronter aux réalisations car, par définition, ces réalisations ne sont pas directement mesurées. Pour pallier ces erreurs dans l'écriture des bilans d'eau (voir Équation (2.1)), des débits de correction sont ajoutés aux apports en eau (voir Équation (2.4)).

Pour une situation donnée, les débits de correction aux aménagements sont calibrés sur des plages passées couvrant la veille et le jour courant. La calibration consiste à définir, pour chaque aménagement, les débits de correction constants sur chaque plage passée qui imposent l'égalité des cotes calculées et celles mesurées en début et fin des plages passées :

$$\forall (i,t) \in \mathcal{I} \times [t_1, t_2], \quad Q_{\text{corr}}(i,t) = \frac{(\tilde{V} - \hat{V})(i,t_2) - (\tilde{V} - \hat{V})(i,t_1)}{t_2 - t_1}, \tag{2.13}$$

où $[t_1, t_2]$ est une plage de temps passée, $\tilde{V}(i,t) \in \mathbb{R}_+$ sont les volumes déduits des mesures de cotes et $\hat{V}(i,t) \in \mathbb{R}_+$ les volumes obtenus par le calcul du modèle hydraulique forcé par les débits mesurés (sans prise en compte de débits de correction).

Les débits de correction à chaque aménagement sur la fenêtre temporelle future sont estimés en prolongeant les valeurs calibrées sur les plages passées. Les opérateurs de gestion de production peuvent, si besoin, modifier les plages de calibrage ou surcharger les prévisions des débits de correction en fonction de leur expertise.

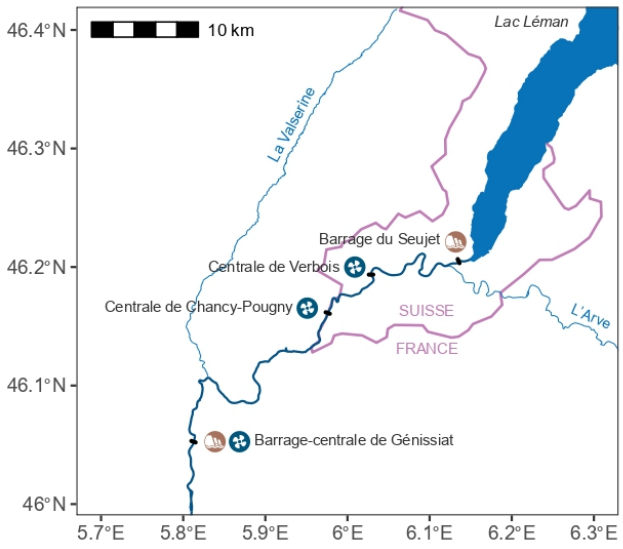


FIGURE 2.10 – Carte du Rhône en amont de l'aménagement de Génissiat.

2.3.4.3 Débit entrant à Génissiat

Trois ouvrages hydroélectriques sont installés sur le Rhône entre le lac Léman et le barrage-centrale de Génissiat (voir Figure 2.10) :

1. le barrage du Seujet (à Genève), qui est l'exutoire du lac Léman,
2. la centrale hydroélectrique de Verbois (en Suisse), en aval du barrage du Seujet et du confluent de l'Arve, sans canal de dérivation, qui fonctionne parfaitement au fil de l'eau (pas de modulation possible du débit sortant par rapport au débit entrant),
3. la centrale hydroélectrique de Chancy-Pougny, en aval de celle de Verbois à cheval entre la Suisse et la France, sans canal de dérivation, qui fonctionne parfaitement au fil de l'eau (pas de modulation possible du débit sortant par rapport au débit entrant).

Le barrage du Seujet et la centrale hydroélectrique de Verbois sont exploités conjointement par les Services Industriels de Genève (SIG). Le débit au barrage du Seujet est modulé en fonction du débit de l'Arve afin de gérer le débit entrant à la centrale hydroélectrique de Verbois. Comme la centrale hydroélectrique de Verbois est parfaitement au fil de l'eau, cette modulation permet de gérer directement le débit sortant et donc la production hydroélectrique. La centrale hydroélectrique de Chancy-Pougny est exploitée par la Société des Forces Motrices de Chancy-Pougny (dont les SIG et CNR sont actionnaires). Comme elle est parfaitement au fil de l'eau, elle n'influence pas les débits du Rhône.

Une prévision des débits au barrage du Seujet et à la centrale-hydroélectrique de Verbois (au pas horaire jusqu'à la fin du lendemain, au pas de 6 h jusqu'à une échéance de 3 jours) est reçue quotidiennement à 8h30 par CNR. Comme la centrale hydroélectrique de Chancy-Pougny est au fil de l'eau, cette prévision permet de calculer le débit entrant à l'aménagement de Génissiat en propageant le débit qui traverse la centrale hydroélectrique de Verbois et en y ajoutant les prévisions du débit de la Valserine (affluent qui conflue avec le Rhône entre la centrale hydroélectrique de Chancy-Pougny et l'aménagement de Génissiat), des bassins versants intermédiaires et des débits de correction (voir Équation (2.3)).

2.3.4.4 Prix de l'électricité

Le calcul du chiffre d'affaires fait intervenir les prix de l'électricité pour mesurer l'optimalité économique d'un programme de production (voir Sous-section 2.3.2). Ainsi, bien que les prix

n'interviennent pas directement dans le modèle de simulation, ils sont un paramètre nécessaire à la planification du programme de production.

À court terme, les prix de l'électricité ne sont pas connus sur toute la fenêtre temporelle. Par exemple, les prix horaires au marché *day-ahead* du lendemain sont publiés quotidiennement vers 12h30, juste après avoir proposé les offres avant 12 h (voir §1.2.2.2). Si l'instant présent est avant 12h30, alors les prix au marché *day-ahead* du lendemain et des jours suivants ne sont pas connus. Si l'instant présent est après 12h30, alors les prix au marché *day-ahead* du lendemain sont connus mais pas ceux des jours suivants.

Une équipe d'opérateurs marché (*traders*) travaille à CNR, essentiellement pour effectuer les transactions sur le marché de l'énergie et anticiper les prix sur les différents horizons du marché de l'énergie. Cette équipe fournit notamment tous les jours avant 10 h une prévision du prix horaire au marché *day-ahead* à un horizon de plus de 7 jours. Cela permet en particulier de définir le paramètre π de l'équation (2.9) sur la fenêtre temporelle. Les opérateurs marché utilisent pour cela leur expertise et des modèles de simulation de marché développés par CNR.

Dans le cadre de notre travail, nous ne considérons pas le marché infrajournalier. En revanche, nous considérons les prix de règlement des écarts que nous estimons par dégradation des prix *day-ahead* :

$$\forall t \in \mathcal{T}, \quad \begin{cases} \pi^+(t) = (1 - \kappa^+(t))\pi(t), \\ \pi^-(t) = (1 + \kappa^-(t))\pi(t), \end{cases} \tag{2.14}$$

où $\pi^+(t) \in \mathbb{R}$ et $\pi^-(t) \in \mathbb{R}$ désignent respectivement les prix de règlement des écarts positifs et négatifs au pas de temps $t \in \mathcal{T}$ de la fenêtre temporelle et $(\kappa^+(t), \kappa^-(t)) \in]0, 1[^2$ sont des paramètres renseignés par les opérateurs. Cette simplification est néanmoins discutable car les erreurs de prévision des prix de règlement des écarts que l'on obtient de cette manière peuvent être assez importantes. La méthodologie que nous présentons dans ce manuscrit peut être facilement adaptée à une évolution de la manière dont les prix de règlement des écarts sont prévus.

2.3.4.5 Autres prévisions et corrélations

D'autres paramètres sont également à prévoir. Par exemple, la disponibilité des groupes de production est planifiée à partir des opérations de maintenance prévues et des avaries connues à l'instant présent. Certains paramètres (temps de propagation et de retard, bornes minimales et maximales à ne pas dépasser etc.) sont généralement les mêmes que ceux de la planification antérieure car ils sont mis à jour uniquement quand cela est nécessaire (par exemple, recalage des temps de propagation grâce aux valeurs mesurées).

Par ailleurs, les différentes prévisions ne sont *a priori* pas indépendantes. L'annexe A.1 présente, par exemple, quelques résultats sur les corrélations entre les apports en eau du Rhône mesurés et les prix de l'électricité réalisés. Ces corrélations peuvent être prises en compte dans les prévisions (au moins en partie) grâce à l'enchaînement des étapes du processus opérationnel. En effet, les prévisions hydro-météorologiques sont généralement effectuées en premier, donc elles peuvent servir de données d'entrée pour effectuer les prévisions des prix.

2.3.5 Outil d'optimisation

2.3.5.1 Gestion semi-manuelle par éclusées énergétiques synchrones

Le défi de la planification du programme de production est de synchroniser les sur- et sous-productions des aménagements hydroélectriques pour que la production totale soit en phase avec les prix de l'électricité, tout en respectant les contraintes d'exploitation.

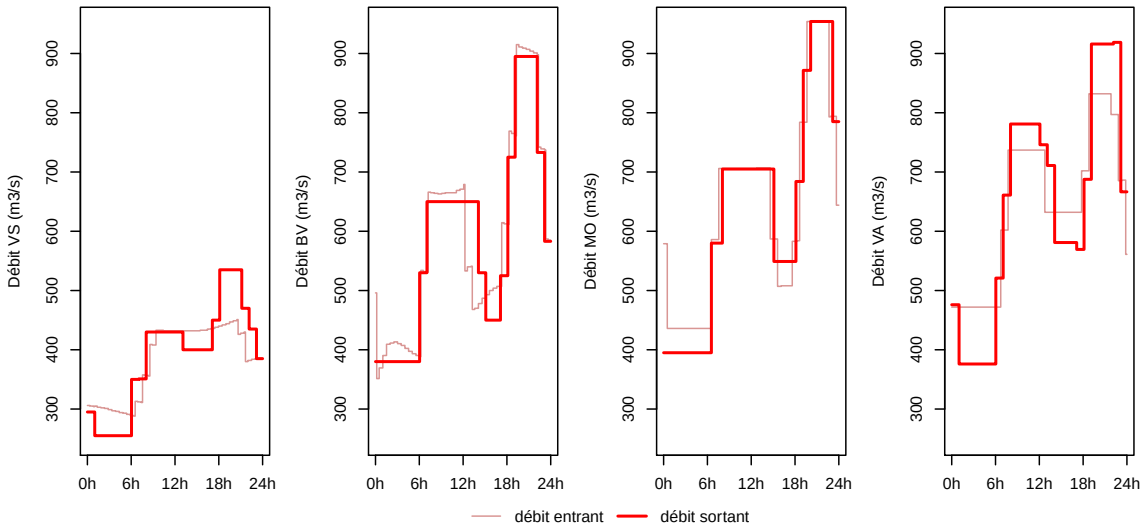


FIGURE 2.11 – Exemple d'éclusées sur 4 aménagements du Bas-Rhône (disposés de l'amont vers l'aval sur les graphiques de gauche à droite) sur une journée type.

L'approche historique employée par CNR consiste à créer des éclusées énergétiques synchrones [Piron *et al.*, 2015], i.e. des débits sortants en « créneaux » synchronisés entre aménagements. Une éclusée initiée à un aménagement se propage aux aménagements en aval et elle s'amplifie progressivement grâce aux stocks d'eau supplémentaires des retenues des aménagements en aval, comme l'illustre la figure 2.11. La production totale est alors maximale aux valeurs hautes et minimale aux valeurs basses des éclusées énergétiques synchrones. Généralement, il y a deux éclusées énergétiques par journée, dont les valeurs hautes sont centrées sur les pics de prix de l'électricité liés aux pics de consommation (vers 8-9 h le matin et 20-21 h le soir, voir Figures 1.4c et 1.7c). De cette manière, la gestion par éclusées énergétiques synchrones permet de moduler la production totale du Rhône pour maximiser le placement de la production sur les heures de forts prix, maximisant ainsi le chiffre d'affaires.

Si les temps de propagation hydrauliques entre les aménagements étaient nuls et si les réservoirs pouvaient amortir entièrement les apports en eau des affluents, alors il serait possible de synchroniser parfaitement les éclusées énergétiques des aménagements. Dans ce cas, la puissance totale de la chaîne hydroélectrique accumulerait les éclusées des débits sortants de chaque aménagement. Il suffirait alors de définir les périodes des éclusées en fonction des prix de l'électricité.

En pratique, les temps de propagation ne sont pas nuls et certains affluents apportent un débit important (par exemple, la Saône) par rapport à la capacité de stockage de l'aménagement en aval, rendant impossible la synchronisation parfaite des éclusées des aménagements. Si les éclusées sont globalement en opposition de phase, alors la courbe de la puissance totale de la chaîne hydroélectrique est globalement plate, ne permettant pas la maximisation du chiffre d'affaires. Pour conserver une synchronisation satisfaisante, l'éclusée à un aménagement est généralement initiée avant que celle de l'aménagement amont n'atteigne l'aménagement. Les éclusées sont alors définies avec des temps de propagation apparents inférieurs aux temps de propagation hydrauliques réels. Ce décalage est inférieur à la durée des éclusées pour permettre leur amplification à chaque aménagement. On forme alors des éclusées énergétiques *presque* synchrones. Une telle approche nécessite de bonnes prévisions des débits d'apports en eau et une bonne connaissance des temps de propagation hydrauliques réels et des marnages disponibles aux aménagements. Il faut en effet que les aménagements soient en mesure d'initier les éclusées (notamment en ayant des retenues suffisamment remplies) et de supporter l'arrivée de l'éclusée de l'aménagement en amont (sans saturer la centrale).

En pratique, les éclusées énergétiques ne sont pas synchronisées sur l'ensemble des aménagements du Rhône en raison du long tronçon sans aménagement entre l'aménagement de Sault-Brénaz et celui de Pierre-Bénite (voir Figure 2.4). Comme le temps de propagation hydraulique réel de ce tronçon est d'environ 6 heures, les modulations de production à l'aménagement de Sault-Brénaz arriveraient à l'aménagement de Pierre-Bénite en opposition de phase avec les prix. Pour éviter ce déphasage, des éclusées énergétiques synchrones sont réalisées indépendamment sur deux parties du Rhône : le Haut-Rhône (les 6 premiers aménagements, de Génissiat à Sault-Brénaz) et le Bas-Rhône (les 12 derniers aménagements, de Pierre-Bénite à Vallabrègues). Le profil du Haut-Rhône permet de créer des éclusées énergétiques synchrones, en phase avec les prix de l'électricité, sur les tous premiers aménagements et de les estomper progressivement avec les derniers aménagements du Haut-Rhône. Le débit sortant du Haut-Rhône est alors démodulé. Cette démodulation est nécessaire pour initier de nouvelles éclusées synchrones sur le Bas-Rhône en raison de la faible flexibilité des aménagements du Bas-Rhône (ils ne pourraient pas réaliser de nouvelles éclusées tout en supportant l'arrivée des éclusées du Haut-Rhône décalées dans le temps). La démodulation du débit sortant du Haut-Rhône permet également d'éviter des amplitudes de débits trop importantes par rapport aux contraintes d'exploitation sur le tronçon entre l'aménagement de Sault-Brénaz et celui de Pierre-Bénite. Les éclusées énergétiques du Bas-Rhône sont progressivement amplifiées d'un aménagement à l'autre, de manière presque synchrone (temps de propagation apparents inférieurs aux temps de propagation hydrauliques) et globalement en phase avec les prix de l'électricité.

L'optimisation par éclusées énergétiques synchrones est effectuée semi-manuellement à l'aide du modèle de simulation de la sous-section 2.2.4, en travaillant un à un sur chaque aménagement de l'amont vers l'aval.

2.3.5.2 Un optimiseur comme outil d'aide à la décision

L'approche d'optimisation semi-manuelle de la production du Rhône par éclusées synchrones tend à maximiser le chiffre d'affaires mais elle ne garantit pas une maximisation stricte du chiffre d'affaires. Autrement dit, le programme de production construit par éclusées synchrones peut être sous-optimal : il peut exister un autre programme de production applicable apportant un chiffre d'affaires plus élevé.

Or, les opérateurs de gestion de la production n'ont généralement pas la possibilité d'étudier manuellement d'autres schémas d'optimisation de la production. L'approche d'optimisation semi-manuelle par éclusées synchrones prend en effet une part importante du travail des opérateurs de gestion de la production, réduisant leurs disponibilités pour apporter leurs expertises, dans un contexte opérationnel en temps contraint.

Par ailleurs, l'approche d'optimisation semi-manuelle par éclusées synchrones est pertinente pour maximiser le chiffre d'affaires lorsqu'elle permet de valoriser la production sur le marché *day-ahead* avec la périodicité actuelle des prix *day-ahead* (pics en début de matinée et de soirée, creux la nuit). Elle risque donc de ne plus être adaptée aux évolutions du marché de l'énergie. Par exemple, CNR se positionne de plus en plus sur les marchés des services système, d'ajustement (voir Sous-section 1.2.3) et de capacité (voir Sous-section 1.2.6), en complément des contrats à long terme et des marchés *day-ahead* et infrajournalier (voir Sous-section 1.2.2). En conséquence, la gestion de la production du Rhône se complexifie et devient de plus en plus difficile à effectuer semi-manuellement par éclusées synchrones. De même, l'approche d'optimisation semi-manuelle par éclusées synchrones risque de ne plus être adaptée à la périodicité des prix *day-ahead* si celle-ci venait à changer du fait notamment des évolutions de la consommation électrique (due à l'électrification des usages nécessaire pour atteindre la neutralité carbone d'ici 2050 [RTE, 2021 ; ADEME, 2015]).

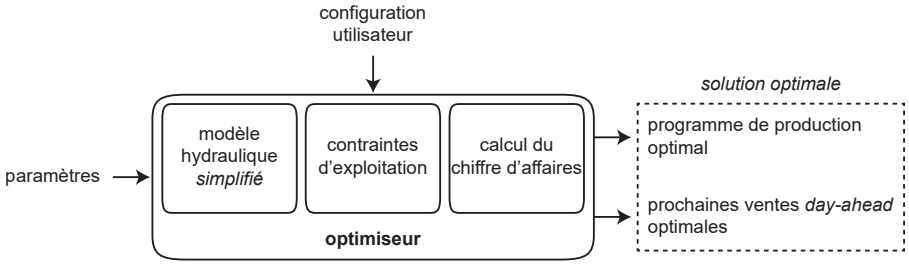


FIGURE 2.12 – Schéma général de l'optimiseur qui repose sur le modèle de simulation du Rhône.

Pour ces raisons, un outil d'aide à la décision a été développé en interne par CNR. Il s'agit d'un optimiseur qui résout numériquement le problème de planification de la production du Rhône, en s'appuyant sur les équations et contraintes du modèle de simulation (voir Sous-section 2.2.4) et le calcul du chiffre d'affaires (voir Sous-section 2.3.3). La figure 2.12, qui illustre le principe général de l'optimiseur, est à comparer avec la figure 2.8 concernant le modèle de simulation. En plus de libérer du temps d'expertise aux opérateurs pendant les calculs, l'optimiseur permet de chercher le programme de production optimal au-delà de ceux qui peuvent être élaborés par les schémas d'éclusées synchrones classiques. Il permet également de proposer les prochaines ventes *day-ahead* qui, associées au programme de production optimal, permettent de maximiser le chiffre d'affaires. L'optimiseur peut facilement être adapté en fonction des évolutions du marché de l'énergie (nouvelles sources de valorisation et possibles changements sur l'évolution temporelle des prix *day-ahead*). En outre, l'optimiseur est plus facile à modifier selon les évolutions des contraintes d'exploitation que l'approche d'optimisation semi-manuelle par éclusées synchrones.

L'optimiseur est utilisé comme un outil d'aide à la décision par les opérateurs de gestion de la production. L'expertise des opérateurs reste indispensable pour paramétrer l'optimiseur à chaque nouvelle planification puis pour retravailler et valider le programme de production optimal renvoyé par l'optimiseur. Les ajustements du programme de production par les opérateurs ont certes un impact sur l'optimalité du point de vue de l'optimiseur, mais ils sont nécessaires pour maîtriser les simplifications utilisées dans la définition de l'optimiseur.

Le temps de calcul de l'optimiseur est très variable selon la situation considérée. Dans certaines situations complexes, il n'est pas certain que l'optimiseur renvoie une solution applicable dans les délais opérationnels (moins d'une heure), et encore moins une solution vraiment optimale en chiffre d'affaires. Par ailleurs, il est difficile de caractériser ces situations complexes, donc il est difficile d'estimer *ex ante* le temps de calcul nécessaire à l'obtention d'une solution optimale pour une situation donnée. C'est pourquoi les opérateurs construisent encore semi-manuellement le programme de production le temps que l'optimiseur fournisse (ou non) une solution optimale. Des travaux sont toujours en cours sur l'optimiseur, notamment pour améliorer les temps de calcul.

En outre, le programme de production ne peut pas être changé en pratique sur la première heure de la fenêtre temporelle en raison du délai nécessaire pour construire le programme de production et pour prendre en compte d'éventuelles modifications dans les consignes de conduite des aménagements. Sur la première heure de la fenêtre temporelle, on impose donc les valeurs des variables de la dernière consigne de conduite à très court terme [8]. Pour l'optimiseur, cela revient à faire commencer la fenêtre temporelle une heure après l'instant présent, et à initialiser les variables sur l'heure entre l'instant présent et le début de la fenêtre temporelle avec la dernière

8. La dernière consigne de conduite des aménagements est issue de la planification antérieure de la production, mais en la confrontant aux valeurs des paramètres (notamment des apports en eau) qui se réalisent en temps réel. Le programme de production de la consigne de conduite des aménagements diverge donc légèrement et progressivement entre la planification antérieure et la planification courante.

consigne de conduite des aménagements.

Le cadre de notre travail repose sur cet optimiseur. En particulier, la définition mathématique du problème d'optimisation sous-jacent et sa résolution numérique sont présentées dans le chapitre 3.

2.4 Vers une gestion du Rhône en univers probabiliste

2.4.1 Impact des erreurs de prévision des apports en eau et des prix

La gestion actuelle du Rhône seul que nous venons de présenter est effectuée en univers déterministe car elle utilise des prévisions déterministes, i.e. des prévisions qui considèrent une seule valeur des paramètres. Elle repose en particulier sur une bonne précision des prévisions des débits d'apports des affluents [Bompart *et al.*, 2009] et des prévisions des prix de l'électricité (voir Sous-section 2.3.4).

Les apports en eau et les prix héritent cependant des incertitudes sur les conditions hydrométéorologiques et les conditions du marché de l'énergie qui peuvent se répercuter sur la planification de la production du Rhône.

— Les apports en eau interviennent dans le calcul des débits entrants (voir Équation (2.3)), donc, à débits sortants fixés, sur les cotes des retenues (d'après le bilan d'eau, voir Équation (2.1)). Or, les capacités de stockage des aménagements hydroélectriques du Rhône sont généralement faibles par rapport à la part des volumes d'eau entrants issus des apports en eau. Ainsi, le respect des cotes minimales et maximales des contraintes d'exploitation sont généralement sensibles aux incertitudes sur les apports en eau. Cette sensibilité aux incertitudes sur les apports en eau se répercute, de manière indirecte, sur le chiffre d'affaires. Le programme de production doit en effet être corrigé s'il s'avère qu'il ne respecte pas les contraintes d'exploitation avec les apports en eau qui se réalisent effectivement en temps réel. C'est une des raisons pour lesquelles le programme de production est mis à jour 4 à 5 fois par jour (voir §2.3.1.2), dont au moins 2 fois avec des nouvelles prévisions hydrologiques. Ces éventuelles corrections entraînent des écarts qui se répercutent sur le chiffre d'affaires aux prix de règlement des écarts (à la hausse ou à la baisse), voir Sous-section 1.4.1.

— Les prix de l'électricité interviennent dans le calcul du chiffre d'affaires. Plus précisément, le chiffre d'affaires est linéaire par rapport aux prix (voir Sous-section 2.3.3), donc les incertitudes sur les prix se propagent linéairement sur le chiffre d'affaires. Ainsi, la valeur économique d'un programme de production est généralement sensible aux incertitudes sur les prix. Il convient de noter que les incertitudes sur les variations temporelles des prix (décalage des pics dans le temps) impactent *a priori* davantage l'optimisation que les incertitudes sur les valeurs des prix (amplitudes des pics). Les prix sont en effet principalement utilisés dans l'optimisation pour placer la production dans le temps, une solution optimale pour une courbe de prix étant également optimale si on ajoute une même constante aux valeurs des prix (translation des valeurs des prix).

Ainsi, l'optimalité des programmes de production renvoyés par l'optimiseur déterministe du §2.3.5.2 peut être fortement remise en question du fait des incertitudes sur les apports en eau et des prix. L'expertise des prévisionnistes permet, en pratique, de donner des informations sur les incertitudes des prévisions aux opérateurs de gestion de la production pour la construction des programmes de production, mais l'optimiseur actuellement utilisé en opérationnel ne dispose pas de ces informations. Il semble donc nécessaire de prendre en compte les incertitudes sur les apports en eau et les prix dans la planification de la production.

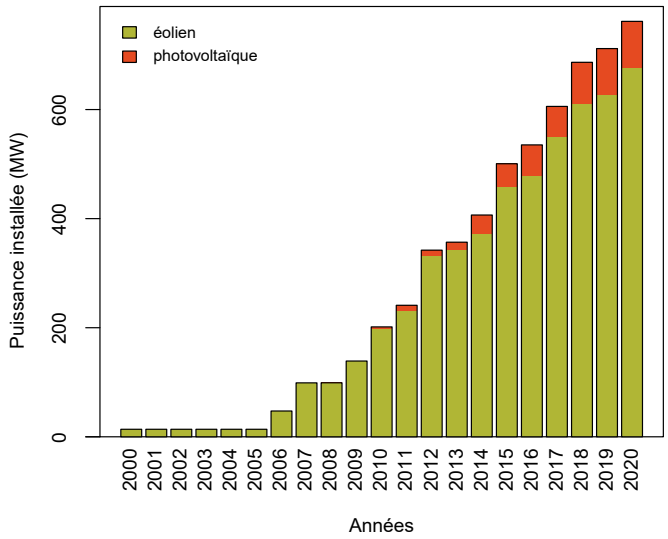


FIGURE 2.13 – Évolution annuelle de la puissance installée des actifs de production variable exploités par CNR.

La sensibilité de l'optimiseur face aux incertitudes sur les apports en eau et sur les prix est en revanche difficile à estimer car elle dépend *a priori* de la situation. Pour les exemples présentés dans [Vardanyan *et al.*, 2013], les incertitudes sur les prix semblent davantage impacter l'optimalité des programmes de production que celles sur les apports en eau. Néanmoins, les actifs hydroélectriques considérés dans [Vardanyan *et al.*, 2013] ont une plus grande flexibilité que ceux des aménagements hydroélectriques du Rhône. Or, plus la flexibilité de la production hydroélectrique est faible, moins le programme de production est *a priori* libre de suivre l'évolution des prix de l'électricité, donc moins les incertitudes sur les prix ont un impact sur l'optimisation. Même si une analyse de sensibilité globale pourrait être effectuée pour la planification du Rhône à différentes situations (voir Sous-section 1.4.4), nous considérons, par expertise, que les incertitudes sur les apports en eau et celles sur les prix ont un impact suffisamment important sur l'optimisation pour qu'il soit intéressant de les prendre en compte.

2.4.2 Gestion conjointe du Rhône et des actifs de production variable

2.4.2.1 Développement des énergies variables en complément du Rhône

En complément de la production du Rhône, CNR agrandit depuis le début des années 2000 sa production d'électricité renouvelable grâce à l'éolien et le solaire, comme l'illustre la figure 2.13. Cette démarche s'inscrit dans le développement de l'éolien et du solaire en France, voir Sous-section 1.1.1. Fin 2020, la puissance installée était d'environ 670 MW pour l'éolien et 110 MWc pour le solaire (voir la carte de la figure 2.14). Lorsque l'on ajoute la puissance installée des énergies variables à celle d'environ 3 100 MW de l'hydraulique (dont environ 3 000 MW sur le Rhône), cela fait de CNR le premier producteur français d'énergie exclusivement renouvelable.

Les premiers parcs éoliens et les premières centrales photovoltaïques exploités par CNR sortent progressivement du dispositif d'obligation d'achat (voir Sous-section 1.1.1) et entrent donc sur le marché dans le périmètre d'équilibre CNR. En outre, CNR renforce depuis 2016 son rôle d'agrégateur d'électricité renouvelable pour le compte de tiers (voir Sous-section 1.2.5). Le périmètre d'équilibre CNR est donc constitué de la chaîne hydroélectrique du Rhône et d'actifs de production variable (voir Figure 1.17). La production totale du périmètre d'équilibre s'écrit alors (voir §2.3.3.2) :

$$\forall t \in \mathbb{R}, \quad P(t) = P_{\text{hydro}}(t) + P_{\text{inter}}(t), \tag{2.15}$$

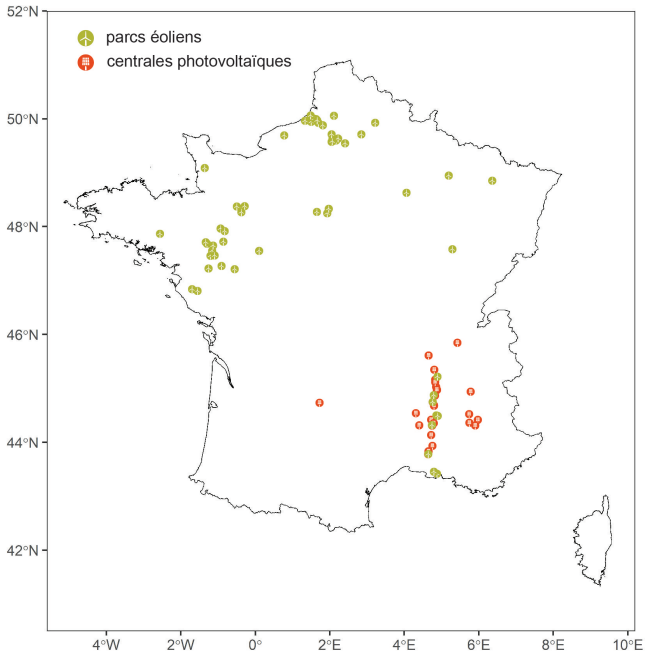


FIGURE 2.14 – Carte des parcs éoliens et des centrales photovoltaïques exploités par CNR (au 31/12/2020).

où $P_{\text{inter}}(t) \in \mathbb{R}_+$ désigne la somme des productions à l'instant t des parcs éoliens et centrales photovoltaïques du périmètre d'équilibre (hors dispositif d'obligation d'achat). De même, les engagements sur l'ensemble du périmètre d'équilibre s'écrivent, dans le cas où l'on peut distinguer les engagements associés au Rhône (voir §2.3.3.2) :

$$\forall t \in \mathbb{R}, \quad c(t) = c_{\text{hydro}}(t) + c_{\text{inter}}(t), \tag{2.16}$$

où $c_{\text{inter}}(t) \in \mathbb{R}$ désigne les engagements à l'instant t associés à la production des parcs éoliens et centrales photovoltaïques du périmètre d'équilibre (hors dispositif d'obligation d'achat).

2.4.2.2 Impact des erreurs de prévision des productions variables

Pour vendre une partie de la production totale sur le marché *day-ahead*, il faut prévoir la production potentielle de tous les actifs de production du périmètre d'équilibre (voir Sous-section 1.3.2). La production potentielle prévue du Rhône est obtenue à partir du programme de production (construit notamment grâce aux prévisions des apports en eau et des prix). Les productions variables, qui dépendent des conditions météorologiques (voir Sous-section 1.1.2), doivent être prévues par des modèles de prévision. C'est pourquoi l'équipe des prévisionnistes de CNR construit les prévisions expertisées des productions variables, en plus des prévisions hydrologiques expertisées du Rhône (voir Sous-section 2.3.4). Le processus de prévision consiste d'abord à construire une prévision météorologique expertisée (vent, rayonnement, température, nébulosité...) sur les zones concernées à un horizon de plusieurs jours (typiquement 4 jours), au pas horaire, à partir d'observations (mesures en temps réel) et de prévisions (modèles des grands centres météorologiques). Ils appliquent ensuite des modèles internes (modèles conceptuels ou modèles par apprentissage supervisé) pour convertir la prévision météorologique en prévision de facteur de charge pour chaque parc (voir Sous-section 1.1.2), puis en prévision de production. L'utilisation des facteurs de charge permet d'utiliser une même méthodologie de prévision pour des parcs ayant des caractéristiques différentes (technologies utilisées, nombre d'actifs disponibles...).

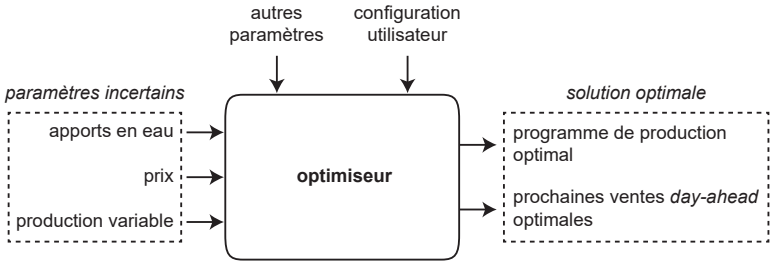


FIGURE 2.15 – Schéma général de l'optimiseur en distinguant les paramètres considérés comme incertains.

En général, les erreurs de prévision des productions variables (de l'ordre de 15% pour l'éolien et 25% pour le solaire à l'échelle d'un parc [Burtin et Silva, 2015]) sont plus élevées que les erreurs des prévisions hydrologiques (de l'ordre de 5% à l'échelle d'un affluent). Néanmoins, à l'échelle de tous les parcs, les erreurs de prévision des productions variables s'atténuent par foisonnement (voir Sous-section 1.2.4) car les parcs sont regroupés dans différentes zones géographiques soumis à des influences climatiques différentes. Tirer profit de ce foisonnement est un des enjeux de l'agrégation (voir Sous-section 1.2.5). Un tel foisonnement est beaucoup moins présent pour les prévisions hydrologiques en raison de la concomitance des débits des affluents. Cependant, les erreurs des prévisions hydrologiques sont atténuées grâce à la modulation des aménagements hydroélectriques.

Dans la planification de la production totale du périmètre d'équilibre, les productions variables interviennent dans la définition des écarts (voir Équation (2.10)). Malgré le foisonnement, les erreurs de prévision des productions variables peuvent donc être des sources d'écarts importants, donc de pénalités financières qui se répercutent sur le chiffre d'affaires (voir §2.3.3.2). La gestion conjointe du Rhône et des énergies variables consiste à construire le programme de production du Rhône en considérant la production totale de tous les actifs de production (voir Équation (2.15)). Ainsi, le cas échéant, une partie de la flexibilité du Rhône pourrait servir à compenser les écarts dus aux incertitudes de prévision des productions variables. Aujourd'hui entièrement dévolue au déplacement d'énergie, la flexibilité du Rhône pourrait être partagée entre compensation des écarts et déplacement d'énergie, en fonction des prix de règlement des écarts prévus et de la différence de prix prévu captée par le déplacement d'énergie.

Cette gestion conjointe peut être effectuée à court terme lors de la planification du programme de production, en adaptant les approches d'optimisation de la sous-section 2.3.5. En raison de la complexité accrue de l'optimisation, il semble plus opportun d'adapter l'optimiseur du §2.3.5.2 que l'appoche semi-manuelle du §2.3.5.1. L'optimiseur pour cette gestion conjointe se déduit de celui utilisé pour le Rhône seul en y ajoutant les productions variables dans le calcul de la puissance totale (voir Équation (2.15)) et en définissant les écarts par rapport à tous les engagements du périmètre d'équilibre (voir Équation (2.16)). Les valeurs de la prévision des productions variables à chaque pas de temps sont alors de nouveaux paramètres pour le problème d'optimisation (voir Figure 1.17).

Dans le cadre de notre étude, nous considérons l'optimiseur pour la gestion conjointe de la production du Rhône et des actifs de production variable du périmètre d'équilibre CNR. Nous considérons également, par expertise, que les incertitudes sur les productions variables ont un impact suffisamment important pour qu'il soit intéressant de les prendre en compte, tout comme celles sur les apports en eau et les prix (voir Sous-section 2.4.1). Le principe général de l'optimiseur pour la gestion conjointe, qui se déduit de celui pour le Rhône seul (voir Figure 2.12), est illustré à la figure 2.15.

2.4.2.3 Revue de la littérature

Dans la littérature sur la gestion conjointe des énergies renouvelables d'origine hydraulique, éolienne et solaire, les actifs hydroélectriques sont principalement des grands réservoirs ou des systèmes de pompage qui offrent une assez grande flexibilité [Jurasz *et al.*, 2018 ; Javed *et al.*, 2020]. Dans ce cas, la production totale (hydraulique, éolienne et solaire) est généralement d'abord vendue sur les marchés puis la grande flexibilité hydroélectrique à disposition est utilisée pour minimiser les écarts, notamment ceux dus aux énergies variables (voir Sous-section 1.3.2). Dans notre cas, la flexibilité de la chaîne hydroélectrique est faible (aménagements au fil de l'eau), donc les ventes de la production totale (hydraulique, éolienne et solaire) sur les marchés et le programme de production de la chaîne hydroélectrique sont décidés en même temps (voir Sous-section 1.3.2). Le problème d'optimisation avec une faible flexibilité hydroélectrique, utilisé dans ce manuscrit, est donc différent de celui avec une grande flexibilité, et dépasse, à notre connaissance, l'état de l'art.

2.4.3 Prise en compte des incertitudes dans l'optimisation

Les incertitudes sur les apports en eau, les prix et les productions variables peuvent être partiellement maîtrisées en univers déterministe, grâce aux mises à jour successives, plusieurs fois par jour, du programme de production du Rhône. Certaines de ces mises à jour quotidiennes du programme de production sont effectuées à partir de prévisions actualisées qui apportent des informations plus précises. Elles peuvent donc permettre d'optimiser le programme de production en fonction des prévisions de moins en moins incertaines à mesure que l'échéance se rapproche. La différence entre la production potentielle mise à jour et les engagements pris à partir des prévisions antérieures est une prévision des écarts sur la période infrajournalière (période où les engagements *day-ahead* ont déjà été pris), comme l'illustre la figure 2.16 sur un exemple fictif. Ces écarts prévus sur la période infrajournalière proviennent de toutes les sources d'incertitudes (voir Sous-section 1.4.1), notamment celles des prévisions que l'on souhaite prendre en compte. Ces écarts prévus sur la période infrajournalière peuvent donc en partie être valorisés grâce aux optimisations successives du programme de production, par exemple en les plaçant du mieux possible sur les demi-heures de la période infrajournalière où les prix de règlement des écarts prévus sont les moins pénalisants.

Les mises à jour successives du programme de production avec un optimiseur en univers déterministe permettent d'ajuster le programme de production face à une partie des incertitudes qui se dévoilent progressivement. Elles permettent de minimiser les pénalités des écarts sur la période infrajournalière où les ventes sur le marché *day-ahead* ont déjà été effectuées. En revanche, elles ne permettent pas de valoriser la production sur le marché *day-ahead* en tenant compte des toutes les incertitudes que l'on souhaite prendre en compte et sur toute la fenêtre temporelle (même la période au-delà de la période infrajournalière). Les décisions des ventes sur le marché *day-ahead* sont en effet proposées par l'optimiseur qui, en univers déterministe, ne dispose pas d'informations sur les incertitudes.

Pour valoriser complètement les incertitudes, il est donc indispensable de faire passer l'optimiseur de l'univers déterministe à l'univers probabiliste (voir Sous-section 1.4.2). Les informations sur les incertitudes que l'on souhaite prendre en compte peuvent alors être utilisées par l'optimiseur lors de la recherche de la solution optimale pour prévoir les possibles écarts et leurs répercutions sur le chiffre d'affaires (voir Sous-section 2.4.1 et §2.4.2.2). Les prochaines mises à jour possibles du programme de production peuvent alors être anticipées lors de la construction du programme de production et de la décision des prochaines ventes sur le marché *day-ahead*, permettant une meilleure maîtrise des incertitudes sur la période infrajournalière et une meilleure gestion des risques sur la période au-delà de la période infrajournalière. La figure 2.17 illustre cette démarche. Contrairement à la démarche en univers déterministe illustré à la figure 2.16 basé sur un seul

scénario (i.e. une unique combinaison des paramètres incertains), on dispose cette fois d'un nouveau scénario. Cela permet d'anticiper les possibles écarts dus aux futures corrections pour rendre applicable et optimal le programme de production avec le nouveau scénario en fixant les prochaines ventes *day-ahead*. Cette anticipation des écarts peut être réalisée en même temps que la construction du programme de production (optimisation dynamique [Bellman, 1966]). Cette démarche est présentée plus en détails dans le chapitre 3.

Dans la suite du manuscrit, nous nous intéressons au passage de l'univers déterministe à l'univers probabiliste de l'optimiseur de la gestion conjointe de la production du Rhône et des actifs de production variable du périmètre d'équilibre CNR.

2.4.4 Modélisation des incertitudes

Une étape indispensable pour faire passer l'optimiseur de l'univers déterministe à l'univers probabiliste est la quantification et la modélisation des incertitudes que l'on souhaite prendre en compte.

2.4.4.1 Sources d'incertitudes

Dans le cadre de la thèse, les incertitudes que nous souhaitons prendre en compte sont celles qui entachent les prévisions sur les apports en eau, les prix et les productions variables (voir Sous-section 2.4.1 et §2.4.2.2).

Nous ne souhaitons pas prendre en compte les incertitudes de modélisation décrites à la sous-section 1.4.1 parce qu'elles correspondent davantage à des erreurs et à des biais systématiques qu'à des incertitudes d'origine aléatoire, si bien que l'on peut essayer d'en diminuer l'impact par l'utilisation de termes correcteurs empiriques déterministes, comme les débits de correction (voir §2.3.4.2). En particulier, nous ne considérons pas les incertitudes sur les débits de correction.

De même, nous ne souhaitons pas prendre en compte les incertitudes concernant les incidents d'exploitation décrites à la sous-section 1.4.1 parce que la fréquence d'occurrence des événements d'exploitation est suffisamment faible pour que leur survenue puisse être traitée en opérationnel comme un événement fortuit.

2.4.4.2 Scénarios

Les apports en eau, les prix de l'électricité et les productions variables sont des paramètres incertains. Leurs valeurs qui seront effectivement observées sont donc issues d'une loi de probabilité jointe effective (continue) sur l'ensemble des réalisations possibles. Même si en pratique la connaissance exacte de cette loi n'est pas accessible, elle est l'expression la plus complète des incertitudes que l'on puisse avoir au moment où les prévisions sont calculées. Pour prendre en compte les incertitudes dans la gestion opérationnelle du Rhône, il faut donc estimer cette loi de probabilité jointe effective, qui est difficile à exprimer et à manipuler en pratique.

Pour notre travail, nous utilisons une approximation discrète par approche ensembliste qui consiste à générer un ensemble fini de scénarios multivariés qui définissent une loi de probabilité jointe discrète des incertitudes (voir §1.4.5.1). Un *scénario* est donc une réalisation possible de la combinaison des apports en eau, des prix de l'électricité et des productions variables. La loi de probabilité jointe discrète ainsi construite est une approximation de la loi de probabilité jointe effective : les dispersions entre les différents scénarios représentent les incertitudes que nous cherchons à modéliser.

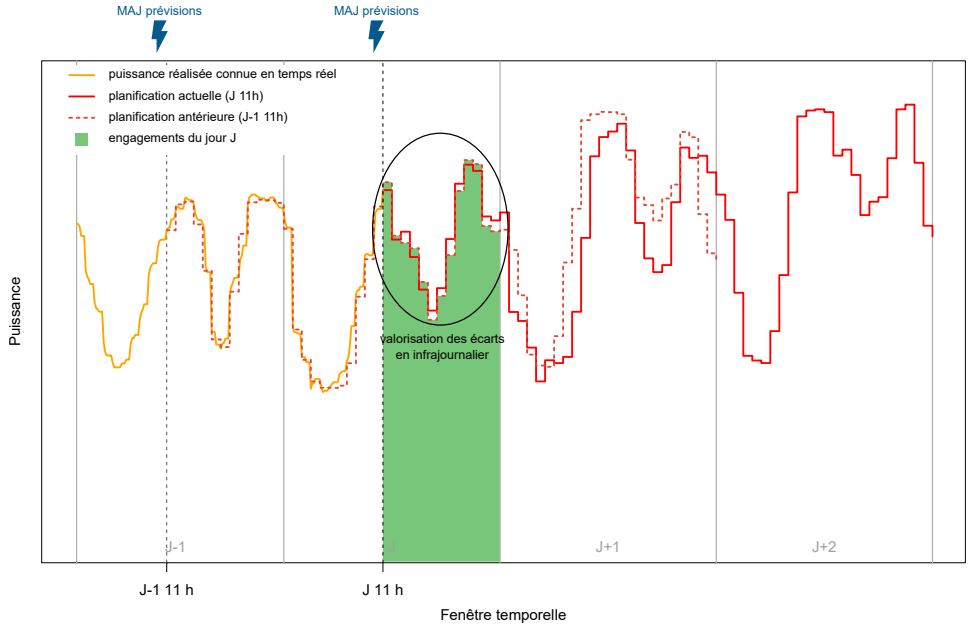


FIGURE 2.16 – Exemple de la valorisation des écarts sur la période infrajournalière pour le cas fictif où la planification est effectuée une fois par jour à 11 h avec une mise à jour des prévisions pour vendre toute la production au marché *day-ahead*. Les écarts infrajournaliers prévus sont la différence entre le programme de production de la planification actuelle (ligne rouge continue) et celui de la planification antérieure (ligne rouge en pointillés) qui a permis de décider les engagements du jour courant (en vert).

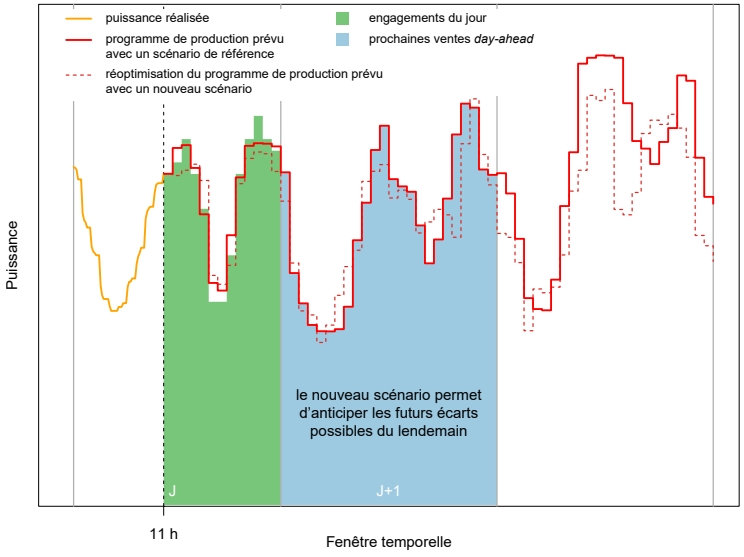


FIGURE 2.17 – Exemple de l'anticipation des écarts possibles en univers probabiliste pour le cas fictif où la planification est effectuée à 11 h pour vendre toute la production au marché *day-ahead*. Le programme de production (ligne rouge continue) est construit à partir d'un scénario de référence, i.e. une combinaison de valeurs des paramètres incertains. Les écarts sont anticipés grâce à la réoptimisation du programme de production avec un autre scénario (ligne rouge en pointillés) en fixant les prochaines ventes *day-ahead* (en bleu).

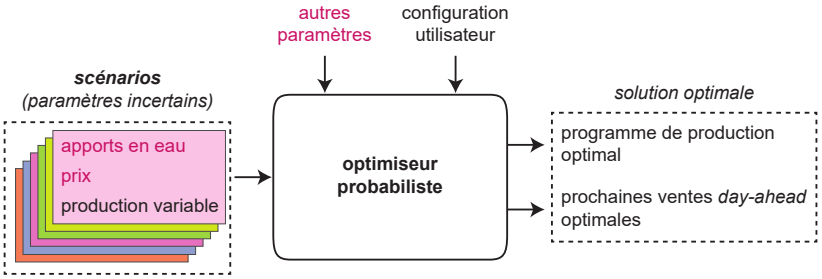


FIGURE 2.18 – Schéma général de l'optimiseur en univers probabiliste prenant plusieurs scénarios multivariés en entrée.

Dans le cadre de la thèse, les trois paramètres incertains sont supposés indépendants. La démarche utilisée consiste à générer des prévisions d'ensemble (voir §1.4.5.1) pour chacun des trois types de paramètres incertains pris individuellement. Les scénarios sont alors les combinaisons d'un membre de chacune des trois prévisions d'ensemble. De cette manière, on respecte la cohérence spatio-temporelle des apports en eau et des productions variables, et la cohérence temporelle des prix de l'électricité. En revanche, cette démarche ne permet pas de considérer directement les éventuelles corrélations entre les trois types de paramètres incertains, d'où l'importance de l'hypothèse d'indépendance pour pouvoir l'appliquer. En outre, les prévisions d'ensemble sont construites de manière à générer des membres équiprobables, donc l'hypothèse d'indépendance se traduit par des scénarios équiprobables. L'hypothèse d'indépendance est légitime pour une première approche car les corrélations entre les apports en eau, les prix et les productions variables (qui proviennent notamment de leur dépendance aux conditions météorologiques) sont assez faibles (voir Annexe A.1). Pour une modélisation plus avancée des incertitudes, les corrélations pourraient néanmoins être reconstruites à partir des trois prévisions d'ensemble en définissant intelligemment les probabilités des scénarios pour représenter au mieux la loi de probabilité jointe (par exemple, à partir d'une copule estimée empiriquement sur un historique passé [Camal *et al.*, 2019]).

Par ailleurs, nous cherchons à définir une méthodologie d'optimisation en univers probabiliste qui soit commune pour les apports en eau, les prix et les productions variables. Malgré l'hypothèse d'indépendance pour la génération des scénarios, nous ne souhaitons pas dissocier les trois types de paramètres incertains dans la formulation mathématique du problème d'optimisation en univers probabiliste présentée au chapitre 3 qui suit. Nous souhaitons considérer les trois types de paramètres incertains de manière couplée grâce aux scénarios multivariés. La raison principale de ce choix est que nous souhaitons prendre en compte les incertitudes des trois types de paramètres incertains pour anticiper leurs répercutions sur le chiffre d'affaires du fait des prochaines mises à jour du programme de production (voir Sous-section 2.4.3). En effet, une partie de toutes les incertitudes se dévoilent progressivement et en même temps à chaque planification successive. Un scénario multivarié est donc associé à une évolution possible de la situation (au sens du §2.3.1.3) sur laquelle les prochaines planifications reposeront. Le choix de combiner les trois types de paramètres incertains permet également d'étendre la méthodologie que nous présentons dans ce manuscrit au cas où l'hypothèse d'indépendance pour la génération des scénarios n'est plus considérée. La figure 2.18 illustre le principe général de l'optimiseur en univers probabiliste en considérant plusieurs scénarios multivariés des paramètres incertains (à comparer avec la figure 2.15).

À notre connaissance, le choix de combiner les différents types de paramètres incertains est peu présent dans la littérature. En général, les incertitudes des différents types de paramètres incertains sont modélisées et utilisées séparément et avec des approches différentes [Fleten et

Kristoffersen, 2008 ; De Ladurantaye *et al.*, 2009 ; Apostolopoulou *et al.*, 2018]. Cela permet, par exemple, d'utiliser les incertitudes sur les prix pour mesurer leurs répercutions sur le chiffre d'affaires (gestion des risques financiers) et celles sur les apports pour définir intelligemment des marges de sécurité sur les bornes minimales et maximales des cotes (gestion des risques sur la ressource en eau). Dans cet exemple, les incertitudes sur les apports en eau sont utilisées pour améliorer la robustesse de la solution, i.e. sa capacité à faire face aux incertitudes sur les apports, et non pour anticiper les répercutions sur le chiffre d'affaires comme ce que nous cherchons à faire dans notre étude.

Dans la suite du manuscrit, on appelle *scénario probabiliste* l'un des scénarios obtenus avec les prévisions d'ensemble et on appelle *scénario déterministe* le scénario obtenu en combinant les prévisions déterministes des apports en eau, des prix et des productions variables. Les notations sont présentées et utilisées à partir du chapitre 3 qui suit.

Bien que les scénarios multivariés ainsi construits peuvent être utilisées directement, il est également possible d'appliquer une étape supplémentaire pour réduire ou augmenter le nombre de scénarios (et modifier les lois de probabilité en conséquence), voir les annexes A.3 et A.4.

2.4.4.3 Prévision d'ensemble des apports en eau

Les prévisions probabilistes des apports en eau ne sont pas encore utilisées en opérationnel à CNR. Néanmoins, les incertitudes sur les apports en eau impactent *a priori* l'optimisation (voir Sous-section 2.4.1). Des études sont menées par CNR pour faire passer en univers probabiliste les outils de prévision des apports utilisés en opérationnel (voir §2.3.4.1). Les études reposent sur une approche ensembliste (voir §1.4.5.1).

Un premier lot issu de ces études, livré avant le début de la thèse, comporte les prévisions d'ensemble des débits des 6 affluents du Haut-Rhône (l'Arve, la Valserine, les Usses, le Fier, le Séran et le Guiers, voir Figure 2.2), à des stations de mesure légèrement en amont de leur confluence avec le Rhône. Ces prévisions hydrologiques d'ensemble sont obtenues par forçage des prévisions d'ensemble météorologiques EPS (*Ensemble Prediction System*) des lancements quotidiens à 0 h UTC du modèle météorologique du ECMWF (*European Centre for Medium-Range Weather Forecasts*) sur la période du 01/01/2011 au 25/12/2014. Elles comportent 50 membres en plus du membre contrôle, définis au pas horaire (valeurs moyennes sur chaque heure) à un horizon de 120 h et commençant à 0 h UTC (il n'y a donc pas de décalage par rapport aux lancements du modèle météorologique du ECMWF). Ces prévisions d'ensemble météorologiques sont post-traitées de manière univariée pour assurer leur fiabilité (voir §1.4.5.2) puis réarrangées pour reproduire leur cohérence spatio-temporelle. Ensuite, ces prévisions sont intégrées dans trois modèles hydrologiques différents. Les hydrogrammes résultants sont post-traités de manière univariée pour combiner les sorties des trois modèles tout en prenant en compte leurs performances relatives afin d'assurer la fiabilité. Enfin, les hydrogrammes sont réarrangés pour reproduire leur cohérence spatio-temporelle. La démarche complète et la qualité des prévisions d'ensemble ainsi obtenues sont décrites dans [Bellier *et al.*, 2018 ; Bellier, 2018]. Dans ce lot, aucune distinction n'est effectuée entre le Fier et les autres affluents, alors que le Fier fait partie des affluents influencés par des aménagements hydroélectriques exploités par EDF (voir §2.2.1.5). Néanmoins, la fiabilité des prévisions d'ensemble est calculée par rapport aux débits réalisés, donc le post-traitement qui corrige la fiabilité tient compte des écarts dus à l'exploitation du débit du Fier mais en sur-estimant, *a priori*, les incertitudes de prévision. La figure 2.19 montre un exemple d'une prévision d'ensemble issue de ce lot.

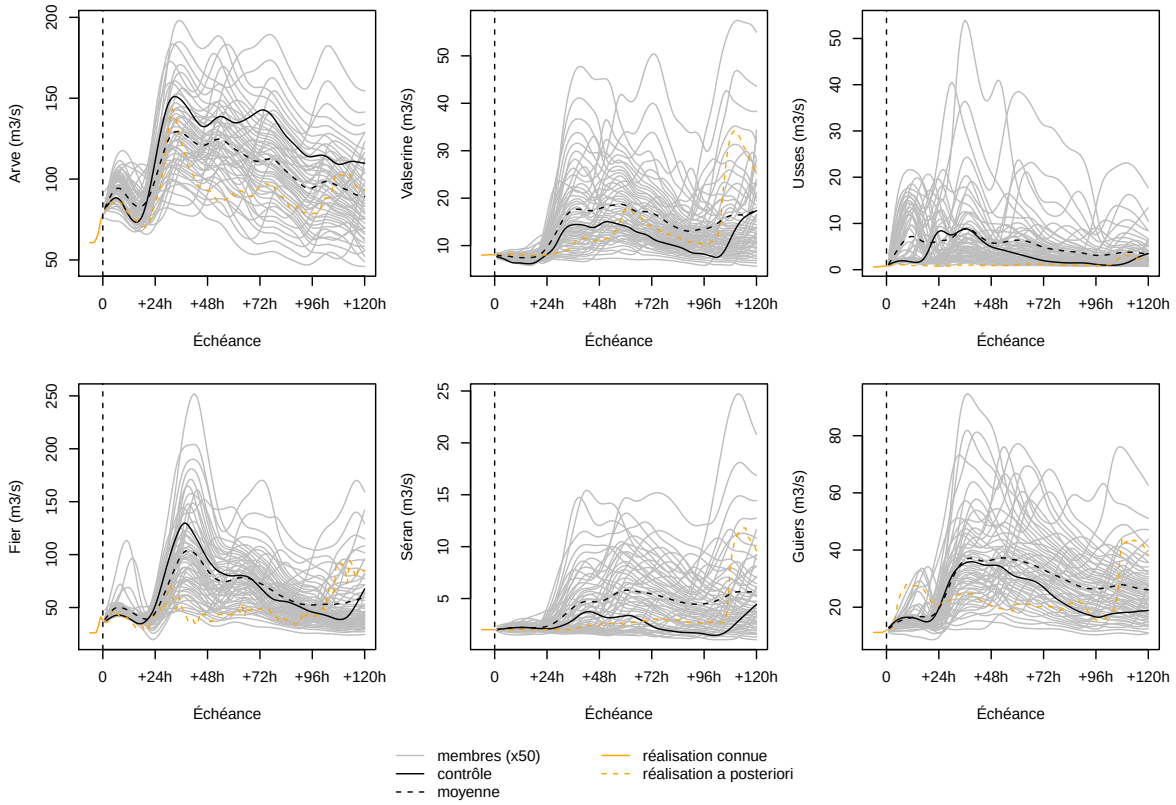


FIGURE 2.19 – Exemple de la prévision d'ensemble brute des débits des affluents du Haut-Rhône émise à 0 h UTC le 26/04/2014 (ligne verticale en tirets).

Les études sur la modélisation des incertitudes des apports en eau ne concernent pas les apports des bassins versants intermédiaires ni celles sur les apports de l'exutoire du lac du Bourget (voir §2.2.1.5). Dans le cadre de notre étude, les incertitudes sur ces apports en eau ne sont donc pas considérées. Cette hypothèse est acceptable car ces débits d'apports et leurs incertitudes sont négligeables devant les débits d'apports des affluents et leurs incertitudes. De même, on rappelle que les incertitudes sur les débits de correction ne sont pas considérées car elles sont liées aux incertitudes de modélisation que nous avons choisies de ne pas considérer (voir §2.4.4.1). Ainsi, les apports des bassins versants intermédiaires (paramètre Q_{BVI}), le débit d'apports de l'exutoire du lac du Bourget (considéré comme un affluent dans le modèle hydraulique) et les débits de correction (paramètre Q_{corr}) sont donnés par les prévisions déterministes utilisées en opérationnel (voir Sous-section 2.3.4). Néanmoins, la méthodologie que nous présentons dans ce manuscrit reste valable pour le cas où la prévision d'ensemble couvre tous les apports en eau.

D'autres lots de données d'apports probabilistes ont été livrés durant la thèse. Ils ont été obtenus par une approche similaire à celle du premier lot mais améliorée. Les données de ces lots couvrent les 21 affluents du Rhône sur un historique plus large (du 01/01/2008 au 31/12/2020 pour le dernier lot) et correspondent à des prévisions lancées à 8 h heure française (au lieu de 0 h UTC) pour correspondre au processus opérationnel. Le dernier lot de données couvre également les apports probabilistes des bassins versants intermédiaires. Des études sont également en cours à CNR pour quantifier les incertitudes sur les débits de correction. Néanmoins, les données de ces lots n'ont pas été utilisées durant la thèse.

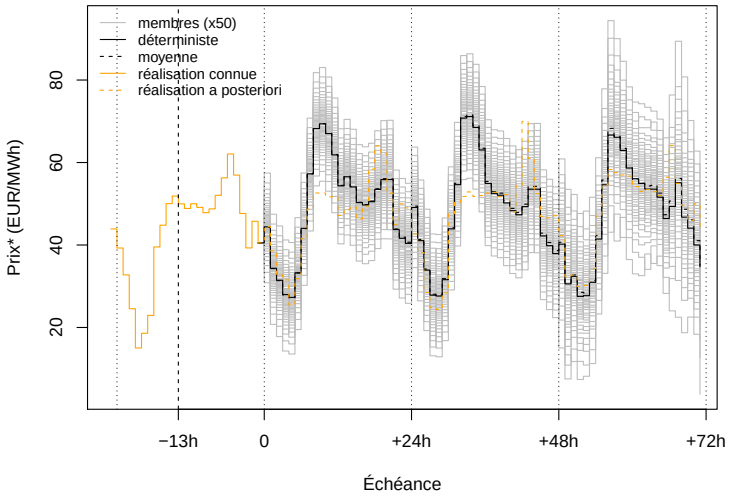
2.4.4.4 Prévision d'ensemble des prix

Les prévisions probabilistes des prix ne sont pas utilisées en opérationnel et, contrairement aux apports en eau, aucune étude préalable n'a été menée à CNR pour en construire. Néanmoins, les incertitudes sur les prix impactent *a priori* l'optimisation du programme de production car elles se répercutent linéairement sur le calcul du chiffre d'affaires (voir Sous-section 2.4.1).

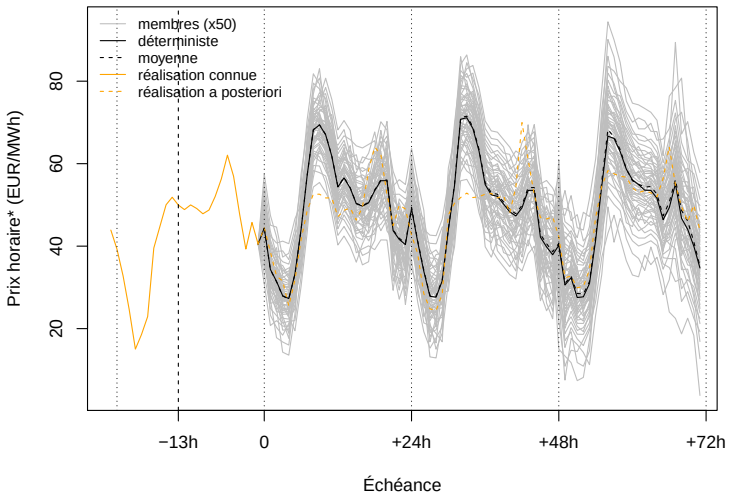
Une approche ensembliste simple a été implémentée durant la thèse. Elle permet d'obtenir une prévision d'ensemble à 50 membres des prix horaires de l'électricité à un horizon de 3 jours pour n'importe quelle situation. Cette approche ensembliste repose sur une approche par habillement (voir §1.4.5.1) qui consiste à entourer la prévision déterministe (voir §2.3.4.4) avec les statistiques des erreurs passées sur un historique de plusieurs années. On forme alors la loi de probabilité univariée empirique à chaque échéance horaire des erreurs des prévisions déterministes utilisées en opérationnel. Les incertitudes ainsi modélisées ne correspondent pas à celles liées aux conditions spécifiques du marché de l'énergie connues à l'instant présent mais celles constatées à grande échelle sur un historique passé. Pour transformer l'approche par habillement en une approche ensembliste à 50 membres équiprobables, nous utilisons la procédure suivante.

1. Pour chaque échéance horaire, on détermine grâce au modèle d'erreur les 50 quantiles empiriques associées aux probabilités $m/51$, pour $m \in \{1, \ldots, 50\}$, des erreurs de prévision observées sur l'historique passé.
2. Ces quantiles sont ajoutés autour de la prévision déterministe des prix horaires utilisée en opérationnel pour la situation donnée.
3. Les prix horaires ainsi obtenus sont combinés entre eux d'une échéance à l'autre par réarrangement de Schaake [Clark *et al.*, 2004] pour reformer la cohérence temporelle.

Pour la thèse, le modèle d'erreur est construit sur l'historique des dernières prévisions des matinées de la période du 01/01/2013 au 31/12/2016 à un horizon de 3 jours. La qualité de l'approche est mesurée sur la période du 01/01/2017 au 31/12/2020. Par construction, cette approche permet d'assurer une bonne fiabilité. En revanche, elle ne permet pas d'avoir une bonne finesse car elle a tendance à surestimer les incertitudes. La figure 2.20 montre un exemple d'une prévision d'ensemble obtenue avec cette approche.



(a) Graphique avec des courbes constantes par morceaux sur chaque heure des journées comme c'est le cas en pratique



(b) Graphique avec des courbes affines sur chaque heure des journées pour une meilleure visualisation des croisements entre les membres de la prévision d'ensemble

FIGURE 2.20 – Exemple de la prévision d'ensemble, émise à 11 h (heure française) le 27/10/2014 (ligne verticale en tirets), des prix du 28/10/2014 au 30/10/2014. L'échéance 0 correspond à l'instant 28/10/2014 0 h (heure française). Les prix du 27/10/2014 entre 11 h et 24 h sont connus au moment où la prévision est émise (prix *day-ahead* fixés la veille). Les lignes verticales en pointillées correspondent aux changements de jour. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les prix prévus.

Le nombre de membres de la prévision d'ensemble des prix, égal à 50, est arbitraire. Il a été choisi pour être égal au nombre de membres de la prévision d'ensemble des apports en eau, même si cette propriété n'est pas nécessaire *a priori*. Le nombre de membres aurait pu être déterminé, par exemple, en sélectionnant celui qui optimise un score de qualité (le CRPS par exemple, voir §1.4.5.2). Pour plus de détails sur la qualité des prévisions probabilistes des prix obtenue avec cette approche ensembliste simple, voir l'annexe A.2.

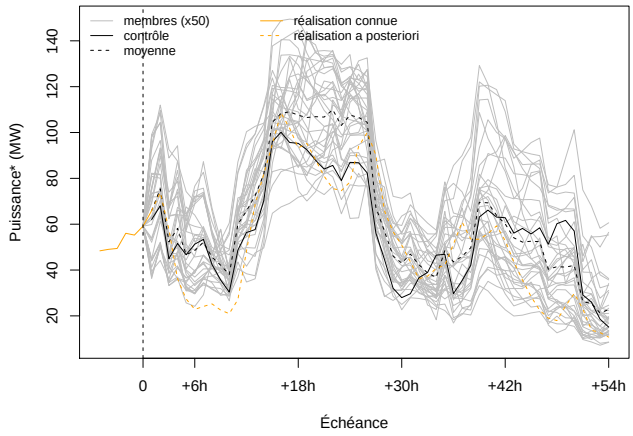
L'approche présentée ci-dessus a été construite uniquement dans le but d'illustrer et tester la méthodologie présentée dans ce manuscrit. En outre, en plus de ne pas représenter les incertitudes sur les conditions du marché de l'énergie connues à l'instant présent, cette approche ne permet pas de suivre les évolutions du marché de l'énergie (car il lui faut un historique passé de plusieurs années). L'approche doit donc être améliorée et validée pour effectuer une validation quantitative de la méthodologie. Des études supplémentaires ont été effectuées à ce sujet, mais elles ne sont pas décrites dans ce manuscrit car elles dépassent le cadre des objectifs de la thèse (voir Section 1.5).

2.4.4.5 Prévision d'ensemble des productions variables

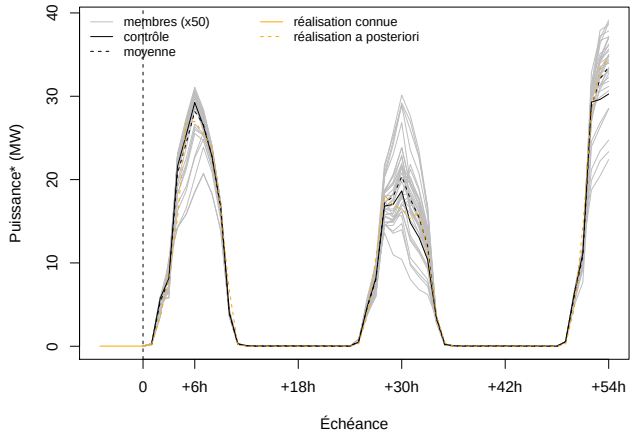
Les prévisions d'ensemble des productions variables ne sont pas encore utilisées en opérationnel. Néanmoins, les incertitudes sur les productions variables impactent *a priori* l'optimisation (voir §2.4.2.2). Tout comme pour les apports en eau, des études sont menées par CNR pour faire passer en univers probabiliste les outils de prévision des productions variables utilisés en opérationnel (voir §2.4.2.2). Les études reposent sur une approche ensembliste similaire à celle utilisée pour les apports en eau (voir §2.4.4.3).

Au début de la thèse, ces études étaient encore peu matures mais quelques données partielles d'une étude préliminaire étaient à disposition. Ces données correspondent à des prévisions d'ensemble de la production horaire d'une partie des actifs de production variable du périmètre d'équilibre CNR. Ces prévisions d'ensemble sont obtenues par forçage des prévisions météorologiques d'ensemble du modèle PEARP de Météo-France sur la période du 11/08/2014 au 16/11/2015. Elles comportent 34 membres en plus du membre contrôle, définis au pas horaire (valeurs moyennes sur chaque heure) à un horizon de 54 h et commençant à 6 h UTC. Elles sont cohérentes en temps et elles respectent les corrélations entre les productions éoliennes et photovoltaïques. La figure 2.21 montre un exemple d'une prévision d'ensemble obtenue avec cette approche. En revanche, par manque de maturité des études au début de la thèse, ces prévisions d'ensemble des productions variables ne sont pas post-traitées. Elles manquent donc de fiabilité et sont ainsi peu pertinentes pour tester la méthodologie d'optimisation en univers probabiliste présentée dans ce manuscrit (voir §1.4.5.2). En outre, l'horizon de 54 h ne permettrait pas de couvrir toute la fenêtre temporelle de l'optimisation. Ces prévisions d'ensemble des productions variable ne sont donc pas utilisées dans la suite. Néanmoins, la théorie de la méthodologie d'optimisation en univers probabiliste est présentée dans un cadre plus général que les données à disposition. Elle pourra donc être utilisée, le moment venu, avec des prévisions d'ensemble des productions variables de qualité satisfaisante.

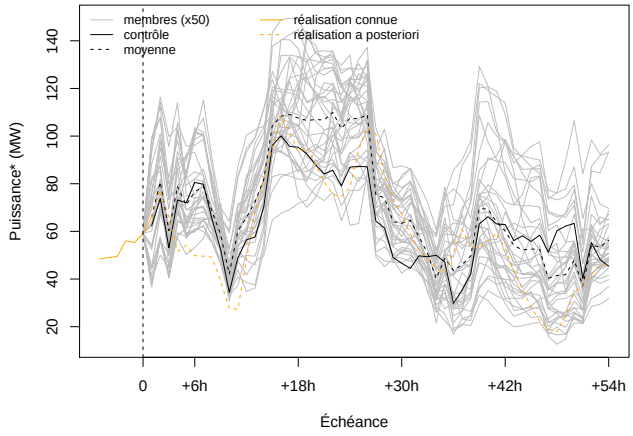
Des études plus avancées ont été réalisées par CNR durant la thèse, notamment concernant les productions photovoltaïques, mais ces nouvelles données n'ont pas été utilisées durant la thèse.



(a) Productions éoliennes



(b) Productions photovoltaïques



(c) Somme des productions éoliennes et photovoltaïques

FIGURE 2.21 – Exemple de la prévision d'ensemble des productions variables émise à 6 h UTC le 27/10/2014 (ligne verticale en tirets). Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les puissances prévues.

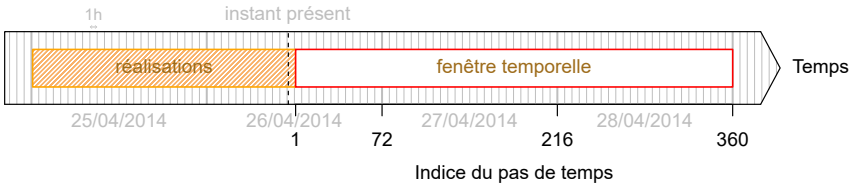


FIGURE 2.22 – Schéma de la fenêtre temporelle du cas d'étude (en rouge) avec la période réalisée permettant de définir les conditions initiales (en orange). Les traits verticaux gris représentent les heures rondes des journées.

2.5 Cas test étudié durant la thèse

La méthodologie que nous présentons dans ce manuscrit est illustrée et testée sur un cas d'étude qui rejoue une situation passée, avec quelques simplifications. Néanmoins, la théorie de la méthodologie proposée est volontairement plus générale que le cas d'étude pour qu'elle puisse être adaptée à n'importe quel cas d'étude moins simplifié, voire pour d'autres chaînes hydroélectriques. Le cas d'étude est choisi en fonction des données probabilistes qui étaient à disposition au début de la thèse (voir Sous-section 2.4.4).

2.5.1 Situation étudiée

L'instant présent de la situation étudiée est le 26/04/2014 à 11 h et l'horizon est la fin du 28/04/2014. Les conditions initiales sont données par les valeurs réalisées sur la période passée du 25/04/2014 0 h au 26/04/2014 11 h. À l'instant présent de ce cas d'étude (26/04/2014 11 h), l'objectif principal de la planification de la production est de définir les prochaines ventes *day-ahead* à effectuer avant 12 h, qui portent sur la production du 27/04/2014.

Comme expliqué dans le §2.3.5.2, la fenêtre temporelle commence une heure après l'instant présent : elle commence le 26/04/2014 à 12 h et elle termine le 28/04/2014 à 24 h (horizon). La fenêtre temporelle est découpée en 320 pas de temps de même durée $\delta t = 10$ min. Pour définir les conditions initiales, on utilise les valeurs réalisées des variables et paramètres sur la période passée du 25/04/2014 à 0 h au début de la fenêtre temporelle (le 26/04/2014 à 12 h). En opérationnel, les valeurs réalisées entre l'instant présent (26/04/2014 11 h) et le début de la fenêtre temporelle (26/04/2014 12h) ne seraient pas disponibles. Sur cette période, il aurait donc fallu utiliser les valeurs des variables des dernières consignes de conduite des aménagements et les dernières valeurs prévues des paramètres (voir §2.3.5.2). Cette simplification est néanmoins acceptable dans le cadre du travail présenté dans ce manuscrit. La figure 2.22 illustre la fenêtre temporelle du cas d'étude.

Le paramétrage utilisateur utilisé pour l'optimiseur est celui par défaut, sans l'utilisation de mode de régulation spécifique. L'optimiseur est initialisé avec le programme de production qui a été construit semi-manuellement sur tous les aménagements du Rhône à cette situation, appelé le *programme de production initial*. En particulier, les valeurs des volumes finaux minimums dans les retenues (paramètres nécessaires pour que l'optimiseur ne vide pas complètement les retenues à la fin de la période temporelle) sont celles du programme de production initial, et les bornes minimales et maximales sur les volumes (considérées comme des paramètres pour l'optimiseur) sont celles associées au programme de production initial.

Pour le cas d'étude, on simplifie également le calcul du chiffre d'affaire en supposant que toute la production est vendue à court terme au marché *day-ahead* (voir Sous-section 2.3.3) et donc que les engagements des contrats à plus long terme sont nuls. Cela revient à soustraire une constante dans le calcul du chiffre d'affaires, ce qui n'a pas d'impact sur l'optimisation.

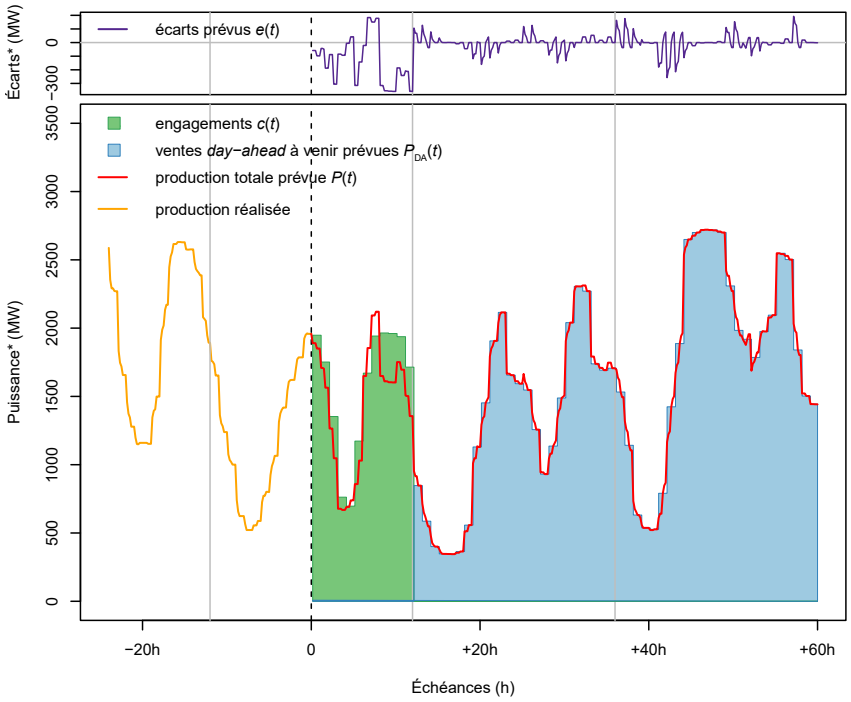


FIGURE 2.23 – Exemple, pour le cas d'étude, de la production potentielle prévue (en rouge), des prochaines ventes *day-ahead* (en bleu) obtenues par la planification semi-manuelle du Rhône (programme de production initial) et les engagements utilisés dans ce manuscrit (en vert). Les écarts sont représentés sur le graphique du haut en violet. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les puissances.

Ainsi, on considère que les engagements connus à l'instant présent sont uniquement ceux issus des ventes *day-ahead* effectuées la veille, qui portent sur la production du jour courant. Comme cela n'est pas le cas en pratique, les valeurs des engagements utilisées pour le cas d'étude sont obtenues à partir de la production potentielle du Rhône prévue la veille à 11 h au pas horaire (on ne considère donc pas les réelles transactions qui ont été faites par les opérateurs marchés). La figure 2.23 illustre, pour le cas d'étude, le programme de production initial (construit par l'approche semi-manuelle du §2.3.5.1) et les écarts qui s'en déduisent à partir des engagements utilisés dans ce manuscrit.

2.5.2 Simplifications

Comme les données d'apports probabilistes ne couvrent que les affluents du Haut-Rhône, l'optimisation est portée uniquement sur le Haut-Rhône, i.e. les 6 premiers aménagements en amont, voir Figure 2.24. Le programme de production pour le Bas-Rhône (les 12 autres aménagements en aval) est fixé par le programme de production initial et il permet de calculer la puissance totale de toute la chaîne hydroélectrique. Cette simplification a l'avantage de réduire la complexité du problème d'optimisation et donc de limiter les temps de calcul pour la thèse. En revanche, elle ne permet pas de tirer profit de toute la flexibilité du Rhône.

Une autre conséquence du choix du cas d'étude est que les incertitudes sur les productions variables ne sont pas considérées par manque de données probabilistes de qualité satisfaisante au début de la thèse. Un scénario est donc une combinaison d'un membre de la prévision d'ensemble des apports et d'un membre de la prévision d'ensemble des prix. Autrement dit, le cas d'étude ne permet pas d'effectuer la gestion conjointe des énergies renouvelables décrite à la sous-section 2.4.2. Or, une des raisons qui nous poussent à faire passer l'optimiseur en

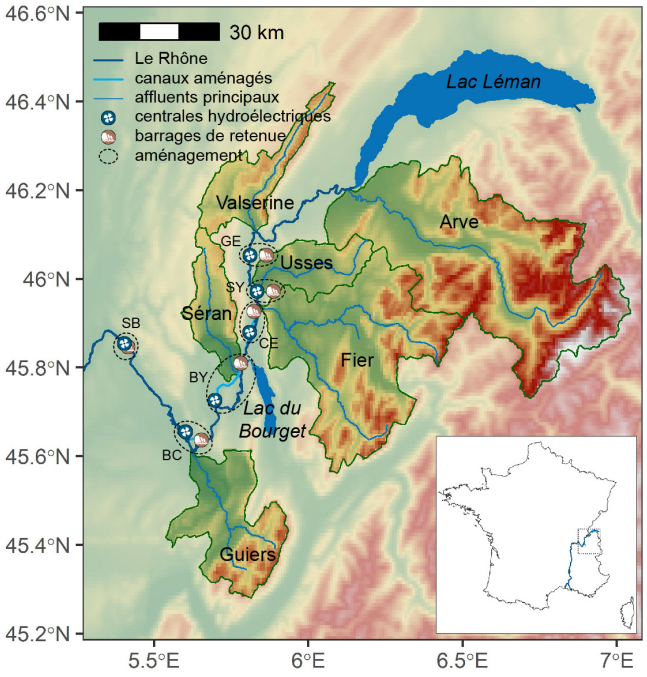


FIGURE 2.24 – Carte des bassins versants et des aménagements hydroélectriques CNR du Haut-Rhône.

univers probabiliste est de pouvoir effectuer une gestion conjointe des énergies renouvelables permettant de gérer les incertitudes qui entachent les prévisions des productions variables (voir Sous-section 2.4.2). Cette simplification ne permet donc pas de tirer profit de tous les avantages de l'optimisation en univers probabiliste. Néanmoins, cette simplification ne concerne pas la méthodologie que nous présentons dans ce manuscrit qui se veut plus générale que le cas d'étude.

2.5.3 Prévisions

2.5.3.1 Prévision des apports en eau

Les débits d'apports en eau des prévisions d'ensemble du §2.4.4.3 sont des données brutes commençant à 0 h UTC (voir §2.4.4.3), ce qui n'est pas adapté aux heures de lancement des planifications de la production en opérationnel. Pour le cas d'étude, la prévision d'ensemble des apports doit commencer le 26/04/2014 à 12 h (heure française), comme c'est le cas pour la prévision déterministe. Pour cela, nous recollons les données brutes de la prévision d'ensemble avec les valeurs réalisées jusqu'à l'instant présent de l'optimisation. La difficulté est que cette modification entraîne une discontinuité des débits d'apports à l'instant présent qui doit être supprimée pour assurer la cohérence physique de la situation que l'on cherche à décrire. Une rampe corrective est alors appliquée sur les premières heures qui suivent l'instant présent. La procédure plus détaillée, qui a été implémentée durant la thèse, est la suivante.

1. On extrait la prévision d'ensemble brute du 26/04/2014 à 0 h UTC ainsi que les débits réalisées du 25/04/2014 0 h au 26/04/2014 12 h (heures françaises). Ces données correspondent aux débits horaires des affluents principaux du Haut-Rhône de la figure 2.19, tronquées à l'horizon considéré (voir Figure 2.25a).
2. On passe les débits d'une granularité horaire à une granularité demi-horaire (voir Figure 2.25b), celle utilisée par les prévisions déterministes (voir §2.3.4.1). Pour cela, les débits horaires sont interpolées à des pas demi-horaires (le débit à une demi-heure stricte

est calculé comme la moyenne des débits des deux heures qui encadrent la demi-heure considérée).

3. Sur la période entre le début de la prévision d'ensemble et l'instant présent du cas d'étude, i.e. la période entre 26/04/2014 0 h (UTC) et 26/04/2014 12 h (heure française), les débits de la prévision d'ensemble sont recollés avec les débits réalisés (voir Figure 2.25c). Pour assurer la continuité des débits à l'instant présent, une rampe corrective affine est appliquée entre l'instant présent et la fin du jour courant, i.e. entre 26/04/2014 12 h et 26/04/2014 24 h (heures françaises), qui impose l'égalité des prévisions avec les réalisations à l'instant présent tout en recollant avec les débits de la prévision d'ensemble à la fin du jour courant (voir Figure 2.25d). Les débits de la prévision d'ensemble du 27/04/2014 et du 28/04/2014 restent inchangés. À ce stade, on dispose des valeurs du paramètre Q_{aff} pour chaque membre de la prévision d'ensemble. Plus précisément, pour tout affluent d'indice $k \in \{1, \dots, 6\}$ et tout membre d'indice $m \in \{1, \dots, 50\}$, on a :

$$\forall t \in \mathbb{R}, \quad Q_{\text{aff}}^{(m)}(k,t) = \begin{cases} Q_{\text{obs}}(k,t) & \text{si } t \leq t_0, \\ Q_{\text{obs}}(k,t_0) + (Q_{\text{ens}}^{(m)}(k,t) - Q_{\text{obs}}(k,t))\dfrac{t-t_0}{t_1-t_0} & \text{si } t_0 < t \leq t_1, \\ Q_{\text{ens}}^{(m)}(k,t) & \text{si } t_1 < t, \end{cases}$$

où, $t_0 \in \mathbb{R}$ et $t_1 > t_0$ désignent respectivement les instants 26/04/2014 12 h et 27/04/2014 0 h, et, pour tout $t \in \mathbb{R}$, $Q_{\text{aff}}^{(m)}(k,t) \in \mathbb{R}_+$ désigne le débit d'apports associé au membre d'indice m pour l'affluent d'indice k à l'instant t, $Q_{\text{obs}}(k,t) \in \mathbb{R}_+$ le débit réalisé de l'affluent d'indice k à l'instant t et $Q_{\text{ens}}^{(m)}(k,t) \in \mathbb{R}_+$ désigne le débit de la prévision d'ensemble à la granularité demi-horaire du membre d'indice m pour l'affluent d'indice k à l'instant t.

4. Les apports en eau aux points de régalage des aménagements sont ensuite calculés pour chaque membre de la prévision d'ensemble grâce à l'équation (2.4). Pour cela, on propage les débits aux points de réglage des aménagements du Haut-Rhône et on ajoute la prévision déterministe Q_{BVI} des apports des bassins versants intermédiaires et la prévision déterministe Q_{corr} des débits de correction (voir Figure 2.25e). Pour la thèse, la somme des deux paramètres Q_{BVI} et Q_{corr} est calculée à partir des apports complets utilisés en opérationnel et la moyenne des débits d'apports des membres de la prévision d'ensemble :

$$\forall (i,t) \in \mathcal{I} \times \mathbb{R}, \quad Q_{\text{BVI}}(i,t) + Q_{\text{corr}}(i,t) \approx \tilde{a}(i,t) - \frac{1}{50}\sum_{m=1}^{50}\sum_{k \in \mathcal{A}(i)} Q_{\text{aff}}^{(m)}(k, t - \tau_{\text{aff}}(k,i)),$$

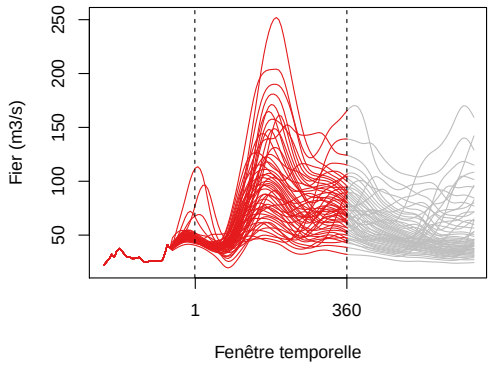
où $\tilde{a}(i,t) \in \mathbb{R}$, pour $(i,t) \in \mathcal{I} \times \mathbb{R}$, désignent les apports complets aux points de régalage des aménagements utilisés en opérationnel.

Cette procédure est acceptable pour tester et illustrer la méthodologie que nous présentons dans ce manuscrit, mais elle doit être améliorée si l'on souhaite faire une validation quantitative de la méthodologie. Les principales limites sont liées aux simplifications faites aux étapes 2 à 4 de la procédure, qui impactent la qualité des prévisions :

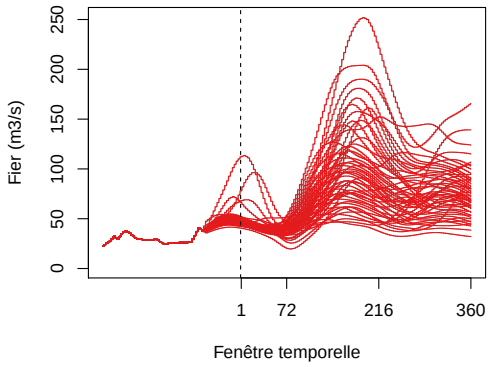
— les valeurs horaires sont des débits moyens sur chaque heure, rendant discutable leur passage à la granularité demi-horaire à l'étape 2,

— la rampe corrective de l'étape 3 déforme les lois de probabilité par rapport aux données brutes sur le jour courant, impactant la qualité des prévisions probabilistes,

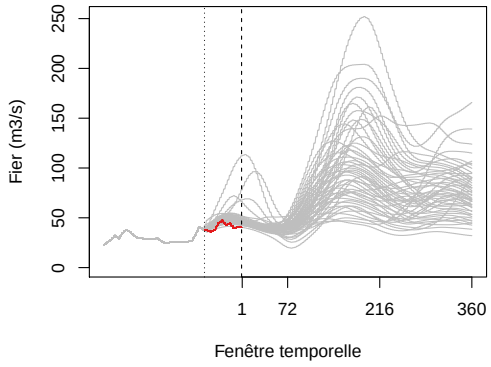
— l'estimation de la somme des deux paramètres Q_{BVI} et Q_{corr} à l'étape 4 est une forte approximation car elle suppose que la moyenne des membres de la prévision d'ensemble correspond à la prévision déterministe utilisée en opérationnel, ce qui n'est pas le cas. Il serait plus pertinent en pratique d'utiliser les paramètres Q_{BVI} et Q_{corr} utilisés en opérationnel.



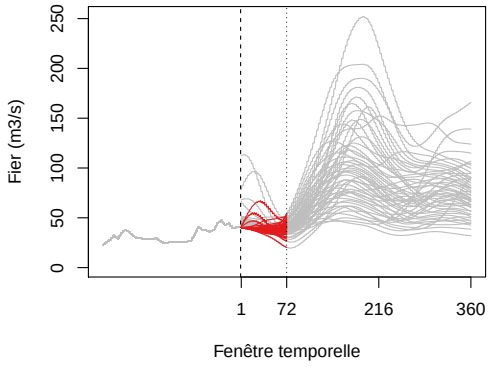
(a) Utilisation de la prévision d'ensemble brute



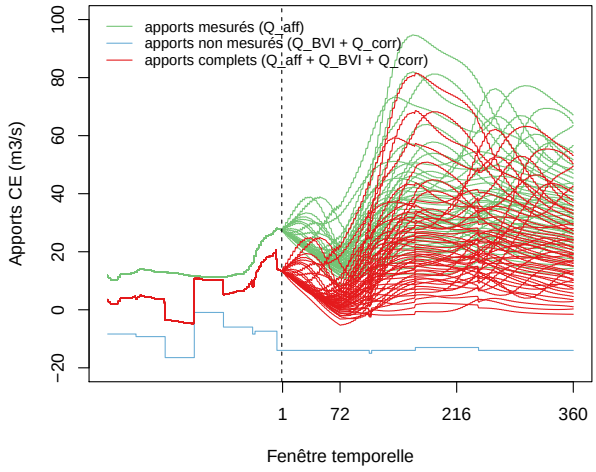
(b) Passage au pas demi-horaire



(c) Recollage avec les valeurs réalisées



(d) Rampe corrective sur assurer la continuité



(e) Calcul des apports complets à l'aménagement

FIGURE 2.25 – Illustration pour le cas d'étude des étapes de la procédure de création des prévisions d'ensemble des apports en eau. Les figures concerne les apports en eau à l'aménagement de Chautagne faisant intervenir le débit du Fier.

Parmi les situations compatibles avec les données d'apports probabilistes qui étaient à disposition au début de la thèse (voir Sous-section 2.4.4), la situation du cas d'étude (dont l'instant présent est le 26/04/2014 à 11 h) a été choisie pour que les membres des apports soient assez dispersés et avec des variations décalées dans le temps (les membres se croisent). Cela permet d'avoir un cas d'étude intéressant pour étudier l'intérêt de l'optimisation en univers probabiliste. La figure 2.26 présente, pour le cas d'étude, la prévision d'ensemble modifiée des apports en eau des affluents du Haut-Rhône à l'issue de l'étape 3 de la procédure ci-dessus. Les apports en eau complets aux aménagements s'en déduisent en appliquant l'étape 4 ci-dessus (opérations « déterministes » car identiques pour chaque membre de la prévision d'ensemble).

Le débit entrant à Génissiat est un paramètre incertain qui dépend des débits d'apports de l'Arve et de la Valserine (voir Sous-section 2.3.4). Les incertitudes sur le débit d'apports de la Valserine se propagent complètement sur le débit entrant à Génissiat. En revanche, seule une partie des incertitudes sur le débit d'apports de l'Arve se propage sur le débit entrant à Génissiat car le débit au barrage du Seujet en Suisse peut être modifié en temps réel sur certaines plages horaires pour compenser les incertitudes sur le débit d'apports de l'Arve sans modifier la production prévue à la centrale hydroélectrique de Verbois. Toutes les incertitudes sur le débit d'apports de l'Arve ne peuvent toutefois pas être compensées car le débit au barrage du Seujet n'est pas modulable sur certaines plages horaires (débit imposé) et sur les plages horaires où il l'est, il doit être compris entre une borne minimale et une borne maximale (contraintes d'exploitation), limitant la modulation possible. Une procédure implémentée durant la thèse permet de calculer le débit entrant à Génissiat pour chaque membre de la prévision d'ensemble hydrologique à partir du programme de production communiqué le 26/04/2014 à 8h30 par les Services Industriels de Genève et des prévisions des débits d'apports de l'Arve et de la Valserine obtenues à l'étape 3 ci-dessus. La prévision d'ensemble des débits entrants à Génissiat ainsi obtenue est présentée à la figure 2.27 pour le cas d'étude.

Les simplifications effectuées sur les membres de la prévision d'ensemble brute des apports en eau ne permettent pas de les comparer à la prévision déterministe utilisée en opérationnel pour la situation étudiée. C'est pourquoi dans la suite du manuscrit, la prévision déterministe des apports en eau est remplacée par le membre contrôle de la prévision d'ensemble (obtenu avec les mêmes simplifications que les membres perturbés), comme présenté sur les figures 2.26 et 2.27.

2.5.3.2 Prévision des prix

La prévision déterministe des prix que nous utilisons n'est pas celle qui correspond à la situation étudiée (dont l'instant présent est le 26/04/2014 à 11 h) mais à la situation passée du 27/10/2014 à 11 h. Cette modification permet d'avoir des valeurs de prix en semaine (le 27/10/2014 était un lundi alors que le 26/04/2014 était un samedi) avec des variations horaires plus intéressantes à étudier pour l'optimisation.

La prévision d'ensemble des prix est construite à partir de cette prévision déterministe grâce au modèle d'erreur du §2.4.4.4 puis elle est mise en correspondance avec la fenêtre temporelle du cas d'étude entre le 28/04/2014 0 h et le 29/04/2014 24 h. Les valeurs des prix entre le 25/04/2014 0 h et le 26/04/2014 24 h sont connus à l'instant présent du cas d'étude, mais on utilise les valeurs réalisées des prix entre le 26/10/2014 0 h et le 27/10/2014 24 h pour correspondre avec la prévision des prix des 2 jours suivants. Ces modifications sur les prix ne changent que le calcul du chiffre d'affaires du programme de production. Elles sont acceptables dans le cadre de la thèse car les apports en eau et les prix sont supposés indépendants (hypothèse d'indépendance du §2.4.4.2). La prévision déterministe et la prévision d'ensemble des prix sont alors mises en correspondance avec la fenêtre temporelle du cas d'étude, comme illustré à la figure 2.28.

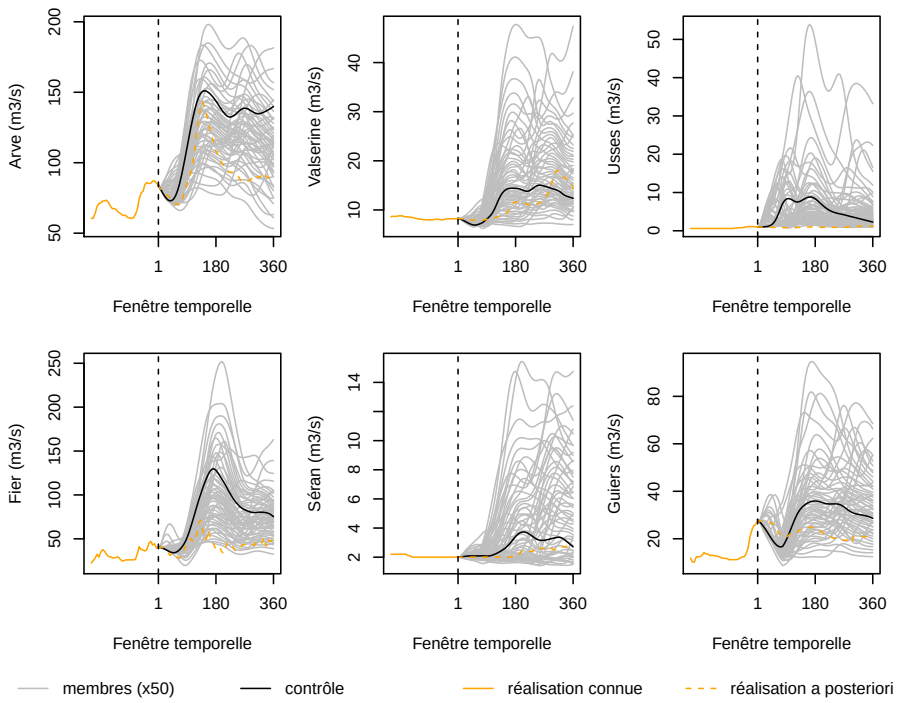


FIGURE 2.26 – Prévision d'ensemble modifiée des apports en eau des affluents du Haut-Rhône pour le cas d'étude. La ligne verticale en tirets correspond à l'instant présent.

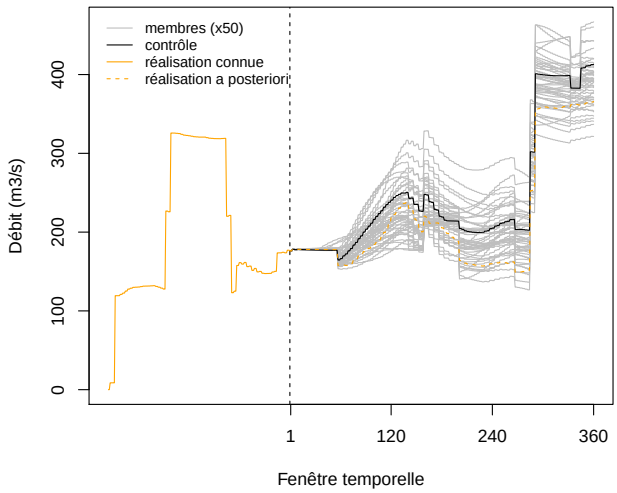


FIGURE 2.27 – Prévision d'ensemble des débits entrants à Génissiat pour le cas d'étude. La ligne verticale en tirets correspond à l'instant présent.

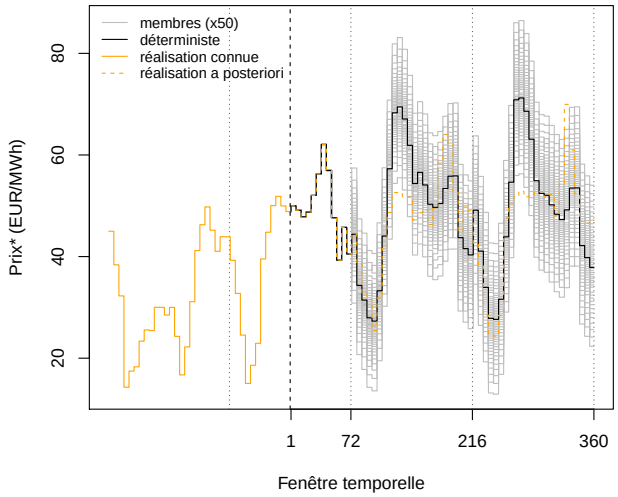


FIGURE 2.28 – Valeurs des paramètres des prix pour le cas d'étude de la prévision déterministe (en noir) et de la prévision d'ensemble (en gris). Les lignes verticales en pointillés correspondent aux changements de jour. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les prix prévus.

En outre, les prix de règlement des écarts pour le cas d'étude sont estimés par : $\forall t \in \mathbb{R},\ \kappa^+(t) = \kappa^-(t) = 0{,}05$ (voir Équation (2.14)). On considère que les erreurs de prévision dues à cette estimation des prix de règlement des écarts sont une source d'incertitudes liées à la modélisation du problème (qui s'ajoutent aux incertitudes de prévision des prix), donc elles ne sont pas prises en compte dans ce manuscrit (voir §2.4.4.1). Il convient néanmoins de noter que le calcul du chiffre d'affaires (et donc l'optimiseur) est *a priori* très sensible à la prévision des prix de règlement des écarts (voir Équation (2.12)).

2.6 Conclusion du chapitre

L'optimiseur qui sert de référence dans notre travail s'inscrit dans le processus opérationnel de CNR pour la gestion de la production hydroélectrique du Rhône, en lien avec le contexte plus général présenté au premier chapitre. Cet optimiseur permet de construire un programme de production du Rhône à court terme qui respecte les contraintes d'exploitation et qui maximise le chiffre d'affaires. Il repose actuellement sur des prévisions déterministes des apports en eau et des prix. Pour améliorer la valorisation de la production du Rhône, il semble important que l'optimiseur prenne en compte les incertitudes qui entachent ces deux prévisions. Par ailleurs, avec le développement des actifs de production variable, la valorisation de la production du Rhône peut être effectuée de manière conjointe avec les productions variables du périmètre d'équilibre CNR. Dans ce cas, il semble important que l'optimiseur prenne en compte les incertitudes qui entachent la prévision de ces productions variables.

Dans le chapitre suivant, qui clôt la première partie de ce manuscrit, nous présentons la formulation mathématique du problème d'optimisation en univers déterministe de la planification à court terme de la production conjointe d'une chaîne hydroélectrique et d'actifs de production variable. Cette formulation mathématique s'appuie sur celle utilisée par l'optimiseur de CNR : elle s'écrit à partir du modèle hydraulique, des contraintes d'exploitation et du calcul du chiffre d'affaires présentés dans ce chapitre. Nous présentons également dans le chapitre suivant la manière dont nous adaptons cette formulation mathématique pour tenir compte des scénarios multivariés qui modélisent les incertitudes. La résolution numérique du problème d'optimisation

en univers probabiliste ainsi obtenu est présentée dans la seconde partie du manuscrit. Le cas d'étude présenté à la section 2.5 de ce chapitre sert de référence pour les figures qui illustrent la méthodologie présentée dans les chapitres suivants.

Chapitre 3

Problème de planification à court terme de la production

Le chapitre précédent a montré l'intérêt d'utiliser un optimiseur pour la gestion de la production à l'échelle du périmètre d'équilibre de CNR. Dans ce chapitre, nous donnons la formulation mathématique du problème d'optimisation à partir des équations du modèle hydraulique, des contraintes d'exploitation et du calcul du chiffre d'affaires présentés au chapitre précédent. Nous nous plaçons néanmoins dans un cadre légèrement plus général et moins spécifique au contexte de CNR que ce qui a été présenté au chapitre précédent. L'objectif est d'aboutir à une écriture simple du problème d'optimisation en univers déterministe, pour servir de base au passage à l'univers probabiliste. Dans la première section de ce chapitre, nous présentons quelques généralités sur l'optimisation, notamment pour mettre en place le vocabulaire approprié. Dans une deuxième section, nous donnons la formulation mathématique du problème de planification à court terme de la production en univers déterministe. Enfin, dans une dernière section, nous présentons un des principaux points de ce manuscrit : l'adaptation de la formulation mathématique du problème d'optimisation pour tenir compte des incertitudes qui entachent les prévisions.

3.1 Généralités sur l'optimisation

Nous présentons dans cette section quelques généralités mathématiques sur les problèmes d'optimisation qui sont utiles pour formuler mathématiquement le problème de planification de la production et mettre en place le vocabulaire approprié. Les éléments de cette section sont principalement issus de [Allaire, 2005 ; Diwekar, 2020 ; Bonnans *et al.*, 1997 ; Culioli, 2012 ; Padberg, 2013].

3.1.1 Problème d'optimisation

Un problème d'optimisation consiste à trouver un point de maximum (ou minimum), s'il existe, d'une fonction $f \in \mathcal{F}(\mathbb{R}^r, \mathbb{R})$ ($r \in \mathbb{N}^*$) dans un ensemble $\mathcal{X} \subset \mathbb{R}^r$. Quitte à remplacer la fonction f par sa fonction opposée $-f \in \mathcal{F}(\mathbb{R}^r, \mathbb{R})$, on peut se ramener au cas de la maximisation. Le problème d'optimisation s'écrit alors :

$$\max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}). \tag{3.1}$$

On dit que f est la *fonction objectif* et que $\mathcal{X}$ est l'*ensemble admissible* du problème d'optimisation (3.1).

Une *solution admissible* du problème d'optimisation (3.1) est par définition un élément de l'ensemble admissible $\mathcal{X}$. Une *solution optimale* du problème d'optimisation (3.1) est par définition un élément $\mathbf{x}^* \in \mathcal{X}$, s'il existe, qui vérifie :

$$\forall \mathbf{x} \in \mathcal{X}, \quad f(\mathbf{x}) \leq f(\mathbf{x}^*).$$

La *valeur optimale* du problème d'optimisation (3.1) est par définition :

$$\nu^* = \sup\{f(\mathbf{x}),\ \mathbf{x} \in \mathcal{X}\} \in \mathbb{R} \cup \{-\infty, +\infty\}.$$

La valeur optimale est égale à $-\infty$ si et seulement si l'ensemble admissible $\mathcal{X}$ est vide (convention). Si l'ensemble admissible $\mathcal{X}$ n'est pas vide, alors la valeur optimale est égale à $+\infty$ si et seulement si la fonction f n'admet pas de point de maximum dans $\mathcal{X}$.

Par définition de la valeur optimale, les propriétés suivantes sont vérifiées.

Propriétés 3.1.

- $\forall \mathbf{x} \in \mathcal{X},\ f(\mathbf{x}) \leq \nu^*$ *(inégalité dans* $\mathbb{R} \cup \{-\infty, +\infty\}$*),*
- *Si* $\mathbf{x}^* \in \mathcal{X}$ *est une solution optimale du problème d'optimisation* (3.1)*, alors* $\nu^* = \max\{f(\mathbf{x}),\ \mathbf{x} \in \mathcal{X}\} = f(\mathbf{x}^*) \in \mathbb{R}$*.*

Une autre propriété importante du problème d'optimisation (3.1) est le résultat suivant qui porte sur l'existence d'une solution optimale (la démonstration est donnée par exemple dans [Gourdon, 2008b]).

Propriété 3.2. *Si la fonction* f *est continue et l'ensemble* $\mathcal{X}$ *est compact (i.e. non vide, fermé et borné), alors le problème d'optimisation* (3.1) *admet (au moins) une solution optimale. Si, en outre, la fonction* f *est strictement concave sur l'ensemble* $\mathcal{X}$*, alors la solution optimale est unique.*

En général, l'ensemble $\mathcal{X} \subset \mathbb{R}^r$ est défini par un ensemble de contraintes, i.e.

$$\mathcal{X} = \bigcap_{1 \leq i \leq s} \{\mathbf{x} \in \mathbb{R}^r\,;\ g_i(\mathbf{x}) \leq 0\},$$

où, pour tout $i \in \{1, \ldots, s\}$, $g_i \in \mathcal{F}(\mathbb{R}^r, \mathbb{R})$ ($s \in \mathbb{N}^*$). Dans ce cas, on dit que le problème d'optimisation (3.1) est sous contraintes, et on peut le noter :

$$\max_{\substack{g_i(\mathbf{x}) \leq 0,\ 1 \leq i \leq s \\ \mathbf{x} \in \mathbb{R}^r}} f(\mathbf{x}). \tag{3.2}$$

Il existe des algorithmes généraux qui permettent de résoudre le problème d'optimisation sous contraintes (3.2). Ils sont généralement de nature itérative, i.e. ils construisent une suite de solutions admissibles qui converge (selon certaines hypothèses sur les fonctions f et g_i, pour $i \in \{1, \ldots, s\}$) vers une solution optimale (si elle existe). Il existe deux grandes catégories d'algorithmes : ceux de type déterministe et ceux de type stochastique (l'analyse de ces derniers fait appel à la théorie des probabilités). Les algorithmes de gradient avec projection, d'Uzawa et de Newton sont déterministes et ils sont généralement efficaces lorsque la fonction objectif f est concave [Allaire, 2005]. Les algorithmes de recuit simulé et génétiques sont de type stochastique et ils sont généralement plus efficaces que les algorithmes déterministes lorsque la fonction objectif f n'est pas concave. Lorsque les hypothèses d'application des algorithmes « classiques » ne sont pas vérifiées, la résolution du problème d'optimisation sous contraintes (3.2) est souvent très difficile. Il est néanmoins parfois possible de construire des heuristiques et méta-heuristiques.

3.1.2 Programmation linéaire

On dit que le problème d'optimisation sous contraintes (3.2) est *linéaire* si la fonction objectif f et les fonctions g_i, pour $i \in \{1, \dots, s\}$, sont linéaires. L'ensemble admissible d'un problème d'optimisation linéaire est alors un polyèdre convexe, i.e.

$$\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^r \,;\; W\mathbf{x} \leq w\},$$

où $W \in \mathcal{M}_{sr}(\mathbb{R})$ et $w \in \mathbb{R}^s$ (l'opérateur $\leq$ porte sur chacune des composantes).

Un *sommet* du polyèdre $\mathcal{X}$ est par définition un point $\hat{\mathbf{x}} \in \mathcal{X}$ qui ne peut pas s'écrire comme combinaison convexe non triviale de deux points de $\mathcal{X}$, i.e. si $\hat{\mathbf{x}} = \theta\mathbf{x} + (1-\theta)\mathbf{y}$ avec $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^2$ et $\theta \in]0, 1[$, alors $\hat{\mathbf{x}} = \mathbf{x} = \mathbf{y}$.

La fonction objectif d'un problème d'optimisation linéaire s'écrit :

$$\forall \mathbf{x} \in \mathbb{R}^r, \quad f(\mathbf{x}) = \gamma^\top \mathbf{x},$$

où $\gamma \in \mathbb{R}^r$ (valeur constante du gradient de la fonction linéaire f).

Le problème d'optimisation linéaire peut alors se noter :

$$\max_{\substack{W\mathbf{x} \leq w \\ \mathbf{x} \in \mathbb{R}^r}} \gamma^\top \mathbf{x}. \tag{3.3}$$

On dit que les coefficients de la matrice $W \in \mathcal{M}_{sr}(\mathbb{R})$ et les composantes des vecteurs $w \in \mathbb{R}^s$ et $\gamma \in \mathbb{R}^r$ sont les *paramètres* du problème d'optimisation linéaire (3.3). On dit également que les composantes des vecteurs $\mathbf{x} \in \mathbb{R}^r$ sont les *variables de décision* du problème d'optimisation linéaire (3.3). Les contraintes du problème d'optimisation linéaire (3.3) sont donc des inégalités larges qui portent sur des combinaisons linéaires des variables de décision. Les contraintes peuvent également être des égalités, en combinant deux inégalités de sens inverse.

Comme la fonction f est continue (car linéaire) et le polyèdre $\mathcal{X}$ est fermé, la propriété 3.2 s'écrit de la manière suivante.

Propriété 3.3. *Si l'ensemble admissible $\mathcal{X}$ est non vide et borné, alors le problème d'optimisation linéaire* (3.3) *admet (au moins) une solution optimale.*

Lorsque l'on parle d'optimisation linéaire, il est courant de remplacer le terme « optimisation » par le terme « programmation ». On parle ainsi de programmation linéaire et de problème de programmation linéaire. Ces termes viennent du terme anglais *Linear Programming* (LP) initié par George Dantzig [Dantzig, 2002].

Les algorithmes de résolution d'un problème LP sont assez efficaces et très bien maîtrisés théoriquement [Vanderbei, 2020]. L'algorithme du simplexe est un des plus connus [Nash, 2000]. Il repose sur un parcours itératif des sommets du polyèdre $\mathcal{X}$ faisant croître la fonction objectif, pour tirer profit de la propriété suivante, dont la démonstration est donnée par exemple dans [Allaire, 2005].

Propriété 3.4. *Si le problème LP* (3.3) *admet une solution optimale, alors il existe une solution optimale qui est un sommet du polyèdre convexe $\mathcal{X}$.*

Les algorithmes de points intérieurs (par exemple l'algorithme de trajectoire centrale) sont aussi très utilisés [Roos *et al.*, 2005]. Lorsque la taille du problème (i.e. le nombre de variables de décision) est assez grande, la résolution numérique peut être assez longue. Des améliorations des algorithmes du simplexe et de points intérieurs permettent d'accélérer la résolution par décomposition du problème initial (décomposition de Benders [Geoffrion, 1972], décomposition de

Dantzig-Wolfe [Lasdon, 2002], par exemple). En outre, ces algorithmes peuvent être appliqués sur le problème primal ou dual (voir la notion de dualité en programmation linéaire dans [Allaire, 2005 ; Gass, 2003]). Les complexités temporelles de ces algorithmes sont généralement polynomiales en la taille du problème d'optimisation. Le solveur CPLEX LP développé par IBM est un des outils qui implémentent ces algorithmes [Cplex, 2009]. Il peut être utilisé avec GAMS, un langage informatique de haut niveau pour les problèmes d'optimisation [Rosenthal, 2004]. Les algorithmes de résolution sont aussi implémentés dans des librairies libres dans la plupart des langages de programmation (Python, R, C, Matlab etc.).

Plus de détails sur la programmation linéaire et les méthodes de résolution numérique sont présentés dans [Solow, 2014 ; Gass, 2003] où elles sont illustrés avec des exemples. En pratique, comme les algorithmes de résolution des problèmes LP sont très efficaces par rapport à ceux des problèmes d'optimisation plus généraux, on préfère souvent linéariser les problèmes d'optimisation si possible [Pshenichnyj, 2012].

3.1.3 Programmation linéaire mixte

On dit que le problème LP (3.3) est *mixte* (ou MILP, pour *Mixed-Integer Linear Programming*) si certaines variables de décision sont discrètes (i.e. entières). L'ensemble admissible d'un problème MILP s'écrit donc :

$$\mathcal{X} = \{\mathbf{x} = (\mathbf{x}_j)_{1\leq j\leq r} \in \mathbb{R}^r \,;\; W\mathbf{x} \leq w \text{ et } \forall j \in \mathcal{Z},\; \mathbf{x}_j \in \mathbb{Z}\},$$

où $\mathcal{Z} \subset \{1, \ldots, r\}$. Le problème MILP peut alors se noter :

$$\max_{\substack{W\mathbf{x}\leq w \\ \mathbf{x}_j\in\mathbb{Z},\ j\in\mathcal{Z} \\ \mathbf{x}_j\in\mathbb{R},\ j\in\{1,\ldots,r\}\setminus\mathcal{Z}}} \gamma^\top \mathbf{x}. \tag{3.4}$$

Une contrainte de type $g(\mathbf{x}) \leq 0$, pour $\mathbf{x} \in \mathbb{R}^r$, avec g une fonction linéaire par morceaux de $\mathbb{R}^r$ dans $\mathbb{R}$ est équivalente aux contraintes linéaires mixtes suivantes [1] :

$$\begin{cases} \sum_{\sigma=1}^{S} \mathbf{y}_\sigma g_\sigma \leq 0, \\ \sum_{\sigma=1}^{S} \mathbf{y}_\sigma x_\sigma = \mathbf{x}, \\ \sum_{\sigma=1}^{S} \mathbf{y}_\sigma = 1, \\ \sum_{\sigma=1}^{S} \mathbf{z}_\sigma = 1, \\ \forall \sigma \in \{2, \ldots, S\},\; \mathbf{y}_\sigma \leq \mathbf{z}_{\sigma-1} + \mathbf{z}_\sigma, \\ \mathbf{y}_1 \leq \mathbf{z}_1, \\ \forall \sigma \in \{2, \ldots, S\},\; \mathbf{y}_\sigma \geq 0,\; \mathbf{z}_\sigma \in \{0, 1\}, \end{cases}$$

où $\mathbf{y}_\sigma$ et $\mathbf{z}_\sigma$, pour $\sigma \in \{1, \ldots, S\}$ ($S \in \mathbb{N}^*$), sont des variables de décision, respectivement continues et binaires, $x_\sigma \in \mathbb{R}^r$ et $g_\sigma = g(x_\sigma) \in \mathbb{R}$, pour $\sigma \in \{1, \ldots, S\}$ ($S \in \mathbb{N}^*$), sont les paramètres qui correspondent respectivement aux points de rupture de la fonction linéaire par morceaux g et les valeurs associées. Ainsi, les problèmes d'optimisation sous contraintes linéaires par morceaux sont des exemples de problèmes MILP.

En partitionnant l'ensemble admissible selon les valeurs des variables discrètes et en utilisant la propriété 3.3, on dispose du résultat d'existence suivant.

1. Voir `https://download.aimms.com/aimms/download/manuals/AIMMS3OM_IntegerProgrammingTricks.pdf`

Propriété 3.5. *Si l'ensemble $\mathcal{X}$ est non vide et borné, alors le problème MILP* (3.4) *admet (au moins) une solution optimale.*

L'utilisation de certaines variables discrètes complexifie le problème d'optimisation et sa résolution numérique [Allaire, 2005]. La *relaxation* du problème MILP (3.4) consiste à remplacer les variables discrètes par des variables continues pour arriver au problème LP (3.3), plus facile à résoudre. La valeur optimale du problème relaxé est donc une borne supérieure de la valeur optimale du problème MILP.

L'algorithme principalement utilisé pour la résolution numérique du problème MILP est celui de coupes et branchements (*branch and cut*) [Mitchell, 2002 ; Albert, 1999]. C'est notamment un algorithme du solveur CPLEX MIP développé par IBM [Cplex, 2009] qui peut être utilisé avec le langage GAMS. L'algorithme revient à résoudre une succession de sous-problèmes LP, chacun d'eux étant résolus avec un algorithme LP (simplexe dual par exemple). Chaque sous-problème correspond à un problème LP relaxé obtenu par la méthode de coupe (*cutting plane*) [Hugues *et al.*, 2002] (ajout de contraintes successives pour réduire la taille de l'ensemble admissible des problèmes LP, sans retirer les solutions mixtes possibles). L'enchaînement des sous-problèmes (création des « branchements », i.e. des partitions de l'ensemble admissible en fonction des valeurs possibles des variables discrètes) et l'arbre de parcours qui fait le lien entre les sous-problèmes est obtenu par le principe de séparation et évaluation (*branch and bound*) [Lawler et Wood, 1966]. Le premier sous-problème (racine de l'arbre de parcours de l'algorithme) correspond au problème totalement relaxé. À chaque nœud de l'arbre, la résolution du sous-problème LP associé permet d'obtenir une valeur optimale (borne supérieure de la valeur optimale du problème MILP) qui est comparée à la valeur de la fonction objectif du problème MILP de la dernière solution admissible (pour le problème MILP) trouvée. On appelle *MIP-gap* (absolu) la différence absolue entre ces deux valeurs. Le MIP-gap relatif est par définition le rapport du MIP-gap absolu sur la valeur de la dernière solution admissible trouvée. Un MIP-gap égal à zéro signifie que la dernière solution admissible trouvée est optimale pour le problème MILP. Le MIP-gap diminue à chaque nouveau niveau de l'arbre de parcours, ce qui permet de mesurer la convergence de l'algorithme. En pratique, on utilise généralement un critère d'arrêt sur le MIP-gap (absolu ou relatif). Il est également possible d'arrêter l'algorithme à la première solution admissible trouvée ou la dernière solution admissible trouvée après un temps de calcul limite. Dans ces deux cas, il est important de mesurer la valeur du MIP-gap pour se faire une idée de l'optimalité de la solution renvoyée. Il convient de noter que le principe de séparation et évaluation (*branch and bound*) suffit, en théorie, à résoudre un problème MILP. Néanmoins, le temps de calcul dépend alors de la qualité des bornes supérieures trouvées à chaque niveau de l'arbre de parcours. Le fait de combiner ce principe avec une méthode de coupe (*cutting plane*) permet d'améliorer les bornes supérieures et donc d'améliorer le temps de calcul.

Les algorithmes de résolution des problèmes MILP sont moins efficaces que ceux des problèmes LP (car résoudre un problème MILP revient à résoudre une succession de problèmes LP) mais ils sont généralement plus efficaces que les algorithmes de résolution des problèmes d'optimisation généraux. Si la résolution du problème LP après linéarisation n'est pas satisfaisante (hypothèse linéaire trop forte), alors il est intéressant de résoudre le problème MILP obtenu par linéarisation par morceaux du problème d'optimisation général.

3.1.4 Optimisation sous incertitudes

Dans les sections précédentes, les problèmes d'optimisation sont présentés dans un cadre déterministe : les décisions (i.e. le choix des variables de décision) sont effectuées en connaissant parfaitement les paramètres du problème d'optimisation (fonction objectif f, fonctions g_i, pour $i \in \{1, \ldots, s\}$, pour l'optimisation générale, matrice W et vecteur w pour l'optimisation linéaire et linéaire mixte). À l'inverse, l'optimisation sous incertitudes consiste à considérer un cadre pro-

babiliste où les décisions sont effectuées lorsque certains paramètres du problème d'optimisation sont incertains, i.e. qui ne sont pas parfaitement connus au moment où le problème d'optimisation est résolu. C'est par exemple le cas lorsque l'optimisation porte sur des événements futurs qui ne peuvent pas être parfaitement prévus.

Nous présentons dans cette sous-section un rapide état de l'art des principales adaptations de la formulation mathématique des problèmes d'optimisation sous contraintes pour les faire passer de l'univers déterministe à l'univers probabiliste. Ces adaptations se distinguent principalement par la manière dont les incertitudes sont prises en compte. Il existe en effet une très grande variété de philosophies de modélisation du problème sous incertitudes, selon notamment si l'on souhaite utiliser les incertitudes pour augmenter la résilience (robustesse) de la solution optimale ou pour gérer les risques en termes de gains ou coûts. Par ailleurs, les algorithmes de résolution sont généralement développés de manière spécifique à chacune des différentes modélisations en adaptant les algorithmes utilisés en univers déterministe. Pour un état de l'art plus complet sur l'optimisation sous incertitudes, voir par exemple [Sahinidis, 2004 ; Bakker *et al.*, 2020 ; Haneveld *et al.*, 2019 ; Shapiro *et al.*, 2021]. Pour plus de détails sur le lien entre la modélisation des incertitudes et leur prise en compte dans l'optimisation, voir par exemple [King et Wallace, 2012].

Dans cette sous-section, on désigne par $\xi \in \mathbb{R}^L$ ($L \in \mathbb{N}^*$) le vecteur des paramètres considérés comme incertains du problème d'optimisation général sous contraintes (3.2). On suppose, de manière générale, que la fonction objectif f et les fonctions g_i, pour $i \in \{s'+1, \ldots, s\}$ (où $s' \in \{0, \ldots, s-1\}$), dépendent du vecteur ξ. En particulier, l'ensemble admissible $\mathcal{X}$ dépend également du vecteur ξ. Les fonctions g_i, pour $i \in \{1, \ldots, s'\}$, sont supposées ne pas dépendre du vecteur ξ. Si le vecteur ξ était parfaitement connu (univers déterministe), alors le problème d'optimisation (3.2) s'écrirait :

$$\max_{\mathbf{x} \in \mathcal{X}(\xi)} f(\mathbf{x}, \xi), \tag{3.5}$$

où

$$\mathcal{X}(\xi) = \left(\bigcap_{1 \leq i \leq s'} \{\mathbf{x} \in \mathbb{R}^r \,;\; g_i(\mathbf{x}) \leq 0\} \right) \cap \left(\bigcap_{s'+1 \leq i \leq s} \{\mathbf{x} \in \mathbb{R}^r \,;\; g_i(\mathbf{x}, \xi) \leq 0\} \right).$$

3.1.4.1 Approche de type Monte-Carlo

Dans certains cas, il est possible de prendre en compte les incertitudes en dehors du problème d'optimisation avec une approche de type Monte-Carlo. On suppose pour cela disposer d'un échantillon équiprobable $(\xi_1, \ldots, \xi_N)$ de réalisations possibles des paramètres incertains ($N \in \mathbb{N}^*$).

En appliquant le problème d'optimisation déterministe (3.5) à chacun des N réalisations possibles des paramètres incertains, on obtient N solutions intermédiaires $\mathbf{x}_n \in \mathcal{X}(\xi_n)$, pour $n \in \{1, \ldots, N\}$, maximisant la fonction objectif $\mathbf{x} \mapsto f(\mathbf{x}, \xi_n)$.

Le principe de l'approche de type Monte-Carlo est de former une solution $\mathbf{x}^* \in \mathbb{R}^r$ à partir des solutions intermédiaires $\mathbf{x}_n \in \mathcal{X}(\xi_n)$, pour $n \in \{1, \ldots, N\}$, voire aussi de leurs valeurs optimales $f(\mathbf{x}_n, \xi_n)$. Il peut s'agir, par exemple, de déterminer des dépendances générales entre les variables de décision (stratégies de décisions anticipatives) en fonction de la réalisation des paramètres incertains grâce à des modèles de régression. Ces dépendances peuvent ensuite être utilisées en tant que nouvelles contraintes, puis le nouveau problème d'optimisation ainsi obtenu peut être résolu avec la moyenne $\overline{\xi}$ de l'échantillon équiprobable $(\xi_1, \ldots, \xi_N)$:

$$\overline{\xi} = \frac{1}{N} \sum_{n=1}^{N} \xi_n \in \mathbb{R}^L.$$

Le nouveau problème d'optimisation ainsi obtenu étant en univers déterministe, il peut être résolu avec les algorithmes présentés aux sous-sections précédentes.

Un exemple simple consiste à définir la solution $\mathbf{x}^* \in \mathbb{R}^r$ comme la moyenne des solutions intermédiaires :

$$\mathbf{x}^* = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_n = \overline{\mathbf{x}} \in \mathbb{R}^r.$$

Pour que cette solution moyenne soit utilisable, il faut qu'elle soit admissible pour au moins une réalisation possible des paramètres incertains. C'est par exemple le cas lorsque les fonctions g_i, pour $i \in \{1, \ldots, s\}$, sont convexes (par rapport à toutes leurs composantes) car, dans ce cas, $\overline{\mathbf{x}} \in \mathcal{X}(\overline{\xi})$, où $\overline{\xi} \in \mathbb{R}^L$ est la moyenne de l'échantillon équiprobable $(\xi_1, \ldots, \xi_N)$. De manière analogue, si la fonction f est concave par rapport à ses deux composantes, alors

$$\frac{1}{N}\sum_{n=1}^{N} f(\mathbf{x}_n, \xi_n) \leq f(\overline{\mathbf{x}}, \overline{\xi}),$$

donc les maximisations des solutions intermédiaires permettent de contrôler l'optimalité de la solution moyenne par rapport à la moyenne $\overline{\xi}$. Cet exemple n'est pas très pertinent pour viser l'optimalité par rapport à la moyenne $\overline{\xi}$ car, pour cela, il serait préférable de résoudre directement le problème d'optimisation déterministe (3.5) avec le scénario moyen $\overline{\xi}$:

$$f(\overline{\mathbf{x}}, \overline{\xi}) \leq \max\{f(\mathbf{x}, \overline{\xi}),\ \mathbf{x} \in \mathcal{X}(\overline{\xi})\}.$$

Néanmoins, la solution moyenne $\overline{\mathbf{x}}$ tient compte des solutions intermédiaires, la rendant *a priori* plus adaptée aux incertitudes que la solution optimale du problème d'optimisation déterministe (3.5) avec le scénario moyen $\overline{\xi}$.

Dans la suite de cette sous-section, nous présentons des approches où les incertitudes sont prises en compte au sein du problème d'optimisation.

3.1.4.2 Optimisation robuste

L'optimisation robuste permet de se prémunir face à un ensemble $\Xi_{\text{rob}} \subset \mathbb{R}^L$ de réalisations possibles des paramètres incertains (parfois appelé *ensemble incertain*), considérées comme défavorables *a priori*, sans regarder leurs probabilités d'occurrence [Ben-Tal *et al.*, 2009]. Les décisions sont alors effectuées de manière à limiter l'impact de ces réalisations défavorables possibles si l'une d'entre elles venait effectivement à se réaliser.

Le principe général de l'optimisation robuste est de fournir des solutions qui font preuve de résilience face aux réalisations défavorables possibles. L'ensemble admissible du problème d'optimisation déterministe (3.5) est alors remplacé par des contraintes dites *robustes* :

$$\bigcap_{\xi \in \Xi_{\text{rob}}} \mathcal{X}(\xi) = \left(\bigcap_{1 \leq i \leq s'} \{\mathbf{x} \in \mathbb{R}^r\,;\ g_i(\mathbf{x}) \leq 0\} \right) \cap \left(\bigcap_{s'+1 \leq i \leq s} \{\mathbf{x} \in \mathbb{R}^r\,;\ \max_{\xi \in \Xi_{\text{rob}}} g_i(\mathbf{x}, \xi) \leq 0\} \right).$$

Ce principe s'applique également pour la fonction objectif. L'optimisation robuste vise à fournir une solution optimale pour les pires valeurs possibles (optimisation pire cas), de manière à s'assurer que le gain de la solution ne prenne pas des valeurs inférieures à la valeur optimale calculée. La fonction objectif du problème d'optimisation déterministe (3.5) est alors remplacée par la fonction :

$$\mathbf{x} \mapsto \min_{\xi \in \Xi_{\text{rob}}} f(\mathbf{x}, \xi).$$

Ainsi, la formulation mathématique de l'optimisation robuste associée au problème d'optimisation déterministe (3.5) est la suivante :

$$\max_{\substack{g_i(\mathbf{x})\leq 0,\ 1\leq i\leq s' \\ g_i(\mathbf{x},\xi)\leq 0,\ \xi\in\Xi_{\text{rob}},\ s'+1\leq i\leq s \\ \mathbf{x}\in\mathbb{R}^r}} \left\{ \min_{\xi\in\Xi_{\text{rob}}} f(\mathbf{x},\xi) \right\}.$$

L'optimisation robuste repose sur le choix de l'ensemble $\Xi_{\text{rob}} \subset \mathbb{R}^L$ des réalisations défavorables des paramètres incertains. Plusieurs approches sont proposées dans la littérature avec des algorithmes spécifiques [Ben-Tal *et al.*, 2009]. Par exemple, si le problème d'optimisation déterministe (3.5) est (MI)LP et si l'ensemble $\Xi_{\text{rob}} = \{\xi_1, \ldots, \xi_N\} \subset \mathbb{R}^L$ est fini ($N \in \mathbb{N}^*$), alors le problème d'optimisation robuste associé est aussi un problème (MI)LP, pouvant être résolu avec un solveur (MI)LP :

$$\max_{\substack{\mathbf{x}\in\mathcal{X}(\xi_n),\ 1\leq n\leq N \\ \mathbf{z}\leq f(\mathbf{x},\xi_n),\ 1\leq n\leq N \\ (\mathbf{x},\mathbf{z})\in\mathbb{R}^r\times\mathbb{R}}} \mathbf{z}.$$

La difficulté dans ce cas est que la taille du problème (MI)LP augmente linéairement avec la taille N de l'ensemble Ξ_{rob}, pouvant rendre la résolution numérique peu accessible en pratique.

3.1.4.3 Optimisation probabiliste

L'optimisation probabiliste ou la programmation probabiliste dans le cas linéaire ou linéaire mixte consiste à considérer que le vecteur ξ des paramètres incertains est une réalisation d'une variable aléatoire Ξ sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathbb{R}^L$. Contrairement à l'optimisation robuste, l'optimisation probabiliste repose donc sur la connaissance de la loi de probabilité des paramètres incertains donnée par la variable aléatoire Ξ.

L'idée de l'optimisation probabiliste est de remplacer les informations données par le vecteur ξ par celles données par la variable aléatoire Ξ. La difficulté est que la fonction $\mathbf{x} \mapsto f(\mathbf{x}, \Xi)$ et les fonctions $\mathbf{x} \mapsto g_i(\mathbf{x}, \Xi)$, pour $i \in \{s'+1, \ldots, s\}$, n'ont pas des valeurs réelles mais des valeurs qui sont des variables aléatoires réelles. Il faut alors préciser ce que signifie des inégalités du type $g_i(\mathbf{x}, \Xi) \leq 0$, pour $i \in \{s'+1, \ldots, s\}$ et $\mathbf{x} \in \mathbb{R}^r$, ou $f(\mathbf{x}, \Xi) \leq f(\mathbf{y}, \Xi)$, pour $(\mathbf{x}, \mathbf{y}) \in (\mathbb{R}^r)^2$, qui n'ont pas de sens *a priori*. Autrement dit, il faut réécrire la fonction objectif et les contraintes de manière à formuler un problème d'optimisation portant sur une fonction objectif et des contraintes non aléatoires (i.e. en univers déterministe mais en utilisant la variable aléatoire Ξ).

Pour prendre en compte les incertitudes qui interviennent dans la fonction objectif, l'approche la plus courante est d'appliquer l'espérance (si elle existe). La nouvelle fonction objectif (réelle) est donc :

$$\mathbf{x} \mapsto \mathbb{E}[f(\mathbf{x}, \Xi)] \in \mathbb{R},$$

où $\mathbb{E}$ désigne l'espérance pour des variables aléatoires réelles sur $(\Omega, \mathcal{F}, \mathbb{P})$. Cela suppose que les variables aléatoires réelles $f(\mathbf{x}, \Xi)$, pour $\mathbf{x} \in \mathbb{R}^r$, sont d'espérance finie. Par exemple, si la variable aléatoire Ξ est discrète et finie avec $\Xi(\Omega) = \{\xi_1, \ldots, \xi_N\} \subset \mathbb{R}^L$ ($N \in \mathbb{N}^*$), alors :

$$\forall \mathbf{x} \in \mathbb{R}^r, \quad \mathbb{E}[f(\mathbf{x}, \Xi)] = \sum_{n=1}^{N} p_n f(\mathbf{x}, \xi_n),$$

où $p_n = \mathbb{P}[\Xi = \xi_n] \in [0, 1]$, pour $n \in \{1, \ldots, N\}$. L'utilisation de l'espérance suppose également une aversion neutre face au risque. Pour tenir compte de l'aversion au risque, il est possible de remplacer l'espérance par un quantile, un expectile ou une mesure de risque, comme la valeur du risque conditionnel (CVaR pour *Conditional Value at Risk* en anglais) qui représente le gain moyen parmi ceux qui sont inférieurs à un quantile donné (i.e. la queue de distribution des pires

cas) [Serraino et Uryasev, 2013]. L'avantage de la valeur du risque conditionnel CVaR est qu'elle peut être linéarisée et écrite sous la forme d'un problème d'optimisation linéaire [Rockafellar et Uryasev, 2002 ; Pflug, 2000]. Certains auteurs suggèrent aussi d'utiliser l'espérance et de pénaliser la variance (si elle existe), en considérant la fonction objectif :

$$\mathbf{x} \mapsto \mathbb{E}[f(\mathbf{x}, \Xi)] - \lambda \operatorname{Var}[f(\mathbf{x}, \Xi)],$$

où $\lambda \in \mathbb{R}_+$ est un coefficient de pénalisation lié à la tolérance au risque et Var désigne la variance pour des variables aléatoires réelles sur $(\Omega, \mathcal{F}, \mathbb{P})$. Plus le coefficient de pénalisation est élevé, plus la variance tolérée sera faible. À l'inverse, plus le coefficient de pénalisation est petit, plus l'espérance de la solution optimale sera élevée. Une autre manière serait d'effectuer une optimisation multi-objectif pour maximiser l'espérance et minimiser la variance.

En ce qui concerne la prise en compte des incertitudes qui interviennent dans les contraintes, l'approche la plus courante est d'utiliser des contraintes en probabilité (*chance constraints* en anglais [Charnes et Cooper, 1959]), en se donnant des probabilités seuils maximales de non respect des contraintes. Ces probabilités seuils correspondent à des valeurs de risque que l'on s'accorde sur les contraintes. Plus elles sont proches de zéro, moins on veut prendre le risque que les contraintes ne soient pas respectées. Le nouvel ensemble admissible est donc défini par :

$$\left(\bigcap_{1 \leq i \leq s'} \{\mathbf{x} \in \mathbb{R}^r \,;\; g_i(\mathbf{x}) \leq 0\} \right) \cap \left(\bigcap_{s'+1 \leq i \leq s} \{\mathbf{x} \in \mathbb{R}^r \,;\; \mathbb{P}[g_i(\mathbf{x}, \Xi) \leq 0] \geq 1 - \varepsilon_i\} \right),$$

où $\varepsilon_i \in [0, 1]$, pour $i \in \{s' + 1, \ldots, s\}$, est la probabilité seuil maximale associée à la contrainte définie avec la fonction g_i. Il convient de noter que si la variable aléatoire Ξ est discrète finie avec $\Xi(\Omega) = \{\xi_1, \ldots, \xi_N\} \subset \mathbb{R}^L$ ($N \in \mathbb{N}^*$) et si $\varepsilon_i = 1$, pour $i \in \{s' + 1, \ldots, s\}$, alors ce nouvel ensemble admissible coïncide avec celui de l'optimisation robuste en considérant que tous les événements possibles sont défavorables. Autrement dit, l'optimisation robuste peut être vue comme l'utilisation de contraintes presque sûres.

Avec ces approches, la formulation mathématique en univers probabiliste du problème d'optimisation déterministe (3.5) est la suivante :

$$\max_{\substack{g_i(\mathbf{x}) \leq 0,\ 1 \leq i \leq s' \\ \mathbb{P}[g_i(\mathbf{x}, \Xi) \leq 0] \geq 1 - \varepsilon_i,\ s'+1 \leq i \leq s \\ \mathbf{x} \in \mathbb{R}^r}} \mathbb{E}[f(\mathbf{x}, \Xi)].$$

La résolution numérique d'un tel problème dépend de la modélisation de la variable aléatoire Ξ. Dans le cas général (loi de probabilité jointe continue quelconque), elle est difficilement accessible [Haneveld et van der Vlerk, 2020]. Un cas simple courant où la résolution numérique est accessible est lorsque les paramètres incertains suivent des lois normales indépendantes et que le problème d'optimisation déterministe (3.5) est (MI)LP [Ben-Tal *et al.*, 2009]. Lorsque les lois de probabilité ne sont pas normales et indépendantes, il peut donc être intéressant de simplifier la modélisation des incertitudes en estimant des lois normales indépendantes qui approchent les lois marginales de la variable aléatoire Ξ [Wu *et al.*, 2014 ; Mesfin et Shuhaimi, 2010].

3.1.4.4 Optimisation stochastique à deux niveaux

Les contraintes en probabilité de l'optimisation probabiliste précédente donnent la possibilité de ne pas respecter certaines contraintes selon une mesure de risque donnée. Cela accorde une souplesse dans le choix de la solution que l'optimisation robuste ne permet pas. Il est possible d'aller encore plus loin en autorisant complètement le non respect des contraintes mais en pénalisant la fonction objectif pour contenir (ou minimiser) les violations des contraintes [Wang et Spall,

2008 ; Haneveld et van der Vlerk, 2020]. À partir du problème d'optimisation déterministe (3.5), on obtient alors la formulation probabiliste suivante :

$$\max_{\substack{g_i(\mathbf{x})\leq 0,\ 1\leq i\leq s' \\ \mathbf{x}\in\mathbb{R}^r}} \left\{ \mathbb{E}[f(\mathbf{x},\Xi)] + \sum_{i=s'+1}^{s} \mathbb{E}[\rho(g_i(\mathbf{x},\Xi))] \right\},$$

où $\rho \in \mathcal{F}(\mathbb{R},\mathbb{R}_-)$ est une fonction de pénalisation. Si la fonction ρ est définie par

$$\forall u \in \mathbb{R}, \quad \rho(u) = \begin{cases} 0 & \text{si } u \leq 0, \\ -\infty & \text{si } u > 0, \end{cases}$$

alors on retrouve le problème d'optimisation robuste où l'ensemble incertain est le support de la variable aléatoire Ξ. Pour conserver une souplesse dans le non respect des contraintes, un exemple simple de fonction de pénalisation est celle définie par

$$\forall u \in \mathbb{R}, \quad \rho(u) = \begin{cases} 0 & \text{si } u \leq 0, \\ -\lambda u & \text{si } u > 0, \end{cases}$$

avec $\lambda \in \mathbb{R}_+$. Cette approche par pénalisation dépend principalement du choix de la fonction de pénalisation, qui est délicat en pratique.

L'optimisation stochastique à deux niveaux (ou l'optimisation avec recours) rejoint l'idée d'autoriser le non respect de certaines contraintes en pénalisant la fonction objectif. En revanche, les pénalités sont déterminées en considérant les corrections qui devront être apportées à la solution pour respecter les contraintes une fois les incertitudes dévoilées. Cela suppose donc d'avoir la possibilité d'adapter les décisions à mesure que les incertitudes se dévoilent. Pour cela, les décisions sont réparties en deux étapes, une première étape avant la réalisation des paramètres incertains et une seconde étape après la réalisation des paramètres incertains. La première étape sert principalement à effectuer les décisions qui ne pourront plus être modifiées en maximisant les gains *a priori*, alors que la seconde étape, appelé le *recours*, sert principalement à effectuer les autres décisions de manière à assurer l'admissibilité de la solution en fonction des décisions de la première étape et de la réalisation des paramètres incertains. L'optimisation avec recours consiste alors à effectuer les décisions de la première étape tout en anticipant les coûts moyens du recours. Généralement, seule la seconde étape dépend des paramètres incertains. On suppose donc que la fonction f ne dépend pas de la variable aléatoire Ξ. La formulation stochastique à deux niveaux du problème d'optimisation déterministe (3.5) est la suivante :

$$\max_{\substack{g_i(\mathbf{x})\leq 0,\ 1\leq i\leq s' \\ \mathbf{x}\in\mathbb{R}^r}} \{ f(\mathbf{x}) + \mathbb{E}[Q(\mathbf{x},\Xi)] \},$$

où, pour tout $(\mathbf{x},\xi) \in \mathbb{R}^r \times \mathbb{R}^L$,

$$Q(\mathbf{x},\xi) = \max_{\mathbf{y}\in\mathcal{Y}(\mathbf{x},\xi)} q(\mathbf{y},\mathbf{x},\xi),$$

avec q une fonction réelle définie sur $\mathbb{R}^{r'} \times \mathbb{R}^r \times \mathbb{R}^L$ ($r' \in \mathbb{N}^*$) et $\mathcal{Y}(\mathbf{x},\xi) \subset \mathbb{R}^{r'}$ est défini par des contraintes faisant intervenir la solution $\mathbf{x}$ du premier niveau et la réalisation ξ des paramètres incertains vérifiant

$$\bigcap_{s'+1\leq i\leq s} \{ \mathbf{y} \in \mathbb{R}^{r'} ;\ g_i(\mathbf{x},\xi) + h_i(\mathbf{y},\xi) \leq 0 \} \subset \mathcal{Y}(\mathbf{x},\xi),$$

h_i, pour $i \in \{s'+1,\ldots,s\}$, étant des fonctions réelles définies sur $\mathbb{R}^{r'} \times \mathbb{R}^L$.

Avec ces notations, $\mathbf{x} \in \mathbb{R}^r$ correspond aux décisions de la première étape et $\mathbf{y} \in \mathbb{R}^{r'}$ à celles de la seconde étape. Les termes $h_i(\mathbf{y}, \xi)$, pour $i \in \{s' + 1, \ldots, s\}$, mesurent les éventuelles violations des contraintes $g_i(\mathbf{x}, \xi) \leq 0$, et $q(\mathbf{y}, \mathbf{x}, \xi)$ représente le gain algébrique (pouvant être négatif) dû à la décision $\mathbf{y} \in \mathbb{R}^{r'}$ de la seconde étape pour compenser ces éventuelles violations des contraintes connaissant la réalisation ξ des paramètres incertains. Bien sûr, il est aussi possible de simplement considérer deux problèmes d'optimisation imbriqués, sans que le second niveau s'interprète comme un recours pour compenser d'éventuelles violations des contraintes du premier niveau.

Si la variable aléatoire Ξ est discrète et finie avec $\Xi(\Omega) = \{\xi_1, \ldots, \xi_N\} \subset \mathbb{R}^L$ ($N \in \mathbb{N}^*$), alors le problème d'optimisation stochastique à deux niveaux peut s'écrire sous une forme déterministe équivalente :

$$\max_{\substack{g_i(\mathbf{x}) \leq 0,\ 1 \leq i \leq s' \\ \mathbf{y} \in \mathcal{Y}(\mathbf{x}, \xi_n),\ 1 \leq n \leq N \\ (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^r \times \mathbb{R}^{r'}}} \left\{ f(\mathbf{x}) + \sum_{n=1}^{N} p_n q(\mathbf{y}, \mathbf{x}, \xi_n) \right\}, \tag{3.6}$$

où $p_n = \mathbb{P}[\Xi = \xi_n] \in [0, 1]$, pour $n \in \{1, \ldots, N\}$. Si la variable aléatoire Ξ n'est pas discrète et finie en pratique, alors il est possible d'appliquer un tirage de type Monte-Carlo (*Sample Average Approximation*) [Ahmed *et al.*, 2002 ; Kleywegt *et al.*, 2002]. La résolution numérique du problème d'optimisation équivalent (3.6) s'effectue avec les algorithmes présentés aux sous-sections précédentes. La difficulté est que la taille du problème déterministe équivalent augmente linéairement avec le nombre N des réalisations possibles des paramètres incertains, pouvant rendre la résolution numérique peu accessible en pratique.

Lorsque les deux niveaux font intervenir des problèmes (MI)LP, on parle de programmation stochastique à deux niveaux (ou de programmation stochastique avec recours). La littérature à ce sujet est très riche car beaucoup de problème d'optimisation sous incertitudes peuvent se formuler de cette façon [Shapiro *et al.*, 2021 ; Sahinidis, 2004]. Dans ce cas, il est possible d'adapter les algorithmes MILP de manière à tirer profit de la structure en scénarios pour rendre la résolution numérique plus efficace (*Progressive Hedging*, décomposition de Dantzig-Wolfe, décomposition de Benders, décomposition duales, ...) [Carøe et Schultz, 1999 ; Rockafellar et Wets, 1991 ; Watson et Woodruff, 2011 ; Carpentier *et al.*, 2018, 2013 ; van Ackooij et Malick, 2016 ; Gröwe-Kuska *et al.*, 2002 ; Tahanan *et al.*, 2015 ; Schulze *et al.*, 2017 ; Finardi et Scuzziato, 2014].

L'optimisation stochastique à deux niveaux peut être étendue à plusieurs niveaux lorsque les décisions peuvent être effectuées à plusieurs étapes à mesure que les incertitudes se dévoilent. On parle alors d'optimisation stochastique multi-niveaux. Lorsque la variable aléatoire Ξ est discrète et finie, on forme ainsi un arbre de décision associé à un arbre de scénarios des paramètres incertains [Defourny *et al.*, 2012].

3.1.4.5 Optimisation dynamique stochastique

L'optimisation stochastique multi-niveaux est généralement définie avec l'hypothèse de non anticipativité : les décisions sont effectuées à chaque niveau en connaissant uniquement la réalisation des paramètres incertains des niveaux précédents mais pas ceux des niveaux suivants. Ainsi, à chaque niveau, les décisions sont seulement effectuées en fonction de l'état du système au niveau considéré (ici, le système désigne ce qui est modélisé par le problème d'optimisation).

Pour le cas de l'optimisation stochastique à deux niveaux, l'hypothèse de non anticipativité entraîne que l'état $\mathbf{s}_2$ du système au second niveau est une fonction de l'état $\mathbf{s}_1$ du système au premier niveau, des décisions $\mathbf{x}$ du premier niveau et de la réalisation $\xi \in \mathbb{R}^L$ des paramètres

incertains (cette réalisation est dévoilée entre les deux niveaux) :

$$\mathbf{s}_2 = \sigma(\mathbf{s}_1, \mathbf{x}, \xi).$$

Cette équation exprime la dynamique entre les états des deux niveaux. Les décisions $\mathbf{x} \in \mathbb{R}^r$ du premier niveau sont effectuées à partir de l'état $\mathbf{s}_1$, dont l'ensemble admissible associé est noté $\mathcal{X}(\mathbf{s}_1) \subset \mathbb{R}^r$. De même, les décisions $\mathbf{y} \in \mathbb{R}^{r'}$ du second niveau sont effectuées à partir de l'état $\mathbf{s}_2$, dont l'ensemble admissible associé est noté $\mathcal{Y}(\mathbf{s}_2) \subset \mathbb{R}^{r'}$. Avec les notations de l'optimisation stochastique à deux niveaux, on a donc :

$$\mathcal{X}(\mathbf{s}_1) = \bigcap_{1 \leq i \leq s'} \{\mathbf{x} \in \mathbb{R}^r \,;\; g_i(\mathbf{x}) \leq 0\} \quad \text{et} \quad \mathcal{Y}(\mathbf{s}_2) = \mathcal{Y}(\mathbf{x}, \xi).$$

De manière analogue, le gain à optimiser au premier niveau s'écrit $f_1(\mathbf{s}_1, \mathbf{x})$ car il dépend uniquement de l'état $\mathbf{s}_1$ et des décisions $\mathbf{x} \in \mathbb{R}^r$. De même, le gain à optimiser au second niveau s'écrit $f_2(\mathbf{s}_2, \mathbf{y}, \xi)$ car il dépend uniquement de l'état $\mathbf{s}_2$, des décisions $\mathbf{y} \in \mathbb{R}^{r'}$ et de la réalisation $\xi \in \mathbb{R}^L$ des paramètres incertains. Avec les notations de l'optimisation stochastique à deux niveaux, on a donc :

$$f_1(\mathbf{s}_1, \mathbf{x}) = f(\mathbf{x}) \quad \text{et} \quad f_2(\mathbf{s}_2, \mathbf{y}, \xi) = q(\mathbf{y}, \mathbf{x}, \xi).$$

Le problème d'optimisation stochastique à deux niveaux peut donc s'écrire sous une formulation dite d'optimisation dynamique stochastique (ou programmation dynamique stochastique dans le cas linéaire ou linéaire mixte) :

$$\max_{\substack{\mathbf{s}_2 = \sigma(\mathbf{s}_1, \mathbf{x}, \Xi) \\ \mathbf{x} \in \mathcal{X}(\mathbf{s}_1) \\ \mathbf{y} \in \mathcal{Y}(\mathbf{s}_2)}} \{f_1(\mathbf{s}_1, \mathbf{x}) + \mathbb{E}[f_2(\mathbf{s}_2, \mathbf{y}, \Xi)]\}. \tag{3.7}$$

Cette écriture permet de faire correspondre l'optimisation stochastique à deux niveaux avec les équations dites de Bellman [Bellman, 1966] qui décrivent, sous certaines hypothèses, la dynamique entre les niveaux en optimisation dynamique stochastique (processus de décision markovien [Howard, 1960]) :

$$\begin{cases} v_2(s_2) = \max\{\mathbb{E}[f_2(s_2, \mathbf{y}, \Xi)],\ \mathbf{y} \in \mathcal{Y}(s_2)\}, \\ v_1(s_1) = \max\{f_1(s_1, \mathbf{x}) + \mathbb{E}[v_2(\sigma(s_1, \mathbf{x}, \Xi))],\ \mathbf{x} \in \mathcal{X}(s_1)\}, \end{cases}$$

où les fonctions v_1 et v_2 donnent les gains moyens algébriques (i.e. la valeur du problème) respectivement du premier et du second niveau et partant d'un état donné du système.

Lorsque les états du système sont échantillonnés (nombre fini d'états), une manière simple de résoudre le problème d'optimisation dynamique stochastique (3.7) est de d'abord calculer la fonction v_2 pour tous les états du système à la seconde étape, puis ensuite de calculer la fonction v_1 pour tous les états du système à la première étape.

L'optimisation dynamique stochastique est une discipline plus générale que ce qui vient d'être présenté (notamment lorsqu'il y a plus de deux étapes) dont plusieurs algorithmes de résolution spécifique sont présents dans la littérature [Ross, 2014]. Ces algorithmes reposent généralement sur la décomposition en niveaux donnée par les équations de Bellman [2], en adaptant les algorithmes du cadre déterministe, de manière à tiret profit de la dynamique entre les niveaux pour rendre la résolution numérique plus efficace (*Stochastic Dynamic Programming*, *Stochastic Dual Dynamic Programming* (SDDP), la décomposition de Benders, la méthode L-Shaped, la méthode Bundle, ...) [Laporte et Louveaux, 1993 ; Carøe et Tind, 1998 ; Sherali et Fraticelli, 2002 ; Ahmed *et al.*, 2004 ; Kong *et al.*, 2006 ; Sherali et Zhu, 2006 ; van der Laan et Romeijnders, 2021 ; Sen, 2005 ; Zou *et al.*, 2019].

2. Les récurrences données par les équations de Bellman sont souvent parcourues dans les deux sens.

3.1.4.6 Optimisation stochastique ou dynamique ?

L'optimisation stochastique multi-niveaux est la formulation naturelle lorsque les niveaux modélisent les adaptations des décisions du niveau précédent en fonction des incertitudes qui se dévoilent. Les problèmes d'optimisation associés aux niveaux peuvent donc être très différents (systèmes différents à chaque niveau) et les variables de décision sont souvent continues et de dimension relativement grande (ensembles admissibles définis par des contraintes). En raison de l'hypothèse de non anticipativité, on s'intéresse surtout aux décisions du premier niveau qui sont applicables quelle que soit la réalisation des paramètres incertains. Les autres niveaux permettent d'anticiper d'éventuelles corrections dans le choix des décisions du premier niveau. Cette approche crée un arbre de décision à partir d'un arbre de scénarios modélisant les incertitudes.

Au contraire, l'optimisation dynamique stochastique est la formulation naturelle lorsque les décisions sont complétées étape par étape à mesure que les incertitudes se dévoilent (souvent les étapes correspondent à des pas de temps), chaque étape apportant un gain « instantané » supplémentaire. Les problèmes d'optimisation associés aux différentes étapes sont donc ici très similaires (même système à chaque étape mais l'état change) et les décisions sont liées aux états du système qui sont souvent discrets ou de dimension relativement petite. En général, on s'intéresse alors à l'anticipation des décisions à chaque étape mais en sachant que l'optimisation sera refaite à chaque étape (mise à jour des décisions) avec de nouvelles informations précisant les incertitudes. Cette approche crée un arbre de décision à partir des états (incertains) du système et les incertitudes sont généralement modélisées par des lois de probabilité indépendantes.

Les deux formulations peuvent se confondre lorsque l'on souhaite effectuer petit à petit les décisions à différentes étapes où les incertitudes se dévoilent tout en cherchant à ajuster les décisions. La différence entre l'optimisation stochastique à plusieurs niveaux et l'optimisation dynamique stochastique réside alors dans le choix de l'algorithme de résolution.

3.2 Formulation mathématique du problème d'optimisation

Dans cette section, nous présentons la formulation mathématique du problème d'optimisation pour la planification à court terme de la production en univers déterministe. Avec le vocabulaire du chapitre précédent, cela revient à définir l'ensemble des solutions admissibles (avec les variables de décision, les paramètres et les contraintes) et la fonction objectif. Nous nous appuyons pour cela sur le modèle de simulation de la sous-section 2.2.4 (modèle hydraulique et contraintes d'exploitation) et le calcul du chiffre d'affaires de la sous-section 2.3.3.

3.2.1 Description

On rappelle que l'on considère la planification de la production conjointe d'une chaîne hydroélectrique et d'actifs de production variable sur une fenêtre temporelle à un horizon court terme (voir Section 1.5). Les variables de décision et les paramètres du problème d'optimisation que l'on cherche à définir peuvent donc dépendre des aménagements de la chaîne hydroélectrique et du temps.

On note $\mathcal{I} = \{1, \dots, I\}$ l'ensemble des indices des aménagements de la chaîne hydroélectrique ($I \in \mathbb{N}^*$). Pour le cas d'étude, $I = 6$ car seuls les 6 aménagements du Haut-Rhône sont considérés pour l'optimisation (voir Sous-section 2.5.2). Néanmoins, les productions des 12 autres aménagements en aval sont prises en compte dans l'optimisation des 6 aménagements du Haut-Rhône pour les deux raisons principales suivantes :

— elles participent à la puissance totale du périmètre d'équilibre CNR,

— elles imposent des contraintes sur le débit sortant du Haut-Rhône pour assurer la continuité de l'eau.

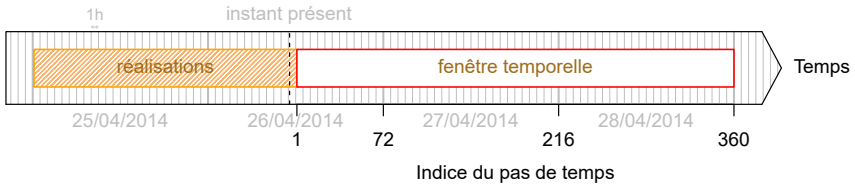


FIGURE 3.1 – Schéma de la fenêtre temporelle du cas d'étude (en rouge) avec la période réalisée permettant de définir les conditions initiales (en orange). Les traits verticaux gris représentent les heures rondes des journées.

Pour le cas d'étude, les productions des 12 autres aménagements sont donc des paramètres pour le problème d'optimisation que l'on cherche à définir. Leurs valeurs sont données par le programme de production initial construit semi-manuellement sur le Rhône entier (voir Sous-section 2.5.1).

On suppose que la fenêtre temporelle est discrétisée en T pas de temps de même durée $\delta t > 0$ ($T \in \mathbb{N}^*$). On note $\mathcal{T} = \{1, \dots, T\}$ l'ensemble des indices des pas de temps. On suppose que la durée δt divise 1 heure et que la fenêtre temporelle couvre au moins le lendemain. En raison des temps de propagation de l'eau, les conditions initiales de la chaîne hydroélectrique sont définies sur une fenêtre temporelle passée qui précède celle de l'optimisation. Certains paramètres du problème d'optimisation sont donc définis sur des pas de temps antérieurs à ceux indexés par l'ensemble $\mathcal{T}$. Nous n'indexons pas ces pas de temps passés dans ce manuscrit car nous n'avons pas besoin d'y faire référence dans la suite, les variables de décision ne les utilisant pas. Pour le cas d'étude, la fenêtre temporelle pour l'optimisation est définie par $\delta t = 10\,\text{min}$ et $T = 360$ sur la période qui débute le 26/04/2014 à 12 h et termine à la fin du 28/04/2014. La fenêtre temporelle passée pour les conditions initiales est la période entre le 25/04/2014 0 h et le 26/04/2014 à 12 h, comme illustré par la figure 3.1 (copie de la figure 2.22 par commodité). On rappelle que l'instant présent réel est le 26/04/2014 à 11 h mais que la fenêtre temporelle débute une heure plus tard (voir Sous-section 2.5.1).

3.2.2 Scénario

Le problème de planification à court terme de la production fait intervenir trois types de paramètres importants (voir Sections 2.3 et 2.4, notamment la figure 2.15) :

— les apports en eau $a = (a(i,t))_{i \in \mathcal{I}, t \in \mathcal{T}} \in \mathbb{R}^{IT}$, somme des débits d'apports des affluents principaux, des débits d'apports des bassins versants intermédiaires et des débits de correction, propagés aux points de réglage des aménagements hydroélectriques et définis sur la fenêtre temporelle, voir Équation (2.4),

— les prix *day-ahead* de l'électricité $\pi = (\pi(t))_{t \in \mathcal{T}} \in \mathbb{R}^T$ définis sur la fenêtre temporelle,

— les productions variables (éoliennes et solaires) $\phi = (\phi(t))_{t \in \mathcal{T}} \in \mathbb{R}^T$ du périmètre d'équilibre, définies sur la fenêtre temporelle.

Ces trois types de paramètres sont des prévisions car ils ne sont pas parfaitement connus à l'instant présent (voir Sous-section 2.3.4). Ces prévisions sont multivariées pour respecter la cohérence spatio-temporelle des apports en eau et la cohérence temporelle des prix *day-ahead* et des productions variables.

Les trois prévisions sont supposées fixées pour la construction du programme de production. Ce sont donc des paramètres pour le problème d'optimisation que l'on cherche à définir. Cette hypothèse est naturelle pour les apports en eau et les productions variables. Elle est justifiée pour les prix *day-ahead* pour le cas d'étude car la production du Rhône n'influe pas le marché (elle est très petite par rapport à toute la production mise en jeu sur le marché *day-ahead*).

Par définition, les prix *day-ahead* sont constants sur chaque heure de la fenêtre temporelle (granularité horaire). Pour le cas d'étude, les apports en eau sont constants sur chaque demi-heure de la fenêtre temporelle (granularité demi-horaire) et les productions variables sont supposées nulles (voir Section 2.5). Un exemple des apports en eau des affluents principaux et des prix *day-ahead* pour le cas d'étude est illustré à la figure 3.2. Pour le cas d'étude, les prévisions des apports en eau commencent le 26/04/2014 à 12 h comme la fenêtre temporelle, et les valeurs entre l'instant présent (26/04/2014 11 h) et le début de la fenêtre temporelle (26/04/2014 12 h) sont les observations. En opérationnel, ces prévisions auraient débuté à l'instant présent.

Dans tout le manuscrit, le triplet $\xi = (a, \pi, \phi) \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ des apports en eau $a \in \mathbb{R}^{IT}$, des prix *day-ahead* $\pi \in \mathbb{R}^T$ et des productions variables $\phi \in \mathbb{R}^T$ est appelé le *scénario* (voir Sous-section 2.4.4).

3.2.3 Solutions admissibles

On suppose que toutes les variables de décision du problème d'optimisation que l'on cherche à définir sont combinées au sein d'un grand vecteur $\mathbf{x} \in \mathbb{R}^r$, où $r \in \mathbb{N}^*$ est le nombre de variables de décision (i.e. la taille du problème), r est en $\mathcal{O}(IT)$. Pour le cas d'étude, $r \approx 500\,000$. Une *solution* du problème de planification à court terme de la production est donc un vecteur $\mathbf{x} \in \mathbb{R}^r$, qui comprend en particulier (voir Section 1.5) :

— les prochaines offres de vente à effectuer au marché *day-ahead*, i.e. la production horaire à mettre en vente sur le marché *day-ahead* et qui sera ajoutée aux prochains engagements,

— le programme de production, i.e. les valeurs allouées à chaque variable de décision qui interviennent dans le modèle de simulation (modèle hydraulique et contraintes d'exploitation) à chaque aménagement hydroélectrique et sur la fenêtre temporelle.

On rappelle que les transactions au marché infrajournalier ne sont pas considérées dans le problème de planification à court terme dans le cadre de notre travail. En outre, les prochaines offres de vente *day-ahead* et le programme de production sont mutuellement dépendants car les écarts entre la production et les engagements sont autorisés mais pénalisés financièrement (voir §2.3.3.2).

L'écriture des contraintes du problème d'optimisation de l'optimiseur du §2.3.5.2 reprend les équations du modèle de simulation de la sous-section 2.2.4. Ainsi, les variables et les paramètres du modèle de simulation sont aussi des variables de décision et des paramètres du problème d'optimisation que l'on cherche à définir. On a donc, par exemple, les contraintes suivantes.

— La contrainte associée au bilan d'eau de l'équation (2.1) s'écrit :

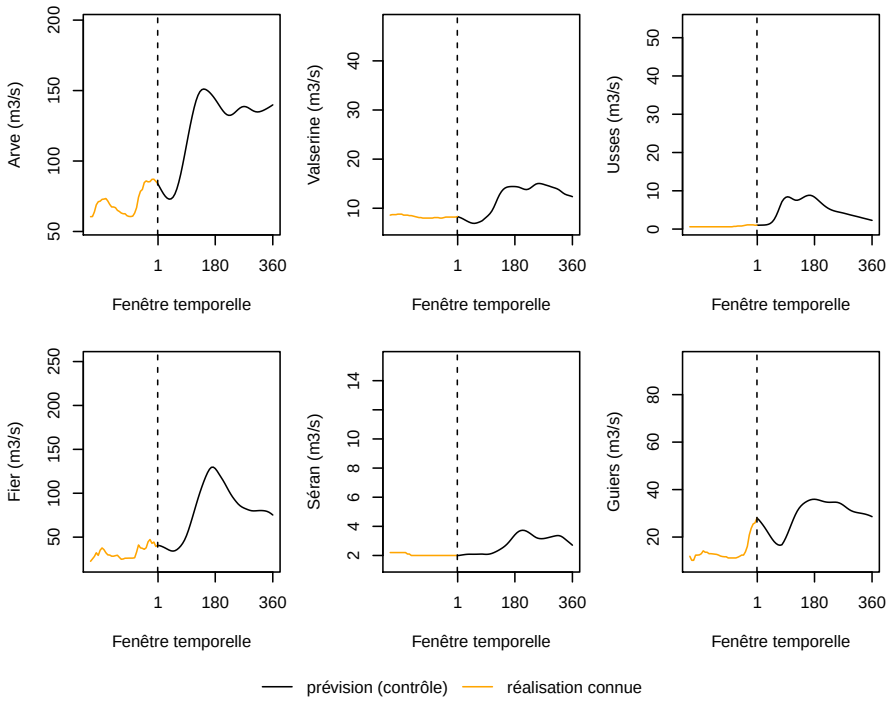
$$\forall (i,t) \in \mathcal{I} \times \mathcal{T}, \quad Q_{\mathrm{ent}\,it}(\mathbf{x}) = Q_{\mathrm{sort}\,it}(\mathbf{x}) + \frac{V_{it}(\mathbf{x}) - V_{i,t-1}(\mathbf{x})}{\delta t}, \tag{3.8}$$

où $Q_{\mathrm{ent}\,it}(\mathbf{x})$, $Q_{\mathrm{sort}\,it}$ et $V_{it}(\mathbf{x})$, pour $(i,t) \in \mathcal{I} \times \mathcal{T}$, sont les composantes du vecteur $\mathbf{x} \in \mathbb{R}^r$ (variables de décision) qui correspondent respectivement au débit entrant, au débit sortant et au volume de retenue à l'aménagement d'indice i et au pas de temps t. Pour $t = 1$, les termes $V_{i,t-1}(\mathbf{x})$, pour $i \in \mathcal{I}$, sont remplacés par les valeurs des paramètres des volumes sur la fenêtre temporelle passée (conditions initiales).

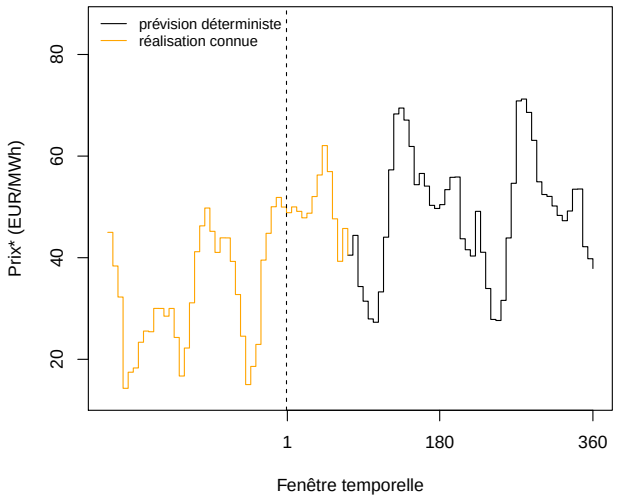
— Le débit sortant de l'équation (2.2) s'écrit :

$$\forall (i,t) \in \mathcal{I} \times \mathcal{T}, \quad Q_{\mathrm{sort}\,it}(\mathbf{x}) = Q_{\mathrm{US}\,i,t-\tau_{\mathrm{US}}(i,i)}(\mathbf{x}) + Q_{\mathrm{BR}\,i,t-\tau_{\mathrm{BR}}(i,i)}(\mathbf{x}),$$

où $Q_{\mathrm{US}\,it}(\mathbf{x})$ et $Q_{\mathrm{BR}\,it}(\mathbf{x})$, pour $(i,t) \in \mathcal{I} \times \mathcal{T}$, sont les composantes du vecteur $\mathbf{x} \in \mathbb{R}^r$ (variables de décision) qui correspondent respectivement au débit au barrage et au débit à la centrale de l'aménagement d'indice i et au pas de temps t. Les termes faisant intervenir des pas de temps d'indices négatifs sont remplacés par les valeurs des paramètres correspondants sur la fenêtre temporelle passée (conditions initiales).



(a) Apports en eau des affluents principaux



(b) Prix *day-ahead*

FIGURE 3.2 – Prévisions déterministes des apports en eau et des prix pour le cas d'étude. Les lignes verticales pointillées représentent le début de la fenêtre temporelle à 12 h le 26/04/2014. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les prix.

— Le débit entrant de l'équation (2.3) s'écrit :

$$\forall (i,t) \in \mathcal{I} \times \mathcal{T}, \quad Q_{\text{ent}it}(\mathbf{x}) = Q_{\text{US}i-1,t-\tau_{\text{US}}(i-1,i)}(\mathbf{x}) + Q_{\text{BR}i-1,t-\tau_{\text{BR}}(i-1,i)}(\mathbf{x}) + a(i,t). \quad (3.9)$$

Les termes faisant intervenir des pas de temps d'indices négatifs sont remplacés par les valeurs des paramètres correspondants sur la fenêtre temporelle passée (conditions initiales). Pour le cas d'étude, on rappelle que l'aménagement d'indice $i = 0$ correspond à celui de Verbois, dont les débits sont des paramètres. On suppose également que les contraintes (3.8) et (3.9) sont appliquées à l'indice $i = 7$ avec le débit sortant de l'aménagement de Pierre-Bénite (en septième position le long du Rhône) fixé par le programme de production initial (afin d'assurer la continuité de l'eau entre le Haut-Rhône et le Bas-Rhône, voir Sous-section 3.2.1).

— La puissance totale produite dans le périmètre d'équilibre de l'équation (2.15) s'écrit :

$$\forall t \in \mathcal{T}, \quad P_t(\mathbf{x}) = \phi(t) + \sum_{i \in \mathcal{I}} P_{\text{hydro}it}(\mathbf{x}), \quad (3.10)$$

où $P_t(\mathbf{x})$, pour $t \in \mathcal{T}$, sont les composantes du vecteur $\mathbf{x} \in \mathbb{R}^r$ (variables de décision) qui correspondent à la puissance totale produite dans le périmètre d'équilibre au pas de temps t et $P_{\text{hydro}it}(\mathbf{x})$, pour $(i,t) \in \mathcal{I} \times \mathcal{T}$, celles qui correspondent à la puissance hydroélectrique produite par l'aménagement i au pas de temps t.

— Les prochaines offres de vente *day-ahead* sont à la granularité horaire, donc

$$\forall t \in \mathcal{T} \setminus \mathcal{H}, \quad b_t(\mathbf{x}) = b_{t-1}(\mathbf{x}),$$

où $b_t(\mathbf{x})$, pour $t \in \mathcal{T}$, sont les composantes du vecteur $\mathbf{x} \in \mathbb{R}^r$ (variables de décision) qui correspondent à la puissance à mettre en vente au marché *day-ahead* pour livraison au pas de temps t, et $\mathcal{H} \subset \mathcal{T}$ désigne l'ensemble des indices des premiers pas de temps de chaque heure de la fenêtre temporelle. Pour le cas d'étude, $\mathcal{H} = \{1 + 6h,\ 0 \leq h \leq 59\} = \{1, 7, 13, \ldots, 349, 355\}$.

— Les prochaines offres de vente *day-ahead* sont seulement définies sur les pas de temps pour lesquels le marché *day-ahead* n'a pas encore été fermé, donc

$$\forall t \in \mathcal{T}_0, \quad b_t(\mathbf{x}) = 0,$$

où $\mathcal{T}_0 \subset \mathcal{T}$ désigne l'ensemble des indices des pas de temps pour lesquels le marché *day-ahead* a déjà été fermé (i.e. pour lesquels les engagements *day-ahead* sont connus à l'instant présent). Pour le cas d'étude, $\mathcal{T}_0 = \{1, \ldots, t_b\}$ avec $t_b = 72$ l'indice du pas de temps qui termine le 26/04/2014 à 24 h.

— Les écarts s'écrivent, si toutes les offres de ventes *day-ahead* sont effectuées et retenues (voir Équation (2.10)) :

$$\forall t \in \mathcal{T}, \quad e_t(\mathbf{x}) = P_t(\mathbf{x}) - (c(t) + b_t(\mathbf{x})),$$

où $e_t(\mathbf{x})$, pour $t \in \mathcal{T}$, sont les composantes du vecteur $\mathbf{x} \in \mathbb{R}^r$ (variables de décision) qui correspondent aux écarts au pas de temps t, et $c(t) \in \mathbb{R}$, pour $t \in \mathcal{T}$, sont les engagements connus à l'instant présent qu'il faut livrer au pas de temps t. Pour le cas d'étude, les engagements $c(t) \in \mathbb{R}$, pour $t \in \mathcal{T}$, connus à l'instant présent et à livrer au pas de temps t sont uniquement les engagements *day-ahead* car les engagements à long terme sont supposés nuls.

— Les écarts positifs et négatifs se déduisent des écarts par les contraintes suivantes (voir Équation (2.11)) :

$$\forall t \in \mathcal{T}, \quad \begin{cases} e_t^+(\mathbf{x}) = \max(0, e_{mt}(\mathbf{x})), \\ e_t^-(\mathbf{x}) = \max(0, -e_{mt}(\mathbf{x})), \end{cases} \tag{3.11}$$

où $e_t^+(\mathbf{x})$ et $e_t^-(\mathbf{x})$, pour $t \in \mathcal{T}$, sont les composantes du vecteur $\mathbf{x} \in \mathbb{R}^r$ (variables de décision) qui correspondent respectivement aux écarts positifs et négatifs au pas de temps t, et $e_{mt}(\mathbf{x})$, pour $t \in \mathcal{T}$, les composantes du vecteur $\mathbf{x} \in \mathbb{R}^r$ (variables de décision) qui correspondent aux écarts moyens sur chaque demi-heure (qui se déduisent entièrement à partir des composantes $e_t(\mathbf{x})$, pour $t \in \mathcal{T}$).

Certaines équations du modèle de simulation sont néanmoins simplifiées, voire omises, pour ne pas trop complexifier le problème d'optimisation. L'objectif est de trouver un bon compromis entre le niveau de simplification et le temps de calcul nécessaire à la résolution numérique du problème d'optimisation. Un problème d'optimisation trop complexe (beaucoup de variables de décision et de contraintes) se résout généralement avec un temps de calcul important (voir Section 3.1). Or, le temps de calcul doit être compatible avec les délais opérationnels. Par exemple, pour le cas d'étude, le problème d'optimisation doit être résolu en moins de 1 h.

Une manière de simplifier le problème est de réduire le nombre de variables de décision en fixant les valeurs de certaines d'entre elles. C'est en ce sens que les hauteurs de chute et les bornes minimales et maximales sur les volumes sont supposées indépendantes du programme de production. Elles sont donc des paramètres pour l'optimiseur, dont les valeurs sont déterminées par le programme de production initial construit semi-manuellement (voir Sous-section 2.5.1).

Une autre manière de réduire les temps de calcul consiste à linéariser les équations du modèle de simulation [Chang *et al.*, 2001 ; Taktak et D'Ambrosio, 2017]. C'est ce qui est appliqué aux courbes de puissance pour notre étude. Plus précisément, les expressions des puissances données par l'équation (2.5) sont linéarisées par morceaux. Cette linéarisation par morceaux doit permettre une bonne approximation des contraintes non linéaires, ce qui peut faire appel à des techniques complexes [Taktak et D'Ambrosio, 2017 ; Borghetti *et al.*, 2008]. De cette manière, les lois de puissance de l'équation (2.6) sont remplacées par :

$$\forall (i,t) \in \mathcal{I} \times \mathcal{T}, \ \forall k \in \mathcal{G}(i), \quad P_{\mathrm{Gr}ikt}(\mathbf{x}) = \phi_{P,ik}(Q_{\mathrm{Gr}ikt}(\mathbf{x})), \tag{3.12}$$

où $P_{\mathrm{Gr}ikt}(\mathbf{x})$ et $Q_{\mathrm{Gr}ikt}(\mathbf{x})$, pour $(i,t) \in \mathcal{I} \times \mathcal{T}$ et $k \in \mathcal{G}(i)$, sont les composantes du vecteur $\mathbf{x} \in \mathbb{R}^r$ (variables de décision) qui correspondent respectivement à la puissance produite et au débit du groupe de production d'indice k au pas de temps t, et $\phi_{P,ik}$ est une fonction linéaire par morceaux concave et continue, calibrée expérimentalement et qui dépend des hauteurs de chute (considérées comme des paramètres).

L'utilisation de fonctions linéaires par morceaux telles que $\phi_{P,ik}$, pour $i \in \mathcal{I}$ et $k \in \mathcal{G}(i)$, nécessite l'ajout de paramètres et variables de décision binaires (donc discrètes) supplémentaires. Par exemple, si $q_{ik\sigma} \in \mathbb{R}$, pour $i \in \mathcal{I}$, $k \in \mathcal{G}(i)$ et $\sigma \in \{1, \ldots, S_i\}$ ($S_i \in \mathbb{N}^*$), désignent les points de rupture de la fonction linéaire par morceaux $\phi_{P,ik}$ et $p_{ik\sigma} = \phi_{P,ik}(q_{ik\sigma}) \in \mathbb{R}$ les valeurs associées, alors l'équation (3.12) est remplacée, de manière équivalente, par les contraintes linéaires mixtes suivantes (voir Sous-section 3.1.3), faisant intervenir des variables de décision continues $\lambda_{ikt\sigma}(\mathbf{x})$ et des variables de décision binaires $\epsilon_{ikt\sigma}(\mathbf{x})$, pour $(i,t) \in \mathcal{I} \times \mathcal{T}$ et $k \in \mathcal{G}(i)$:

$$\forall (i,t) \in \mathcal{I} \times \mathcal{T},\ \forall k \in \mathcal{G}(i), \quad \begin{cases} P_{\mathrm{Gr}ikt}(\mathbf{x}) = \sum_{\sigma=1}^{S_i} \lambda_{ikt\sigma}(\mathbf{x}) p_{ik\sigma}, \\ Q_{\mathrm{Gr}ikt}(\mathbf{x}) = \sum_{\sigma=1}^{S_i} \lambda_{ikt\sigma}(\mathbf{x}) q_{ik\sigma}, \\ \sum_{\sigma=1}^{S_i} \lambda_{ikt\sigma}(\mathbf{x}) = 1, \\ \sum_{\sigma=1}^{S_i} \epsilon_{ikt\sigma}(\mathbf{x}) = 1, \\ \forall \sigma \in \{2, \dots, S_i\},\ \lambda_{ikt\sigma}(\mathbf{x}) \leq \epsilon_{i,k,t,\sigma-1}(\mathbf{x}) + \epsilon_{ikt\sigma}(\mathbf{x}), \\ \lambda_{ik1}(\mathbf{x}) \leq \epsilon_{ikt1}(\mathbf{x}), \\ \forall \sigma \in \{1, \dots, S_i\},\ \lambda_{ikt\sigma}(\mathbf{x}) \geq 0,\ \epsilon_{ikt\sigma}(\mathbf{x}) \in \{0,1\}, \end{cases} \tag{3.13}$$

En pratique, ces simplifications sur les lois de puissance ont un impact non négligeable sur l'évaluation de la puissance hydroélectrique produite (et donc le chiffre d'affaires) d'un programme de production. On suppose néanmoins que les erreurs dues à ces hypothèses simplificatrices sont acceptables pour l'optimisation. Des travaux sont en cours sur l'optimiseur de CNR pour améliorer ces approximations.

Une autre simplification utilisée dans notre étude est de considérer que les débits qui traversent les centrales se répartissent équitablement entre les groupes de production disponibles :

$$\forall (i,t) \in \mathcal{I} \times \mathcal{T},\ \forall k \in \mathcal{G}(i), \quad Q_{\mathrm{Gr}ikt}(\mathbf{x}) = \frac{Q_{\mathrm{US}it}(\mathbf{x}) - Q_{\mathrm{dech}}(i,t)}{n_{\mathrm{Gr}}(i,t)},$$

où $n_{\mathrm{Gr}}(i,t) \in \mathbb{N}^*$ désigne le nombre de groupes de production disponibles à l'aménagement $i \in \mathcal{I}$ au pas de temps $t \in \mathcal{T}$ (paramètres).

Des contraintes supplémentaires sont aussi ajoutées pour que l'optimiseur puisse modéliser certaines opérations manuelles. Par exemple, une contrainte de volume minimum en fin d'horizon permet de forcer le programme de production à garder des réserves au-delà de l'horizon :

$$\forall i \in \mathcal{I}, \quad V_{iT}(\mathbf{x}) \geq \underline{V}_f(i), \tag{3.14}$$

où $\underline{V}_f(i) \in \mathbb{R}_+$, pour $i \in \mathcal{I}$, sont des paramètres. Pour le cas d'étude, les volumes minimaux en fin d'horizon sont les volumes du programme de production initial construit semi-manuellement (voir Sous-section 2.5.1). Des contraintes minimales et maximales sur les écarts (voir Équation (2.10)) sont également ajoutées afin de restreindre l'utilisation des écarts :

$$\forall t \in \mathcal{T}, \quad \underline{e}(t) \leq e_t(\mathbf{x}) \leq \overline{e}(t), \tag{3.15}$$

où $\underline{e}(t)$ et $\overline{e}(t)$ désignent respectivement les bornes minimales et maximales des écarts à ne pas dépasser, supposées constantes par rapport au temps pour le cas d'étude. La contrainte (3.15) permet aussi d'écrire les contraintes (3.11) sous une forme linéaire mixte équivalente :

$$\forall t \in \mathcal{T}, \quad \begin{cases} e_{mt}(\mathbf{x}) = e_t^+(\mathbf{x}) - e_t^-(\mathbf{x}), \\ e_t^+(\mathbf{x}) \leq \overline{e}(t) \beta_{e,t}(\mathbf{x}), \\ e_t^-(\mathbf{x}) \leq \underline{e}(t)(\beta_{e,t}(\mathbf{x}) - 1), \\ e_t^-(\mathbf{x}) \geq 0,\ e_t^+(\mathbf{x}) \geq 0,\ \beta_{e,t}(\mathbf{x}) \in \{0,1\}, \end{cases} \tag{3.16}$$

ce qui nécessite l'ajout des variables binaires $\beta_{e,t}(\mathbf{x}) \in \{0,1\}$, pour $t \in \mathcal{T}$.

Le programme de production d'une solution *admissible* pour le problème d'optimisation que l'on cherche à définir est donc, aux simplifications ci-dessus près, applicable (i.e. cohérent avec le modèle hydraulique et les contraintes d'exploitation, voir §2.3.1.1) et cohérent avec les prochaines offres de vente *day-ahead* de la solution. L'admissibilité d'une solution dépend du scénario $\xi = (a, \pi, \phi)$ car certains paramètres qui entrent en jeu dans les contraintes du problème d'optimisation dépendent du scénario (voir Équations (3.9) et (3.10)).

On note $\mathcal{X}(\xi) \subset \mathbb{R}^r$ l'ensemble des solutions admissibles pour le scénario $\xi = (a, \pi, \phi) \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$. On suppose que l'ensemble (fermé) $\mathcal{X}(\xi)$ est borné (cette hypothèse est assez facile à imposer en pratique). Ainsi, l'ensemble $\mathcal{X}(\xi)$ est compact si et seulement s'il est non vide. Comme les contraintes sont linéarisées par morceaux, elles sont compatibles avec une formulation MILP (voir Sous-section 3.1.3). On peut donc écrire :

$$\forall \xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T, \quad \mathcal{X}(\xi) = \{\mathbf{x} = (\mathbf{x}_j)_{1 \leq j \leq r} \ ; \ W\mathbf{x} \leq w(\xi), \ \forall j \in \mathcal{Z}, \ \mathbf{x}_j \in \mathbb{Z}\},$$

où $W \in \mathcal{M}_{sr}(\mathbb{R})$ est une matrice indépendante du scénario dans notre étude [3], $\mathcal{Z} \subset \{1, \ldots, r\}$ et, pour tout $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$, $w(\xi) \in \mathbb{R}^s$, avec $s \in \mathbb{N}^*$ le nombre de contraintes (les contraintes d'égalité sont doublées en deux contraintes d'inégalité larges de sens inverse). En pratique, la plupart des variables discrètes sont des variables supplémentaires à celles du modèle de simulation qui ne correspondent pas à des grandeurs physiques mais qui permettent de représenter les morceaux des fonctions linéaires par morceaux (voir Équations (3.13) et (3.16)).

3.2.4 Chiffre d'affaires

La définition du chiffre d'affaires d'un programme de production fait intervenir trois types de variables de décision dépendantes : les prochaines offres de vente *day-ahead*, la puissance produite et les écarts (voir Sous-section 2.3.3). Plus précisément, le chiffre d'affaires $f(\mathbf{x}, \xi) \in \mathbb{R}$ d'une solution $\mathbf{x} \in \mathbb{R}^r$ pour le scénario $\xi = (a, \pi, \phi) \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ est défini par :

$$f(\mathbf{x}, \xi) = \delta t \sum_{t \in \mathcal{T}} \Big(\pi(t) b_t(\mathbf{x}) + \pi^+(t) e_t^+(\mathbf{x}) - \pi^-(t) e_t^-(\mathbf{x}) \Big), \tag{3.17}$$

où $\pi^+ = \big(\pi^+(t)\big)_{t \in \mathcal{T}} \in \mathbb{R}^T$ et $\pi^- = (\pi^-(t))_{t \in \mathcal{T}} \in \mathbb{R}^T$ sont respectivement les prix de règlement des écarts positifs et négatifs, supposés définis par dégradation des prix *day-ahead*, i.e., pour tout $t \in \mathcal{T}$, $\pi^+(t) = (1 - \kappa^+(t))\pi(t)$ et $\pi^-(t) = (1 + \kappa^-(t))\pi(t)$ avec $(\kappa^+(t), \kappa^-(t)) \in]0, 1[^2$. Pour le cas d'étude : $\forall t \in \mathcal{T}, \ \kappa^+(t) = \kappa^-(t) = 0{,}05$. Les erreurs d'approximation dues à cette estimation des prix de règlement des écarts font partie des incertitudes de modélisation de la section 1.4, au même titre que les erreurs d'approximation dues à l'estimation des débits de correction (voir §2.3.4.2).

Si la solution $\mathbf{x} \in \mathbb{R}^r$ est admissible pour le scénario $\xi = (a, \pi, \phi) \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ (i.e. $\mathbf{x} \in \mathcal{X}(\xi)$), alors l'équation (3.17) reprend le calcul du chiffre d'affaires de la sous-section 2.3.3. Le chiffre d'affaires de l'équation (3.17) n'est pas celui effectivement réalisé mais une estimation du gain associé au programme de production de la solution $\mathbf{x} \in \mathbb{R}^r$ compte tenu des prévisions du scénario $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$. Il sert principalement d'indicateur d'optimalité.

Il convient de noter que le chiffre d'affaires total défini à l'équation (3.17) est linéaire avec la solution.

3. Les apports en eau et les productions variables n'interviennent que dans les contraintes (3.9) et (3.10), où elles ne sont pas multipliées par des variables de décision. Les prix n'interviennent pas dans les contraintes.

3.2.5 Optimiseur

Le problème de planification à court terme de la production pour le scénario $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ consiste à trouver une solution admissible $\mathbf{x} \in \mathcal{X}(\xi)$ qui maximise le chiffre d'affaires total $f(\mathbf{x}, \xi)$ (voir Sous-section 2.3.2). Comme le chiffre d'affaires est linéaire avec la solution (voir Sous-section 3.2.4) et les contraintes sont écrites avec une formulation MILP (voir Sous-section 3.2.3), la formulation mathématique est le problème MILP suivant :

$$\max_{\mathbf{x} \in \mathcal{X}(\xi)} f(\mathbf{x}, \xi). \tag{3.18}$$

En pratique, un terme de pénalité linéaire de la forme $\mathbf{x} \mapsto \lambda_{\text{pen}}^\top \mathbf{x}$ (où $\lambda_{\text{pen}} \in \mathbb{R}^r$) est soustrait à la fonction objectif $\mathbf{x} \mapsto f(\mathbf{x}, \xi)$, notamment pour pénaliser les variations sur les débits et les volumes. En effet, un programme de production présentant des variations trop importantes et fréquentes sur les débits et les volumes pourrait difficilement être accepté en opérationnel, même s'il respecte les contraintes d'exploitation. Un tel programme de production pourrait, par exemple, entraîner une usure prématurée des groupes de production ou une plus grande complexité des consignes de conduite des aménagements en temps réel. Bien que des contraintes pourraient être ajoutées pour limiter les variations sur les débits et les volumes, il est plus commode de garder une certaine souplesse dans l'optimisation en pénalisant ces variations dans la fonction objectif. Cela permet d'augmenter les chances de trouver un programme de production applicable car les variations sur les débits et les volumes sont inévitablement importantes et fréquentes dans certaines situations. La principale difficulté réside dans le choix des pénalités, i.e. les composantes du paramètre $\lambda_{\text{pen}} \in \mathbb{R}^r$. Des pénalités trop faibles par rapport au chiffre d'affaires ne permettraient pas de discriminer efficacement les programmes de production à fortes variations. Au contraire, des pénalités trop élevées par rapport au chiffre d'affaires ne permettraient pas de maximiser efficacement le chiffre d'affaires. Dans l'idéal, les pénalités sont déterminées en chiffrant les coûts associés aux variations sur les débits et les volumes. Dans la suite du manuscrit, le terme de pénalité est omis dans la formulation mathématique par commodité mais il est bien présent dans les simulations numériques.

La fonction objectif $\mathbf{x} \mapsto f(\mathbf{x}, \xi)$ est continue (car linéaire) sur $\mathbb{R}^r$ et $\mathcal{X}(\xi)$ est une partie fermée et bornée de $\mathbb{R}^r$, donc le problème MILP (3.18) admet (au moins) une solution si et seulement si l'ensemble $\mathcal{X}(\xi)$ est non vide (voir Propriété 3.5).

Pour le cas d'étude, le problème MILP (3.18) est implémenté dans le langage GAMS et il est résolu numériquement à l'aide du solveur CPLEX MIP (voir Sous-section 3.1.3). Cet optimiseur (déterministe) fait partie des outils utilisés en opérationnel par CNR pour la gestion de la production hydroélectrique du Rhône (voir §2.3.5.2). La taille $r \approx 500\ 000$ du problème MILP (3.18) est assez grande pour le solveur CPLEX MIP. La solution renvoyée par l'optimiseur n'est pas nécessairement une solution optimale du problème, mais la première solution trouvée vérifiant le critère d'arrêt défini par une tolérance sur le MIP-gap (voir Sous-section 3.1.3). Pour le cas d'étude, on utilise un MIP-gap relatif seuil de 2% et un temps de calcul maximal de 8 000 s (bien plus important que le temps disponible en opérationnel qui est de 1 h).

D'autres approches de résolution sont présentes dans la littérature pour résoudre ce type de problème d'optimisation (ou plus généralement le problème d'attribution d'unités de production) sans nécessairement linéariser les contraintes non linéaires [van Ackooij *et al.*, 2018 ; Saravanan *et al.*, 2013 ; Taktak et D'Ambrosio, 2017]. Ces approches sont généralement classées dans les catégories suivantes : programmation dynamique, méthode de décomposition, relaxation lagrangienne, heuristiques et méta-heuristiques. Néanmoins, ces approches sont très dépendantes du problème étudié et elles sont souvent peu précises et peu efficaces en grande dimension. À l'inverse, les approches MILP deviennent populaires dans la littérature grâce à l'augmentation de l'efficacité des solveurs numériques pour résoudre les problèmes MILP [van Ackooij *et al.*, 2018].

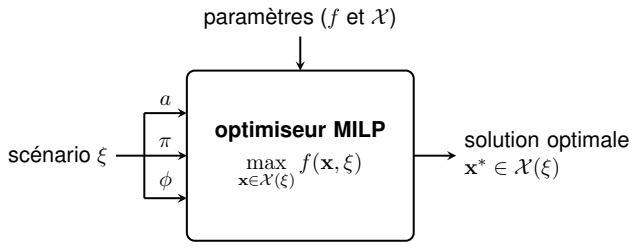


FIGURE 3.3 – Schéma de l'optimiseur en univers déterministe.

Le temps de calcul de la résolution est un des principaux inconvénients des approches MILP. Il augmente rapidement si on ajoute des contraintes ou des variables de décision (notamment discrètes). En outre, il est difficile de l'estimer en avance pour une situation donnée. C'est pourquoi certains auteurs proposent une combinaison de différentes approches de résolution [van Ackooij *et al.*, 2018].

Avec les notations des sous-sections précédentes, le schéma de la figure 2.15 devient donc celui de la figure 3.3.

3.2.6 Solution déterministe

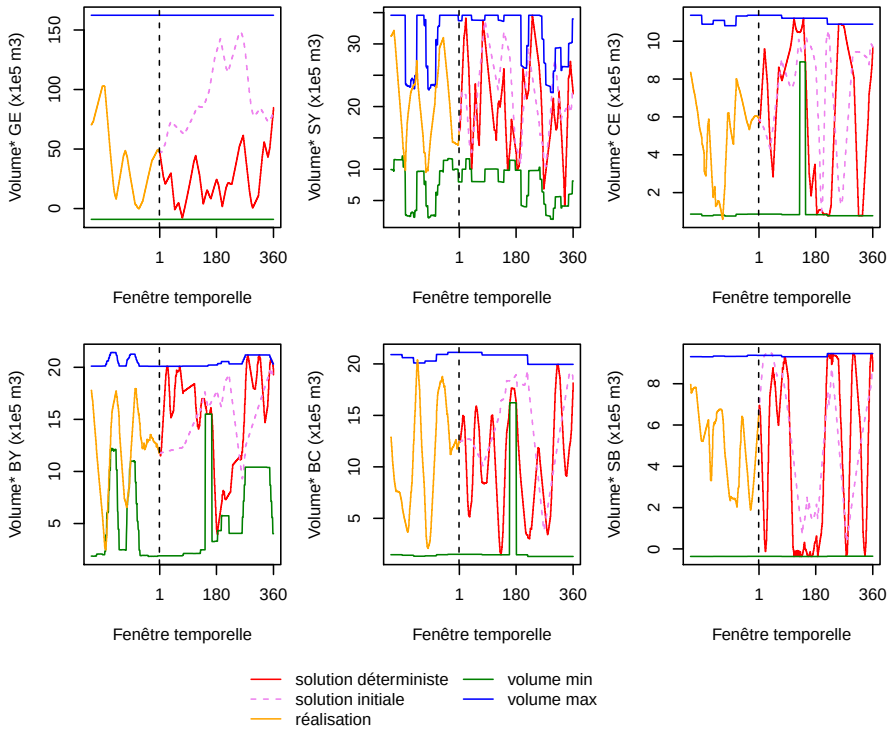
On appelle *scénario déterministe* celui composé des prévisions déterministes des apports en eau, des prix et des productions variables. On rappelle que l'on considère, pour le cas d'étude, que la prévision déterministe des apports est celle obtenue à partir du membre contrôle de la prévision d'ensemble et non la prévision déterministe effectivement utilisée en opérationnel (voir Sous-section 2.5.3). La figure 3.2 illustre la prévision déterministe des apports en eau des principaux affluents du Haut-Rhône (qui interviennent dans les apports en eau complets) et la prévision déterministe des prix pour le cas d'étude. Dans la suite du manuscrit, le scénario déterministe est noté $\xi_0 = (a_0, \pi_0, \phi_0) \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$.

Pour la suite du manuscrit, la résolution en univers déterministe du problème MILP (3.18) avec le scénario déterministe sert de référence pour l'analyse des résultats de la résolution en univers probabiliste. On appelle *solution déterministe*, notée $\mathbf{x}_0^* \in \mathcal{X}(\xi_0)$, la solution optimale du problème de planification de la production en univers déterministe avec le scénario déterministe ξ_0, i.e. :

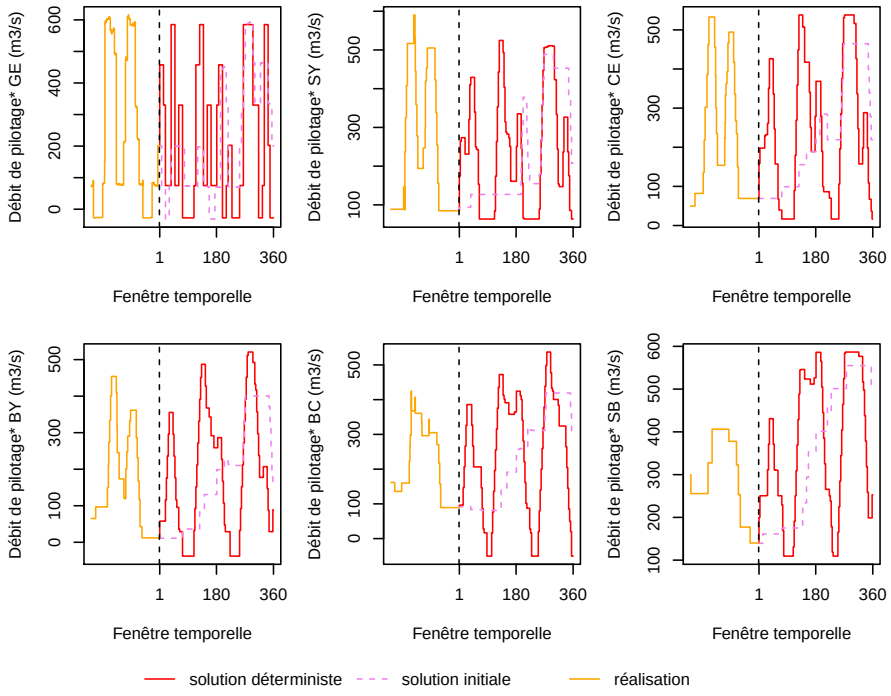
$$\max_{\mathbf{x}_0\in\mathcal{X}(\xi_0)} f(\mathbf{x}_0, \xi_0). \tag{3.19}$$

La figure 3.4 présente les valeurs de deux types de variable de décision de la solution déterministe pour le cas d'étude (courbes en rouge) : les volumes (Figure 3.4a) et les débits de pilotage (Figure 3.4b). Les courbes violettes en tirets correspondent aux valeurs de ces variables de décision pour le programme de production initial construit semi-manuellement. On rappelle que le programme de production initial et celui de la solution déterministe sont construits avec des hypothèses et paramètres légèrement différents :

- le programme de production initial est construit avec un scénario différent du scénario déterministe (voir Sous-section 2.5.3),
- le programme de production initial est construit par éclusées énergétiques synchrones en démodulant le débit sortant du Haut-Rhône (voir §2.3.5.1 et Figure 3.4b), alors que celui de la solution déterministe est construit par optimisation et uniquement sur le Haut-Rhône, ce qui incite à créer des modulations importantes (voir Sous-section 2.5.2 et Figure 3.4b),
- le programme de production initial est construit avec les équations du modèle de simulation, alors que celui de la solution déterministe est construit avec les simplifications de l'optimiseur (voir Sous-section 3.2.3).



(a) Volumes des retenues



(b) Débits de pilotage

FIGURE 3.4 – Exemple de la solution déterministe et du programme de production initial pour le cas d'étude. Les lignes verticales en tirets correspondent au début de la fenêtre temporelle. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les volumes et les débits de pilotage.

La comparaison entre les deux solutions ne peut donc pas servir à tirer des conclusions sur le gain de l'optimiseur déterministe par rapport à l'optimisation semi-manuelle par éclusées énergétiques synchrones pour le cas d'étude. L'analyse du gain de l'optimiseur déterministe n'entre de toute manière pas dans le cadre de ce manuscrit.

3.3 Prise en compte des incertitudes sur le scénario

La section précédente aboutit à la formulation mathématique du problème d'optimisation en univers déterministe. Dans cette section, nous présentons un des points clefs de ce manuscrit : le passage de l'univers déterministe à l'univers probabiliste pour tenir compte des incertitudes sur le scénario, comme évoqué dans les objectifs de la thèse à la section 1.5. Nous commençons par présenter les notations liées à la modélisation des incertitudes en scénarios multivariés et par étudier le comportement de l'optimiseur déterministe face aux scénarios multivariés du cas d'étude. Cette étude permet de quantifier l'impact des incertitudes sur l'optimisation déterministe et de motiver l'utilisation du cadre probabiliste pour le cas d'étude. Nous analysons ensuite la méthode d'optimisation sous incertitudes la plus adaptée à la stratégie de valorisation présentée à la section 2.4 dans le contexte de CNR, en lien avec l'état de l'art de la sous-section 3.1.4. Cela nous conduit à la formulation mathématique du problème d'optimisation en univers probabiliste. Enfin, dans une dernière sous-section, nous donnons quelques éléments de discussion sur la complémentarité de l'optimisation en univers déterministe et en univers probabiliste.

3.3.1 Modélisation des incertitudes sur le scénario

Pour tenir compte des incertitudes sur le scénario, on modélise les apports en eau, les prix *day-ahead* et les productions variables par des variables aléatoires notées respectivement A, Π et Φ, définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans $\mathbb{R}^{IT}$, $\mathbb{R}^T$ et $\mathbb{R}^T$ respectivement (munis de leur tribu borélienne). Ainsi, le scénario est modélisé par la variable aléatoire jointe $\Xi = (A, \Pi, \Phi)$ sur $(\Omega, \mathcal{F}, \mathbb{P})$. Dans tout le manuscrit, sauf indication contraire, le symbole $\mathbb{E}$ désigne l'espérance pour les variables aléatoires définies sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

En pratique, la variable aléatoire Ξ n'est pas connue mais elle est estimée grâce à des modèles de prévision probabiliste (voir Sous-section 2.4.4). Pour notre étude, on utilise une approche ensembliste qui consiste à construire un échantillon $(\xi_1, \dots, \xi_N) \in \Xi(\Omega)^N$ de scénarios possibles ($N \in \mathbb{N}^*$), appelés *scénarios probabilistes* (par opposition au scénario déterministe ξ_0 de la sous-section 3.2.6), avec leurs probabilités associées $(p_1, \dots, p_N) \in [0,1]^N$, $p_1 + \dots + p_N = 1$.

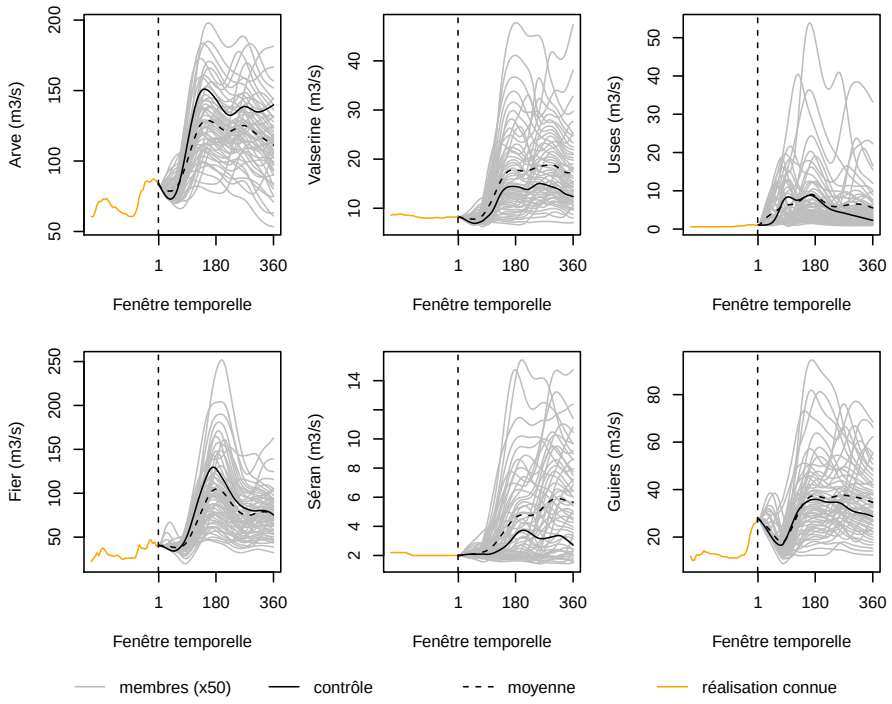
Pour le cas d'étude, les apports en eau et les prix *day-ahead* sont échantillonnés séparément et les productions variables ne sont pas considérées (voir Sous-sections 2.4.4 et 2.5.3). Un échantillon équiprobable $(a_1, \dots, a_{N_a}) \in A(\Omega)^{N_a}$ ($N_a = 50$) d'apports en eau possibles des affluents principaux est obtenu à partir de prévisions météorologiques d'ensemble, voir Figure 3.5a. Un échantillon équiprobable $(\pi_1, \dots, \pi_{N_\pi}) \in \Pi(\Omega)^{N_\pi}$ ($N_\pi = 50$) de prix *day-ahead* possibles est obtenu à partir d'un modèle d'erreurs simple par quantiles, voir Figure 3.5b. En outre, les variables aléatoires A et Π sont supposées indépendantes pour le cas d'étude, donc l'échantillon $(\xi_1, \dots, \xi_N)$ des scénarios possibles et leurs probabilités associées $(p_1, \dots, p_N)$ s'en déduisent en considérant tous les couples $\xi_n = (a_{k_n}, \pi_{l_n})$ avec les probabilités $p_n = 1/N$, pour $n \in \{1, \dots, N\}$, où $N = N_a N_\pi$ ($N = 2\ 500$), $k_n \in \{1, \dots, N_a\}$ et $l_n \in \{1, \dots, N_\pi\}$.

On considère ainsi la variable aléatoire finie $\hat{\Xi}$ définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ par :

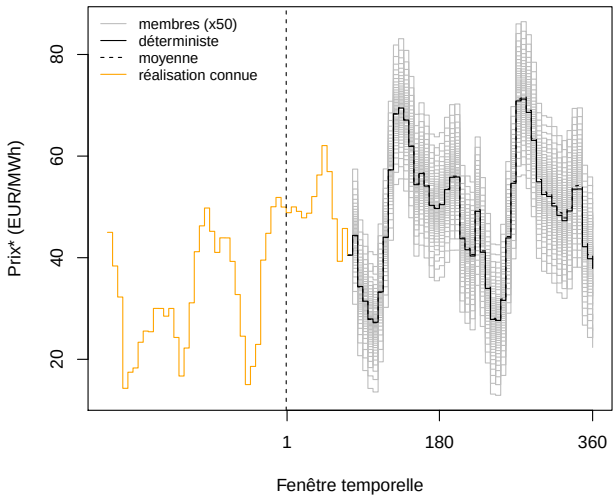
— $\hat{\Xi}(\Omega) = \{\xi_1, \dots, \xi_N\}$,

— $\forall n \in \{1, \dots, N\},\ \mathbb{P}[\hat{\Xi} = \xi_n] = p_n$.

Si les prévisions probabilistes sont assez fiables (voir Sous-section 1.4.5), alors $\hat{\Xi}$ est une « bonne » approximation de la variable aléatoire Ξ.

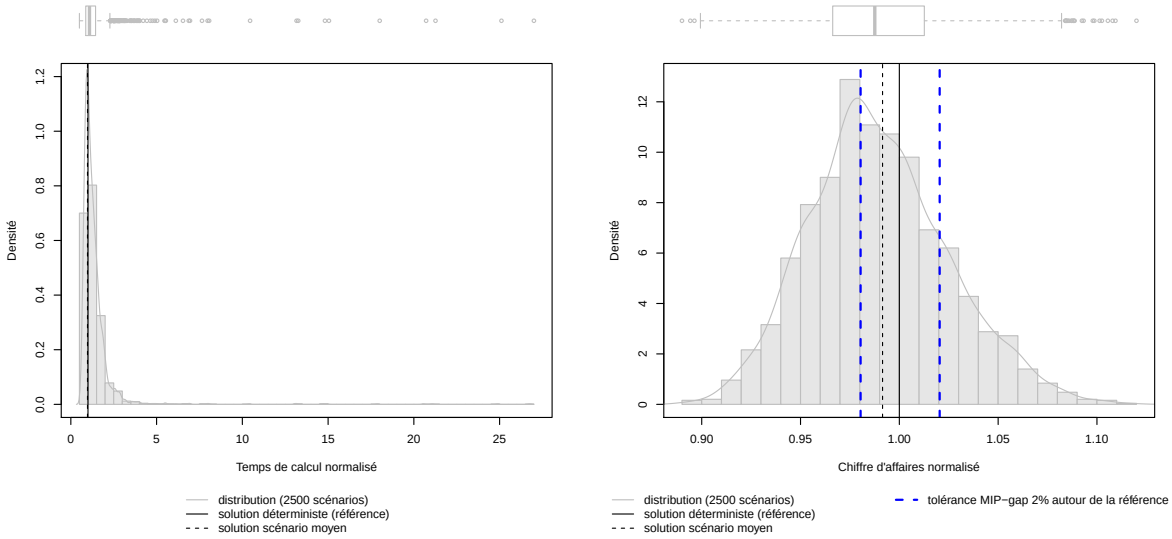


(a) Apports en eau des affluents principaux



(b) Prix *day-ahead*

FIGURE 3.5 – Prévisions d'ensemble des apports en eau et des prix pour le cas d'étude. Les lignes verticales pointillées représentent le début de la fenêtre temporelle à 12 h le 26/04/2014. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les prix.



(a) Distribution des temps de calcul (b) Distribution des chiffres d'affaires optimaux

FIGURE 3.6 – Sensibilité du temps de calcul et des chiffres d'affaires optimaux de l'optimiseur déterministe face aux incertitudes pour le cas d'étude. Les temps de calcul et les valeurs optimales sont normalisés par rapport à ceux obtenus avec le scénario déterministe. Les fonctions de densité sont obtenues par estimation par noyaux.

On note également $\overline{\xi} = (\overline{a}, \overline{\pi}, \overline{\phi}) \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ le scénario moyen :

$$\overline{\xi} = (\overline{a}, \overline{\pi}, \overline{\phi}) = \mathbb{E}[\hat{\Xi}] = \sum_{n=1}^{N} p_n \xi_n.$$

3.3.2 Comportement de l'optimiseur déterministe face aux incertitudes

Avant d'adapter la formulation mathématique du problème d'optimisation (3.18) pour tenir compte des incertitudes, nous présentons dans cette sous-section le comportement de l'optimiseur déterministe face aux incertitudes pour le cas d'étude. Cela permet d'étudier l'impact des incertitudes sur l'optimiseur déterministe et de motiver le passage à l'univers probabiliste.

On lance successivement l'optimiseur déterministe de l'équation (3.18) pour chacun des N scénarios probabilistes ξ_1, ..., ξ_N, les autres paramètres étant inchangés. On obtient alors N solutions optimales intermédiaires, notées $\mathbf{x}_n^* \in \mathcal{X}(\xi_n)$, pour $n \in \{1, \ldots, N\}$. Ces N solutions optimales intermédiaires peuvent être comparées à la solution déterministe de la sous-section 3.2.6.

La figure 3.6 présente la distribution de deux covariables en fonction des scénarios probabilistes : les temps de calcul et les chiffres d'affaires optimaux. On observe que les temps de calcul sont en moyenne 14,23% plus élevés avec les scénarios probabilistes qu'avec le scénario déterministe. Il existe même plusieurs scénarios extrêmes qui conduisent à des temps de calcul 5 à 25 fois supérieurs au temps de calcul de la solution déterministe, illustrant la problématique des temps de calcul de l'optimiseur (voir §2.3.5.2). On observe également que les chiffres d'affaires optimaux sont en moyenne 1,33% moins élevés avec les scénarios probabilistes qu'avec le scénario déterministe. Néanmoins, cette valeur est inférieure au MIP-gap relatif seuil de 2% utilisé par l'optimiseur, donc on peut considérer que le chiffre d'affaires optimal moyen est égal au chiffre d'affaire de la solution optimale déterministe. En revanche, l'écart-type des chiffres d'affaires optimaux normalisés est de 0,035, ce qui est supérieur au MIP-gap relatif seuil de 2%. Autrement dit, les incertitudes ont un impact non négligeable sur la valeur optimale de la solution.

Type paramètre	*Temps de calcul*	*Chiffre d'affaires optimal*
Apports	0,19	0,72
Prix	0,03	0,28
Interactions	0,78	0,00

TABLEAU 3.1 – Indices de Sobol du premier ordre des apports et des prix pour le cas d'étude.

À ce stade, il est possible d'effectuer une analyse de sensibilité pour mesurer l'influence des deux types de paramètres (apports et prix) sur le temps de calcul et le chiffre d'affaires optimal. Par exemple, on peut calculer les indices de Sobol du premier ordre, voir Équation (1.2). Pour notre cas, on a deux covariables, le temps de calcul et le chiffre d'affaires optimal, et deux facteurs explicatifs, les apports en eau et les prix des scénarios. Comme les apports en eau et les prix sont donnés par des variables aléatoires discrètes finies et que les scénarios sont toutes les combinaisons possibles entre les apports en eau et les prix supposés indépendants (voir Sous-section 2.4.4), la formule de l'équation (1.2) peut être estimée directement. On obtient les résultats du tableau 3.1. Pour le cas d'étude, on observe que les apports en eau seuls ont plus d'influence que les prix seuls sur le temps de calcul et sur le chiffre d'affaires optimal de l'optimiseur déterministe. La faible influence des prix par rapport aux apports en eau pourrait être due aux formes similaires des membres de la prévision d'ensemble des prix (en raison du modèle d'erreur, voir §2.4.4.4), ce qui impacte peu l'optimisation (c'est surtout les différences dans les variations des prix qui impactent l'optimisation, voir Sous-section 2.4.1). Ces résultats pourraient servir, par exemple, à ne considérer que les incertitudes sur les apports en eau dans un premier temps. Néanmoins, nous ne faisons pas cette hypothèse pour les raisons suivantes :

— il est difficile de transposer ces résultats à d'autres cas d'étude,

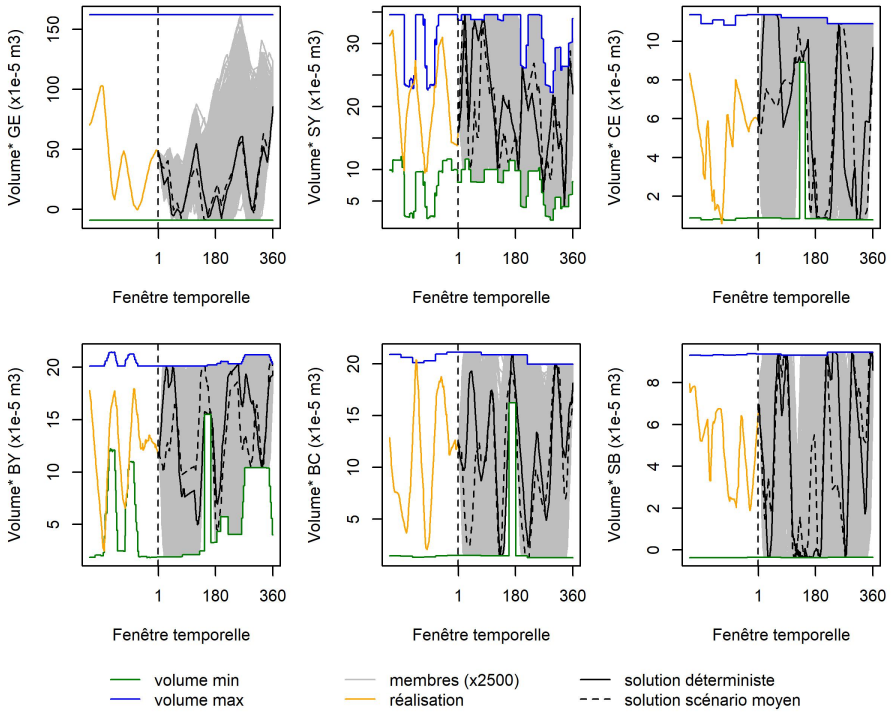
— l'objectif de la thèse est de définir une méthodologie qui puisse tenir compte des éventuelles corrélations entre les apports et les prix, auquel cas les apports et les prix sont combinés et ne peuvent donc pas être séparés (voir §2.4.4.2).

En outre, la forte valeur des interactions pour le temps de calcul illustre là encore la problématique des temps de calcul de l'optimiseur (voir §2.3.5.2).

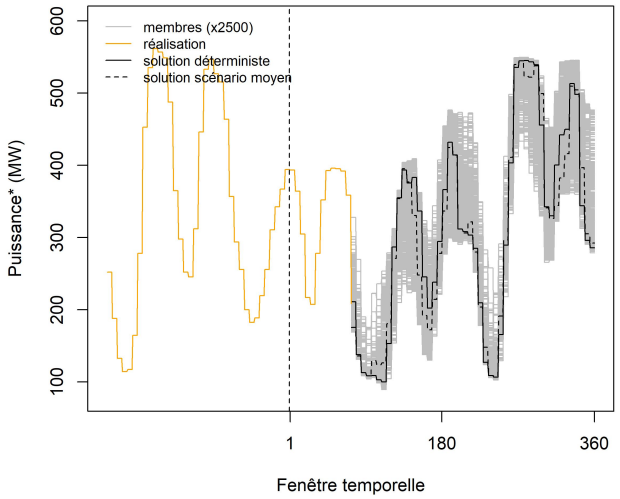
La figure 3.7a présente les volumes des solutions optimales intermédiaires à chaque aménagement. On observe que les solutions optimales intermédiaires couvrent presque tout le domaine admissible, malgré les hypothèses simplificatrices qui réduisent les degrés de liberté (voir Sous-section 2.5.2). La figure 3.7b présente les ventes *day-ahead* associées aux solutions optimales intermédiaires. On observe que les prochaines ventes *day-ahead* ont globalement la même allure temporelle, peut être en raison des formes similaires des membres de la prévision d'ensemble des prix. Néanmoins, les différences d'amplitude des ventes entre chaque solution optimale intermédiaire ne sont pas négligeables car elles sont globalement du même ordre de grandeur que les variations temporelles des ventes. Les solutions optimales intermédiaires sont donc globalement assez différentes.

Par ailleurs, l'admissibilité et l'optimalité des solutions optimales intermédiaires sont analysées avec les autres scénarios qui n'ont pas servi à leurs optimisations. Une étude d'admissibilité montre qu'aucune solution optimale intermédiaire n'est admissible pour un autre scénario que celui avec lequel elle est optimisée, i.e.

$$\forall (n, n') \in \{1, \dots, N\}^2,\ n \neq n', \quad \mathbf{x}_n^* \notin \mathcal{X}(\xi_{n'}).$$



(a) Volumes des retenues



(b) Ventes *day-ahead*

FIGURE 3.7 – Exemples des solutions optimales intermédiaires pour le cas d'étude. Les lignes verticales en tirets correspondent au début de la fenêtre temporelle. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les volumes et les puissances.

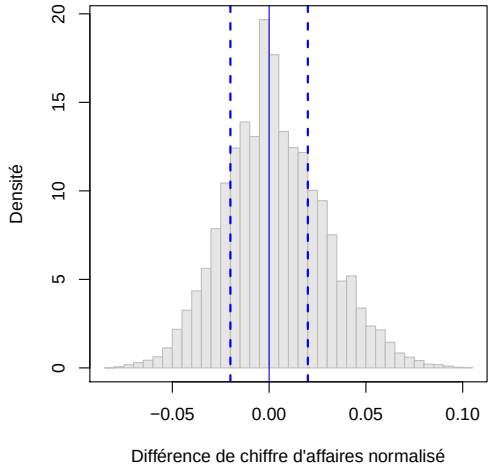


FIGURE 3.8 – Distribution des différences entre les valeurs optimales des solutions optimales intermédiaires et leurs chiffres d'affaires calculés avec les autres scénarios. Les lignes verticales bleues en tirets correspondent aux valeurs seuils de 2% au-delà desquels les différences sont considérées comme significatives.

Pour le cas d'étude, les résultats montrent donc que les solutions optimales intermédiaires sont très spécifiques aux scénarios probabilistes avec lesquels elles sont optimisées, donc peu robustes face aux incertitudes. Concernant l'optimalité, la figure 3.8 présente la distribution des N^2 réels

$$\frac{f(\mathbf{x}_n^*, \xi_n) - f(\mathbf{x}_n^*, \xi_{n'})}{f(\mathbf{x}_0^*, \xi_0)} \in \mathbb{R}, \quad \text{pour } (n, n') \in \{1, \dots, N\}^2,$$

où on rappelle que $\xi_0 \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ désigne le scénario déterministe et $\mathbf{x}_0^* \in \mathcal{X}(\xi_0)$ la solution déterministe (voir Sous-section 3.2.6). On observe que les différences de chiffre d'affaires sont significatives pour environ 42,64% des couples $(n, n') \in \{1, \dots, N\}^2$, proportion qui n'est pas négligeable. Cette étude d'optimalité montre donc que les solutions optimales intermédiaires sont globalement peu optimales avec les autres scénarios qui n'ont pas servi à leurs optimisations.

Ainsi, les solutions optimales intermédiaires sont suffisamment différentes entre elles et spécifiques aux scénarios avec lesquels elles sont optimisées pour conclure que les incertitudes ont un impact non négligeable sur l'optimiseur déterministe pour le cas d'étude. D'où l'intérêt de faire passer l'optimiseur en univers probabiliste pour le cas d'étude.

3.3.3 Quelle méthode d'optimisation sous incertitudes ?

Nous présentons dans cette sous-section le raisonnement qui aboutit au choix de la méthode d'optimisation sous incertitudes que nous utilisons dans ce manuscrit pour le problème de planification de la production en univers probabiliste. Nous nous appuyons notamment sur la stratégie de valorisation de la production initiée à la sous-section 2.4.3 et sur l'état de l'art général de la sous-section 3.1.4.

3.3.3.1 Besoin d'une solution unique

Malgré un cadre probabiliste, l'optimiseur doit renvoyer une unique solution à la fois optimale (qui maximise le chiffre d'affaires) et applicable (qui respecte le modèle hydraulique et les contraintes d'exploitation). Or, certaines contraintes dépendent des scénarios considérés (voir Sous-section 3.2.3). L'unique programme de production renvoyé par l'optimiseur doit donc respecter les contraintes d'exploitation pour au moins un scénario, appelé *scénario cible* dans la

suite de ce paragraphe. Le besoin d'une solution unique est contraignant car il exclut d'office certaines méthodes d'optimisation sous incertitudes, comme par exemple les méthodes type Monte-Carlo (voir Sous-section 3.1.4).

Dans un univers déterministe, le scénario cible correspond au seul scénario déterministe à disposition. En revanche, dans un cadre probabiliste où l'on dispose d'un ensemble de scénarios possibles qui viennent s'additionner au scénario déterministe, le choix du scénario cible pose question. On peut ainsi envisager, en amont de l'optimisation, trois possibilités :

1. choisir le scénario déterministe,
2. choisir un des scénarios probabilistes,
3. construire un nouveau scénario.

Avec la première possibilité, la solution renvoyée par l'optimiseur déterministe et celle renvoyée par l'optimiseur probabiliste sont compatibles dans le sens où elles respectent toutes les deux les contraintes pour le scénario déterministe. La deuxième possibilité permet de ne pas se restreindre à la prévision déterministe, mais le choix du scénario cible n'est pas évident pour autant. Pour garder une optimisation globalement neutre face aux risques, il semble plus pertinent de choisir un scénario proche du scénario moyen (défini par la moyenne des prévisions d'ensemble) plutôt qu'un scénario extrême. Cela peut se faire par exemple en sélectionnant le scénario le plus proche du scénario moyen grâce à une distance donnée ou encore celui qui permet de mieux représenter la distribution initiale (obtenu, par exemple, par une itération de l'algorithme de sélection de scénarios, voir Annexe A.3). La troisième possibilité n'utilise ni le scénario déterministe, ni un des scénarios probabilistes, mais le scénario moyen, toujours pour rester neutre face aux risques. Néanmoins, le scénario moyen peut être trop « lisse » et ne pas correspondre à un scénario physiquement possible.

Dans le cadre de la thèse, le scénario cible est le scénario déterministe $\xi_0 = (a_0, \pi_0, \phi_0)$ construit à partir des prévisions déterministes $a_0 \in A(\Omega)$ des apports en eau (membre contrôle), $\pi_0 \in \Pi(\Omega)$ des prix de l'électricité et $\phi_0 \in \Phi(\Omega)$ des productions variables (voir Figure 3.2). Ce choix est motivé par le fait que les prévisions déterministes sont complétées par expertise en opérationnel, alors que l'ajout d'expertise sur les prévisions probabilistes est une recherche en cours [Celie *et al.*, 2019]. La méthodologie présentée dans ce manuscrit peut néanmoins être adaptée aux deux autres possibilités du choix du scénario cible.

3.3.3.2 Vers un problème d'optimisation stochastique à deux niveaux

Le programme de production effectivement appliqué en temps réel est le résultat de l'enchaînement de plusieurs planifications successives (voir Sous-section 2.3.1). Ces planifications coïncident avec les mises à jour des prévisions et les prises de décision du processus opérationnel (i.e. le choix des ventes sur le marché *day-ahead* et du programme de production). Elles permettent de mettre à jour le programme de production, notamment en fonction du scénario qui se réalise effectivement. Ces modifications créent des écarts qui viennent grever les chiffres d'affaires estimés lors des optimisations (voir Sous-section 2.4.3). Autrement dit, si on considère, pour une planification donnée, la solution optimale $\mathbf{x}^* \in \mathcal{X}(\xi)$ du problème d'optimisation (3.18) avec un scénario $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$, alors le chiffre d'affaires réel associé à cette solution s'écrit

$$f(\mathbf{x}^*, \xi) + \mathcal{R}(\mathbf{x}^*), \tag{3.20}$$

où $\mathcal{R}(\mathbf{x}^*) \in \mathbb{R}$ représente les gains algébriques (pouvant être positifs ou négatifs) apportés par les mises à jour de la solution $\mathbf{x}^*$ aux planifications ultérieures. Le terme de gauche $f(\mathbf{x}^*, \xi)$ est le chiffre d'affaires obtenu si le scénario ξ se réalise effectivement. Le terme de droite $\mathcal{R}(\mathbf{x}^*)$ est un terme de recours qui tient compte de toutes les opérations effectuées entre le moment où la solution $\mathbf{x}^*$ est calculée et le moment où le programme de production modifié est appliqué

en temps réel. Ce terme est nul si aucune modification n'est apportée et si le scénario ξ se réalise effectivement. Bien sûr, il n'est entièrement connu qu'*a posteriori*, lorsque l'on connaît la réalisation de toutes les incertitudes et les modifications apportées. Il n'est donc pas accessible lors du calcul de la solution $\mathbf{x}^*$.

En raison de l'enchaînement des décisions au cours du temps, les variables de décision d'une planification peuvent être classées en deux catégories, selon que les décisions associées sont considérées comme définitives ou provisoires.

— Les décisions définitives sont celles qui sont appliquées une fois que le problème de planification est résolu et qui ne pourront plus être modifiées par la suite. Il peut s'agir, par exemple, de la partie du programme de production entre l'instant présent et l'instant de la prochaine planification (i.e. sur les premières heures de la fenêtre temporelle) ou des ventes *day-ahead* à effectuer à la prochaine session du marché *day-ahead* pour la dernière planification de la matinée (les ventes *day-ahead* sont effectuées chaque jour avant midi).

— Les décisions provisoires sont celles qui pourront être modifiées aux planifications ultérieures pour que le programme de production soit toujours applicable et optimal du mieux possible en fonction du scénario qui se réalise. Les décisions provisoires peuvent être le programme de production au-delà de l'instant de la planification ultérieure (quelques heures après l'instant présent) ou les prochaines ventes *day-ahead* qui ne sont pas encore définitives.

En univers déterministe, l'optimiseur agit en considérant implicitement que tous les paramètres de la planification sont parfaitement justes (en particulier, en considérant que le scénario déterministe va parfaitement se réaliser), et donc qu'aucune mise à jour de la solution ne sera nécessaire (le terme $\mathcal{R}$ de l'équation (3.20) est considéré comme nul). Les décisions provisoires sont alors considérées comme définitives (avec une connaissance parfaite du futur, les décisions peuvent être parfaitement déterminées en avance).

En univers probabiliste, les incertitudes sur les paramètres de la planification sont prises en compte. Les décisions définitives doivent être admissibles quel que soit le scénario qui se réalise effectivement. En revanche, on considère que les décisions provisoires peuvent être changées selon le scénario qui se réalise effectivement. Ainsi, pour chaque scénario probabiliste, il est possible de définir les décisions provisoires à partir des décisions définitives et sous l'hypothèse que le scénario probabiliste considéré se réalise effectivement. De cette manière, les valeurs économiques des décisions provisoires pour chaque scénario probabiliste peuvent servir à anticiper une partie des modifications ultérieures du programme de production pour s'adapter au scénario qui se réalise effectivement. Ces modifications sont estimées en considérant les réoptimisations du programme de production pour chaque scénario probabiliste. Autrement dit, en univers probabiliste, on peut estimer, au moment de l'optimisation, une partie des gains algébriques $\mathcal{R}$ de l'équation (3.20), permettant d'améliorer l'estimation du chiffre d'affaires. On retrouve alors l'idée fondamentale de l'optimisation stochastique associé à l'arbre décision à deux niveaux représenté à la figure 3.9.

— La racine de l'arbre est associée au premier niveau et aux décisions définitives valables pour tous les scénarios probabilistes. Ces décisions définitives sont liées aux variables de décision d'intérêt de la planification, i.e. celles qui permettent de définir la solution renvoyée par l'optimiseur (voir §3.3.3.1).

— Les feuilles de l'arbre représentent le second niveau et aux décisions provisoires pour chaque scénario probabiliste à partir des décisions définitives du premier niveau. Ces décisions provisoires sont liées aux variables de décision des réoptimisations possibles de la solution renvoyée par l'optimiseur. Même si elles servent pour l'optimisation en univers probabiliste de la solution renvoyée par l'optimiseur, ces variables de décision sont « cachées » car elles ne seront pas directement renvoyées par l'optimiseur.

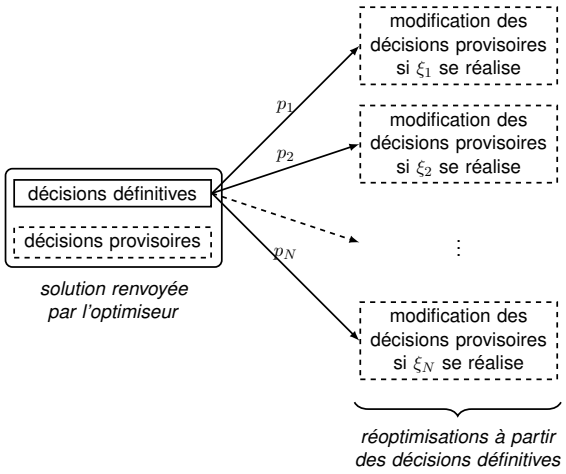


FIGURE 3.9 – Schéma général de l'arbre de décision associé à l'optimisation stochastique à deux niveaux.

L'hypothèse de non-anticipativité de l'optimisation stochastique à deux niveaux est portée sur les décisions définitives qui doivent être valables quel que soit le scénario car elles sont effectuées avant de connaître la réalisation du scénario.

La stratégie de valorisation de la sous-section 2.4.3 peut donc se modéliser avec un problème d'optimisation stochastique à deux niveaux. Par ailleurs, cette approche est très adaptée à la modélisation des incertitudes en scénarios multivariés, qui forme naturellement un arbre de scénarios. Comme il n'y a pas de distinction entre les trois types de prévision qui interviennent dans les scénarios, il serait même possible de prendre en compte d'autres sources d'incertitudes dans les scénarios, sans que cela n'impacte la modélisation par optimisation stochastique à deux niveaux (le nombre de scénarios doit toutefois rester limité). Il convient néanmoins de noter que le premier niveau fait intervenir un scénario car les prochaines ventes *day-ahead* (et plus généralement la solution renvoyée par l'optimiseur) sont déterminées à partir du scénario déterministe. Il s'agit d'une particularité qui n'est généralement pas présente dans la littérature. En général, le premier niveau ne fait pas intervenir les paramètres incertains : les scénarios interviennent uniquement dans le second niveau. Dans le cadre de la thèse, aucune décision n'est effectuée sans incertitudes. On peut alors voir le scénario déterministe comme la meilleure estimation possible des paramètres incertains avec laquelle on souhaite prendre les décisions définitives. Les scénarios probabilistes sont ceux qui prennent le rôle des scénarios « classiques » de la littérature.

La suite de cette sous-section donne davantage d'éléments sur les rôles de ces deux niveaux et leur complémentarité. La formulation mathématique du problème d'optimisation en univers probabiliste est présentée à la sous-section 3.3.4 qui suit.

3.3.3.3 Premier niveau

Le premier niveau permet de définir les décisions définitives qui doivent être valables pour tous les scénarios probabilistes. Deux décisions définitives se présentent naturellement pour la planification de la production à court terme :

— le programme de production sur les premières heures de la fenêtre temporelle jusqu'à l'instant de la prochaine planification,

— les offres de vente à effectuer à la prochaine session du marché *day-ahead* si on considère la dernière planification de la matinée (comme c'est le cas pour le cas d'étude où l'instant présent est à 11 h le 26/04/2014).

En effet, le programme de production sur les premières heures de la fenêtre temporelle ne peut pas être corrigé (sauf corrections des consignes de conduite des aménagements hydroélectriques en temps réel, ce qui dépasse le cadre de ce manuscrit). Il doit donc être applicable quelle que soit la réalisation des paramètres incertains (les apports en eau notamment). De même, pour la dernière planification de la matinée, les offres de vente *day-ahead* ne pourront plus être modifiées avant la planification ultérieure car les sessions du marché *day-ahead* ferment tous les jours à midi. Les offres de vente *day-ahead* retenues s'additionneront aux engagements pour la production du lendemain (voir §1.2.2.2).

Néanmoins, le choix des décisions définitives est assez subjectif et dépend de la modélisation souhaitée. Dans le cadre de la thèse, nous considérons une seule décision définitive : les offres de vente à effectuer à la prochaine session du marché *day-ahead*.

Le programme de production sur les premières heures de la fenêtre temporelle n'est pas utilisé dans le premier niveau car il est en général déjà fixé en grande partie sur les premières heures (la fenêtre temporelle commence au moins une heure après l'instant présent, voir §2.3.5.2). En outre, pour forcer le programme de production renvoyé par l'optimiseur à être applicable sur les premières heures, il semble plus pertinent d'ajouter une contrainte de robustesse qui impose l'applicabilité du programme de production pour tous les scénarios sur les premières heures de la fenêtre temporelle. Néanmoins, on suppose que les incertitudes sur les premières heures sont suffisamment faibles pour ne pas considérer cette contrainte de robustesse dans la méthodologie que nous présentons dans ce manuscrit. La méthodologie peut toutefois être facilement adaptée pour ajouter la contrainte de robustesse.

Le choix d'utiliser les offres de vente à effectuer à la prochaine session du marché *day-ahead* au premier niveau concerne toutes les planifications, même celles qui ne sont pas en fin de matinée et dont les prochaines offres de vente *day-ahead* ne sont pas réellement définitives. Cela permet d'avoir une méthodologie applicable à toutes les planifications de la journée. On souhaite ainsi construire un arbre de décision associé aux ventes sur le marché *day-ahead*.

Néanmoins, l'optimiseur probabiliste que l'on cherche à construire doit renvoyer une unique solution applicable pour le scénario déterministe (voir §3.3.3.1). Les variables de décision d'intérêt de la planification ne sont donc pas seulement les prochaines ventes *day-ahead* que l'on considère comme définitives, mais une solution complète, notée $\mathbf{x}_0 \in \mathcal{X}(\xi_0)$, comprenant aussi le programme de production cohérent avec ces prochaines ventes *day-ahead* et le scénario déterministe. En particulier, le premier niveau fait intervenir les prochaines offres de ventes *day-ahead* atteignables avec une solution $\mathbf{x}_0 \in \mathcal{X}(\xi_0)$ admissible pour le scénario déterministe. Autrement dit, les contraintes qui caractérisent le premier niveau sont celles données par l'ensemble admissible $\mathcal{X}(\xi_0)$ (le même que celui du problème d'optimisation déterministe (3.19)).

3.3.3.4 Second niveau

Le second niveau permet de définir les décisions provisoires pour chaque scénario probabiliste à partir des décisions définitives du premier niveau. Dans le cadre de la thèse, les décisions provisoires sont :

— le programme de production (pour un scénario probabiliste donné) sur toute la fenêtre temporelle,

— les prochaines offres de vente *day-ahead* qui ne correspondent pas à la prochaine session du marché *day-ahead*.

Ces décisions provisoires sont utilisées pour déterminer le chiffre d'affaires atteignable à partir des offres de vente de la prochaine session du marché *day-ahead* du premier niveau en fonction des scénarios qui peuvent se réaliser.

Elles sont donc obtenues à partir du problème d'optimisation (3.18) avec chaque scénario probabiliste, en considérant que :

— les offres de vente de la prochaine session du marché *day-ahead*, fixées au premier niveau, seront retenues,

— le scénario probabiliste considéré se réalise effectivement.

Une partie des contraintes qui caractérisent le second niveau reprend donc les contraintes des ensembles admissibles $\mathcal{X}(\xi_n)$, pour $n \in \{1, \dots, N\}$. De même, les fonctions objectifs qui caractérisent le second niveau sont données par les fonctions $\mathbf{x} \mapsto f(\mathbf{x}, \xi_n)$, pour $n \in \{1, \dots, N\}$. Les corrections apportées aux planifications ultérieures sont alors anticipées au moment de l'optimisation, mais avec les informations connues à l'instant présent (les situations associées aux planifications ultérieures n'étant pas accessibles à l'instant présent).

Comme le premier et le second niveau font intervenir des contraintes similaires, ils modélisent un même système : seul l'état de ce système change d'un niveau à l'autre. On retrouve alors l'idée de l'optimisation dynamique stochastique, avec laquelle le problème d'optimisation probabiliste que l'on cherche à définir peut aussi s'interpréter.

3.3.3.5 Revue de la littérature

L'optimisation stochastique multi-niveaux et l'optimisation dynamique stochastique sont les méthodes d'optimisation sous incertitudes les plus courantes dans la littérature pour la gestion d'un système de production hydroélectrique (parfois couplée avec des productions variables) [Römisch et Vigerske, 2010]. Bien que la littérature à ce sujet soit très riche, les contextes sont souvent très différents de celui de notre étude, notamment avec une stratégie de valorisation différente. En général, les niveaux ne correspondent pas à l'enchaînement des décisions des ventes, mais à différents pas de temps, formant un arbre de décision à plusieurs niveaux. Les incertitudes sont alors modélisées à partir d'un arbre de scénarios à plusieurs niveaux, construit par exemple à partir des prévisions d'ensemble [Raso *et al.*, 2013, 2014 ; Faber et Stedinger, 2001 ; Séguin *et al.*, 2017]. Les références [Fleten et Kristoffersen, 2008 ; Haneveld et van der Vlerk, 2020 ; Ruszczyński et Shapiro, 2003 ; Conejo *et al.*, 2010 ; Morales *et al.*, 2014 ; Wallace et Fleten, 2003 ; Nürnberg et Römisch, 2002] exposent néanmoins le rôle des incertitudes dans la gestion d'un système énergétique et utilisent une formulation mathématique du problème d'optimisation probabiliste assez similaire à celle proposée dans ce manuscrit.

Pour la gestion de centrales hydroélectriques, il est aussi courant d'utiliser des méthodes par optimisation robuste ou des contraintes en probabilité (voir Sous-section 3.1.4), de manière à limiter les futures corrections du programme de production (la solution étant ainsi construite pour être applicable pour la plupart des scénarios qui peuvent se réaliser) [van Ackooij, 2017]. Les contraintes concernées sont généralement celles qui limitent le volume de retenue. Ces approches sont surtout utiles lorsque la robustesse du programme de production est cruciale en pratique, ce qui n'est pas le cas pour notre étude où nous considérons que le programme de production peut être corrigé et réoptimisé plusieurs fois par jour. L'optimisation robuste n'est pas adaptée pour le cas d'étude car elle fournirait une solution « plate » largement sous-optimale, voire aucune solution, car l'ensemble $\bigcap_{1 \leq n \leq N} \mathcal{X}(\xi_n)$ est peu riche, voire vide pour le cas d'étude. Par ailleurs, avec les contraintes en probabilité, les paramètres incertains sont généralement considérés comme indépendants et parfois définis selon une loi de probabilité paramétrique donnée (loi normale par exemple). Dans ce cas, on perdrait la richesse offerte par les prévisions d'ensemble qui forment les scénarios multivariés (perte des cohérences spatio-temporelles et déformation des lois de probabilité).

Des états de l'art plus complets sont disponibles par exemple dans [van Ackooij *et al.*, 2018 ; Labadie, 2004 ; Osório *et al.*, 2013 ; Tahanan *et al.*, 2015].

3.3.4 Problème d'optimisation sous incertitudes

Dans cette sous-section, nous formalisons la méthode d'optimisation sous incertitudes présentée à la sous-section 3.3.3.

Le premier niveau du problème d'optimisation stochastique que l'on cherche à définir fait intervenir les variables de décision associées aux offres de vente à effectuer à la prochaine session du marché *day-ahead*. On considère donc la fonction linéaire $b : \mathbb{R}^r \to \mathbb{R}^{24}$ définie par [4] :

$$\forall \mathbf{x} \in \mathbb{R}^r, \quad b(\mathbf{x}) = \begin{pmatrix} b_{h_1}(\mathbf{x}) \\ \vdots \\ b_{h_{24}}(\mathbf{x}) \end{pmatrix} \in \mathbb{R}^{24},$$

où $(h_1, \ldots, h_{24}) \in \mathcal{T}^{24}$ désignent les premiers pas de temps de chaque heure de la journée de livraison associée à la prochaine session du marché *day-ahead*. Pour le cas d'étude, il s'agit des 24 heures du lendemain : $\forall j \in \{1, \ldots, 24\},\ h_j = 73 + 6(j-1)$.

Les variables de décision du premier niveau doivent être compatibles avec le scénario déterministe ξ_0. Leur ensemble de définition est donc l'ensemble des choix possibles des offres de vente à effectuer à la prochaine session du marché *day-ahead* qui sont compatibles avec le scénario déterministe ξ_0, i.e.

$$\mathcal{B}_0 = b(\mathcal{X}(\xi_0)) = \{b(\mathbf{x}_0),\ \mathbf{x}_0 \in \mathcal{X}(\xi_0)\} \subset \mathbb{R}^{24}.$$

Le second niveau correspond aux chiffres d'affaires potentiels que l'on peut atteindre à partir de la décision des prochaines ventes *day-ahead* et en suivant chacun des scénarios probabilistes sous l'hypothèse qu'ils se réalisent effectivement. Ces chiffres d'affaires potentiels correspondent aux valeurs optimales du problème (3.18) de la planification de la production en univers déterministe en fixant les variables de décision associées aux prochaines ventes *day-ahead* et le scénario. Plus précisément, étant donnés une décision $b_0 \in \mathcal{B}_0$ du premier niveau et un scénario $\xi_n \in \hat{\Xi}(\Omega)$, le second niveau correspond au problème d'optimisation (3.18) avec le scénario ξ_n en fixant les prochaines ventes *day-ahead* b_0, i.e.

$$\max_{\substack{\mathbf{x}_n \in \mathcal{X}(\xi_n) \\ b(\mathbf{x}_n) = b_0}} f(\mathbf{x}_n, \xi_n).$$

On considère donc, plus généralement, la fonction $Q : \mathbb{R}^{24} \times (\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T) \to [-\infty, +\infty[$ des valeurs optimales du second niveau définie par :

$$\forall (b_0, \xi) \in \mathbb{R}^{24} \times (\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T), \quad Q(b_0, \xi) = \max_{\substack{\mathbf{x} \in \mathcal{X}(\xi) \\ b(\mathbf{x}) = b_0}} f(\mathbf{x}, \xi) \in \mathbb{R} \cup \{-\infty\}. \tag{3.21}$$

Comme les ensembles $\mathcal{X}(\xi)$, pour $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$, sont supposés bornés, la fonction Q n'atteint pas la valeur $+\infty$ et elle atteint la valeur $-\infty$ si et seulement si aucune solution n'existe (i.e. pour $(b_0, \xi) \in \mathbb{R}^{24} \times (\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T)$, $Q(b_0, \xi) \in \mathbb{R}$ si et seulement si $\mathcal{X}(\xi) \neq \emptyset$ et $b_0 \in b(\mathcal{X}(\xi))$. Dans la suite du manuscrit, quitte à relâcher les contraintes qui limitent les écarts (voir Équation (3.15)), on suppose que le problème d'optimisation (3.21) admet une solution optimale, i.e. $Q(b_0, \xi) \in \mathbb{R}$, avec les prochaines ventes *day-ahead* $b_0 \in \mathcal{B}_0$ et avec les scénarios $\xi \in \{\xi_0, \xi_1, \ldots, \xi_N\}$.

4. La notation b fait référence au terme anglais *bids* pour désigner des offres en bourse.

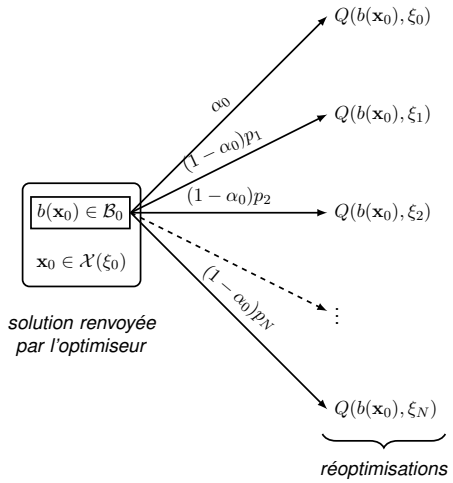


FIGURE 3.10 – Schéma de l'arbre de décision avec les notations du problème de planification de la production. Le scénario déterministe est ici associé à une feuille supplémentaire de l'arbre de poids α_0.

Il convient de noter que la fonction Q dépend de la situation et des paramètres de l'utilisateur pour configurer l'optimiseur, tout comme le scénario et l'ensemble des solutions admissibles. En outre, en appliquant la fonction Q avec les prochaines ventes *day-ahead* de la solution déterministe $\mathbf{x}_0^*$ et le scénario déterministe ξ_0, on a l'égalité (vérifiée en pratique au MIP-gap relatif seuil de 2% près) :

$$Q(b(\mathbf{x}_0^*), \xi_0) = f(\mathbf{x}_0^*, \xi_0). \tag{3.22}$$

La valeur d'une décision $b_0 \in \mathbb{R}^{24}$ des prochaines ventes *day-ahead* pourrait être évaluée par la moyenne des chiffres d'affaires portée sur les scénarios probabilistes :

$$\mathbb{E}[Q(b_0, \hat{\Xi})] = \sum_{n=1}^{N} p_n Q(b_0, \xi_n) \in \mathbb{R}. \tag{3.23}$$

Dans ce cas, le scénario déterministe serait utilisé pour définir la décision $b_0 \in \mathbb{R}^{24}$ mais pas pour évaluer sa valeur économique. Pour tenir compte du scénario déterministe dans l'évaluation de la valeur économique des prochaines ventes *day-ahead* $b_0 \in \mathbb{R}^{24}$, on considère donc plus généralement l'expression

$$\alpha_0 Q(b_0, \xi_0) + (1-\alpha_0)\mathbb{E}[Q(b_0, \hat{\Xi})] \in \mathbb{R}, \tag{3.24}$$

où $\alpha_0 \in [0,1]$ est un poids arbitraire du scénario déterministe ξ_0 par rapport aux scénarios probabilistes. Si $\alpha_0 = 0$, alors on retrouve l'équation (3.23). Si $\alpha_0 = 1$, alors les scénarios probabilistes ne sont pas utilisés, donc on se retrouve dans un cadre déterministe. La nouvelle expression (3.24) revient à remplacer la variable aléatoire $\hat{\Xi}$ dans l'équation (3.23) par la nouvelle variable aléatoire discrète finie $\hat{\Xi}_0$ définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ par :

— $\hat{\Xi}_0(\Omega) = \{\xi_0, \xi_1, \ldots, \xi_N\}$,
— $\mathbb{P}[\hat{\Xi}_0 = \xi_0] = \alpha_0$ et $\forall n \in \{1, \ldots, N\},\ \mathbb{P}[\hat{\Xi}_0 = \xi_n] = (1-\alpha_0)p_n$.

La figure 3.10 illustre l'arbre de décision présenté à la figure 3.9 avec les notations précédentes et en ajoutant une feuille supplémentaire de poids α_0 pour le scénario déterministe.

Le poids $\alpha_0 \in [0,1]$ est arbitraire mais il peut être défini, par exemple, par le poids d'un *cluster* central des scénarios probabilistes. Le *cluster* central pourrait être obtenu suivant le principe de l'algorithme *tubing*, utilisé dans la littérature pour visualiser et interpréter les incertitudes

des prévisions météorologiques d'ensemble [Atger, 1999] (l'algorithme *tubing* est présenté plus en détails dans la sous-section 4.3.6 du chapitre suivant). Dans ce cas, le *cluster* central serait le plus petit ensemble constitué des scénarios les plus proches du scénarios moyen (selon une distance à définir) dont la variance expliquée est supérieure ou égale à un seuil fixé (ou un autre critère d'arrêt). Plus le *cluster* central contient une grande proportion des scénarios probabilistes, plus les incertitudes sont faibles (prévisions probabilistes fines) et plus le poids α_0 est proche de 1, rapprochant la prévision probabiliste à une prévision déterministe. Le *cluster* central pour le cas d'étude est présenté dans la sous-section 4.3.6 du chapitre suivant. Le poids α_0 pourrait également être défini comme la probabilité du scénario déterministe obtenue par application de la redistribution optimale de Kantorovich de l'annexe A.3.3 après l'ajout du scénario déterministe (ce qui changerait aussi les probabilités des scénarios probabilistes).

L'importance des incertitudes dans l'évaluation de la valeur économique de la décision $b_0 \in \mathbb{R}^{24}$ des prochaines ventes *day-ahead* n'est pas seulement représentée par le poids α_0, mais aussi par la différence entre les réels $Q(b_0, \xi_0)$ et $\mathbb{E}[Q(b_0, \hat{\Xi})]$. En effet, si $Q(b_0, \xi_0) \approx \mathbb{E}[Q(b_0, \hat{\Xi})]$, alors

$$\forall \alpha_0 \in [0,1], \quad \alpha_0 Q(b_0, \xi_0) + (1-\alpha_0)\mathbb{E}[Q(b_0, \hat{\Xi})] \approx Q(b_0, \xi_0),$$

ce qui revient à ne plus utiliser les scénarios probabilistes et donc de considérer un cadre déterministe (même avec $\alpha_0 < 1$).

Le problème d'optimisation de la planification de la production en univers probabiliste consiste alors à déterminer les prochaines ventes *day-ahead* $b_0 = b(\mathbf{x}_0) \in \mathcal{B}_0$, compatibles avec une solution $\mathbf{x}_0 \in \mathcal{X}(\xi_0)$ admissible pour le scénario déterministe, et qui maximisent la valeur économique en moyenne sur les scénarios probabilistes. On forme alors le problème de programmation stochastique à deux niveaux suivant :

$$\max_{b_0 \in \mathcal{B}_0} \mathbb{E}[Q(b_0, \hat{\Xi}_0)]. \tag{3.25}$$

On parle ici de « programmation » plutôt que d'« optimisation » stochastique car les deux niveaux correspondent à des problèmes MILP (contraintes et fonctions objectifs linéaires mixtes). Le problème d'optimisation (3.25) s'écrit également, en séparant le scénario déterministe des scénarios probabilistes, de la manière suivante :

$$\max_{b_0 \in \mathcal{B}_0} \left\{ \alpha_0 Q(b_0, \xi_0) + (1-\alpha_0)\mathbb{E}[Q(b_0, \hat{\Xi})] \right\}. \tag{3.26}$$

On utilise l'espérance dans la fonction objectif du problème d'optimisation de l'équation (3.26) pour être neutre face au risque dans la prise de décision des prochaines offres de vente *day-ahead*. Si les prévisions probabilistes sont suffisamment fiables, alors, pour tout $b_0 \in \mathcal{B}_0$, l'espérance $\mathbb{E}[Q(b_0, \hat{\Xi})]$ représente approximativement le gain « réel » moyen que l'on peut atteindre avec la décision des prochaines offres de vente *day-ahead* b_0. D'un point de vue économique, le choix d'un programme de production qui fournit une très grande moyenne mais une grande variance (dispersion) du chiffre d'affaires pourrait ne pas constituer une bonne décision (risque important). Il pourrait alors être nécessaire d'ajouter une aversion au risque dans l'optimisation afin de faire un compromis entre la maximisation de la moyenne et la minimisation de la variance du chiffre d'affaires. Même si elle n'est pas considérée dans note étude, l'aversion au risque est une problématique courante dans la littérature sur la gestion en univers probabiliste d'un système de production électrique, notamment lorsque l'on considère les incertitudes sur les prix [García-González *et al.*, 2007 ; Castillo *et al.*, 2019]. La mesure de risque généralement utilisée est la valeur du risque conditionnel CVaR qui a l'avantage de bien s'intégrer dans un problème d'optimisation linéaire (voir §3.1.4.3) [Castillo *et al.*, 2019].

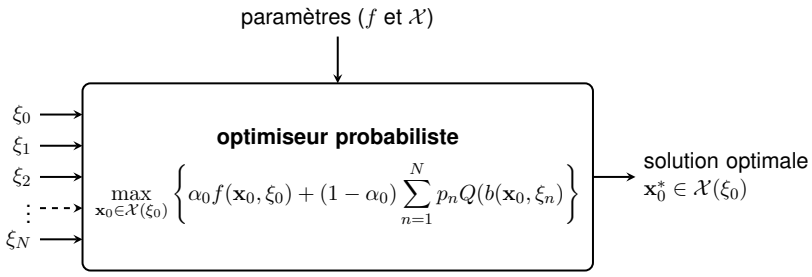


FIGURE 3.11 – Schéma de l'optimiseur en univers probabiliste.

Le problème d'optimisation probabiliste (3.26) s'interprète par la programmation stochastique à deux niveaux : la fonction Q, valeur optimale d'un problème MILP, quantifie le chiffre d'affaires potentiel atteignable en s'adaptant à la décision des prochaines ventes *day-ahead* et au scénario qui peut se réaliser. Néanmoins, qu'il soit en univers déterministe ou probabiliste, l'optimiseur doit renvoyer une solution du problème de planification de la production, comme définie à la sous-section 3.2.3, qui comprend toutes les prochaines offres de ventes au marché *day-ahead* (et pas seulement celles de la prochaine session) et, surtout, le programme de production. Il faut donc réécrire le problème d'optimisation (3.26) de manière à expliciter cette solution optimale, qui est celle cachée par l'ensemble $\mathcal{B}_0$. En utilisant la définition de l'ensemble $\mathcal{B}_0$, le problème d'optimisation (3.26) est en effet équivalent au problème d'optimisation

$$\max_{\substack{b_0 = b(\mathbf{x}_0) \\ \mathbf{x}_0 \in \mathcal{X}(\xi_0)}} \left\{ \alpha_0 \left(\max_{\substack{\mathbf{y}_0 \in \mathcal{X}(\xi_0) \\ b(\mathbf{y}_0) = b_0}} f(\mathbf{y}_0, \xi_0) \right) + (1-\alpha_0) \mathbb{E}[Q(b_0, \hat{\Xi})] \right\},$$

i.e., en combinant les deux problèmes d'optimisation explicités,

$$\max_{\substack{b_0 = b(\mathbf{x}_0) \\ \mathbf{x}_0 \in \mathcal{X}(\xi_0)}} \max_{\substack{\mathbf{y}_0 \in \mathcal{X}(\xi_0) \\ b(\mathbf{y}_0) = b_0}} \left\{ \alpha_0 f(\mathbf{y}_0, \xi_0) + (1-\alpha_0) \mathbb{E}[Q(b_0, \hat{\Xi})] \right\},$$

ou encore, en retirant les variables de décision redondantes :

$$\max_{\mathbf{x}_0 \in \mathcal{X}(\xi_0)} \left\{ \alpha_0 f(\mathbf{x}_0, \xi_0) + (1-\alpha_0) \mathbb{E}[Q(b(\mathbf{x}_0), \hat{\Xi})] \right\}. \tag{3.27}$$

Cette dernière formulation équivalente du problème d'optimisation en univers probabiliste est celle qui est principalement utilisée dans la suite du manuscrit. Sous cette forme, le problème d'optimisation s'interprète davantage avec la programmation dynamique stochastique : les deux niveaux reposent sur la construction d'une solution, le second niveau permettant d'anticiper la planification ultérieure en fonction de la solution optimale et du scénario qui peut se réaliser. On modélise ainsi l'enchaînement des planifications de la production du processus opérationnel. Le problème d'optimisation (3.21) qui définit la fonction Q est donc une estimation de la planification ultérieure, à la différence que la fenêtre temporelle n'est pas décalée dans le temps, la situation associée à la planification ultérieure (mise à jour des prévisions, valeurs réalisées du programme de production etc.) n'étant pas accessible à la planification courante. Avec les notations de cette sous-section, le schéma de la figure 2.18 devient donc celui de la figure 3.11.

Le problème d'optimisation (3.27) est une généralisation du problème d'optimisation (3.19) car ils sont tous les deux identiques lorsque $\alpha_0 = 1$, i.e. en univers déterministe. En outre, la prise en compte des incertitudes revient à pondérer la fonction objectif, de manière à modéliser les opportunités financières liées aux réoptimisations possibles du programme de production à la planification ultérieure lorsqu'une partie des incertitudes auront été dévoilées.

Le problème d'optimisation (3.27) peut également se réécrire de la manière suivante :

$$\max_{\mathbf{x}_0 \in \mathcal{X}(\xi_0)} \left\{ f(\mathbf{x}_0, \xi_0) + \tilde{\mathcal{R}}(\mathbf{x}_0, \xi_0) \right\},$$

où, pour tout $\mathbf{x}_0 \in \mathcal{X}(\xi_0)$, $\tilde{\mathcal{R}}(\mathbf{x}_0, \xi_0) = \mathbb{E}[(1-\alpha_0)(Q(b(\mathbf{x}_0), \hat{\Xi}) - f(\mathbf{x}_0, \xi_0))]$ est une estimation du gain algébrique moyen (pouvant être positif ou négatif) apporté par les mises à jour de la solution $\mathbf{x}_0$ aux planifications ultérieures. La moyenne est portée sur les scénarios probabilistes de la planification courante, qui modélisent qu'une partie des sources d'incertitudes. Le terme $\tilde{\mathcal{R}}$ est donc une estimation partielle du terme $\mathcal{R}$ de l'équation (3.20). Néanmoins, si les scénarios probabilistes sont suffisamment fiables, alors le terme $\tilde{\mathcal{R}}$ permet d'améliorer l'estimation du chiffre d'affaires réel par rapport au cadre déterministe (où seule la fonction f est utilisée).

3.3.5 Optimisation en univers déterministe ou probabiliste ?

Avant de nous intéresser à la recherche d'une solution optimale pour le problème d'optimisation probabiliste (3.27), nous analysons dans cette sous-section le comportement de la fonction objectif du problème d'optimisation probabiliste (3.27) avec la solution déterministe $\mathbf{x}_0^* \in \mathcal{X}(\xi_0)$ de la sous-section 3.2.6. On rappelle que la solution déterministe $\mathbf{x}_0^* \in \mathcal{X}(\xi_0)$ est la solution optimale du problème d'optimisation déterministe (3.19), et elle est admissible pour le problème d'optimisation probabiliste (3.27). On note $b_0^* = b(\mathbf{x}_0^*) \in \mathcal{B}_0$ les prochaines ventes *day-ahead* de la solution déterministe.

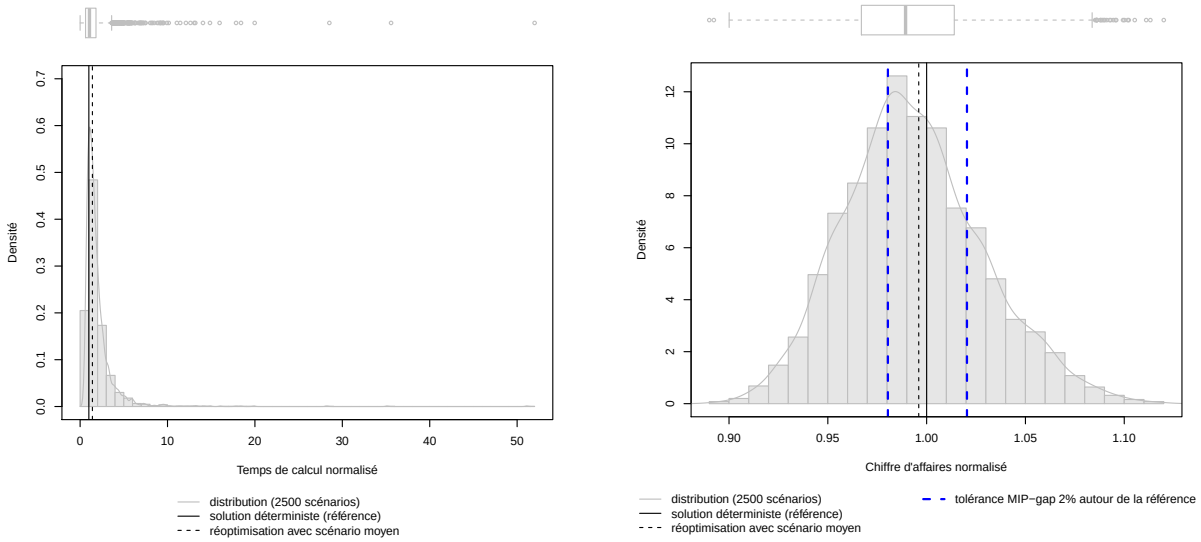
Pour chaque scénario probabiliste $\xi_n \in \hat{\Xi}(\Omega)$, la valeur $Q(b_0^*, \xi_n) \in \mathbb{R}$ est obtenue par résolution numérique du problème MILP (3.21). Les temps de calcul et les chiffres d'affaires optimaux sont présentés à la figure 3.12. On observe sur la figure 3.12a que les temps de calcul sont en moyenne 98,65% plus élevés que celui de la solution déterministe, ce qui est supérieur aux résultats observés à la sous-section 3.3.2. Cela s'explique par le fait que le problème d'optimisation (3.21) est plus contraint que le problème d'optimisation (3.18). On observe sur la figure 3.12b que les valeurs de la variable aléatoire discrète réelle $Q(b_0^*, \hat{\Xi})$ sont environ 0,6% inférieures au chiffre d'affaires de la solution déterministe, ce qui est inférieur au MIP-gap relatif seuil de 2%. Ainsi, pour le cas d'étude, on observe :

$$\mathbb{E}[Q(b_0^*, \hat{\Xi})] \approx f(\mathbf{x}_0^*, \xi_0) \approx Q(b_0^*, \xi_0), \tag{3.28}$$

la dernière approximation venant de l'équation (3.22) en tenant compte du MIP-gap relatif seuil de 2% utilisé par le solveur MILP. En outre, l'écart-type de la variable aléatoire discrète finie réelle $Q(b_0^*, \hat{\Xi})$ est un indicateur du risque associé aux opportunités offertes par les réoptimisations possibles de la solution déterministe. On observe que cet écart-type est de 0,035, légèrement supérieur à celui obtenu à la sous-section 3.3.2, et supérieur au MIP-gap relatif seuil de 2%. Ainsi, même en fixant les prochaines ventes *day-ahead*, les incertitudes ont toujours un impact non négligeable sur le chiffre d'affaires.

Le tableau 3.2 présente les résultats de l'analyse de sensibilité analogue à celle effectuée à la sous-section 3.3.2 mais pour le problème d'optimisation (3.21) avec les ventes *day-ahead* b_0^* de la solution déterministe. On observe que la sensibilité du temps de calcul est similaire à celle de la sous-section 3.3.2.

Ainsi, le comportement de l'optimiseur semble ne pas changer si l'on impose les prochaines ventes *day-ahead*. Il est difficile d'interpréter ce résultat sans réaliser des études complémentaires avec d'autres cas d'études.



(a) Distribution des temps de calcul de $Q(b_0^*, \cdot)$ (b) Distribution de $Q(b_0^*, \hat{\Xi})/f(\mathbf{x}_0^*, \xi_0)$

FIGURE 3.12 – Temps de calcul et chiffre d'affaires du second niveau à partir de la solution déterministe pour le cas d'étude. Les valeurs sont normalisées par rapport à celles du problème d'optimisation déterministe. Les fonctions de densité sont obtenues par estimation par noyaux.

Type paramètre	*Temps de calcul*	*Chiffre d'affaires optimal*
Apports	0,15	0,74
Prix	0,06	0,26
Interactions	0,79	0,00

TABLEAU 3.2 – Indices de Sobol du premier ordre des apports et des prix de la fonction $Q(b_0^*, \cdot)$.

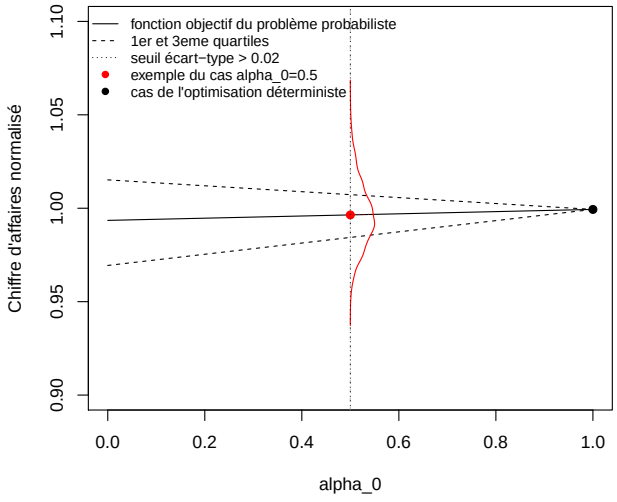


FIGURE 3.13 – Évolution en fonction du paramètre α_0 de la fonction objectif normalisée du problème d'optimisation probabiliste appliquée à la solution déterministe $\mathbf{x}_0^*$ du cas d'étude.

La figure 3.13 représente l'évolution, en fonction du paramètre α_0, de la valeur de la fonction objectif normalisée du problème d'optimisation probabiliste (3.27) pour la solution déterministe $\mathbf{x}_0^*$. Plus précisément, la ligne noire continue est la courbe représentative de la fonction affine

$$\alpha_0 \mapsto \frac{\alpha_0 f(\mathbf{x}_0^*, \xi_0) - (1-\alpha_0)\mathbb{E}[Q(b_0^*, \hat{\Xi})]}{f(\mathbf{x}_0^*, \xi_0)} = E_0^* + (1 - E_0^*)\alpha_0, \text{ avec } E_0^* = \frac{\mathbb{E}[Q(b_0^*, \hat{\Xi})]}{f(\mathbf{x}_0^*, \xi_0)} \in \mathbb{R},$$

et les lignes noires en tirets sont obtenues en remplaçant l'espérance par le premier et le troisième quartile dans l'équation précédente. Ces quartiles sont un indicateur du risque associé aux répercutions sur le chiffre d'affaires des réoptimisations possibles de la solution déterministe. Pour le cas d'étude, on a $\mathbb{E}[Q(b_0^*, \hat{\Xi})] \leq f(\mathbf{x}_0^*, \xi_0)$, i.e. $E_0^* \leq 1$, d'où la croissante de la courbe noire continue de la figure 3.13. La fonction de densité et le point en rouge illustrent le cas particulier où $\alpha_0 = 0{,}5$. On rappelle que le cas $\alpha_0 = 1$ revient à considérer le problème d'optimisation déterministe (3.19), d'où un chiffre d'affaires normalisé égal à 1. Au contraire, le chiffre d'affaires normalisé pour le cas $\alpha_0 = 0$ est identique à celui qui suit la distribution de la figure 3.12b.

L'écart-type du chiffre d'affaires décroît linéairement de 0,035 (écart-type de la figure 3.12b) à 0 lorsque le paramètre α_0 croît de 0 à 1. Il existe donc un seuil $\alpha_0^*(b_0^*) \in [0, 1]$ sur le paramètre α_0, représenté par la ligne verticale en pointillées sur la figure 3.13 à environ $\alpha_0^*(b_0^*) = 0{,}5$ pour le cas d'étude, au-delà duquel l'écart-type devient inférieur au MIP-gap seuil de 2%. Autrement dit, avec un paramètre α_0 supérieur au seuil $\alpha_0^*(b_0^*)$, le risque associé aux répercutions sur le chiffre d'affaires des réoptimisations possibles de la solution déterministe devient négligeable, si bien que la fonctions objectif déterministe et celle probabiliste sont équivalentes (aux erreurs du MIP-gap près). Bien sûr, le seuil $\alpha_0^*(b_0^*)$ dépend *a priori* de la solution $\mathbf{x}_0^*$ (via les prochaines ventes *day-ahead* $b_0^* = b(\mathbf{x}_0^*)$), donc il n'est *a priori* pas accessible avant de lancer l'optimisation. Une estimation globale de ce seuil pourrait être obtenue à partir de l'analyse de plusieurs cas d'étude.

Le choix du paramètre α_0 n'est pas évident en pratique. Si le paramètre α_0 est défini à partir d'un *cluster* central (voir Sous-section 3.3.4), alors il mesure l'importance des incertitudes : plus il est proche de 1, plus le *cluster* central est gros, donc plus les incertitudes sont faibles. Ainsi, le seuil global du paramètre α_0 permettrait de déterminer la version du problème d'optimisation (déterministe ou probabiliste) à utiliser en fonction des incertitudes de la situation. Si les incertitudes sont faibles (paramètre α_0 du *cluster* central supérieur au seuil global), alors il est préférable d'utiliser le problème d'optimisation déterministe (le problème d'optimisation probabiliste, plus complexe à résoudre, n'apportant pas d'informations supplémentaires). Au contraire, si les incertitudes sont fortes (paramètre α_0 du *cluster* central inférieur au seuil global), alors il est préférable d'utiliser le problème d'optimisation probabiliste.

Néanmoins, comme expliqué dans la sous-section 3.3.4, l'importance des incertitudes dans la fonction objectif probabiliste appliquée à la solution déterministe est aussi liée à la différence entre les réels $Q(b_0^*, \xi_0)$ et $\mathbb{E}[Q(b_0^*, \hat{\Xi})]$. Or, pour le cas d'étude, cette différence est négligeable (voir Équation (3.28)) : le paramètre α_0 n'a pas d'impact sur la fonction objectif probabiliste appliquée à la solution déterministe $\mathbf{x}_0^*$, au MIP-gap relatif seuil de 2% près. C'est la raison par laquelle la ligne noire continue de la figure 3.13 est toujours comprise dans l'intervalle [0,98, 1,02]. En particulier, la fonction objectif déterministe (pour $\alpha_0 = 1$) et celle probabiliste (pour $\alpha_0 \in [0, 1[$) sont égales (au MIP-gap relatif seuil de 2% près) lorsqu'elles sont appliquées à la solution déterministe $\mathbf{x}_0^*$. Si cette observation se vérifiait pour une grande partie des solutions admissibles de $\mathcal{X}(\xi_0)$ autres que la solution déterministe, alors le passage de l'univers déterministe à l'univers probabiliste pour le cas d'étude ne serait pas pertinent. Ce point sera abordé plus en détails à la section 6.2.

3.4 Conclusion de la première partie

Nous nous sommes intéressés dans cette première partie à la formulation mathématique du problème de planification à court terme de la production conjointe d'une chaîne hydroélectrique et d'actifs de production variable. Ce problème consiste à effectuer, en même temps, la construction du programme de production hydroélectrique et la décision des prochaines ventes au marché *day-ahead*, de manière à maximiser le chiffre d'affaires. Il s'agit donc d'un problème d'optimisation. Nous nous sommes appuyés sur l'exemple du contexte particulier de CNR pour donner une formulation mathématique simple et assez générale de ce problème d'optimisation.

Même à court terme, la gestion conjointe d'une chaîne hydroélectrique et d'actifs de production variable est enclin aux incertitudes, notamment celles qui entachent les prévisions météorologiques, hydrologiques et des conditions du marché de l'énergie. Nous avons donc consacré cette première partie du manuscrit au premier objectif de la thèse : faire passer la formulation mathématique du problème d'optimisation de l'univers déterministe vers l'univers probabiliste. Nous avons abouti à un problème de programmation stochastique à deux niveaux à partir d'une modélisation des incertitudes en scénarios multivariés. Cette formulation probabiliste repose notamment sur le fait que les planifications de la production sont mises à jour régulièrement au cours du temps lors du processus opérationnel.

Dans cette première partie, nous n'avons cependant pas évoqué la résolution numérique du problème d'optimisation en univers probabiliste. Cela fait l'objet de la seconde partie du manuscrit, dans laquelle nous présentons une méthodologie de résolution par métamodèle.

Deuxième partie

Résolution numérique par métamodèle

Chapitre 4

Méthodologie générale et transformation des données d'entrée

La première partie du manuscrit a été consacrée au premier objectif de la thèse : la formulation mathématique du problème d'optimisation en univers probabiliste de la planification à court terme de la production d'une chaîne hydroélectrique et d'actifs de production variable. Dans cette seconde partie, nous nous intéressons au second objectif de la thèse : la résolution numérique de ce problème d'optimisation probabiliste. La méthodologie générale, qui repose sur l'utilisation d'un métamodèle, est présentée dans la première section de ce chapitre. La deuxième section de ce chapitre présente la réduction de la dimension des données d'entrée fonctionnelles, une étape préliminaire pour mettre en place la méthodologie générale proposée. Enfin, la troisième section de ce chapitre présente la réduction du nombre de scénarios, une des étapes de la méthodologie générale. Les étapes de la construction du métamodèle seront présentées dans le chapitre suivant.

4.1 Description méthodologique

4.1.1 Rappel du problème d'optimisation probabiliste

On rappelle que l'on s'intéresse à la planification à court terme de la production d'une chaîne hydroélectrique et d'actifs de production variable en univers probabiliste. La première partie du manuscrit conduit à la formulation du problème de programmation dynamique stochastique à deux niveaux (3.27) que l'on rappelle ci-dessous par commodité :

$$\max_{\mathbf{x}_0 \in \mathcal{X}(\xi_0)} \left\{ \alpha_0 f(\mathbf{x}_0, \xi_0) + (1-\alpha_0)\mathbb{E}[Q(b(\mathbf{x}_0), \hat{\Xi})] \right\}, \tag{4.1}$$

où (voir Sections 3.2 et 3.3) :

- $\xi_0 = (a_0, \pi_0, \phi_0) \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ est le scénario déterministe composé des prévisions déterministes des apports en eau $a_0 \in \mathbb{R}^{IT}$ (membre contrôle de la prévision d'ensemble), des prix $\pi_0 \in \mathbb{R}^T$ et des productions variables $\phi_0 \in \mathbb{R}^T$ ($I \in \mathbb{N}^*$ désigne le nombre d'aménagements de la chaîne hydroélectrique et T le nombre de pas de temps de la fenêtre temporelle sur laquelle la planification de la production est considérée),
- $\hat{\Xi}$ est une variable aléatoire discrète et finie sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$, représentant les scénarios probabilistes $\xi_n = (a_n, \pi_n, \phi_n) \in \hat{\Xi}(\Omega)$, pour $n \in \{1, \dots, N\}$, et leurs probabilités $p_n = \mathbb{P}[\hat{\Xi} = \xi_n] \in]0, 1[$, $p_1 + \cdots + p_N = 1$, ($N \in \mathbb{N}^*$ étant le nombre de scénarios probabilistes),
- $\mathcal{X}(\xi) \subset \mathbb{R}^r$ est l'ensemble des solutions admissibles pour un scénario $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ (partie fermée et bornée de $\mathbb{R}^r$, non vide si $\xi \in \{\xi_0, \xi_1, \dots, \xi_N\}$), défini à partir de contraintes MILP (i.e. linéaires mixtes),

— $\alpha_0 \in [0,1]$ est un paramètre arbitraire, qui peut être défini par le poids d'un *cluster* central des scénarios probabilistes (dans ce cas, le poids α_0 est un indicateur de la dispersion des scénarios probabilistes par rapport au scénario déterministe : plus α_0 est proche de 0, plus les incertitudes sont fortes),

— f est une fonction linéaire réelle définie sur $\mathbb{R}^r \times (\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T)$ qui représente le chiffre d'affaires d'une solution pour un scénario,

— b est une fonction linéaire de $\mathbb{R}^r$ dans $\mathbb{R}^{24}$ qui donne les prochaines ventes *day-ahead* associées à une solution,

— Q est une fonction réelle, coûteuse à évaluer numériquement, définie sur $\mathbb{R}^{24} \times (\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T)$ qui donne le chiffre d'affaires optimal que l'on peut atteindre pour une décision de ventes *day-ahead* $b_0 \in \mathbb{R}^{24}$ et un scénario $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ sous l'hypothèse qu'il se réalise effectivement (problème MILP) :

$$\forall (b_0, \xi) \in \mathbb{R}^{24} \times (\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T), \quad Q(b_0, \xi) = \max_{\substack{\mathbf{x} \in \mathcal{X}(\xi) \\ b(\mathbf{x}) = b_0}} f(\mathbf{x}, \xi) \in \mathbb{R} \cup \{-\infty\}. \tag{4.2}$$

On rappelle également que $\mathcal{B}_0 = \{b(\mathbf{x}_0),\ \mathbf{x}_0 \in \mathcal{X}(\xi_0)\} \subset \mathbb{R}^{24}$ est l'ensemble des prochaines ventes *day-ahead* compatibles avec le scénario déterministe ξ_0.

Une particularité importante du problème stochastique à deux niveaux (4.1) est que les deux problèmes d'optimisation imbriqués ont une formulation MILP.

4.1.2 Résolution numérique du problème d'optimisation probabiliste

Comme la variable aléatoire $\hat{\Xi}$ est discrète et finie, le problème de programmation stochastique à deux niveaux (4.1) peut s'écrire sous une forme MILP équivalente (voir Sous-section 3.1.4) :

$$\max_{\substack{\mathbf{x}_0 \in \mathcal{X}(\xi_0) \\ \mathbf{x}_n \in \mathcal{X}(\xi_n),\ 1 \le n \le N \\ b(\mathbf{x}_n) = b(\mathbf{x}_0),\ 1 \le n \le N}} \left\{ \alpha_0 f(\mathbf{x}_0, \xi_0) + (1 - \alpha_0) \sum_{n=1}^{N} p_n f(\mathbf{x}_n, \xi_n) \right\}. \tag{4.3}$$

Le problème MILP (4.3) peut être théoriquement résolu à l'aide d'un solveur MILP. Néanmoins, le nombre de variables de décision et de contraintes est alors en $\mathcal{O}(NIT)$ (duplication pour chaque scénario probabiliste), faisant exploser la mémoire allouée et la complexité temporelle de la résolution numérique. Cette explosion de la complexité du problème d'optimisation probabiliste en fonction du nombre de scénarios est une problématique générale de l'optimisation stochastique à deux niveaux et de l'optimisation dynamique stochastique, dont fait partie le problème d'optimisation (4.1). C'est pourquoi d'autres algorithmes de résolution sont présentés dans la littérature (voir Sous-section 3.3.4), notamment dans le domaine de la gestion d'actifs de production électrique. Ils sont généralement répartis en deux grandes catégories : les méthodes par décomposition et les méthodes par relaxation lagrangienne. Dans la suite de cette sous-section, nous donnons quelques éléments sur ces méthodes de résolution et leurs limites pour notre problème, mais pour plus de détails, voir par exemple [Römisch et Vigerske, 2010 ; Torres *et al.*, 2019 ; Sherali et Zhu, 2006].

Les méthodes par décomposition sont des améliorations des algorithmes utilisés pour la résolution des problèmes LP et MILP en univers déterministe présentés à la section 3.1 : décomposition de Benders, coupes et branchements, séparation et évaluation. Ces améliorations permettent de tenir compte de la structure en niveaux d'un problème de programmation stochastique à deux niveaux [Sen, 2005]. Les méthodes qui reposent sur la décomposition de Benders s'appliquent lorsque le second niveau est linéaire (problème LP) et ils construisent de manière itérative une approximation explicite de la valeur optimale du second niveau (convexe ou concave car le second

niveau est un problème LP) par séparation d'hyperplans [Rahmaniani *et al.*, 2017]. En plus de nécessiter la linéarité du second niveau, leur principal défaut est le temps de calcul, notamment avec des problèmes de grande dimension (fléau de la dimension) [Powell, 2007 ; van der Laan et Romeijnders, 2021]. Un exemple est l'algorithme SDDP (*Stochastic Dual Dynamic Programming*), très utilisé dans la littérature pour résoudre les problèmes d'optimisation stochastique multi-niveaux (notamment pour la gestion d'actifs hydroélectriques) [Pereira et Pinto, 1991 ; Pereira, 1989 ; Shapiro, 2011 ; Zou *et al.*, 2019]. On peut également évoquer la méthode *Integer L-Shaped* qui repose sur le principe des coupes et branchements. Elle s'applique lorsque le premier niveau ne comprend que des variables binaires et elle permet de définir la valeur optimale du second niveau par des inégalités linéaires ajoutées itérativement [Laporte et Louveaux, 1993 ; Carøe et Tind, 1998]. En plus de nécessiter uniquement des variables binaires au premier niveau, la méthode *Integer L-Shaped* conduit à des approximations de la valeur optimale du second niveau souvent peu satisfaisantes [Römisch et Vigerske, 2010]. Pour plus de détails sur les méthodes de décomposition pour le problème d'attribution d'unités de production (dont fait partie le problème d'optimisation (4.1)), voir par exemple [van Ackooij et Malick, 2016 ; van Ackooij *et al.*, 2018].

Les méthodes par relaxation lagrangienne consistent à pénaliser certaines contraintes dans la fonction objectif, permettant de décomposer le problème MILP (4.3) de grande dimension en plusieurs sous-problèmes MILP de plus petite dimension (donc plus faciles à résoudre). En général, les méthodes de relaxation lagrangienne sont complétées par des heuristiques pour combiner les solutions des sous-problèmes et gérer les éventuelles violations des contraintes relaxées, qui sont les principaux inconvénients de ces méthodes. Il existe deux grandes familles de méthodes par relaxation lagrangienne, selon qu'elles reposent sur une décomposition par scénarios ou sur une décomposition spatiale (lorsque le système physique modélisé comporte plusieurs sous-systèmes physiques, par exemple plusieurs actifs de production).

— Avec la décomposition par scénarios, la relaxation lagrangienne est appliquée avec la contrainte de non-anticipativité. Cela permet de décomposer le problème d'optimisation probabiliste en plusieurs sous-problèmes d'optimisation déterministes, un sous-problème par scénario. Pour le problème (4.3), la contrainte de non-anticipativité est donnée par $b(\mathbf{x}_n) = b(\mathbf{x}_0)$, pour $n \in \{1, \dots, N\}$, avec $(\mathbf{x}_0, \mathbf{x}_n) \in \mathcal{X}(\xi_0) \times \mathcal{X}(\xi_n)$, et chaque sous-problème revient à résoudre un problème similaire au problème (3.18) en univers déterministe. La approches dites de *Progressive Hedging* [Rockafellar et Wets, 1991 ; Watson et Woodruff, 2011] ou de décomposition duale [Carøe et Schultz, 1999] reposent sur cette approche. La principale différence entre les deux concerne la combinaison des solutions des sous-problèmes et la gestion des éventuelles violations de la contrainte de non-anticipativité. Pour des exemples d'utilisation dans le domaine de la gestion d'actifs de production électrique, voir par exemple [Takriti *et al.*, 1996 ; Finardi et Scuzziato, 2014 ; Gröwe-Kuska *et al.*, 2002 ; Schulze *et al.*, 2017].

— Avec la décomposition spatiale, la relaxation lagrangienne est appliquée avec les contraintes qui créent la dépendance entre les sous-systèmes du système physique modélisé par le problème d'optimisation (par exemple les différentes centrales d'une chaîne hydroélectrique). Cela permet d'avoir des sous-problèmes, un pour chaque sous-système, plus faciles à résoudre (même si, dans ce cas, les sous-problèmes sont en univers probabiliste). Dans la littérature, cette approche est par exemple utilisée pour la gestion de la production d'une chaîne hydroélectrique [Carpentier *et al.*, 2018, 1996]. Pour un problème d'attributions d'unités de production avec des actifs hydroélectriques, une comparaison effectuée dans [Scuzziato *et al.*, 2017] montre que la décomposition spatiale offre un meilleur compromis entre le temps de calcul et l'optimalité que la décomposition par scénarios.

Les problèmes d'optimisation stochastique à deux niveaux peuvent également être résolus à l'aide d'algorithmes par approximations successives qui consistent à remplacer la valeur optimale du second niveau (la fonction Q pour le problème d'optimisation probabiliste (4.1)) par un émulateur statistique (métamodèle) construit itérativement, dans le même esprit que les méthodes par décomposition. La littérature à ce sujet est assez pauvre et elle concerne surtout des problèmes d'optimisation quadratiques sans variables discrètes [Longstaff et Schwartz, 2001 ; Deák *et al.*, 2012 ; Deák, 2011 ; Flammarion, 2017 ; van der Vlerk, 2010 ; Romeijnders *et al.*, 2016 ; Bernreuther, 2017]. Ces algorithmes sont, à notre connaissance, peu utilisés pour la gestion d'actifs de production électrique.

Les algorithmes généraux présentés et utilisés dans la littérature ne semblent pas adaptés pour résoudre le problème de programmation stochastique à deux niveaux (4.1) car les deux niveaux sont associés à des problèmes MILP de grande dimension. Même si certains de ces algorithmes pourraient être appliqués avec le problème (4.1) (mais sans assurer leur convergence vers une solution satisfaisante, les hypothèses n'étant pas toutes respectées), il existe une autre problématique qui est cruciale dans notre contexte : le temps de calcul. L'optimiseur que l'on cherche à faire passer en univers probabiliste a en effet vocation à être un outil d'aide à la décision pour libérer du temps d'expertise en opérationnel lors de la planification et de la construction du programme de production (voir §2.3.5.2). Afin d'être compatible avec les délais opérationnels, l'optimiseur doit renvoyer une solution optimale en moins d'une heure. En pratique, cet objectif est difficile à atteindre, même en univers déterministe (voir §2.3.5.2). Comme le passage de l'univers déterministe à l'univers probabiliste complexifie *a priori* l'optimiseur, le risque est de faire exploser les temps de calcul. Il faut donc être vigilant sur le choix des méthodes d'optimisation probabiliste. Pour le processus opérationnel, il est préférable de rester en univers déterministe (au risque d'être moins optimal face aux incertitudes), voire de continuer à utiliser l'optimisation semi-manuelle (au risque d'être moins optimal en chiffre d'affaires), que de dépasser le délai maximal pour la planification de la production. Bien que cette question des temps de calcul soit cruciale en opérationnel, elle se pose moins pour la thèse où il n'y a pas de délais opérationnels à respecter. Dans notre étude, la priorité est donc donnée à la méthodologie d'optimisation sous incertitudes, mais le temps de calcul est un des critères pour mesurer la qualité de la méthodologie que nous présentons dans ce manuscrit (voir Sous-section 6.2.3).

4.1.3 Méthodologie générale proposée

Contrairement à la plupart des problèmes rencontrés dans la littérature, les deux niveaux du problème d'optimisation probabiliste (4.1) sont linéaires mixtes et de grande dimension. Cette spécificité entraîne les trois principales difficultés suivantes [Ahmed, 2010] :

1. L'évaluation de l'espérance $\mathbb{E}[Q(b_0, \hat{\Xi})] = \sum_{n=1}^{N} p_n Q(b_0, \xi_n)$ pour une décision $b_0 \in \mathcal{B}_0$ nécessite N évaluations de la fonction coûteuse Q.
2. L'évaluation de $Q(b_0, \xi_n)$ pour une décision $b_0 \in \mathcal{B}_0$ et un scénario probabiliste $\xi_n \in \hat{\Xi}(\Omega)$ est coûteuse en temps et en mémoire car elle nécessite la résolution d'un problème MILP de grande dimension.
3. La fonction $\mathcal{E} : b_0 \in \mathcal{B}_0 \mapsto \mathbb{E}[Q(b_0, \hat{\Xi})]$ n'a pas des propriétés mathématiques qui pourraient faciliter l'optimisation globale du problème (4.1) (concavité par exemple).

Pour résoudre numériquement le problème d'optimisation probabiliste (4.1), il faut pouvoir surmonter ces trois difficultés, comme le font (au moins en partie) les algorithmes évoqués à la sous-section 4.1.2 sous certaines hypothèses. Nous proposons quelques solutions adaptées au problème d'optimisation probabiliste (4.1) pour surmonter chacune des trois difficultés et qui peuvent être combinées entre elles.

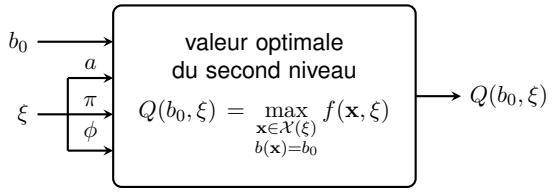


FIGURE 4.1 – Schéma type « boîte noire » de la valeur optimale du second niveau.

— La première difficulté peut être surmontée en réduisant le nombre N de scénarios (réduction en amont, ou tirage type Monte-Carlo). Il faut alors veiller à ne pas trop déformer la loi de probabilité des scénarios pour conserver une fiabilité suffisante de la modélisation des incertitudes. La solution optimale du problème d'optimisation probabiliste (4.1) serait en effet peu pertinente avec une mauvaise fiabilité (voir §1.4.5.2).

— La deuxième difficulté peut être surmontée en utilisant une approximation $\hat{Q}$ de la fonction Q rapide à évaluer, obtenue, par exemple, par dégradation du problème MILP (4.2) (suppression de certaines contraintes, élargissement de la résolution temporelle etc.), relaxation des variables discrètes ou par émulation statistique (métamodèle).

— La troisième difficulté peut être surmontée en utilisant une approximation $\hat{\mathcal{E}}$ de la fonction $\mathcal{E}$ ayant des propriétés mathématiques qui facilitent l'optimisation globale (concavité, linéarité par morceaux pour utiliser un solveur MILP etc.). Néanmoins, ces simplifications peuvent avoir un impact non négligeable sur le problème d'optimisation. Selon certaines hypothèses ou simplifications sur la fonction Q, l'approximation $\hat{\mathcal{E}}$ peut être obtenue grâce à une enveloppe concave linéaire par morceaux ou en adaptant les méthodes par décomposition et par approximations successives habituellement utilisées pour la résolution des problèmes d'optimisation à deux niveaux lorsque le second niveau est linéaire (voir Sous-section 4.1.2). Dans ce dernier cas, l'approximation $\hat{\mathcal{E}}$ est construite itérativement par émulations statistiques linéaires par morceaux successives.

Dans notre étude, nous n'étudions pas toutes ces solutions proposées mais une approche de résolution du problème d'optimisation probabiliste (4.1) qui repose sur :

— la réduction des scénarios,

— la construction d'un métamodèle $\hat{Q}$ de la fonction Q.

La réduction des scénarios est utilisée, le cas échéant, pour l'évaluation de l'espérance (première difficulté ci-dessus) et pour la construction du métamodèle $\hat{Q}$ (deuxième difficulté ci-dessus). Des méthodes de réduction des scénarios sont étudiées dans la section 4.3 de ce chapitre. Elles reposent notamment toutes sur une définition des similarités entre les scénarios.

Le métamodèle $\hat{Q}$ est construit pour que la fonction $\hat{\mathcal{E}} : b_0 \in \mathcal{B}_0 \mapsto \mathbb{E}[\hat{Q}(b_0, \hat{\Xi})]$ ait directement les propriétés mathématiques adaptées à l'optimisation, sans avoir besoin d'utiliser une approximation supplémentaire $\tilde{\mathcal{E}}$ de la fonction $\hat{\mathcal{E}}$. En outre, la méthode de construction du métamodèle $\hat{Q}$ présentée dans le chapitre 5 suivant consiste en un *apprentissage supervisé* [Friedman *et al.*, 2001], i.e. un calibrage automatique à partir d'un jeu de données composé d'un nombre fini de valeurs des données d'entrée et des valeurs des données de sortie associées. Dans notre cas, la fonction Q à métamodéliser est à valeurs réelles, donc on parle de *régression* (par opposition à la classification pour des variables de sortie quantitatives, i.e. représentant des catégories ou des valeurs discrètes). Les performances de l'apprentissage supervisé reposent notamment sur la faible dimension des données d'entrée.

Du point de vue de l'apprentissage supervisé, le problème d'optimisation (4.2) du second niveau est considéré comme une « boîte noire », i.e. un outil numérique, qui renvoie une donnée de

sortie à partir de données d'entrée, comme l'illustre la figure 4.1. La donnée de sortie est un scalaire qui représente la valeur optimale du second niveau donnée par la fonction Q. Les données d'entrée sont les arguments de la fonction Q, i.e. des prochaines ventes *day-ahead* $b_0 \in \mathbb{R}^{24}$ et un scénario $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$. En dissociant les apports en eau, les prix de l'électricité et les productions variables qui forment le scénario $\xi = (a, \pi, \phi)$, on a donc quatre données d'entrée. La représentation de la fonction Q par une boîte noire permet d'occulter le problème d'optimisation sous-jacent, que le métamodèle $\hat{Q}$ vise à remplacer.

La section 4.2 suivante présente la réduction de la dimension et la comparaison des données d'entrée, une étape préliminaire mais nécessaire à la réduction des scénarios (définition des similarités entre les scénarios) et la construction du métamodèle (amélioration des performances de l'apprentissage supervisé). La section 4.3 présente les résultats concernant la réduction des scénarios. La construction du métamodèle est présentée dans le chapitre 5 suivant.

4.2 Réduction de la dimension des données fonctionnelles

4.2.1 Problématique

Les quatre données d'entrée de la boîte noire de la figure 4.1 ont une structure fonctionnelle (cohérence temporelle ou spatio-temporelle), donc de grande dimension (voir par exemple la figure 3.5 qui illustre les scénarios probabilistes du cas d'étude). En revanche, on rappelle que, dans le cadre de la thèse, les apports en eau, les prix et les productions variables sont supposés indépendants entre eux (voir §2.4.4.2).

En pratique, ces données d'entrée de grande dimension sont délicates à comparer et à manipuler, notamment pour l'apprentissage du métamodèle. Il est donc important d'appliquer une réduction de la dimension des données d'entrée fonctionnelles, comme cela a été introduit dans plusieurs articles de la littérature [Amri *et al.*, 2020 ; Antoniadis *et al.*, 2012 ; Nanty *et al.*, 2016]. L'apprentissage du métamodèle est alors effectué sur les données d'entrée réduites, permettant *a priori* de limiter le temps de calcul (éviter le fléau de la dimension) et d'améliorer les performances de l'apprentissage (éviter notamment le sur-apprentissage) [Friedman *et al.*, 2001]. Cela est abordé dans le chapitre 5 suivant. La réduction de la dimension des données d'entrée fonctionnelles peut également servir pour faciliter la comparaison des données d'entrée. En effet, il est souvent plus facile de choisir une mesure de similarités sur les données d'entrée réduites plutôt que sur les données d'entrée fonctionnelles. Ces mesures de similarités peuvent servir, par exemple, pour la réduction des scénarios. Cela est abordé dans la section 4.3 suivante.

Excepté les prochaines ventes *day-ahead* qui sont définie uniquement sur une journée, les trois autres données d'entrée sont des valeurs discrétisées sur toute la fenêtre temporelle de l'optimisation pour la planification de la production. Comme l'instant présent et la taille de la fenêtre temporelle varient d'une planification à l'autre (par exemple, le nombre T de pas de temps n'est pas toujours égal à 360 comme pour le cas d'étude), la dimension des données d'entrée dépend de la planification considérée. Nous cherchons donc à définir une méthodologie de réduction de la dimension des données d'entrée qui ne dépend que des paramètres de la planification courante. En particulier, nous ne souhaitons pas utiliser des valeurs passées.

Dans la suite du manuscrit, en plus de supposer que les apports en eau a, les prix π et les productions variables ϕ qui composent le scénario $\xi = (a, \pi, \phi)$ sont indépendants entre eux (voir Sous-section 2.4.4), on suppose également qu'ils sont chacun indépendants des prochaines ventes *day-ahead* b_0. Cette hypothèse est acceptable dans une première approche car la fonction Q peut *a priori* être évaluée pour n'importe quelles valeurs $b_0 \in \mathbb{R}^{24}$ (indépendamment du scénario $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$), tant que les valeurs optimales du second niveau associées sont finies. L'avantage de cette hypothèse est que la réduction de la dimension peut être appliquée

séparément sur chacune des quatre données d'entrée fonctionnelles.

C'est pourquoi, dans la suite de cette section, les éléments théoriques de la réduction de la dimension sont présentés à partir d'une donnée d'entrée fonctionnelle générique, notée z, définie par ses valeurs en L points, où $L \in \mathbb{N}^*$ est la dimension de z. Ainsi, $z = (z(\ell))_{1 \leq \ell \leq L}$ est un vecteur de $\mathbb{R}^L$. Pour le cas d'étude,

— si la donnée d'entrée fonctionnelle z correspond aux prochaines ventes *day-ahead* b_0, alors elle est entièrement définie par les ventes pour les 24 heures du lendemain, donc $L = 24$,

— si la donnée d'entrée fonctionnelle z correspond aux apports en eau a, alors elle est entièrement définie par les débits des 6 affluents du Haut-Rhône aux 60 échéances horaires de la fenêtre temporelle, donc $L = 6 \times 60$,

— si la donnée d'entrée fonctionnelle z correspond aux prix *day-ahead* π, alors elle est entièrement définie par les prix horaires (les prix sont définis à la granularité horaire, voir §2.4.4.4) des deux derniers jours de la fenêtre temporelle, donc $L = 48$.

Pour le cas d'étude, nous utilisons les débits horaires des affluents car la prévision d'ensemble des apports en eau aux aménagements hydroélectriques s'obtient à partir de la prévision d'ensemble brute des débits horaires des affluents (voir Sous-section 2.5.3) en appliquant des opérations « déterministes » (voir Équation (2.4), les incertitudes des débits d'apports des bassins versants intermédiaires et des débits de correction n'étant pas prises en compte). Dans un cadre plus général où les scénarios font intervenir des débits d'apports en eau aux aménagements à la granularité demi-horaire (comme ceux de la prévision déterministe utilisée en opérationnel, voir §2.3.4.1), la dimension de la donnée d'entrée fonctionnelle des apports en eau serait égale à $|\mathcal{A}(\mathcal{I})| \times |\mathcal{T}_{1/2}|$, avec $\mathcal{A}(\mathcal{I}) = \bigcup_{i \in \mathcal{I}} \mathcal{A}(i) \subset \{1, \ldots, 21\}$ l'ensemble des indices des affluents de la planification considérée et $\mathcal{T}_{1/2} \subset \mathcal{T}$ l'ensemble des indices des pas de temps à la granularité demi-horaire.

4.2.2 Approches de réduction de la dimension

Nous présentons dans cette sous-section les deux approches utilisées dans la suite de ce manuscrit pour réduire la dimension de la donnée d'entrée fonctionnelle générique $z \in \mathbb{R}^L$: la représentation par un scalaire et la réduction par projection orthogonale sur les directions principales (analyse en composantes principales).

4.2.2.1 Représentation par un scalaire

Une réduction « brutale » est de représenter la donnée d'entrée fonctionnelle $z = (z(\ell))_{1 \leq \ell \leq L}$ par un scalaire. Pour cela, on utilise généralement la moyenne des composantes [Betancourt *et al.*, 2020], i.e.

$$\frac{1}{L} \sum_{\ell=1}^{L} z(\ell) \in \mathbb{R}.$$

Excepté pour les prix *day-ahead*, la moyenne de z a une interprétation physique. Pour les prochaines ventes *day-ahead* et les productions variables, la moyenne est liée aux puissances cumulées sur la fenêtre temporelle, i.e. des énergies. De même, pour les débits d'apports en eau, la moyenne est liée aux débits cumulés sur la fenêtre temporelle, i.e. des volumes d'eau.

Dans le cas des apports en eau, il serait intéressant de pondérer la moyenne selon les affluents pour tenir compte de leurs influences sur la production totale de la chaîne hydroélectrique (l'eau issue d'un affluent est turbinée à tous les aménagements hydroélectriques en aval, donc plus l'affluent est en amont, plus il a d'influence sur la production à débit équivalent).

4.2.2.2 Réduction par projection orthogonale

Les approches habituelles de réduction de la dimension consistent à remplacer la donnée d'entrée fonctionnelle $z \in \mathbb{R}^L$ par les coefficients $\tilde{z} = (\tilde{z}_1, \dots, \tilde{z}_K) \in \mathbb{R}^K$ de la décomposition de $z \in \mathbb{R}^L$ dans une famille orthonormale $(u_1, \dots, u_K)$ de K vecteurs de $\mathbb{R}^L$ ($K \in \{1, \dots, L\}$) :

$$z \approx \sum_{k=1}^{K} \tilde{z}_k u_k.$$

Bien sûr, si $K = L$, alors l'approximation de l'équation précédente est en fait une égalité (car $(u_1, \dots, u_L)$ est une base de $\mathbb{R}^L$) et, dans ce cas, aucune réduction de dimension n'est effectuée.

Le caractère orthonormal de la famille $(u_1, \dots, u_K)$ permet d'avoir l'unicité des coefficients $\tilde{z} = (\tilde{z}_1, \dots, \tilde{z}_K) \in \mathbb{R}^K$ (famille libre) et de les définir simplement par projection orthogonale :

$$\forall k \in \{1, \dots, K\}, \quad \tilde{z}_k = \langle z, u_k \rangle = z^\top u_k.$$

La réduction de la dimension revient donc à construire la famille orthonormale $(u_1, \dots, u_K)$. Plusieurs approches sont présentées dans la littérature (ondelettes, splines, analyse en composantes principales, décomposition orthogonale aux valeurs propres etc.) [Ramsay et Silverman, 2002 ; Torrence et Compo, 1998 ; De Boor, 1978 ; Nanty *et al.*, 2017 ; Jolliffe, 2005 ; Chatterjee, 2000]. Dans la suite de cette sous-section, nous décrivons l'approche utilisée pour notre étude : l'analyse en composantes principales.

4.2.2.3 Analyse en composantes principales

Une propriété souvent intéressante pour l'apprentissage d'un métamodèle ou la comparaison de données fonctionnelles est de définir la famille orthonormale $(u_1, \dots, u_K)$ de manière à décorréler les coefficients de $\tilde{z}$. Cette propriété peut être obtenue par analyse en composantes principales (ACP) à partir d'une matrice de données propre à la situation étudiée. Comme la donnée d'entrée z est fonctionnelle, on dépasse le cadre classique d'utilisation de l'ACP de la littérature. Dans la suite de cette sous-section, nous présentons brièvement la théorie de réduction par ACP pour la donnée d'entrée fonctionnelle z.

On suppose avoir à disposition une matrice de données, notée $Z = (z_{j\ell})_{1 \leq j \leq J, 1 \leq \ell \leq L} \in \mathcal{M}_{JL}(\mathbb{R}) \setminus \{0_{JL}\}$, regroupant J valeurs de la donnée d'entrée fonctionnelle z pour la situation étudiée ($J \in \mathbb{N}^*$). Pour les apports en eau, les prix *day-ahead* et les productions variables, la matrice de données est construite à partir de la prévision d'ensemble associée qui est utilisée pour la génération des scénarios probabilistes de la situation considérée (voir Sous-section 2.4.4). Dans ce cas, J correspond au nombre de membres (supposés équiprobables) de la prévision d'ensemble. Pour les prochaines ventes *day-ahead*, la matrice de données est obtenue en discrétisant l'ensemble $\mathcal{B}_0$ (c'est l'objet de la section 5.2 du chapitre suivant).

La matrice de covariance de la matrice de données Z est par définition la matrice carrée symétrique positive $\Sigma = \frac{1}{J-1} Z_0^\top Z_0 \in S_L^+(\mathbb{R})$, où $Z_0 = (z_{0j\ell})_{1 \leq j \leq J, 1 \leq \ell \leq L} \in \mathcal{M}_{JL}(\mathbb{R})$ est la matrice obtenue à partir de la matrice de données Z en centrant les colonnes, i.e.

$$\forall (j, \ell) \in \{1, \dots, J\} \times \{1, \dots, L\}, \quad z_{0j\ell} = z_{j\ell} - \overline{z}_\ell,$$

où $\overline{z} = (\overline{z}_\ell)_{1 \leq \ell \leq L} \in \mathbb{R}^L$ est le vecteur moyen de la matrice de données :

$$\forall \ell \in \{1, \dots, L\}, \quad \overline{z}_\ell = \frac{1}{J} \sum_{j=1}^{J} z_{j\ell} \in \mathbb{R}.$$

Dans le cadre classique de l'ACP, les colonnes devraient aussi être normalisées par les écarts-types pour que les variances de chaque colonne aient un poids équivalent et avec une unité comparable (on obtient alors ce qu'on appelle parfois les *z-scores*). Dans le cadre de notre travail, les colonnes sont centrées mais elles ne sont pas normalisées par les écarts-types afin de conserver les variations de la donnée fonctionnelle : les différences de variance entre les colonnes ont une importance.

Comme la matrice carrée réelle $Z_0^\top Z_0 \in S_L^+(\mathbb{R})$ est symétrique positive, elle est diagonalisable dans une base orthonormale de vecteurs propres associés à des valeurs propres positives ou nulles [Gourdon, 2008a], i.e.

$$Z_0^\top Z_0 = U \begin{pmatrix} \lambda_1^2 & & (0) \\ & \ddots & \\ (0) & & \lambda_L^2 \end{pmatrix} U^\top,$$

où $(\lambda_1^2, \ldots, \lambda_L^2) \in (\mathbb{R}_+)^L$ sont les valeurs propres et $U = (u_1 \cdots u_L) \in \mathcal{O}_L(\mathbb{R})$ la matrice des vecteurs propres orthogonaux $u_k \in \mathbb{R}^L$, pour $k \in \{1, \ldots, L\}$, appelés les *directions principales* (ou *axes principaux*). Quitte à intervertir l'ordre des vecteurs propres, on suppose que les valeurs propres sont rangées dans l'ordre décroissant : $\lambda_L^2 \leq \cdots \leq \lambda_1^2$.

À ce stade, la donnée d'entrée fonctionnelle $z \in \mathbb{R}^L$ peut être exprimée dans la base orthonormale $(u_1, \ldots, u_K)$ mais sans réduction de la dimension :

$$z = \sum_{k=1}^{L} \langle z, u_k \rangle u_k = \overline{z} + \sum_{k=1}^{L} \langle z - \overline{z}, u_k \rangle u_k.$$

Les coefficients $\langle z, u_1 \rangle$, ..., $\langle z, u_L \rangle$ de z sont décorrélés car les directions principales forment une base de $\mathbb{R}^L$ dans laquelle la matrice de covariance est diagonale.

Pour réduire la dimension, il suffit de sélectionner un nombre $K \in \{1, \ldots, L\}$ de directions principales qui permettent d'expliquer une partie suffisante de la variance totale. Pour cela, on remarque que la proportion de variance expliquée d'une direction principale $u_k \in \mathbb{R}^L$ par rapport à la variance totale, pour $k \in \{1, \ldots, L\}$, est liée à la valeur propre λ_k^2 :

$$\frac{\lambda_k^2}{\lambda_1^2 + \cdots + \lambda_L^2} \in [0, 1]. \tag{4.4}$$

Ainsi, comme les valeurs propres λ_k^2, pour $k \in \{1, \ldots, L\}$, sont rangées dans l'ordre décroissant, les directions principales sont rangées dans l'ordre décroissant de leur variance expliquée. La réduction de la dimension revient alors à ne sélectionner que les K premières directions principales, i.e.

$$z \approx \overline{z} + \sum_{k=1}^{K} \langle z - \overline{z}, u_k \rangle u_k. \tag{4.5}$$

Le choix du nombre $K \in \{1, \ldots, L\}$ de directions principales à sélectionner est généralement effectué à partir d'un seuil $v_{\min} \in]0, 1[$ sur la proportion de la variance expliquée par les directions principales sélectionnées par rapport à la variance totale : K est le plus petit nombre de directions principales qui permet d'atteindre le seuil, i.e.

$$\begin{cases} \dfrac{\lambda_1^2 + \cdots + \lambda_K^2}{\lambda_1^2 + \cdots + \lambda_L^2} \geq v_{\min}, \\ \dfrac{\lambda_1^2 + \cdots + \lambda_{K-1}^2}{\lambda_1^2 + \cdots + \lambda_L^2} < v_{\min}. \end{cases}$$

Comme $v_{\min} > 0$, on a : $\forall k \in \{1, \ldots, K\},\ \lambda_k^2 > 0$.

4.2.3 Comparaison des données fonctionnelles

Pour comparer les valeurs de la donnée d'entrée fonctionnelle générique z, il suffit de définir une distance sur $\mathbb{R}^L$, ou plus simplement une mesure de similarités sur $\mathbb{R}^L$. Dans cette sous-section, nous commençons par présenter les distances que nous utilisons sur les données d'entrée fonctionnelles non réduites, puis nous présentons une mesure de similarités basée sur la réduction par analyse en composantes principales.

4.2.3.1 Distance sur les données non réduites

Comme des puissances cumulées sur la fenêtre temporelle correspondent à des énergies, il semble pertinent de comparer les prochaines ventes *day-ahead* et les productions variables grâce à la distance d_1 associée à la norme L^1 dans $\mathbb{R}^L$, i.e. celle définie par : pour tous $z_1 = (z_1(\ell))_{1\leq\ell\leq L} \in \mathbb{R}^L$ et $z_2 = (z_2(\ell))_{1\leq\ell\leq L} \in \mathbb{R}^L$,

$$\mathrm{d}_1(z_1, z_2) = \sum_{\ell=1}^{L} |z_1(\ell) - z_2(\ell)| \in \mathbb{R}_+.$$

Pour comparer les prix, on utilise la distance euclidienne d_2 sur $\mathbb{R}^L$ définie par : pour tous $z_1 = (z_1(\ell))_{1\leq\ell\leq L} \in \mathbb{R}^L$ et $z_2 = (z_2(\ell))_{1\leq\ell\leq L} \in \mathbb{R}^L$,

$$\mathrm{d}_2(z_1, z_2) = \sqrt{\sum_{\ell=1}^{L} (z_1(\ell) - z_2(\ell))^2} \in \mathbb{R}_+.$$

L'intérêt de la distance euclidienne est qu'elle donne plus d'importance aux composantes où les différences sont les plus importantes et moins d'importance aux composantes où les différences sont les plus faibles. Or, les prix sont utilisés dans la fonction objectif du problème d'optimisation : des courbes de prix ayant des moyennes similaires peuvent fournir des valeurs optimales très différentes. La distance d_2 semble donc plus adaptée que la distance d_1 pour comparer les prix.

Pour les débits d'apports en eau, la tâche est plus complexe en raison de la cohérence spatiale. Comme des débits cumulés correspondent à des volumes, il semble intéressant de considérer la distance euclidienne des différences des débits cumulés de chaque affluent sur la fenêtre temporelle, i.e. la distance d_{2-1} sur $\mathbb{R}^L$ ($L = 6 \times 60$) définie par : pour tous $z_1 = (z_1(i,h))_{\substack{1\leq i\leq 6\\1\leq h\leq 60}} \in \mathbb{R}^L$ et $z_2 = (z_2(i,h))_{\substack{1\leq i\leq 6\\1\leq h\leq 60}} \in \mathbb{R}^L$,

$$\mathrm{d}_{2-1}(z_1, z_2) = \sqrt{\sum_{i=1}^{6} \left(\sum_{h=1}^{60} |z_1(i,h) - z_2(i,h)| \right)^2} \in \mathbb{R}_+,$$

(il s'agit bien d'une distance sur $\mathbb{R}^L$, par combinaison de la distance associée à la norme L^1 dans $\mathbb{R}^{60}$ et de la distance euclidienne dans $\mathbb{R}^6$). Là encore, il serait intéressant d'ajouter des poids aux affluents dans la définition de la distance d_{2-1} pour tenir compte de leurs influences sur la production totale de la chaîne hydroélectrique.

4.2.3.2 Distance de Mahalanobis

Une autre façon de comparer des valeurs de la donnée d'entrée fonctionnelle z est de considérer sa réduction par analyse en composantes principales grâce à l'application d_M définie par :

$$\forall (z_1, z_2) \in (\mathbb{R}^L)^2, \quad \mathrm{d}_M(z_1, z_2) = \sqrt{\sum_{k=1}^{K} \frac{\langle z_1 - z_2, u_k \rangle^2}{\lambda_k^2}} \in \mathbb{R}_+.$$

Si $K = L$ (et donc $\forall k \in \{1, \dots, L\},\ \lambda_k \neq 0$), alors l'application d_M ainsi définie est une distance sur $\mathbb{R}^L$, appelée la *distance de Mahalanobis* [De Maesschalck *et al.*, 2000], car on a la relation

$$\forall (z_1, z_2) \in (\mathbb{R}^L)^2, \quad \mathrm{d}_M(z_1, z_2) = \|\phi(z_1 - z_2)\|_2,$$

où $\|\cdot\|_2$ est la norme euclidienne dans $\mathbb{R}^L$ et $\phi \in \mathrm{GL}(\mathbb{R}^L)$ l'isomorphisme sur $\mathbb{R}^L$ défini par :

$$\forall z \in \mathbb{R}^L, \quad \phi(z) = \begin{pmatrix} \langle z, u_1 \rangle / \lambda_1 \\ \vdots \\ \langle z, u_L \rangle / \lambda_L \end{pmatrix} \in \mathbb{R}^L,$$

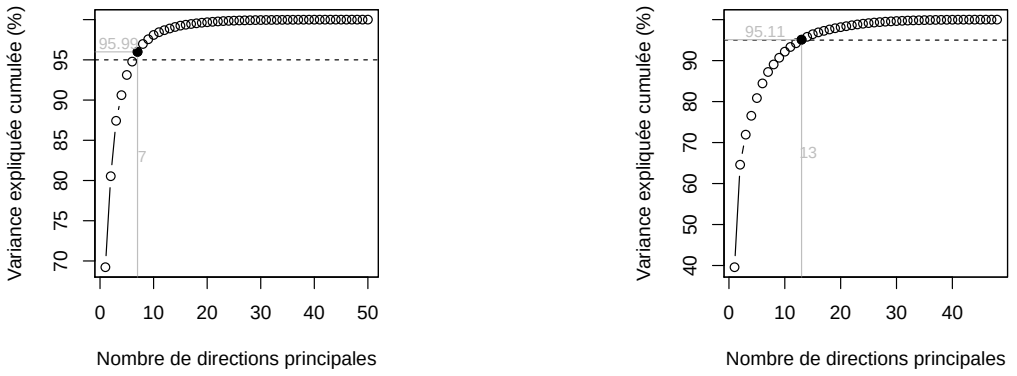
(la linéarité vient de la linéarité à gauche du produit scalaire et la bijectivité vient du fait que $(u_1, \dots, u_L)$ est une base orthonormale de $\mathbb{R}^L$). Autrement dit, la distance de Mahalanobis est la distance euclidienne des composantes principales (i.e. des coefficients de la décomposition dans la base orthonormale des directions principales) et pondérées par les inverses des variances expliquées.

En revanche, si $K < L$, alors l'application d_M n'est pas une distance sur $\mathbb{R}^L$ car elle ne vérifie pas la propriété de séparation (l'endomorphisme ϕ, tronquée aux K premières composantes, n'est pas un isomorphisme). L'application d_M peut s'annuler pour deux valeurs différentes $z_1 \in \mathbb{R}^L$ et $z_2 \in \mathbb{R}^L$, $z_1 \neq z_2$, dont la différence de variance expliquée est inférieure à $1 - v_{\min}$. Néanmoins, si le seuil $v_{\min} \in [0, 1]$ est suffisamment proche de 1 (en pratique $v_{\min} = 0{,}90$ ou $v_{\min} = 0{,}95$), alors les différences de variance expliquée inférieures à $1 - v_{\min}$ peuvent être considérées comme négligeables. Le seuil $v_{\min} \in [0, 1]$ pourrait, par exemple, être défini par rapport aux erreurs de mesure connues de la donnée fonctionnelle z. Dans ce cas, les différences de variance expliquée inférieures à $1 - v_{\min}$ sont comprises dans l'intervalle de confiance associé aux erreurs de mesure, donc elles sont physiquement négligeables. L'application d_M est alors suffisante pour mesurer les similarités entre des valeurs de la donnée d'entrée fonctionnelle. Par abus de langage, nous continuons à dire que d_M est la distance de Mahalanobis, même s'il ne s'agit pas d'une distance mais d'une mesure de similarités.

4.2.4 Application aux scénarios du cas d'étude

Pour la réduction des apports en eau et des prix *day-ahead* du cas d'étude, les courbes des variances expliquées en fonction du nombre K de directions principales retenues sont présentées à la figure 4.2. On remarque que la réduction des apports en eau est plus efficace que celle des prix *day-ahead* (les variances expliquées cumulées augmentent plus rapidement). Cela peut s'expliquer par le fait que la cohérence spatio-temporelle est une caractéristique importante des apports en eau (les valeurs sont donc très corrélées), rendant la réduction efficace. À l'inverse, la cohérence temporelle est une caractéristique moins importante pour les prix (malgré leur structure journalière, les prix sont plus « chaotiques » que des grandeurs physiques continues comme les débits), rendant la réduction moins efficace que celle des apports en eau.

Pour le cas d'étude, la réduction est effectuée avec un seuil de la variance expliquée fixé arbitrairement à $v_{\min} = 0{,}95$ pour chaque donnée d'entrée. On utilise alors $K = 7$ directions principales pour les apports en eau (expliquant 95,99% de la variance totale) et $K = 13$ directions principales pour les prix *day-ahead* (expliquant 95,11% de la variance totale), voir Figure 4.2. Les allures des directions principales retenues pour les apports en eau et les prix *day-ahead* sont représentées à la figure 4.3. Pour les prix *day-ahead*, les allures des directions principales, qui illustrent la cohérence temporelle, sont assez « chaotiques » et peu différentes, rendant les interprétations difficiles (il n'y a pas de forme particulière des incertitudes qui se dessine au cours du temps). Pour les apports en eau, les allures des directions principales, les allures des directions principales sont plus faciles à interpréter car elles illustrent la cohérence spatio-temporelle : les



(a) Apports en eau

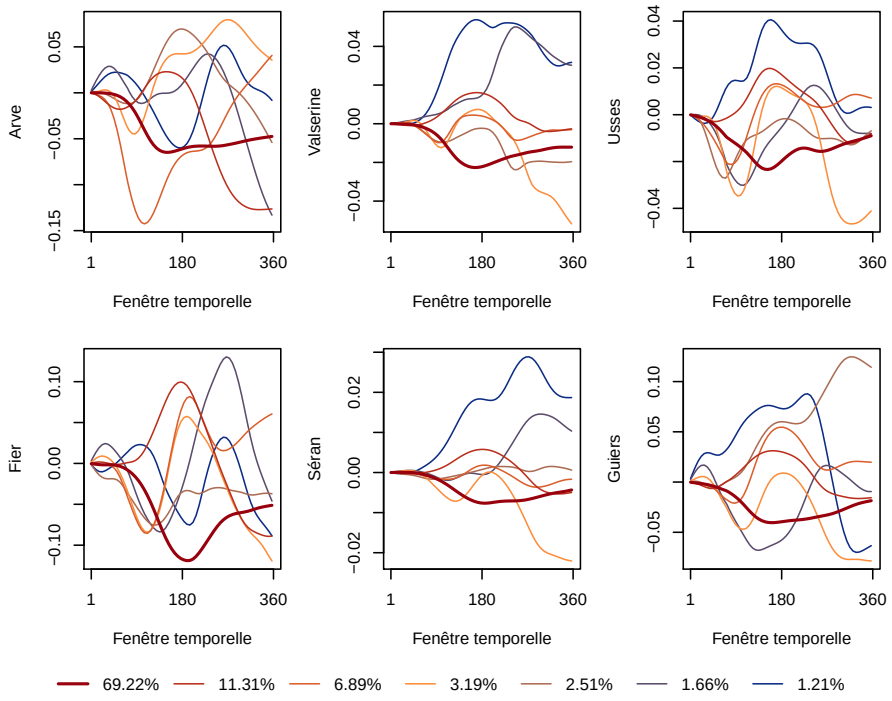
(b) Prix *day-ahead*

FIGURE 4.2 – Variance expliquée en fonction du nombre de directions principales pour le cas d'étude.

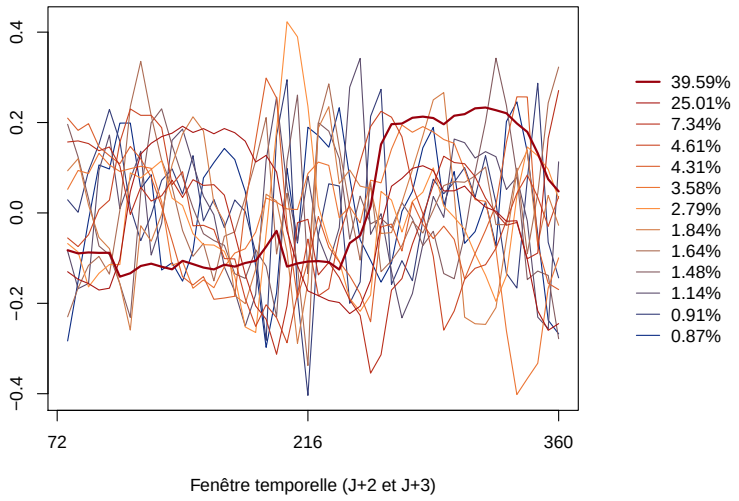
« pics » ou « creux » des directions principales correspondent aux affluents et aux pas de temps où les incertitudes sont les plus élevées. Par exemple, l'allure de la première direction principale (celle ayant la variance expliquée la plus élevée) montre qu'une variation (augmentation ou diminution) du débit de l'Arve vers le pas de temps d'indice $t = 120$ se propage sur les autres affluents avec un décalage temporel correspondant au temps de propagation de l'eau.

Comme $K < L$, la réduction ne permet pas de reconstituer parfaitement les données d'entrée fonctionnelles initiales. Les erreurs dues à la réduction pour le cas d'étude sont représentées à la figure 4.4. On observe que la réduction a tendance à atténuer les membres extrêmes de la prévision d'ensemble, surtout pour les apports, mais les erreurs semblent acceptables visuellement. Pour aller plus loin, une étude des corrélations croisées aurait permis de mesurer les erreurs concernant le respect des cohérences temporelles et spatio-temporelles.

Les données d'entrée fonctionnelles réduites du cas d'étude sont utilisées dans le chapitre 5 qui suit pour illustrer et tester les étapes de la construction du métamodèle. Elles servent également à définir les distances de Mahalanobis pour le cas d'étude. Les figures 4.5 et 4.6 illustrent respectivement les distances sur les données non réduites et les distances de Mahalanobis pour les apports en eau et les prix *day-ahead* du cas d'étude. Il est difficile de comparer quantitativement ces deux distances à ce stade car le choix de la distance dépend de l'utilisation. Ces deux distances sont utilisées et comparées dans la suite du manuscrit, notamment dans la section 4.3 qui suit, consacrée à la réduction des scénarios.

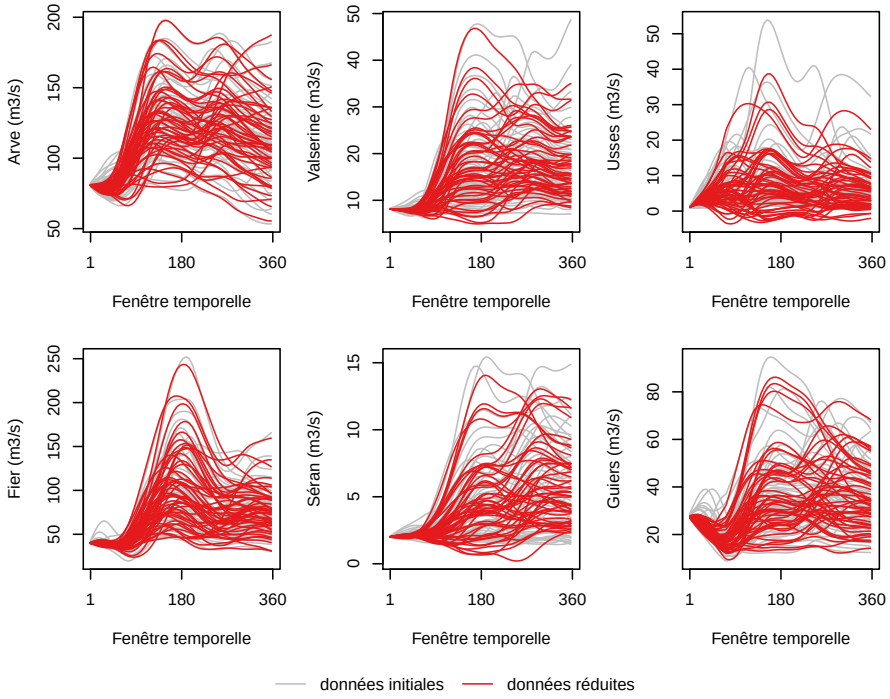


(a) Apports en eau

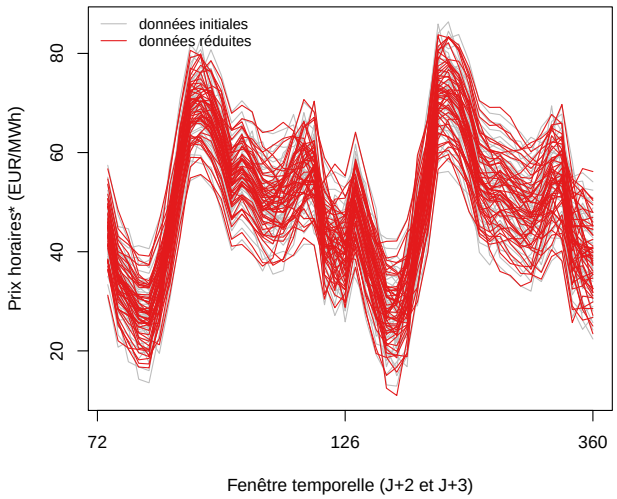


(b) Prix *day-ahead*

FIGURE 4.3 – Directions principales retenues pour le cas d'étude. Les pourcentages sont les variances expliquées associées aux directions principales. Les courbes épaisses en rouge correspondent à la première direction principale.

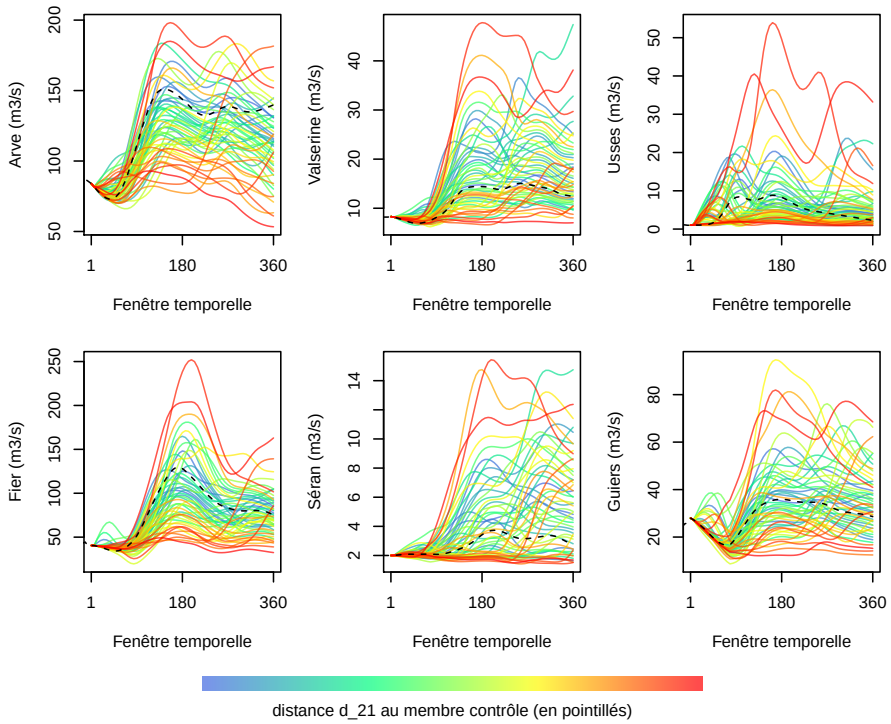


(a) Apports en eau

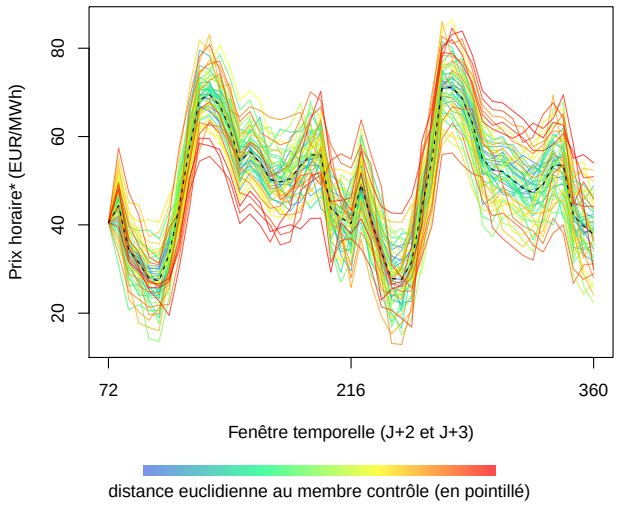


(b) Prix *day-ahead*

FIGURE 4.4 – Reconstitution des données d'entrée après réduction pour le cas d'étude. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les prix.

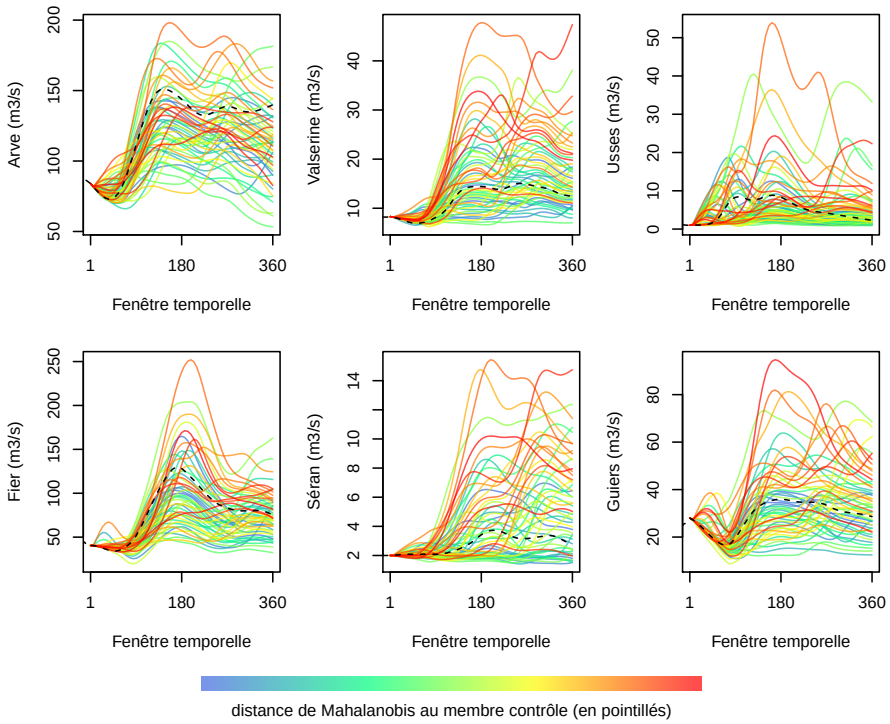


(a) Apports en eau des affluents principaux

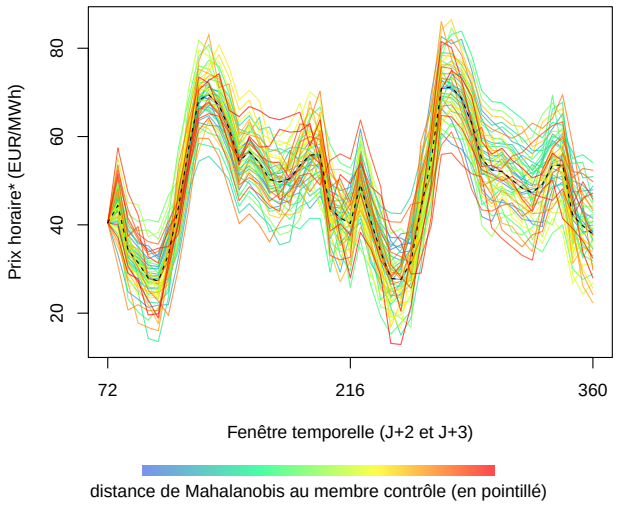


(b) Prix *day-ahead*

FIGURE 4.5 – Courbes des apports en eau et des prix *day-ahead* du cas d'étude, mises en couleur en fonction de la valeur de leurs distances calculées sur les données non réduites avec les courbes du scénario moyen (trait noir en pointillés). Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les prix.



(a) Apports en eau des affluents principaux



(b) Prix *day-ahead*

FIGURE 4.6 – Courbes des apports en eau et des prix *day-ahead* du cas d'étude, mises en couleur en fonction de la valeur de leurs distances de Mahalanobis (sur les données réduites) avec les courbes du scénario moyen (trait noir en pointillés). Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les prix.

4.3 Réduction des scénarios

La méthodologie proposée à la sous-section 4.1.3 pour résoudre le problème d'optimisation probabiliste (4.1) repose sur la réduction des scénarios. L'objectif est de ne considérer qu'un nombre restreint de scénarios probabilistes et de définir les probabilités des scénarios considérés de manière à approcher la loi de probabilité des scénarios initiaux. Cette réduction des scénarios peut être utilisée dans le chapitre 5 qui suit pour faciliter l'évaluation des espérances de la fonction Q et faciliter la construction du jeu de données nécessaire à l'apprentissage du métamodèle $\hat{Q}$. Pour ces deux utilisations, la réduction des scénarios ne doit pas trop déformer les lois de probabilité des scénarios initiaux. Nous présentons dans cette section plusieurs approches possibles de réduction des scénarios qui reposent toutes sur la définition des similarités (distance) entre les scénarios. Ces approches sont testées et comparées sur le cas d'étude.

4.3.1 Problématique

Mathématiquement, la réduction des scénarios consiste à remplacer la variable aléatoire discrète et finie $\hat{\Xi}$ des scénarios probabilistes (définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans $\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$) par une variable aléatoire discrète et finie $\hat{\Xi}'$ sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ et ayant un plus petit support que $\hat{\Xi}$ (i.e. $N' = |\{\xi' \in \hat{\Xi}'(\Omega),\ \mathbb{P}[\hat{\Xi}' = \xi'] > 0\}| \leq N$, idéalement $N' \ll N$). Dans cette section, on dit que $\hat{\Xi}'$ est la variable aléatoire réduite obtenue à partir de la variable aléatoire initiale $\hat{\Xi}$. L'objectif est que $\hat{\Xi}'$ soit une approximation de $\hat{\Xi}$, notamment par rapport à leur loi de probabilité.

En particulier, la variable aléatoire discrète et finie $\hat{\Xi}'$ permet d'approcher les espérances :

$$\forall b_0 \in \mathcal{B}_0, \quad \mathbb{E}[Q(b_0, \hat{\Xi}')] = \sum_{k=1}^{N'} p'_k Q(b_0, \xi'_k) \approx \sum_{n=1}^{N} p_n Q(b_0, \xi_n) = \mathbb{E}[Q(b_0, \hat{\Xi})], \tag{4.6}$$

où $\{\xi'_k,\ 1 \leq k \leq N'\} = \hat{\Xi}'(\Omega)$ et $p'_k = \mathbb{P}[\hat{\Xi}' = \xi'_k] \in]0, 1[$, pour $k \in \{1, \ldots, N'\}$.

Ainsi, l'utilisation de la variable aléatoire finie $\hat{\Xi}'$ au lieu de $\hat{\Xi}$ permet, en particulier, de réduire le nombre d'évaluations de la fonction coûteuse Q pour l'estimation de l'espérance (N' au lieu de N). Le problème d'optimisation probabiliste (4.1) est alors approché par le problème d'optimisation

$$\max_{\mathbf{x}_0 \in \mathcal{X}(\xi_0)} \left\{ \alpha_0 f(\mathbf{x}_0, \xi_0) + (1 - \alpha_0) \mathbb{E}[Q(b(\mathbf{x}_0), \hat{\Xi}')] \right\}. \tag{4.7}$$

De manière analogue à la sous-section 4.1.2, le problème d'optimisation (4.7) peut se résoudre numériquement à partir de sa formulation MILP équivalente

$$\max_{\substack{\mathbf{x}_0 \in \mathcal{X}(\xi_0) \\ \mathbf{x}_k \in \mathcal{X}(\xi'_k),\ 1 \leq k \leq N' \\ b(\mathbf{x}_k) = b(\mathbf{x}_0),\ 1 \leq k \leq N'}} \left\{ \alpha_0 f(\mathbf{x}_0, \xi_0) + (1 - \alpha_0) \sum_{k=1}^{N'} p'_k f(\mathbf{x}_k, \xi'_k) \right\}, \tag{4.8}$$

similaire au problème MILP (4.3) mais avec une dimension beaucoup plus petite (car $N' \ll N$), donc *a priori* plus facile à résoudre.

Au-delà de son utilisation pour l'évaluation des espérances, la réduction des scénarios pourrait également être utilisée pour former le jeu de données nécessaire à l'apprentissage du métamodèle $\hat{Q}$ (voir Section 5.3 du chapitre suivant).

Il y a deux possibilités pour le choix de l'image de la variable aléatoire finie $\hat{\Xi}'$:

1. imposer $\hat{\Xi}'(\Omega) \subset \hat{\Xi}(\Omega)$, i.e. construire la variable aléatoire $\hat{\Xi}'$ par quantification de la variable aléatoire $\hat{\Xi}$ de manière à considérer un petit nombre de scénarios parmi les scénarios probabilistes déjà existants mais en laissant la possibilité de redistribuer les probabilités des scénarios conservés,
2. ne pas imposer $\hat{\Xi}'(\Omega) \subset \hat{\Xi}(\Omega)$, i.e. donner la possibilité de construire des nouveaux scénarios (et des nouvelles probabilités) qui peuvent être différents des scénarios probabilistes déjà existants.

Néanmoins, la seconde possibilité ci-dessus semble donc plus difficile à mettre en œuvre que la première en raison des corrélations temporelles ou spatio-temporelles des apports en eau, des prix *day-ahead* et des production variables qui composent les scénarios (voir Sous-section 2.4.4). C'est pourquoi, pour notre étude, nous nous intéressons uniquement à la première possibilité.

On rappelle néanmoins que les apports en eau, les prix et les productions variables qui composent les scénarios sont supposés indépendants (voir Sous-section 2.4.4). De manière analogue à la réduction de la dimension des données fonctionnelles (voir Sous-section 4.2.1), les approches de réductions des scénarios peuvent donc être appliquées séparément sur les apports en eau, les prix et les productions variables. La réduction des scénarios pour la variable aléatoire $\hat{\Xi}$ s'en déduit en considérant le triplet des variables aléatoires réduites indépendantes des apports en eau, des prix et des productions variables. Cette simplification est discutée à la sous-section 6.3.2.

Dans la suite de cette section, on reprend donc la notation générique $z = (z(\ell))_{1 \leq \ell \leq L} \in \mathbb{R}^L$ ($L \in \mathbb{N}^*$) de la sous-section 4.2.1, qui peut désigner les apports en eau, les prix ou les productions variables (mais pas les prochaines ventes *day-ahead* dans cette section). Les incertitudes sur la donnée fonctionnelle générique z sont modélisées par la prévision d'ensemble associée, notée $(z_1, \ldots, z_J) \in (\mathbb{R}^L)^J$ ($J \in \mathbb{N}^*$), comportant J membres supposés équiprobables (voir Sous-section 2.4.4). On considère alors la variable aléatoire discrète et finie $\hat{Z}$ sur $(\Omega, \mathcal{F}, \mathbb{P})$ associée à la prévision d'ensemble (loi uniforme sur $\{z_1, \ldots, z_J\}$) :

— $\hat{Z}(\Omega) = \{z_1, \ldots, z_J\}$,
— $\forall j \in \{1, \ldots, J\},\ \mathbb{P}[\hat{Z} = z_j] = 1/J$.

La réduction des scénarios est alors appliquée sur la variable aléatoire discrète et finie $\hat{Z}$ (et non pas $\hat{\Xi}$). Dans la suite de cette section, il faut donc comprendre le terme « scénario » comme un élément de $\hat{Z}(\Omega) = \{z_1, \ldots, z_J\}$ (et non pas comme un élément de $\hat{\Xi}(\Omega) = \{\xi_1, \ldots, \xi_N\}$). Nous faisons cette confusion pour reprendre le vocabulaire de la littérature en lien avec la réduction des scénarios. Par ailleurs, les scénarios sélectionnés par la réduction des scénarios sont appelés les *scénarios conservés*.

4.3.2 Algorithmes testés

Dans notre étude, la réduction des scénarios est effectuée à l'aide d'un des cinq algorithmes suivants issus de la littérature :

— (SBR) *simultaneous backward reduction* [Gröwe-Kuska *et al.*, 2003],
— (FFS) *fast forward selection* [Gröwe-Kuska *et al.*, 2003],
— (RS) *random search* [Armstrong *et al.*, 2013],
— (GM) *greedy maximin* [El Amri *et al.*, 2020],
— (WC) *Ward's clustering* [Murtagh et Legendre, 2014].

Ces algorithmes reposent sur deux étapes :

1. le choix des scénarios à conserver,
2. le calcul des probabilités des scénarios conservés.

Nous considérons dans notre étude que la seconde étape est effectuée par redistribution optimale de Kantorovich (voir Annexe A.3.3) : la loi de probabilité des scénarios conservés est celle qui est la plus proche de loi de probabilité des scénarios initiaux selon la distance de Kantorovich (distance sur l'ensemble des probabilités sur $(\Omega, \mathcal{F})$ définie à partir d'une distance sur les scénarios, voir Annexe A.3.2). Les algorithmes se différencient donc par la première étape.

Les deux premiers algorithmes (SBR et FFS), sont très populaires dans la littérature dans des contextes analogues à celui de ce manuscrit [Bruninx et Delarue, 2016 ; Larsen *et al.*, 2015]. Il s'agit d'algorithmes qui combinent les deux étapes précédentes de manière itérative. L'algorithme SBR (respectivement FFS) retire (respectivement sélectionne) un à un les scénarios qui permettent de minimiser les erreurs de la redistribution de Kantorovich des scénarios conservés à chaque itération. Pour les aspects théoriques de ces algorithmes, voir Annexe A.3.

L'algorithme RS consiste à tirer aléatoirement plusieurs ensembles de scénarios à conserver (recherche aléatoire) et applique la redistribution optimale de Kantorovich avec chacun d'eux. L'algorithme renvoie l'ensemble des scénarios conservés et leurs probabilités redistribuées qui donnent les plus faibles erreurs lors de la redistribution de Kantorovich.

Les algorithmes GM et WC permettent de regrouper les scénarios initiaux en différents *clusters*, par quantification (discrète) pour l'algorithme GM (dont le lien avec les algorithmes basés sur la distance de Kantorovich est décrit à l'annexe A.3.4) ou par *clustering* hiérarchique pour l'algorithme WC. Dans les deux cas, les *clusters* sont calculés à partir d'une distance (ou une mesure de similarités) sur les scénarios. Pour choisir un scénario représentatif par *cluster*, on applique une itération de l'algorithme FFS à chaque cluster (les scénarios représentatifs sont donc ceux qui expliquent le mieux chaque *cluster*). La redistribution optimale de Kantorovich est ensuite appliquée pour définir les probabilités des scénarios représentatifs.

En raison notamment de l'utilisation de la distance de Kantorovich, tous ces algorithmes reposent sur une distance sur $\mathbb{R}^L$ qui permet de comparer deux éléments de la donnée fonctionnelle z. Pour cela, on utilise les deux distances définies de la sous-section 4.2.3 pour la donnée fonctionnelle z : celle basée sur les données de dimension non réduite L et celle de Mahalanobis basée sur les données de dimension réduite $K < L$ par analyse en composantes principales.

Appliquée à la variable aléatoire discrète et finie $\hat{Z}$, la réduction des scénarios avec un des ces algorithmes permet d'affecter la probabilité nulle aux scénarios supprimés et de nouvelles probabilités aux scénarios conservés. Autrement dit, cela permet de construire une variable aléatoire discrète et finie $\hat{Z}'$ sur $(\Omega, \mathcal{F}, \mathbb{P})$ telle que :

— $\hat{Z}'(\Omega) = \{z_1, \ldots, z_J\}$,

— pour tout $j \in \{1, \ldots, J\}$, $\mathbb{P}[\hat{Z}' = z_j] \in [0, 1]$ désigne la probabilité réduite du membre z_j, nulle si et seulement s'il n'est pas conservé par la réduction.

L'image de la variable aléatoire réduite $\hat{Z}'$ est identique à celle de la variable aléatoire initiale $\hat{Z}$ mais sa loi de probabilité est différente avec un support non vide (la taille du support correspond au nombre de scénarios conservés).

4.3.3 Critères de qualité de la réduction

Nous définissons dans cette sous-section des critères de qualité pour évaluer les performances des algorithmes de réduction de scénarios de la sous-section précédente appliqués à la variable aléatoire discrète et finie $\hat{Z}$. L'objectif est double :

— choisir le meilleur compromis entre le nombre de scénarios conservés et la qualité de la réduction,

— choisir l'algorithme le plus performant parmi les cinq de la sous-section 4.3.2 précédente.

Quatre critères de qualité sont définis dans cette sous-section : la précision relative, la moyenne des variances expliquées, la fiabilité et finesse, et la stabilité. Ils mesurent la qualité de la réduction des scénarios pour une situation donnée. Comme les performances des algorithmes dépendent *a priori* de la situation, il peut être nécessaire d'évaluer ces critères de qualité sur un historique de vérification composé de plusieurs situations. Cela permet, par exemple, de choisir un seul algorithme de réduction des scénarios à appliquer à chaque nouvelle situation. Néanmoins, si le temps de calcul le permettait, alors il serait aussi possible de tester les performances des cinq algorithmes de la sous-section 4.3.2 à chaque nouvelle situation et de choisir le plus performant pour cette situation.

4.3.3.1 Précision relative

La précision relative, inspirée de [Dupačová *et al.*, 2000], permet de comparer la loi de probabilité de la variable aléatoire réduite $\hat{Z}'$ avec celle de la variable aléatoire initiale $\hat{Z}$. On calcule pour cela leur distance de Kantorovich car c'est celle qui est utilisée par la redistribution optimale des scénarios conservés.

Comme les valeurs de la distance de Kantorovich sont difficilement interprétables, elles sont comparées à une valeur de référence qui correspond au cas où la variable aléatoire $\hat{Z}$ est réduite à un élément (cadre déterministe). On considère que cet élément est le scénario sélectionné par la première itération de l'algorithme FFS, noté z_{j^*} avec $j^* \in \{1, \dots, J\}$, dont la loi de probabilité associée est la mesure de Dirac $\delta_{z_{j^*}}$ définie sur $(\Omega, \mathcal{F})$ et centrée sur z_{j^*}. Parmi tous les choix d'un unique scénario, le scénario z_{j^*} est celui dont la mesure de Dirac associée est la plus proche de la loi de probabilité initiale (au sens de la distance de Kantorovich).

La précision relative est alors mesurée par le réel

$$\eta = 1 - \frac{\mathrm{d}_K(\mathbb{P}_{\hat{Z}}, \mathbb{P}_{\hat{Z}'})}{\mathrm{d}_K(\mathbb{P}_{\hat{Z}}, \delta_{z_{j^*}})} \in \left]-\infty, 1\right],$$

où $\mathbb{P}_{\hat{Z}}$ et $\mathbb{P}_{\hat{Z}'}$ sont les lois des variables aléatoires $\hat{Z}$ et $\hat{Z}'$, et d_K désigne la distance de Kantorovich associée à la distance sur les données de dimension non réduite de la sous-section 4.2.3.

Plus la précision relative est proche de 1, plus la probabilité réduite est proche de la probabilité initiale, donc plus la réduction des scénarios est précise.

4.3.3.2 Moyenne des variances expliquées

Au lieu de s'intéresser à la comparaison des lois des variables aléatoires, on peut seulement s'intéresser à la comparaison de leurs variances. C'est en effet la variance qui représente le plus l'effet des incertitudes modélisées par les variables aléatoires $\hat{Z}$ et $\hat{Z}'$. La variance aléatoire $\hat{Z}$ correspond aux scénarios initiaux, donc à la variance totale que l'on peut représenter par réduction des scénarios. La comparaison des variances peut alors s'effectuer en calculant la variance expliquée de la variable aléatoire $\hat{Z}'$, i.e. la variance par rapport à la variance totale, de manière analogue à l'analyse en composantes principales (voir §4.2.2.3).

Comme les variables aléatoires $\hat{Z}$ et $\hat{Z}'$ sont multivariées (cohérence temporelle ou spatio-temporelle), on calcule la moyenne des variances expliquées de chaque composante de la donnée fonctionnelle z, i.e.

$$\nu = \frac{1}{L} \sum_{\ell=1}^{L} \frac{\mathrm{Var}[\hat{Z}'_\ell]}{\mathrm{Var}[\hat{Z}_\ell]} \in \mathbb{R}_+,$$

où $\hat{Z}_\ell$ et $\hat{Z}'_\ell$, pour $\ell \in \{1, \dots, L\}$, sont les variables aléatoires marginales des variables aléatoires $\hat{Z}$ et $\hat{Z}'$, et Var désigne la variance pour des variables aléatoires réelles sur $(\Omega, \mathcal{F}, \mathbb{P})$.

Plus la variance expliquée est proche de 1, plus la dispersion des scénarios conservés est proche de celle des scénarios initiaux, donc plus la réduction des scénarios est précise (au sens des variances expliquées).

4.3.3.3 Fiabilité et finesse

La réduction des scénarios déforme les lois de probabilité, donc elle a un impact sur la qualité des prévisions d'ensemble représentées par les scénarios conservés. Or, la qualité de la modélisation des incertitudes est un point crucial en optimisation en univers probabiliste (voir §1.4.5.2).

On propose donc d'analyser les performances de la réduction des scénarios avec un critère associé à la fiabilité et la finesse de la prévision d'ensemble représentée par la variable aléatoire discrète et finie $\hat{Z}'$. On utilise pour cela le CRPS moyen défini au §1.4.5.2. La moyenne est ici portée sur les composantes de la donnée fonctionnelle z, donc le CRPS moyen s'écrit

$$\mathrm{CRPS}_{\hat{Z}'} = \frac{1}{L}\sum_{\ell=1}^{L} \mathrm{crps}(F'_\ell, y_\ell) = \frac{1}{L}\sum_{\ell=1}^{L} \int_{-\infty}^{+\infty} \left(F'_\ell(x) - H(x - y_\ell)\right)^2 \mathrm{d}x \in \mathbb{R}_+, \tag{4.9}$$

où F'_ℓ, pour $\ell \in \{1, \ldots, L\}$, sont les fonctions de répartition des lois marginales de la variable aléatoire réduite $\hat{Z}'$, $y_\ell \in \mathbb{R}$ la réalisation associée (connue *a posteriori*), et $H : \mathbb{R} \to \{0, 1\}$ la fonction d'Heaviside (le CRPS individuel est défini à l'équation (1.3)). Comme les valeurs des CRPS moyen sont difficilement interprétables, elles sont comparées à une valeur de référence qui correspond ici au CRPS moyen de la prévision d'ensemble associée à la variable aléatoire initiale $\hat{Z}$ (formule analogue à l'équation (4.9)). Cela revient à calculer le score de compétence [Ahrens et Walser, 2008] associé au CRPS moyen, le CRPSS (*Continuous Ranked Probability Skill Score*) défini par :

$$\mathrm{CRPSS} = 1 - \frac{\mathrm{CRPS}_{\hat{Z}'}}{\mathrm{CRPS}_{\hat{Z}}} \in]-\infty, 1],$$

où $\mathrm{CRPS}_{\hat{Z}} \in \mathbb{R}_+$ est le CRPS moyen de la prévision d'ensemble associée à la variable aléatoire initiale $\hat{Z}$ (obtenu de manière analogue par l'équation (4.9) mais avec $\hat{Z}$ au lieu de $\hat{Z}'$).

Un CRPSS égal à 0 signifie que la qualité de la prévision d'ensemble associée à la variable aléatoire réduite $\hat{Z}$ est similaire à celle associée à la variable aléatoire initiale $\hat{Z}$, donc que la réduction des scénarios n'impacte globalement pas la qualité des prévisions. Un CRPSS égal à 1 ne peut être obtenu qu'en ne conservant qu'un seul scénario et que ce scénario se réalise effectivement (ce qui n'arrive bien sûr jamais).

4.3.3.4 Stabilité

Les trois critères ci-dessus sont intrinsèques car ils dépendent uniquement de la donnée fonctionnelle z pour la situation considérée. Comme la réduction des scénarios peut être appliquée pour l'estimation des espérances dans le problème d'optimisation probabiliste (4.1) (voir Sous-section 4.1.2), il faut également s'intéresser à l'impact de la réduction des scénarios sur le problème d'optimisation. On utilise pour cela un critère de *stabilité* [Rachev et Römisch, 2002 ; Dupačová *et al.*, 2000 ; Römisch, 2003], qui quantifie l'impact sur la valeur optimale des changements sur la loi de probabilité des scénarios.

On considère que la réduction des scénarios a été appliquée aux apports en eau, aux prix et aux productions variables. En combinant les variables aléatoires réduites associées (supposées indépendantes), on forme la variable aléatoire discrète et finie $\hat{\Xi}'$ de la sous-section 4.3.1 qui correspond à la réduction de la variable aléatoire $\hat{\Xi}$.

Dans note étude, la qualité d'estimation de l'espérance de la valeur optimale Q du second niveau pour des prochaines ventes *day-ahead* $b_0 \in \mathcal{B}_0$ de la variable aléatoire réduite $\hat{\Xi}'$ est comparée par rapport à celle obtenue avec la variable aléatoire initiale $\hat{\Xi}$.

Plus précisément, on utilise le critère suivant (voir Équation (4.6)) :

$$\forall b_0 \in \mathcal{B}_0, \quad \sigma(b_0) = \frac{\mathbb{E}[Q(b_0, \hat{\Xi}')]}{\mathbb{E}[Q(b_0, \hat{\Xi})]} \in \mathbb{R}.$$

On cherche à avoir des valeurs de σ proche de 1 pour réduire les erreurs d'estimation de l'espérance. La difficulté est que ce critère de stabilité n'est pas connu *a priori* sans évaluer la fonction coûteuse Q sur tous les points de $\mathcal{B}_0 \times \hat{\Xi}(\Omega)$, ce qui n'est pas accessible en pratique. Autrement dit, le critère de stabilité σ ne peut être calculé qu'*a posteriori* sur seulement quelques points de $\mathcal{B}_0$ et demande un temps de calcul important.

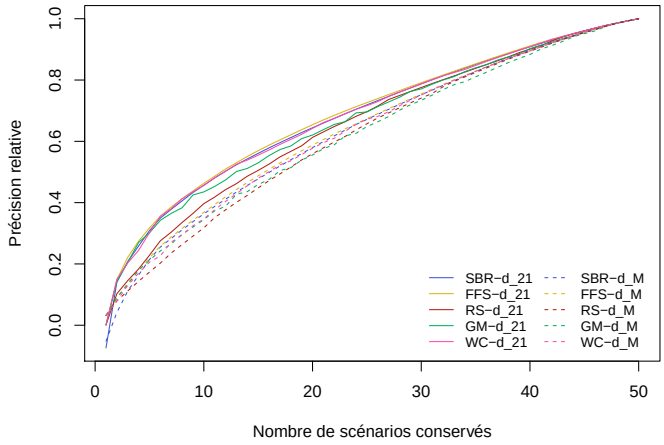
Sous certaines hypothèses (notamment avec des scénarios non discrets et une distance bien choisie), certains critères de stabilité définis dans la littérature peuvent être encadrés par la distance de Kantorovich (et donc la précision relative) [Dupačová *et al.*, 2000], évitant ainsi les nombreuses évaluations de la fonction coûteuse Q. Malheureusement, ces hypothèses ne sont pas satisfaites dans notre cas, donc la stabilité est évaluée sur seulement quelques points de $\mathcal{B}_0$.

4.3.4 Application au cas d'étude

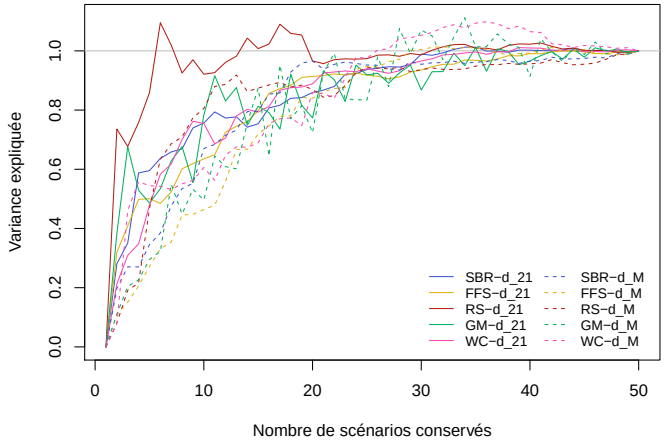
Nous présentons dans cette sous-section les résultats de la qualité de la réduction des scénarios pour le cas d'étude, en considérant uniquement les apports en eau pour simplifier. Plus précisément, nous appliquons la réduction des scénarios sur la prévision d'ensemble brute à 50 membres des débits horaires des 6 affluents du Haut-Rhône (voir Sous-section 2.5.3). Nous n'utilisons donc pas la prévision d'ensemble des apports en eau aux aménagements qui s'en déduisent par des opérations « déterministes » (voir Équation (2.4), les incertitudes des débits d'apports des bassins versants intermédiaires et des débits de correction n'étant pas prises en compte). En outre, nous utilisons les 120 échéances horaires de la prévision d'ensemble brute (voir §2.4.4.3) au lieu des 60 échéances horaires de la fenêtre temporelle de l'optimisation. Cela permet de tester les algorithmes de réduction des scénarios dans un cadre plus général que celui de la planification de la production hydroélectrique à court terme.

La donnée fonctionnelle générique $z = (z(\ell))_{1 \leq \ell \leq L} \in \mathbb{R}^L$ associée à ces données est donc de dimension $L = 6 \times 120$ et le nombre de scénarios initiaux est $J = 50$ (nombre de membres de la prévision d'ensemble brute). En outre, les algorithmes de la sous-section 4.3.2 sont appliqués avec les deux distances de la sous-section 4.2.3 associées aux apports (mais avec 120 échéances horaires au lieu de 60) : la distance d_{2-1} sur les données de dimension non réduite et la distance de Mahalanobis d_M sur les données de dimension réduite par analyse en composantes principales (avec un seuil de variance expliquée minimale égale à 95%).

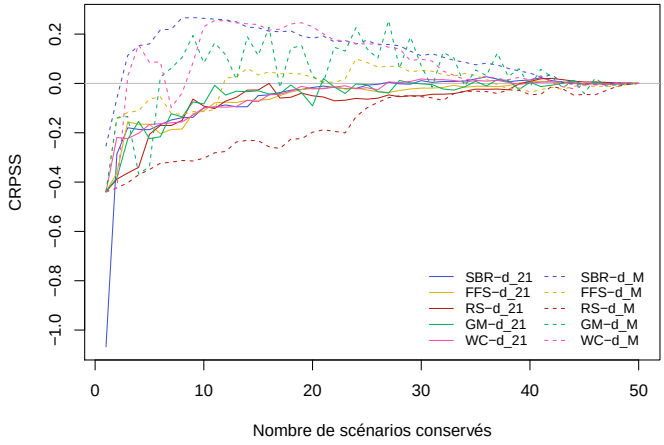
Les quatre critères de qualité de la sous-section 4.3.3 sont calculés pour les différents algorithmes de réduction des scénarios en considérant la prévision d'ensemble brute du cas d'étude (émise le 26/04/2014 à 0 h UTC). Cela permet de comparer les algorithmes en fonction de la distance d_{2-1} ou d_M utilisée et du nombre de scénarios conservés. Les résultats des critères intrinsèques sont présentés à la figure 4.7 et ceux du critère de stabilité sont présentés à la figure 4.8.



(a) Précision relative



(b) Variance expliquée



(c) CRPSS

FIGURE 4.7 – Qualité de la réduction de la prévision d'ensemble brutes des débits horaires des affluents du Haut-Rhône à l'horizon 120 h émise le 26/04/2014 à 0 h UTC. Les courbes continues correspondent à la distance d_{2-1} et les courbes en tirets correspondent à la distance de Mahalanobis d_M.

On observe que la précision relative (voir Figure 4.7a) a globalement la même allure pour tous les algorithmes, avec une croissance de type logarithmique. Par exemple, en conservant environ 15 scénarios, on obtient une précision relative d'environ 50% et en conservant environ 40 scénarios, on obtient une précision relative d'environ 90%. Néanmoins, une réduction efficace qui ne conserve qu'un petit nombre de scénarios (moins de 5) dégrade beaucoup la distribution, limitant l'application de la réduction des scénarios dans la résolution du problème d'optimisation probabiliste (4.1). Les différences de précision relative entre les algorithmes sont surtout présentes dans le « coude » des courbes, pour un nombre de scénarios conservés compris entre 1 et 20. Plus le nombre de scénarios conservés est proche de 50 (i.e. moins la réduction est importante), moins les algorithmes se différencient. On observe que les algorithmes donnent des précisions relatives plus élevées avec la distance d_{2-1} qu'avec la distance de Mahalanobis. En outre, les précisions relatives les plus élevées sont globalement obtenues avec les algorithmes FFS, SBR et WC, et les plus faibles avec l'algorithme RS.

Les allures des courbes de la variance expliquée (voir Figure 4.7b) sont plus chaotiques que celles des courbes de la précision relative, mais avec une croissance globale plus rapide. Par exemple, en conservant environ 15 scénarios, on obtient une variance expliquée d'environ 75% et en conservant environ 25 scénarios (la moitié), on obtient une variance expliquée d'environ 90%. L'algorithme RS avec la distance d_{2-1} est celui qui donne la meilleure variance expliquée pour un faible nombre de scénarios conservés. Pour la situation du cas d'étude, il est possible d'atteindre 80% de variance expliquée avec seulement 5 scénarios conservés, ce qui est intéressant pour l'application de la réduction des scénarios dans la résolution du problème d'optimisation probabiliste (4.1) (mais dans ce cas, la précision relative est plutôt mauvaise). L'algorithme FFS est celui qui donne globalement les plus faibles variances expliquées, en désaccord avec les résultats du premier critère. Les autres algorithmes se différencient peu. On observe également que certains algorithmes dépassent (de moins de 10%) les 100% de variance expliquée : il est possible d'obtenir une distribution réduite sur-dispersée par rapport à la distribution initiale en raison de la redistribution des scénarios conservés.

Les allures des CRPSS (voir Figure 4.7c) sont également plus chaotiques avec une très faible croissance globale. Par exemple, en conservant un nombre de scénarios entre 15 et 40, le CRPSS est globalement le même. On observe que les algorithmes donnent des CRPSS plus proches de 0 avec la distance d_{2-1} qu'avec la distance de Mahalanobis. On observe également qu'il est difficile de départager les algorithmes entre eux pour chacune des deux distances d_{2-1} et d_M car leurs CRPSS sont similaires.

Le critère de stabilité est calculé pour le cas d'étude uniquement avec les prochaines ventes *day-ahead* $b_0^* = b(\mathbf{x}_0^*) \in \mathcal{B}_0$ de la solution déterministe $\mathbf{x}_0^* \in \mathcal{X}(\xi_0)$ de la sous-section 3.2.6. Les éléments de l'ensemble $\mathcal{B}_0 = \{b(\mathbf{x}_0),\ \mathbf{x}_0 \in \mathcal{X}(\xi_0)\}$ sont en effet difficiles à extraire en raison des nombreuses contraintes MILP qui définissent l'ensemble admissible $\mathcal{X}(\xi_0)$. Par ailleurs, on rappelle que la réduction des scénarios est uniquement appliquée sur les apports et pas sur les prix. Les résultats sont présentés à la figure 4.8. On observe que les valeurs des critères de stabilité avec la distance d_{2-1} sont toujours compris dans l'intervalle $[0{,}98, 1{,}02]$ en conservant au moins 3 scénarios, donc les erreurs d'estimation des espérances sont plus faibles que le MIP-gap relatif seuil de 2% utilisé pour les résolutions numériques des problèmes d'optimisation. En revanche, les algorithmes avec la distance de Mahalanobis sont moins concluants, surtout avec un petit nombre de scénarios conservés.

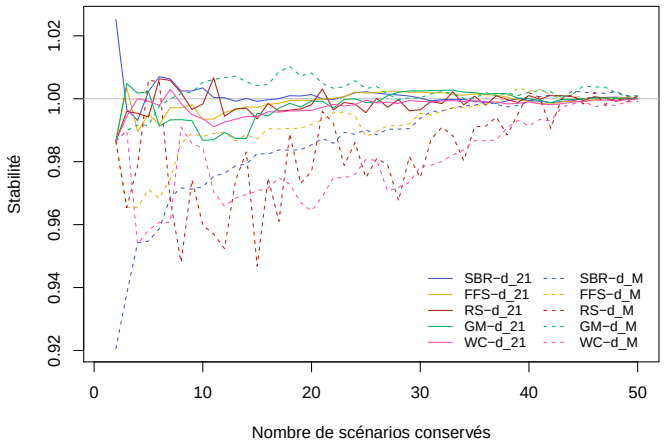


FIGURE 4.8 – Stabilité évaluée aux prochaines ventes *day-ahead* de la solution déterministe pour le cas d'étude. Les courbes continues correspondent à la distance d_{2-1} et les courbes en tirets correspondent à la distance de Mahalanobis d_M.

Algorithme	Distance	Précision relative	Variance expliquée	CRPSS	Stabilité
SBR	d_{2-1}	Bonne	Moyenne	Bon	Bonne
FFS	d_{2-1}	Bonne	Moyenne	Bon	Bonne
RS	d_{2-1}	Mauvaise	Bonne	Moyen	Bonne
GM	d_{2-1}	Moyenne	Moyenne	Moyen	Bonne
WC	d_{2-1}	Bonne	Moyenne	Bon	Bonne
SBR	d_M	Mauvaise	Moyenne	Mauvais	Mauvaise
FFS	d_M	Mauvaise	Mauvaise	Mauvais	Mauvaise
RS	d_M	Mauvaise	Moyenne	Mauvais	Mauvaise
GM	d_M	Mauvaise	Mauvaise	Mauvais	Mauvaise
WC	d_M	Mauvaise	Mauvaise	Mauvais	Mauvaise

TABLEAU 4.1 – Récapitulatif des résultats qualitatifs des algorithmes de réduction des scénarios pour le cas d'étude.

Le tableau 4.1 récapitule de manière qualitative les résultats obtenus. On remarque qu'il est difficile de comparer les algorithmes pour la situation du cas d'étude. Les algorithmes semblent être performants selon un critère au détriment des autres. Néanmoins, le compromis qui nous semble le meilleur entre les critères de qualité de la réduction des scénarios pour le cas d'étude serait obtenu avec :

— l'algorithme SBR (*simultaneous backward reduction*),

— la distance d_{2-1} sur les données de dimension non réduite,

— un nombre de scénarios conservés d'environ 20.

La réduction est alors moyennement efficace (i.e. elle ne supprime pas suffisamment de scénarios) car elle conserve encore 40% des membres (20 sur 50). La distribution des membres conservés de la prévision d'ensemble ainsi obtenue est représentée à la figure 4.9. On observe que les membres ayant les probabilités les plus élevées sont ceux les plus proches de la moyenne. Cela s'explique par le fait que les membres initiaux les plus proches sont localisés autour de la moyenne (voir Figure 3.5a), donc ils se retrouvent fusionnés entre eux par l'algorithme SBR (voir Annexe A.3.6).

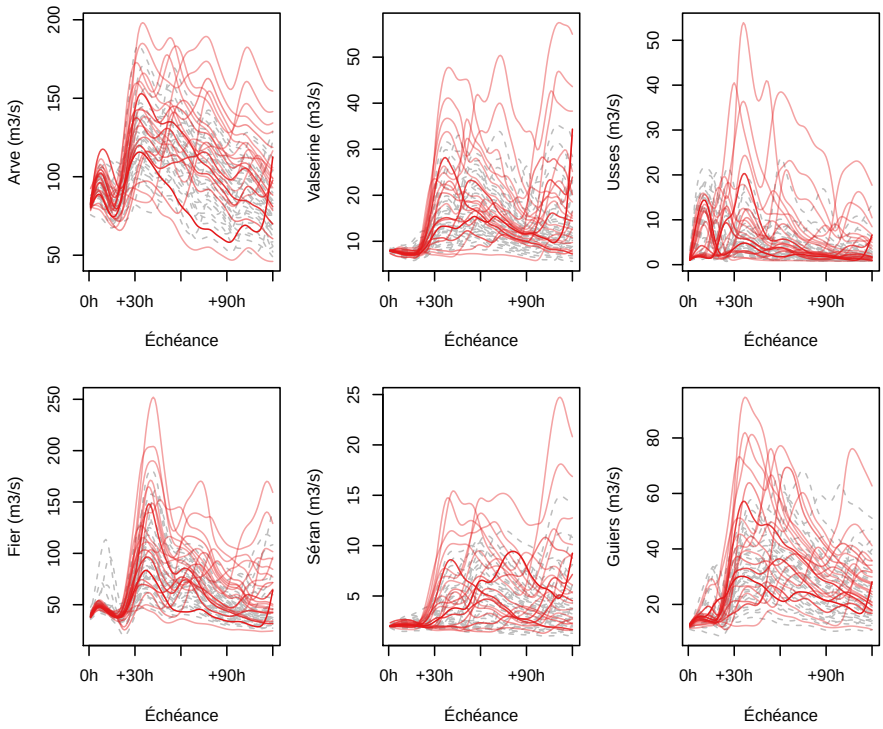


FIGURE 4.9 – Application de l'algorithme SBR pour la situation du cas d'étude avec la distance d_{2-1} et 20 membres conservés sur 50. Les courbes grises en pointillés correspondent aux membres supprimés et les courbes rouges continues aux membres conservés de la prévision d'ensemble brute. L'intensité du rouge indique les probabilités des membres conservés (comprises entre 0,02 et 0,18).

Pour une utilisation dans le problème d'optimisation probabiliste (4.1), le critère de stabilité est crucial car l'estimation de l'espérance intervient directement dans la fonction objectif. Si on pouvait étendre les résultats de la figure 4.8 à n'importe quelles prochaines ventes *day-ahead*, alors le nombre de scénarios à conserver pourrait être abaissé à 3 pour le critère de stabilité, ce qui serait suffisamment efficace pour la résolution numérique du problème d'optimisation MILP (4.7). Néanmoins, on rappelle que le critère de stabilité n'est pas accessible en pratique sans lancer les évaluations de la fonction coûteuse Q.

4.3.5 Comparaison sur un historique de vérification

Pour éviter de comparer les algorithmes de réduction à chaque nouvelle situation, nous préférons choisir celui qui donne les meilleurs résultats globaux sur un historique de vérification. Dans cette sous-section, nous comparons donc les algorithmes de réduction sur l'historique de vérification des prévisions d'ensemble des débits horaires des 6 affluents du Haut-Rhône, émises quotidiennement à 0 h UTC entre le 01/01/2011 et le 25/12/2014 (voir §2.4.4.3).

Les critères de qualité globaux sont obtenus en considérant la moyenne pour les deux premiers critères de la sous-section 4.3.3. Pour le troisième critère, on utilise les CRPS moyen sur l'historique de vérification avant de comparer leur rapport, de manière à correspondre à la définition usuelle des scores de compétences (*skill scores*) en prévision probabiliste [Roulston et Smith, 2002]. Plus précisément, on le critère associé au CRPSS s'écrit

$$\overline{\mathrm{CRPSS}} = 1 - \frac{\overline{\mathrm{CRPS}}'}{\overline{\mathrm{CRPS}}} \in]-\infty, 1],$$

où $\overline{\text{CRPS}} \in \mathbb{R}_+$ et $\overline{\text{CRPS}}' \in \mathbb{R}_+$ sont les CRPS moyens sur l'historique de vérification pour les prévisions d'ensemble initiales et réduites. Le critère de stabilité n'est pas utilisé sur l'historique de vérification.

Les résultats en fonction du nombre de scénarios conservés sont présentés à la figure 4.10. Les allures des courbes de la précision relative moyenne sur l'historique de vérification (voir Figure 4.10a) sont très similaires à celles pour le cas d'étude (voir Figure 4.7a) : les algorithmes FFS, SBR et WC avec la distance d_{2-1} sont ceux qui permettent d'obtenir les précisions relatives moyennes les plus élevées.

Les courbes de la variance expliquée moyenne sur l'historique de vérification (voir Figure 4.10b) sont plus lisses que celles pour le cas d'étude (voir Figure 4.7b) mais les interprétations sont similaires. Comme pour le cas d'étude, la croissance des courbes est plus rapide avec la variance expliquée qu'avec la précision relative moyenne. On observe également que l'algorithme RS avec la distance d_{2-1} est celui qui permet d'avoir une variance expliquée la plus proche de 100%, surtout avec un petit nombre de scénarios conservés. Les algorithmes qui ressortent pour la précision relative moyenne (FFS, SBR et WC avec la distance d_{2-1}) sont dans la moyenne des algorithmes pour la variance expliquée moyenne.

Enfin, les courbes des CRPSS sur l'historique de vérification (voir Figure 4.10c) sont également plus lisses que celles pour le cas d'étude (voir Figure 4.7c) mais les interprétations sont similaires. La différence avec le cas d'étude est que les valeurs des CRPSS sur l'historique de vérification sont toujours négatifs : en moyenne, les prévisions d'ensemble réduites sont moins fiables et moins fines que les prévisions d'ensemble initiales. On observe que les algorithmes donnent des CRPSS plus proches de 0 avec la distance d_{2-1} qu'avec la distance de Mahalanobis. Les algorithmes qui ressortent pour la précision relative moyenne (FFS, SBR et WC avec la distance d_{2-1}) sont également ceux qui ressortent pour le CRPSS.

Ainsi, de manière qualitative, la meilleure réduction des scénarios pour les apports en eau est globalement obtenue avec :

— l'algorithme SBR (*simultaneous backward reduction*),

— la distance d_{2-1},

— un nombre de scénarios conservés entre 10 et 20.

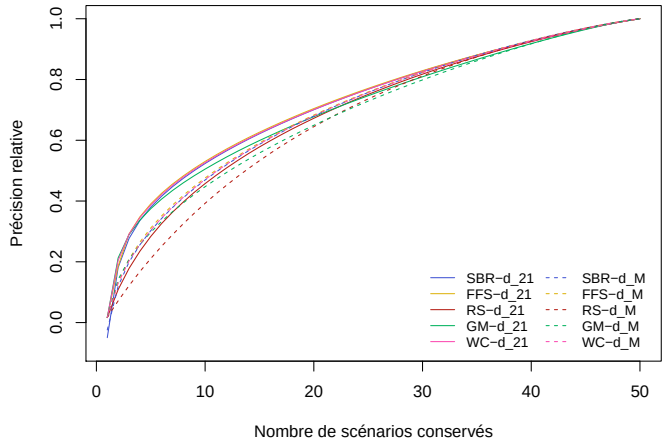
Même si les résultats présentés dans cette sous-section sont parfois prometteurs, nous choisissons de ne pas appliquer de réduction des scénarios pour l'estimation de l'espérance (i.e. pour considérer le problème d'optimisation probabiliste réduit (4.7)) pour les raisons suivants :

— le temps n'est pas contraint durant la thèse (phase d'étude),

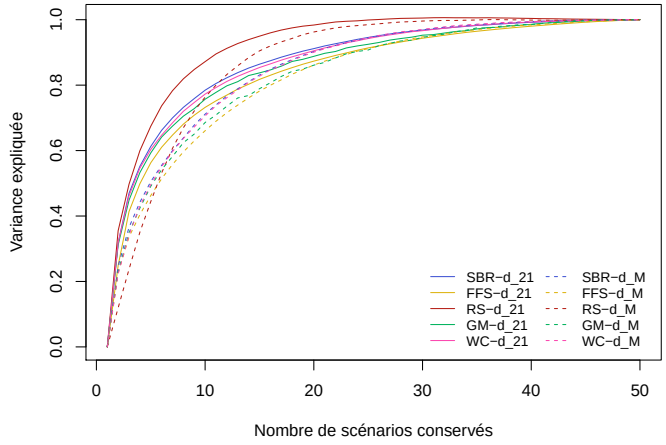
— la réduction des scénarios n'a pas été testée sur les prix *day-ahead*, ni sur la combinaison des apports en eau et des prix,

— la réduction des scénarios sur les apports en eau est moyennement efficace, ne rendant pas la résolution numérique du problème MILP réduit équivalent (4.8) accessible en pratique.

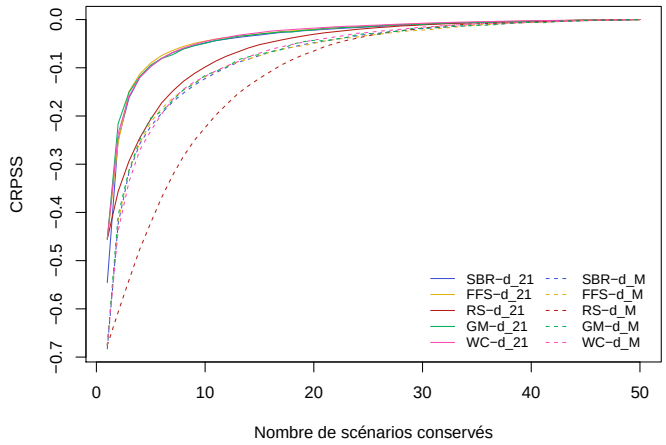
Néanmoins, la réduction des scénarios peut être appliquée pour la construction du métamodèle dans le chapitre 5 qui suit.



(a) Précision relative



(b) Variance expliquée



(c) CRPSS

FIGURE 4.10 – Qualité de la réduction des prévisions d'ensemble brutes des débits horaires des affluents du Haut-Rhône à l'horizon 120 h sur l'historique de vérification des lancements quotidiens à 0 h UTC entre le 01/01/2011 et le 25/12/2014. Les courbes continues correspondent à la distance d_{2-1} et les courbes en tirets correspondent à la distance de Mahalanobis d_M.

4.3.6 Algorithme *tubing*

En plus des algorithmes de réduction des scénarios présentés dans cette section, nous présentons dans cette sous-section l'algorithme *tubing* qui permet de classifier les membres d'une prévision météorologique d'ensemble [Atger, 1999]. En pratique, l'algorithme *tubing* n'est pas utilisé pour effectuer une réduction des scénarios en tant que telle, mais pour aider les prévisionnistes à visualiser et interpréter toutes les informations modélisées par une prévision météorologique d'ensemble. La classification des membres obtenue avec l'algorithme *tubing* permet en effet de considérer les membres de la prévision d'ensemble par familles types plutôt qu'individuellement.

Le principe de l'algorithme *tubing* est de mettre en avant (voir Figure 4.11) :

— un *cluster* central au « centre » de la prévision d'ensemble,

— des *tubes*, autour de la moyenne, dont les axes sont dirigés vers les membres extrêmes.

Le *cluster* central est calculé en regroupant les membres les plus proches de la moyenne de la prévision d'ensemble. Pour cela, il faut définir une distance sur les membres (voir Sous-section 4.2.3) et un critère sur la taille maximale du *cluster* central (seuil sur le nombre de membres ou la variance expliquée par exemple). Le rayon du *cluster* central est alors la distance maximale entre la moyenne et le membre du *cluster* central le plus éloigné de la moyenne.

Les axes des tubes représentent les directions où la réalisation est susceptible de s'écarter de la moyenne de la prévision d'ensemble. Les tubes sont construits itérativement. L'axe du premier tube correspond à la droite entre la moyenne et le membre le plus éloigné de la moyenne (premier membre extrême). Le premier tube est calculé est regroupant tous les membres hors du *cluster* central dont la distance à l'axe du tube est inférieure au rayon du *cluster* central. Le processus est itéré tant qu'il reste des membres extrêmes qui ne sont pas dans (au moins) un tube.

L'algorithme fournit également des poids au *cluster* central (en fonction de sa taille) et aux tubes (en fonction de la distance des membres extrêmes). Tous les membres de la prévision d'ensemble appartiennent soit au *cluster* central, soit à un ou plusieurs tubes (les tubes peuvent s'intersecter). La moyenne du *cluster* central constitue un élément pour une prévision déterministe (prévision la plus probable). Le nombre de tubes est une échelle de confiance de cette prévision déterministe. Les tubes sont représentés non pas par leur moyenne mais par leurs membres extrêmes qui indiquent des possibles variantes à la prévision déterministe.

La figure 4.12 illustre les résultats de l'algorithme *tubing* appliqué à la prévision d'ensemble brute des débits horaires des 6 affluents du Haut-Rhône utilisée pour le cas d'étude. Le *cluster* central est obtenu avec un seuil minimal de variance expliquée égale à 50%. Le poids du *cluster* central obtenu est de 0,74.

Comme le poids du *cluster* central est lié à la performance de la moyenne de la prévision d'ensemble (cadre déterministe), il peut servir à la définition du paramètre α_0 qui intervient dans le problème d'optimisation probabiliste (4.1) (voir Sous-section 3.3.4). Par exemple, avec l'application sur le cas d'étude, on obtient $\alpha_0 = 0{,}74$ (voir Figure 4.12). Dans le cadre de la thèse, il s'agit du principal intérêt de l'algorithme *tubing* car on ne cherche pas à interpréter et à donner de l'expertise aux prévisions d'ensemble. Le poids $\alpha = 0{,}74$ obtenu pour le cas d'étude est donc supérieur au seuil $\alpha_0^*(b_0^*) = 0{,}5$ à partir duquel la fonction objectif probabiliste appliquée à la solution déterministe est considérée comme égale à celle de la fonction déterministe, rendant l'optimisation probabiliste peu pertinente (voir Sous-section 3.3.5). Si cette observation venait à être confirmée pour une grande partie des solutions admissibles (et pas seulement la solution déterministe), alors l'utilisation du cadre probabiliste pour le cas d'étude serait peu pertinente (l'optimisation en univers déterministe suffirait). Ce point est abordé plus en détails dans la section 6.2.

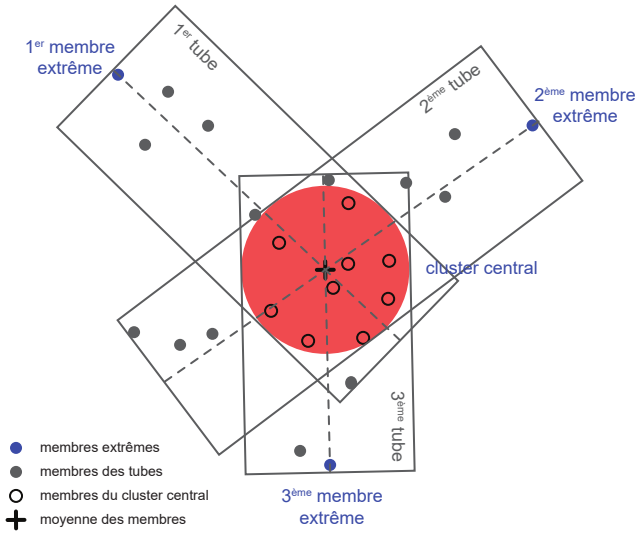


FIGURE 4.11 – Schéma du fonctionnement de l'algorithme *tubing* pour une prévision d'ensemble en deux dimensions.

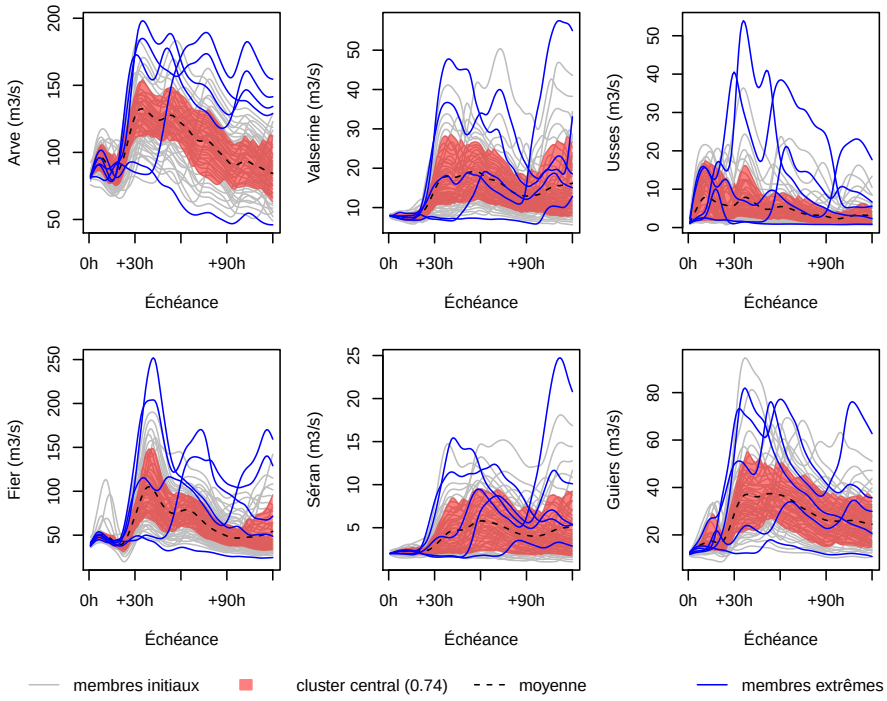


FIGURE 4.12 – Application de l'algorithme *tubing* pour la prévision d'ensemble des débits horaires des 6 affluents du Haut-Rhône à l'horizon 120 h émise le 26/04/2014 à 0 h UTC.

4.4 Conclusion du chapitre

La méthodologie que nous proposons dans ce manuscrit pour résoudre le problème d'optimisation de la planification de la production en univers probabiliste est basée sur deux points : la réduction des scénarios et la construction d'un métamodèle par apprentissage surpervisé. Une des difficultés pour l'apprentissage du métamodèle réside dans la structure fonctionnelle des données d'entrée. Pour contourner cette difficulté, nous proposons d'appliquer une réduction de la dimension des données d'entrée par analyse en composantes principales. Ces données d'entrée réduites peuvent être utilisées pour définir une distance sur les données d'entrée, servant notamment à la réduction des scénarios. Néanmoins, les algorithmes de réduction des scénarios semblent plus performants avec une distance sur les données d'entrée de dimension non réduite qu'avec une distance sur les données d'entrée de dimension réduite. Par ailleurs, les résultats montrent que la réduction des scénarios semble moyennement efficace, même avec les approches les plus performantes.

Dans le chapitre suivant, nous présentons les étapes de la construction du métamodèle. Certaines de ces étapes reposent sur les résultats présentés dans ce chapitre : la réduction de la dimension des données fonctionnelles et la réduction des scénarios.

Chapitre 5

Plan d'expérience et métamodèle

La première partie du manuscrit aboutit à la formulation mathématique du problème de planification de la production en univers probabiliste par programmation stochastique à deux niveaux. Pour résoudre numériquement ce problème, la méthodologie décrite dans le chapitre précédent repose notamment sur l'utilisation d'un métamodèle pour remplacer la valeur optimale du second niveau. Ce chapitre est consacré à la construction de ce métamodèle. La première section de ce chapitre sert d'introduction en rappelant la problématique et en décrivant la démarche proposée. Les sections qui suivent décrivent une à une les étapes successives de la démarche. Les deuxième et troisième sections permettent de construire le jeu de données pour l'apprentissage du métamodèle, en présentant respectivement la discrétisation des données d'entrée et le plan d'expérience. Ces deux sections font appel aux résultats du chapitre précédent concernant la réduction de la dimension des données d'entrée et la réduction des scénarios. Le jeu de données ainsi construit sert à l'apprentissage du métamodèle présenté dans la quatrième section.

5.1 Approximation par métamodèle

5.1.1 Rappel de la méthodologie

On rappelle que l'on s'intéresse à la planification à court terme de la production d'une chaîne hydroélectrique et d'actifs de production variable en univers probabiliste, dont la formulation mathématique est le problème de programmation dynamique stochastique à deux niveaux (voir Équation (3.27)) :

$$\max_{\mathbf{x}_0 \in \mathcal{X}(\xi_0)} \left\{ \alpha_0 f(\mathbf{x}_0, \xi_0) + (1-\alpha_0)\mathbb{E}[Q(b(\mathbf{x}_0), \hat{\Xi})] \right\}, \tag{5.1}$$

où (voir Sections 3.2 et 3.3) :

- $\xi_0 = (a_0, \pi_0, \phi_0) \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ est le scénario déterministe composé des prévisions déterministes des apports en eau $a_0 \in \mathbb{R}^{IT}$ (membre contrôle de la prévision d'ensemble), des prix $\pi_0 \in \mathbb{R}^T$ et des productions variables $\phi_0 \in \mathbb{R}^T$ ($I \in \mathbb{N}^*$ désigne le nombre d'aménagements de la chaîne hydroélectrique et T le nombre de pas de temps de la fenêtre temporelle sur laquelle la planification de la production est considérée),
- $\hat{\Xi}$ est une variable aléatoire discrète et finie sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$, représentant les scénarios probabilistes $\xi_n = (a_n, \pi_n, \phi_n) \in \hat{\Xi}(\Omega)$, pour $n \in \{1, \dots, N\}$, et leurs probabilités $p_n = \mathbb{P}[\hat{\Xi} = \xi_n] \in]0, 1[$, $p_1 + \cdots + p_N = 1$, ($N \in \mathbb{N}^*$ étant le nombre de scénarios probabilistes),
- $\mathcal{X}(\xi) \subset \mathbb{R}^r$ est l'ensemble des solutions admissibles pour un scénario $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ (partie fermée et bornée de $\mathbb{R}^r$, non vide si $\xi \in \{\xi_0, \xi_1, \dots, \xi_N\}$), défini à partir de contraintes MILP (i.e. linéaires mixtes),

— $\alpha_0 \in [0,1]$ est un paramètre arbitraire, qui peut être défini par le poids d'un *cluster* central (voir Sous-section 4.3.6) des scénarios probabilistes (dans ce cas, le poids α_0 est un indicateur de la dispersion des scénarios probabilistes par rapport au scénario déterministe : plus α_0 est proche de 0, plus les incertitudes sont fortes),

— f est une fonction linéaire réelle définie sur $\mathbb{R}^r \times (\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T)$ qui représente le chiffre d'affaires d'une solution pour un scénario,

— b est une fonction linéaire de $\mathbb{R}^r$ dans $\mathbb{R}^{24}$ qui donne les prochaines ventes *day-ahead* associées à une solution,

— Q est une fonction réelle, coûteuse à évaluer numériquement, définie sur $\mathbb{R}^{24} \times (\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T)$ qui donne le chiffre d'affaires optimal que l'on peut atteindre pour une décision de ventes *day-ahead* $b_0 \in \mathbb{R}^{24}$ et un scénario $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ sous l'hypothèse qu'il se réalise effectivement (problème MILP) :

$$\forall (b_0, \xi) \in \mathbb{R}^{24} \times (\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T), \quad Q(b_0, \xi) = \max_{\substack{\mathbf{x} \in \mathcal{X}(\xi) \\ b(\mathbf{x}) = b_0}} f(\mathbf{x}, \xi) \in \mathbb{R} \cup \{-\infty\}. \tag{5.2}$$

On rappelle également que $\mathcal{B}_0 = \{b(\mathbf{x}_0),\ \mathbf{x}_0 \in \mathcal{X}(\xi_0)\} \subset \mathbb{R}^{24}$ est l'ensemble des prochaines ventes *day-ahead* compatibles avec le scénario déterministe ξ_0. Cet ensemble est borné et défini par les contraintes MILP qui interviennent dans l'ensemble admissible $\mathcal{X}(\xi_0)$.

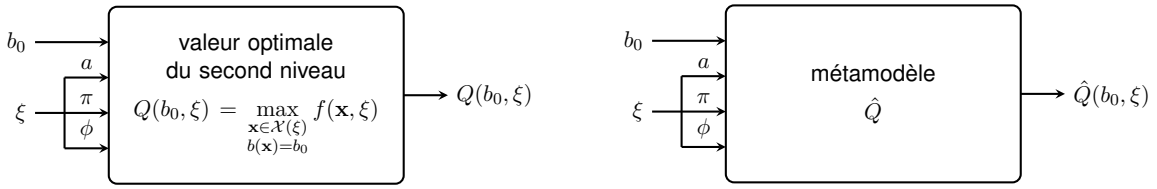
La méthodologie décrite dans le chapitre précédent (voir Sous-section 4.1.3) repose sur la construction d'un métamodèle (émulation statistique) $\hat{Q}$ qui remplace la fonction Q (coûteuse à évaluer) dans le problème d'optimisation probabiliste (5.1). On cherche donc à occulter le problème d'optimisation du second niveau. La fonction Q est alors considérée comme une boîte noire, dont les données d'entrée sont des prochaines ventes *day-ahead* $b_0 \in \mathbb{R}^{24}$ et un scénario $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$, et la donnée de sortie est le scalaire qui représente la valeur optimale $Q(b_0, \xi) \in \mathbb{R} \cup \{-\infty\}$ (voir Figure 5.1a). Dans l'écriture du problème d'optimisation probabiliste (5.1), cette boîte noire n'est pas utilisée avec tous les éléments de $\mathbb{R}^{24} \times (\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T)$ en entrée, mais seulement avec ceux de $\mathcal{B}_0 \times \hat{\Xi}(\Omega)$, pour lesquels la fonction Q est supposée à valeurs réelles (voir Sous-section 3.3.4). Les deux données d'entrée ont en revanche des rôles bien distincts dans le problème d'optimisation probabiliste (5.1) : les prochaines ventes *day-ahead* $b_0 \in \mathcal{B}_0$ sont des variables de décision, alors que les scénarios $\xi_n \in \hat{\Xi}(\Omega)$ sont des paramètres utilisés pour évaluer les espérances.

Ainsi, on cherche à construire un métamodèle $\hat{Q}$ sur $\mathbb{R}^{24} \times (\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T)$ qui est rapide à évaluer et qui est une approximation de la fonction Q sur $\mathcal{B}_0 \times \hat{\Xi}(\Omega)$:

$$\forall (b_0, \xi_n) \in \mathcal{B}_0 \times \hat{\Xi}(\Omega), \quad \hat{Q}(b_0, \xi_n) \approx Q(b_0, \xi_n). \tag{5.3}$$

La substitution de la fonction Q par le métamodèle $\hat{Q}$ dans le problème d'optimisation probabiliste (5.1) permet donc d'éviter les évaluations de la fonction Q responsables de l'explosion du temps de calcul par rapport au problème d'optimisation déterministe (voir Sous-section 4.1.2).

Cela revient à remplacer la boîte noire de la fonction Q par celle du métamodèle $\hat{Q}$, comme illustré à la figure 5.1. Il convient de noter qu'en dissociant les apports en eau, les prix et les productions variables qui forment les scénarios, on peut également considérer que la boîte noire de la fonction Q et celle du métamodèle $\hat{Q}$ ont chacune quatre données d'entrée (voir Figure 5.1). Cela est surtout pertinent du fait de l'hypothèse d'indépendance entre les apports en eau, les prix et les productions variables utilisée dans le cadre de la thèse (voir Sous-section 2.4.4).



(a) Fonction coûteuse Q (b) Métamodèle $\hat{Q}$

FIGURE 5.1 – Schéma type « boîte noire » de la valeur optimale du second niveau et du métamodèle.

Selon son temps d'évaluation, le métamodèle $\hat{Q}$ peut être utilisé en complément ou à la place de la réduction des scénarios pour l'estimation de l'espérance (voir Sous-section 4.3.1). En effet, la condition de l'équation (5.3) assure en particulier que

$$\forall b_0 \in \mathcal{B}_0, \quad \begin{cases} \mathbb{E}[\hat{Q}(b_0,\hat{\Xi})] = \sum_{n=1}^{N} p_n \hat{Q}(b_0,\xi_n) \approx \sum_{n=1}^{N} p_n Q(b_0,\xi_n) = \mathbb{E}[Q(b_0,\hat{\Xi})], \\ \mathbb{E}[\hat{Q}(b_0,\hat{\Xi}')] = \sum_{k=1}^{N'} p'_k \hat{Q}(b_0,\xi'_k) \approx \sum_{k=1}^{N'} p'_k Q(b_0,\xi'_k) = \mathbb{E}[Q(b_0,\hat{\Xi}')], \end{cases} \tag{5.4}$$

(car on impose $\hat{\Xi}'(\Omega) \subset \hat{\Xi}(\Omega)$). Néanmoins, on rappelle que la réduction des scénarios n'est pas considérée pour l'estimation des espérances (voir Sous-section 4.3.5), donc on suppose dans la suite que le métamodèle $\hat{Q}$ est utilisé avec les N scénarios probabilistes. Ainsi, le problème d'optimisation probabiliste (5.1) est approché par le problème d'optimisation

$$\max_{\mathbf{x}_0\in\mathcal{X}(\xi_0)} \left\{ \alpha_0 f(\mathbf{x}_0,\xi_0) + (1-\alpha_0)\mathbb{E}[\hat{Q}(b(\mathbf{x}_0),\hat{\Xi})] \right\}. \tag{5.5}$$

Le problème d'optimisation approché (5.5) est celui qui permettra de résoudre le problème de planification de la production en univers probabiliste.

5.1.2 Construction du métamodèle

Le métamodèle $\hat{Q}$ peut être construit par apprentissage supervisé (i.e. par un calibrage automatique à partir d'un jeu de données) selon différents modes opératoires :

1. de manière globale (i.e. une fois pour toute), pour être utilisé pour toutes les situations (éventuellement en distinguant les situations par catégories),
2. de manière locale dans une phase de pré-traitement de l'optimisation pour la situation courante,
3. de manière locale et imbriquée à la phase de résolution du problème d'optimisation pour la situation courante.

Un métamodèle global est plus difficile à construire qu'un métamodèle local car il doit prendre en compte une donnée d'entrée supplémentaire liée à la situation, la fonction Q changeant beaucoup à chaque nouvelle situation et les données d'entrée (prochaines ventes *day-ahead* et scénarios) étant dépendantes de la situation considérée. Cette donnée d'entrée supplémentaire est également difficile à définir car les situations font intervenir de nombreux paramètres variés qui sont difficiles à décrire correctement. C'est pourquoi nous nous limitons à la construction d'un métamodèle local. Si les résultats obtenus ne sont pas concluants, alors il est inutile de chercher à construire un métamodèle global. S'ils s'avèrent en revanche concluants, alors la question de la construction d'un métamodèle global pourra se poser.

La construction d'un métamodèle local pendant la résolution du problème d'optimisation (troisième point ci-dessus) correspond aux méthodes par approximations successives présentées à

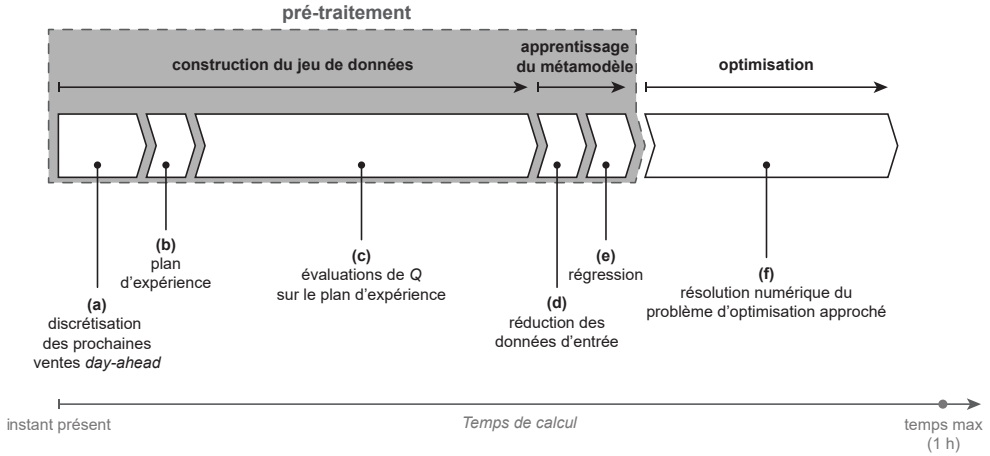


FIGURE 5.2 – Enchaînement des étapes de la méthodologie qui repose sur le calibrage du métamodèle dans une phase de pré-traitement.

la sous-section 4.1.2. Ces méthodes rejoignent le principe des méthodes de décomposition où une approximation de la valeur optimale Q du second niveau est construite itérativement en modifiant certaines contraintes. La différence est que les méthodes par approximations successives remplacent le problème d'optimisation du second niveau par un métamodèle au lieu de modifier certaines contraintes. Le métamodèle est construit de manière itérative grâce aux solutions optimales approchées obtenues à chaque itération. Néanmoins, la littérature à ce sujet est assez pauvre et elle concerne surtout des problèmes d'optimisation quadratiques sans variables discrètes [Longstaff et Schwartz, 2001 ; Deák *et al.*, 2012 ; Deák, 2011 ; Flammarion, 2017 ; van der Vlerk, 2010 ; Romeijnders *et al.*, 2016 ; Bernreuther, 2017]. Par ailleurs, chaque itération des méthodes par approximations successives revient à construire un métamodèle à partir d'un jeu de données, ce qui est similaire à la problématique de construction d'un métamodèle dans une phase de pré-traitement (deuxième point ci-dessus). La différence entre les deux approches locales réside dans la définition du jeu de données. Avec les méthodes par approximations successives, le jeu de données est enrichi grâce aux itérations. Dans une phase de pré-traitment, le jeu de données doit être défini en amont de l'optimisation. Néanmoins, l'étude de la construction du métamodèle dans une phase de pré-traitement est une étape nécessaire avant de commencer à étudier les méthodes par approximations successives : la mise en œuvre des méthodes par approximation successives dépend des conclusions obtenues et des difficultés rencontrées avec la construction du métamodèle dans une phase de pré-traitement. C'est pourquoi nous nous intéressons à la construction du métamodèle dans une phase de pré-traitement. Les grandes lignes d'une méthode par approximations successives sont néanmoins présentées à la sous-section 6.3.3 du chapitre suivant.

Comme l'apprentissage du métamodèle $\hat{Q}$ est effectué lors d'une phase de pré-traitement avant de résoudre le problème d'optimisation probabiliste approché (5.5) pour la situation courante (voir Figure 5.2), il doit être répété à chaque nouvelle situation. Le temps de calcul pour le calibrage du métamodèle $\hat{Q}$ et la résolution du problème d'optimisation (5.5) doit donc être limité pour rester compatible aux exigences du processus opérationnel (résolution totale en moins de 1 h).

5.1.3 Étapes de la phase de pré-traitement

La phase de pré-traitement est divisée en deux sous-phases principales : la construction du jeu de données (étapes (a) à (c) de la figure 5.2) et l'apprentissage du métamodèle (étapes (d) et (e) de la figure 5.2).

Le jeu de données nécessaire à l'apprentissage du métamodèle $\hat{Q}$ est un ensemble fini de valeurs des données d'entrée $\mathcal{B}_0 \times \hat{\Xi}(\Omega)$ et des évaluations de la fonction Q associées. La difficulté de l'apprentissage dans une phase de pré-traitement est que le jeu de données doit être construit à chaque nouvelle situation. Cela nécessite de choisir un échantillon de points des données d'entrée $\mathcal{B}_0 \times \hat{\Xi}(\Omega)$ (ce que l'on appelle un *plan d'expérience*) puis d'évaluer numériquement la fonction coûteuse Q sur chacun de ces points.

La difficulté pour construire le plan d'expérience est que l'ensemble $\mathcal{B}_0 = \{b(\mathbf{x}_0),\ \mathbf{x}_0 \in \mathcal{X}(\xi_0)\} \subset \mathbb{R}^{24}$ des prochaines ventes *day-ahead* n'est pas connu sous une forme explicite mais à partir des contraintes MILP de l'ensemble admissible $\mathcal{X}(\xi_0)$. Il est donc difficile de caractériser les éléments de $\mathcal{B}_0$, ni même de tirer aléatoirement un certain nombre d'entre eux. C'est pourquoi nous cherchons à approcher l'ensemble $\mathcal{B}_0$ par un ensemble $\hat{\mathcal{B}}_0 \subset \mathbb{R}^{24}$. C'est l'objet de la section 5.2 qui suit, où la démarche proposée permet d'obtenir un ensemble $\hat{\mathcal{B}}_0$ fini (étape (a) de la figure 5.2). L'ensemble $\mathcal{B}_0 \times \hat{\Xi}(\Omega)$ des données d'entrée est donc approché par l'ensemble fini $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$ (l'ensemble $\hat{\Xi}(\Omega)$ des scénarios probabilistes est fini).

Comme l'ensemble $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$ est fini, il serait possible d'appliquer la fonction coûteuse Q à chacun de ses points, permettant de construire un jeu de données complet. Néanmoins, le nombre d'éléments de l'ensemble $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$ est *a priori* trop élevé pour envisager une telle approche. La fonction Q est en effet coûteuse à évaluer et la construction du jeu de données est une des étapes de la résolution numérique pour laquelle le temps de calcul total est limité pour rester compatible aux exigences du processus opérationnel (temps de calcul total inférieur à 1 h). C'est pourquoi nous cherchons à construire un plan d'expérience à partir de l'ensemble $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$. Cela consiste à sélectionner un petit nombre de points de l'ensemble fini $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$ (étape (b) de la figure 5.2) sur lesquels la fonction coûteuse Q est évaluée pour former le jeu de données (étape (c) de la figure 5.2). C'est l'objet de la section 5.3 de ce chapitre. La difficulté est que les données d'entrée sont fonctionnelles et que les points sélectionnés doivent correctement décrire l'ensemble $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$.

Le type du métamodèle doit être choisi en fonction de plusieurs considérations, notamment :

- son implémentation dans le problème d'optimisation (5.5),
- la dimension des données d'entrée,
- la taille du jeu de données,
- la structure de la fonction Q,
- le temps de calcul disponible pour l'apprentissage.

C'est pourquoi la section 5.4 de ce chapitre présente et compare plusieurs type de métamodèle pour les étapes (d) et (e) de la figure 5.2.

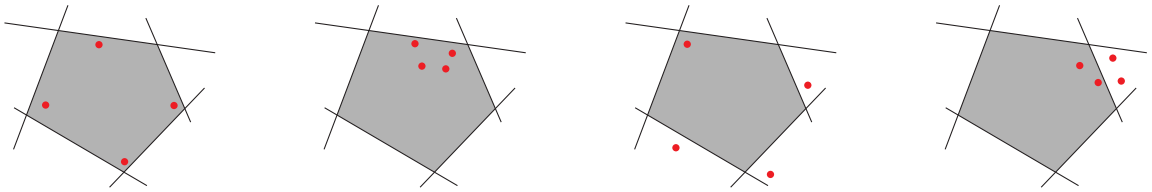
La validation de toute la méthodologie proposée dans ce manuscrit est présentée dans la section 6.2 du chapitre suivant.

5.2 Discrétisation des prochaines ventes *day-ahead*

5.2.1 Problématique et méthodologie

La première phase pour le calibrage du métamodèle $\hat{Q}$ dans une phase de pré-traitement à l'optimisation est la construction du jeu de données pour la situation étudiée (étapes (a) à (c) de la figure 5.2), ce qui nécessite :

- de construire un plan d'expérience (i.e. de choisir un nombre fini de points des données d'entrée, étape (b) de la figure 5.2),
- d'évaluer numériquement la fonction Q sur chacun des points du plan d'expérience (étape (c) de la figure 5.2).



(a) $\hat{\mathcal{B}}_0$ idéal

(b) $\hat{\mathcal{B}}_0$ compatible mais mal réparti

(c) $\hat{\mathcal{B}}_0$ bien réparti mais pas compatible

(d) $\hat{\mathcal{B}}_0$ ni compatible ni bien réparti

FIGURE 5.3 – Illustration des critères de qualité de l'approximation $\hat{\mathcal{B}}_0$ de l'ensemble $\mathcal{B}_0$. Les lignes noires représentent, en 2 dimensions, les hyperplans de contraintes supposées linéaires pour les schémas. La partie en gris représente le polyèdre convexe $\mathcal{B}_0$ des prochaines ventes *day-ahead*. Les points en rouge représentent les éléments de l'ensemble approché $\hat{\mathcal{B}}_0$ supposé fini.

Le plan d'expérience est idéalement défini à partir de l'ensemble des données d'entrée d'intérêt, i.e. $\mathcal{B}_0 \times \hat{\Xi}(\Omega)$. L'ensemble $\hat{\Xi}(\Omega) = \{\xi_1, \dots, \xi_N\}$ est connu car il s'agit des scénarios probabilistes. En revanche, l'ensemble $\mathcal{B}_0 = \{b(\mathbf{x}_0),\ \mathbf{x}_0 \in \mathcal{X}(\xi_0)\} \subset \mathbb{R}^{24}$ n'est pas connu sous une forme explicite mais à partir de l'ensemble admissible $\mathcal{X}(\xi_0)$ défini par un grand nombre de contraintes MILP (voir Sous-section 3.2.3). Il est donc difficile de caractériser les éléments de $\mathcal{B}_0$, ni même de tirer aléatoirement un certain nombre d'entre eux. L'ensemble $\mathcal{B}_0 \times \hat{\Xi}(\Omega)$ des données d'entrée n'est donc pas accessible pour former le plan d'expérience.

C'est pourquoi nous ajoutons une étape préliminaire qui consiste à approcher l'ensemble $\mathcal{B}_0$ des prochaines ventes *day-ahead* par un ensemble $\hat{\mathcal{B}}_0 \subset \mathbb{R}^{24}$ dont les éléments sont explicites (étape (a) de la figure 5.2). Le plan d'expérience est alors défini à partir de l'ensemble $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$ au lieu de l'ensemble $\mathcal{B}_0 \times \hat{\Xi}(\Omega)$. Deux critères semblent importants à respecter pour que l'approximation $\hat{\mathcal{B}}_0$ de l'ensemble $\mathcal{B}_0$ puisse servir à définir le plan d'expérience (voir Figure 5.3) :

— la compatibilité des éléments de $\hat{\mathcal{B}}_0$ avec le scénario déterministe ξ_0 : dans l'idéal, on souhaite $\hat{\mathcal{B}}_0 \subset \mathcal{B}_0$,

— la répartition des éléments de $\hat{\mathcal{B}}_0$ par rapport à l'ensemble $\mathcal{B}_0$: dans l'idéal, on souhaite que les distances entre deux éléments de $\hat{\mathcal{B}}_0$ ne soient pas trop petites par rapport au diamètre [1] de l'ensemble borné $\mathcal{B}_0$.

Il convient de noter que le critère de compatibilité peut être évalué *a posteriori* pour un élément de $\hat{\mathcal{B}}_0$ (vérification du respect des contraintes MILP) mais pas le critère de répartition car le diamètre de l'ensemble $\mathcal{B}_0$ n'est pas accessible.

Dans la suite, nous proposons de construire l'ensemble $\hat{\mathcal{B}}_0$ de manière qu'il soit fini. Il semble en effet plus difficile de construire une approximation « continue » de l'ensemble $\mathcal{B}_0$, surtout que les données d'entrée devraient ensuite être discrétisées pour former le plan d'expérience. Nous cherchons alors à construire une famille finie $(b_1, \dots, b_M)$ d'éléments distincts de $\mathbb{R}^{24}$ ($M \in \mathbb{N}^*$) telle que l'ensemble $\hat{\mathcal{B}}_0 = \{b_1, \dots, b_M\} \subset \mathbb{R}^{24}$ soit une « bonne » approximation de l'ensemble $\mathcal{B}_0$ selon les deux critères ci-dessus.

1. Le diamètre d'un ensemble borné est par définition la borne supérieure des distances entre deux éléments de cet ensemble.

Pour construire la famille finie $(b_1, \ldots, b_M)$, on peut citer (de manière non-exhaustive) les trois approches suivantes :

— la sélection de $M \leq N$ éléments de $\{b(\mathbf{x}_n^*),\ 1 \leq n \leq N\}$, où $\mathbf{x}_n^* \in \mathcal{X}(\xi_n)$, pour $n \in \{1, \ldots, N\}$, sont les solutions optimales intermédiaires obtenues en appliquant l'optimiseur déterministe (problème MILP (3.18)) à chacun des N scénarios probabilistes (comme ce qui a été effectué à la sous-section 3.3.2),

— la prévision probabiliste par régression, i.e. l'estimation des prochaines ventes *day-ahead* à partir de la série temporelle des ventes *day-ahead* effectivement réalisées sur une période passée,

— la prévision probabiliste par analogie (voir §1.4.5.1), i.e. l'utilisation des ventes *day-ahead* effectivement réalisées à M situations analogues à la situation courante (plus proches voisins).

La sélection des M solutions optimales intermédiaires dans la première approche peut être effectuée en échantillonnant selon la distribution de l'ensemble $\{b(\mathbf{x}_n^*),\ 1 \leq n \leq N\}$ des ventes *day-ahead* intermédiaires ou de l'ensemble $\{Q(b(\mathbf{x}_n^*), \xi_n),\ 1 \leq n \leq N\}$ des valeurs optimales intermédiaires. Cette première approche permet d'obtenir des prochaines ventes *day-ahead* cohérentes avec la situation considérée mais pas forcément compatibles avec le scénario déterministe ni bien réparties dans l'ensemble $\mathcal{B}_0$. En effet, les solutions optimales intermédiaires sont *a priori* admissibles qu'avec les scénarios probabilistes pour lesquelles elles sont définies. En outre, cette première approche nécessiterait le calcul de M solutions intermédiaires, ce qui peut être coûteux en temps de calcul sans calcul parallèle (mémoire distribuée).

Les deux autres approches sont définies à partir d'un historique passé des ventes *day-ahead* : l'approche par régression repose sur le choix d'un modèle statistique issu de la théorie des séries temporelles, tandis que l'approche par analogie repose sur le choix des situations analogues. Elles ne permettent donc pas de garantir le respect du critère de compatibilité et du critère de répartition. Pour la même raison, elles ne permettent pas non plus d'obtenir des prochaines ventes *day-ahead* cohérentes avec la situation considérée, contrairement à la première approche. Le critère de compatibilité risque donc d'être moins bien respecté qu'avec la première approche. En revanche, leur atout est qu'elles demandent un temps de calcul beaucoup plus faible que la première approche. En outre, l'approche par régression semble plus difficile à mettre en œuvre que l'approche par analogie car elle suppose que les ventes *day-ahead* futures puissent être expliquées par les ventes *day-ahead* passées, ce qui n'est pas évident *a priori*[2].

Ainsi, on s'intéresse à l'approche par analogie dans la suite de cette section. Elle consiste à construire la famille finie $(b_1, \ldots, b_M) \in (\mathbb{R}^{24})^M$ en sélectionnant les ventes *day-ahead* effectivement réalisées pour M situations passées les plus proches de la situation courante (situations analogues) parmi une archive de situations passées. Les éléments $b_m \in \mathbb{R}^{24}$, pour $m \in \{1, \ldots, M\}$, sont considérés comme équiprobables. Le principe général de l'approche par analogie est illustré à la figure 5.4.

Le choix des situations analogues est contrôlé par des paramètres, appelés *hyperparamètres* dans la suite de cette section. Les hyperparamètres principaux, qui sont détaillés dans la sous-section 5.2.2 qui suit, sont les suivants :

— l'archive des situations candidates dans laquelle les situations analogues sont sélectionnées,

— les variables d'analogie et les critères d'analogie pour comparer les situations entre elles,

— la taille d'échantillonnage pour définir le nombre d'analogues sélectionnés.

2. Une étude statistique pourrait permettre d'établir d'éventuelles relations sur un historique passé. Néanmoins, les possibles évolutions futures du marché de l'énergie pourraient modifier ces relations.

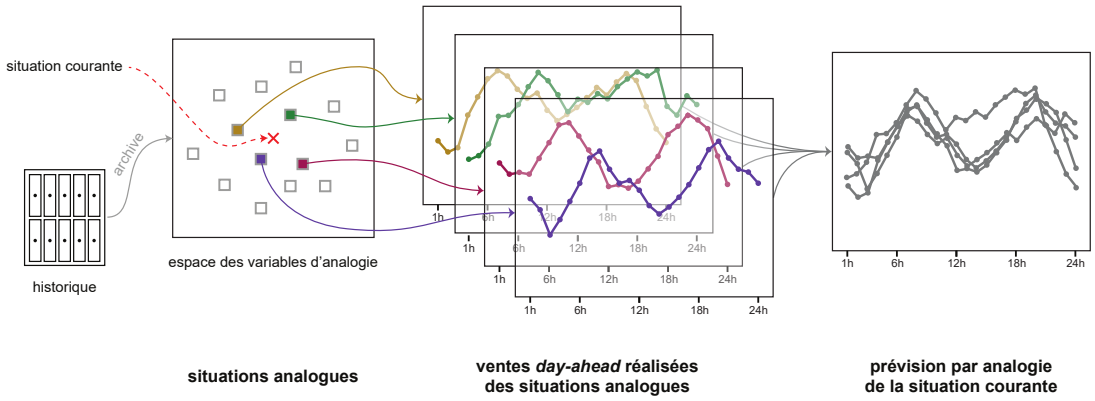


FIGURE 5.4 – Schéma de la construction d'une prévision par analogie.

Il est difficile de définir la meilleure combinaison des valeurs de ces hyperparamètres sans mesurer au préalable leur qualité de discrétisation pour les comparer. La meilleure combinaison des valeurs des hyperparamètres pourrait être déterminée à chaque nouvelle situation, en appliquant toutes les combinaisons possibles puis en choisissant celle qui permet d'avoir la meilleure qualité pour la situation courante. Néanmoins, pour limiter le temps de calcul pour une nouvelle situation, nous cherchons à définir la combinaison des hyperparamètres qui fournit la meilleure qualité globale de discrétisation sur un historique de vérification (et pas seulement la situation courante). C'est l'objet de la sous-section 5.2.3 de ce chapitre.

5.2.2 Hyperparamètres pour la prévision par analogie

5.2.2.1 Archive des situations candidates

L'approche par analogie repose sur l'utilisation d'une archive historique. Pour pouvoir comparer les situations historiques avec la situation courante, l'archive historique doit contenir les valeurs des variables d'analogie et des ventes *day-ahead* effectivement réalisées aux situations passées. On rappelle que les ventes *day-ahead* sont la différence entre la production totale horaire prévue en fin de matinée (les ventes *day-ahead* sont effectuées chaque jour avant midi) et les engagements à long terme pour le lendemain (voir Sous-section 1.3.2). Or, les engagements à long terme sont moins spécifiques aux situations car ils sont généralement définis sur des périodes longues (voir §1.2.2.1). Il semble donc plus intéressant de définir l'archive historique avec les prévisions à 11 h de la production totale du lendemain au lieu de l'historique des ventes *day-ahead* effectivement réalisées (qui portent sur la production du lendemain). Les ventes *day-ahead* effectivement réalisées s'en déduisent alors en second temps, en considérant la différence entre les productions totales horaires prévues pour le lendemain et les engagements des ventes à long terme. On rappelle que les engagements des ventes à long terme sont supposés nuls pour le cas d'étude (voir Sous-section 2.5.1). Avec les données dont nous disposions, l'archive historique est constitué des situations quotidiennes (fin de matinée) entre le 01/01/2014 et le 31/12/2019 (inclus).

Par ailleurs, on considère que les situations analogues à une situation considérée ne sont cherchées que dans une partie de l'archive historique, appelée dans la suite l'archive des situations candidates (pour la situation considérée). Cela permet de réduire le nombre de comparaisons et donc de limiter les temps de calcul. L'archive des situations candidates dépend de la situation considérée mais on suppose qu'elle est de taille fixe et qu'elle se décale de manière dynamique dans l'archive historique à chaque nouvelle situation, comme illustré à la figure 5.5. Cela permet d'être adapté au contexte opérationnel où les données s'ajoutent de manière séquentielle.

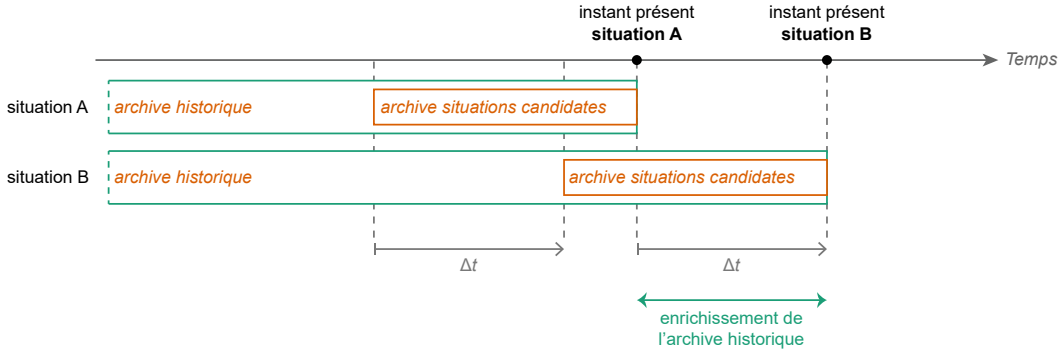


FIGURE 5.5 – Schéma de la dynamique de l'archive des situations candidates.

Le choix de l'archive des situations candidates influe *a priori* sur la qualité de l'échantillon obtenu par analogie. C'est pourquoi nous proposons de tester séparément trois archives différentes, chacune constituée d'un nombre fixe de situations passées consécutives antérieures à la situation courante, mais en tenant compte de catégories différentes :

1. sans catégorie particulière,
2. distinction par rapport à la saison (on ne considère que les situations passées de la même saison que celle de la situation courante),
3. distinction par rapport à la catégorie semaine/weekend (on ne considère que les situations passées appartenant à la même catégorie semaine/weekend que celle de la situation courante).

Les deux derniers cas peuvent être nécessaires pour capter les variations hebdomadaires et saisonnaires de la production de la chaîne hydroélectrique (et donc des ventes *day-ahead*).

La taille des archives des situations candidates est arbitraire. Elle est choisie de manière à être suffisamment grande pour refléter la diversité de toutes les situations de l'archive historique [3] et suffisamment petite pour réduire les temps de calcul des comparaisons avec la situation courante. Dans le cadre de la thèse, nous utilisons une taille fixe égale à 1 000 pour les archives des deux premiers cas (sans catégorie ou distinction par saisons) et une taille fixe égale à 500 ou 1 000 pour l'archive du dernier cas (distinction semaine/weekend) si la situation courante est en weekend ou en semaine. L'archive est réduite à 500 pour les situations en weekend (samedi ou dimanche) car l'archive historique est trop restreinte pour choisir une taille plus grande. Une analyse plus poussée serait nécessaire en pratique pour déterminer la « meilleure » taille de l'archive selon un attribut qu'il faudrait définir.

5.2.2.2 Variables d'analogie et critères d'analogie

L'approche par analogie consiste principalement à comparer les situations les unes avec les autres. La difficulté est qu'un nombre important de paramètres changent d'une situation à l'autre. Il est donc crucial de déterminer avec attention :

— la ou les variables d'analogie, i.e. les prédicteurs qui expliquent le mieux les ventes *day-ahead*,

— le critère d'analogie, i.e. la distance qui permet de mesurer les similitudes entre les variables d'analogie.

3. Ce que l'on appellerait la climatologie pour des prévisions météorologiques.

Nous avons fait le choix, par expertise parmi les données que nous disposions, de tester séparément les trois variables d'analogie suivantes :

1. la puissance totale réalisée la veille par la chaîne hydroélectrique à la granularité horaire (car les ventes *day-ahead* dépendent des conditions du marché de l'énergie et des conditions météorologiques qui changent généralement peu d'une journée à l'autre),
2. la prévision déterministe des débits d'apports cumulés de tous les affluents du Rhône du lendemain à la granularité horaire (car elle détermine globalement la production hydroélectrique du lendemain, donc les ventes *day-ahead*),
3. la prévision déterministe des prix *day-ahead* du lendemain à la granularité horaire (car elle détermine les variations temporelles de la production du lendemain, donc aussi des ventes *day-ahead*).

La première variable d'analogie est liée aux conditions initiales des situations alors que les deux autres sont liées aux scénarios déterministes des situations. Pour la deuxième variable d'analogie, on considère les débits d'apports cumulés et non les apports de chaque affluent pris individuellement car on cherche à expliquer globalement la production hydroélectrique totale du Rhône (sachant que les débits des affluents sont corrélés, voir §2.2.1.5). Il serait également possible d'affecter des poids différents à chaque affluent de manière à tenir compte de leurs influences sur la production totale. Pour chacune de ces variables d'analogie, une situation est donc représentée par un élément de $\mathbb{R}^{24}$. Ces variables d'analogie sont testées séparément, i.e. on ne considère pas une combinaison de plusieurs d'entre elles.

Chacune des variables d'analogie est associée à un critère d'analogie défini par une distance sur les valeurs des variables d'analogie. Pour cela, nous reprenons les distances définies à la sous-section 4.2.3 du chapitre précédent, mais pour des données fonctionnelles de dimension 24. Comme le choix du critère d'analogie influe *a priori* sur la qualité de discrétisation, nous choisissons de tester et comparer les distances sur les données non réduites (voir §4.2.3.1) et les distances de Mahalanobis sur les données réduites (voir §4.2.3.2).

— Pour les distances sur les données non réduites, on utilise donc la distance d_1 (associée à la norme L^1 dans $\mathbb{R}^{24}$) pour la première variable d'analogie (puissance) et la distance d_2 (associée à la norme L^2 dans $\mathbb{R}^{24}$) pour la troisième variable d'analogie (prix). Pour la deuxième variable d'analogie (apports), on utilise la distance d_{2-1} mais qui se réduit ici à la distance d_1 (associée à la norme L^1 dans $\mathbb{R}^{24}$) car on considère les apports cumulés (et non les apports à chaque affluent pris individuellement).

— Les distances de Mahalanobis d_M sont obtenues pour chaque variable d'analogie après réduction de la dimension par analyse en composantes principales avec un seuil de variance expliquée minimale égale à 95%.

Pour aller plus loin, il aurait été possible de combiner les trois variables d'analogie ci-dessus en pondérant les distances associées à chacune d'entre elles (dans ce cas, une situation serait représentée par un élément de $\mathbb{R}^{24} \times \mathbb{R}^{24} \times \mathbb{R}^{24}$). Cela nécessiterait néanmoins des temps de calcul plus importants qu'avec une seule variable d'analogie et demanderait de définir les poids des variables d'analogie, ce qui n'est pas évident *a priori*.

Les variables et critères d'analogie servent à comparer les situations de l'archive des situations candidates avec la situation courante. Cela permet de sélectionner les M situations les plus proches de la situation courante (plus proches voisins) : c'est le principe de la sélection par analogie. Les figures 5.6, 5.7 et 5.8 représentent les variables d'analogie de 100 situations analogues pour le cas d'étude en fonction de l'archive des situations candidates, de la variable d'analogie et du critère d'analogie, ainsi que les ventes *day-ahead* associées aux situations analogues. On observe que la sélection des analogues diffère selon les hyperparamètres. On observe également que les variables d'analogie des situations analogues sont moins dispersées autour de celles de la

situation étudiée avec les critères d'analogie associés aux distances d_1 ou d_2 sans catégorie ou par saison (les sous-figures a et b) qu'avec les autres combinaisons (les sous-figures c à f). Les mêmes observations sont valables pour les ventes *day-ahead* associées aux situations analogues. On observe également que la dispersion des variables d'analogie des puissances réalisées la veille et des apports cumulés du lendemain se répercute sur celle des ventes *day-ahead*. En revanche, avec la variable d'analogie des prix du lendemain, les échantillons des ventes *day-ahead* ont une dispersion similaire pour les deux critères d'analogie, ce qui n'est pas le cas pour les prix du lendemain.

5.2.2.3 Taille d'échantillonnage

La taille d'échantillonnage M est arbitraire et elle doit également être choisie avec attention car elle influe *a priori* sur la qualité de la prévision par analogie. Dans le cadre de la thèse, nous testons et comparons les résultats obtenus avec toutes les tailles d'échantillonnage M comprises entre $M = 1$ et $M = 100$. Nous ne considérons pas des tailles d'échantillonnage supérieures à $M = 100$ pour les raisons suivantes :

— la taille de l'archive des situations candidates est égale à 1 000 (voire 500 pour celles définies par distinction semaine/weekend) donc plus la taille d'échantillonnage M est grande, moins la sélection est pertinente (avec $M = 100$, les situations sélectionnées représentent au moins 10% de l'archive des situations candidates, ce qui est suffisamment grand),

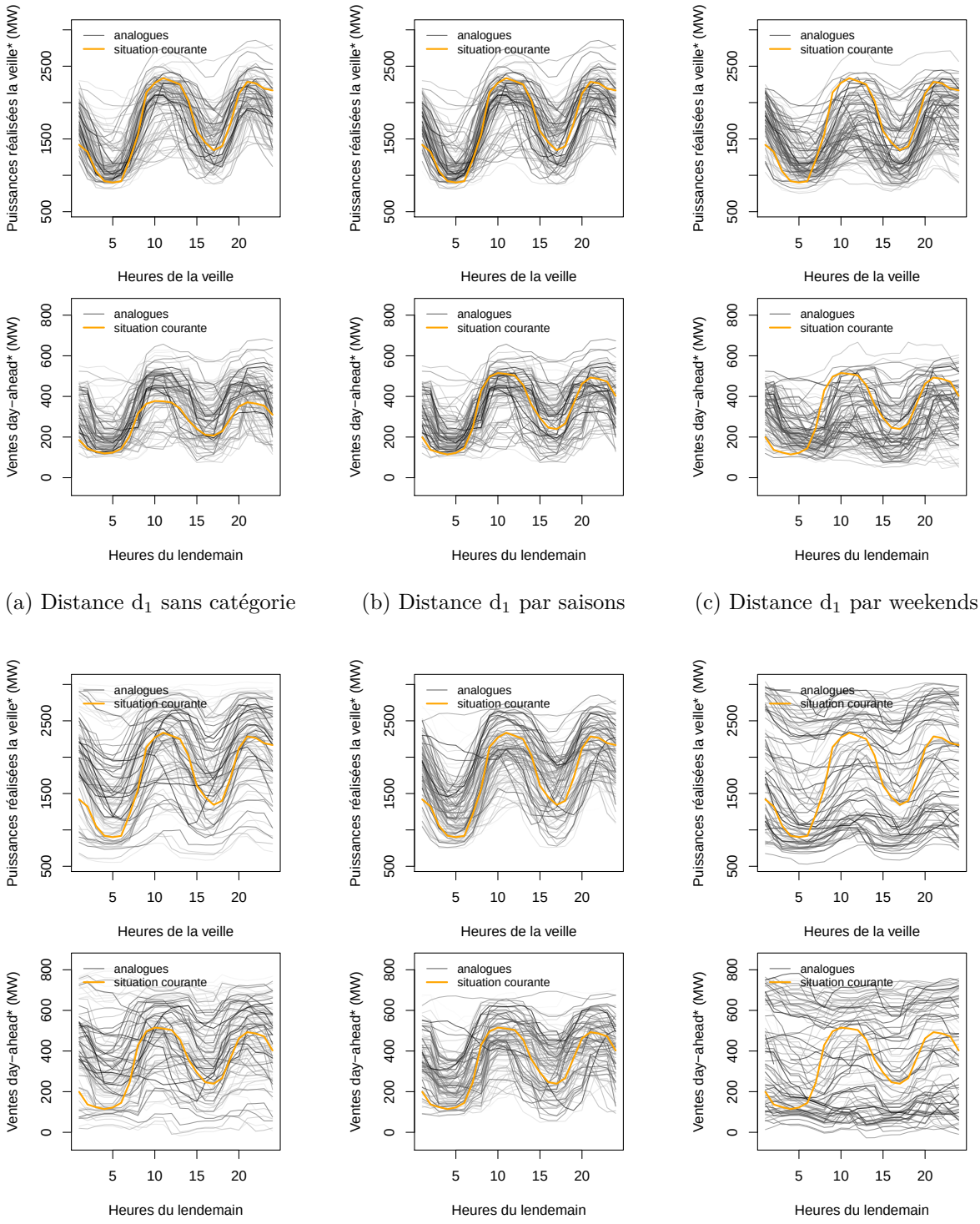
— plus on sélectionne de situations analogues éloignées de la situation courante, plus le critère de compatibilité de la sous-section 5.2.1 avec l'ensemble $\hat{\mathcal{B}}_0 = \{b_1, \ldots, b_M\}$ risque de ne pas être respecté (où $b_m \in \mathbb{R}^{24}$, pour $m \in \{1, \ldots, M\}$, sont les ventes *day-ahead* effectivement réalisées aux situations analogues sélectionnées).

5.2.3 Qualité de discrétisation

Les différents choix possibles présentés à la sous-section 5.2.2 précédente pour chaque hyperparamètre permettent de définir 1 800 combinaisons des hyperparamètres (3 choix pour l'archive des situations candidates, 3 choix pour la variable d'analogie, 2 choix pour le critère d'analogie et 100 choix pour la taille, soit $3 \times 3 \times 2 \times 100 = 1\,800$ combinaisons). Dans cette sous-section, nous cherchons à déterminer la combinaison qui fournit la meilleure qualité de discrétisation, globalement sur un historique de vérification. La combinaison sélectionnée est celle qui sera effectivement utilisée en pratique pour construire l'ensemble $\hat{\mathcal{B}}_0$ approchant l'ensemble $\mathcal{B}_0$ des prochaines ventes *day-ahead* pour une situation donnée. Comme l'étude de la qualité est effectuée de manière globale sur un historique de vérification, il n'y a pas besoin de la refaire à chaque nouvelle planification de la production. Nous gagnons ainsi du temps de calcul.

Compte tenu de l'archive historique à disposition (les 2 191 situations quotidiennes entre le 01/01/2014 et le 31/12/2019) et de la taille de l'archive des situations candidates (500 ou 1 000 situations), les méthodes de prévision par analogie associées aux trois archives des situations candidates du §5.2.2.1 peuvent être étudiées sur l'historique de vérification composé des situations quotidiennes de l'année 2019. Si l'archive historique était plus petit ou si la taille fixe de l'archive des situations candidates était plus grande, alors une validation croisée serait plus pertinente.

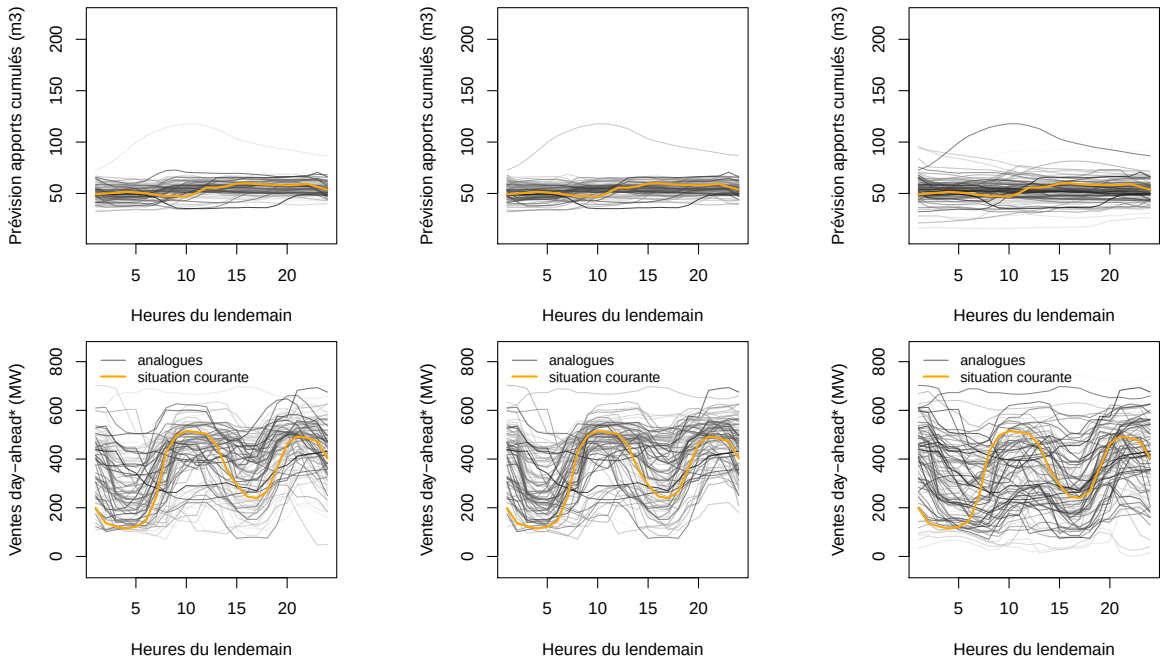
Dans le cadre de la thèse, la qualité de discrétisation de l'ensemble $\mathcal{B}_0$ pour une situation fait référence à la cohérence statistique de l'échantillon $(b_1, \ldots, b_M)$ avec les ventes *day-ahead* $\tilde{b} \in \mathbb{R}^{24}$ effectivement réalisées pour la situation considérée. Pour vérification, on calcule le CRPS moyen, un score classique utilisé dans le domaine des prévisions météorologiques probabilistes pour quantifier leur qualité. Il mesure conjointement deux attributs des prévisions : leur fiabilité et leur finesse (voir §1.4.5.2).



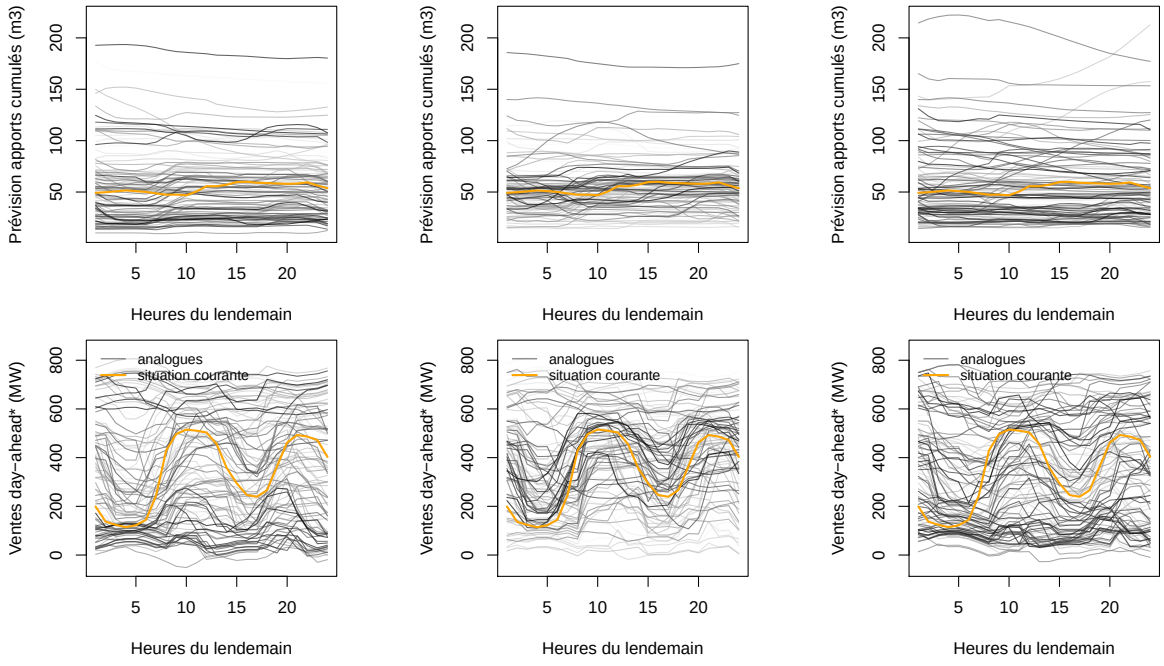
(a) Distance d_1 sans catégorie (b) Distance d_1 par saisons (c) Distance d_1 par weekends

(d) Distance d_M sans catégorie (e) Distance d_M par saisons (f) Distance d_M par weekends

FIGURE 5.6 – Représentation de la variable d'analogie correspondant à la puissance réalisée la veille (graphiques du haut) avec les ventes *day-ahead* associées (graphiques du bas) pour les 100 situations analogues à la situation du cas d'étude (situation courante) en fonction de l'archive des situations candidates et du critère d'analogie (distance d_1 ou d_M). Pour le cas d'étude, l'archive historique comporte également les situations ultérieures qui ne seraient pas accessibles en pratique. Plus le niveau de gris est proche du noir, plus la similarité est forte. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les courbes.

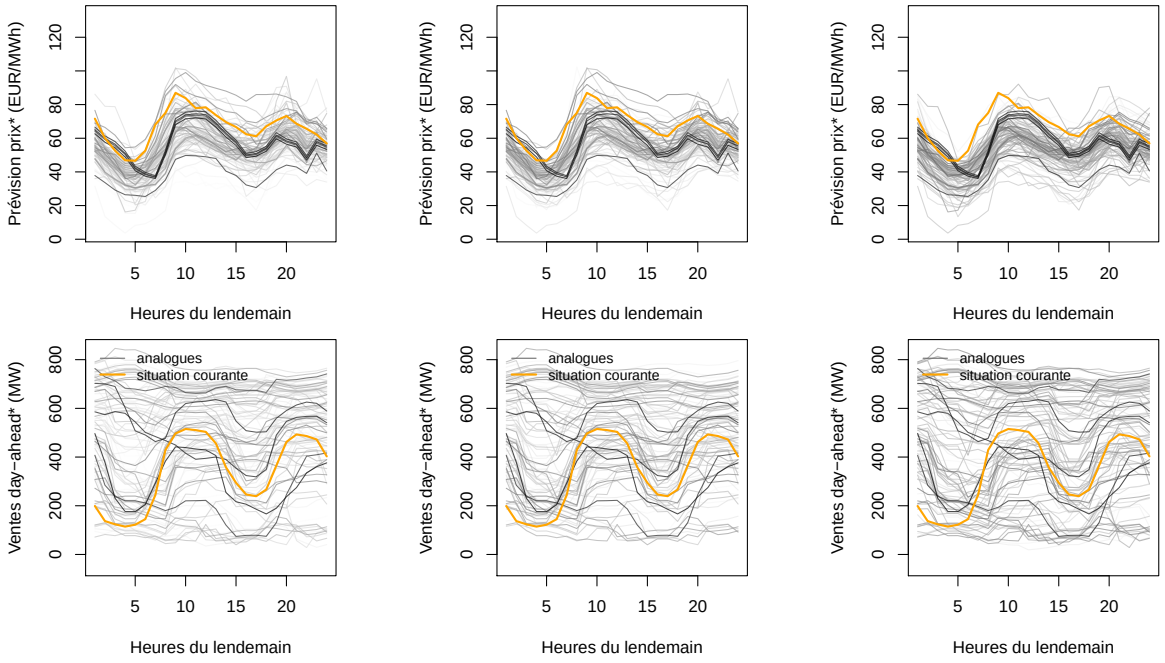


(a) Distance d_1 sans catégorie (b) Distance d_1 par saisons (c) Distance d_1 par weekends

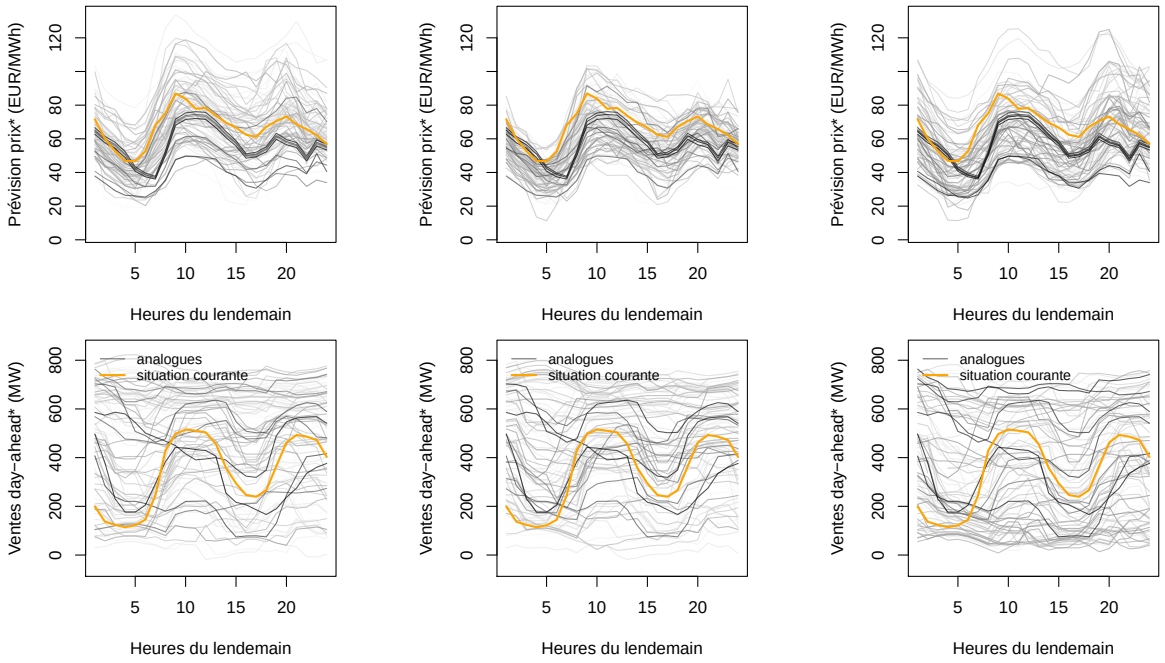


(d) Distance d_M sans catégorie (e) Distance d_M par saisons (f) Distance d_M par weekends

FIGURE 5.7 – Représentation de la variable d'analogie correspondant à la prévision des apports cumulés du lendemain (graphiques du haut) avec les ventes *day-ahead* associées (graphiques du bas) pour les 100 situations analogues à la situation du cas d'étude (situation courante) en fonction de l'archive des situations candidates et du critère d'analogie (distance d_1 ou d_M). Pour le cas d'étude, l'archive historique comporte également les situations ultérieures qui ne seraient pas accessibles en pratique. Plus le niveau de gris est proche du noir, plus la similarité est forte. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les ventes *day-ahead*.



(a) Distance d_2 sans catégorie (b) Distance d_2 par saisons (c) Distance d_2 par weekends



(d) Distance d_M sans catégorie (e) Distance d_M par saisons (f) Distance d_M par weekends

FIGURE 5.8 – Représentation de la variable d'analogie correspondant à la prévision des prix horaires du lendemain (graphiques du haut) avec les ventes *day-ahead* associées (graphiques du bas) pour les 100 situations analogues à la situation du cas d'étude (situation courante) en fonction de l'archive des situations candidates et du critère d'analogie (distance d_1 ou d_M). Pour le cas d'étude, l'archive historique comporte également les situations ultérieures qui ne seraient pas accessibles en pratique. Plus le niveau de gris est proche du noir, plus la similarité est forte. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les courbes.

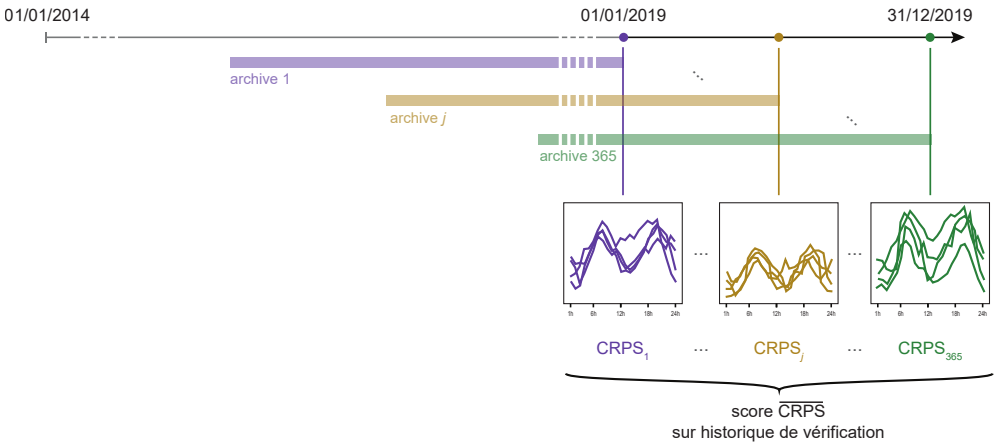


FIGURE 5.9 – Schéma de la vérification de la qualité de discrétisation pour une combinaison des hyperparamètres de la prévision par analogie.

On rappelle que le CRPS moyen compare la fonction de répartition empirique de l'échantillon avec celle de l'observation, i.e. les ventes *day-ahead* réalisées dans notre cas. Plus précisément, le CRPS moyen pour la situation considérée est le score positif défini par :

$$\mathrm{CRPS}\left((b_1,\dots,b_M),\tilde{b}\right) = \frac{1}{24}\sum_{h=1}^{24}\int_{-\infty}^{+\infty}\left(F_{(b_1,\dots,b_M),h}(x) - H(x-\tilde{b}_h)\right)^2 \mathrm{d}x \in \mathbb{R}_+,$$

où $H : \mathbb{R} \to \mathbb{R}$ est la fonction d'Heaviside et $F_{(b_1,\dots,b_M),h} \in \mathcal{F}(\mathbb{R},\mathbb{R})$, pour $h \in \{1,\dots,24\}$, est la fonction de répartition de l'échantillon $(b_1,\dots,b_M) \in (\mathbb{R}^{24})^M$ pour l'heure h, i.e.

$$\forall x \in \mathbb{R}, \quad F_{(b_1,\dots,b_M),h}(x) = \frac{1}{M}\sum_{m=1}^{M} H(x-b_{mh}).$$

Plus le CRPS moyen est proche de zéro, plus la qualité d'échantillonnage est bonne.

La qualité globale de discrétisation sur l'historique de vérification est donc évaluée par la moyenne des CRPS moyens des situations quotidiennes de l'année 2019, notée $\overline{\mathrm{CRPS}} \in \mathbb{R}_+$, comme l'illustre la figure 5.9. L'objectif de la suite de cette sous-section est de déterminer la combinaison des valeurs des hyperparamètres de la sous-section 5.2.2 (choix de l'archive des situations candidates, de la variable d'analogie, du critère d'analogie et de la taille d'échantillonnage) qui minimise le score $\overline{\mathrm{CRPS}}$.

Les résultats sont présentés sur la figure 5.10. Les courbes des scores $\overline{\mathrm{CPRS}}$ sont globalement continues en fonction du nombre M d'analogues. On observe globalement que les scores $\overline{\mathrm{CPRS}}$ diminuent fortement lorsque le nombre d'analogues augmente entre $M = 1$ et $M \approx 10$, puis ils continuent de diminuer plus légèrement jusqu'à environ $M \approx 25$ avant d'augmenter légèrement lorsque M augmente au-delà de $M \approx 25$. Les scores $\overline{\mathrm{CPRS}}$ présentent donc un minimum autour de $M \approx 25$. On observe que les choix de la variable d'analogie (couleur des courbes) et du critère d'analogie (styles des traits) ont plus d'impact sur la qualité globale de discrétisation que le choix de la catégorie de l'archive des situations candidates (épaisseurs des courbes). On remarque que les deux critères d'analogie se valent pour chaque variable d'analogie. Les différences du score $\overline{\mathrm{CPRS}}$ sont principalement dues au choix de la variable d'analogie. La variable d'analogie associée aux prix donne les moins bons résultats ($\overline{\mathrm{CPRS}}$ les plus élevés) tandis que celle associée aux apports cumulés donne les meilleurs résultats ($\overline{\mathrm{CPRS}}$ les plus faibles).

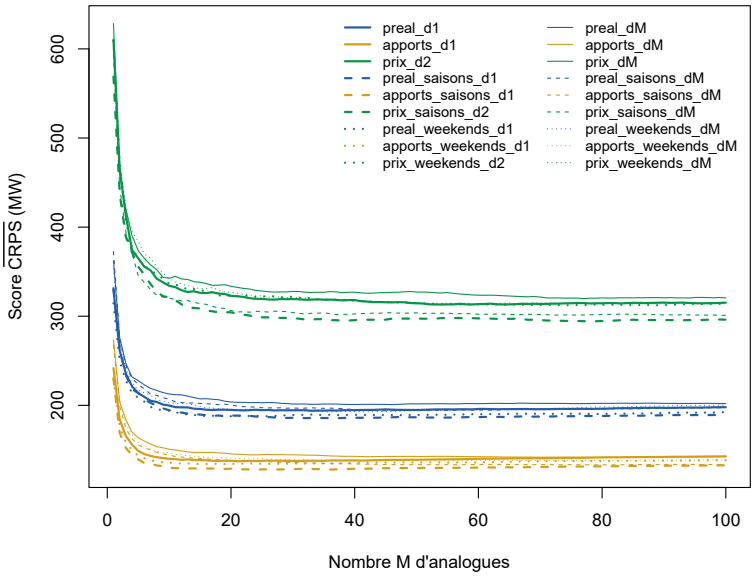


FIGURE 5.10 – Qualité globale de discrétisation évaluée par le score $\overline{\text{CRPS}}$ sur l'historique de vérification en fonction de la taille d'échantillonnage pour les différentes combinaisons de l'archive des situations candidates, de la variable d'analogie et du critère d'analogie.

La combinaison des hyperparamètres qui permet de minimiser le score $\overline{\text{CRPS}}$ est la suivante :

- la variable d'analogie définie par la prévision déterministe des débits d'apports cumulés du lendemain,
- l'archive par saisons,
- le critère d'analogie associé à la distance sur les données non réduites (distance d_1),
- la taille d'échantillonnage $M = 24$.

5.2.4 Application au cas d'étude

La méthode d'échantillonnage par analogie construite à partir de la « meilleure » combinaison des valeurs des hyperparamètres sur l'historique de vérification n'est en revanche pas utilisable en l'état pour le cas d'étude. En effet, le cas d'étude correspond à la situation du 26/04/2014 pour laquelle l'historique des données que nous avions à disposition durant la thèse ne permet pas de construire des archives des situations candidates suffisamment grandes de situations antérieures (il n'y a que 116 jours entre le 01/01/2014 et le 26/04/2014). Ainsi, pour le cas d'étude, les archives des situations candidates sont définies avec tout l'archive historique global, donc elles couvrent des situations postérieures à la situation du 26/04/2014 qui ne seraient pas accessibles en pratique.

La méthode d'échantillonnage pour le cas d'étude est utilisée avec la variable d'analogie associée à la prévision déterministe des débits d'apports cumulés du lendemain et le critère d'analogie associé à la distance sur les données non réduites (distance d_1), comme suggérés par les résultats de la sous-section 5.2.3. En revanche, contrairement aux résultats de la sous-section 5.2.3, on ne considère pas de catégorie dans l'archive des situations candidates (au lieu de la distinction par saisons) et on utilise une taille d'échantillonnage égale à $M = 20$ (au lieu de $M = 24$). La raison est que l'étude de la qualité présentée à la sous-section 5.2.3 a été effectuée *a posteriori*. Néanmoins, la combinaison des hyperparamètres considérée pour le cas d'étude est acceptable car la différence de $\overline{\text{CRPS}}$ avec la combinaison optimale obtenue à la sous-section 5.2.3 est faible comparée aux différences entre toutes les combinaisons (voir Figure 5.10).

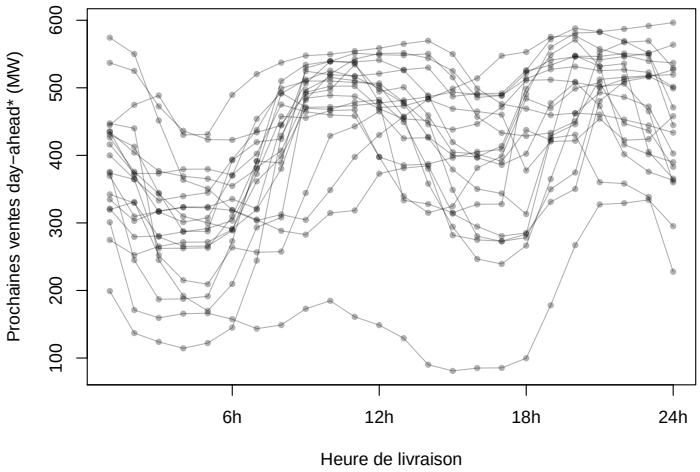
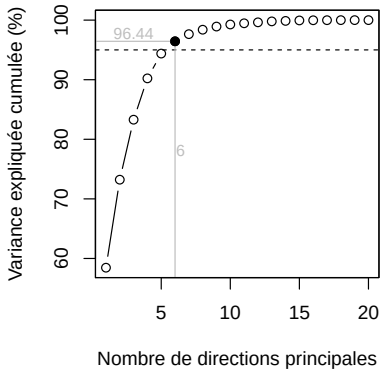


FIGURE 5.11 – Échantillonnage des prochaines ventes *day-ahead* pour le cas d'étude. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les puissances.

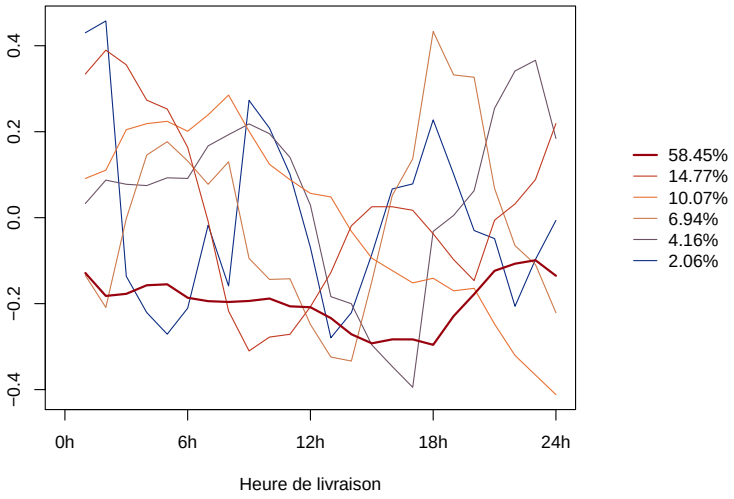
L'échantillon $(b_1, \ldots, b_M)$ des prochaines ventes *day-ahead* ainsi obtenu pour le cas d'étude est présenté à la figure 5.11. On observe que les valeurs échantillonnées ont globalement la même forme (pics autour de 9 h et 21 h comme les prix *day-ahead*) car les ventes *day-ahead* effectivement réalisées aux situations analogues ont été principalement déterminées à partir des prix *day-ahead* (placement de la production aux heures les plus chères, voir Sous-section 2.3.2). On observe également qu'une courbe de ventes *day-ahead* se distingue des autres car ses valeurs et ses variations aux heures de pointes sont plus faibles.

Comme les prochaines ventes *day-ahead* sont une donnée d'entrée fonctionnelle de la boîte noire de la figure 5.1, la réduction de la dimension de la section 4.2 du chapitre précédent peut s'appliquer. La réduction par analyse en composantes principales est effectuée avec le seuil de $v_{\min} = 0{,}95$ sur la variance expliquée. Cela permet de retenir que les $K = 6$ premières directions principales sur les 20 (voir Figure 5.12a), expliquant 96,44% de la variance totale. Les allures des directions principales retenues (voir Figure 5.12b) représentent les corrélations temporelles des prochaines ventes *day-ahead*. Par exemple, la première direction principale (ayant la variance expliquée la plus élevée) montre qu'une baisse des puissances vendues en milieu d'après-midi entraîne une augmentation des puissances vendues en soirée. Cela est une conséquence de la gestion de la production des aménagements hydroélectriques du Rhône (l'eau est globalement retenue en milieu d'après-midi pour placer la production en soirée où les prix sont plus élevés, voir Sous-section 2.3.2). Cette observation est aussi une conséquence du fait que toute la production est supposée être vendue au marché *day-ahead* (voir Sous-section 2.5.1). Enfin, on observe sur la figure 5.12c que les erreurs dues à la réduction (3,56%) sont visuellement acceptables pour le cas d'étude.

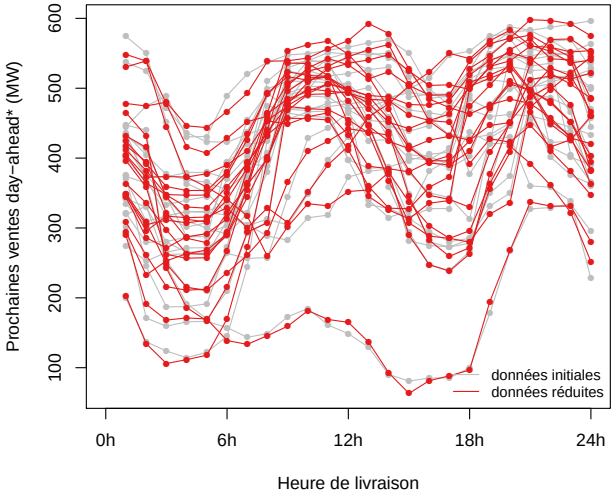
On note $\hat{\mathcal{B}}_0 = \{b_1, \ldots, b_M\} \subset \mathbb{R}^{24}$ l'approximation de l'ensemble $\mathcal{B}_0 \subset \mathbb{R}^{24}$. On rappelle que le critère de compatibilité n'est pas forcément respecté car l'ensemble $\hat{\mathcal{B}}_0$ n'est pas forcément inclus dans $\mathcal{B}_0$. Les éléments $b_m \in \mathbb{R}^{24}$, pour $m \in \{1, \ldots, M\}$, sont considérés comme équiprobables. En particulier, nous n'appliquons pas de poids aux prochaines ventes *day-ahead* en fonction des similarités entre la situation courante et les situations analogues dont elles sont originaires.



(a) Variance expliquée



(b) Directions principales retenues (le pourcentage indique la variance expliquée)



(c) Reconstitution après réduction

FIGURE 5.12 – Réduction de la dimension de l'échantillon des prochaines ventes *day-ahead* par analyse en composantes principales au seuil de 95% pour le cas d'étude. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les puissances.

Pour le cas d'étude, une analyse du respect des contraintes MILP montre même que :

$$\hat{\mathcal{B}}_0 \not\subset \bigcup_{0\leq n\leq N} b(\mathcal{X}(\xi_n)).$$

Cette observation signifie que les courbes de la figure 5.11 sont trop éloignées de ce qu'il est réellement possible de produire pour la situation étudiée, de sorte que la contrainte sur les écarts de l'équation (3.15) ne peut pas être respectée en pratique. Cette observation est bloquante pour la suite de la méthodologie car elle empêche les évaluations de la fonction Q sur les éléments de $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$ (voir Équation (5.2)) : la fonction Q n'est pas à valeurs réelles sur certains éléments de $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$. En conséquence, pour pouvoir continuer l'étude, nous supprimons à partir de maintenant les contraintes encadrant les écarts (voir Équation (3.15)) qui interviennent dans le problème d'optimisation (5.2), uniquement pour la partie de la fenêtre temporelle correspondant au jour de livraison des prochaines ventes *day-ahead* (le lendemain pour le cas d'étude). Ces contraintes sont néanmoins conservées sur le reste de la fenêtre temporelle (jour courant et sur-lendemain pour le cas d'étude). Ainsi, pour tout $(b_m, \xi_n) \in \hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$, $Q(b_m, \xi_n) \in \mathbb{R}$, i.e. $b_m \in b(\mathcal{X}(\xi_n))$. Ce relâchement des contraintes n'a en revanche pas été effectué dans la sous-section 3.3.5 dans laquelle la fonction objectif du problème d'optimisation probabiliste (5.1) a été analysée avec la solution déterministe. Il convient de noter que ce relâchement des contraintes retire de l'importance à la contrainte $\{b(\mathbf{x}_n) = b_m,\ \mathbf{x}_n \in \mathcal{X}(\xi_n)\}$, pour $(b_m, \xi_n) \in \hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$, dans la définition du problème d'optimisation (5.2) qui défini la fonction Q. Or, cette contrainte définit la dynamique essentielle entre les deux niveaux du problème d'optimisation probabiliste (5.1). Ainsi, si cette simplification s'avérait indispensable pour d'autres cas d'étude, alors la méthodologie de résolution numérique par métamodèle que nous proposons perdraient en intérêt pour résoudre le problème d'optimisation (5.1).

5.3 Plan d'expérience

5.3.1 Problématique

On rappelle que l'on cherche à construire un métamodèle $\hat{Q}$ de la fonction Q sur l'ensemble des données d'entrée $\mathcal{B}_0 \times \hat{\Xi}(\Omega)$ dans une phase de pré-traitement à la résolution du problème d'optimisation approché (5.5). L'ensemble $\mathcal{B}_0 \subset \mathbb{R}^{24}$ est le domaine de la donnée d'entrée associée aux prochaines ventes *day-ahead* possibles et $\hat{\Xi}(\Omega) = \{\xi_1, \ldots, \xi_N\}$ celui associé aux scénarios probabilistes. Pour le cas d'étude, on rappelle que $N = N_a N_\pi = 2\ 500$ car on considère toutes les combinaisons possibles entre les $N_a = 50$ membres de la prévision d'ensemble des apports en eau et les $N_\pi = 50$ membres de la prévision d'ensemble des prix (voir Sous-section 3.3.1).

Comme le métamodèle $\hat{Q}$ est construit dans une phase de pré-traitement, il est nécessaire de construire un jeu de données pour la situation considérée en amont de l'optimisation. La difficulté est que l'ensemble $\mathcal{B}_0$ n'est pas accessible en pratique, donc il est remplacé par l'ensemble fini $\hat{\mathcal{B}}_0 = \{b_1, \ldots, b_M\}$ issu de l'échantillonnage par analogie des prochaines ventes *day-ahead* effectué à la section 5.2 précédente. On rappelle que $M = 20$ pour le cas d'étude.

Le nouvel ensemble des données d'entrée que l'on considère est donc $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega) = \{b_1, \ldots, b_M\} \times \{\xi_1, \ldots, \xi_N\}$. Comme cet ensemble est fini, il serait possible de calculer les valeurs $Q(b_m, \xi_n)$, pour $(m, n) \in \{1, \ldots, M\} \times \{1, \ldots, N\}$ (variables à expliquer). La difficulté est que la taille MN de l'ensemble fini $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$ est trop élevée pour évaluer la fonction Q à chaque point (pour le cas d'étude, $MN = 20 \times 2\ 500 = 50\ 000$). En effet, le temps de calcul disponible en opérationnel est limité (temps de calcul total inférieur à 1 h) alors qu'une évaluation de la fonction Q nécessite la résolution d'un problème MILP de grande dimension.

C'est pourquoi nous cherchons dans cette section à construire un plan d'expérience, de manière à évaluer la fonction Q sur seulement quelques points judicieusement sélectionnés $(b_{m_d}, \xi_{n_d}) \in \hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$, pour $d \in \{1, \dots, D\}$, avec $D \ll MN$, $m_d \in \{1, \dots, M\}$ et $n_d \in \{1, \dots, N\}$.

Une première approche possible pour sélectionner les points (b_{m_d}, ξ_{n_d}), pour $d \in \{1, \dots, D\}$, serait d'appliquer une réduction de chacune des données d'entrée discrétisées et de former toutes les combinaisons des données d'entrée réduites. Les résultats de la section 4.3 du chapitre précédent pourraient alors être utilisés pour les scénarios et adaptés pour les prochaines ventes *day-ahead* possibles. Autrement dit, l'ensemble des scénarios probabilistes $\hat{\Xi}(\Omega)$ serait remplacé par le sous-ensemble $\hat{\Xi}'(\Omega) = \{\xi'_1, \dots, \xi'_{N'}\} \subset \hat{\Xi}(\Omega)$ des scénarios réduits, avec $N' \ll N$, et l'ensemble $\hat{\mathcal{B}}_0 = \{b_1, \dots, b_M\}$ des prochaines ventes *day-ahead* possibles serait remplacé par un sous-ensemble $\hat{\mathcal{B}}'_0 = \{b'_1, \dots, b'_{M'}\} \subset \hat{\mathcal{B}}_0$, avec $M' \ll M$. Les points sélectionnés seraient alors définis par $D = M'N' \ll MN$ et $\{(b_{m_d}, \xi_{n_d}),\ 1 \leq d \leq D\} = \hat{\mathcal{B}}'_0 \times \hat{\Xi}'(\Omega)$. Ces points sélectionnés ne seraient en revanche pas équiprobables : leurs probabilités seraient données par les redistributions optimales de Kantorovich de la section 4.3.

Cette approche par réduction permet de quantifier la distribution des données d'entrée : les points sélectionnés et leurs probabilités permettent d'approcher au mieux la distribution initiale au sens de la distance de Kantorovich. Néanmoins, pour définir un jeu de données, il est souvent d'usage de construire le plan d'expérience en considérant des points équiprobables et qui respectent des critères de répartition dans le domaine des données d'entrée. Le caractère équiprobable des points est important car les algorithmes d'apprentissage supervisé sont généralement définis lorsque les observations du jeu de données ont toutes le même poids. Les critères de répartition sont importants car on cherche à construire le métamodèle $\hat{Q}$ de manière « robuste » pour qu'il soit une bonne approximation de la fonction Q sur tout le domaine des données d'entrée avec un petit nombre de points. Autrement dit, les points sélectionnés doivent être bien répartis dans l'ensemble $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$ pour assurer que les valeurs numériques $Q(b_{m_d}, \xi_{n_d}) \in \mathbb{R}$, pour $d \in \{1, \dots, D\}$, donnent suffisamment d'information sur la structure globale de la fonction Q sur $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$. Par ailleurs, les algorithmes de réduction des scénarios semblent moyennement efficaces dans le cadre de notre travail (voir Section 4.3). C'est pourquoi nous proposons dans la sous-section 5.3.2 suivante de construire le plan d'expérience à l'aide d'un échantillonnage par hypercube latin appliqué à l'ensemble $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega) = \{b_1, \dots, b_M\} \times \{\xi_1, \dots, \xi_N\}$ des données d'entrée. Il s'agit d'une méthode très utilisée dans la littérature pour construire des plans d'expériences.

5.3.2 Échantillonnage par hypercube latin

L'ensemble des données d'entrée discrétisées est

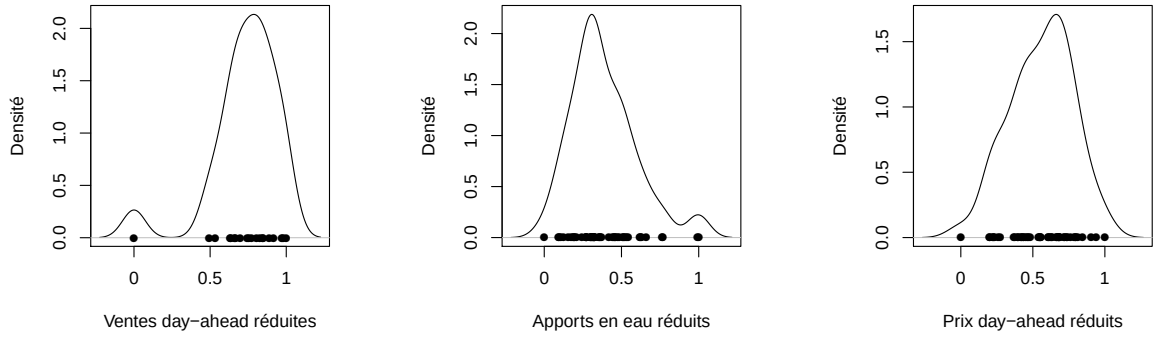
$$\begin{aligned}\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega) &= \{(b_m, \xi_n),\ (m,n) \in \{1, \dots, M\} \times \{1, \dots, N\}\} \\ &= \{(b_m, (a_n, \pi_n, \phi_n)),\ (m,n) \in \{1, \dots, M\} \times \{1, \dots, N\}\}.\end{aligned}$$

En dissociant les composantes des scénarios, on a donc quatre données d'entrée : les engagements *day-ahead*, les apports en eau, les prix de l'électricité et les productions variables, comme illustré par la schéma boîte noire de la figure 5.1. Ces données d'entrée sont fortement corrélées car elles ont une structure fonctionnelle (voir Figure 3.5) avec possiblement des corrélations entre elles (voir Sous-section 3.3.1). L'échantillonnage de telles données d'entrée étant difficile, nous faisons les hypothèses simplificatrices suivantes dans la suite de cette sous-section :

— les prochaines ventes *day-ahead* sont supposées indépendantes des scénarios,

— les prochaines ventes *day-ahead*, les apports en eau, les prix *day-ahead* et les productions variables sont chacun réduits à un scalaire pour supprimer leur structure fonctionnelle.

Nous rappelons également que les apports en eau, les prix *day-ahead* et les productions variables qui composent les scénarios sont supposés indépendants entre eux (voir Sous-section 2.4.4).



(a) Prochaines ventes *day-ahead* (b) Apports en eau (c) Prix *day-ahead*

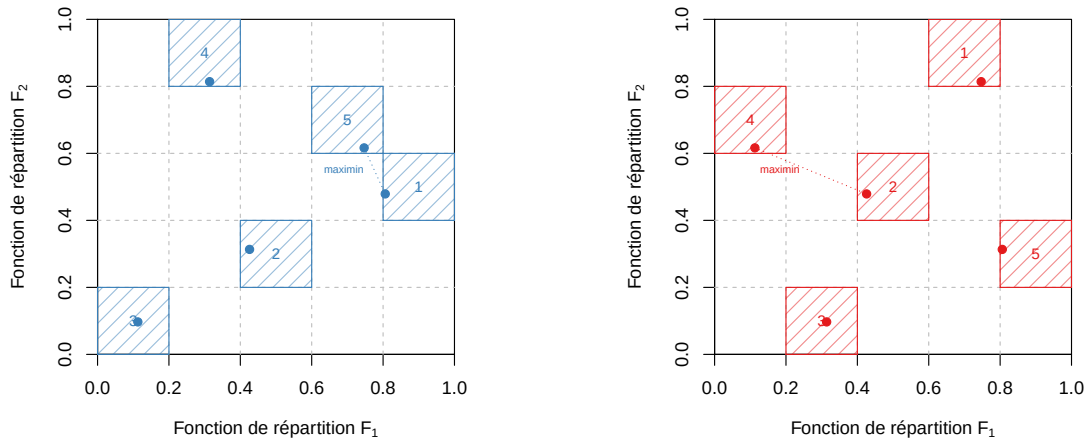
FIGURE 5.13 – Exemple de la représentation scalaire dans $[0,1]$ des données d'entrée pour le cas d'étude.

Après une mise à l'échelle (facultative), ces hypothèses conduisent à quatre données d'entrée indépendantes dans $[0,1]$, dont nous disposons MN observations. Pour le cas d'étude, la réduction à un scalaire des données d'entrée fonctionnelles est effectuée en utilisant les valeurs moyennes, comme au §4.2.2.1 (voir Figure 5.13). On rappelle également que les productions variables ne sont pas considérées pour le cas d'étude, enlevant une donnée d'entrée. Néanmoins, nous continuons de présenter la méthodologie avec les productions variables.

Nous proposons de construire le plan d'expérience par échantillonnage d'un hypercube latin (LHS, pour *Latin Hypercube Sampling*) de quatre dimensions avec D points et optimal par rapport au critère maximin [Santner *et al.*, 2003].

Un plan hypercube latin de taille D est par définition un plan d'expérience à D points où chaque donnée d'entrée est représentée par D quantiles équiprobables. Pour toutes les données d'entrée, chaque point du plan hypercube latin est donc associé à un quantile différent. Les D quantiles équiprobables de chaque donnée d'entrée forment une quantification de la distribution de la donnée d'entrée considérée. Les projections des points du plan hypercube latin sur chaque donnée d'entrée permettent alors d'approcher les distributions marginales des données d'entrée. Par exemple, si on considère que les données d'entrée suivent des lois uniformes, alors les projections des points du plan hypercube latin sur chaque donnée d'entrée forment une subdivision régulière du domaine de la donnée d'entrée considérée. Le plan hypercube latin permet donc une bonne répartition des points par rapport aux données d'entrée prises individuellement. Des algorithmes efficaces permettent de générer aléatoirement un plan hypercube latin, dont les étapes générales pour notre problème sont les suivantes.

1. On considère le cube unité $[0,1]^4$ de dimension 4 (une dimension par donnée d'entrée indépendante) où les 4 directions représentent les images des fonctions de répartition des 4 données d'entrée indépendantes.
2. On applique une subdivision régulière en D segments aux 4 directions du cube unité $[0,1]^4$. Pour tout $j \in \{1,2,3,4\}$, la j-ième direction est donc représentée par les D segments $I_{dj} = [(d-1)/D, d/D]$, pour $d \in \{1,\dots,D\}$. Comme les directions correspondent aux images des fonctions de répartition des données d'entrée, cette subdivision régulière permet, par inversion, d'obtenir une quantification des distributions marginales des données d'entrée.
3. On tire uniformément au hasard D points $\mathbf{y}_d = (y_{d1}, y_{d2}, y_{d3}, y_{d4}) \in I_{\sigma_1(d)1} \times I_{\sigma_2(d)2} \times I_{\sigma_3(d)3} \times I_{\sigma_4(d)4}$, pour $d \in \{1,\dots,D\}$, où σ_j, pour $j \in \{1,2,3,4\}$, sont des permutations aléatoires de $\{1,\dots,D\}$. On pose $\mathbb{Y} = (\mathbf{y}_d)_{1\leq d\leq D} \in ([0,1]^4)^D$. La figure 5.14a illustre les trois premières étapes.



(a) Aléatoire

(b) Optimisé selon le critère maximin

FIGURE 5.14 – Exemple de deux plans hypercubes latins en dimension 2 avec $D = 5$ points. Celui de gauche est obtenu aléatoirement et celui de droite par optimisation du critère maximin. Les carrés en couleurs sont ceux obtenus aléatoirement par les permutations. Les points sont ceux tirés aléatoirement dans chacun des 5 carrés. Le segment en pointillés représente la distance minimale entre deux points du plan (critère maximin).

4. On considère le plan d'expérience $\mathbb{X}' = (\mathbf{x}'_d)_{1\leq d\leq D} \in ([0,1]^4)^D$ où, pour tout $d \in \{1,\dots,D\}$, $\mathbf{x}'_d = (F_1^{-1}(y_{d1}), F_2^{-1}(y_{d2}), F_3^{-1}(y_{d3}), F_4^{-1}(y_{d4})) \in [0,1]^4$ (F_j^{-1}, pour $j \in \{1,2,3,4\}$, désigne la fonction quantile de la j-ième donnée d'entrée). Le plan d'expérience $\mathbb{X}'$ est alors défini sur les données d'entrée réduites.
5. On considère le plan d'expérience $\mathbb{X} = ((b_{m_d}, \xi_{n_d}))_{1\leq d\leq D} \in (\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega))^D$ obtenu à partir du plan d'expérience $\mathbb{X}'$ mais en considérant les données d'entrée fonctionnelles associées (pour le cas d'étude, il y a une bijection entre les données d'entrée fonctionnelles et les données réduites dans $[0,1]^4$).

La génération aléatoire du plan hypercube latin est due au tirage des permutations et des points à l'étape 3. Pour plus de détails sur les algorithmes pour générer des plans hypercubes latin, voir par exemple [McKay *et al.*, 2000 ; Stein, 1987].

Bien que les points du plan hypercube latin soient bien répartis selon chaque donnée d'entrée prise individuellement, ils ne sont pas nécessairement bien répartis dans le domaine joint des données d'entrée. Pour assurer une bonne répartition des points du plan d'expérience dans le domaine joint des données d'entrée, une solution est de chercher un plan hypercube latin qui maximise la distance minimale entre deux points distincts du plan d'expérience, comme illustré à la figure 5.14b. On dit qu'un tel plan hypercube latin est optimal par rapport au critère maximin [Morris et Mitchell, 1995]. Pour cela, les étapes 1 à 3 précédentes sont itérées par recuit simulé [4] pour maximiser le critère maximin

$$\forall \mathbb{Y} = (\mathbf{y}_d)_{1\leq d\leq D} \in ([0,1]^4)^D, \quad \Phi_{Mm}(\mathbb{Y}) = \min_{1\leq d<d'\leq D} \|\mathbf{y}_d - \mathbf{y}_{d'}\|_2,$$

où $\|\ \|_2$ désigne la norme euclidienne dans $\mathbb{R}^4$. Pour plus de détails sur l'optimisation du critère maximin, voir par exemple [Pronzato et Müller, 2012 ; Damblin *et al.*, 2013]. Pour le cas d'étude, le plan hypercube latin uniforme $\mathbb{Y}^*$ optimal pour le critère maximin est présenté à la figure 5.15.

4. Le recuit simulé est une méthode d'optimisation type stochastique pour résoudre un problème d'optimisation général, voir Sous-section 3.1.1.

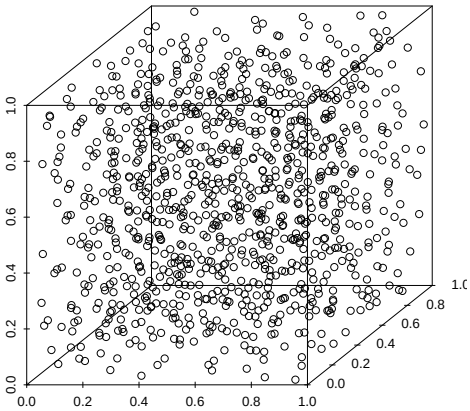


FIGURE 5.15 – Représentation de l'hypercube latin $\mathbb{Y}^*$ pour le cas d'étude avec $D = 1\,000$ et optimisé selon le critère maximin (en 3 dimensions car les productions variables ne sont pas considérées dans le cas d'étude).

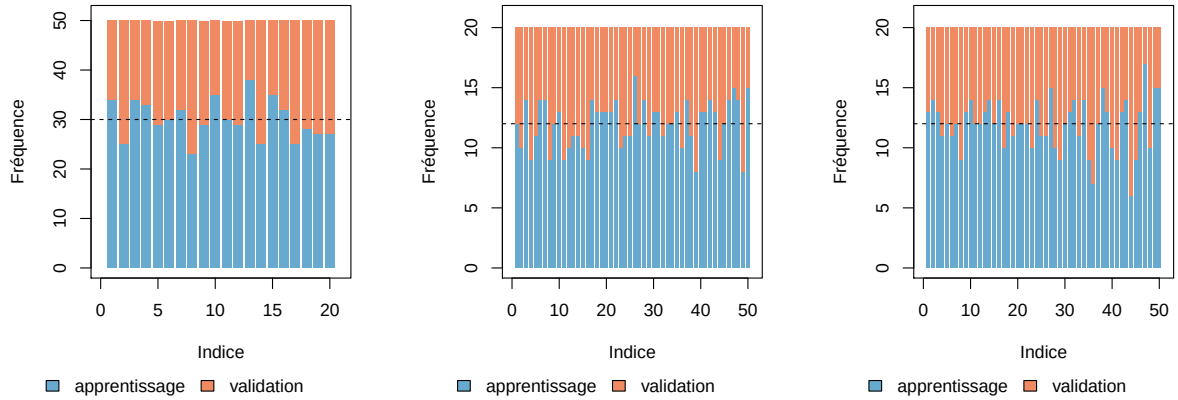
Les étapes 4 et 5 ci-dessus sont ensuite appliquées avec le plan hypercube latin uniforme $\mathbb{Y}^*$ optimal pour le critère maximin.

On note $(b_{m_d}, \xi_{n_d}) \in \hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$, pour $d \in \{1, \ldots, D\}$, les points sélectionnés avec leurs structures fonctionnelles. Comme les données d'entrée réduites sont supposées indépendantes, les points sélectionnés sont bien distribués dans le domaine joint $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$ des données d'entrée fonctionnelles.

La taille D du plan d'expérience est arbitraire. En pratique, elle est choisie empiriquement en fonction de la puissance de calcul à disposition et de la complexité (degrés de liberté) du métamodèle. Pour le cas d'étude, on choisit $D = 1\,000$, une taille très grande par rapport à la complexité du métamodèle qui sera utilisé dans la sous-section 5.4.1 de ce chapitre. Ce choix n'est adapté que pour le travail de la thèse (étude de la méthodologie) où le temps de calcul n'est pas autant limité qu'en opérationnel. Comme $D = 1\,000$ est un multiple de $M = 20$ (respectivement $N_a = N_\pi = 50$) et que les prochaines ventes *day-ahead* et les scénarios sont équiprobables, les éléments de $\hat{\mathcal{B}}_0$ (respectivement de $\hat{\Xi}(\Omega)$) sont tous utilisés le même nombre de fois [Santner *et al.*, 2003]. En outre, comme les prochaines ventes *day-ahead* et les scénarios sont supposés indépendants, les points du plan d'expérience sont équiprobables.

5.3.3 Jeu de données

Le jeu de données est obtenu en évaluant la fonction Q à chaque point du plan d'expérience. Pour tout $d \in \{1, \ldots, D\}$, on note $Q_d = Q(b_{m_d}, \xi_{n_d}) \in \mathbb{R}$. Cela nécessite de résoudre indépendamment D problèmes MILP de grande dimension (voir Équation (5.2)). Le temps de calcul de cette étape dépend donc de la taille D du jeu de données et la puissance de calcul à disposition (accès au calcul distribué notamment), sachant qu'une évaluation de la fonction Q nécessite un temps de calcul de l'ordre de grandeur de celui de la résolution du problème d'optimisation déterministe (3.19) (qui est non négligeable par rapport au temps limite de 1 h pour rester compatible aux exigences du processus opérationnel). On rappelle que les contraintes qui encadrent les écarts (voir Équation (3.15)) sont supprimées pour la période de la fenêtre temporelle correspondant au lendemain de manière à assurer l'existence des solutions du problème MILP (5.2)



(a) Prochaines ventes *day-ahead* (b) Apports en eau (c) Prix *day-ahead*

FIGURE 5.16 – Répartition du jeu de données entre l'ensemble d'apprentissage et l'ensemble de validation pour le cas d'étude.

avec l'échantillon $(b_1, \ldots, b_M)$ des prochaines ventes *day-ahead* (voir Sous-section 5.2.4). C'est la raison pour laquelle les valeurs de la fonction Q sur le plan d'expérience sont réelles.

Le jeu de données est divisé en un ensemble d'apprentissage avec les indices $\mathcal{D}_c \subset \{1, \ldots, D\}$ et un ensemble de validation avec les indices $\mathcal{D}_v = \{1, \ldots, D\} \setminus \mathcal{D}_c$. La taille de l'ensemble d'apprentissage est arbitraire. Pour le cas d'étude, le ratio apprentissage/validation est de 60% et la répartition entre l'apprentissage et la validation est effectuée aléatoirement de manière uniforme.

La figure 5.16 illustre la répartition du jeu de données entre l'ensemble d'apprentissage et l'ensemble de validation. Comme la répartition est effectuée aléatoirement, les histogrammes ne sont pas plats : les valeurs des données d'entrée discrétisées ne sont pas représentées le même nombre de fois dans l'ensemble d'apprentissage (donc pas non plus dans l'ensemble de validation). Pour obtenir une répartition uniforme, une approche possible serait d'effectuer un nouvel échantillonnage par hypercube latin avec $600 = 0{,}6 \times 1\,000 = 0{,}6 \times D$ points (600 est un multiple de 20 et 50) et optimal selon le critère maximin (voir Sous-section 5.3.2). Nous n'utilisons pas cette approche pour les raisons suivantes :

— nous ne souhaitons pas imposer dans la méthodologie que la taille de l'ensemble d'apprentissage soit un multiple du nombre de prochaines ventes *day-ahead* M et du nombre de membres N_a (respectivement N_π) de la prévision d'ensemble des apports en eau (respectivement des prix) qui composent les scénarios,

— l'ensemble de validation n'est à utiliser que dans le cadre des études de la méthodologie et non pas dans le cadre d'une utilisation en opérationnel.

Néanmoins, les distributions de la variable de sortie Q sur les ensembles d'apprentissage et de validation semblent proches de celle sur tout le jeu de données (voir Figure 5.17). Pour aller plus loin, il serait possible d'effectuer des tests statistiques pour confirmer cette observation de manière quantitative.

On observe sur la figure 5.17 que les valeurs de la fonction Q sur le jeu de données sont en moyenne 1,32% inférieures au chiffre d'affaires $f(\mathbf{x}_0^*, \xi_0)$ de la solution déterministe $\mathbf{x}_0^* \in \mathcal{X}(\xi_0)$ de la sous-section 3.2.6, ce qui est inférieur au MIP-gap relatif seuil de 2%. On peut donc considérer que la moyenne des chiffres d'affaires sur le jeu de données correspond au chiffre d'affaires de la solution déterministe. On rappelle que les résultats de la sous-section 3.3.5 (étude de la fonction objectif du problème d'optimisation probabiliste (5.1) avec la solution déterministe) montrent que cette observation est valable pour les chiffres d'affaires obtenus à partir des prochaines ventes

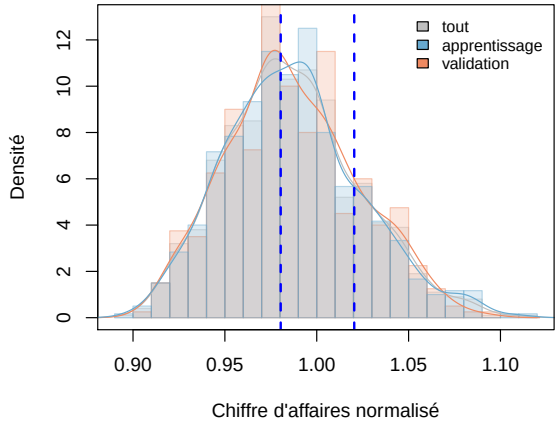


FIGURE 5.17 – Distribution de la variable de sortie Q sur l'ensemble d'apprentissage et l'ensemble de validation pour le cas d'étude. Les chiffres d'affaires sont normalisés par la valeur de référence $f(\mathbf{x}_0^*, \xi_0)$, le chiffre d'affaires de la solution déterministe $\mathbf{x}_0^* \in \mathcal{X}(\xi_0)$ calculé avec le scénario déterministe ξ_0. Les lignes verticales bleues en tirets représentent la zone de tolérance autour de la valeur de référence $f(\mathbf{x}_0^*, \xi_0)$ avec un MIP-gap relatif seuil de 2%.

Données d'entrée	Q
Ventes *day-ahead*	0,02
Apports	0,67
Prix	0,31
Interactions	0,00

TABLEAU 5.1 – Indices de Sobol du premier ordre des données d'entrée sur le jeu de données pour le cas d'étude.

day-ahead $b_0^* = b(\mathbf{x}_0^*) \in \mathcal{B}_0$ associées à la solution déterministe $\mathbf{x}_0^* \in \mathcal{X}(\xi_0)$ en faisant varier les scénarios $(\xi_1, \ldots, \xi_N)$. Sur la figure 5.17, les distributions portent à la fois sur les scénarios $(\xi_1, \ldots, \xi_N)$ et sur l'échantillon $(b_1, \ldots, b_M)$ des prochaines ventes *day-ahead*. En revanche, on observe sur la figure 5.17 que l'écart-type de la fonction Q sur le plan d'expérience est supérieur aux erreurs d'approximation dues au MIP-gap relatif seuil de 2%. La fonction Q prend donc des valeurs suffisamment distinctes sur le plan d'expérience : au moins une partie des données d'entrée explique la fonction Q.

Plus généralement, on observe que les distributions de la figure 5.17 sont similaires à celle obtenue à la sous-section 3.3.5 du chapitre précédent consacrée à l'étude de la fonction Q pour les prochaines ventes *day-ahead* $b_0^* = b(\mathbf{x}_0^*) \in \mathcal{B}_0$ de la solution déterministe $\mathbf{x}_0^* \in \mathcal{X}(\xi_0)$. Pourtant, la fonction Q de ce chapitre n'est pas comparable avec la fonction Q de la sous-section 3.3.5 en raison du relâchement des contraintes qui encadrent les écarts appliqué après la discrétisation des prochaines ventes *day-ahead* (voir Sous-section 5.2.4). La figure 5.17 montre donc que, pour le cas d'étude, le relâchement des contraintes qui encadrent les écarts et le changement de ventes *day-ahead* semblent avoir peu d'impact sur les chiffres d'affaires.

Le tableau 5.1 présente les résultats de l'analyse de sensibilité portée sur le jeu de donnée, selon une démarche analogue à celles effectuées aux sous-sections 3.3.2 et 3.3.5 (calcul des indices de Sobol). Ces résultats peuvent se visualiser en comparant la taille des boîtes à moustaches de la figure 5.18 entre les trois données d'entrée du cas d'étude. Les ordres de grandeur des indices de Sobol des apports en eau et des prix sont similaires à ceux observés en fixant les prochaines ventes *day-ahead* par la solution déterministe (voir Sous-section 3.3.5). En outre, on

observe que la sensibilité des prochaines ventes *day-ahead* sur la fonction Q est très faible pour le cas d'étude par rapport à celles des apports en eau et des prix *day-ahead*. Néanmoins, dans le problème d'optimisation probabiliste (5.1), la fonction Q intervient par sa moyenne portée sur les scénarios (voir courbe en rouge sur la figure 5.18a) et non pas ses valeurs à chaque scénario. Ainsi, même si la fonction Q est peu sensible aux prochaines ventes *day-ahead*, les indices de Sobol ne permettent pas de donner des informations sur les variations de la fonction $\mathcal{E} : b_0 \mapsto \mathbb{E}[Q(b_0, \hat{\Xi})]$ qui intervient dans le problème d'optimisation probabiliste (5.1). Néanmoins, on observe sur la figure 5.18a que la courbe rouge a globalement des valeurs comprises dans l'intervalle [0,98, 1,02], ce qui signifie que la fonction $\mathcal{E}$ est globalement constante et ses valeurs sont égales, au MIP-gap relatif seuil de 2% près, au chiffre d'affaires $f(\mathbf{x}_0^*, \xi_0)$ de la solution déterministe $\mathbf{x}_0^*$. L'observation faite lorsque nous avons étudié la fonction objectif probabiliste avec la solution déterministe à la fin de la sous-section 3.3.5 est donc aussi valable pour l'échantillon $(b_1, \ldots, b_M)$ des prochaines ventes *day-ahead* et le relâchement des contraintes qui encadrent les écarts. La modification de la fonction objectif du problème d'optimisation qui permet de passer de l'univers déterministe à l'univers probabiliste est alors trop faible par rapport au MIP-gap relatif seuil de 2% pour le cas d'étude. Autrement dit, le passage de l'univers déterministe à l'univers probabiliste n'est pas pertinent pour le cas d'étude. Malgré tout, la méthodologie que nous proposons peut s'avérer pertinente pour d'autres cas d'étude, notamment sans les simplifications faites pour le cas d'étude (voir Sous-section 2.5.2). C'est pourquoi nous poursuivons la présentation de la méthodologie en l'illustrant grâce au cas d'étude, même si les résultats présentés doivent être mis en perspective avec ces observations.

5.4 Régression

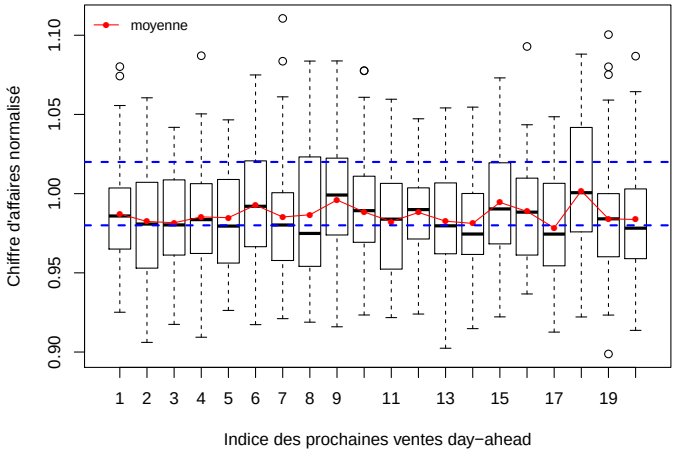
Le jeu de données obtenu avec les démarches décrites aux sections 5.2 et 5.3 peut être utilisé pour l'apprentissage du métamodèle $\hat{Q}$ dans une phase de pré-traitement à la résolution l'optimisation approché (5.5). C'est l'objet de cette section (étapes (d) et (e) de la figure 5.2). Comme la donnée de sortie de la fonction Q que l'on cherche à remplacer par le métamodèle $\hat{Q}$ est réelle, l'apprentissage supervisé du métamodèle $\hat{Q}$ consiste en une régression (par opposition à la classification pour des variables de sortie quantitatives, i.e. représentant des catégories ou des valeurs discrètes).

5.4.1 Choix de la méthode d'apprentissage

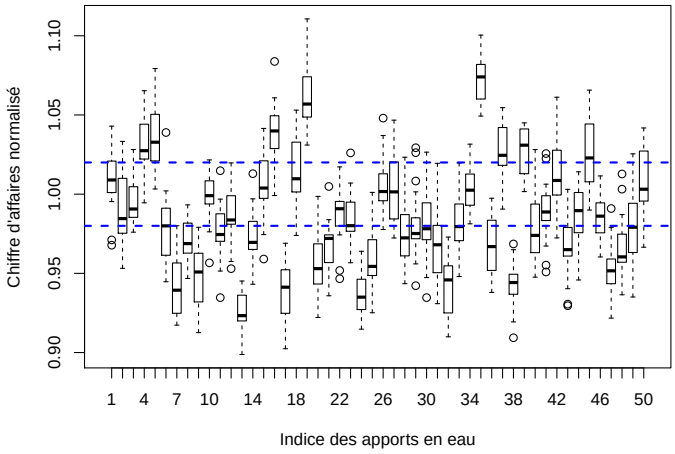
Il existe de nombreuses méthodes d'apprentissage par régression. On distingue souvent deux catégories de méthodes de régression : celles qui sont paramétriques et celles qui ne sont pas paramétriques.

— Les méthodes de régression paramétriques consistent à choisir une forme particulière *a priori* de la relation entre les entrées et les sorties du métamodèle. Dans ce cas, la phase d'apprentissage permet de déterminer les valeurs des paramètres associées à la forme choisie de manière à minimiser les erreurs de prédiction sur l'ensemble d'apprentissage. La régression linéaire (et ses variantes régularisées, comme la régression LASSO par exemple) et la régression par un réseau de neurones sont des exemples de méthodes de régression paramétriques.

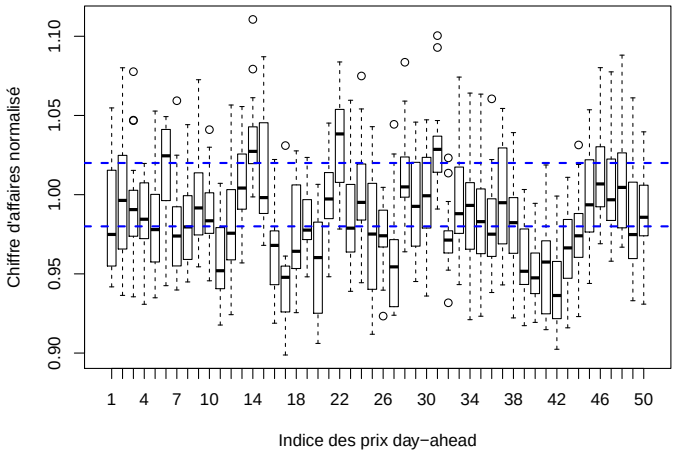
— À l'inverse, avec les méthodes de régression non paramétriques, la forme de la relation entre les entrées et les sorties du métamodèle n'est pas déterminée *a priori* mais durant la phase d'apprentissage (minimisation des erreurs de prédiction sur l'ensemble d'apprentissage). La régression par noyau, la régression par processus gaussien (krigeage) [Rasmussen, 2003], la régression par arbre de décision et la régression par splines (MARS) sont des exemples de méthodes de régression non paramétriques.



(a) Prochaines ventes *day-ahead*



(b) Apports en eau



(c) Prix *day-ahead*

FIGURE 5.18 – Boîtes à moustaches de la fonction Q normalisée par le chiffre d'affaires $f(\mathbf{x}_0^*, \xi_0)$ de la solution déterministe $\mathbf{x}_0^*$ pour le cas d'étude en fixant une donnée d'entrée. Les lignes horizontales bleues en tirets représentent la zone de tolérance autour de la valeur de référence $f(\mathbf{x}_0^*, \xi_0)$ avec un MIP-gap relatif seuil de 2%.

Pour plus de détails sur ces approches, voir par exemple [Rasmussen, 2003] pour la régression par processus gaussien et [Friedman *et al.*, 2001] pour les autres approches.

Pour notre problème, la fonction Q que l'on souhaite remplacer par le métamodèle $\hat{Q}$ est linéaire par morceaux. En effet, pour un scénario $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ fixé, $b_0 \mapsto Q(b_0, \xi)$ est la fonction objectif du problème d'optimisation MILP (5.2), donc elle est linéaire par morceaux avec un nombre fini de points de discontinuité [Haneveld et van der Vlerk, 2020 ; Haneveld *et al.*, 2019 ; Blair et Jeroslow, 1977]. Plus de détails sur la structure de ce type de fonction sont présentés dans [Ralphs et Hassanzadeh, 2014]. La fonction $b_0 \mapsto \mathbb{E}[Q(b_0, \hat{\Xi})]$ qui intervient dans le problème d'optimisation probabiliste (5.1) est donc aussi linéaire par morceaux (combinaison linéaire de fonctions linéaires par morceaux).

Dans le cadre de la thèse, nous souhaitons que le problème d'optimisation approché (5.5) puissent se résoudre avec le solveur MILP pour les deux raisons principales suivantes :

— il semble peu pertinent de rendre le problème d'optimisation approché (5.5) plus complexe qu'un problème MILP sachant que le temps de calcul du problème MILP déterministe (3.19) est déjà important,

— nous souhaitons que le problème d'optimisation probabiliste (5.1) reste une extension du problème d'optimisation déterministe (3.19) (voir Sous-section 3.3.4), les deux problèmes faisant intervenir l'ensemble admissible $\mathcal{X}(\xi_0)$ défini avec des contraintes MILP (voir Sous-section 3.2.3).

Une manière d'imposer une forme MILP au problème d'optimisation approché (5.5) est de choisir le métamodèle $\hat{Q}$ parmi les fonctions linéaires par morceaux par rapport à son premier argument (celui associé aux prochaines ventes *day-ahead* $b_0 \in \mathbb{R}^{24}$, des variables de décision du problème d'optimisation). En effet, avec un tel métamodèle $\hat{Q}$, la fonction $\hat{\mathcal{E}} : b_0 \in \mathcal{B}_0 \mapsto \mathbb{E}[\hat{Q}(b_0, \hat{\Xi})]$ est aussi linéaire par morceaux (l'espérance étant linéaire), donc le problème d'optimisation (5.5) a une formulation MILP (voir Sous-section 3.1.3), avec une dimension similaire au problème MILP (3.18) en univers déterministe. Dans ce cas, il n'est pas indispensable de construire une approximation supplémentaire $\tilde{\mathcal{E}}$ de la fonction $\hat{\mathcal{E}}$ (voir Sous-section 4.1.2) car le problème d'optimisation (5.5) peut se résoudre numériquement à partir de l'optimiseur déterministe de la sous-section 3.2.5 en modifiant l'expression de la fonction objectif. La résolution numérique ne devrait pas avoir un impact significatif sur le temps de calcul par rapport à l'optimiseur en univers déterministe (associé au problème MILP (3.19)). Au contraire, si le métamodèle $\hat{Q}$ n'était pas choisi parmi les fonctions linéaires par morceaux par rapport aux prochaines ventes *day-ahead* $b_0 \in \mathbb{R}^{24}$, alors la fonction $\hat{\mathcal{E}} : b_0 \in \mathcal{B}_0 \mapsto \mathbb{E}[\hat{Q}(b_0, \hat{\Xi})]$ ne serait pas nécessairement linéaire par morceaux, rendant difficile la résolution du problème d'optimisation (5.5). Dans ce cas, il serait préférable de construire une approximation linéaire par morceaux $\tilde{\mathcal{E}}$ de la fonction $\hat{\mathcal{E}}$ (voir Sous-section 4.1.2) et de résoudre numériquement le problème d'optimisation approché

$$\max_{\mathbf{x}_0 \in \mathcal{X}(\xi_0)} \left\{ \alpha_0 f(\mathbf{x}_0, \xi_0) + (1 - \alpha_0) \tilde{\mathcal{E}}(b(\mathbf{x}_0)) \right\}$$

avec un solveur MILP. En revanche, les erreurs d'approximation seraient alors le cumul des erreurs d'approximation du métamodèle $\hat{Q}$ et celles du métamodèle $\tilde{\mathcal{E}}$.

Même si la forme du métamodèle est imposée *a priori*, les méthodes de régression non paramétriques ne sont pas à exclure d'office pour autant si elles peuvent fournir des métamodèles linéaires par morceaux (en choisissant correctement les splines dans la régression par splines par exemple).

Dans le cadre de la thèse, en revanche, nous simplifions le problème en choisissant de construire le métamodèle $\hat{Q}$ par régression linéaire, ce qui impose que le métamodèle $\hat{Q}$ soit linéaire par rapport aux prochaines ventes *day-ahead* $b_0 \in \mathbb{R}^{24}$ mais aussi par rapport au scénario $\xi \in \mathbb{R}^{IT} \times$

$\mathbb{R}^T \times \mathbb{R}^T$. Une régression linéaire est en effet beaucoup plus facile (notamment en temps de calcul) à mettre en place qu'une régression linéaire par morceaux par rapport aux prochaines ventes *day-ahead* $b_0 \in \mathbb{R}^{24}$. Bien que cette simplification ne permette pas de représenter les différents morceaux sur lesquels la fonction $b_0 \mapsto \mathbb{E}[Q(b_0, \hat{\Xi})]$ est linéaire, ni même le comportement de la fonction Q par rapport au scénario, il est possible qu'elle soit suffisante en pratique. Si les résultats s'avéraient peu concluants avec un modèle linéaire, alors cela signifierait qu'il faut chercher le métamodèle $\hat{Q}$ parmi un ensemble plus vaste de fonctions linéaires par morceaux.

5.4.2 Choix des entrées du métamodèle linéaire

Dans le cadre de la thèse, les données d'entrée sont fonctionnelles, donc de grande dimension par rapport à la taille de l'ensemble d'apprentissage qu'il est possible de créer dans la limite de temps pour rester compatible aux exigences du processus opérationnel (résolution totale en moins de 1 h). La régression linéaire classique (estimateur des moindres carrés) n'est alors pas directement applicable sur l'ensemble d'apprentissage car la somme des dimensions des données d'entrée est plus grande que la taille de l'ensemble d'apprentissage. C'est pourquoi nous testons et comparons les trois approches d'apprentissage suivantes :

1. la régression linéaire classique [Friedman *et al.*, 2001] appliquée aux données d'entrée réduites à un scalaire (voir §4.2.2.1),
2. la régression linéaire classique appliquée aux données d'entrée réduites par analyses en composantes principales (en utilisant les résultats de la section 4.2 du chapitre précédent),
3. la régression linéaire régularisée par LASSO [Tibshirani, 1996 ; Friedman *et al.*, 2001] appliquée aux données d'entrée complètes.

Ces trois approches d'apprentissage sont effectuées sur l'ensemble d'apprentissage. Par rapport à la première, la deuxième approche permet de déterminer si la structure fonctionnelle des données d'entrée ajoute des informations supplémentaires au métamodèle ou si, au contraire, la représentation scalaire est suffisante [Betancourt *et al.*, 2020]. En outre, la deuxième approche est proche de la méthode de régression sur composantes principales (*Principal Components Regression* en anglais [Jolliffe, 1986]) à la différence qu'avec la deuxième approche, la réduction par analyse en composantes principales est effectuée indépendamment sur chaque donnée d'entrée. La régularisation LASSO de la troisième approche permet de ne sélectionner que quelques composantes des données d'entrée complètes [Tibshirani, 1996].

5.4.3 Critères de qualité intrinsèque du métamodèle

La qualité intrinsèque du métamodèle $\hat{Q}$, calibré sur l'ensemble d'apprentissage, est évaluée à partir des erreurs de prédiction sur l'ensemble de validation (donné par les indices de $\mathcal{D}_v$) selon les deux points de vue suivants :

— l'approximation de l'évaluation point par point de la fonction Q à données d'entrée fixées (voir Équation (5.3)),

— l'approximation de l'évaluation des espérances $\mathcal{E} : b_0 \in \mathcal{B}_0 \mapsto \mathbb{E}[Q(b_0, \hat{\Xi})]$ en fixant les prochaines ventes *day-ahead* (voir Équation (5.4)).

Le point de vue point par point est similaire à celui utilisé pour calibrer le métamodèle $\hat{Q}$ sur l'ensemble d'apprentissage. Il permet donc de mesurer la qualité prédictive du métamodèle $\hat{Q}$ sur l'ensemble de validation. Le point de vue par espérance est lié à l'utilisation du métamodèle dans le problème d'optimisation approché (5.5). Ce point de vue est important car c'est celui qui est réellement utile, sachant que le métamoèdle $\hat{Q}$ n'est pas calibré sur l'ensemble d'apprentissage par rapport aux espérances. Bien sûr, une bonne approximation point par point entraîne une bonne approximation des espérances.

Pour chacun de ces deux points de vue, la qualité du métamodèle est mesurée par deux scores :

— le coefficient classique Q^2, que nous appelons le *coefficient de prédictivité*[5], qui représente le coefficient de détermination (variance expliquée) sur l'ensemble de validation,

— l'erreur absolue moyenne MAE (*Mean Absolute Error*) des chiffres d'affaires normalisés par la valeur de référence $f(\mathbf{x}_0^*, \xi_0) \in]0, +\infty[$ de la solution déterministe $\mathbf{x}_0^* \in \mathcal{X}(\xi_0)$.

Le métamodèle $\hat{Q}$ est acceptable si le coefficient de prédictivité Q^2 est proche de 1 (typiquement supérieur à 0,95) et si l'erreur absolue moyenne MAE est inférieure au MIP-gap relatif seuil de 2% utilisé par le solveur MILP. Ces scores sont explicités dans la suite de cette sous-section. Il convient de noter que le coefficient de prédictivité Q^2 et l'erreur absolue moyenne MAE peuvent être évalués par validation croisée sur l'ensemble de validation [Refaeilzadeh *et al.*, 2009]. Cela serait surtout utile si la taille D du jeu de données était petite.

Pour tout $d \in \mathcal{D}$, on note $\hat{Q}_d = \hat{Q}(b_{m_d}, \xi_{n_d}) \in \mathbb{R}$ la prédiction du métamodèle $\hat{Q}$ au point $(b_{m_d}, \xi_{n_d}) \in \hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$ d'indice d du plan d'expérience. Comme le métamodèle $\hat{Q}$ est très rapide à évaluer, le temps de calcul de ces évaluations est négligeable.

5.4.3.1 Évaluation point par point

Pour le premier point de vue concernant l'évaluation point par point de la fonction Q, le coefficient de prédictivité $\mathrm{Q}^2_{\mathrm{ind}}$ est

$$\mathrm{Q}^2_{\mathrm{ind}} = 1 - \frac{\sum\limits_{d \in \mathcal{D}_v} (Q_d - \hat{Q}_d)^2}{\sum\limits_{d \in \mathcal{D}_v} (Q_d - \overline{Q})^2} \in]-\infty, 1],$$

avec

$$\overline{Q} = \frac{1}{|\mathcal{D}_v|} \sum_{d \in \mathcal{D}_v} Q_d \in \mathbb{R}.$$

Par ailleurs, l'erreur absolue moyenne $\mathrm{MAE}_{\mathrm{ind}}$ des chiffres d'affaires normalisés est

$$\mathrm{MAE}_{\mathrm{ind}} = \frac{1}{|\mathcal{D}_v|} \sum_{d \in \mathcal{D}_v} \frac{|Q_d - \hat{Q}_d|}{f(\mathbf{x}_0^*, \xi_0)} \in \mathbb{R}_+.$$

5.4.3.2 Évaluation des espérances

Pour le second point de vue concernant l'évaluation des espérances, les formules sont analogues mais en utilisant la fonction $\mathcal{E}$ au lieu de la fonction Q. La difficulté est que tous les scénarios ne sont pas représentés dans l'ensemble de validation. En revanche, comme le plan d'expérience est bien distribué par rapport aux données d'entrée et l'ensemble de validation est tiré aléatoirement de manière uniforme, on peut considérer les approximations suivantes : pour tout $d \in \mathcal{D}_v$, $\mathbb{E}[Q(b_{m_d}, \hat{\Xi})] \approx \mathcal{E}_d$ et $\mathbb{E}[\hat{Q}(b_{m_d}, \hat{\Xi})] \approx \hat{\mathcal{E}}_d$, où

$$\mathcal{E}_d = \frac{1}{|\mathcal{D}_v(d)|} \sum_{d' \in \mathcal{D}_v(d)} Q_{d'} \in \mathbb{R} \quad \text{et} \quad \hat{\mathcal{E}}_d = \frac{1}{|\mathcal{D}_v(d)|} \sum_{d' \in \mathcal{D}_v(d)} \hat{Q}_{d'} \in \mathbb{R},$$

avec $\mathcal{D}_v(d) = \{d' \in \mathcal{D}_v \,;\; m_{d'} = m_d\}$ les indices des points l'ensemble de validation associés aux mêmes prochaines ventes *day-ahead* que le point d'indice $d \in \mathcal{D}_v$.

5. En hydrologie, lorsque l'on cherche à évaluer la qualité prédictive d'un modèle hydrologique, le coefficient de prédictivité est utilisé sur un ensemble de validation de pas de temps successifs. On forme alors le critère de Nash-Sutcliffe, très utilisé en pratique [Nash et Sutcliffe, 1970].

Ainsi, le coefficient de prédictivité $\mathrm{Q}^2_{\mathrm{esp}}$ est donné par

$$\mathrm{Q}^2_{\mathrm{esp}} = 1 - \frac{\sum\limits_{d\in\mathcal{D}_v} (\mathcal{E}_d - \hat{\mathcal{E}}_d)^2}{\sum\limits_{d\in\mathcal{D}_v} (\mathcal{E}_d - \overline{\mathcal{E}})^2} \in \left]-\infty, 1\right],$$

où

$$\overline{\mathcal{E}} = \frac{1}{|\mathcal{D}_v|} \sum_{d\in\mathcal{D}_v} \mathcal{E}_d \in \mathbb{R}.$$

Par ailleurs, l'erreur absolue moyenne $\mathrm{MAE}_{\mathrm{esp}}$ des chiffres d'affaires normalisés est donnée par

$$\mathrm{MAE}_{\mathrm{esp}} = \frac{1}{|\mathcal{D}_v|} \sum_{d\in\mathcal{D}_v} \frac{|\mathcal{E}_d - \hat{\mathcal{E}}_d|}{f(\mathbf{x}_0^*, \xi_0)} \in \mathbb{R}_+.$$

5.4.4 Apprentissage du métamodèle

À l'aide des critères que nous venons de définir, nous pouvons maintenant évaluer et comparer les résultats obtenus avec les trois approches d'apprentissage de la sous-section 5.4.2. On rappelle la notation $\overline{\xi} = (\overline{a}, \overline{\pi}, \overline{\phi}) \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ du scénario moyen (voir Sous-section 4.2.1) :

$$(\overline{a}, \overline{\pi}, \overline{\phi}) = \left(\sum_{n=1}^N p_n a_n, \sum_{n=1}^N p_n \pi_n, \sum_{n=1}^N p_n \phi_n \right) = \sum_{n=1}^N p_n \xi_n = \overline{\xi} \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T.$$

On note également $\overline{b}_0 \in \mathbb{R}^{24}$ la moyenne de l'échantillon $(b_1, \ldots, b_M)$ des prochaines ventes *day-ahead*, i.e.

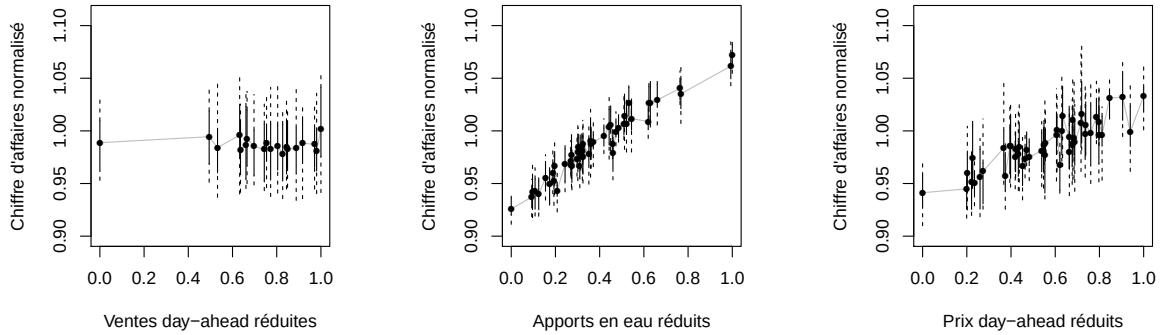
$$\overline{b}_0 = \frac{1}{M} \sum_{m=1}^M b_m \in \mathbb{R}^{24}.$$

5.4.4.1 Approche sur les données d'entrée scalaires

On note $\varphi_1 \in \mathcal{F}(\mathbb{R}^{24}, \mathbb{R})$, $\varphi_2 \in \mathcal{F}(\mathbb{R}^{IT}, \mathbb{R})$, $\varphi_3 \in \mathcal{F}(\mathbb{R}^T, \mathbb{R})$ et $\varphi_4 \in \mathcal{F}(\mathbb{R}^T, \mathbb{R})$ les applications qui donnent les représentations scalaires des quatre données d'entrée, respectivement les prochaines ventes *day-ahead*, les apports en eau, les prix *day-ahead* et les productions variables, avec la méthode décrite à la sous-section 5.3.2 puis une mise à l'échelle dans l'intervalle $[0, 1]$ (voir Figure 5.13). Pour tout $j \in \{1, \ldots, 4\}$, l'application $\varphi_j \in \mathcal{F}(\mathbb{R}^{L_j}, \mathbb{R})$ (où $L_j = 24$ si $j = 1$, $L_j = IT$ si $j = 2$ et $L_j = T$ si $j \in \{3, 4\}$, sachant que $I = 6$ et $T = 360$) est affine, i.e. il existe une constante $v_j \in \mathbb{R}$ et un vecteur $w_j \in \mathbb{R}^{L_j}$ tels que :

$$\forall z \in \mathbb{R}^{L_j}, \quad \varphi_j(z) = v_j + w_j^\top z.$$

La figure 5.19 représente l'évolution de la variable de sortie Q normalisée (par le chiffre d'affaires de la solution déterministe) dans l'espace des données d'entrée scalaires réduites pour le cas d'étude. On observe sur la figure 5.19b que la fonction Q a une tendance linéaire croissante en fonction des apports en eau normalisés : la fonction Q augmente lorsque les apports en eau augmentent. De même, la figure 5.19c montre que la fonction Q augmente lorsque les prix augmentent. En revanche, la figure 5.19a montre que la fonction Q est globalement constante par rapport aux prochaines ventes *day-ahead*, comme ce qui a été observé lors de l'étude du jeu de données à la sous-section 5.3.3.



(a) Prochaines ventes *day-ahead* (b) Apports en eau (c) Prix *day-ahead*

FIGURE 5.19 – Exemple pour le cas d'étude de l'évolution de la fonction Q en fonction de la représentation scalaire des données d'entrée pour les points du jeu de données. Les productions variables ne sont pas considérées dans le cas d'étude. Les boîtes à moustaches représentent les distributions lorsqu'une donnée d'entrée est fixée et les points représentent les moyennes.

Le métamodèle $\hat{Q}$ défini par régression linéaire sur les données d'entrée scalaires normalisées s'écrit : pour tous $b_0 \in \mathbb{R}^{24}$ et $\xi = (a, \pi, \phi) \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$,

$$\hat{Q}(b_0, \xi) = \beta_0 + \beta_1 \varphi_1(b_0) + \beta_2 \varphi_2(a) + \beta_3 \varphi_3(\pi) + \beta_4 \varphi_4(\phi),$$

où $\beta_j \in \mathbb{R}$, pour $j \in \{0, \dots, 4\}$, sont les paramètres de la régression linéaire qui sont calibrés sur l'ensemble d'apprentissage (donné par les indices $\mathcal{D}_c$, de taille 600 pour le cas d'étude) pour minimiser [6] les moindres carrés [Friedman *et al.*, 2001]

$$\sum_{d \in \mathcal{D}_c} (Q_d - \hat{Q}(b_{m_d}, \xi_{n_d}))^2.$$

Le métamodèle $\hat{Q}$ ainsi construit est affine (donc linéaire par rapport aux données d'entrée) car les applications φ_j, pour $j \in \{1, \dots, 4\}$, sont affines. En outre, le caractère affine des applications φ_2, φ_3 et φ_4 permet d'écrire :

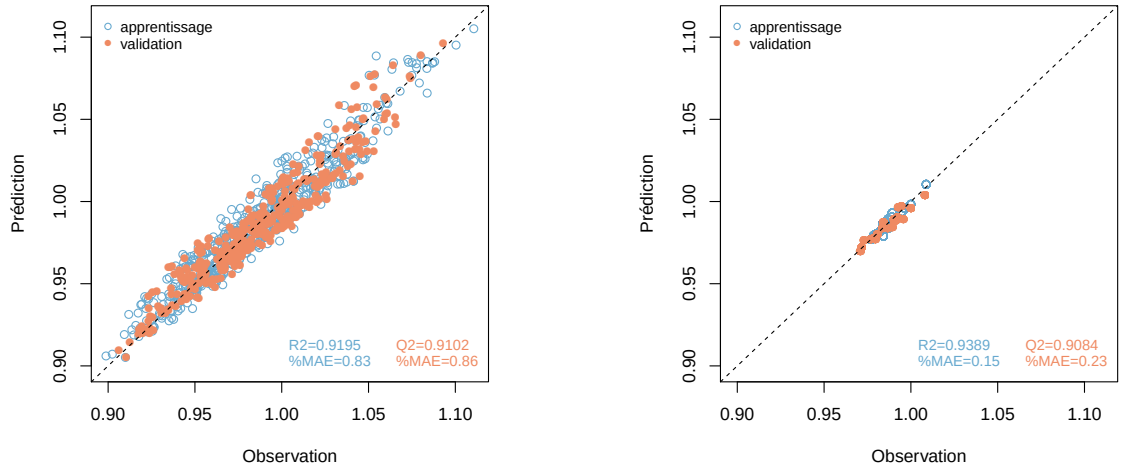
$$\forall b_0 \in \mathbb{R}^{24}, \quad \mathbb{E}[\hat{Q}(b_0, \hat{\Xi})] = \sum_{n=1}^{N} p_n \hat{Q}(b_0, \xi_n) = \tilde{\beta}_0 + \beta_1 \varphi_1(b_0),$$

où

$$\tilde{\beta}_0 = \beta_0 + \beta_2 \varphi_2(\overline{a}) + \beta_3 \varphi_3(\overline{\pi}) + \beta_4 \varphi_4(\overline{\phi}) \in \mathbb{R}.$$

Les résultats de l'apprentissage et de la qualité sur l'ensemble de validation sont présentés à la figure 5.20. On observe que le métamodèle $\hat{Q}$ est satisfaisant pour l'approximation point par point et pour l'approximation des espérances car les erreurs absolues moyennes sur l'ensemble de validation sont respectivement de $\mathrm{MAE}_{\mathrm{ind}} = 0{,}86\%$ et $\mathrm{MAE}_{\mathrm{esp}} = 0{,}23\%$, des valeurs inférieures au MIP-gap relatif seuil de 2%. En revanche, les coefficients de prédictivité sont respectivement $\mathrm{Q}^2_{\mathrm{ind}} = 0{,}91$ et $\mathrm{Q}^2_{\mathrm{esp}} = 0{,}91$, des valeurs inférieures au seuil arbitraire de 0,95. Il convient de noter que les résultats concernant l'approximation des espérances doivent être utilisés avec prudence car la taille de l'ensemble de validation ne permet pas d'obtenir beaucoup de points sur la figure 5.20b (maximum 20 points car la taille M de l'échantillon des prochaines ventes *day-ahead* est égale à $M = 20$).

6. L'algorithme repose sur une méthode de décomposition QR.



(a) Évaluation point par point

(b) Évaluation des espérances

FIGURE 5.20 – Résultats de la régression linéaire sur les données d'entrée scalaires pour le cas d'étude. L'abscisse correspond aux valeurs de la fonction Q et l'ordonnée à celles prédites par le métamodèle $\hat{Q}$, divisées par le chiffre d'affaires de la solution déterministe.

5.4.4.2 Approche sur les données d'entrée réduites

On note $\varphi_1 \in \mathcal{F}(\mathbb{R}^{24}, \mathbb{R}^{K_1})$, $\varphi_2 \in \mathcal{F}(\mathbb{R}^{IT}, \mathbb{R}^{K_2})$, $\varphi_3 \in \mathcal{F}(\mathbb{R}^{T}, \mathbb{R}^{K_3})$ et $\varphi_4 \in \mathcal{F}(\mathbb{R}^{T}, \mathbb{R}^{K_4})$ les applications qui réduisent les quatre données d'entrée, respectivement les prochaines ventes *day-ahead*, les apports en eau, les prix *day-ahead* et les productions variables, avec la méthode décrite à la section 4.2 du chapitre précédent. Pour tout $j \in \{1, \dots, 4\}$, l'entier $K_j \in \mathbb{N}^*$ est le nombre de directions principales retenues lors de la réduction de la donnée d'entrée d'indice j. On rappelle que, pour le cas d'étude, $K_1 = 6$ (voir Sous-section 5.2.4), $K_2 = 7$ et $K_3 = 13$ (voir Sous-section 4.2.4, sachant que $I = 6$ et $T = 360$). La dimension totale des données d'entrée réduites est donc de $7 + 13 + 6 = 26$.

Pour tout $j \in \{1, \dots, 4\}$, l'application $\varphi_j \in \mathcal{F}(\mathbb{R}^{L_j}, \mathbb{R}^{K_j})$ (où $L_j = 24$ si $j = 1$, $L_j = IT$ si $j = 2$ et $L_j = T$ si $j \in \{3, 4\}$) est affine, i.e. il existe une constante $v_j \in \mathbb{R}^{K_j}$ et une matrice $W_j \in \mathcal{M}_{K_j L_j}(\mathbb{R})$ telles que :

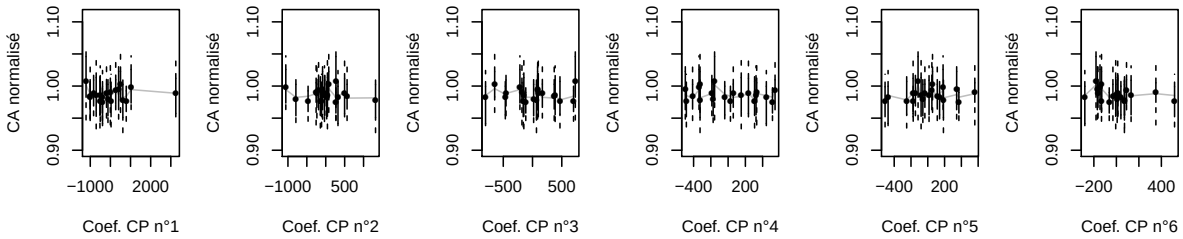
$$\forall z \in \mathbb{R}^{L_j}, \quad \varphi_j(z) = v_j + W_j z.$$

La figure 5.21 représente l'évolution de la variable de sortie Q normalisée (par le chiffre d'affaires de la solution déterministe) dans l'espace des données d'entrée réduites pour le cas d'étude. La seule tendance linéaire remarquable est celle obtenue pour la première direction principale des apports en eau.

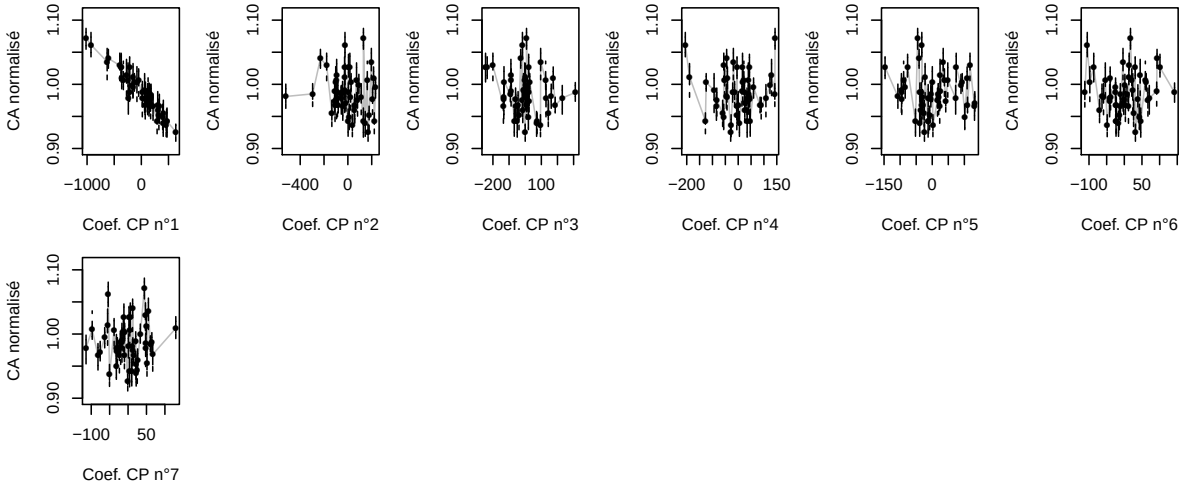
Le métamodèle $\hat{Q}$ défini par régression linéaire sur les données d'entrée réduites s'écrit : pour tous $b_0 \in \mathbb{R}^{24}$ et $\xi = (a, \pi, \phi) \in \mathbb{R}^{IT} \times \mathbb{R}^{T} \times \mathbb{R}^{T}$,

$$\hat{Q}(b_0, \xi) = \beta_0 + \beta_1^\top \varphi_1(b_0) + \beta_2^\top \varphi_2(a) + \beta_3^\top \varphi_3(\pi) + \beta_4^\top \varphi_4(\phi),$$

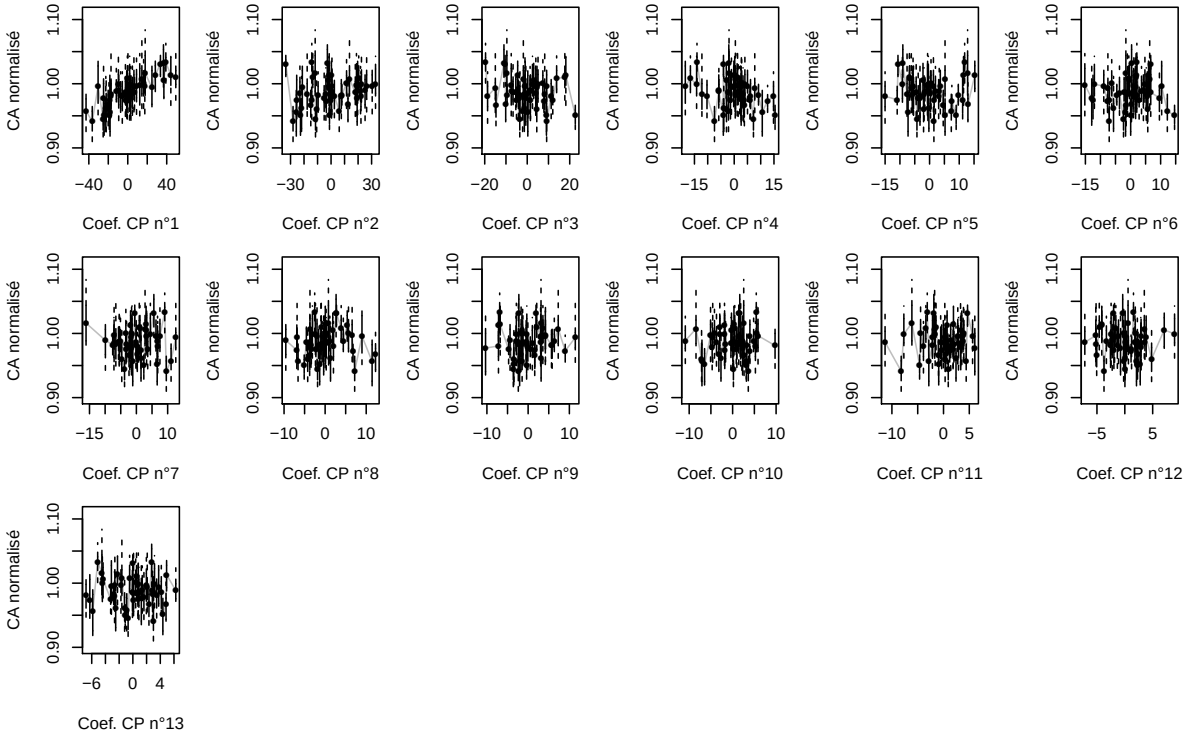
où $\beta_0 \in \mathbb{R}$ et $\beta_j \in \mathbb{R}^{K_j}$, pour $j \in \{1, \dots, 4\}$, sont les paramètres de la régression linéaire qui sont calibrés sur l'ensemble d'apprentissage (donné par les indices $\mathcal{D}_c$, de taille 600 pour le cas d'étude, ce qui devrait être suffisant pour calibrer les $27 = 26 + 1$ paramètres de la régression linéaire).



(a) Prochaines ventes *day-ahead*

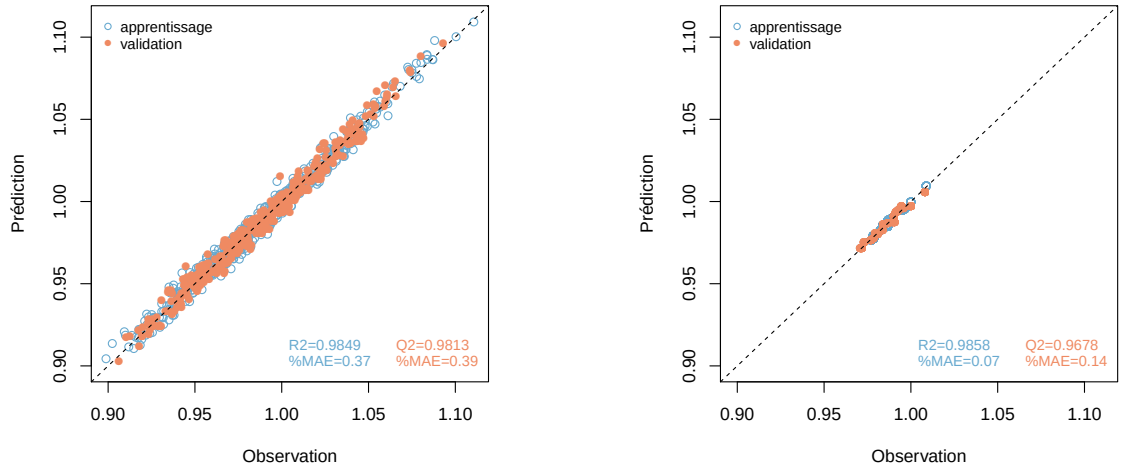


(b) Apports en eau



(c) Prix *day-ahead*

FIGURE 5.21 – Exemple pour le cas d'étude de l'évolution de la fonction Q en fonction des composantes des données d'entrée réduites par analyse en composantes principales. Les productions variables ne sont pas considérées dans le cas d'étude. Les boîtes à moustaches représentent les distributions lorsqu'une donnée d'entrée est fixée et les points représentent les moyennes.



(a) Évaluation point par point

(b) Évaluation des espérances

FIGURE 5.22 – Résultats de la régression linéaire sur les données d'entrée réduites pour le cas d'étude. L'abscisse correspond aux valeurs de la fonction Q et l'ordonnée à celles prédites par le métamodèle $\hat{Q}$, divisées par le chiffre d'affaires de la solution déterministe.

Le métamodèle $\hat{Q}$ ainsi construit est affine (donc linéaire par rapport aux données d'entrée) car les applications φ_j, pour $j \in \{1, \dots, 4\}$, sont affines. En outre, le caractère affine des applications φ_2, φ_3 et φ_4 permet d'écrire :

$$\forall b_0 \in \mathbb{R}^{24}, \quad \mathbb{E}[\hat{Q}(b_0, \hat{\Xi})] = \sum_{n=1}^{N} p_n \hat{Q}(b_0, \xi_n) = \tilde{\beta}_0 + \beta_1^\top \varphi_1(b_0),$$

où

$$\tilde{\beta}_0 = \beta_0 + \beta_2^\top \varphi_2(\overline{a}) + \beta_3^\top \varphi_3(\overline{\pi}) + \beta_4^\top \varphi_4(\overline{\phi}) \in \mathbb{R}.$$

Les résultats de l'apprentissage et de la qualité sur l'ensemble de validation sont présentés à la figure 5.22. On observe que le métamodèle $\hat{Q}$ est acceptable à la fois pour l'approximation point par point et pour l'approximation des espérances car les erreurs absolues moyennes sont respectivement $\text{MAE}_{\text{ind}} = 0{,}39\%$ et $\text{MAE}_{\text{esp}} = 0{,}14\%$, des valeurs inférieures au MIP-gap relatif seuil de 2%. De même, les coefficients de prédictivité sont respectivement $\text{Q}^2_{\text{ind}} = 0{,}98$ et $\text{Q}^2_{\text{esp}} = 0{,}97$, des valeurs supérieures au seuil arbitraire de 0,95. Comme le métamodèle $\hat{Q}$ est acceptable pour l'approximation point par point, nous pouvons faire confiance aux résultats concernant l'approximation des espérances, même si la taille de l'ensemble de validation ne permet pas d'obtenir beaucoup de points sur la figure 5.22b.

5.4.4.3 Approche sur les données d'entrée complètes

La régression LASSO porte sur les données d'entrée complètes. Néanmoins, les composantes des données d'entrée qui sont redondantes (voir Sous-section 4.2.1) sont supprimées. On note $z_1 \in \mathcal{L}(\mathbb{R}^{24}, \mathbb{R}^{L_1})$, $z_2 \in \mathcal{L}(\mathbb{R}^{IT}, \mathbb{R}^{L_2})$, $z_3 \in \mathcal{L}(\mathbb{R}^{T}, \mathbb{R}^{L_3})$ et $z_4 \in \mathcal{L}(\mathbb{R}^{T}, \mathbb{R}^{L_4})$ les applications linéaires qui donnent les composantes non redondantes pour les quatre données d'entrée, respectivement les prochaines ventes *day-ahead*, les apports en eau, les prix *day-ahead* et les productions variables. Pour le cas d'étude, $L_1 = 24$, $L_2 = 6 \times 60$ et $L_3 = 48$ (voir Sous-section 4.2.1) (sachant que $I = 6$ et $T = 360$). La dimension totale des données d'entrée complètes non redondantes est donc $24 + 6 \times 60 + 48 = 432$ (sur les $24 + 6 \times 360 + 360 = 2\,544$ en tout).

Le métamodèle $\hat{Q}$ défini par régression LASSO sur les données d'entrée complètes s'écrit : pour tous $b_0 \in \mathbb{R}^{24}$ et $\xi = (a, \pi, \phi) \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$,

$$\hat{Q}(b_0, \xi) = \beta_0 + \beta_1^\top z_1(b_0) + \beta_2^\top z_2(a) + \beta_3^\top z_3(\pi) + \beta_4^\top z_4(\phi), \tag{5.6}$$

où $\beta_0 \in \mathbb{R}$ et $\beta_j = (\beta_{j\ell})_{1 \leq \ell \leq L_j} \in \mathbb{R}^{L_j}$, pour $j \in \{1, \dots, 4\}$, sont les paramètres de la régression linéaire qui sont calibrés sur l'ensemble d'apprentissage (donné par les indices $\mathcal{D}_c$, de taille 600 pour le cas d'étude) avec les données d'entrée complètes non redondantes, pour minimiser les moindres carrés régularisés [Tibshirani, 1996 ; Friedman *et al.*, 2001]

$$\frac{1}{2} \sum_{d \in \mathcal{D}_c} (Q_d - \hat{Q}(b_{m_d}, \xi_{n_d}))^2 - \sum_{j=1}^{4} \lambda_j \sum_{\ell=1}^{L_j} |\beta_{j\ell}|, \tag{5.7}$$

avec $\lambda_j \in \mathbb{R}$, pour $j \in \{1, \dots, 4\}$, les coefficients de régularisation. On obtiendrait une régression linéaire classique sur les données d'entrée complètes non redondantes sans le second terme dans l'équation (5.7). L'expression de l'équation (5.7) est minimisée en pratique par maximum de vraisemblance pénalisée [Friedman *et al.*, 2010], dont les coefficients de régularisation sont des hyperparamètres (i.e. ils doivent être définis *a priori*). Les valeurs des coefficients de régularisation sont alors obtenues par validation croisée [Refaeilzadeh *et al.*, 2009] sur l'ensemble d'apprentissage à partir d'une famille de 500 valeurs initiales.

En pratique, la minimisation des moindres carrés régularisés de l'équation (5.7) entraîne qu'une partie des composantes des paramètres β_j, pour $j \in \{1, \dots, 4\}$, sont nulles. Ainsi, seules certaines composantes des données d'entrée complètes sont retenues par le LASSO. On peut alors définir les applications linéaires $\varphi_1 \in \mathcal{L}(\mathbb{R}^{24}, \mathbb{R}^{K_1})$, $\varphi_2 \in \mathcal{L}(\mathbb{R}^{IT}, \mathbb{R}^{K_2})$, $\varphi_3 \in \mathcal{L}(\mathbb{R}^T, \mathbb{R}^{K_3})$ et $\varphi_4 \in \mathcal{L}(\mathbb{R}^T, \mathbb{R}^{K_4})$ qui donnent les composantes retenues par la régularisation LASSO pour les quatre données d'entrée, respectivement les prochaines ventes *day-ahead*, les apports en eau, les prix *day-ahead* et les productions variables. Pour le cas d'étude, on obtient $K_1 = 11$, $K_2 = 100$ et $K_3 = 33$, ce qui donne des dimensions supérieures à celles obtenues après réduction par analyse en composantes principales (voir §5.4.4.2). Les $11 + 100 + 33 = 144 < 432$ composantes des données d'entrée retenues par la régularisation LASSO sont présentées à la figure 5.23. On remarque qu'elles correspondent principalement aux pics des prochaines ventes *day-ahead* et des prix *day-ahead*. L'affluent des Usses est peu utilisé par la régression LASSO.

En supprimant les composantes non retenues dans l'équation (5.6), le métamodèle $\hat{Q}$ de la régression LASSO sur les données d'entrée complètes s'écrit plus simplement : pour tous $b_0 \in \mathbb{R}^{24}$ et $\xi = (a, \pi, \phi) \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$,

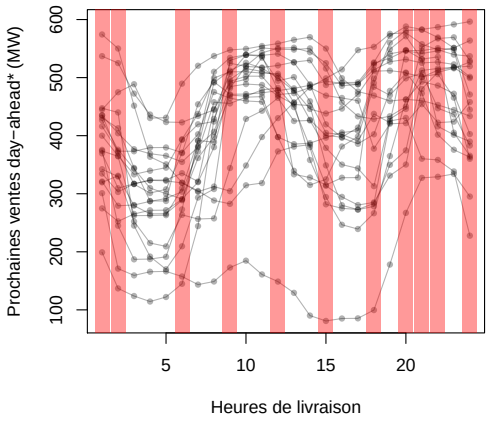
$$\hat{Q}(b_0, \xi) = \beta_0 + \tilde{\beta}_1^\top \varphi_1(b_0) + \tilde{\beta}_2^\top \varphi_2(a) + \tilde{\beta}_3^\top \varphi_3(\pi) + \tilde{\beta}_4^\top \varphi_4(\phi),$$

où, pour tout $j \in \{1, \dots, 4\}$, $\tilde{\beta}_j \in \mathbb{R}^{K_j}$ est obtenu à partir de $\beta_j \in \mathbb{R}^{L_j}$ en retirant les composantes nulles. Le métamodèle $\hat{Q}$ ainsi construit est affine (donc linéaire par rapport aux données d'entrée) car les applications φ_j, pour $j \in \{1, \dots, 4\}$, sont linéaires. En outre, la linéarité des applications φ_2, φ_3 et φ_4 permet d'écrire :

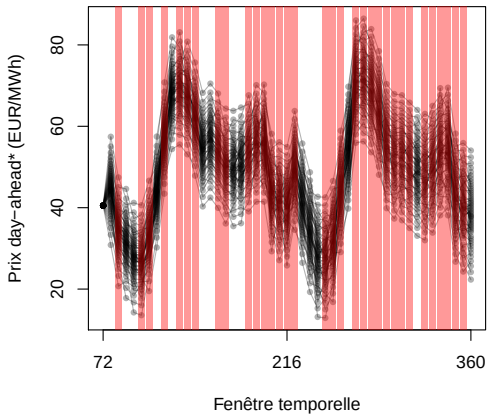
$$\forall b_0 \in \mathbb{R}^{24}, \quad \mathbb{E}[\hat{Q}(b_0, \hat{\Xi})] = \sum_{n=1}^{N} p_n \hat{Q}(b_0, \xi_n) = \tilde{\beta}_0 + \tilde{\beta}_1^\top \varphi_1(b_0),$$

où

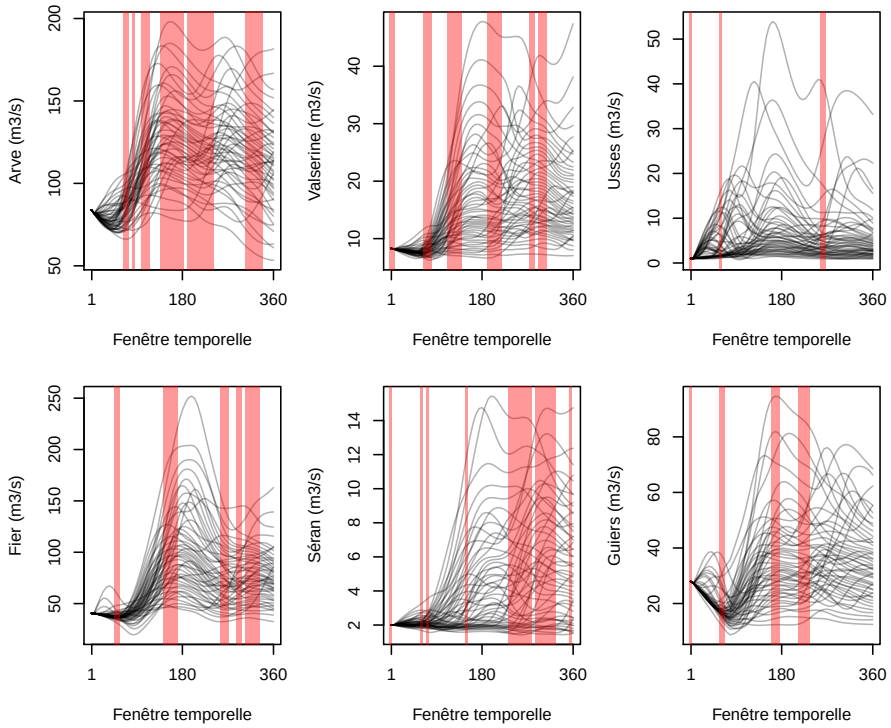
$$\tilde{\beta}_0 = \beta_0 + \tilde{\beta}_2^\top \varphi_2(\overline{a}) + \tilde{\beta}_3^\top \varphi_3(\overline{\pi}) + \tilde{\beta}_4^\top \varphi_4(\overline{\phi}) \in \mathbb{R}.$$



(a) Prochaines ventes *day-ahead*

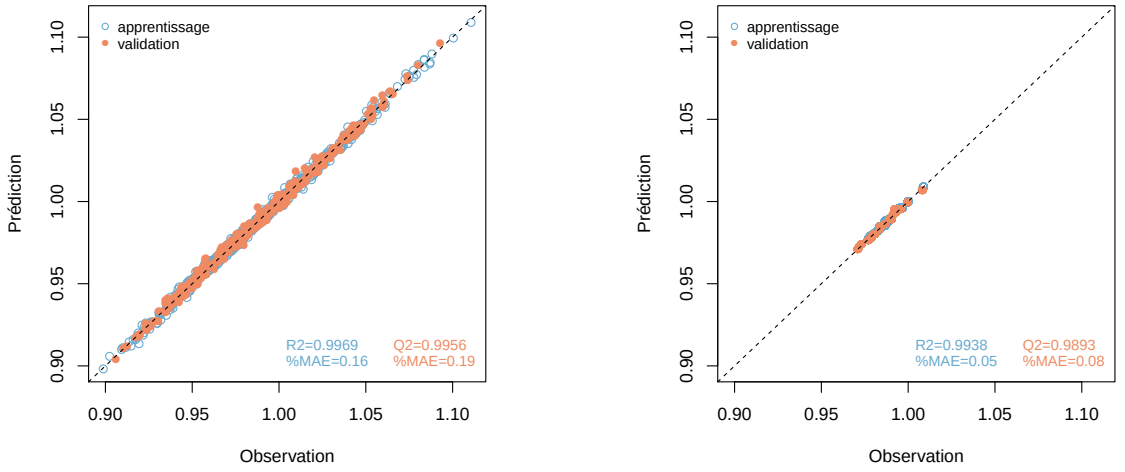


(b) Prix *day-ahead*



(c) Apports en eau

FIGURE 5.23 – Composantes des données d'entrée retenues (en rouge) par la régularisation LASSO pour le cas d'étude. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les puissances et les prix.



(a) Évaluation point par point

(b) Évaluation des espérances

FIGURE 5.24 – Résultats de la régression LASSO sur les données d'entrée complètes pour le cas d'étude. L'abscisse correspond aux valeurs de la fonction Q et l'ordonnée à celles prédites par le métamodèle $\hat{Q}$, divisées par le chiffre d'affaires de la solution déterministe.

Les résultats de l'apprentissage et de la qualité sur l'ensemble de validation sont présentés à la figure 5.24. Les résultats sont analogues à ceux obtenus par régression linéaire sur les données d'entrée réduites (voir §5.4.4.2). On observe que le métamodèle $\hat{Q}$ est acceptable à la fois pour l'approximation point par point et pour l'approximation des espérances car les erreurs absolues moyennes sont respectivement $\text{MAE}_{\text{ind}} = 0{,}19\%$ et $\text{MAE}_{\text{esp}} = 0{,}08\%$, des valeurs inférieures au MIP-gap relatif seuil de 2%. De même, les coefficients de prédictivité sont respectivement $\text{Q}^2_{\text{ind}} = 1{,}00$ et $\text{Q}^2_{\text{esp}} = 0{,}99$, ce qui est très supérieur au seuil arbitraire de 0,95. Comme le métamodèle $\hat{Q}$ est acceptable pour l'approximation point par point, nous pouvons faire confiance aux résultats concernant l'approximation des espérances, même si la taille de l'ensemble de validation ne permet pas d'obtenir beaucoup de points sur la figure 5.24b.

5.4.4.4 Comparaison des approches d'apprentissage

Le tableau 5.2 récapitule les résultats obtenus avec les métamodèles des trois approches d'apprentissage testées sur le cas d'étude. Ces trois approches sont *a priori* acceptables par rapport à l'approximation des espérances. Elles peuvent donc toutes les trois être utilisées dans le problème d'optimisation probabiliste approché (5.5). D'après les résultats, il semble que la fonction $\mathcal{E} : b_0 \mapsto \mathbb{E}[Q(b_0, \hat{\Xi})]$ est très proche d'être linéaire sur le jeu de données (le jeu de données semble donc compris dans un des morceaux où la fonction $\mathcal{E}$ est linéaire).

Approche	Q^2_{ind}	MAE_{ind}	Q^2_{esp}	MAE_{esp}
entrées scalaires	0,9102	0,86%	0,9080	0,23%
entrées réduites	0,9813	0,39%	0,9678	0,14%
entrées complètes	0,9956	0,19%	0,9893	0,08%

TABLEAU 5.2 – Récapitulatif de la qualité des métamodèles des trois approches d'apprentissage testées sur le cas d'étude.

Néanmoins, les critères de qualité associés à l'approximation des espérances sont obtenus avec peu de points. Il est donc préférable de comparer les approches par les critères de qualité associés à l'approximation point par point car ils sont obtenus sur davantage de points et ils sont plus forts que ceux associés à l'approximation des espérances (un métamodèle acceptable pour l'approximation point par point est acceptable pour l'approximation des espérances). Ainsi, l'approche sur les données d'entrée scalaires est à proscrire (mauvais coefficient de prédictivité et plus grande erreur absolue moyenne).

Parmi les deux autres approches, il est préférable d'utiliser l'approche sur les données d'entrée réduites car elle est plus facile à manipuler et interpréter que celle sur les données d'entrée complètes par régression LASSO. En outre, comme les composantes des données d'entrée sont fortement corrélées, la régression LASSO peut ne pas être suffisamment consistante en pratique [Zhao et Yu, 2006].

Les temps de calcul des trois approches pour l'apprentissage du métamodèle sont négligeables par rapport au temps limite de 1 h, même si la régression LASSO nécessite un temps de calcul légèrement plus élevé que ceux des deux autres approches (en raison notamment de la validation croisée pour calibrer les coefficients de régularisation).

Le meilleur compromis entre les méthodes de régression pour le cas d'étude est donc la régression sur les données d'entrée réduites par analyse en composantes principales.

5.5 Conclusion du chapitre

Nous avons détaillé dans ce chapitre la démarche proposée pour résoudre numériquement le problème d'optimisation probabiliste à deux niveaux de la planification à court terme de la production. Cette démarche consiste à remplacer la valeur optimale du second niveau par un métamodèle calibré dans une phase de pré-traitement à l'optimisation. Un jeu de données est alors créé en amont de l'optimisation, à partir d'un plan d'expérience sur le lequel la valeur optimale du second niveau est évaluée. Comme le domaine de la donnée d'entrée associée aux prochaines ventes *day-ahead* n'est pas connu sous une forme explicite, il est d'abord approché par un ensemble fini obtenu par analogie. Parmi les approches par analogie testées, les meilleurs résultats sont globalement obtenus avec celle dont la variable d'analogie est la prévision déterministe des apports cumulés, le critère d'analogie la distance sur les données complètes et avec une taille d'échantillonnage égale à 24. Un échantillonnage par hypercube latin est ensuite appliqué pour construire le plan d'expérience. Le jeu de données ainsi obtenu permet de calibrer le métamodèle par régression linéaire. Parmi les méthodes d'apprentissage testées sur le cas d'étude, la régression linéaire sur les données d'entrée réduites par analyses en composantes principales est la plus adaptée.

Dans le chapitre suivant, nous donnons quelques compléments sur le choix des hyperparmètres de la régression linéaire sur les données d'entrée réduites, puis nous validons l'ensemble de la méthodologie en considérant la résolution numérique du problème d'optimisation probabiliste approché avec le métamodèle. Quelques perspectives sont ensuite présentées, notamment les grandes lignes d'une approche par approximations successives où le métamodèle est construit pendant la phase d'optimisation et non dans une phase de pré-traitement.

Chapitre 6

Conclusions méthodologiques et perspectives

Dans le chapitre précédent, nous avons présenté la construction du métamodèle dans une phase de pré-traitement pour résoudre le problème d'optimisation de la planification de la production d'une chaîne hydroélectrique et d'actifs de production variable en univers probabiliste. Ce chapitre permet de conclure sur l'intérêt de la méthodologie proposée pour formuler mathématiquement et résoudre numériquement le problème d'optimisation en univers probabiliste. Cette méthodologie est résumée dans la première section. Dans la deuxième section, nous validons la méthodologie sur le cas d'étude en considérant la résolution numérique du problème d'optimisation probabiliste approché avec le métamodèle. Dans la dernière section, nous évoquons quelques perspectives générales.

6.1 Résumé de la méthodologie proposée

6.1.1 Problématique

Le problème abordé dans ce manuscrit est celui de la planification à court terme de la production conjointe d'une chaîne hydroélectrique au fil de l'eau et d'actifs de production variable. Ce problème consiste à déterminer en même temps les prochaines offres de vente à effectuer au marché *day-ahead* (pour tout le périmètre d'équilibre) et le programme de production de la chaîne hydroélectrique sur la fenêtre temporelle à court terme (résolution temporelle fine). La solution du problème est déterminée de manière à respecter les contraintes d'exploitation et à maximiser le chiffre d'affaires obtenu par les ventes de la production de l'ensemble du périmètre d'équilibre au marché *day-ahead* et par la pénalisation des écarts. On agit ainsi sur le placement de la production hydroélectrique à la fois dans le temps et entre les centrales hydroélectriques. Cela est possible grâce à la légère flexibilité de production offerte par la chaîne hydroélectrique au fil de l'eau.

En univers déterministe, le problème de planification à court terme de la production s'écrit mathématiquement sous la forme du problème MILP (i.e. problème de programmation linéaire mixte) suivant :

$$\max_{\mathbf{x}_0 \in \mathcal{X}(\xi_0)} f(\mathbf{x}_0, \xi_0), \tag{6.1}$$

où

— $\xi_0 = (a_0, \pi_0, \phi_0) \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ est le scénario déterministe composé des prévisions déterministes des apports en eau $a_0 \in \mathbb{R}^{IT}$, des prix $\pi_0 \in \mathbb{R}^T$ et des productions variables $\phi_0 \in \mathbb{R}^T$ ($I \in \mathbb{N}^*$ désigne le nombre d'aménagements de la chaîne hydroélectrique et T

le nombre de pas de temps de la fenêtre temporelle sur laquelle la planification de la production est considérée),

— $\mathcal{X}(\xi_0) \subset \mathbb{R}^r$ est l'ensemble des solutions admissibles pour un scénario déterministe $\xi_0 \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ (partie fermée, bornée et non vide de $\mathbb{R}^r$), défini à partir de contraintes MILP (i.e. linéaires mixtes) qui reprennent les équations du modèle hydraulique et les contraintes d'exploitation de la chaîne hydroélectrique,

— $\mathbf{x}_0 \mapsto f(\mathbf{x}_0, \xi_0)$ est une fonction linéaire réelle définie sur $\mathbb{R}^r$ qui représente le chiffre d'affaires d'une solution pour le scénario déterministe.

Une solution admissible $\mathbf{x}_0 \in \mathcal{X}(\xi_0)$ comporte à la fois les prochaines offres de vente à effectuer au marché *day-ahead* et le programme de production de le chaîne hydroélectrique.

Le scénario déterministe ξ_0 est composé de prévisions déterministes, alors qu'il est sujet en réalité à des incertitudes. L'objectif de la méthodologie proposée est de faire passer le problème d'optimisation (6.1) de l'univers déterministe à l'univers probabiliste, en prenant en compte ces incertitudes de manière à améliorer la valorisation de la production. Nous nous appuyons pour cela sur une modélisation ensembliste des incertitudes : les incertitudes sont modélisées par une variable aléatoire discrète et finie $\hat{\Xi}$ sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ avec $\hat{\Xi}(\Omega) = \{\xi_1, \ldots, \xi_N\} \subset \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ ($N \in \mathbb{N}^*$). Les éléments $\xi_n \in \hat{\Xi}(\Omega)$ sont les scénarios probabilistes (N désigne donc le nombre de scénarios probabilistes).

6.1.2 Formulation mathématique du problème d'optimisation probabiliste

Dans le processus opérationnel de la gestion de la chaîne hydroélectrique, le problème de la planification à court terme de la production est résolu à des instants réguliers au cours du temps. La durée entre deux planifications successives est petit par rapport à l'horizon de la planification. Une partie de la solution obtenue à une planification n'est donc pas définitive : elle est réoptimisée aux planifications ultérieures. En revanche, pour les planifications en fin de matinée (juste avant la fermeture à midi de la session quotidienne du marché *day-ahead* portant sur la production du lendemain), les offres de vente au marché *day-ahead* pour la production du lendemain sont définitives car elles ne pourront plus être modifiées par la suite.

Cet enchaînement des planifications et des décisions des ventes au marché *day-ahead* peut être utilisé pour faire passer le problème d'optimisation (6.1) de l'univers déterministe à l'univers probabiliste. En effet, les scénarios probabilistes peuvent servir à anticiper les prochaines réoptimisations possibles de la solution et leurs impacts sur le chiffre d'affaires.

Le problème de la planification à court terme de la production en univers probabiliste s'écrit alors sous la forme du problème de programmation dynamique stochastique à deux niveaux suivant :

$$\max_{\mathbf{x}_0 \in \mathcal{X}(\xi_0)} \left\{ \alpha_0 f(\mathbf{x}_0, \xi_0) + (1 - \alpha_0) \mathbb{E}[Q(b(\mathbf{x}_0), \hat{\Xi})] \right\}, \tag{6.2}$$

où

— $\alpha_0 \in [0, 1]$ est un paramètre arbitraire, qui peut être défini par le poids d'un *cluster* central des scénarios probabilistes (dans ce cas, le poids α_0 est un indicateur de la dispersion des scénarios probabilistes par rapport au scénario déterministe : plus α_0 est proche de 0, plus les incertitudes sont fortes),

— b est une fonction linéaire de $\mathbb{R}^r$ dans $\mathbb{R}^{24}$ qui donne les offres de vente à effectuer à la prochaine session du marché *day-ahead* associées à une solution,

— Q est la valeur optimale du second niveau, coûteuse à évaluer numériquement, qui donne le chiffre d'affaires optimal que l'on peut atteindre pour une décision de ventes *day-ahead*

$b_0 \in \mathbb{R}^{24}$ et un scénario $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ sous l'hypothèse qu'il se réalise effectivement (problème MILP) :

$$\forall (b_0, \xi) \in \mathbb{R}^{24} \times (\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T), \quad Q(b_0, \xi) = \max_{\substack{\mathbf{x} \in \mathcal{X}(\xi) \\ b(\mathbf{x}) = b_0}} f(\mathbf{x}, \xi) \in \mathbb{R} \cup \{-\infty\}, \tag{6.3}$$

avec $\mathcal{X}(\xi) \subset \mathbb{R}^r$ l'ensemble des solutions admissibles pour un scénario $\xi \in \mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T$ (partie fermée et bornée de $\mathbb{R}^r$, non vide si $\xi \in \{\xi_0, \xi_1, \ldots, \xi_N\}$), défini à partir des mêmes contraintes MILP que $\mathcal{X}(\xi_0)$ mais en considérant le scénario ξ au lieu du scénario déterministe ξ_0, et f est une fonction réelle définie sur $\mathbb{R}^r \times (\mathbb{R}^{IT} \times \mathbb{R}^T \times \mathbb{R}^T)$ qui représente le chiffre d'affaires d'une solution pour un scénario (la fonction f est linéaire par rapport à son premier argument).

Le problème d'optimisation probabiliste (6.2) est identique au problème déterministe lorsque $\alpha_0 = 1$ (i.e. lorsque les scénarios probabilistes ne sont pas considérés).

6.1.3 Résolution numérique à l'aide d'un métamodèle

La résolution numérique du problème d'optimisation probabiliste (6.2) n'est pas accessible en pratique, surtout pour une utilisation en opérationnel où le temps maximal disponible pour construire une solution est limité (1 h pour notre cas d'étude).

Nous proposons donc de résoudre le problème d'optimisation probabiliste approché

$$\max_{\mathbf{x}_0 \in \mathcal{X}(\xi_0)} \left\{ \alpha_0 f(\mathbf{x}_0, \xi_0) + (1 - \alpha_0) \mathbb{E}[\hat{Q}(b(\mathbf{x}_0), \hat{\Xi})] \right\}, \tag{6.4}$$

où $\hat{Q}$ est un métamodèle linéaire construit par apprentissage supervisé dans une phase de pré-traitement à l'optimisation. Le problème d'optimisation (6.4) est MILP et il est très similaire au problème MILP déterministe (6.1) (seulement quelques modifications dans la fonction objectif). En particulier, le problème MILP (6.4) se résout avec le même solveur que le problème MILP déterministe (6.1) et les temps de calcul devraient *a priori* être du même ordre de grandeur.

La difficulté est alors de construire le métamodèle $\hat{Q}$ dans la phase de pré-traitement. Comme l'ensemble $\mathcal{B}_0 = \{b(\mathbf{x}_0),\ \mathbf{x}_0 \in \mathcal{X}(\xi_0)\} \subset \mathbb{R}^{24}$ n'est pas accessible en pratique (écriture sous une forme non explicite à partir de contraintes MILP), il est d'abord approché par un ensemble $\hat{\mathcal{B}}_0 = \{b_1, \ldots, b_M\} \subset \mathbb{R}^{24}$ ($M \in \mathbb{N}^*$), où $(b_1, \ldots, b_M)$ est un échantillon des prochaines ventes *day-ahead* obtenu par analogie (voir Section 5.2). Un plan d'expérience est ensuite construit de manière à sélectionner quelques points $(b_{m_d}, \xi_{n_d}) \in \hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$, pour $d \in \{1, \ldots, D\}$ ($D \in \mathbb{N}^*$), sur lesquels la fonction coûteuse Q est évaluée pour former le jeu de données (voir Section 5.3). Le métamodèle $\hat{Q}$ est alors construit grâce au jeu de données par régression linéaire sur les données réduites par analyse en composantes principales (voir Section 5.4). Les étapes de la résolution numérique du problème d'optimisation probabiliste sont représentées sur la figure 6.1 (copie par commodité de la figure 5.2).

6.2 Validation de la méthodologie proposée

6.2.1 Résolution du problème d'optimisation probabiliste approché

Dans cette sous-section, nous étudions la résolution du problème d'optimisation probabiliste approché (6.4) pour le cas d'étude. Plus précisément, le problème MILP (6.4) est résolu pour le cas d'étude avec cinq valeurs du poids α_0 : $\alpha_{0_k}^* = (k-1)/4$, pour $k \in \{1, \ldots, 5\}$. Pour tout $k \in \{1, \ldots, 5\}$, on note $\mathbf{x}_{0_k}^* \in \mathcal{X}(\xi_0)$ la solution optimale du problème MILP (6.4) avec le poids $\alpha_0 = \alpha_{0_k}^* = (k-1)/4$. On rappelle que la solution optimale $\mathbf{x}_{0_5}^* \in \mathcal{X}(\xi_0)$ obtenue avec le poids $\alpha_0 = \alpha_{0_5}^* = 1$ est identique à la solution déterministe $\mathbf{x}_0^* \in \mathcal{X}(\xi_0)$ de la sous-section 3.2.6 qui sert de référence pour les comparaisons.

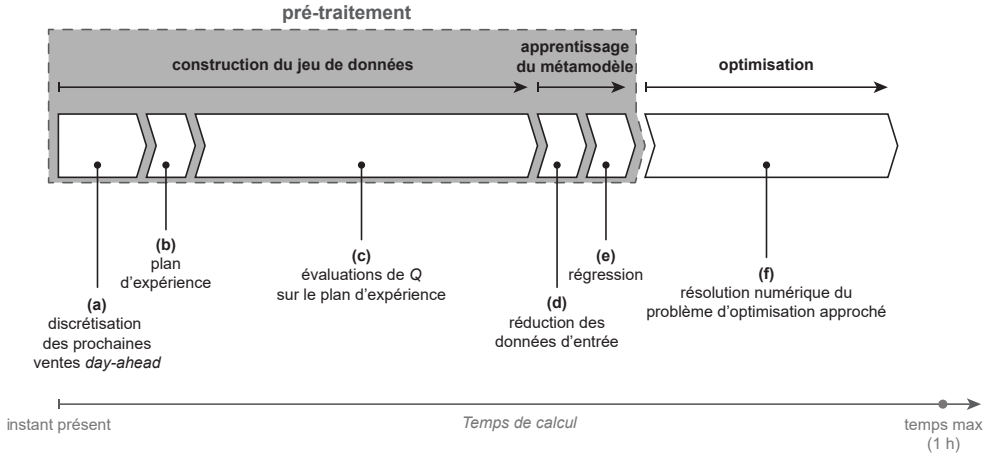
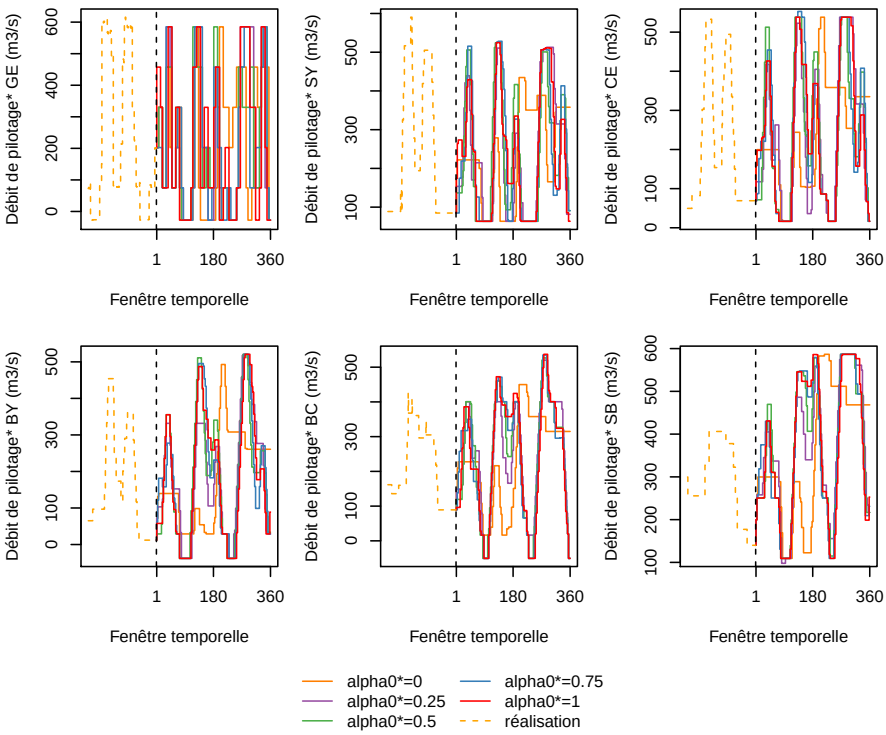


FIGURE 6.1 – Étapes de la résolution numérique du problème d'optimisation probabiliste.

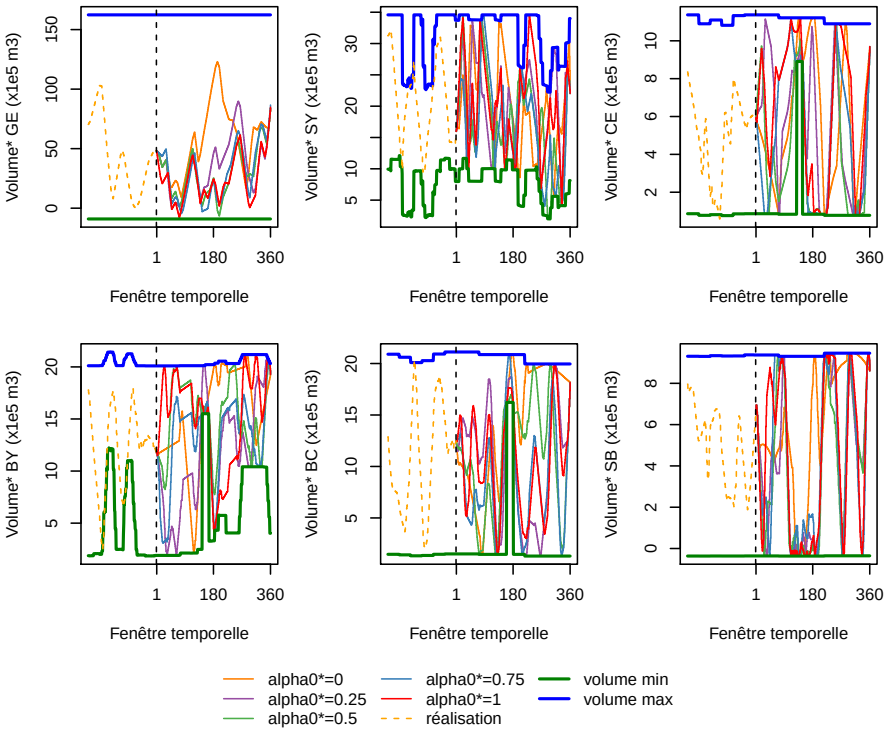
La figure 6.2 présente deux types de variables de décision des solutions optimales $\mathbf{x}_{0_k}^* \in \mathcal{X}(\xi_0)$, pour $k \in \{1, \ldots, 5\}$, associées au programme de production : les débits de pilotage et les volumes de retenue. On observe que les débits de pilotage des solutions optimales sont assez proches, sauf ceux de la solution optimale associée au poids $\alpha_0 = \alpha_{0_1}^* = 0$ qui correspond au cas où le scénario déterministe n'est pas du tout utilisé. Le fait de tenir compte du scénario déterministe, même avec un poids très faible, semble donc conditionner le programme de production. Les faibles différences de débit de pilotage se retrouvent sur les différences entre les volumes de retenue des solutions optimales (on rappelle que les volumes de retenue se déduisent des débits de pilotage et des apports en eau considérés (ici ceux associés à la prévision déterministe), grâce au bilan d'eau de l'équation (2.1)).

La figure 6.3 illustre l'évolution temporelle et la représentation scalaire réduite des prochaines ventes *day-ahead* $b_{0_k}^* = b(\mathbf{x}_{0_k}^*) \in \mathcal{B}_0$ associées aux solutions optimales $\mathbf{x}_{0_k}^*$, pour $k \in \{1, \ldots, 5\}$. Elles sont comparées avec les prochaines ventes *day-ahead* $b_m \in \mathbb{R}^{24}$, pour $m \in \{1, \ldots, M\}$ ($M = 20$ pour le cas d'étude) obtenues par analogie (courbes en gris, qui correspondent aux figures 5.11 et 5.13a du chapitre précédent) et qui forment l'ensemble approché $\hat{\mathcal{B}}_0 \subset \mathbb{R}^{24}$ utilisé pour construire le plan d'expérience. On observe que les différences entre les prochaines ventes *day-ahead* des cinq solutions optimales $\mathbf{x}_{0_k}^*$, pour $k \in \{1, \ldots, 5\}$, sont assez faibles comparées aux différences entre les prochaines ventes *day-ahead* de l'ensemble $\hat{\mathcal{B}}_0$. Les prochaines ventes *day-ahead* associées aux poids $\alpha_0 = \alpha_{0_1}^* = 0$ et $\alpha_0 = \alpha_{0_2}^* = 0{,}25$ ont néanmoins un creux plus accentué en milieu de journée que les trois autres. Celles associées aux poids $\alpha_0 = \alpha_{0_1}^* = 0$ a même un pic en fin de matinée plus fin et en avance (vers 10 h) ainsi qu'un pic en fin de journée plus tardif (vers 22 h) par rapport aux autres (vers 11 h pour le pic du matin et 20 h pour le pic du soir). Globalement, plus le poids α_0 est proche de 1, plus les prochaines ventes *day-ahead* sont visuellement proches de celles associées à la solution déterministe (obtenue pour $\alpha_0 = \alpha_{0_5}^* = 1$). Il semble donc y avoir une certaine « continuité » entre le choix du poids α_0 et la distance des prochaines ventes *day-ahead* par rapport à celles de la solution déterministe.

On observe également que les prochaines ventes *day-ahead* des cinq solutions optimales $\mathbf{x}_{0_k}^*$, pour $k \in \{1, \ldots, 5\}$, sont à la frontière inférieure de la zone délimitée par l'ensemble approché $\hat{\mathcal{B}}_0$. Elles sont même en dehors de la zone sur les premières heures de livraison. Autrement dit, les prochaines ventes *day-ahead* des solutions optimales $\mathbf{x}_{0_k}^*$, pour $k \in \{1, \ldots, 5\}$, sont situées à la limite de la zone définie par l'ensemble d'apprentissage sur lequel le métamodèle $\hat{Q}$ est calibré et par l'ensemble de validation sur lequel la qualité du métamodèle $\hat{Q}$ est évaluée.

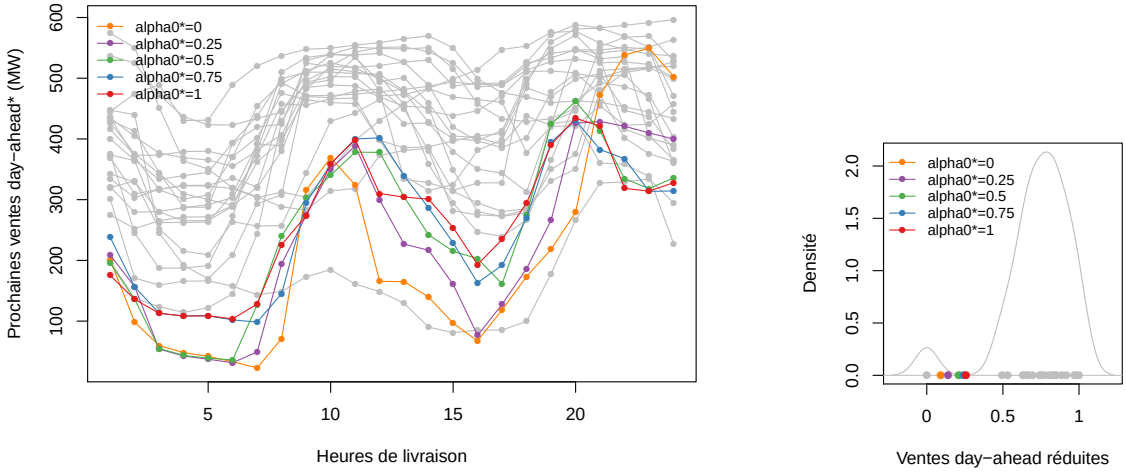


(a) Débits de pilotage



(b) Volumes des retenues

FIGURE 6.2 – Débits de pilotage et volumes des retenues des solutions optimales du problème d'optimisation probabiliste approché du cas d'étude avec les poids $\alpha_{0_k}^* = (k-1)/4$, pour $k \in \{1, \ldots, 5\}$. Les lignes verticales en tirets correspondent au début de la fenêtre temporelle. Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les volumes et les débits.



(a) Évolution temporelle

(b) Représentation scalaire

FIGURE 6.3 – Prochaines ventes *day-ahead* des solutions optimales du problème d'optimisation probabiliste approché du cas d'étude (en couleurs) avec les poids $\alpha_{0_k}^* = (k-1)/4$, pour $k \in \{1, \ldots, 5\}$, en comparaison avec l'échantillon des prochaines ventes *day-ahead* utilisé pour l'apprentissage du métamodèle (en gris). Pour des raisons de confidentialité, une transformation affine arbitraire est appliquée sur les puissances.

Comme $\hat{\mathcal{B}}_0 \neq \mathcal{B}_0$ (pour le cas d'étude, on a même $\hat{\mathcal{B}}_0 \cap \mathcal{B}_0 = \emptyset$), la bonne qualité du métamodèle $\hat{Q}$ sur l'ensemble $\hat{\mathcal{B}}_0 \times \hat{\Xi}(\Omega)$ (voir Sous-section 5.4.4) ne se retrouve pas nécessairement sur l'ensemble $\mathcal{B}_0 \times \hat{\Xi}(\Omega)$ réellement utilisé dans le problème MILP (6.4). On ne connait donc pas les erreurs d'approximation de la fonction approchée $b_0 \mapsto \mathbb{E}[\hat{Q}(b_0, \hat{\Xi})]$ par rapport à la fonction $b_0 \mapsto \mathbb{E}[Q(b_0, \hat{\Xi})]$ sur l'ensemble $\mathcal{B}_0 \setminus \hat{\mathcal{B}}_0$. Ces éventuelles erreurs se répercutent sur le calcul de la fonction objectif, donc sur l'optimalité des solutions renvoyées par l'optimiseur probabiliste. Les deux risques principaux sont :

— si la fonction approchée $b_0 \mapsto \mathbb{E}[\hat{Q}(b_0, \hat{\Xi})]$ sur-estime la fonction $b_0 \mapsto \mathbb{E}[Q(b_0, \hat{\Xi})]$ sur une partie de l'ensemble $\mathcal{B}_0 \setminus \hat{\mathcal{B}}_0$, alors ces erreurs d'approximation peuvent « faussement » conduire à une solution optimale dont les prochaines ventes *day-ahead* sont dans cette partie sur-estimée,

— si la fonction approchée $b_0 \mapsto \mathbb{E}[\hat{Q}(b_0, \hat{\Xi})]$ sous-estime la fonction $b_0 \mapsto \mathbb{E}[Q(b_0, \hat{\Xi})]$ sur une partie de l'ensemble $\mathcal{B}_0 \setminus \hat{\mathcal{B}}_0$, alors ces erreurs d'approximation peuvent « faussement » conduire à une solution optimale dont les prochaines ventes *day-ahead* sont en-dehors de cette partie sous-estimée.

Comme l'ensemble $\mathcal{B}_0 = \{b(\mathbf{x}_0),\ \mathbf{x}_0 \in \mathcal{X}(\xi_0)\}$ n'est pas accessible en pratique (défini sous une forme non explicite à partir de contraintes MILP), il n'est pas possible d'évaluer la qualité du métamodèle $\hat{Q}$ sur l'ensemble $\mathcal{B}_0 \times \hat{\Xi}(\Omega)$. En revanche, nous pouvons utiliser les solutions optimales $\mathbf{x}_{0_k}^*$, pour $k \in \{1, \ldots, 5\}$, pour évaluer la qualité du métamodèle $\hat{Q}$ sur le sous-ensemble $\{b_{0_k}^*,\ 1 \leq k \leq 5\} \times \hat{\Xi}(\Omega) \subset \mathcal{B}_0 \times \hat{\Xi}(\Omega)$ ($\{b_{0_k}^*,\ 1 \leq k \leq 5\} \subset \mathcal{B}_0$) obtenu à partir des prochaines ventes *day-ahead* des cinq solutions optimales (compatibles avec le scénario déterministe) et des N scénarios probabilistes. Les critères de qualité utilisés sont les mêmes que ceux utilisés à la sous-section 5.4.3 pour évaluer la qualité du métamodèle sur l'ensemble de validation. On mesure donc le coefficient de prédictivité Q^2 et l'erreur absolue moyenne MAE des chiffres d'affaires normalisés pour l'approximation point par point et pour l'approximation des espérances. Cela nécessite l'évaluation de la fonction coûteuse Q aux $5N$ points de l'ensemble $\{b_{0_k}^*,\ 1 \leq k \leq 5\} \times \hat{\Xi}(\Omega)$.

Pour l'approximation point par point, le coefficient de prédictivité s'écrit

$$\mathrm{Q}^{2}_{\mathrm{ind}}{}^{*} = 1 - \frac{\sum_{k=1}^{5}\sum_{n=1}^{N}\left(Q(b_{0k}^{*},\xi_n) - \hat{Q}(b_{0k}^{*},\xi_n)\right)^2}{\sum_{k=1}^{5}\sum_{n=1}^{N}\left(Q(b_{0k}^{*},\xi_n) - \overline{Q}^{*}\right)^2} \in]-\infty, 1],$$

où

$$\overline{Q}^{*} = \frac{1}{5N}\sum_{k=1}^{5}\sum_{n=1}^{N} Q(b_{0k}^{*},\xi_n) \in \mathbb{R},$$

et l'erreur absolue moyenne des chiffres d'affaires normalisés s'écrit

$$\mathrm{MAE}^{*}_{\mathrm{ind}} = \frac{1}{5N}\sum_{k=1}^{5}\sum_{n=1}^{N}\frac{|Q(b_{0k}^{*},\xi_n) - \hat{Q}(b_{0k}^{*},\xi_n)|}{f(\mathbf{x}_0^{*},\xi_0)} \in \mathbb{R}_+.$$

Pour l'approximation des espérances, on considère les différences entre les fonctions $\mathcal{E} : b_0 \mapsto \mathbb{E}[Q(b_0,\hat{\Xi})]$ et $\hat{\mathcal{E}} : b_0 \mapsto \mathbb{E}[\hat{Q}(b_0,\hat{\Xi})]$. Le coefficient de prédictivité s'écrit alors

$$\mathrm{Q}^{2}_{\mathrm{esp}}{}^{*} = 1 - \frac{\sum_{k=1}^{5}\left(\mathcal{E}(b_{0k}^{*}) - \hat{\mathcal{E}}(b_{0k}^{*})\right)^2}{\sum_{k=1}^{5}\left(\mathcal{E}(b_{0k}^{*}) - \overline{\mathcal{E}}^{*}\right)^2},$$

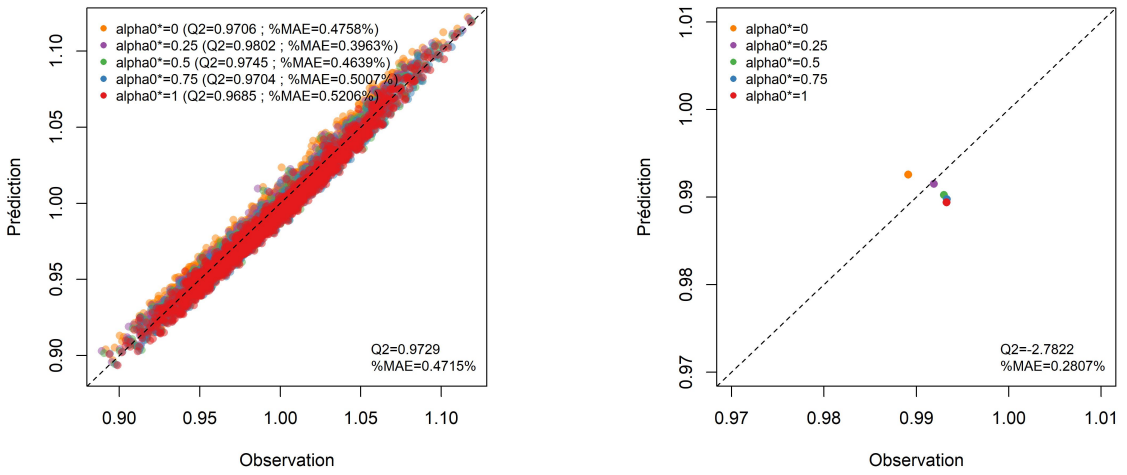
où

$$\overline{\mathcal{E}}^{*} = \frac{1}{5}\sum_{k=1}^{5}\mathcal{E}(b_{0k}^{*}) \in \mathbb{R},$$

et l'erreur absolue moyenne des chiffres d'affaires normalisés s'écrit

$$\mathrm{MAE}^{*}_{\mathrm{esp}} = \frac{1}{5}\sum_{k=1}^{5}\frac{|\mathcal{E}(b_{0k}^{*}) - \hat{\mathcal{E}}(b_{0k}^{*})|}{f(\mathbf{x}_0^{*},\xi_0)} \in \mathbb{R}_+.$$

Les résultats sont présentés à la figure 6.4. On observe que la qualité du métamodèle $\hat{Q}$ sur l'ensemble $\{b_{0k}^{*},\ 1 \leq k \leq 5\} \times \hat{\Xi}(\Omega)$ est acceptable pour l'approximation point par point et l'approximation des espérances car les erreurs absolues moyennes sont respectivement $\mathrm{MAE}^{*}_{\mathrm{ind}} = 0{,}47\%$ et $\mathrm{MAE}^{*}_{\mathrm{esp}} = 0{,}28\%$, des valeurs inférieures au MIP-gap relatif seuil de 2%. Le coefficient de prédictivité pour l'approximation point par point $\mathrm{Q}^{2}_{\mathrm{ind}}{}^{*} = 0{,}97$ est également supérieur au seuil arbitraire de 0,95. En revanche, celui pour l'approximation des espérances est $\mathrm{Q}^{2}_{\mathrm{esp}}{}^{*} = -2{,}78$, très inférieure au seuil arbitraire de 0,95. Néanmoins, les valeurs concernant l'approximation des espérances doivent être utilisées avec prudence car elles sont obtenues avec seulement cinq points. Mis à part le coefficient de prédictivité des espérances, les critères de qualité sont comparables (mais légèrement moins bons) à ceux obtenus sur l'ensemble de validation (voir Sous-section 5.4.4). Le métamodèle $\hat{Q}$ est donc de bonne qualité pour les solutions optimales $\mathbf{x}_{0k}^{*}$, pour $k \in \{1,\ldots,5\}$. La fonction objectif n'est donc pas sur-estimée pour des prochaines ventes *day-ahead* des solutions optimales. En revanche, rien ne garantit que la fonction objectif n'est pas sous-estimée pour des solutions ayant des prochaines ventes *day-ahead* différentes des $b_{0k}^{*} \in \mathcal{B}_0$, pour $k \in \{1,\ldots,5\}$.



(a) Approximation point par point

(b) Approximation des espérances

FIGURE 6.4 – Qualité du métamodèle pour les prochaines ventes *day-ahead* associées aux solutions optimales du problème d'optimisation probabiliste approché du cas d'étude avec les poids $\alpha_{0_k}^* = (k-1)/4$, pour $k \in \{1, \ldots, 5\}$.

6.2.2 Erreurs d'approximation de la fonction objectif

Nous cherchons dans cette sous-section à étudier l'optimalité réelle des solutions optimales $\mathbf{x}_{0_k}^*$, pour $k \in \{1, \ldots, 5\}$. Pour cela, nous comparons les fonctions objectifs probabilistes en fonction du poids $\alpha_0 \in [0, 1]$. La fonction objectif du problème d'optimisation probabiliste (6.2) est donnée, pour un poids $\alpha_0 \in [0, 1]$, par

$$\forall \mathbf{x}_0 \in \mathbb{R}^r, \quad F(\alpha_0, \mathbf{x}_0) = \alpha_0 f(\mathbf{x}_0, \xi_0) + (1 - \alpha_0)\mathbb{E}[Q(b(\mathbf{x}_0), \hat{\Xi})],$$

et son approximation est obtenue en substituant la fonction Q par le métamodèle $\hat{Q}$:

$$\forall \mathbf{x}_0 \in \mathbb{R}^r, \quad \hat{F}(\alpha_0, \mathbf{x}_0) = \alpha_0 f(\mathbf{x}_0, \xi_0) + (1 - \alpha_0)\mathbb{E}[\hat{Q}(b(\mathbf{x}_0), \hat{\Xi})].$$

Pour une solution admissible $\mathbf{x}_0 \in \mathcal{X}(\xi_0)$ donnée, les fonctions $\alpha_0 \mapsto F(\alpha_0, \mathbf{x}_0)$ et $\alpha_0 \mapsto \hat{F}(\alpha_0, \mathbf{x}_0)$ sont affines (leurs courbes représentatives sont des droites), dont les coefficients directeurs sont donnés par les différences respectives $f(\mathbf{x}_0, \xi_0) - \mathbb{E}[Q(b(\mathbf{x}_0), \hat{\Xi})]$ et $f(\mathbf{x}_0, \xi_0) - \mathbb{E}[\hat{Q}(b(\mathbf{x}_0), \hat{\Xi})]$.

La figure 6.5 présente, pour le cas d'étude, l'évolution de la fonction objectif F et son approximation $\hat{F}$, normalisées par rapport au chiffre d'affaires $f(\mathbf{x}_0^*, \xi_0)$ de la solution déterministe, pour les cinq solutions optimales $\mathbf{x}_{0_k}^*$, pour $k \in \{1, \ldots, 5\}$, en fonction du poids α_0. Comme la solution optimale $\mathbf{x}_{0_5}^*$ est la solution déterministe $\mathbf{x}_0^*$ de la sous-section 3.2.6, la courbe en rouge de la figure 6.5 est aussi celle présentée à la figure 3.13 de la sous-section 3.3.5 où la fonction objectif est étudiée pour la solution déterministe. Les valeurs du graphique obtenues pour $\alpha_0 = 0$ correspondent au second terme de la fonction objectif probabiliste (celui associé à la fonction Q ou au métamodèle $\hat{Q}$). Elles sont également données dans la troisième et la quatrième colonne du tableau 6.1 avec respectivement la fonction Q et le métamodèle $\hat{Q}$. De même, les valeurs du graphique obtenues pour $\alpha_0 = 1$ correspondent au premier terme de la fonction objectif probabiliste (celui qui ne dépend pas de la fonction Q ou du métamodèle $\hat{Q}$). Elles sont également données dans la deuxième colonne du tableau 6.1. Les points sur le graphique correspondent aux valeurs optimales des solutions optimales $\mathbf{x}_{0_k}^*$, pour $k \in \{1, \ldots, 5\}$, dont les valeurs sont données dans la dernière colonne du tableau 6.1. L'avant dernière colonne du tableau 6.1 correspond aux valeurs de la fonction objectif non approchée F des solutions optimales $\mathbf{x}_{0_k}^*$ pour leurs poids $\alpha_{0_k}^*$, pour $k \in \{1, \ldots, 5\}$.

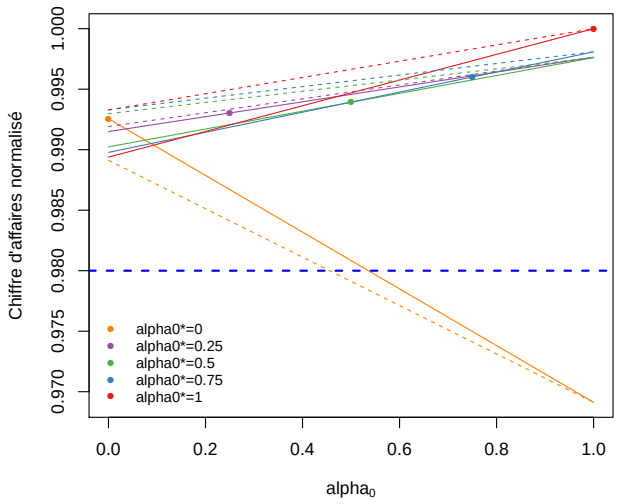


FIGURE 6.5 – Évolution de la fonction objectif F (lignes en tirets) et de son approximation $\hat{F}$ (lignes continues) en fonction du poids α_0 pour les solutions optimales du problème d'optimisation probabiliste approché du cas d'étude avec les poids $\alpha_{0_k}^* = (k-1)/4$, pour $k \in \{1, \ldots, 5\}$. Les valeurs sont normalisées par le chiffre d'affaires $f(\mathbf{x}_0^*, \xi_0)$ de la solution déterministe. Les points sur le graphique indiquent les valeurs optimales, i.e. celles obtenues avec la fonction objectif approché $\hat{F}$ pour les poids $\alpha_0 = \alpha_{0_k}^*$, pour $k \in \{1, \ldots, 5\}$. La ligne horizontale bleue indique le seuil en-dessous duquel les différences avec le chiffre d'affaires de la solution déterministe sont considérées comme significatives.

$\alpha_{0_k}^*$	$\frac{f(\mathbf{x}_{0_k}^*, \xi_0)}{f(\mathbf{x}_0^*, \xi_0)}$	$\frac{\mathbb{E}[Q(b_{0_k}^*, \hat{\Xi})]}{f(\mathbf{x}_0^*, \xi_0)}$	$\frac{\mathbb{E}[\hat{Q}(b_{0_k}^*, \hat{\Xi})]}{f(\mathbf{x}_0^*, \xi_0)}$	$\frac{F(\alpha_{0_k}^*, \mathbf{x}_{0_k}^*)}{f(\mathbf{x}_0^*, \xi_0)}$	$\frac{\hat{F}(\alpha_{0_k}^*, \mathbf{x}_{0_k}^*)}{f(\mathbf{x}_0^*, \xi_0)}$
0	0,9691	0,9891	0,9926	0,9891	0,9926
0,25	0,9976	0,9919	0,9915	0,9933	0,9930
0,5	0,9976	0,9930	0,9902	0,9953	0,9939
0,75	0,9981	0,9933	0,9898	0,9969	0,9960
1,0	1,0000	0,9933	0,9894	1,0000	1,0000

TABLEAU 6.1 – Décomposition des deux termes de la fonction objectif probabiliste pour les solutions optimales du problème d'optimisation probabiliste du cas d'étude avec les poids $\alpha_{0_k}^* = (k-1)/4$, pour $k \in \{1, \ldots, 5\}$. Les valeurs sont données à 10^{-4} près.

On observe que les droites sont toutes croissantes sauf pour la solution optimale $\mathbf{x}_{0_1^*}$ associée au poids $\alpha_0 = \alpha_{0_1^*} = 0$ (cas où le scénario déterministe n'est pas utilisé dans la fonction objectif). Cela signifie que pour les quatre autres solutions optimales $\mathbf{x}_{0_k^*}$, pour $k \in \{2, \ldots, 5\}$, le second terme de la fonction objectif probabiliste est pénalisant par rapport à la fonction objectif déterministe. En revanche, pour la solution optimale $\mathbf{x}_{0_1^*}$, le second terme de la fonction objectif est avantageux par rapport à la fonction objectif déterministe. Cette différence peut s'expliquer par le fait que le scénario déterministe n'est pas utilisé dans le calcul de l'optimalité de la solution optimale $\mathbf{x}_{0_1^*}$, si bien que l'optimisation favorise uniquement les solutions par rapport à leur réoptimisation pour les scénarios probabilistes. Cela montre l'importance, pour le cas d'étude, de tenir compte du chiffre d'affaires déterministe des solutions dans la fonction objectif probabiliste (voir Sous-section 3.3.4).

Mise à part celle associée à la solution optimale $\mathbf{x}_{0_1^*}$, on observe également que toutes les droites sont comprises dans l'intervalle [0,98, 1,02]. Cela signifie que, pour tout $\alpha_0 \in [0, 1]$, les valeurs des fonctions objectifs probabilistes associées aux quatre solutions optimales $\mathbf{x}_{0_k^*}$, pour $k \in \{2, \ldots, 5\}$, sont équivalentes au chiffre d'affaires de la solution déterministe, au MIP-gap relatif seuil de 2% près. Cela est dû au fait que, pour tout $k \in \{2, \ldots, 5\}$, $\mathbb{E}[Q(b_{0_k^*}, \hat{\Xi})] \approx \mathbb{E}[\hat{Q}(b_{0_k^*}, \hat{\Xi})] \approx f(\mathbf{x}_{0_k^*}, \xi_0)$ (au MIP-gap relatif seuil de 2% près, voir Tableau 6.1), impliquant

$$\forall \alpha_0 \in [0, 1], \quad F(\alpha_0, \mathbf{x}_{0_k^*}) \approx \hat{F}(\alpha_0, \mathbf{x}_{0_k^*}) \approx f(\mathbf{x}_{0_k^*}, \xi_0).$$

Le poids α_0 n'a donc pas d'impact sur la fonction objectif probabiliste appliquée aux solutions optimales $\mathbf{x}_{0_k^*}$, pour $k \in \{2, \ldots, 5\}$. Cela généralise donc les observations faites à la sous-section 3.3.5 où la fonction objectif probabiliste est étudiée pour la solution déterministe. Ainsi la solution optimale $\mathbf{x}_{0_k^*}$, où $k \in \{2, \ldots, 5\}$, est aussi optimale pour $\alpha_{0_{k'}^*}$, avec $k' \in \{2, \ldots, 5\}$, $k' \neq k$: les solutions optimales $\mathbf{x}_{0_k^*}$, pour $k \in \{2, \ldots, 5\}$, ont une optimalité équivalente (au MIP-gap relatif seuil de 2% près). En particulier, la solution déterministe $\mathbf{x}_0^* = \mathbf{x}_{0_5^*}$ est aussi une solution optimale du problème d'optimisation probabiliste (6.4) avec les poids $\alpha_0 = \alpha_{0_k^*} = (k-1)/4$, pour $k \in \{2, 3, 4\}$. Cette équivalence entre les solutions optimales $\mathbf{x}_{0_k^*}$, pour $k \in \{2, \ldots, 5\}$, est également la raison pour laquelle leurs valeurs optimales (les points sur la figure 6.5) ne sont pas toujours supérieures aux valeurs des autres droites pour le poids $\alpha_0 = \alpha_{0_k^*}$.

Les différences négligeables entre le chiffre d'affaires déterministe $f(\mathbf{x}_{0_k^*}, \xi_0)$ et les moyennes des chiffres d'affaires réoptimisés $\mathbb{E}[Q(b_{0_k^*}, \hat{\Xi})]$ et $\mathbb{E}[\hat{Q}(b_{0_k^*}, \hat{\Xi})]$, pour $k \in \{2, \ldots, 5\}$, rejoignent les mêmes observations faites sur le jeu de données (voir Sous-section 5.3.3) : les fonctions $b_0 \mapsto \mathbb{E}[Q(b_0, \hat{\Xi})]$ et $b_0 \mapsto \mathbb{E}[\hat{Q}(b_0, \hat{\Xi})]$ sont globalement constantes (au MIP-gap relatif seuil de 2% près) sur $\hat{\mathcal{B}}_0 \cup \{b_{0_k^*},\ 2 \leq k \leq 5\}$ (et pas seulement sur $\hat{\mathcal{B}}_0$ comme observé à la sous-section 5.3.3), comme l'illustre la figure 6.6. Cela peut être un signe que les fonctions $b_0 \mapsto \mathbb{E}[Q(b_0, \hat{\Xi})]$ et $b_0 \mapsto \mathbb{E}[\hat{Q}(b_0, \hat{\Xi})]$ sont globalement constantes (au MIP-gap relatif seuil de 2% près) sur tout l'ensemble $\mathcal{B}_0$, si bien que le second terme de la fonction objectif probabiliste est « transparent » pour l'optimisation (il ne dépend pas des variables de décision). Dans ce cas, l'optimisation en univers probabiliste serait équivalent à l'optimisation en univers déterministe. En revanche, comme l'illustre la figure 6.6, le support des distributions des variables aléatoires réelles $Q(b_0, \hat{\Xi})$ pour $b_0 \in \hat{\mathcal{B}}_0 \cup \{b_{0_k^*},\ 1 \leq k \leq 5\}$ sort de la zone de tolérance au MIP-gap relatif seuil de 2% autour du chiffre d'affaires de la solution déterministe. Autrement dit, les incertitudes ajoutent un risque financier non négligeable mais, en moyenne, elles n'ont pas un impact significatif sur le chiffre d'affaires pour le cas d'étude. Si cette observation venait à se vérifier sur d'autres cas d'étude, alors il serait intéressant de prendre en compte une aversion face au risque dans la formulation mathématique (6.2).

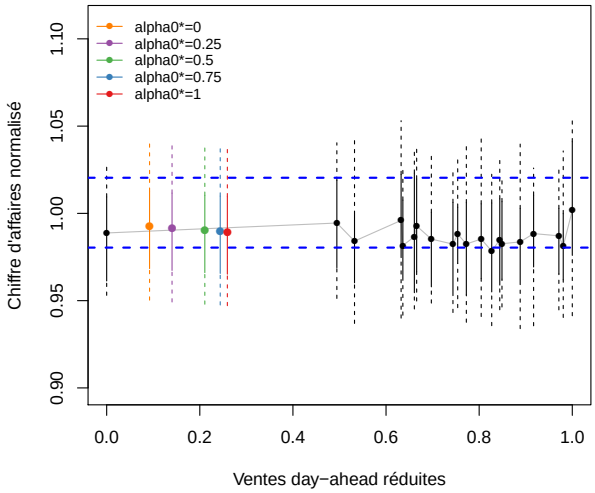


FIGURE 6.6 – Boîtes à moustaches des variables aléatoires réelles $Q(b_0, \hat{\Xi})$ pour les prochaines ventes *day-ahead* $b_0 \in \hat{\mathcal{B}}_0$ (en noir) et $b_0 \in \{b_{0k}^*,\ 1 \leq k \leq 5\}$ (en couleurs) représentées par leurs valeurs réduites. Les extrémités des segments continus sont les premiers et troisièmes quartiles et ceux des segments en pointillés sont les premiers et neuvièmes déciles. Les points correspondent aux moyennes des distributions. Les lignes horizontales bleues en tirets représentent la zone de tolérance au MIP-gap relatif seuil de 2% autour du chiffre d'affaires de la solution déterministe.

La fonction objectif probabiliste F et son approximation $\hat{F}$ pour la solution optimale $\mathbf{x}_{01}^*$ associée au poids $\alpha_0 = \alpha_{01}^* = 0$ sont particulières car, en plus d'être décroissantes en fonction de α_0, elles sortent de l'intervalle $[0{,}98, 1{,}02]$ à partir d'environ $\alpha_0 \approx 0{,}5$. Cela signifie encore que, contrairement aux quatre autres solutions optimales, $\mathbf{x}_{01}^*$ est très spécifique au cas $\alpha_0 = \alpha_{01}^* = 0$ pour lequel elle est optimisée. Dès que l'on prend en compte le scénario déterministe pour le calcul de la fonction objectif, alors la solution $\mathbf{x}_{01}^*$ devient sous-optimale. Cela est encore dû au fait que la solution $\mathbf{x}_{01}^*$ est optimisée sans tenir compte du scénario déterministe dans la fonction objectif.

Pour le cas d'étude, le choix du poids α_0 pour l'optimisation probabiliste (6.4) semble donc avoir peu d'importance. Si on choisit le poids $\alpha_0 = 0{,}74$ obtenu avec le *cluster* central de la sous-section 4.3.6, alors on devrait avoir des résultats comparables à la solution optimale $\mathbf{x}_{04}^*$ optimisée pour $\alpha_0 = \alpha_{04}^* = 0{,}75$.

6.2.3 Temps de calcul

Pour une utilisation de l'optimiseur probabiliste dans le processus opérationnel, la problématique des temps de calcul est cruciale. Dans le cadre de la thèse, le temps de calcul limite au bout duquel l'optimiseur doit fournir une solution est de 1 h.

Le temps de calcul de la méthodologie proposée est la somme du temps de calcul de la phase de pré-traitement et celui de la résolution numérique du problème MILP approché (6.4).

Les temps de calcul des résolutions numériques du problème MILP approché (6.4) pour les poids $\alpha_0 = \alpha_{0k}^* = (k-1)/4$, pour $k \in \{1, \ldots, 5\}$, sont présentés dans le tableau 6.2. On remarque qu'ils sont comparables à celui de la solution déterministe, ce qui n'est pas étonnant car le problème MILP (6.4) est très similaire à celui du problème MILP en univers déterministe, grâce à la linéarité du métamodèle $\hat{Q}$. Le cas $\alpha_0 = \alpha_{02}^* = 0{,}25$ est néanmoins associé à un temps de calcul presque moitié plus élevé que celui de la solution déterministe, ce qui n'est pas si négligeable en pratique.

$\alpha_{0_k}^*$	*Temps de calcul normalisé*
0	1,0680
0,25	1,4891
0,5	1,0547
0,75	0,9434
1	1,0000

TABLEAU 6.2 – Temps de calcul des solutions optimales $\mathbf{x}_{0_k}^*$, pour $k \in \{1, \ldots, 5\}$, par rapport au temps de calcul de la solution déterministe.

Il reste à considérer le temps de calcul de la phase de pré-traitement qui est principalement déterminé par le temps de calcul de l'étape d'évaluation de la fonction coûteuse Q sur le plan d'expérience pour former le jeu de données (étape (c) de la figure 6.1). Les autres étapes de la phase de pré-traitement ont en effet un temps de calcul négligeable.

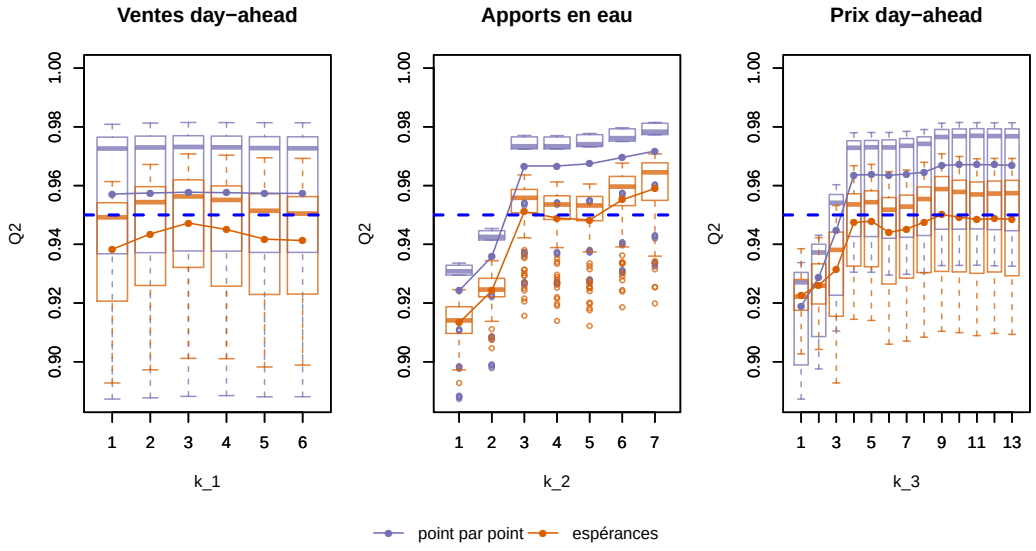
Le temps de calcul de l'étape associée à la création du jeu de données dépend du nombre D de points du plan d'expérience et de l'accès au calcul parallèle (mémoire distribuée car les évaluations de la fonction Q sur le plan d'expérience sont indépendantes). Plus précisément, le nombre maximal de points du plan d'expérience qu'il faut considérer pour garder un temps de calcul de la création du jeu de données inférieur à $\Delta t_{\text{lim}} > 0$ est

$$D_{\max} = n_c \frac{\Delta t_{\text{lim}}}{\Delta t_{\text{eval}}},$$

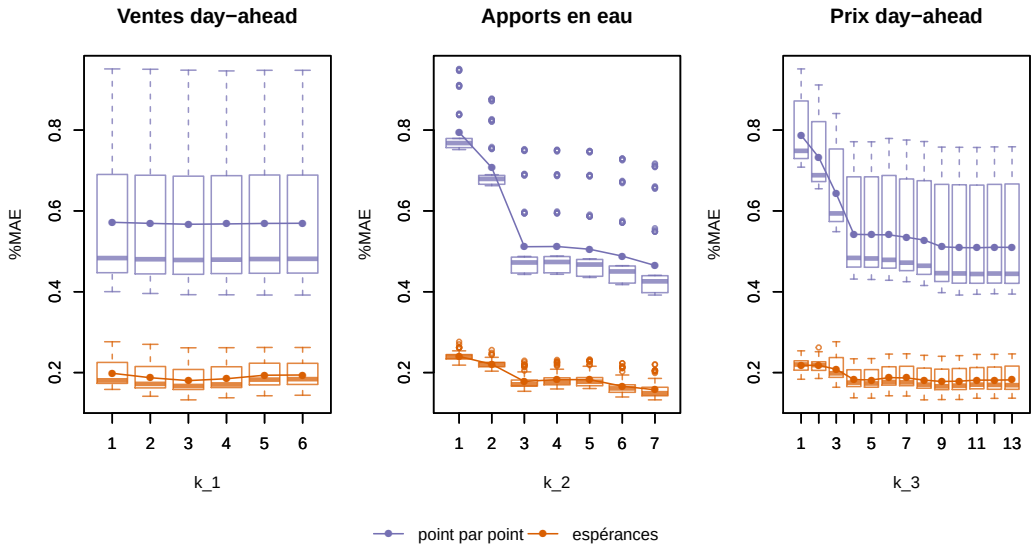
où $\Delta t_{\text{eval}} > 0$ est le temps de calcul moyen d'une évaluation de la fonction Q et $n_c \in \mathbb{N}^*$ le nombre d'unités de calcul (cœurs) que l'on peut lancer en parallèle.

Pour le cas d'étude, nous avons utilisé $D = 1\,000$ points, ce qui ne permet pas un temps de calcul acceptable pour une utilisation en opérationnel sans utiliser plusieurs centaines d'unités de calcul en parallèle. Il faut donc chercher à réduire le nombre D de points du plan d'expérience, sachant qu'il doit également être suffisamment élevé pour calibrer les paramètres du métamodèle. Comme le nombre de paramètres du métamodèle dépend de la dimension des données d'entrée réduites (nombre de directions principales retenues par la réduction par analyse en composantes principales), il faut trouver un compromis entre le nombre D de points du plan d'expérience et la dimension des données d'entrée réduites. Ce compromis ne doit pas détériorer la qualité du métamodèle mesurée par les critères définis à la sous-section 5.4.3 du chapitre précédent.

La figure 6.7 présente les critères de qualité du métamodèle en fonction du nombre de directions principales retenues de chaque donnée d'entrée. Sans surprise, on observe globalement que la qualité est meilleure lorsque le nombre de directions principales retenues augmente. Avec le nombre maximal de directions principales retenues, on retrouve les résultats de la sous-section 5.4.3 du chapitre précédent. On observe également que les erreurs absolues moyennes sont acceptables pour toutes les dimensions des données d'entrée (car elles sont inférieures à 2%). En revanche, les coefficients de prédictivité sont acceptables uniquement lorsque la dimension réduite des apports en eau est supérieure à 3 et celle des prix *day-ahead* est supérieure à 4. Pour les prochaines ventes *day-ahead*, l'erreur absolue moyenne et le coefficient de prédictivité ne dépendent pas de la dimension réduite. Cela n'est pas surprenant car la fonction Q s'avère presque indépendante des prochaines ventes *day-ahead* (voir Sous-section 5.4.4).



(a) Coefficient de prédictivité



(b) Erreur absolue moyenne

FIGURE 6.7 – Évolution des critères de qualité point par point et des espérances du métamodèle en fonction du nombre de directions principales retenues pour le cas d'étude. Les boîtes à moustaches représentent les distributions des critères de qualité en fixant la dimension d'une donnée d'entrée mais en faisant varier la dimension des autres. Les lignes horizontales bleues en tirets représentent le seuil arbitraire à 0,95 sur le coefficient de prédictivité. Les critères sont calculés sur l'ensemble de validation.

La figure 6.8 présente les critères de qualité en fonction de la dimension totale des données d'entrée réduites, en combinant les résultats précédents. Sans surprise, on observe que la qualité est meilleure lorsque la dimension totale des données d'entrée augmente. Avec la dimension totale maximale de 26, on retrouve les résultats de la sous-section 5.4.3 du chapitre précédent. On observe également que les erreurs absolues moyennes sont acceptables pour toutes les dimensions totales des données d'entrée réduites (car elles sont inférieures à 2%). En revanche, les coefficients de prédictivité sont acceptables à partir d'une dimension totale égale à 12 pour l'approximation point par point et 17 pour l'approximation en espérance (sur les 26 dimensions réduites utilisées pour l'apprentissage du métamodèle dans la sous-section 5.4.4 du chapitre précédent).

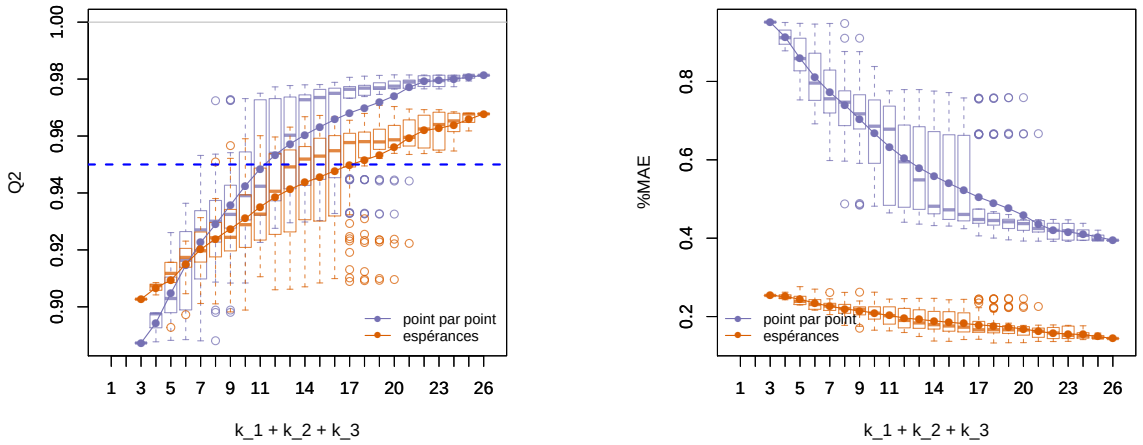
La figure 6.9 présente les critères de qualité en fonction de la taille de l'ensemble d'apprentissage sans toucher à la dimension des données d'entrée réduites (traits pleins) ou en considérant uniquement une dimension totale égale à 12 (2 pour les prochaines ventes *day-ahead* sur les 6, 4 pour les apports en eau sur les 7 et 6 pour les prix sur les 13). Sans surprise, on observe globalement que la qualité est meilleure lorsque la taille de l'ensemble d'apprentissage augmente puis elle stagne à partir d'une certaine taille de l'ensemble d'apprentissage (entre 100 et 200 points). Avec la taille maximale égale à 600, on retrouve les résultats de la sous-section 5.4.3 du chapitre précédent. On observe également que les erreurs absolues moyennes sont acceptables avec seulement 30 points dans l'ensemble d'apprentissage avec les données d'entrée réduites de dimension 26 et un peu moins avec celles de dimension 12. Les résultats sont analogues concernant les coefficients de prédictivité point par point. En revanche, pour obtenir un coefficient de prédicitivité sur les espérances supérieur à 0,95, il faut au moins entre 100 et 150 points dans l'ensemble d'apprentissage.

Ainsi, pour le cas d'étude, le jeu de données doit contenir au minimum une centaine d'évaluations de la fonction coûteuse Q. Le nombre minimal d'unités de calcul en parallèle qu'il faut avoir à disposition pour garder un temps de calcul de la création du jeu de données inférieur à $\Delta t_{\text{lim}} > 0$ est donc

$$n_c \approx 100 \frac{\Delta t_{\text{eval}}}{\Delta t_{\text{lim}}},$$

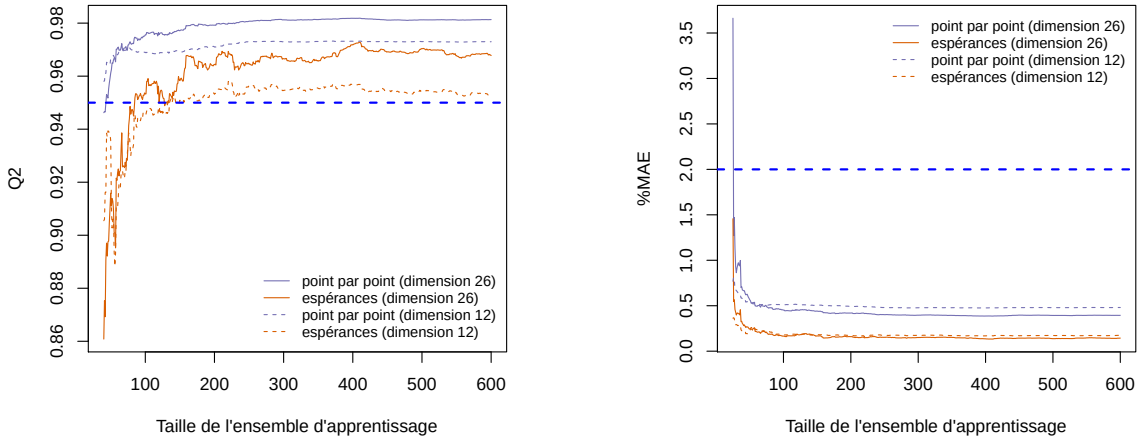
où $\Delta t_{\text{eval}} > 0$ est le temps de calcul moyen d'une évaluation. En pratique, compte tenu du court délai de 1 h par rapport au temps de calcul de la résolution du problème d'optimisation en univers déterministe, le rapport $\Delta t_{\text{eval}}/\Delta t_{\text{lim}}$ n'est pas si petit. Le nombre d'unités de calcul en parallèle peut alors engendrer un coût financier qu'il faudrait considérer dans la comparaison des gains financiers entre l'optimiseur déterministe et l'optimiseur probabiliste.

Il convient également de noter que cette étude de la qualité du métamodèle en fonction de la taille de l'ensemble d'apprentissage est effectuée *a posteriori*. En pratique, le court délai de 1 h ne permet pas de faire une telle étude à chaque nouvelle situation dans un contexte opérationnel. Il serait alors indispensable de construire une méthodologie par enrichissement du jeu de données. Partant d'un petit plan d'expérience, des points supplémentaires seraient ajoutés progressivement au plan d'expérience (voire des dimensions supplémentaires pour les données d'entrées réduites) tant que la qualité du métamodèle (évaluée dans ce cas par validation croisée) n'est pas acceptable. On obtiendrait alors la configuration minimale du jeu de données pour assurer une qualité suffisante du métamodèle, limitant le nombre d'évaluations superflues de la fonction coûteuse Q et donc le temps de calcul.



(a) Coefficient de prédictivité (b) Erreur absolue moyenne

FIGURE 6.8 – Évolution des critères de qualité point par point et des espérances du métamodèle en fonction de la dimension totale des données d'entrée pour le cas d'étude. Les boîtes à moustaches représentent les distributions portées par toutes les combinaisons des dimensions de chaque donnée d'entré qui donnent la même dimension totale. La ligne horizontale bleue en tirets représente le seuil arbitraire à 0,95 sur le coefficient de prédictivité. Les critères sont calculés sur l'ensemble de validation.



(a) Coefficient de prédictivité (b) Erreur absolue moyenne

FIGURE 6.9 – Évolution des critères de qualité point par point et des espérances du métamodèle en fonction de la taille de l'ensemble d'apprentissage. L'ensemble de validation reste inchangé et l'ensemble d'apprentissage est enrichi en ajoutant aléatoirement un à un les 600 points de l'ensemble d'apprentissage initial. Les lignes horizontales bleues en tirets correspondent au seuil d'acceptabilité de la qualité.

6.2.4 La méthodologie proposée est-elle satisfaisante ?

Les résultats de cette section montrent que la méthodologie proposée pour faire passer l'optimiseur de l'univers déterministe à l'univers probabiliste n'est pas satisfaisante en l'état pour le cas d'étude à cause des deux limites principales suivantes :

— le terme ajouté dans la fonction objectif probabiliste (valeur optimale du second niveau) semble négligeable par rapport au chiffre d'affaires déterministe, si bien qu'avec un poids α_0 non nul, les problèmes d'optimisation probabiliste et déterministe sont équivalents au MIP-gap relatif seuil de 2% près,

— le temps de calcul de la phase de pré-traitement ne permet pas de respecter la limite de temps de 1 h pour une utilisation en opérationnel, sans utiliser des gros moyens de calcul parallèle.

La première limite est liée à la formulation mathématique du problème d'optimisation en univers probabiliste, alors que la seconde relève de la résolution numérique à l'aide d'un métamodèle dans une phase de pré-traitement. Des études supplémentaires sur d'autres cas d'étude restent néanmoins nécessaires pour déterminer si ces résultats sont spécifiques au cas d'étude considéré ou s'ils se généralisent (voir Sous-section 6.3.1). Il convient également de rappeler que le cas d'étude ne prend pas en compte les incertitudes sur les productions variables, ce qui pourrait expliquer la première limite ci-dessus. Cette simplification du cas d'étude réduit en effet l'intérêt de la formulation mathématique que nous avons proposée dans la thèse, qui est basée sur la gestion conjointe de la chaîne hydroélectrique et des actifs de production variable. Les études supplémentaires nécessiteraient donc la prise en compte des prévisions probabilistes des productions variables.

Par ailleurs, la méthodologie proposée repose sur des hypothèses et simplifications qui limitent son utilisation.

— La formulation mathématique du problème d'optimisation probabiliste en deux niveaux est surtout pertinente lorsque la solution peut être totalement réoptimisée en période infrajournalière (i.e. la période où la production a déjà été vendue sur le marché *day-ahead*). En pratique, comme les prix de règlement des écarts sont très incertains, il est souvent préférable de faire un compromis entre les modifications du programme de production et les volumes prévus des écarts. Ce compromis repose sur une gestion manuelle ou sur des heuristiques, mais pas nécessairement sur une réoptimisation de la solution au sens du problème de planification de la production. Dans ce cas, la fonction Q ne permet pas une bonne estimation des chiffres d'affaires que l'on peut atteindre à partir d'une décision de ventes *day-ahead* et d'un scénario. Néanmoins, la stratégie de valorisation en période infrajournalière pourrait évoluer vers une réoptimisation totale de la solution grâce à l'utilisation d'un optimiseur efficace et à des prévisions performantes des prix de règlement des écarts.

— De manière analogue, la formulation mathématique du problème d'optimisation probabiliste en deux niveaux est surtout pertinente lorsque la décision des prochaines ventes *day-ahead* est définitive, i.e. pour les planifications qui ont lieu en fin de matinée. Par exemple, pour les planifications durant l'après-midi, la décision des prochaines ventes *day-ahead* (à effectuer le lendemain avant midi portant sur la production du sur-lendemain) n'est pas définitive car elle pourra être modifiée le lendemain matin. Dans ce cas, il serait peut être souhaitable d'utiliser le problème d'optimisation déterministe ou de porter la dynamique entre les deux niveaux sur d'autres variables de décision que les prochaines ventes *day-ahead* (le programme de production sur les premières heures par exemple). Néanmoins, la planification en fin de matinée est souvent la plus cruciale dans le processus opérationnel (du fait notamment des décisions des offres de ventes *day-ahead*). Nous pourrions alors utiliser l'optimiseur probabiliste défini dans la thèse pour la planification en fin

de matinée et l'optimiseur déterministe pour les autres planifications où les incertitudes ont *a priori* moins d'impact.

— L'approximation $\hat{\mathcal{B}}_0 = \{b_1, \ldots, b_M\}$ de l'ensemble $\mathcal{B}_0$ des prochaines ventes *day-ahead* possibles obtenue par analogie ne permet pas de respecter le critère de compatibilité avec le scénario déterministe $\hat{\mathcal{B}}_0 \subset \mathcal{B}_0$. Pour le cas d'étude, on a même $\hat{\mathcal{B}}_0 \cap \mathcal{B}_0 = \emptyset$. Certaines ventes *day-ahead* de $\hat{\mathcal{B}}_0$ peuvent donc être trop éloignées de ce qu'il est réellement possible de produire et les bornes minimales et maximales sur les écarts peuvent ne pas être respectées. La fonction Q ne peut alors pas être évaluée sur le plan d'expérience (elle donnerait la valeur $-\infty$). Pour pallier cette difficulté, les contraintes sur les écarts minimums et maximums sont supprimées dans le problème d'optimisation du second niveau (6.3) sur la période de livraison des prochaines ventes *day-ahead* (voir Sous-section 5.2.4). Néanmoins, cette simplification enlève de l'importance aux contraintes $\{b(\mathbf{x}_n) = b_0,\ \mathbf{x}_n \in \mathcal{X}(\xi_n)\}$, pour $(b_0, \xi_n) \in \mathcal{B}_0 \times \hat{\Xi}(\Omega)$, du problème MILP (6.3). Au lieu de supprimer ces contraintes sur les écarts, il serait plus souhaitable de revoir leur définition, en considérant par exemple les bornes minimales et maximales obtenues à partir d'une première estimation des écarts que l'on subirait si l'on ne faisait rien (par simulation, donc avec des temps de calcul abordables).

6.3 Perspectives générales

6.3.1 Validation robuste de la méthodologie

La méthodologie proposée a été testée sur un seul cas d'étude qui rejoue une planification passée de la production hydroélectrique du Rhône, en ne considérant que la partie Haut-Rhône (les 6 premiers aménagements hydroélectriques sur les 18 au total) sans production variable. Pour effectuer une validation plus robuste de la méthodologie, il convient donc de la tester sur d'autres cas d'étude plus récents en considérant l'ensemble du Rhône entier et les productions variables du périmètre d'équilibre de CNR. Les nouveaux historiques des prévisions probabilistes qui ont été construits par CNR en parallèle de la thèse permettraient de former ces nouveaux cas d'étude (voir Sous-section 2.4.4).

La validation consisterait à comparer les temps de calcul et les chiffres d'affaires de la méthodologie d'optimisation en univers probabiliste par rapport à celle en univers déterministe (comme ce qui a été effectué à la section 6.2) pour tous les cas d'étude. Pour que la validation soit suffisamment robuste, il faudrait que les cas d'étude testés couvrent une grande partie de la diversité des situations. En plus de la validation des performances de la méthodologie d'optimisation probabiliste, les résultats sur les cas d'étude permettraient d'effectuer une analyse de sensibilité pour déterminer les paramètres des situations qui impactent le plus les performances de la méthodologie.

Pour tenir compte du caractère probabiliste de la méthodologie dans la validation, les solutions optimales des cas d'étude doivent être confrontées avec les réalisations des apports en eau, des prix et des productions variables. La fiabilité des chiffres d'affaires prévus lors des optimisations pourrait alors être comparée à celle des prévisions probabilistes. Si la fiabilité des chiffres d'affaires prévus était avérée, alors cela validerait la méthodologie d'optimisation probabiliste.

Les cas d'études historiques poseraient néanmoins des difficultés pour la validation car ils seraient définis avec des conditions initiales imposées par le programme de production effectivement réalisé jusqu'alors. Or, ce programme effectivement réalisé est lui-même le résultat d'une optimisation effectuée en opérationnel. La validation sur les cas d'études risquerait alors de limiter le champ d'expression de la méthodologie d'optimisation testée, laquelle se verrait imposer des conditions initiales qui ne correspondraient pas forcément à celles qui seraient advenues si cette

méthodologie avait été utilisée lors des planifications antérieures. Pour pallier ces défauts, un protocole de validation sur un horizon roulant (*rolling*) pourrait être mis en place. Le principe serait d'estimer le chiffre d'affaires effectif cumulé que l'on obtiendrait, sur une période passée de plusieurs mois, en utilisant toujours la même méthodologie d'optimisation et en calquant le processus opérationnel confronté au scénario observé en temps réel sur l'ensemble de la période. En pratique, cette démarche consisterait à effectuer itérativement toutes les planifications successives en considérant les conditions initiales données par le résultat des itérations précédentes ainsi que les mises à jour des observations et des prévisions (probabilistes ou déterministes, selon la méthodologie d'optimisation testée). À chaque itération, la planification serait portée sur la même fenêtre temporelle que celle qui serait utilisée en opérationnel. Cette démarche permettrait de construire itérativement une solution sur toute la période considérée puis de calculer le chiffre d'affaires effectif cumulé. Les méthodologies d'optimisation pourraient alors être comparées par rapport à leur chiffre d'affaires effectif cumulé sur la même période. La mise en place d'un tel protocole de validation ne serait cependant pas trivial car le processus opérationnel et l'expertise des opérateurs de gestion de la production sont très difficiles à approcher de manière automatique sans faire d'hypothèses simplificatrices, impactant ainsi l'estimation du chiffre d'affaires effectif cumulé.

6.3.2 Corrélations entre les données d'entrée

La formulation mathématique du problème d'optimisation probabiliste (6.2) a l'avantage d'être valable pour toute variable aléatoire $\hat{\Xi}$ discrète et finie. En particulier, elle est aussi valable lorsque la loi de la variable aléatoire $\hat{\Xi}$ tient compte des corrélations entre les apports en eau, les prix *day-ahead* et les productions variables. En revanche, la construction du métamodèle $\hat{Q}$ dans la phase de pré-traitement repose sur l'hypothèse d'indépendance entre les apports en eau, les prix *day-ahead* et les productions variables qui forment les scénarios. Plus précisément, cette hypothèse d'indépendance intervient dans :

- la réduction de la dimension des données d'entrée fonctionnelles par analyse en composantes principales (voir Sous-section 4.2.2),
- la définition des distances entre les données d'entrée fonctionnelles (voir Sous-section 4.2.3),
- la réduction des scénarios (voir Section 4.3),
- la construction du plan d'expérience (voir Section 5.3).

Or, cette hypothèse d'indépendance est discutable car les apports en eau, les prix *day-ahead* et les productions variables sont légèrement corrélés en pratique (voir Annexe A.1). Il serait donc intéressant d'étendre la méthodologie pour retirer l'hypothèse d'indépendance. Bien sûr, pour cela, il faudrait que la modélisation des incertitudes tiennent compte des corrélations. Une manière d'y parvenir à partir des prévisions d'ensemble obtenues séparément serait de redéfinir les probabilités de chaque combinaison des membres des prévisions d'ensemble entre eux (au lieu de les considérer comme équiprobables). Ces nouvelles probabilités pourraient être calculées à partir d'une copule estimée empiriquement sur un historique passé [Camal *et al.*, 2019], de manière à représenter au mieux la loi de probabilité jointe des scénarios. Une autre manière serait de définir directement les prévisions d'ensemble de manière multivariée.

La réduction de la dimension des scénarios sans l'hypothèse d'indépendance reviendrait à projeter orthogonalement les scénarios sur une famille orthonormale de directions principales communes aux apports en eau, aux prix *day-ahead* et aux productions variables, au lieu de réduire séparément la dimension de chacun de ces trois types de paramètres. La difficulté serait alors de mettre à l'échelle la matrice de données commune pour rendre les valeurs comparables dans l'analyse en composantes principales, sans détériorer la structure fonctionnelle des trois types de paramètres. Une autre difficulté est que le nombre de colonnes de la matrice de données serait très grande (somme des dimensions initiales des trois types de paramètres qui forment les

scénarios), pouvant rendre les calculs numériques de l'analyse en composantes principales peu efficaces en pratique.

La réduction de la dimension des scénarios sans l'hypothèse d'indépendance permettrait de définir la distance de Mahalanobis sur les scénarios (distance euclidienne sur les composantes principales). En revanche, la tâche serait plus ardue pour définir une distance sur les scénarios sans réduction de la dimension (il faudrait appliquer une mise à l'échelle qui fasse sens). La définition d'une telle distance pourrait servir à considérer une variable d'analogie multivariée (notamment aussi pour considérer les apports en eau de chaque affluent au lieu des seuls apports en eau cumulés) pour l'échantillonnage par analogie des prochaines ventes *day-ahead* (voir Section 5.2). Cette distance multivariée sur les scénarios servirait également pour la réduction des scénarios.

Sans l'hypothèse d'indépendance, le plan d'expérience serait construit en deux dimensions : une dimension pour la représentation scalaire des prochaines ventes *day-ahead* (utilisation de la moyenne, voir Sous-section 5.3.2) et une seconde dimension pour la représentation scalaire des scénarios qui serait donnée par la première composante principale commune (projection sur la première direction principale commune). En particulier, l'échantillonnage par hypercube latin serait effectué en dimension 2 (au lieu de la dimension 4 dans la sous-section 5.3.2), faisant gagner légèrement en temps de calcul (l'échantillonnage par hypercube latin n'est pas l'étape la plus gourmande en temps de calcul, même à 4 dimensions).

6.3.3 Choix du métamodèle

Le métamodèle $\hat{Q}$ est choisi parmi les fonctions linéaires. Or, la fonction Q que le métamodèle $\hat{Q}$ remplace est linéaire par morceaux par rapport aux prochaines ventes *day-ahead* avec un nombre fini de points de discontinuité (voir Sous-section 5.4.1). Même si le modèle linéaire donne des résultats très acceptables pour le cas d'étude (voir Sous-section 5.4.4), il peut s'avérer insuffisant pour d'autres cas d'étude. Il serait alors intéressant de chercher le métamodèle $\hat{Q}$ parmi les fonctions linéaires par morceaux (par exemple, en effectuant une régression sur des splines bien choisies [Friedman *et al.*, 2001]). Néanmoins, le risque est d'augmenter le temps de calcul de l'apprentissage du métamodèle et de la résolution numérique du problème MILP approché (6.4) (le problème serait toujours MILP mais l'ajout des variables de décision binaires pour caractériser les morceaux du métamodèle complexifierait le problème d'optimisation).

Une limite de la méthodologie proposée est que la création du jeu de données dans la phase de pré-traitement demande un temps de calcul important (ou des gros moyens en calcul parallèle), voir Sous-section 6.2.4. Une manière de contourner cette difficulté serait de construire le jeu de données, non pas à partir de la fonction coûteuse Q mais à partir de la valeur optimale dégradée $\tilde{Q}$ d'un problème d'optimisation approché obtenu à partir du problème MILP (6.3) en relâchant certaines contraintes ou en relaxant les variables discrètes. La fonction $\tilde{Q}$ serait alors plus rapide à évaluer que la fonction Q. Le métamodèle $\hat{Q}$, encore plus rapide à évaluer, serait donc construit à partir du jeu de données obtenu avec la fonction $\tilde{Q}$ évaluée sur le plan d'expérience. Une analyse de la qualité du métamodèle par rapport à la fonction Q pour plusieurs cas d'étude permettrait d'évaluer la pertinence d'une telle approche en pratique (car les erreurs d'approximation se cumulent à deux endroits : le passage de Q à $\tilde{Q}$ et le passage de $\tilde{Q}$ à $\hat{Q}$).

Une autre limite de la méthodologie proposée est que l'échantillonnage par analogie des prochaines ventes *day-ahead* pour obtenir l'approximation $\hat{\mathcal{B}}_0$ de l'ensemble $\mathcal{B}_0$ n'est pas satisfaisante (voir Sous-section 6.2.4). Une manière de contourner cette difficulté serait de ne plus utiliser la phase de pré-traitement mais l'une des deux approches suivantes :

— une approche de décomposition (comme la SDDP), après relaxation des variables de décision discrètes du problème MILP du second niveau (6.3) (voir Sous-section 3.1.4), où le

métamodèle $\hat{Q}$ ne serait pas obtenu par apprentissage supervisé mais à l'aide des coupes ajoutées itérativement par l'approche de décomposition,

— en utilisant une approche par approximations successives (voir Sous-section 4.1.2) où le métamodèle $\hat{Q}$ serait obtenu par apprentissage supervisé à partir d'un jeu de données enrichi itérativement.

Dans les deux approches, le métamodèle $\hat{Q}$ ne serait pas construit en amont de l'optimisation, mais de manière imbriquée à la résolution du problème d'optimisation par des itérations successives.

Nous présentons ci-dessous les grandes lignes d'une approche par approximations successives qui s'effectuerait en complément de la réduction des scénarios présentée à la section 4.3. La structure générale de l'algorithme serait la suivante.

- *Initialisation* :
  - calculer numériquement la solution déterministe $\mathbf{x}_0^* \in \mathcal{X}(\xi_0)$ et le chiffre d'affaires $\nu_0 \in \mathbb{R}$ associé (voir Équation (6.1)),
  - initialiser le jeu de données $\mathcal{D}_0 = \emptyset$.
- *Itération* $j \geq 1$:
  - pour chaque scénario réduit $\xi'_k \in \hat{\Xi}'(\Omega)$, calculer numériquement

$$Q_j(\xi'_k) = Q(b(\mathbf{x}^*_{j-1}), \xi'_k) \in \mathbb{R},$$

  - enrichir le jeu de données $\mathcal{D}_j = \mathcal{D}_{j-1} \cup \{((b(\mathbf{x}^*_j), \xi'_k), Q_j(\xi'_k))\}$,
  - construire un métamodèle $\hat{Q}_j$ de Q à partir du jeu de données $\mathcal{D}_j$,
  - construire un métamodèle linéaire par morceaux $\tilde{\mathcal{E}}_j$ de $\hat{\mathcal{E}}_j : b_0 \in \mathcal{B}_0 \mapsto \mathbb{E}[\hat{Q}_j(b_0, \hat{\Xi}')]$,
  - calculer numériquement la solution optimale $\mathbf{x}^*_j \in \mathcal{X}(\xi_0)$ et la valeur optimale $\nu_j \in \mathbb{R}$ du problème MILP

$$\max_{\mathbf{x}_0 \in \mathcal{X}(\xi_0)} \left\{ \alpha_0 f(\mathbf{x}_0, \xi_0) + (1 - \alpha_0) \tilde{\mathcal{E}}_j(b(\mathbf{x}_0)) \right\}.$$

Un critère d'arrêt possible pourrait porter sur l'écart entre les valeurs optimales : la dernière itération $j \in \mathbb{N}^*$, si elle existe, vérifierait la condition $|\nu_j - \nu_{j-1}| \leq \varepsilon$, pour un $\varepsilon \in \mathbb{R}_+$ donné (par exemple égal à un MIP-gap absolu seuil). La solution renvoyée par l'algorithme serait alors la dernière solution optimale $\mathbf{x}^*_j \in \mathcal{X}(\xi_0)$ calculée avant l'arrêt de l'algorithme.

Si les métamodèles $\hat{Q}_j$, pour $j \in \mathbb{N}$, sont choisis parmi les fonctions linéaires par morceaux (ce qui serait souhaitable d'après la structure de la fonction Q), alors les métamodèles $\tilde{\mathcal{E}}_j$ peuvent être définis par :

$$\forall b_0 \in \mathcal{B}_0, \quad \tilde{\mathcal{E}}_j(b_0) = \sum_{k=1}^{N'} p'_k \hat{Q}_j(b_0, \xi'_k).$$

Dans le cas contraire, les métamodèles $\tilde{\mathcal{E}}_j$, pour $j \in \mathbb{N}$, seraient obtenus par approximation linéaire par morceaux de $\hat{\mathcal{E}}_j$.

Cette approche par approximations successives présente néanmoins des difficultés :

— une itération de l'algorithme nécessite N' résolutions numériques indépendantes du problème MILP (6.3), donc le temps de calcul d'une itération n'est pas négligeable, sauf si la réduction des scénarios est très efficace (i.e. $N' \ll N$) et avec la possibilité de recourir au calcul parallèle (mémoire distribuée),

— la construction des métamodèles $\hat{Q}_j$, pour $j \in \mathbb{N}^*$, reprend les problématiques de la méthode de construction dans une phase de pré-traitement,

— la convergence n'est *a priori* pas garantie et le nombre d'itérations (et donc le temps de calcul) peut être important.

La première difficulté pourrait être contournée en utilisant une version dégradée $\tilde{Q}$ de la fonction Q par relaxation. La fonction $\tilde{Q}$ serait alors la valeur optimale d'un problème LP plus rapide à résoudre, au prix d'erreurs d'approximation qu'il faudrait analyser. Le travail présenté dans les chapitres 4 et 5 du manuscrit, consacré à la méthode de construction du métamodèle dans une phase de pré-traitement, est une étape préliminaire pour étudier la deuxième difficulté ci-dessus. La troisième difficulté ci-dessus nécessiterait quant à elle des études approfondies.

6.3.4 Généralisation de la méthodologie

Dans le cadre de la thèse, les prévisions des prix de règlement des écarts sont effectuées par dégradation des prix *day-ahead* (voir §2.3.4.4). Leurs incertitudes sont donc issues de celles sur les prix *day-ahead*. Si nous avions des prévisions probabilistes des prix de règlement des écarts positifs π^+ et négatifs π^-, alors la méthodologie proposée pourrait s'appliquer avec des scénarios de la forme $\xi_n = (a_n, \pi_n, \pi_n^+, \pi_n^-, \phi_n)$, pour $n \in \{1, \dots, N\}$. Cela ajouterait néanmoins des dimensions supplémentaires pour la réduction de la dimension, la réduction des scénarios et la construction du plan d'expérience. La prévision d'ensemble des apports en eau aux aménagements hydroélectriques pourraient également tenir compte des débits non mesurés, en plus de ceux des affluents principaux, si nous avions les prévisions probabilistes associées (voir §2.4.4.3).

Une difficulté qui n'a pas été abordée dans ce manuscrit concerne la contrainte de volume minimum aux aménagements hydroélectriques en fin d'horizon (voir Équation (3.14)). En effet, dans la méthodologie proposée, la même valeur minimale est utilisée dans le problème MILP (6.3) pour tous les scénarios probabilistes, ce qui peut poser question car cette valeur a été définie à partir d'une solution initiale construite à partir d'une prévision déterministe des apports en eau (voir Sous-section 3.2.3). Une manière de contourner cette difficulté serait d'ajouter une valeur économique dans le calcul du chiffre d'affaires f pour valoriser les quantités d'eau restantes dans les retenues en fin d'horizon (analogue à la valeur de l'eau pour les centrales hydroélectrique de haute chute).

En outre, le problème d'optimisation déterministe (6.1) sur lequel repose le problème d'optimisation probabiliste (6.2) pourrait être généralisé pour tenir compte d'autres marchés que le marché *day-ahead* (marché infrajournalier, marché des services systèmes par exemple), voire des actifs de stockage ou de flexibilité supplémentaires (batteries, hydrogène renouvelable...) qui pourraient compléter la flexibilité offerte par la chaîne hydroélectrique.

Conclusion générale

Dans ce manuscrit de thèse, nous avons abordé le problème de la planification à court terme de la production conjointe d'une chaîne hydroélectrique et d'actifs de production variable pour une utilisation dans un contexte opérationnel. Plus précisément, nous nous sommes intéressés au problème d'optimisation qui consiste en la construction du programme de production détaillé de la chaîne hydroélectrique pour maximiser le chiffre d'affaires obtenu par la vente sur le marché *day-ahead* de la production de l'ensemble des actifs et par la pénalisation des écarts. Une telle gestion conjointe permet de partager la flexibilité de la chaîne hydroélectrique entre le déplacement d'énergie et la compensation des écarts dus aux incertitudes de prévision.

La gestion conjointe d'une chaîne hydroélectrique et d'actifs de production variable repose sur des prévisions performantes des apports en eau de la chaîne hydroélectrique, des prix de l'électricité et des productions variables. Dans un cadre déterministe, les incertitudes qui entachent ces prévisions sont principalement gérées grâce à l'enchaînement des planifications à des instants réguliers et fréquents au cours du temps à partir de la connaissance des valeurs qui se réalisent effectivement et des mises à jour des prévisions. Néanmoins, il est possible de valoriser les incertitudes de prévision en considérant la planification dans un cadre probabiliste à partir de leur estimation. Nous avons choisi de modéliser ces incertitudes de prévision en scénarios multivariés construits à partir de prévisions d'ensemble. L'intérêt principal de ce choix est de respecter la cohérence spatio-temporelle des prévisions, notamment celle des apports en eau qui est une caractéristique importante à prendre en compte dans la gestion d'une chaîne hydroélectrique. Dans ce manuscrit de thèse, nous avons proposé une méthodologie pour considérer ce problème d'optimisation dans un univers probabiliste.

Le premier objectif de la thèse était de formuler mathématiquement le problème d'optimisation dans un univers probabiliste adapté aux scénarios multivariés. Nous nous sommes appuyés pour cela sur l'optimiseur déterministe et le processus opérationnel de CNR pour la gestion de la seule chaîne hydroélectrique le long du Rhône. Le problème d'optimisation probabiliste s'écrit sous la forme d'un problème de programmation dynamique stochastique à deux niveaux linéaires mixtes. Le premier niveau correspond au problème d'optimisation en univers déterministe, tandis que le second niveau quantifie l'impact sur le chiffre d'affaires des prochaines modifications possibles du programme de production en fonction de l'évolution effective du scénario. La liaison entre les deux niveaux est donnée par les prochaines ventes *day-ahead*. Cette formulation permet de mettre à profit la flexibilité offerte par la production hydroélectrique. Néanmoins, une telle formulation s'appuie sur la simplification qui consiste à considérer que la valorisation infrajournalière (i.e. une fois les ventes *day-ahead* connues) est une réoptimisation complète du programme de production, pouvant proposer des modifications sur l'ensemble de la fenêtre temporelle, le jour-même compris. Une autre limite de cette formulation probabiliste est que la valeur optimale du second niveau doit avoir un impact significatif sur le chiffre d'affaires. Dans le cas contraire, comme avec le cas d'étude testé dans ce manuscrit, la formulation en univers probabiliste n'apporte aucune information significative supplémentaire par rapport à celle en univers déterministe en raison de l'absence des productions variables dans le cas d'étude.

Le second objectif de la thèse était de résoudre numériquement le problème d'optimisation en univers probabiliste. En raison de la dimension du problème d'optimisation et de la formulation linéaire mixte des contraintes des deux niveaux, les algorithmes de la littérature ne sont pas adaptés. Nous avons donc proposé de résoudre le problème d'optimisation approché dans lequel la valeur optimale du second niveau, coûteuse à évaluer numériquement, est remplacée par un métamodèle rapide à évaluer et calibré par apprentissage supervisé dans une phase de pré-traitement à l'optimisation. Le domaine de la donnée d'entrée associée aux prochaines ventes *day-ahead* issues des variables de décision du premier niveau n'étant pas accessible, il est approché par un ensemble fini obtenu par échantillonnage par analogie. Les données d'entrée fonctionnelles sont ensuite réduites par analyse en composantes principales et le nombre de scénarios peut éventuellement être réduit. Le jeu de données est alors construit à partir d'un plan d'expérience par échantillonnage par hypercube latin sur lequel la valeur optimale du second niveau est évaluée. Plusieurs métamodèles linéaires sont ensuite testés. Cette construction du métamodèle n'est acceptable en pratique que si les erreurs d'approximation cumulées à chaque étape sont suffisamment petites. Néanmoins, la réduction des erreurs d'approximation conduit à une augmentation des temps de calcul de la phase de pré-traitement. Pour le cas d'étude testé dans ce manuscrit, le métamodèle obtenu par régression linéaire sur les données d'entrée fonctionnelles réduites donne des performances acceptables. Il permet d'obtenir une solution du problème d'optimisation en univers probabiliste mais avec un temps de calcul qui n'est pas compatible avec un usage en opérationnel. Pour une utilisation dans un contexte opérationnel à temps restreint, un compromis acceptable entre les erreurs d'approximation et le temps de calcul à disposition doit donc être trouvé, ce qui ne semble pas évident en pratique. Par exemple, pour le cas d'étude, il semble difficile de rendre acceptable le temps de calcul de la phase pré-traitement pour une utilisation opérationnelle sans fortement dégrader la qualité du métamodèle et sans recourir au calcul parallèle (mémoire distribuée).

Une autre limite de la méthodologie proposée concerne l'échantillonnage par analogie pour approcher l'ensemble des prochaines ventes *day-ahead* possibles car il ne permet pas de quantifier les erreurs d'approximation. Pour le cas d'étude testé dans ce manuscrit, certaines contraintes du problème d'optimisation du second niveau, pourtant importantes dans la formulation probabiliste, ont dû être relâchées pour assurer l'existence d'au moins une solution. Cette limite pourrait être contournée en construisant le métamodèle, non pas dans une phase de pré-traitement comme cela est présenté dans ce manuscrit, mais de manière itérative avec la résolution du problème d'optimisation probabiliste approché. On formerait alors une approche itérative par approximations successives. Le jeu de données serait enrichi à chaque itération grâce aux solutions optimales des problèmes d'optimisation approchés des itérations précédentes. Une telle approche serait acceptable en pratique si les performances du métamodèle s'amélioraient à chaque itération (convergence de l'algorithme) et si le nombre d'itérations pour obtenir une qualité acceptable du métamodèle était compatible avec le temps de calcul à disposition.

Par ailleurs, avec le développement du stockage d'électricité renouvelable (notamment avec l'hydrogène vert) et d'actifs de flexibilité, l'optimiseur pourrait servir, plus largement, pour la gestion conjointe en univers probabiliste de l'ensemble des actifs de production, de stockage et de flexibilité.

Bibliographie

Hamdi Abdi : Profit-based unit commitment problem : A review of models, methods, challenges, and future directions. *Renewable and Sustainable Energy Reviews*, page 110504, 2020.

ADEME : *Un mix électrique 100% renouvelable ? Analyses et optimisations*, 2015.

Shabbir Ahmed : Two-stage stochastic integer programming : A brief introduction. *Wiley encyclopedia of operations research and management science*, 2010.

Shabbir Ahmed, Alexander Shapiro et Er Shapiro : The sample average approximation method for stochastic programs with integer recourse. *Submitted for publication*, pages 1–24, 2002.

Shabbir Ahmed, Mohit Tawarmalani et Nikolaos V Sahinidis : A finite branch-and-bound algorithm for two-stage stochastic integer programs. *Mathematical Programming*, 100(2):355–377, 2004.

Bodo Ahrens et André Walser : Information-based skill scores for probabilistic forecasts. *Monthly Weather Review*, 136(1):352–363, 2008.

Shon Albert : Solving mixed integer linear programs using branch and cut algorithm. *Ph. D. dissertation*, 1999.

Laurent Alfandari et Jérôme Monnot : A note on the clustered set covering problem. *Discrete Applied Mathematics*, 164:13–19, 2014.

Grégoire Allaire : *Analyse numérique et optimisation : une introduction à la modélisation mathématique et à la simulation numérique*. Editions Ecole Polytechnique, 2005.

Reda El Amri, Céline Helbert, Miguel Munoz Zuniga, Clémentine Prieur et Delphine Sinoquet : Set inversion under functional uncertainties with joint meta-models. Working paper or preprint, 2020.

Anestis Antoniadis, Céline Helbert, Clémentine Prieur et Laurence Viry : Spatio-temporal metamodeling for west african monsoon. *Environmetrics*, 23(1):24–36, 2012.

Dimitra Apostolopoulou, Zacharie De Greve et Malcolm McCulloch : Robust optimization for hydroelectric system operation under uncertainty. *IEEE Transactions on Power Systems*, 33(3):3337–3348, 2018.

Alicia Arce, Takaaki Ohishi et Sérgio Soares : Optimal dispatch of generating units of the itaipú hydroelectric plant. *IEEE Transactions on power systems*, 17(1):154–158, 2002.

Margaret Armstrong, Aziz Ndiaye, Rija Razanatsimba et Alain Galli : Scenario reduction applied to geostatistical simulations. *Mathematical Geosciences*, 45(2):165–182, 2013.

Frédéric Atger : Tubing : An alternative to clustering for the classification of ensemble forecasts. *Weather and forecasting*, 14(5):741–757, 1999.

Hannah Bakker, Fabian Dunke et Stefan Nickel : A structuring review on multi-stage optimization under uncertainty : Aligning concepts from theory and practice. *Omega*, 96: 102080, 2020.

Peter Bauer, Alan Thorpe et Gilbert Brunet : The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015.

Joseph Bellier : *Prévisions hydrologiques probabilistes dans un cadre multivarié : quels outils pour assurer fiabilité et cohérence spatio-temporelle ?* Thèse de doctorat, Université Grenoble Alpes (ComUE), 2018.

Joseph Bellier, Guillaume Bontron et Isabella Zin : Using meteorological analogues for reordering postprocessed precipitation ensembles in hydrological forecasting. *Water Resources Research*, 53(12):10085–10107, 2017.

Joseph Bellier, Isabella Zin et Guillaume Bontron : Generating coherent ensemble forecasts after hydrological postprocessing : Adaptations of ecc-based methods. *Water Resources Research*, 54(8):5741–5762, 2018.

Joseph Bellier, Isabella Zin, Stanislas Siblot et Guillaume Bontron : Probabilistic flood forecasting on the Rhone river : evaluation with ensemble and analogue-based precipitation forecasts. *In E3S Web of Conferences*, volume 7, page 18011. EDP Sciences, 2016.

Richard Bellman : Dynamic programming. *Science*, 153(3731):34–37, 1966.

Aharon Ben-Tal, Laurent El Ghaoui et Arkadi Nemirovski : *Robust optimization*. Princeton university press, 2009.

Jacob Benesty, Jingdong Chen, Yiteng Huang et Israel Cohen : Pearson correlation coefficient. *In Noise reduction in speech processing*, pages 1–4. Springer, 2009.

Marco Bernreuther : Solving mixed-integer programming problems using piecewise linearization methods. Mémoire de D.E.A., Universität Konstanz, Konstanz, 2017.

Dimitri P Bertsekas, Gregory S Lauer, Nils R Sandell et Thomas A Posbergh : Optimal short-term scheduling of large-scale power systems. *IEEE Transactions on Automatic Control*, 28(1):1–11, 1983.

José Betancourt, François Bachoc, Thierry Klein, Déborah Idier, Rodrigo Pedreros et Jérémy Rohmer : Gaussian process metamodeling of functional-input code for coastal flood hazard assessment. *Reliability Engineering & System Safety*, 198:106870, 2020.

Charles Eugene Blair et Robert G Jeroslow : The value function of a mixed integer program : I. *Discrete Mathematics*, 19(2):121–138, 1977.

Pierre Bompart, Guillaume Bontron, Sabrina Celie et Muriel Haond : Une chaîne opérationnelle de prévision hydrométéorologique pour les besoins de la production hydroélectrique de la CNR. *La Houille Blanche*, 95(5):54–60, 2009.

J Frédéric Bonnans, Jean Charles Gilbert, Claude Lemaréchal et Claudia Sagastizábal : *Optimisation Numérique : aspects théoriques et pratiques*, volume 27. Springer Berlin, Germany, 1997.

Guillaume Bontron : *Prévision quantitative des précipitations : Adaptation probabiliste par recherche d'analogues. Utilisation des Réanalyses NCEP/NCAR et application aux précipitations du Sud-Est de la France*. Thèse de doctorat, Institut National Polytechnique Grenoble (INPG), 2004.

Alberto Borghetti, Claudia D'Ambrosio, Andrea Lodi et Silvano Martello : An MILP approach for short-term hydro scheduling and unit commitment with head-dependent reservoir. *IEEE Transactions on power systems*, 23(3):1115–1124, 2008.

Marie-Amélie Boucher, D Tremblay, Louis Delorme, L Perreault et F Anctil : Hydro-economic assessment of hydrological forecasting systems. *Journal of Hydrology*, 416:133–144, 2012.

Ronald N Bracewell : Pentagram notation for cross correlation. The Fourier transform and its applications. *New York : McGraw-Hill*, 46:243, 1965.

Kenneth Bruninx et Erik Delarue : Scenario reduction techniques and solution stability for stochastic unit commitment problems. *In IEEE International Energy Conference (ENERGYCON)*, pages 1–7. IEEE, 2016.

Alain Burtin et Vera Silva : Technical and economic analysis of the european electricity system with 60% RES. Rapport technique, EDF R&D, 2015.

Simon Camal, Fei Teng, Andrea Michiorri, Georges Kariniotakis et Luis Badesa : Scenario generation of aggregated wind, photovoltaics and small hydro production for power systems applications. *Applied Energy*, 242:1396–1406, 2019.

Claus C Carøe et Rüdiger Schultz : Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24(1-2):37–45, 1999.

Claus C Carøe et Jørgen Tind : L-shaped decomposition of two-stage stochastic programs with integer recourse. *Mathematical Programming*, 83(1):451–464, 1998.

Pierre Carpentier, J-Ph Chancelier, Vincent Leclère et François Pacaud : Stochastic decomposition applied to large-scale hydro valleys management. *European Journal of Operational Research*, 270(3):1086–1098, 2018.

Pierre Carpentier, Michel Gendreau et Fabian Bastin : Long-term management of a hydroelectric multireservoir system under uncertainty using the progressive hedging algorithm. *Water Resources Research*, 49(5):2812–2827, 2013.

Pierre Carpentier, Guy Gohen, J-C Culioli et Arnaud Renaud : Stochastic optimization of unit commitment : a new decomposition framework. *IEEE Transactions on Power Systems*, 11(2):1067–1073, 1996.

Miguel Carrión et José Manuel Arroyo : A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem. *IEEE Transactions on power systems*, 21(3):1371–1378, 2006.

Manon Cassagnole, Maria-Helena Ramos, Ioanna Zalachori, Guillaume Thirel, Rémy Garçon, Joël Gailhard et Thomas Ouillon : Impact of the quality of hydrological forecasts on the management and revenue of hydroelectric reservoirs–a conceptual approach. *Hydrology and Earth System Sciences*, 25(2):1033–1052, 2021.

Anya Castillo, Jack Flicker, Clifford W Hansen, Jean-Paul Watson et Jay Johnson : Stochastic optimisation with risk aversion for virtual power plant operations : a rolling horizon control. *IET Generation, Transmission & Distribution*, 13(11):2063–2076, 2019.

Sabrina Celie, Guillaume Bontron, David Ouf et Evelyne Pont : Apport de l'expertise dans la prévision hydro-météorologique opérationnelle. *La Houille Blanche*, 2:55–62, 2019.

CETIM : *Récupération de l'énergie des rivières et des océans*. Centre techniques des industries mécaniques, 2009.

Jean-Philippe Chancelier et Arnaud Renaud : Daily generation scheduling : decomposition methods to solve the hydraulic problems. *International Journal of Electrical Power & Energy Systems*, 16(3):175–181, 1994.

Gary W Chang, Mohamed Aganagic, James G Waight, José Medina, Tony Burton, Steve Reeves et Milton Christoforidis : Experiences with mixed integer linear programming based approaches on short-term hydro scheduling. *IEEE Transactions on power systems*, 16 (4):743–749, 2001.

Abraham Charnes et William Wager Cooper : Chance-constrained programming. *Management science*, 6(1):73–79, 1959.

Anindya Chatterjee : An introduction to the proper orthogonal decomposition. *Current science*, pages 808–817, 2000.

Martyn Clark, Subhrendu Gangopadhyay, Lauren Hay, Balaji Rajagopalan et Robert Wilby : The schaake shuffle : A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology*, 5(1):243–262, 2004.

Antonio J Conejo, José Manuel Arroyo, Javier Contreras et Francisco Apolinar Villamor : Self-scheduling of a hydro producer in a pool-based electricity market. *IEEE Transactions on power systems*, 17(4):1265–1272, 2002.

Antonio J Conejo, Miguel Carrión et Juan M Morales : *Decision making under uncertainty in electricity markets*, volume 1. Springer, 2010.

Sylvain Corlay et Gilles Pagès : Functional quantization-based stratified sampling methods. *Monte Carlo Methods and Applications*, 21(1):1–32, 2015.

IBM ILOG Cplex : V12. 1 : User's manual for CPLEX. *International Business Machines Corporation*, 46(53):157, 2009.

Jean-Christophe Culioli : *Introduction à l'optimisation*. Ellipses, 2012.

Thibaut Cuvelier, Pierre Archambeau, Benjamin Dewals et Quentin Louveaux : Comparison between robust and stochastic optimisation for long-term reservoir management under uncertainty. *Water resources management*, 32(5):1599–1614, 2018.

Guillaume Damblin, Mathieu Couplet et Bertrand Iooss : Numerical studies of space-filling designs : optimization of latin hypercube samples and subprojection properties. *Journal of Simulation*, 7(4):276–289, 2013.

George B Dantzig : Linear programming. *Operations research*, 50(1):42–47, 2002.

Carl De Boor : *A practical guide to splines*, volume 27. springer-verlag New York, 1978.

Daniel De Ladurantaye, Michel Gendreau et Jean-Yves Potvin : Optimizing profits from hydroelectricity production. *Computers & Operations Research*, 36(2):499–529, 2009.

Roy De Maesschalck, Delphine Jouan-Rimbaud et Désiré Luc Massart : The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, 50(1):1–18, 2000.

István Deák : Testing successive regression approximations by large-scale two-stage problems. *Annals of Operations Research*, 186(1):83–99, 2011.

István Deák, Imre Pólik, András Prékopa et Tamás Terlaky : Convex approximations in stochastic programming by semidefinite programming. *Annals of Operations Research*, 200 (1):171–182, 2012.

Boris Defourny, Damien Ernst et Louis Wehenkel : Multistage stochastic programming : A scenario tree based approach to planning under uncertainty. *In Decision theory models for applications in artificial intelligence : concepts and solutions*, pages 97–143. IGI Global, 2012.

Luca Delle Monache, F Anthony Eckel, Daran L Rife, Badrinath Nagarajan et Keith Searight : Probabilistic weather prediction with an analog ensemble. *Monthly Weather Review*, 141(10):3498–3516, 2013.

Timothy R Derrick et Joshua M Thomas : *Time series analysis : the cross-correlation function*, pages 189–205. Human Kinetics, 01 2004.

F Javier Diaz, Javier Contreras, José Ignacio Muñoz et David Pozo : Optimal scheduling of a price-taker cascaded reservoir system in a pool-based electricity market. *IEEE Transactions on Power Systems*, 26(2):604–615, 2010.

Urmila M Diwekar : *Introduction to applied optimization*, volume 22. Springer Nature, 2020.

Jitka Dupačová : Stability and sensitivity-analysis for stochastic programming. *Annals of operations research*, 27(1):115–142, 1990.

Jitka Dupačová, Nicole Gröwe-Kuska et Werner Römisch : *Scenario reduction in stochastic programming : An approach using probability metrics.* Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2000.

Jitka Dupačová, Nicole Gröwe-Kuska et Werner Römisch : Scenario reduction in stochastic programming. *Mathematical programming*, 95(3):493–511, 2003.

Reda El Amri, Céline Helbert, Olivier Lepreux, Miguel Munoz Zuniga, Clémentine Prieur et Delphine Sinoquet : Data-driven stochastic inversion via functional quantization. *Statistics and Computing*, 30, 05 2020.

Beth A Faber et JR Stedinger : Reservoir optimization using sampling sdp with ensemble streamflow prediction (ESP) forecasts. *Journal of Hydrology*, 249(1-4):113–133, 2001.

Fernando Mainardi Fan, Dirk Schwanenberg, Rodolfo Alvarado, Alberto Assis Dos Reis, Walter Collischonn et Steffi Naumman : Performance of deterministic and probabilistic hydrological forecasts for the short-term optimization of a tropical hydropower reservoir. *Water Resources Management*, 30(10):3609–3625, 2016.

Erlon Cristian Finardi et Murilo Reolon Scuzziato : A comparative analysis of different dual problems in the lagrangian relaxation context for solving the hydro unit commitment problem. *Electric Power Systems Research*, 107:221–229, 2014.

Nicolas Flammarion : *Stochastic approximation and least-squares regression, with applications to machine learning.* Thèse de doctorat, Paris Sciences et Lettres (ComUE), 2017.

Stein-Erik Fleten et Trine Krogh Kristoffersen : Stochastic programming for optimizing bidding strategies of a nordic hydropower producer. *European Journal of Operational Research*, 181(2):916–928, 2007.

Stein-Erik Fleten et Trine Krogh Kristoffersen : Short-term hydropower production planning by stochastic programming. *Computers & Operations Research*, 35(8):2656–2671, 2008.

Jerome Friedman, Trevor Hastie et Rob Tibshirani : Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.

Jerome Friedman, Trevor Hastie et Robert Tibshirani : *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

Tomas Gal et Harvey J Greenberg : *Advances in sensitivity analysis and parametric programming*, volume 6. Springer Science & Business Media, 2012.

Javier García-González, Ernesto Parrilla et Alicia Mateo : Risk-averse profit-based optimal scheduling of a hydro-chain in the day-ahead electricity market. *European Journal of Operational Research*, 181(3):1354–1369, 2007.

Saul I Gass : *Linear programming : methods and applications*. Courier Corporation, 2003.

Xiaolin Ge, Shu Xia et Wei-Jen Lee : An efficient stochastic algorithm for mid-term scheduling of cascaded hydro systems. *Journal of Modern Power Systems and Clean Energy*, 7(1):163–173, 2019.

Claude Gemaehling : Quelques aspects des études hydrauliques des aménagements du Bas-Rhône. *La Houille Blanche*, pages 837–904, 1962.

Arthur M Geoffrion : Generalized benders decomposition. *Journal of optimization theory and applications*, 10(4):237–260, 1972.

Roger Ginocchio et Pierre-Louis Viollet : *L'énergie hydraulique*. Lavoisier, 2012.

Tilmann Gneiting, Fadoua Balabdaoui et Adrian E Raftery : Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 69(2):243–268, 2007.

Tilmann Gneiting et Adrian E Raftery : Weather forecasting with ensemble methods. *Science*, 310(5746):248–249, 2005.

Tilmann Gneiting et Adrian E Raftery : Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

Tilmann Gneiting, Adrian E Raftery, Anton H Westveld III et Tom Goldman : Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133(5):1098–1118, 2005.

Xavier Gourdon : *Les maths en tête - Algèbre*. Ellipses, 2008a.

Xavier Gourdon : *Les maths en tête - Analyse*. Ellipses, 2008b.

Siegfried Graf et Harald Luschgy : Quantization for probability measures in the prokhorov metric. *Theory of Probability & Its Applications*, 53(2):216–241, 2009.

Laëtitia Grimaldi, Guillaume Bontron et Pierre Balayn : Hydraulic modelling for Rhône river operation. *In Advances in Hydroinformatics*, pages 35–45. Springer, 2014.

Nicole Gröwe-Kuska, Holger Heitsch et Werner Römisch : Scenario reduction and scenario tree construction for power management problems. *In IEEE Bologna Power Tech Conference Proceedings,*, volume 3, pages 7–pp. IEEE, 2003.

Nicole Gröwe-Kuska, Krzysztof C Kiwiel, Matthias P Nowak, Werner Römisch et Isabel Wegner : Power management in a hydro-thermal system under uncertainty by lagrangian relaxation. *In Decision making under uncertainty*, pages 39–70. Springer, 2002.

Thomas M Hamill : Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, 129(3):550–560, 2001.

Klein Haneveld et Maarten H van der Vlerk : *Stochastic programming.* Springer, 2020.

Willem K Klein Haneveld, Maarten H van der Vlerk et Ward Romeijnders : *Stochastic Programming : Modeling Decision Problems Under Uncertainty.* Springer Nature, 2019.

Grace Hechme-Doukopoulos, Sandrine Brignol-Charousset, Jérôme Malick et Claude Lemaréchal : The short-term electricity production management problem at EDF. *Optima Newsletter*, 84:2–6, 2010.

Holger Heitsch et Werner Römisch : Scenario reduction algorithms in stochastic programming. *Computational optimization and applications*, 24(2):187–206, 2003.

Hans Hersbach : Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570, 2000.

Ronald A Howard : *Dynamic programming and markov processes.* John Wiley, 1960.

Yuping Huang, Panos M Pardalos et Qipeng P Zheng : *Electrical power unit commitment : deterministic and two-stage stochastic programming models and algorithms.* Springer, 2017.

Marchand Hugues, Martin Alexander, Weismantel Robert et Wolsey Laurence : Cutting planes in integer and mixed integer programming. *Discrete Applied Mathematics*, 123(1):397–446, 2002.

Muhammad Shahzad Javed, Tao Ma, Jakub Jurasz et Muhammad Yasir Amin : Solar and wind power generation systems with pumped hydro storage : Review and future perspectives. *Renewable Energy*, 148:176–192, 2020.

Ian Jolliffe : Principal component analysis. *Encyclopedia of statistics in behavioral science*, 2005.

Ian T Jolliffe : Principal components in regression analysis. *In Principal component analysis*, pages 129–155. Springer, 1986.

Ian T Jolliffe et David B Stephenson : *Forecast verification : a practitioner's guide in atmospheric science.* John Wiley & Sons, 2012.

Jakub Jurasz, Jerzy Mikulik, Magdalena Krzywda, Bartłomiej Ciapała et Mirosław Janowski : Integrating a wind- and solar-powered hybrid to the power system by coupling it with a hydroelectric power station with pumping installation. *Energy*, 144:549–563, 2018.

Alan J King et Stein W Wallace : *Modeling with stochastic programming.* Springer Science & Business Media, 2012.

Anton J Kleywegt, Alexander Shapiro et Tito Homem-de Mello : The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.

Nan Kong, Andrew J Schaefer et Brady Hunsaker : Two-stage integer programs with stochastic right-hand sides : a superadditive dual approach. *Mathematical Programming*, 108 (2):275–296, 2006.

Wolfgang Kreitmeier : Optimal vector quantization in terms of wasserstein distance. *Journal of multivariate analysis*, 102(8):1225–1239, 2011.

John W Labadie : Optimal operation of multireservoir systems : State-of-the-art review. *Journal of water resources planning and management*, 130(2):93–111, 2004.

Anne Labouret et Michel Villoz : *Energie solaire photovoltaïque*, volume 3. Dunod, 2006.

Gilbert Laporte et François V Louveaux : The integer L-shaped method for stochastic integer programs with complete recourse. *Operations research letters*, 13(3):133–142, 1993.

Camilla Thorrud Larsen, Gerard L Doorman et Birger Mo : Evaluation of scenario reduction methods for stochastic inflow in hydro scheduling models. *In IEEE Eindhoven PowerTech*, pages 1–6. IEEE, 2015.

Leon S Lasdon : *Optimization theory for large systems*. Courier Corporation, 2002.

Eugene L Lawler et David E Wood : Branch-and-bound methods : A survey. *Operations research*, 14(4):699–719, 1966.

Martin Leutbecher et Timothy Noel Palmer : Ensemble forecasting. *Journal of computational physics*, 227(7):3515–3539, 2008.

Chao-an Li, Rui Yan et Jing-yan Zhou : Stochastic optimization of interconnected multireservoir power systems. *IEEE Transactions on Power Systems*, 5(4):1487–1496, 1990.

Xiang Li, Tiejian Li, Jiahua Wei, Guangqian Wang et William W-G Yeh : Hydro unit commitment via mixed integer linear programming : A case study of the three gorges project, China. *IEEE Transactions on Power Systems*, 29(3):1232–1241, 2013.

Francis A Longstaff et Eduardo S Schwartz : Valuing american options by simulation : a simple least-squares approach. *The review of financial studies*, 14(1):113–147, 2001.

Edward N Lorenz : Atmospheric predictability as revealed by naturally occurring analogues. *Journal of Atmospheric Sciences*, 26(4):636–646, 1969.

James E Matheson et Robert L Winkler : Scoring rules for continuous probability distributions. *Management science*, 22(10):1087–1096, 1976.

Michael D McKay, Richard J Beckman et William J Conover : A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1):55–61, 2000.

Xiao-Li Meng, Robert Rosenthal et Donald B Rubin : Comparing correlated correlation coefficients. *Psychological bulletin*, 111(1):172, 1992.

Getu Mesfin et Mahadzir Shuhaimi : A probabilistic optimization approach for a gas processing plant under uncertain feed conditions and product requirements. *International Journal of Chemical and Molecular Engineering*, 4(2):192–197, 2010.

John E Mitchell : Branch-and-cut algorithms for combinatorial optimization problems. *Handbook of applied optimization*, 1:65–77, 2002.

Juan Manuel Morales, Antonio J Conejo, Henrik Madsen, Pierre Pinson et Marco Zugno : Virtual power plants. *In Integrating renewables in electricity markets*, pages 243–287. Springer, 2014.

Max D Morris et Toby J Mitchell : Exploratory designs for computational experiments. *Journal of statistical planning and inference*, 43(3):381–402, 1995.

Fionn Murtagh et Pierre Legendre : Ward's hierarchical agglomerative clustering method : which algorithms implement ward's criterion ? *Journal of classification*, 31(3):274–295, 2014.

Simon Nanty, Céline Helbert, Amandine Marrel, Nadia Pérot et Clémentine Prieur : Sampling, metamodelling and sensitivity analysis of numerical simulators with functional stochastic inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):636–659, mai 2016.

Simon Nanty, Céline Helbert, Amandine Marrel, Nadia Pérot et Clémentine Prieur : Uncertainty quantification for functional dependent random variables. *Computational Statistics*, 32(2):559–583, juin 2017.

James Eamonn Nash et John V Sutcliffe : River flow forecasting through conceptual models part I—a discussion of principles. *Journal of hydrology*, 10(3):282–290, 1970.

John C Nash : The (Dantzig) simplex method for linear programming. *Computing in Science & Engineering*, 2(1):29–31, 2000.

Henrik Aalborg Nielsen, Henrik Madsen, Torben Skov Nielsen, Jake Badger, Gregor Giebel, Lars Landberg, Kai Sattler et Henrik Feddersen : Wind power ensemble forecasting. *In Proceedings of the 2004 Global Windpower Conference and Exhibition*, 2004.

Lukasz Bartosz Nikonowicz et Jaroslaw Milewski : Virtual power plants-general review : structure, application and optimization. *Journal of power technologies*, 92(3):135, 2012.

Jakub Nowotarski et Rafał Weron : Recent advances in electricity price forecasting : A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81:1548–1568, 2018.

Robert Nürnberg et Werner Römisch : A two-stage planning model for power scheduling in a hydro-thermal system under uncertainty. *Optimization and Engineering*, 3(4):355–378, 2002.

Gerardo J Osório, João Matias et João PS Catalão : A review of short-term hydro scheduling tools. *In 48th International Universities' Power Engineering Conference (UPEC)*, pages 1–6. IEEE, 2013.

Mahmoud M Othman, Yasser G Hegazy et Almoataz Y Abdelaziz : A review of virtual power plant definitions, components, framework and optimization. *International Electrical Engineering Journal*, 6(9):2010–2024, 2015.

Manfred Padberg : *Linear optimization and extensions*, volume 12. Springer Science & Business Media, 2013.

Gilles Pagès : Introduction to optimal vector quantization and its applications for numerics. Rapport technique, LPMA, juillet 2014.

Hrvoje Pandžić, Igor Kuzle et Tomislav Capuder : Virtual power plant mid-term dispatch optimization. *Applied energy*, 101:134–141, 2013.

Florian Pappenberger, Maria-Helena Ramos, Hannah L Cloke, Fredrik Wetterhall, Lorenzo Alfieri, Konrad Bogner, Anna Mueller et Peter Salamon : How do I know if my forecasts are better ? Using benchmarks in hydrological ensemble prediction. *Journal of Hydrology*, 522:697–713, 2015.

Mario VF Pereira : Optimal stochastic operations scheduling of large hydroelectric systems. *International Journal of Electrical Power & Energy Systems*, 11(3):161–169, 1989.

Mario VF Pereira, Nora Campodónico et Rafael Kelman : Long-term hydro scheduling based on stochastic models. *EPSOM*, 98:23–25, 1998.

Mario VF Pereira et Leontina MVG Pinto : Multi-stage stochastic optimization applied to energy planning. *Mathematical programming*, 52(1):359–375, 1991.

Georg Ch Pflug : Some remarks on the value-at-risk and the conditional value-at-risk. *In Probabilistic constrained optimization*, pages 272–281. Springer, 2000.

Marian R Piekutowski, Tadeusz Litwinowicz et Rod Frowd : Optimal short-term scheduling for a large-scale cascaded hydro system. *In Conference Proceedings Power Industry Computer Application Conference*, pages 292–298. IEEE, 1993.

Peeter Pikk et Marko Viiding : The dangers of marginal cost based electricity pricing. *Baltic journal of economics*, 13(1):49–62, 2013.

Vincent Piron, Guillaume Bontron et Martin Pochat : Operating a hydropower cascade to optimize energy management. *International Journal on Hydropower and Dams*, 22, 2015.

Efstratios N Pistikopoulos : *Multi-parametric programming : Theory, Algorithms and Applications*. Wiley-vch, 2007.

Warren B Powell : *Approximate Dynamic Programming : Solving the curses of dimensionality*, volume 703. John Wiley & Sons, 2007.

Luc Pronzato et Werner G Müller : Design of computer experiments : space filling and beyond. *Statistics and Computing*, 22(3):681–701, 2012.

Boris N Pshenichnyj : *The linearization method for constrained optimization*, volume 22. Springer Science & Business Media, 2012.

Danny Pudjianto, Charlotte Ramsay et Goran Strbac : Virtual power plant and system integration of distributed energy resources. *IET Renewable power generation*, 1(1):10–16, 2007.

Svetlozar T Rachev et Werner Römisch : Quantitative stability in stochastic programming : The method of probability metrics. *Mathematics of Operations Research*, 27(4):792–818, 2002.

Svetlozar T Rachev et Ludger Rüschendorf : *Mass Transportation Problems : Volume I : Theory*, volume 1. Springer Science & Business Media, 1998.

Adrian E Raftery, Tilmann Gneiting, Fadoua Balabdaoui et Michael Polakowski : Using bayesian model averaging to calibrate forecast ensembles. *Monthly weather review*, 133(5): 1155–1174, 2005.

Ragheb Rahmaniani, Teodor Gabriel Crainic, Michel Gendreau et Walter Rei : The benders decomposition algorithm : A literature review. *European Journal of Operational Research*, 259 (3):801–817, 2017.

Ted K Ralphs et Anahita Hassanzadeh : On the value function of a mixed integer linear optimization problem and an algorithm for its construction. *COR@L Technical Report 14T–004*, 2014.

James O Ramsay et Bernard W Silverman : *Applied functional data analysis : methods and case studies*, volume 77. Springer, 2002.

Marc Rapin et Jean-Marc Noël : *Energie éolienne*. Dunod, Paris, 2010.

Carl Edward Rasmussen : Gaussian processes in machine learning. *In Summer school on machine learning*, pages 63–71. Springer, 2003.

Luciano Raso, Dirk Schwanenberg, Nick van de Giesen et Peter Jules van Overloop : Short-term optimal operation of water systems using ensemble forecasts. *Advances in water resources*, 71:200–208, 2014.

Luciano Raso, Nick van de Giesen, Philip Stive, Dirk Schwanenberg et Peter Jules van Overloop : Tree structure generation from ensemble forecasts for real time control. *Hydrological Processes*, 27(1):75–82, 2013.

Payam Refaeilzadeh, Lei Tang et Huan Liu : Cross-validation. *Encyclopedia of database systems*, 5:532–538, 2009.

Shafiqur Rehman, Luai M Al-Hadhrami et Md Mahbub Alam : Pumped hydro energy storage system : A technological review. *Renewable and Sustainable Energy Reviews*, 44:586–598, 2015.

R Tyrrell Rockafellar et Roger J-B Wets : Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of operations research*, 16(1):119–147, 1991.

Ralph Tyrrell Rockafellar et Stanislav Uryasev : Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.

Ward Romeijnders, Rüdiger Schultz, Maarten H van der Vlerk et Willem K Klein Haneveld : A convex approximation for two-stage mixed-integer recourse models with a uniform error bound. *SIAM Journal on Optimization*, 26(1):426–447, 2016.

Werner Römisch : Stability of stochastic programming problems. *Handbooks in operations research and management science*, 10:483–554, 2003.

Werner Römisch et Stefan Vigerske : Recent progress in two-stage mixed-integer stochastic programming with applications to power production planning. *Handbook of power systems I*, pages 177–208, 2010.

Cornelis Roos, Tamás Terlaky et Jean-Philippe Vial : *Interior point methods for linear optimization*. Springer Science & Business Media, 2005.

Richard E Rosenthal : *GAMS – a user's guide*, 2004.

Sheldon M Ross : *Introduction to stochastic dynamic programming*. Academic press, 2014.

Mark S Roulston et Leonard A Smith : Evaluating probabilistic forecasts using information theory. *Monthly Weather Review*, 130(6):1653–1660, 2002.

RTE : *Règles relatives à la Programmation, au Mécanismes d'Ajustement et au Recouvrement des charges d'ajustement*, 2020.

RTE : *Futurs énergétiques 2050 - Rapport complet*, 2021.

Andrzej Ruszczyński et Alexander Shapiro : Stochastic programming models. *Handbooks in operations research and management science*, 10:1–64, 2003.

Hedayat Saboori, Mohammad Mohammadi et R Taghe : Virtual power plant (VPP), definition, concept, components and types. *In Asia-Pacific Power and Energy Engineering Conference*, pages 1–4. IEEE, 2011.

Nikolaos V Sahinidis : Optimization under uncertainty : state-of-the-art and opportunities. *Computers & Chemical Engineering*, 28(6-7):971–983, 2004.

Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto et Stefano Tarantola : Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer physics communications*, 181(2):259–270, 2010.

Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana et Stefano Tarantola : *Global sensitivity analysis : the primer*. John Wiley & Sons, 2008.

Thomas J Santner, Brian J Williams, William I Notz et Brain J Williams : *The design and analysis of computer experiments*, volume 1. Springer, 2003.

Balasubramanian Saravanan, Siddharth Das, Surbhi Sikri et Dwarkadas Kothari : A solution to the unit commitment problem—a review. *Frontiers in Energy*, 7(2):223–236, 2013.

Roman Schefzik, Thordis L Thorarinsdottir et Tilmann Gneiting : Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical science*, 28 (4):616–640, 2013.

Tim Schulze, Andreas Grothey et Ken McKinnon : A stabilised scenario decomposition algorithm applied to stochastic unit commitment problems. *European Journal of Operational Research*, 261(1):247–259, 2017.

Murilo Reolon Scuzziato, Erlon Cristian Finardi et Antonio Frangioni : Comparing spatial and scenario decomposition for stochastic hydrothermal unit commitment problems. *IEEE Transactions on Sustainable Energy*, 9(3):1307–1317, 2017.

Sara Séguin, Stein-Erik Fleten, Pascal Côté, Alois Pichler et Charles Audet : Stochastic short-term hydropower planning with inflow scenario trees. *European Journal of Operational Research*, 259(3):1156–1168, 2017.

Suvrajeet Sen : Algorithms for stochastic mixed-integer programming models. *Handbooks in operations research and management science*, 12:515–558, 2005.

Gaia Serraino et Stan Uryasev : Conditional value-at-risk (CVaR). *Encyclopedia of operations research and management science*, pages 258–266, 2013.

Alexander Shapiro : Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research*, 209(1):63–72, 2011.

Alexander Shapiro, Darinka Dentcheva et Andrzej Ruszczyński : *Lectures on stochastic programming : modeling and theory*. SIAM, 2021.

Hanif D Sherali et Barbara MP Fraticelli : A modification of benders' decomposition algorithm for discrete subproblems : An approach for stochastic programs with integer recourse. *Journal of Global Optimization*, 22(1):319–342, 2002.

Hanif D Sherali et Xiaomei Zhu : On solving discrete two-stage stochastic programs having mixed-integer first-and second-stage variables. *Mathematical programming*, 108(2):597–616, 2006.

VR Sherkat, R Campo, K Moslehi et EO Lo : Stochastic long-term hydrothermal optimization for a multireservoir system. *IEEE Transactions on Power Apparatus and Systems*, 8:2040–2050, 1985.

Pierluigi Siano : Demand response and smart grids—a survey. *Renewable and sustainable energy reviews*, 30:461–478, 2014.

Thomas K Siu, Garth A Nash et Ziad K Shawwash : A practical hydro, dynamic unit commitment and loading model. *IEEE transactions on power systems*, 16(2):301–306, 2001.

Ilya M Sobol : Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280, 2001.

Sobrina Sobri, Sam Koohi-Kamali et Nasrudin Abd Rahim : Solar photovoltaic generation forecasting methods : A review. *Energy Conversion and Management*, 156:459–497, 2018.

Daniel Solow : *Linear programming : An introduction to finite improvement algorithms*. Courier Corporation, 2014.

Michael Stein : Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):143–151, 1987.

Milad Tahanan, Wim van Ackooij, Antonio Frangioni et Fabrizio Lacalandra : Large-scale unit commitment under uncertainty. *4OR*, 13(2):115–171, 2015.

Samer Takriti, John R Birge et Erik Long : A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems*, 11(3):1497–1508, 1996.

Raouia Taktak et Claudia D'Ambrosio : An overview on mathematical programming approaches for the deterministic unit commitment problem in hydro valleys. *Energy Systems*, 8 (1):57–79, 2017.

Olivier Talagrand : Evaluation of probabilistic prediction systems. *In Workshop proceedings "Workshop on predictability", 20-22 October 1997, ECMWF, Reading, UK*, 1999.

James W Taylor, Patrick E McSharry et Roberto Buizza : Wind power density forecasting using ensemble predictions and time series models. *IEEE Transactions on Energy Conversion*, 24(3):775–782, 2009.

Robert Tibshirani : Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society : Series B (Methodological)*, 58(1):267–288, 1996.

Christopher Torrence et Gilbert P Compo : A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1):61–78, 1998.

Juan J Torres, Can Li, Robert M Apap et Ignacio E Grossmann : A review on the performance of linear and mixed integer two-stage stochastic programming algorithms and software. *Optimization Online*, 2019.

Nenad Tufegdzic, Rod Frowd et Walter Olaf Stadlin : A coordinated approach for real-time short term hydro scheduling. *IEEE Transactions on Power Systems*, 11(4):1698–1704, 1996.

Wim van Ackooij : A comparison of four approaches from stochastic programming for large-scale unit-commitment. *EURO Journal on Computational Optimization*, 5(1-2):119–147, 2017.

Wim van Ackooij, Irene Danti Lopez, Antonio Frangioni, Fabrizio Lacalandra et Milad Tahanan : Large-scale unit commitment under uncertainty : an updated literature survey. *Annals of Operations Research*, 271(1):11–85, 2018.

Wim van Ackooij et Jérôme Malick : Decomposition algorithm for large-scale two-stage unit-commitment. *Annals of Operations Research*, 238(1-2):587–613, 2016.

Niels van der Laan et Ward Romeijnders : A loose benders decomposition algorithm for approximating two-stage mixed-integer recourse models. *Mathematical Programming*, 190 (1):761–794, 2021.

Maarten H van der Vlerk : Convex approximations for a class of mixed-integer recourse models. *Annals of Operations Research*, 177(1):139–150, 2010.

Robert J Vanderbei : *Linear programming : foundations and extensions*, volume 285. Springer Nature, 2020.

Yelena Vardanyan et Mikael Amelin : A sensitivity analysis of short-term hydropower planning using stochastic programming. *In IEEE Power and Energy Society General Meeting*, pages 1–7. IEEE, 2012.

Yelena Vardanyan, Mikael Amelin et Mohammad Hesamzadeh : Short-term hydropower planning with uncertain wind power production. *In IEEE Power & Energy Society General Meeting*, pages 1–5. IEEE, 2013.

William N Venables et Brian D Ripley : *Modern applied statistics with S-PLUS*. Springer Science & Business Media, 2013.

Cédric Villani : *Optimal transport : old and new*, volume 338. Springer, 2009.

Stein W Wallace et Stein-Erik Fleten : Stochastic programming models in energy. *Handbooks in operations research and management science*, 10:637–677, 2003.

I-Jeng Wang et James C Spall : Stochastic optimisation with inequality constraints using simultaneous perturbations and penalty functions. *International Journal of Control*, 81(8): 1232–1238, 2008.

Jean-Paul Watson et David L Woodruff : Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. *Computational Management Science*, 8(4):355–370, 2011.

Hongyu Wu, Mohammad Shahidehpour, Zuyi Li et Wei Tian : Chance-constrained day-ahead scheduling in stochastic power system operation. *IEEE Transactions on Power Systems*, 29 (4):1583–1591, 2014.

Peng Zhao et Bin Yu : On model selection consistency of lasso. *The Journal of Machine Learning Research*, 7:2541–2563, 2006.

Yanjia Zhao, Xi Chen, Qing-Shan Jia, Xiaohong Guan, Shuanghu Zhang et Yunzhong Jiang : Long-term scheduling for cascaded hydro energy systems with annual water consumption and release constraints. *IEEE Transactions on Automation Science and Engineering*, 7(4):969–976, 2010.

Feilin Zhu, Ping-an Zhong, Bin Xu, Weifeng Liu, Wenzhuo Wang, Yimeng Sun, Juan Chen et Jieyu Li : Short-term stochastic optimization of a hydro-wind-photovoltaic hybrid system under multiple uncertainties. *Energy Conversion and Management*, 214:112902, 2020.

Jikai Zou, Shabbir Ahmed et Xu Andy Sun : Stochastic dual dynamic integer programming. *Mathematical Programming*, 175(1):461–502, 2019.

Annexe A

Compléments sur les scénarios

A.1 Étude des corrélations

Pour simplifier la génération des scénarios du §2.4.4.2, nous faisons l'hypothèse que les apports en eau du Rhône, les prix de l'électricité et les productions variables du périmètre d'équilibre CNR sont indépendants. Dans cette section, nous proposons de confronter cette hypothèse aux valeurs observées en pratique.

Les valeurs réalisées des prix et des productions variables ne dépendent que du temps, donc elles forment des séries temporelles. En revanche, les apports en eau réalisés dépendent des affluents et du temps, donc ils forment une série spatio-temporelle. Pour faciliter l'étude des corrélations de cette section, nous considérons que les apports en eau sont les apports cumulés des affluents, formant ainsi une série temporelle.

Nous cherchons donc à étudier les corrélations entre trois séries temporelles : les apports en eau (cumulés), les prix et les productions variables. Pour tenir compte de l'aspect temporel, nous calculons les *corrélations croisées*, une extension des coefficients de Pearson qui tient compte des décalages temporels (appelés *lags*) entre les séries temporelles.

A.1.1 Théorie sur les corrélations croisées

Nous donnons dans cette sous-section les formules qui définissent les corrélations croisées entre deux séries temporelles. Nous présentons d'abord un point de vue issu du traitement du signal avant d'aborder un point de vue statistique avec des données discrétisées sur une période donnée.

Soient $u \in \mathcal{L}^2(\mathbb{R}, \mathbb{R})$ et $v \in \mathcal{L}^2(\mathbb{R}, \mathbb{R})$ des signaux temporels de carrés intégrables. On pose $\mu_u \in \mathbb{R}$ et $\mu_v \in \mathbb{R}$ les réels définis par

$$\mu_u = \int_{\mathbb{R}} u(t)\,\mathrm{d}t \in \mathbb{R} \quad \text{et} \quad \mu_v = \int_{\mathbb{R}} v(t)\,\mathrm{d}t \in \mathbb{R}.$$

On pose également $\sigma_u \in \mathbb{R}_+$ et $\sigma_v \in \mathbb{R}_+$ les réels positifs ou nuls définis par

$$\sigma_u = \sqrt{\int_{\mathbb{R}} (u(t) - \mu_u)^2\,\mathrm{d}t} \in \mathbb{R}_+ \quad \text{et} \quad \sigma_v = \sqrt{\int_{\mathbb{R}} (v(t) - \mu_v)^2\,\mathrm{d}t} \in \mathbb{R}.$$

La fonction $\mathrm{CCF}_{uv} \in \mathcal{F}(\mathbb{R}, \mathbb{R})$ des corrélations croisées (normalisées) entre les signaux temporels u et v est définie par [Venables et Ripley, 2013 ; Derrick et Thomas, 2004] :

$$\forall \tau \in \mathbb{R}, \quad \mathrm{CCF}_{uv}(\tau) = \frac{1}{\sigma_u \sigma_v} \int_{\mathbb{R}} (u(t) - \mu_u)(v(t+\tau) - \mu_v)\,\mathrm{d}t \in \mathbb{R}.$$

D'après l'inégalité de Cauchy-Schwartz pour des fonctions de carrés intégrables (puis changement de variable linéaire $s = t + \tau$), on a : pour tout $\tau \in \mathbb{R}$,

— $\mathrm{CCF}_{uv}(\tau) \in [-1, 1]$,

— $\mathrm{CCF}_{uv}(\tau) \in \{-1, 1\}$ si et seulement s'il existe $C \in \mathbb{R}$ tel que : $\forall t \in \mathbb{R},\ u(t) - \mu_u = C(v(t+\tau) - \mu_v)$ (corrélations parfaites au sens du traitement du signal [Bracewell, 1965]).

Ainsi, la fonction CCF_{uv} donne une mesure normalisée des corrélations (au sens du traitement du signal [Bracewell, 1965]) entre les signaux temporels u et v pour tous les décalages temporels (appelés *lags*).

En pratique, les signaux temporels sont mesurés sur des pas de temps d'une période donnée, formant des séries temporelles discrétisées. Dans ce cas, faute de données disponibles en dehors de la période considérée, les décalages temporels ne sont pas entièrement définis. Nous choisissons dans la suite d'ignorer les valeurs qui sortent de la période considérées.

Les formules ci-dessus s'adaptent dans le cas où les séries temporelles u et v sont définies sur une grille $(t_1, \ldots, t_n)$ de n pas de temps réguliers (où $n \in \mathbb{N}^*$) :

$$\forall \tau \in \{1, \ldots, n\}, \quad \mathrm{CCF}_{uv}(\tau) = \frac{1}{(n-\tau)\sigma_u \sigma_v} \sum_{k=1}^{n-\tau} (u(t_k) - \mu_u)(v(t_{k+\tau}) - \mu_v) \in [-1, 1],$$

où $\mu_u \in \mathbb{R}$ et $\mu_v \in \mathbb{R}$ désignent cette fois les moyennes empiriques, i.e.

$$\mu_u = \frac{1}{n} \sum_{k=1}^{n} u(t_k) \in \mathbb{R}, \quad \mu_v = \frac{1}{n} \sum_{k=1}^{n} v(t_k) \in \mathbb{R},$$

et $\sigma_u \in \mathbb{R}_+$ et $\sigma_v \in \mathbb{R}_+$ les écarts-types empiriques, i.e.

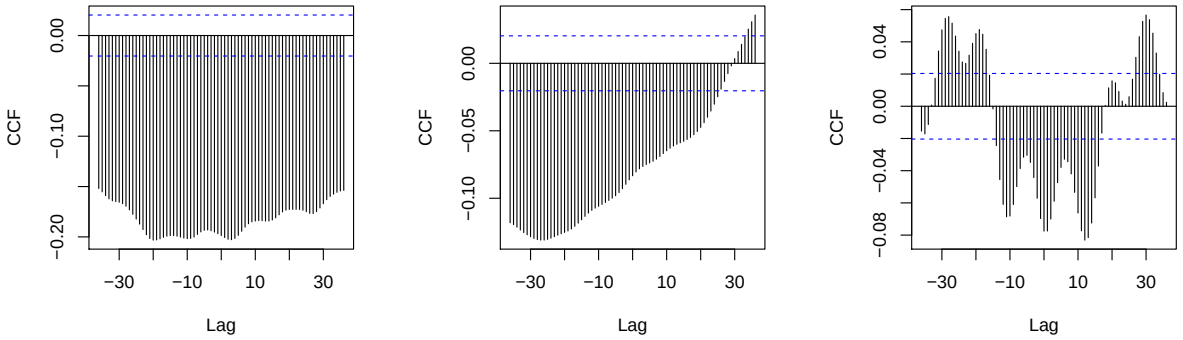
$$\sigma_u = \sqrt{\frac{1}{n} \sum_{k=1}^{n} (u(t_k) - \mu_u)^2} \in \mathbb{R}_+, \quad \sigma_v = \sqrt{\frac{1}{n} \sum_{k=1}^{n} (v(t_k) - \mu_v)^2} \in \mathbb{R}_+.$$

On retrouve ainsi le sens statistique des corrélations croisées : la fonction CCF_{uv} détermine dans quelle mesure les séries temporelles u et v varient ensemble par rapport au cas où elles varient séparément. Pour un *lag* nul (i.e. $\tau = 0$), on retrouve la définition du coefficient de corrélation de Pearson [Benesty *et al.*, 2009]. Il convient de noter que le nombre de valeurs utilisées pour calculer les corrélations croisées diminue donc lorsque le *lag* augmente, réduisant la qualité d'estimation des corrélations pour des *lags* élevés proches de n.

Il existe un seuil critique au-delà duquel les corrélations croisées sont considérées comme statistiquement significatives : si les séries temporelles u et v sont indépendantes (i.e. qu'elles sont considérées comme des réalisations de processus stochastiques indépendants), alors un intervalle de confiance à 95% des corrélations croisées est $[-1.96/\sqrt{n}, 1.96/\sqrt{n}]$ (1.96 est la valeur, à 10^{-2} près, du quantile à 2.5% de la loi normale centrée réduite) [Meng *et al.*, 1992].

A.1.2 Application

La fonction des corrélations croisées est calculée pour les apports en eau (cumulés), les prix et les productions variables pris deux à deux. Les données sont les réalisations au pas horaire sur la période du 01/01/2013 au 25/12/2014 pour les apports en eau cumulés des 6 affluents du Haut-Rhône (observations associées aux données du §2.4.4.3), du 01/01/2013 au 16/11/2015 pour les prix et du 11/08/2014 au 16/11/2015 pour les productions variables d'une partie des actifs de production du périmètre CNR (observations associées aux données du §2.4.4.5). Le calcul de la fonction des corrélations croisées entre deux séries temporelles s'effectue en utilisant l'intersection des périodes sur lesquelles les séries temporelles sont définies.



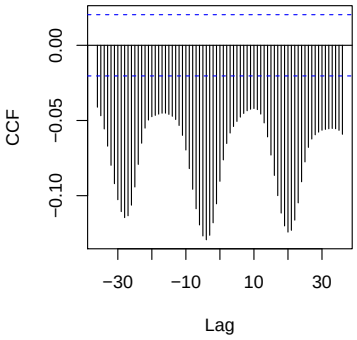
(a) Apports - Prix (b) Apports - Prod. variables (c) Prix - Prod. variables

FIGURE A.1 – Corrélations croisées entre les apports en eau, les prix et les productions variables. Les lignes bleues en tirets représentent les seuils critiques au-delà desquels les corrélations croisées sont statistiquement significatives. Un *lag* correspond à un décalage temporel de 1 h.

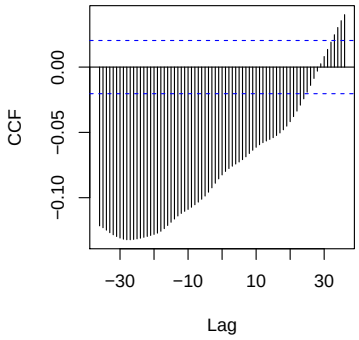
Les résultats sont présentés à la figure A.1. On observe que les corrélations croisées sont statistiquement significatives pour les trois combinaisons possibles. Il convient néanmoins de noter que les corrélations présentées à la figure A.1 sont souvent indirectes. Par exemple, les prix étant définis par rapport à la consommation, ils varient selon l'heure de la journée (voir Sous-section 1.2.1), tout comme la production solaire (production plus forte lorsque le soleil est au zénith et nulle en pleine nuit) et, dans une moindre mesure, les productions éoliennes (certains courants d'air, comme la brise du matin, dépendent de l'heure de la journée). Il n'est donc pas surprenant d'observer les corrélations et la périodicité des pics de la figure A.1c.

Les corrélations croisées sont plus importantes en valeurs absolues entre les apports et les prix que dans les deux autres combinaisons. On remarque que les corrélations croisées sur la figure A.1a sont négatifs, ce qui signifie globalement que les prix diminuent lorsque les apports en eau cumulés du Haut-Rhône augmentent. On remarque également sur la figure A.1a que les pics des corrélations sont régulièrement espacés (d'environ 6 h). Cette observation pourrait être due au fait que le Fier, un des affluents utilisés dans les apports en eau, a un débit influencé par la conduite d'aménagements hydroélectriques effectuée par Électricité De France (EDF) en fonction des prix. Cette périodicité peut également être due à des conditions hydrologiques qui dépendent de l'heure de la journée, comme les prix. L'inversion des corrélations croisées observées sur la figure A.1b entre les apports et les productions variables à partir d'un délai d'environ 1 jour est difficilement explicable. Cette observation pourrait venir du fait que les épisodes de pluie, qui apportent de l'eau, sont généralement précédées ou suivies d'épisodes de vent ou de soleil (alternance entre dépressions et anticyclones qui traversent la France avec les mouvements d'air qui en résultent) faisant augmenter les productions variables et baisser les apports en eau.

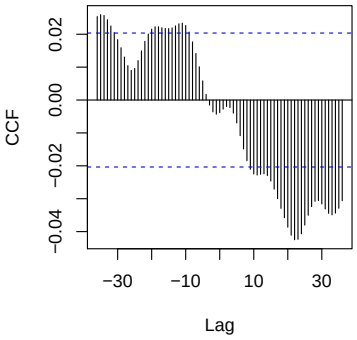
Par curiosité, la figure A.2 donnent les corrélations croisées en distinguant les productions éoliennes et solaires. En particulier, on observe sur la figure A.2a que les productions éoliennes et solaires sont significativement corrélées (du même ordre de grandeur que celles entre les apports et la production éolienne de la figure A.2b) et on retrouve les cycles journaliers dus à l'absence de production solaire en pleine nuit. Les corrélations croisées étant négatives, une augmentation de la production éolienne est globalement associée à une diminution de la production solaire. On observe également que les apports en eau ont des corrélations croisées plus élevées avec les productions éoliennes (Figure A.2b) qu'avec les productions solaires (Figure A.2c). C'est l'inverse pour les prix (Figures A.2d et A.2e) mais cela est biaisé par le fait que les prix et la production solaire sont fortement dépendant de l'heure de la journée.



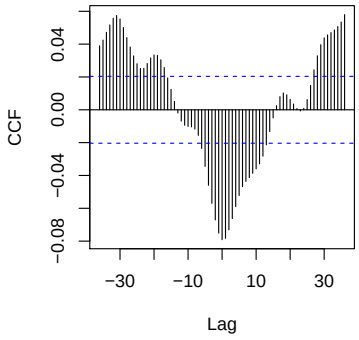
(a) Prod. éolienne - Prod. solaire



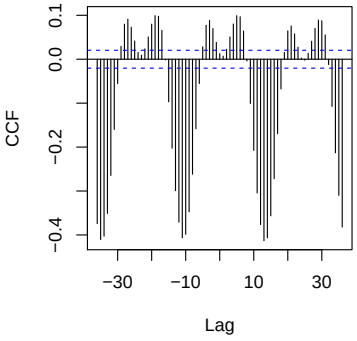
(b) Apports - Prod. éolienne



(c) Apports - Prod. solaire



(d) Prix - Prod. éolienne



(e) Prix - Prod. solaire

FIGURE A.2 – Corrélations croisées entre les apports en eau, les prix et les productions variables en distinguant les productions éoliennes et solaires. Les lignes bleues en tirets représentent les seuils critiques au-delà desquels les corrélations croisées sont statistiquement significatives. Un *lag* correspond à un décalage temporel de 1 h.

En conclusion, les résultats des corrélations croisées montrent que l'hypothèse d'indépendance entre les apports en eau, les prix et les productions variables est contestable. Néanmoins, ces résultats doivent être confirmés sur un historique plus grand, en considérant une période de plusieurs années avec tous les affluents du Rhône et tous les actifs de production variable du périmètre d'équilibre CNR.

A.2 Qualité de la prévision probabiliste des prix

Nous nous intéressons dans cette section à la qualité de la méthode de prévision probabiliste des prix présentée au §2.4.4.4. Nous rappelons que cette méthode repose sur un modèle d'erreurs calibré sur l'historique entre le 01/01/2013 et le 31/12/2016. Ce modèle d'erreurs permet de quantifier les incertitudes autour d'une prévision déterministe à un horizon de 72 h. Une prévision d'ensemble peut alors être obtenue par échantillonnage et réarrangement de Schaake (voir §2.4.4.4).

L'étude de la qualité de la méthode de prévision probabiliste des prix s'effectue sur l'historique de vérification entre le 01/01/2017 et le 31/12/2020, en s'appuyant sur deux attributs : la fiabilité et la finesse (voir §1.4.5.2).

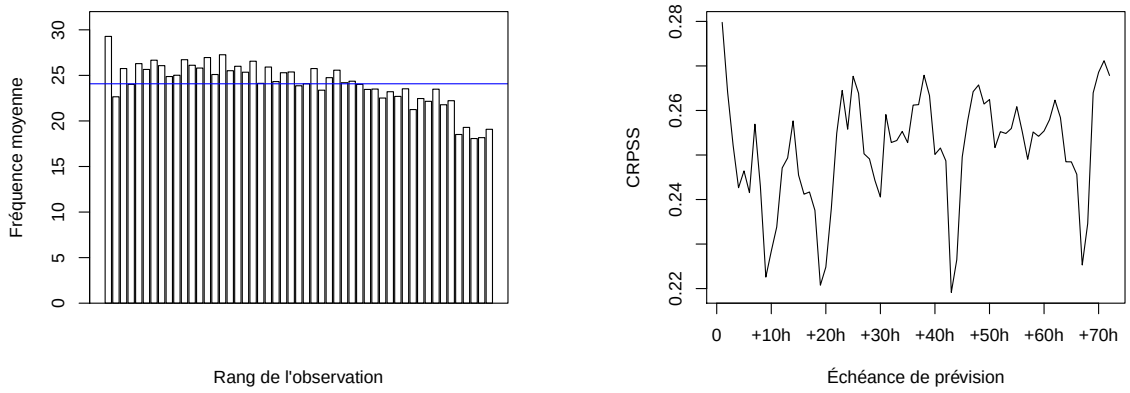
La fiabilité peut être évaluée graphiquement à l'aide de l'histogramme de rang (diagramme de Talagrand) [Hamill, 2001 ; Talagrand, 1999]. Il s'agit de l'histogramme, construit sur l'historique de vérification, du rang de l'observation par rapport aux membres de la prévision d'ensemble pour chaque échéance de prévision. Si la prévision d'ensemble est suffisamment fiable, alors l'histogramme de rang est presque plat (les observations de l'historique sont alors réparties équitablement entre les membres des prévisions d'ensemble de l'historique). Si au contraire la prévision d'ensemble présente respectivement des biais ou des défauts de dispersion, alors l'histogramme de rang est respectivement penché ou en forme de $\cap$ ou $\cup$. Les histogrammes de rang permettent donc de vérifier graphiquement des conditions nécessaires de fiabilité ou de défaut de fiabilité pour une méthode de modélisation des incertitudes donnée.

La figure A.3a donne l'histogramme de rang moyen pour la méthode de prévision probabiliste des prix sur l'historique de vérification. Cet histogramme est obtenu en prenant la moyenne des fréquences des histogrammes de rang de chaque échéance. La ligne bleue représente la fréquence que l'on aurait obtenue si la méthode de prévision probabiliste était parfaitement fiable. On observe que les premiers rang sont légèrement sur-représentés par rapport aux derniers. Cela peut être un signe d'un léger biais négatif de la méthode de prévision probabiliste. Malgré ce défaut, la fiabilité de la méthode de prévision probabiliste des prix est globalement satisfaisante.

La fiabilité et la finesse peuvent être mesurées par le CRPS moyen sur l'historique de vérification (voir §1.4.5.2). Comme il est difficile d'interpréter les valeurs des CRPS d'une méthode de prévision, le CRPS est généralement utilisé en comparaison avec un CRPS d'une méthode de prévision de référence (par exemple, une autre méthode de prévision ou celle obtenue à une étape antérieure). On utilise alors le score de compétence [Ahrens et Walser, 2008] associé au CRPS moyen, le CRPSS (*Continuous Ranked Probability Skill Score*) défini par :

$$\mathrm{CRPSS} = 1 - \frac{\mathrm{CRPS}}{\mathrm{CRPS}_{\mathrm{ref}}} \in]-\infty, 1],$$

où $\mathrm{CRPS} \in \mathbb{R}_+$ est le CRPS moyen de la méthode de prévision probabiliste étudiée et $\mathrm{CRPS}_{\mathrm{ref}} \in \mathbb{R}_+$ le CRPS moyen de la méthode de prévision de référence.



(a) Histogramme de rang

(b) CRPSS

FIGURE A.3 – Qualité de la méthode de prévision probabiliste des prix sur l'historique de vérification.

Le CRPSS est un score orienté négativement :

- un CRPSS égal à 1 correspond à des prévisions parfaites (CRPS moyen nul),
- un CRPSS compris entre 0 et 1 indique que les prévisions de la méthode étudiée sont de meilleure qualité que celles de la méthode de référence,
- un CRPS égal à 0 indique que la qualité de la méthode de prévision étudiée est similaire à celle de la méthode de référence,
- un CRPS négatif indique que les prévisions de la méthode étudiée sont de moins bonne qualité que celles de la méthode de référence.

En pratique, le choix de la méthode de référence n'est pas évident car il faut veiller à ce qu'il ne surévalue pas artificiellement le CRPSS [Pappenberger *et al.*, 2015]. Dans notre cas, nous considérons que la méthode de prévision de référence est la prévision déterministe des prix autour de laquelle le modèle d'erreurs est appliqué. Le CRPS de référence est alors l'erreur absolue moyenne (MAE) de la méthode de prévision déterministe des prix sur l'historique de vérification (voir §1.4.5.2).

La figure A.3b donne l'évolution du CRPSS en fonction de l'échéance de prévision. On observe que les CRPSS sont compris dans l'intervalle $[0.2, 0.3]$, donc la prévision probabiliste des prix permet d'obtenir une meilleure qualité que celle associée à la prévision déterministe [1]. On observe également une légère baisse du CRPSS aux échéances qui correspondent aux pics de consommation du matin et du soir où les prix sont généralement plus élevés (l'échéance 0 correspond au début du lendemain).

En conclusion, cette méthode de prévision probabiliste des prix a une qualité suffisante pour illustrer et tester la méthodologie d'optimisation présentée dans ce manuscrit. Des études pourraient être menées pour améliorer la qualité, par exemple en effectuant le réarrangement de Schaake avec des situations analogues ou en calibrant un modèle d'erreurs paramétrique. En revanche, l'utilisation d'un modèle d'erreurs calibré sur un historique ne permet pas de quantifier les incertitudes propres aux situations. Pour cela, il faudrait mettre en place des approches de prévision probabiliste des prix qui utilisent les incertitudes sur la situation considérée (principe similaire aux prévisions météorologiques d'ensemble par exemple). Un état de l'art des approches rencontrées dans la littérature est par exemple disponible dans [Nowotarski et Weron, 2018].

1. Il convient de noter que la finesse n'a pas de sens pour une prévision déterministe.

A.3 Réduction des scénarios par distance de Kantorovich

Nous présentons dans cette section la théorie générale de la réduction des scénarios dans le cadre discret et fini, dont la problématique est la suivante. Étant donné un nombre fini de scénarios et leurs probabilités (distribution initiale), nous cherchons à supprimer des scénarios et à redéfinir les probabilités des scénarios conservés (distribution réduite) d'une telle manière que la distribution réduite soit « proche » de la distribution initiale. L'objectif est de pouvoir approcher la distribution initiale des scénarios seulement à partir d'un petit nombre de scénarios.

La littérature en lien avec ce problème est très riche. Dans cette section, nous présentons une approche très populaire, notamment pour l'optimisation sous incertitudes d'actifs de production électrique, qui utilise la distance de Kantorovich pour définir la proximité entre deux lois de probabilité à partir d'une fonction coût. Une large partie de cette section provient donc de [Dupačová *et al.*, 2000, 2003] (pour la théorie originale) et [Heitsch et Römisch, 2003]. Contrairement à la littérature, nous nous restreignons au cadre discret et fini.

Le problème général est divisé en sous-problèmes qui visent à répondre aux trois questions suivantes.

1. Si l'ensemble des scénarios à supprimer est fixé, comment redistribuer les probabilités sur l'ensemble des scénarios conservés sans trop changer la distribution initiale ?
2. Si le nombre de scénarios à supprimer est fixé mais avec la liberté de choisir les scénarios à supprimer, comment déterminer l'ensemble des scénarios à supprimer sur lequel appliquer la redistribution ?
3. Quel algorithme numérique utiliser ? Comment déterminer le nombre de scénarios à supprimer ?

Ces trois problématiques sont respectivement traitées dans les sous-sections A.3.3, A.3.5 et A.3.6. Avant de traiter ces problématiques, la sous-section A.3.1 présente les notations générales et la sous-section A.3.2 définit la distance de Kantorovich pour mesurer la proximité entre deux probabilités finies. La sous-section intermédiaire A.3.4 permet de faire le lien entre la redistribution de la première problématique avec la quantification (discrète) et la classification de Voronoï.

A.3.1 Notations générales

Les notations qui suivent reprennent très largement les notations originales de [Dupačová *et al.*, 2000, 2003]. Il n'y a donc pas de correspondance entre les notations de cette annexe et celles du reste du manuscrit.

Soit $\Omega = \{\omega_1, \dots, \omega_N\}$ un univers fini (au sens de la théorie des probabilités), avec $N \in \mathbb{N}^*$. Les éléments $\omega_i \in \Omega$, pour $i \in \{1, \dots, N\}$, sont les *scénarios*. On rappelle que $\mathfrak{P}(\Omega)$ désigne l'ensemble des parties de l'univers Ω.

Dans cette section, nous considérons que les scénarios sont des éléments d'un espace métrique non vide $(\mathcal{X}, c)$. Ainsi : $\Omega \subset \mathcal{X}$. En général, pour des calculs numériques, $\mathcal{X} = \mathbb{R}^r$ avec $r \in \mathbb{N}^*$. La distance c permet de comparer les scénarios entre eux. Pour tout $(i, j) \in \{1, \dots, N\}^2$, on note $c_{ij} = c(\omega_i, \omega_j) \in \mathbb{R}_+$ la distance entre les scénarios ω_i et ω_j. Il convient de noter que le choix de la distance c n'est pas toujours évident en pratique.

Dans la suite, nous manipulons à la fois les probabilités en tant que mesures sur l'espace probabilisable $(\Omega, \mathfrak{P}(\Omega))$ ou en tant que poids sur les scénarios. Plus précisément, nous utilisons la proposition suivante.

Proposition A.1. *L'ensemble de toutes les probabilités sur* $(\Omega, \mathfrak{P}(\Omega))$ *est en bijection avec l'ensemble*

$$P(\Omega) = \left\{ (p_\omega)_{\omega\in\Omega} \in [0,1]^\Omega \,;\; \sum_{\omega\in\Omega} p_\omega = 1 \right\} = \left\{ (p_i)_{1\leq i\leq N} \in [0,1]^N \,;\; \sum_{i=1}^N p_i = 1 \right\}.$$

Justification. Comme l'univers Ω est fini, définir une probabilité $\mathbb{P}$ sur $(\Omega, \mathfrak{P}(\Omega))$ revient, de manière équivalente, à choisir un vecteur $(p_i)_{1\leq i\leq N} \in P(\Omega)$ en posant $\mathbb{P}[\{\omega_i\}] = p_i$, pour $i \in \{1, \ldots, N\}$ (écriture unique), i.e.

$$\mathbb{P} = \sum_{i=1}^N p_i \delta_{\omega_i},$$

où, pour tout $i \in \{1, \ldots, N\}$, δ_{ω_i} est la masse de Dirac en ω_i, i.e. la probabilité sur $(\Omega, \mathfrak{P}(\Omega))$ définie par :

$$\forall A \in \mathfrak{P}(\Omega), \quad \delta_{\omega_i}(A) = \begin{cases} 1 & \text{si } \omega_i \in A, \\ 0 & \text{sinon.} \end{cases}$$

□

A.3.2 Distance de Kantorovich entre deux probabilités finies

Nous présentons dans cette sous-section la distance de Kantorovich qui permet de mesurer la proximité entre deux probabilités finies. Nous commençons par quelques notations.

Soient $\mathbb{P}$ et $\mathbb{Q}$ deux probabilités sur $(\Omega, \mathfrak{P}(\Omega))$. On écrit

$$\mathbb{P} = \sum_{i=1}^N p_i \delta_{\omega_i} \quad \text{et} \quad \mathbb{Q} = \sum_{j=1}^N q_j \delta_{\omega_j},$$

avec $(p_i)_{1\leq i\leq N} \in P(\Omega)$ et $(q_j)_{1\leq j\leq N} \in P(\Omega)$ (écritures uniques).

Pour toute probabilité π sur $(\Omega\times\Omega, \mathfrak{P}(\Omega\times\Omega))$, on pose :

$$F(\pi) = \mathbb{E}_\pi[c] = \sum_{\gamma\in c(\Omega\times\Omega)} \gamma\pi[c = \gamma] \in \mathbb{R}_+.$$

La fonction F est à valeurs finies et positives ou nulles car l'ensemble $\Omega\times\Omega$ est fini et la fonction c est positive ou nulle.

On note $\Pi(\mathbb{P}, \mathbb{Q})$ l'ensemble de toutes les probabilités π sur $(\Omega\times\Omega, \mathfrak{P}(\Omega\times\Omega))$ ayant les lois marginales $\mathbb{P}$ et $\mathbb{Q}$ selon respectivement la première et la seconde composante, i.e. telles que :

(i) $\forall i \in \{1, \ldots, N\},\ \pi[\{\omega_i\}\times\Omega] = \mathbb{P}[\{\omega_i\}] = p_i,$
(ii) $\forall i \in \{1, \ldots, N\},\ \pi[\Omega\times\{\omega_j\}] = \mathbb{Q}[\{\omega_j\}] = q_j.$

Les notations précédentes permettent de définir la distance de Kantorovich.

Définition A.2. La *distance de Kantorovich* (ou la *distance de Kantorovich-Rubinstein*, ou la *distance de Wasserstein* d'ordre 1) entre les probabilités $\mathbb{P}$ et $\mathbb{Q}$ (sur l'espace probabilisable $(\Omega, \mathfrak{P}(\Omega))$) est par définition le réel positif ou nul

$$\mathrm{d}_K(\mathbb{P}, \mathbb{Q}) = \inf_{\pi\in\Pi(\mathbb{P},\mathbb{Q})} F(\pi) \in \mathbb{R}_+. \tag{A.1}$$

Justification de la cohérence de la définition A.2. La définition A.2 est cohérente car $\{F(\pi),\ \pi \in \Pi(\mathbb{P}, \mathbb{Q})\}$ est une partie non vide [2] et minorée (par 0) de $\mathbb{R}$. □

2. Par exemple, on peut définir $\pi \in \Pi(\mathbb{P}, \mathbb{Q})$ par : $\forall (i,j) \in \{1, \ldots, N\}^2,\ \pi[\{\omega_i\}\times\{\omega_j\}] = p_i q_j \in [0,1]$.

La distance de Kantorovich entre deux probabilités $\mathbb{P}$ et $\mathbb{Q}$ est alors le coût moyen minimal parmi toutes les probabilités jointes conservant les lois marginales $\mathbb{P}$ et $\mathbb{Q}$. Elle est donc la solution du problème de transport optimal (ou le problème de Monge-Kantorovich, ou le problème de transport de masse) avec la fonction coût c. En effet, si les probabilités représentent des quantités d'unités de masse, la distance de Kantorovich permet de trouver une manière de transporter les masses de la probabilité $\mathbb{P}$ vers la configuration de la probabilité $\mathbb{Q}$ sur le même espace Ω tout en minimisant le coût moyen (les poids $c_{ij} \in \mathbb{R}_+$, pour $(i,j) \in \{1,\dots,N\}^2$, représentent alors le coût pour transporter une unité de masse de ω_i vers ω_j). Pour plus de détails sur le transport optimal, voir [Villani, 2009 ; Rachev et Rüschendorf, 1998].

La suite de cette sous-section vise à démontrer que la distance de Kantorovich est bien une distance sur l'espace probabilisable $(\Omega, \mathfrak{P}(\Omega))$. Nous présentons d'abord le lien entre les probabilités jointes en tant que mesures ou en tant que poids sur les couples de scénarios, de manière analogue à la proposition A.1. Cela permet ensuite de donner une autre formulation de la distance de Kantorovich, plus facile à manipuler, avec laquelle nous démontrons les propriétés qui définissent une distance.

Pour toute probabilité π sur $(\Omega \times \Omega, \mathfrak{P}(\Omega \times \Omega))$, on pose $h(\pi) = (\pi_{ij})_{1\leq i,j\leq N} \in P(\Omega \times \Omega)$ avec $\pi_{ij} = \pi[\{\omega_i\} \times \{\omega_j\}] \in [0,1]$, pour $(i,j) \in \{1,\dots,N\}^2$. Par le même argument que celui de la proposition A.1, la fonction h est bijective et, pour tout $\hat{\pi} = (\pi_{ij})_{1\leq i,j\leq N} \in P(\Omega \times \Omega)$,

$$h^{-1}(\hat{\pi}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_{ij} \delta_{(\omega_i, \omega_j)}.$$

On pose :

$$\Gamma(\mathbb{P}, \mathbb{Q}) = \Bigg\{ (\pi_{ij})_{1\leq i,j\leq N} \in P(\Omega \times \Omega) \ ; \ \forall i \in \{1,\dots,N\}, \ \sum_{j=1}^{N} \pi_{ij} = p_i, \\ \forall j \in \{1,\dots,N\}, \ \sum_{i=1}^{N} \pi_{ij} = q_j \Bigg\} \subset P(\Omega \times \Omega).$$

La proposition suivante permet de donner une formulation de la fonction F en utilisant les poids de la probabilité jointe.

Proposition A.3.

(i) $\Gamma(\mathbb{P}, \mathbb{Q}) = h(\Pi(\mathbb{P}, \mathbb{Q}))$.

(ii) On pose $\hat{F} = F \circ h^{-1}$*, alors, pour tout* $\hat{\pi} = (\pi_{ij})_{1\leq i,j\leq N} \in P(\Omega \times \Omega)$,

$$\hat{F}(\hat{\pi}) = F(h^{-1}(\hat{\pi})) = \sum_{i=1}^{N} \sum_{j=1}^{N} \pi_{ij} c_{ij} \in \mathbb{R}_+.$$

Preuve. (i) Pour $\hat{\pi} = (\pi_{ij})_{1\leq i,j\leq N} \in P(\Omega\times\Omega)$, posant $\pi = h^{-1}(\hat{\pi}) = \sum_{i=1}^{N}\sum_{j=1}^{N}\pi_{ij}\delta_{(\omega_i,\omega_j)}$, il vient :

$$\begin{aligned}
\hat{\pi}\in\Gamma(\mathbb{P},\mathbb{Q}) &\iff \begin{cases}\forall i\in\{1,\dots,N\},\ \sum_{j=1}^{N}\pi_{ij}=p_i,\\ \forall j\in\{1,\dots,N\},\ \sum_{i=1}^{N}\pi_{ij}=q_j,\end{cases}\\
&\iff \begin{cases}\forall i\in\{1,\dots,N\},\ \sum_{j=1}^{N}\pi[\{\omega_i\}\times\{\omega_j\}]=p_i,\\ \forall j\in\{1,\dots,N\},\ \sum_{i=1}^{N}\pi[\{\omega_i\}\times\{\omega_j\}]=q_j,\end{cases}\\
&\iff \begin{cases}\forall i\in\{1,\dots,N\},\ \pi[\{\omega_i\}\times\Omega]=p_i,\\ \forall j\in\{1,\dots,N\},\ \pi[\Omega\times\{\omega_j\}]=q_j,\end{cases}\\
&\iff h^{-1}(\hat{\pi})=\pi\in\Pi(\mathbb{P},\mathbb{Q}).
\end{aligned}$$

(ii) Soit $\hat{\pi} = (\pi_{ij})_{1\leq i,j\leq N} \in P(\Omega\times\Omega)$. On pose $\pi = h^{-1}(\hat{\pi}) = \sum_{i=1}^{N}\sum_{j=1}^{N}\pi_{ij}\delta_{(\omega_i,\omega_j)}$. Alors,

$$\hat{F}(\hat{\pi}) = F(\pi) = \sum_{\gamma\in c(\Omega\times\Omega)}\gamma\pi[c=\gamma] = \sum_{\gamma\in c(\Omega\times\Omega)}\gamma\sum_{\substack{1\leq i,j\leq N\\ c_{ij}=\gamma}}\pi_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{N}c_{ij}\pi_{ij}.$$

□

Les notations et résultats précédents permettent de donner une nouvelle expression de la distance de Kantorovich plus facile à manipuler.

Théorème A.4. *La borne inférieure de l'équation* (A.1) *est atteinte et on a l'égalité :*

$$\mathrm{d}_K(\mathbb{P},\mathbb{Q}) = \min_{\hat{\pi}\in\Gamma(\mathbb{P},\mathbb{Q})}\hat{F}(\hat{\pi})\in\mathbb{R}_+. \tag{A.2}$$

Preuve. D'après la proposition A.3 et la bijectivité de h,

$$\mathrm{d}_K(\mathbb{P},\mathbb{Q}) = \inf_{\pi\in\Pi(\mathbb{P},\mathbb{Q})}F(\pi) = \inf_{\hat{\pi}\in h(\Pi(\mathbb{P},\mathbb{Q}))}F(h^{-1}(\hat{\pi})) = \inf_{\hat{\pi}\in\Gamma(\mathbb{P},\mathbb{Q})}\hat{F}(\hat{\pi}).$$

Pour tout $\hat{x} = (x_{ij})_{1\leq i,j\leq N}\in\mathbb{R}^{N^2}$, on pose :

$$\begin{cases}\forall i\in\{1,\dots,N\},\ f_i(\hat{x}) = \sum_{j=1}^{N}x_{ij}\in\mathbb{R},\\ \forall j\in\{1,\dots,N\},\ g_j(\hat{x}) = \sum_{i=1}^{N}x_{ij}\in\mathbb{R}.\end{cases}$$

Alors, pour tout $i\in\{1,\dots,N\}$ (respectivement $j\in\{1,\dots,N\}$), la fonction f_i (respectivement g_i) est continue et l'ensemble $\{p_i\}$ (respectivement $\{q_j\}$) est fermé dans $\mathbb{R}$, donc l'ensemble $f_i^{-1}(\{p_i\})$ (respectivement $g_j^{-1}(\{q_j\})$) est fermé dans $\mathbb{R}^{N^2}$. Ainsi, l'ensemble

$$\Gamma(\mathbb{P},\mathbb{Q}) = \left(\bigcap_{1\leq i\leq N}f_i^{-1}(\{p_i\})\right)\cap\left(\bigcap_{1\leq j\leq N}g_j^{-1}(\{q_j\})\right)$$

est fermé dans $\mathbb{R}^{N^2}$. En outre, $\Gamma(\mathbb{P},\mathbb{Q}) \subset [0,1]^{N^2}$, donc l'ensemble $\Gamma(\mathbb{P},\mathbb{Q})$ est borné (dans $\mathbb{R}^{N^2}$). Le $\mathbb{R}$-espace vectoriel $\mathbb{R}^{N^2}$ étant de dimension finie, on en déduit que l'ensemble $\Gamma(\mathbb{P},\mathbb{Q})$ est un compact de $\mathbb{R}^{N^2}$.

Par ailleurs, l'application $\hat{F}_0$ définie sur le $\mathbb{R}$-espace vectoriel $\mathbb{R}^{N^2}$ par

$$\forall \hat{x} = (x_{ij})_{1\leq i,j\leq N} \in \mathbb{R}^{N^2},\ \hat{F}_0(\hat{x}) = \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}x_{ij} \in \mathbb{R}$$

est linéaire, donc continue.

Ainsi, la borne inférieure de $\hat{F}_0$ sur $\Gamma(\mathbb{P},\mathbb{Q})$ est atteinte : il existe $\hat{\pi}^* \in \Gamma(\mathbb{P},\mathbb{Q})$ tel que

$$\hat{F}_0(\hat{\pi}^*) = \inf_{\hat{\pi}\in\Gamma(\mathbb{P},\mathbb{Q})} \hat{F}_0(\hat{\pi}).$$

Comme $\hat{F}_0 = \hat{F} = F \circ h^{-1}$ sur $\Gamma(\mathbb{P},\mathbb{Q})$, il vient

$$\mathrm{d}_K(\mathbb{P},\mathbb{Q}) = \min_{\hat{\pi}\in\Gamma(\mathbb{P},\mathbb{Q})} \hat{F}(\hat{\pi}) = F(\pi^*),$$

avec $\pi^* = h^{-1}(\hat{\pi}^*) \in \Pi(\mathbb{P},\mathbb{Q})$. □

Cette nouvelle expression permet de démontrer que la distance de Kantorovich est bien une distance.

Proposition A.5. *L'application* d_K *est une distance sur l'ensemble de toutes les probabilités sur* $(\Omega,\mathfrak{P}(\Omega))$.

Preuve. D'après l'équation (A.2), pour toutes probabilités $\mathbb{P}$ et $\mathbb{Q}$ sur $(\Omega,\mathfrak{P}(\Omega))$, on peut écrire :

$$\mathrm{d}_K(\mathbb{P},\mathbb{Q}) = \min_{(\pi_{ij})_{i,j}\in\Gamma(\mathbb{P},\mathbb{Q})} \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\pi_{ij}.$$

— L'application d_K est bien définie et à valeurs dans $\mathbb{R}_+$.

— Soient $\mathbb{P}$ et $\mathbb{Q}$ des probabilités sur $(\Omega,\mathfrak{P}(\Omega))$. L'application T définie sur $\Gamma(\mathbb{P},\mathbb{Q})$ en échangeant les indices, i.e. telle que, pour tout $\hat{\pi} = (\pi_{ij})_{1\leq i,j\leq N} \in \Gamma(\mathbb{P},\mathbb{Q})$,

$$T(\hat{\pi}) = \left(\pi'_{ij}\right)_{1\leq i,j\leq N} \in P(\Omega\times\Omega), \quad \text{avec } \forall (i,j)\in\{1,\dots,N\}^2,\ \pi'_{ij} = \pi_{ji},$$

est une bijection de $\Gamma(\mathbb{P},\mathbb{Q})$ dans $\Gamma(\mathbb{Q},\mathbb{P})$. Ainsi,

$$d_K(\mathbb{P},\mathbb{Q}) = \min_{(\pi_{ij})_{i,j}\in\Gamma(\mathbb{P},\mathbb{Q})} \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\pi_{ij} = \min_{(\pi'_{k\ell})_{k,\ell}\in\Gamma(\mathbb{Q},\mathbb{P})} \sum_{\ell=1}^{N}\sum_{k=1}^{N} c_{\ell k}\pi'_{k\ell},$$

donc, par symétrie de l'application c,

$$d_K(\mathbb{P},\mathbb{Q}) = \min_{(\pi'_{ij})_{i,j}\in\Gamma(\mathbb{Q},\mathbb{P})} \sum_{k=1}^{N}\sum_{\ell=1}^{N} c_{k\ell}\pi'_{k\ell} = \mathrm{d}_K(\mathbb{Q},\mathbb{P}).$$

— La preuve qui suit de l'inégalité triangulaire est adaptée de la définition 6.1 de [Villani, 2009].

Soient $\mathbb{P}_1$, $\mathbb{P}_2$ et $\mathbb{P}_3$ des probabilités sur $(\Omega, \mathfrak{P}(\Omega))$. On écrit (voir Proposition A.1) :

$$\mathbb{P}_1 = \sum_{i=1}^{N} p_i^{(1)} \delta_{\omega_i}, \quad \mathbb{P}_2 = \sum_{j=1}^{N} p_j^{(2)} \delta_{\omega_j} \quad \text{et} \quad \mathbb{P}_3 = \sum_{k=1}^{N} p_k^{(3)} \delta_{\omega_k},$$

avec $\left(p_i^{(1)}\right)_{1\leq i\leq N} \in P(\Omega)$, $\left(p_j^{(2)}\right)_{1\leq j\leq N} \in P(\Omega)$ et $\left(p_k^{(3)}\right)_{1\leq k\leq N} \in P(\Omega)$ (écritures uniques).
D'après l'équation (A.2), on peut écrire :

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\pi_{ij}^{(12)} = \mathrm{d}_K(\mathbb{P}_1, \mathbb{P}_2) \quad \text{et} \quad \sum_{j=1}^{N}\sum_{k=1}^{N} c_{jk}\pi_{jk}^{(23)} = \mathrm{d}_K(\mathbb{P}_2, \mathbb{P}_3),$$

avec $\hat{\pi}^{(12)} = \left(\pi_{ij}^{(12)}\right)_{1\leq i,j\leq N} \in \Gamma(\mathbb{P}_1, \mathbb{P}_2)$ et $\hat{\pi}^{(23)} = \left(\pi_{jk}^{(23)}\right)_{1\leq j,k\leq N} \in \Gamma(\mathbb{P}_2, \mathbb{P}_3)$.
Pour tout $(i,j,k) \in \{1,\dots,N\}^3$, on pose :

$$\mu_{ijk} = \begin{cases} \frac{\pi_{ij}^{(12)}\pi_{jk}^{(23)}}{p_j^{(2)}} & \text{si } p_j^{(2)} > 0, \\ 0 & \text{sinon.} \end{cases}$$

Alors $(\mu_{ijk})_{1\leq i,j,k\leq N} \in P(\Omega\times\Omega\times\Omega)$ et

$$\begin{cases} \forall (i,j) \in \{1,\dots,N\}^2, & \sum\limits_{k=1}^{N} \mu_{ijk} = \pi_{ij}^{(12)}, \\ \forall (j,k) \in \{1,\dots,N\}^2, & \sum\limits_{i=1}^{N} \mu_{ijk} = \pi_{jk}^{(23)}. \end{cases}$$

On pose également $\hat{\pi}^{(13)} = \left(\pi_{ik}^{(3)}\right)_{1\leq i,k\leq N} \in P(\Omega\times\Omega)$, où

$$\forall (i,k) \in \{1,\dots,N\}^2, \ \pi_{ik}^{(13)} = \sum_{j=1}^{N} \mu_{ijk}.$$

Alors $\hat{\pi}^{(13)} \in \Gamma(\mathbb{P}_1, \mathbb{P}_3)$ (mais pas nécessairement optimale pour la distance de Kantorovich entre $\mathbb{P}_1$ et $\mathbb{P}_3$). En outre, comme c est une distance,

$$\forall (i,j,k) \in \{1,\dots,N\}^3, \quad c_{ik} = c(\omega_i,\omega_k) \leq c(\omega_i,\omega_j) + c(\omega_j,\omega_k) = c_{ij} + c_{jk}.$$

D'où

$$\begin{aligned} \mathrm{d}_K(\mathbb{P}_1,\mathbb{P}_3) \leq \sum_{i=1}^{N}\sum_{k=1}^{N} c_{ik}\pi_{ik}^{(13)} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} c_{ik}\mu_{ijk} &\leq \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} (c_{ij}+c_{jk})\mu_{ijk} \\ &\leq \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\sum_{k=1}^{N}\mu_{ijk} + \sum_{j=1}^{N}\sum_{k=1}^{N} c_{jk}\sum_{i=1}^{N}\mu_{ijk} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\pi_{ij}^{(12)} + \sum_{j=1}^{N}\sum_{k=1}^{N} c_{jk}\pi_{jk}^{(23)} \\ &= \mathrm{d}_K(\mathbb{P}_1,\mathbb{P}_2) + \mathrm{d}_K(\mathbb{P}_2,\mathbb{P}_3). \end{aligned}$$

— Soient $\mathbb{P}$ et $\mathbb{Q}$ des probabilités sur $(\Omega, \mathfrak{P}(\Omega))$ telles que $\mathrm{d}_K(\mathbb{P},\mathbb{Q}) = 0$.
D'après l'équation (A.2), il existe $\hat{\pi} = (\pi_{ij})_{i,j} \in \Gamma(\mathbb{P},\mathbb{Q})$ tel que

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \underbrace{c_{ij}\pi_{ij}}_{\geq 0} = \mathrm{d}_K(\mathbb{P},\mathbb{Q}) = 0.$$

Or, $\forall i \in \{1, \dots, N\},\ c_{ii} = 0$. Donc

$$\forall (i,j) \in \{1, \dots, N\}^2,\ i \neq j, \quad \pi_{ij} = 0.$$

En outre, d'après la proposition A.3, $\pi = h^{-1}(\hat{\pi}) \in \Pi(\mathbb{P}, \mathbb{Q})$, donc

$$\begin{cases} \forall i \in \{1, \dots, N\}, \quad p_i = \pi[\{\omega_i\} \times \Omega] = \sum_{j=1}^{N} \pi_{ij} = \pi_{ii}, \\ \forall j \in \{1, \dots, N\}, \quad q_j = \pi[\Omega \times \{\omega_j\}] = \sum_{i=1}^{N} \pi_{ij} = \pi_{jj}. \end{cases}$$

D'où $\mathbb{P} = \mathbb{Q}$.

□

Les formulations de la distance de Kantorovich aux équations (A.1) et (A.2) ne sont pas toujours faciles à manipuler en raison des probabilités jointes. Le théorème suivant donne une formulation équivalente de la distance de Kantorovich, basée sur le problème d'optimisation dual du problème d'optimisation linéaire (A.2). L'énoncé et la preuve sont adaptés du chapitre 5 de [Villani, 2009] pour le cadre fini et discret. Pour plus de détails sur la théorie de la dualité des problèmes d'optimisation linéaires (programmation linéaire), voir par exemple [Allaire, 2005].

Théorème A.6 (Dualité de Kantorovich)**.** *On considère l'ensemble*

$$\Gamma^*(\mathbb{P}, \mathbb{Q}) = \left\{ (u_i, v_j)_{1 \leq i,j \leq N} \in (\mathbb{R}^2)^N\ ;\ \forall (i,j) \in \{1, \dots, N\}^2,\ u_i + v_j \leq c_{ij} \right\}.$$

Alors on a la représentation primale-duale de la distance de Kantorovich suivante :

$$\mathrm{d}_K(\mathbb{P}, \mathbb{Q}) = \min_{(\pi_{ij})_{i,j} \in \Gamma(\mathbb{P}, \mathbb{Q})} \sum_{i=1}^{N} \sum_{j=1}^{N} c_{ij} \pi_{ij} = \max_{(u_i, v_j)_{i,j} \in \Gamma^*(\mathbb{P}, \mathbb{Q})} \left\{ \sum_{i=1}^{N} p_i u_i + \sum_{j=1}^{N} q_j v_j \right\}.$$

Cette représentation primale-duale peut se retrouver directement et facilement par la théorie de dualité en programmation linéaire [Allaire, 2005]. Nous présentons dans la suite de cette sous-section une preuve *ad hoc*, beaucoup plus laborieuse mais qui a l'avantage de pouvoir se généraliser pour des scénarios qui ne sont pas discrets.

Cette preuve *ad hoc* s'appuie sur le lemme suivant.

Lemme A.7. *Soit $\hat{\pi}^* = (\pi_{ij})_{1 \leq i,j \leq N} \in \Gamma(\mathbb{P}, \mathbb{Q})$ tel que (voir Équation (A.2)) :*

$$\mathrm{d}_K(\mathbb{P}, \mathbb{Q}) = \hat{F}(\hat{\pi}^*) = \sum_{i=1}^{N} \sum_{j=1}^{N} c_{ij} \pi_{ij}.$$

On note $S = \{(i,j) \in \{1, \dots, N\}^2\ ;\ \pi_{ij} > 0\}$ le support de $\hat{\pi}^*$.

Alors, pour toute famille finie $(i_k, j_k)_{1 \leq k \leq m} \in S^m$ (avec $m \in \mathbb{N}^*$),

$$\sum_{k=1}^{m} c_{i_k j_k} \leq \sum_{k=1}^{m} c_{i_{k+1} j_k},$$

où, par convention, $i_{m+1} = i_1$.

Preuve du lemme. On suppose par contradiction qu'il existe une famille finie $(i_k, j_k)_{1\leq k\leq m} \in S^m$ (avec $m \in \mathbb{N}^*$) telle que

$$\sum_{k=1}^{m} c_{i_k j_k} > \sum_{k=1}^{m} c_{i_{k+1} j_k}, \tag{A.3}$$

avec la convention $i_{m+1} = i_1$. Pour tout $(i, j) \in \{1, \dots, N\}^2$, on pose

$$\forall k \in \{1, \dots, m\}, \quad \gamma_{ij}(k) = \delta_{i_{k+1}, j_k}(i, j) - \delta_{i_k, j_k}(i, j) \in \{-1, 0, 1\}.$$

On utilise les unions *disjointes* suivantes :

$$\{1, \dots, m\} = \bigcup_{1\leq i\leq N} \{k \in \{1, \dots, m\};\ i = i_{k+1}\} \tag{A.4}$$

$$= \bigcup_{1\leq i\leq N} \{k \in \{1, \dots, m\};\ i = i_k\} \tag{A.5}$$

$$= \bigcup_{1\leq j\leq N} \{k \in \{1, \dots, m\};\ j = j_k\}. \tag{A.6}$$

Soit $\tilde{\pi} = (\tilde{\pi}_{ij})_{1\leq i,j\leq N} \in \mathbb{R}^{N^2}$ tel que

$$\forall (i, j) \in \{1, \dots, N\}^2, \quad \tilde{\pi}_{ij} = \pi_{ij} + a \sum_{k=1}^{m} \gamma_{ij}(k),$$

avec $a = \frac{1}{m} \min_{1\leq k\leq m} \pi_{i_k j_k} \in \mathbb{R}_+^*$ (par définition de S).

Les points suivants montrent que $\tilde{\pi} \in \Gamma(\mathbb{P}, \mathbb{Q})$.

— Soit $j \in \{1, \dots, N\}$. On a

$$\sum_{i=1}^{N} \sum_{k=1}^{m} \gamma_{ij}(k) = \sum_{i=1}^{N} \sum_{\substack{1\leq k\leq m \\ i_{k+1}=i}} \delta_{j_k}(j) - \sum_{i=1}^{N} \sum_{\substack{1\leq k\leq m \\ i_k=i}} \delta_{j_k}(j) = \sum_{k=1}^{m} \delta_{j_k}(j) - \sum_{k=1}^{m} \delta_{j_k}(j) = 0,$$

où la seconde égalité vient des réunions disjointes (A.4) et (A.5). Ainsi :

$$\sum_{i=1}^{N} \tilde{\pi}_{ij} = \sum_{i=1}^{N} \pi_{ij} + a \sum_{i=1}^{N} \sum_{k=1}^{m} \gamma_{ij}(k) = \sum_{i=1}^{N} \pi_{ij} = q_j.$$

— Soit $i \in \{1, \dots, N\}$. On a

$$\sum_{j=1}^{N} \sum_{k=1}^{m} \gamma_{ij}(k) = \sum_{j=1}^{N} \sum_{\substack{1\leq k\leq m \\ j_k=j}} \left(\delta_{i_{k+1}}(i) - \delta_{i_k}(i)\right) = \sum_{k=1}^{m} \left(\delta_{i_{k+1}}(i) - \delta_{i_k}(i)\right) = \delta_{i_{m+1}}(i) - \delta_{i_1}(i),$$

où la seconde égalité vient de la réunion disjointe (A.6).

Comme $i_{m+1} = i_1$, il vient

$$\sum_{j=1}^{N} \sum_{k=1}^{m} \gamma_{ij}(k) = 0.$$

Ainsi :

$$\sum_{j=1}^{N} \tilde{\pi}_{ij} = \sum_{j=1}^{N} \pi_{ij} + a \sum_{j=1}^{N} \sum_{k=1}^{m} \gamma_{ij}(k) = \sum_{j=1}^{N} \pi_{ij} = p_i.$$

— Par définition de a,

$$\forall k \in \{1,\dots,m\}, \quad a \leq \frac{1}{m}\pi_{i_k j_k},$$

donc, pour tout $(i,j) \in \{1,\dots,N\}^2$,

$$\tilde{\pi}_{ij} \geq \pi_{ij} - \sum_{k=1}^{m} a\delta_{i_k j_k}(i,j) \geq \pi_{ij} - \frac{1}{m}\sum_{k=1}^{m} \underbrace{\pi_{i_k j_k}\delta_{i_k j_k}(i,j)}_{\leq \pi_{ij}} \geq 0.$$

En outre, d'après le point précédent,

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \tilde{\pi}_{ij} = \sum_{i=1}^{N} p_i = 1.$$

Ainsi : $\forall (i,j) \in \{1,\dots,N\}^2,\ \tilde{\pi}_{ij} \in [0,1]$.

D'où $\tilde{\pi} \in \Gamma(\mathbb{P},\mathbb{Q})$.

En outre,

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\sum_{k=1}^{N}\gamma_{ij}(k) &= \sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{k=1}^{m} c_{ij}\delta_{i_{k+1}j_k}(i,j) - \sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{k=1}^{m} c_{ij}\delta_{i_k j_k}(i,j) \\
&= \sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{\substack{1\leq k\leq m \\ i_{k+1}=i}} c_{i_{k+1}j}\delta_{j_k}(j) - \sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{\substack{1\leq k\leq m \\ j_k=i}} c_{i_k j}\delta_{j_k}(j) \\
&\overset{(a)}{=} \sum_{j=1}^{N}\sum_{k=1}^{m} c_{i_{k+1}j}\delta_{j_k}(j) - \sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{k=1}^{m} c_{i_k j}\delta_{j_k}(j) \\
&= \sum_{j=1}^{N}\sum_{\substack{1\leq k\leq m \\ j_k=j}} c_{i_{k+1}j_k} - \sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{\substack{1\leq k\leq m \\ j_k=j}} c_{i_k j_k} \\
&\overset{(b)}{=} \sum_{k=1}^{m} c_{i_{k+1}j_k} - \sum_{k=1}^{m} c_{i_k j_k} \overset{(c)}{<} 0,
\end{aligned}$$

où les égalités (a) et (b) viennent des réunions disjointes (A.4)-(A.5)-(A.6) et l'inégalité (c) vient de l'équation (A.3). Ainsi :

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\tilde{\pi}_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\pi_{ij} + \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\sum_{k=1}^{N}\gamma_{ij}(k) < \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\pi_{ij} = \mathrm{d}_K(\mathbb{P},\mathbb{Q}),$$

d'où la contradiction avec la définition de $\mathrm{d}_K(\mathbb{P},\mathbb{Q})$. □

Preuve du théorème de dualité de Kantorovich.
D'après l'équation (A.2), il existe $\hat{\pi}^* = (\pi_{ij})_{1\leq i,j\leq N} \in \Gamma(\mathbb{P},\mathbb{Q})$ telle que

$$\mathrm{d}_K(\mathbb{P},\mathbb{Q}) = \hat{F}(\hat{\pi}) = \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\pi_{ij}.$$

— Soit $(u_i, v_j)_{1\leq i,j\leq N} \in \Gamma^*(\mathbb{P},\mathbb{Q})$. Par définition de $\Gamma(\mathbb{P},\mathbb{Q})$,

$$\sum_{i=1}^{N} p_i u_i + \sum_{j=1}^{N} q_j v_j = \sum_{i=1}^{N} u_i\left(\sum_{j=1}^{N}\pi_{ij}\right) + \sum_{j=1}^{N} v_j\left(\sum_{i=1}^{N}\pi_{ij}\right) = \sum_{i=1}^{N}\sum_{j=1}^{N}(u_i+v_j)\pi_{ij}, \qquad \text{(A.7)}$$

donc, par définition de $\Gamma^*(\mathbb{P},\mathbb{Q})$, on obtient

$$\sum_{i=1}^{N} p_i u_i + \sum_{j=1}^{N} q_j v_j \leq \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\pi_{ij} = \mathrm{d}_K(\mathbb{P},\mathbb{Q}).$$

D'où

$$\forall\, (u_i, v_j)_{1\leq i,j\leq N} \in \Gamma^*(\mathbb{P},\mathbb{Q}), \quad \sum_{i=1}^{N} p_i u_i + \sum_{j=1}^{N} q_j v_j \leq \mathrm{d}_K(\mathbb{P},\mathbb{Q}).$$

— Soit $S = \{(i,j) \in \{1,\dots,N\}^2\,;\ \pi_{ij} > 0\}$ le support de $\hat{\pi}^*$. L'ensemble S est non vide car $\Gamma(\mathbb{P},\mathbb{Q}) \in P(\Omega\times\Omega)$.
Soit $(i_0, j_0) \in S$.
Soit $i \in \{1,\dots,N\}$. D'après le lemme A.7, pour tout $m \in \mathbb{N}^*$ et tout $(i_k, j_k)_{1\leq k\leq m} \in S^m$,

$$\begin{aligned}\sum_{k=0}^{m-1}(c_{i_{k+1}j_k} - c_{i_k j_k}) + (c_{ij_m} - c_{i_m j_m}) &= \sum_{k=1}^{m}(c_{i_{k+1}j_k} - c_{i_k j_k}) + c_{i_1 j_0} + c_{ij_m} - c_{i_0 j_0} - c_{i_m j_m}\\ &\geq c_{i_1 j_0} + c_{ij_m} - c_{i_0 j_0} - c_{i_m j_m} \geq -2C,\end{aligned}$$

avec $C = \max\limits_{1\leq i,j\leq m} c_{ij} \in \mathbb{R}_+$ (indépendant de m et $(i_k,j_k)_{1\leq k\leq m}$).
Ainsi, pour tout $i \in \{1,\dots,N\}$, la borne inférieure suivante est réelle :

$$u_i = \inf_{m\in\mathbb{N}^*}\left\{\max_{(i_k,j_k)_{1\leq k\leq m}\in S^m}\left(\sum_{k=0}^{m-1}(c_{i_{k+1}j_k} - c_{i_k j_k}) + (c_{ij_m} - c_{i_m j_m})\right)\right\} \in \mathbb{R}. \tag{A.8}$$

Pour tout $j \in \{1,\dots,N\}$, on pose également :

$$v_j = \min_{1\leq \ell\leq N}(c_{\ell j} - u_\ell) \in \mathbb{R}. \tag{A.9}$$

D'après l'équation (A.9),

$$\forall (i,j) \in \{1,\dots,N\}^2,\ u_i + v_j \leq c_{ij}, \tag{A.10}$$

donc $(u_i, v_j)_{1\leq i,j\leq N} \in \Gamma^*(\mathbb{P},\mathbb{Q})$.
Soit $(i,j) \in \{1,\dots,N\}^2$.
Pour tout $\ell \in \{1,\dots,N\}$, en forçant $i_m = i$ et $j_m = j$ dans la définition de u_ℓ (voir l'équation (A.8)), on obtient :

$$\begin{aligned}u_\ell &\leq \inf_{m\geq 2}\left\{\max_{(i_k,j_k)_{1\leq k\leq m-1}\in S^{m-1}}\left(\sum_{k=0}^{m-2}(c_{i_{k+1}j_k} - c_{i_k j_k}) + (c_{ij_{m-1}} - c_{i_{m-1}j_{m-1}}) + (c_{\ell j} - c_{ij})\right)\right\}\\
&= \inf_{m\geq 2}\left\{\max_{(i_k,j_k)_{1\leq k\leq m-1}\in S^{m-1}}\left(\sum_{k=0}^{m-2}(c_{i_{k+1}j_k} - c_{i_k j_k}) + (c_{ij_{m-1}} - c_{i_{m-1}j_{m-1}})\right)\right\} + c_{\ell j} - c_{ij}\\
&= \inf_{n\in\mathbb{N}^*}\left\{\max_{(i_k,j_k)_{1\leq k\leq n}\in S^{n}}\left(\sum_{k=0}^{n-1}(c_{i_{k+1}j_k} - c_{i_k j_k}) + (c_{ij_n} - c_{i_n j_n})\right)\right\} + c_{\ell j} - c_{ij}\\
&= u_i + c_{\ell j} - c_{ij}.\end{aligned}$$

Ainsi, $\forall \ell \in \{1,\dots,N\},\ c_{\ell j} - u_\ell \geq c_{ij} - u_i$.
D'où $v_j = \min\limits_{1\leq \ell\leq N}(c_{\ell j} - u_\ell) \geq c_{ij} - u_i$, i.e. $u_i + v_j \geq c_{ij}$.

Avec l'équation (A.10), on a donc

$$\forall (i,j) \in \{1,\dots,N\}^2, \quad u_i + v_j = c_{ij},$$

puis, en injectant cette égalité dans l'équation (A.7),

$$\sum_{i=1}^{N} p_i u_i + \sum_{j=1}^{N} q_j v_j = \sum_{i=1}^{N}\sum_{j=1}^{N} (u_i + v_j)\pi_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\pi_{ij} = \mathrm{d}_K(\mathbb{P},\mathbb{Q}).$$

Les deux points précédents montrent que :

$$\mathrm{d}_K(\mathbb{P},\mathbb{Q}) = \max_{(u,v)\in\Gamma^*(\mathbb{P},\mathbb{Q})} \left\{ \sum_{i=1}^{N} p_i u_i + \sum_{j=1}^{N} q_j v_j \right\}.$$

□

A.3.3 Redistribution optimale

Dans cette sous-section, nous fixons l'ensemble des scénarios à supprimer et nous nous intéressons au problème de redistribution optimale des scénarios conservés. Nous continuons à utiliser les notations précédentes.

Pour le problème de réduction des scénarios, les éléments de Ω correspondent aux scénarios initiaux. On considère que les scénarios initiaux sont l'image d'une variable aléatoire discrète finie X d'un ensemble probabilisé $(\mathcal{U},\mathcal{A},\mathbf{P})$ dans l'ensemble probabilisable fini $(\Omega,\mathfrak{P}(\Omega))$. On note $\mathbb{P}_X = \mathbf{P}\circ X^{-1}$ la loi de la variable aléatoire X. Ainsi, $\mathbb{P}_X$ est une probabilité sur l'ensemble probabilisable fini $(\Omega,\mathfrak{P}(\Omega))$, appelée la *probabilité initiale*. Les réels $p_i = \mathbb{P}_X[\{\omega_i\}] \in [0,1]$, pour $i \in \{1,\dots,N\}$, désignent alors les probabilités des scénarios initiaux. D'après la proposition A.1, on a donc $(p_i)_{1\le i\le N} \in P(\Omega)$ et (écriture unique)

$$\mathbb{P}_X = \sum_{i=1}^{N} p_i \delta_{\omega_i}.$$

Quitte à réduire la taille de l'univers, on peut supposer que les probabilités initiales sont strictes :

$$\forall i \in \{1,\dots,N\}, \quad p_i = \mathbb{P}_X[\{\omega_i\}] \in]0,1[.$$

Soit J un sous-ensemble strict et non vide de $\{1,\dots,N\}$ qui correspond aux indices des scénarios que l'on choisit de supprimer. On note :

- $\overline{J} = \{1,\dots,N\} \setminus J \neq \emptyset$, les indices des scénarios conservés,
- $\Omega_J = \{\omega_i,\ i \in J\} \subset \Omega$, les scénarios supprimés,
- $\Omega_{\overline{J}} = \{\omega_j,\ j \in \overline{J}\} \subset \Omega$, les scénarios conservés.

On note $\mathcal{Q}_J$ l'ensemble des probabilités $\mathbb{Q}$ sur $(\Omega,\mathfrak{P}(\Omega))$ dont le support est égal à $\overline{J}$, i.e. telles que, pour $j \in \{1,\dots,N\}$, $\mathbb{Q}[\{\omega_j\}] = 0 \iff j \in J$. Une probabilité $\mathbb{Q} \in \mathcal{Q}_J$ peut être vue comme la loi d'une variable aléatoire Y de $(\mathcal{U},\mathcal{A},\mathbf{P})$ dans $(\Omega,\mathfrak{P}(\Omega))$ avec $Y(\mathcal{U}) = \Omega_{\overline{J}}$. Avec le même argument que celui à la fin de la sous-section A.3.1, l'ensemble $\mathcal{Q}_J$ est en bijection avec l'ensemble

$$\hat{\mathcal{Q}}_J = \Big\{ (q_j)_{1\le j\le N} \in P(\Omega)\,;\ \forall j \in J,\ q_j = 0 \text{ et } \forall j \in \overline{J},\ q_j > 0 \Big\}.$$

Plus précisément, pour tout $\hat{\mathbb{Q}} = (q_j)_{1\le j\le N} \in P(\Omega)$, $\hat{\mathbb{Q}} \in \hat{\mathcal{Q}}_J \iff \mathbb{Q} = \sum_{j=1}^{N} q_j \delta_{\omega_j} \in \mathcal{Q}_J$.

Comparée à la probabilité initiale $\mathbb{P}_X$, une probabilité $\mathbb{Q} \in \mathcal{Q}_J$ est *réduite* dans le sens où les scénarios ω_i, pour $i \in J$, sont supprimés (leurs probabilités sont mises à zéro) et que les autres scénarios ω_j, pour $j \in \overline{J}$ sont conservés et distribués par les nouvelles probabilités q_j.

Avec ces notations, la redistribution optimale pour l'ensemble J fixé consiste à construire une probabilité $\mathbb{Q} \in \mathcal{Q}_J$ qui est la plus proche possible (au sens de la distance de Kantorovich) de la probabilité initiale $\mathbb{P}_X$, i.e. qui minimise la fonction D_J définie sur $\mathcal{Q}_J$ par :

$$\forall \mathbb{Q} \in \mathcal{Q}_J, \quad D_J(\mathbb{Q}) = \mathrm{d}_K(\mathbb{P}_X, \mathbb{Q}) \in \mathbb{R}_+.$$

Il s'agit d'un problème de transport optimal particulier dont la solution explicite est formulée par le théorème et la preuve qui suivent (très largement repris du théorème 2 de [Dupačová *et al.*, 2003]).

Théorème A.8 (Redistribution optimale)**.**
Pour tout $i \in J$, on pose : $\alpha(i) \in \underset{k \in \overline{J}}{\arg\min}\, c_{ik} \in \overline{J}$.

Pour tout $j \in \{1, \dots, N\}$, on pose également :

$$q_j^* = \begin{cases} 0 & \text{si } j \in J, \\ p_j + \sum\limits_{i \in J_j} p_i & \text{si } j \in \overline{J}, \end{cases} \tag{A.11}$$

où $J_j = \{i \in J\,;\ j = \alpha(i)\} \subset J$. On note $\mathbb{Q}^* = \sum\limits_{j=1}^{N} q_j^* \delta_{\omega_j} = \sum\limits_{j \in \overline{J}} q_j^* \delta_{\omega_j}$. Alors :

(i) $\left(q_j^*\right)_{1 \leq j \leq N} \in \hat{\mathcal{Q}}_J$ *and* $\mathbb{Q}^* \in \mathcal{Q}_J$,

(ii) $\mathbb{Q}^*$ *est une* redistribution optimale *pour l'ensemble J, i.e. $\mathbb{Q}^*$ est un minimum de la fonction D_J dans $\mathcal{Q}_J$, et*

$$D_J(\mathbb{Q}^*) = \min_{\mathbb{Q} \in \mathcal{Q}_J} D_J(\mathbb{Q}) = \sum_{i \in J} p_i \min_{j \in \overline{J}} c_{ij}. \tag{A.12}$$

Le théorème A.8 assure l'existence d'une redistribution optimale pour l'ensemble J et en donne une construction explicite : la nouvelle probabilité q_j^* d'un scénario conservé ω_j (avec $j \in \overline{J}$) est la somme de la probabilité initiale p_j et des probabilités initiales des scénarios supprimés qui lui sont les plus proches par rapport à la distance c (voir Équation (A.11)). En général, il peut exister plusieurs façons de définir l'application α, donc la redistribution optimale pour l'ensemble J n'est pas nécessairement unique. En revanche, la valeur optimale de l'équation (A.12) est indépendante de la définition de l'application α.

Preuve.

(i) Pour tout $i \in J$, on a $i \in J_j$ où $j = \alpha(i) \in \overline{J}$. En outre, soit $i \in J_j \cap J_{j'}$ (avec $(j, j') \in \overline{J}^2$), alors $j = \alpha(i) = j'$. On en déduit la réunion *disjointe* suivante : $J = \bigcup\limits_{j \in \overline{J}} J_j$.

Ainsi,

$$\mathbb{P}_X[\{\omega_i,\ i \in J\}] = \mathbb{P}_X\left[\bigcup_{j \in \overline{J}} \{\omega_i,\ i \in J_j\}\right] = \sum_{j \in \overline{J}} \mathbb{P}_X[\{\omega_i,\ i \in J_j\}] = \sum_{j \in \overline{J}} \sum_{i \in J_j} p_i. \tag{A.13}$$

Pour tout $j \in J$, $q_j^* \in \mathbb{R}_+$ et, d'après l'équation (A.13),

$$\sum_{j=1}^{N} q_j^* = \sum_{j \in \overline{J}} p_j + \sum_{j \in \overline{J}} \sum_{i \in J_j} p_i = \mathbb{P}_X[\{\omega_j,\ j \in \overline{J}\}] + \mathbb{P}_X[\{\omega_i,\ i \in J\}] = \mathbb{P}_X[\Omega] = 1,$$

donc $\left(q_j^*\right)_{1\leq j\leq N} \in P(\Omega)$, i.e. $\mathbb{Q}^*$ est une probabilité sur $(\Omega, \mathfrak{P}(\Omega))$.
En outre, pour $j \in \{1,\dots,J\}$, $q_j^* = 0 \iff j \in J$ (car $p_j > 0$),
donc $\left(q_j^*\right)_{1\leq j\leq N} \in \hat{\mathcal{Q}}_J$ et, par bijectivité, $\mathbb{Q}^* \in \mathcal{Q}_J$.

(ii) — Soit $\mathbb{Q} = \sum_{j=1}^{N} q_j \delta_{\omega_j} \in \mathcal{Q}_J$, où $(q_j)_{1\leq 1\leq N} \in \hat{\mathcal{Q}}_J$ (écriture unique).
D'après le théorème A.6, on a la représentation primale-duale de $D_J(\mathbb{Q})$ suivante :

$$D_J(\mathbb{Q}) = \min_{(\pi_{ij})_{ij}\in\Gamma(\mathbb{P}_X,\mathbb{Q})} \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\pi_{ij} = \max_{(u_i,v_j)_{ij}\in\Gamma^*(\mathbb{P}_X,\mathbb{Q})} \left\{ \sum_{i=1}^{N} p_i u_i + \sum_{j=1}^{N} q_j v_j \right\}. \tag{A.14}$$

— Pour tout $(i,j) \in \{1,\dots,N\}^2$, on pose :

$$\pi_{ij} = \begin{cases} p_i & \text{si } i \in J_j \text{ et } j \in \overline{J}, \\ p_i & \text{si } i = j \text{ et } j \in \overline{J}, \\ 0 & \text{sinon.} \end{cases}$$

Alors :

· $\forall (i,j) \in \{1,\dots,N\}^2, \quad \pi_{ij} \geq 0,$

· $\forall i \in \{1,\dots,N\}, \quad \sum_{j=1}^{N} \pi_{ij} = \begin{cases} \sum_{\substack{1\leq j\leq N \\ j=\alpha(i)}} p_i & \text{si } i \in J, \\ p_i & \text{si } i \in \overline{J}, \end{cases} = p_i,$

· $\forall j \in \{1,\dots,N\}, \quad \sum_{i=1}^{N} \pi_{ij} = \begin{cases} 0 & \text{si } j \in J, \\ p_j + \sum_{i\in J_j} p_i & \text{si } j \in \overline{J}, \end{cases} = q_j^*.$

Donc $(\pi_{ij})_{1\leq i,j\leq N} \in \Gamma(\mathbb{P}_X, \mathbb{Q}^*)$. On en déduit, d'après la représentation primale (le problème de minimisation de l'équation (A.14)) pour $\mathbb{Q}^* \in \mathcal{Q}_J$,

$$D_J(\mathbb{Q}^*) = \mathrm{d}_K(\mathbb{P}_X, \mathbb{Q}^*) \leq \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\pi_{ij}.$$

En outre,

$$\forall i \in \{1,\dots,N\}, \quad \sum_{j=1}^{N} c_{ij}\pi_{ij} = \begin{cases} \sum_{\substack{1\leq j\leq N \\ j=\alpha(i)}} p_i c_{ij} & \text{si } i \in J, \\ p_i c_{ii} & \text{si } i \in \overline{J}, \end{cases} = \begin{cases} p_i c_{i\alpha(i)} & \text{si } i \in J, \\ 0 & \text{si } i \in \overline{J}. \end{cases}$$

Ainsi

$$D_J(\mathbb{Q}^*) \leq \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\pi_{ij} = \sum_{i\in J} p_i c_{i\alpha(i)} = \sum_{i\in J} p_i \min_{j\in\overline{J}} c_{ij}. \tag{A.15}$$

— Soit $\mathbb{Q} = \sum_{j=1}^{N} q_j \delta_{\omega_j} \in \mathcal{Q}_J$, où $(q_j)_{1\leq 1\leq N} \in \hat{\mathcal{Q}}_J$ (écriture unique).
Pour tout $i \in \{1,\dots,N\}$, on pose :

$$u_i = \begin{cases} 0 & \text{si } i \in \overline{J}, \\ c_{i\alpha(i)} & \text{si } i \in J, \end{cases} \in \mathbb{R}.$$

Pour tout $j \in \{1, \dots, N\}$, on pose également $v_j = \min_{1 \leq \ell \leq N} (c_{\ell j} - u_\ell) \in \mathbb{R}$. Alors,

$$\forall (i,j) \in \{1, \dots, N\}^2, \quad u_i + v_j \leq c_{ij}.$$

En outre, pour tout $j \in \overline{J}$, $v_j = 0$ et, pour tout $j \in J$, $q_j = 0$. Donc, d'après la représentation duale (le problème de maximisation de l'équation (A.14)) pour $\mathbb{Q} \in \mathcal{Q}_J$,

$$\sum_{i \in J} p_i \min_{j \in \overline{J}} c_{ij} = \sum_{i \in J} p_i c_{i\alpha(i)} = \sum_{i=1}^{N} p_i u_i = \sum_{i=1}^{N} p_i u_i + \sum_{j=1}^{N} q_j v_j \leq D_J(\mathbb{Q}).$$

Ainsi,

$$\forall \mathbb{Q} \in \mathcal{Q}_J, \quad \sum_{i \in J} p_i \min_{j \in \overline{J}} c_{ij} \leq D_J(\mathbb{Q}). \tag{A.16}$$

Les équations (A.15) et (A.16) montrent que $\mathbb{Q}^*$ est un point de minimum de la fonction D_J dans $\mathcal{Q}_J$, et que l'on a l'équation (A.12).

□

A.3.4 Lien avec la quantification et la classification de Voronoï

Dans cette sous-section, nous montrons qu'il existe un lien entre le théorème A.8 de redistribution optimale et la quantification. Une large partie de cette sous-section est issue de [Pagès, 2014 ; Corlay et Pagès, 2015] mais en conservant les notations précédentes. Dans la suite, $\mathbb{E}$ désigne l'espérance pour des variables aléatoires à valeurs dans l'espace probabilisé $(\mathcal{U}, \mathcal{A}, \mathbf{P})$.

Soit $\mathcal{C}_J = (C_j)_{j \in \overline{J}}$ une partition de Voronoï de Ω engendrée par $\Omega_{\overline{J}}$ par rapport à la distance c, i.e. telle que

$$\forall j \in \overline{J},\ C_j \subset \left\{ \omega \in \Omega\,;\ c(\omega, \omega_j) \leq \min_{k \in \overline{J}} c(\omega, \omega_k) \right\}.$$

Pour tout $i \in \{1, \dots, N\}$, on pose :

$$q_{\mathcal{C}_J}(\omega_i) = \sum_{j \in \overline{J}} \omega_j \chi_{C_j}(\omega_i) \in \Omega,$$

où, pour tout $j \in \overline{J}$, χ_{C_j} est l'application caractéristique de l'ensemble $C_j \subset \Omega$, i.e. l'application de Ω dans $\{0, 1\}$ définie par

$$\forall \omega \in \Omega, \quad \chi_{C_j}(\omega) = \begin{cases} 1 & \text{si } \omega_i \in C_j, \\ 0 & \text{sinon.} \end{cases}$$

Remarque A.9. Comme $\mathcal{C}_J = (C_j)_{j \in \overline{J}}$ est une partition de Ω, pour tout $i \in \{1, \dots, N\}$, il existe un unique $\alpha(i) \in \overline{J}$ tel que $\omega_i \in C_{\alpha(i)}$ et $\forall j \in \overline{J} \setminus \{\alpha(i)\}$, $\omega_i \notin C_j$, donc $q_{\mathcal{C}_J}(\omega_i) = \omega_{\alpha(i)} \in \Omega$. Ainsi, l'application $q_{\mathcal{C}_J}$ est bien à valeurs dans Ω.

Proposition A.10.

(i) L'application (discrète) $q_{\mathcal{C}_J}$ est mesurable de $(\Omega, \mathfrak{P}(\Omega))$ dans $(\Omega, \mathfrak{P}(\Omega))$.

(ii) On a : $q_{\mathcal{C}_J}(\Omega) = \Omega_{\overline{J}}$.

(iii) Pour tout $i \in \{1, \dots, N\}$, $\operatorname{dist}(\omega_i, \Omega_{\overline{J}}) = \min_{j \in \overline{J}} c_{ij} = c(\omega_i, q_{\mathcal{C}_J}(\omega_i))$.

Preuve.

(i) Le premier point est vrai car l'application $q_{\mathcal{C}_J}$ est une combinaison linéaire de fonctions mesurables de $(\Omega, \mathfrak{P}(\Omega))$ dans $(\Omega, \mathfrak{P}(\Omega))$.

(ii) Avec la définition de l'application α de la remarque A.9, pour tout $i \in \{1, \dots, N\}$, $\alpha(i) \in \overline{J}$, donc $q_{\mathcal{C}_J}(\omega_i) = \omega_{\alpha(i)} \in \Omega_{\overline{J}}$. Ainsi, $q_{\mathcal{C}_J}(\Omega) \subset \Omega_{\overline{J}}$.
Inversement, soit $\omega_j \in \Omega_{\overline{J}}$ (où $j \in \overline{J}$). Alors, $\omega_j \in C_j \subset \Omega$ et, pour tout $k \in \overline{J} \setminus \{j\}$, $\omega_j \notin C_k$, donc $q_{\mathcal{C}_J}(\omega_j) = \omega_j$. Ainsi, $\Omega_{\overline{J}} \subset q_{\mathcal{C}_J}(\Omega)$.

(iii) Soit $i \in \{1, \dots, N\}$.

— Comme l'ensemble $\overline{J}$ est fini, il vient :

$$\operatorname{dist}(\omega_i, \Omega_{\overline{J}}) = \inf_{j \in \overline{J}} c(\omega_i, \omega_j) = \min_{j \in \overline{J}} c_{ij}.$$

— Comme $\omega_i \in C_{\alpha(i)}$ et $q_{\mathcal{C}_J}(\omega_i) = \omega_{\alpha(i)}$, il vient, par définition de $C_{\alpha(i)}$:

$$c(\omega_i, q_{\mathcal{C}_J}(\omega_i)) = c(\omega_i, \omega_{\alpha(i)}) \leq \min_{j \in \overline{J}} c(\omega_i, \omega_j) = \min_{j \in \overline{J}} c_{ij}.$$

Or $\alpha(i) \in \overline{J}$, donc l'inégalité précédente est une égalité. D'où

$$c(\omega_i, q_{\mathcal{C}_J}(\omega_i)) = \min_{j \in \overline{J}} c_{ij}.$$

□

Définition A.11 (Quantification de Voronoï)**.**

— La variable aléatoire $q_{\mathcal{C}_J}(X) = q_{\mathcal{C}_J} \circ X$, de $(\mathcal{U}, \mathcal{A}, \mathbf{P})$ dans $(\Omega, \mathfrak{P}(\Omega))$, est appelée la *quantification de Voronoï* de X associée à la partition de Voronoï $\mathcal{C}_J$.

— L'*erreur de quantification* d'ordre 1 induite par la quantification de Voronoï $q_{\mathcal{C}_J}(X)$ est par définition le réel positif ou nul

$$e(X, \mathcal{C}_J) = \mathbb{E}[c(X, q_{\mathcal{C}_J}(X))] = \sum_{i=1}^{N} p_i c(\omega_i, q_{\mathcal{C}_J}(\omega_i)) \in \mathbb{R}_+.$$

La proposition suivante est adaptée de la proposition 3.1 de [Kreitmeier, 2011] et de la proposition 1.3 de [Pagès, 2014]. Dans ce manuscrit, nous donnons une nouvelle preuve à partir du théorème A.8 (redistribution optimale) et la remarque suivante.

Remarque A.12. Avec la définition de l'application α dans la remarque A.9, le point (iii) de la proposition A.10 montre que, pour tout $i \in J$, $\alpha(i) \in \arg\min_{k \in \overline{J}} c_{ik}$, comme dans les hypothèses du théorème A.8. En revanche, contrairement au théorème A.8, l'application α définie dans la remarque A.9 est aussi définie sur $\overline{J}$ par : $\forall i \in \overline{J},\ \alpha(i) = i$ (voir la preuve du point (ii) de la proposition A.10).

Proposition A.13. *On considère la redistribution optimale $\mathbb{Q}^*$ pour l'ensemble J définie par le théorème A.8 (licite d'après la remarque précédente). Alors :*

(i) $e(X, \mathcal{C}_J) = \mathbb{E}[\operatorname{dist}(X, \Omega_{\overline{J}})] = \sum_{i \in J} p_i \min_{j \in \overline{J}} c_{ij} = D_J(\mathbb{Q}^*)$,

(ii) $\mathbb{Q}^* = \mathbb{P}_X \circ q_{\mathcal{C}_J}^{-1}$.

Preuve.

(i) D'après le point (iii) de la proposition A.10,

$$e(X, \mathcal{C}_J) = \sum_{i=1}^{N} p_i c(\omega_i, q_{\mathcal{C}_J}(\omega_i)) = \sum_{i=1}^{N} p_i \operatorname{dist}(\omega_i, \Omega_{\overline{J}}) = \mathbb{E}[\operatorname{dist}(X, \Omega_{\overline{J}})],$$

et

$$e(X, \mathcal{C}_J) = \sum_{i=1}^{N} p_i \operatorname{dist}(\omega_i, \Omega_{\overline{J}}) = \sum_{i=1}^{N} p_i \min_{j \in \overline{J}} c_{ij} = \sum_{i \in J} p_i \min_{j \in \overline{J}} c_{ij}.$$

Ainsi, d'après le théorème A.8, il vient :

$$e(X, \mathcal{C}_J) = \min_{\mathbb{Q} \in \mathcal{Q}_J} D_J(\mathbb{Q}) = D_J(\mathbb{Q}^*).$$

(ii) — L'application $q_{\mathcal{C}_J}$ est mesurable de $(\Omega, \mathfrak{P}(\Omega))$ dans $(\Omega, \mathfrak{P}(\Omega))$, donc $\mathbb{P}_X \circ q_{\mathcal{C}_J}^{-1}$ est une probabilité sur $(\Omega, \mathfrak{P}(\Omega))$.

— Soit $j \in J$. Comme $\omega_j \notin \Omega_{\overline{J}}$, il vient, d'après le point (ii) de la proposition A.10 :

$$q_{\mathcal{C}_J}^{-1}(\{\omega_j\}) = \emptyset.$$

Ainsi :

$$(\mathbb{P}_X \circ q_{\mathcal{C}_J}^{-1})[\{\omega_j\}] = \mathbb{P}_X[q_{\mathcal{C}_J}^{-1}(\{\omega_j\})] = \mathbb{P}_X[\emptyset] = 0 = q_j^*.$$

— Soit $j \in \overline{J}$. D'après la remarque A.9,

$$\begin{aligned} q_{\mathcal{C}_J}^{-1}(\{\omega_j\}) &= \{\omega_i\,;\ i \in \{1, \dots, N\} \text{ et } \omega_j = \underbrace{q(\omega_i)}_{=\omega_{\alpha(i)}}\} \\ &= \{\omega_i\,;\ i \in \{1, \dots, N\} \text{ et } j = \alpha(i)\}. \end{aligned}$$

Comme $\alpha(i) = i$ pour $i \in \overline{J}$, les égalités précédentes permettent d'écrire la réunion *disjointe* suivante :

$$q_{\mathcal{C}_J}^{-1}(\{\omega_j\}) = \{\omega_j\} \cup \{\omega_i,\ i \in J_j\} = \{\omega_j\} \cup \left(\bigcup_{i \in J_j} \{\omega_i\} \right),$$

avec l'ensemble $J_j \subset J$ défini comme dans le théorème A.8. Ainsi :

$$(\mathbb{P}_X \circ q_{\mathcal{C}_J}^{-1})[\{\omega_j\}] = \mathbb{P}_X[q_{\mathcal{C}_J}^{-1}(\{\omega_j\})] = p_j + \sum_{i \in J_j} p_i = q_j^*.$$

D'où : $\mathbb{Q}^* = \mathbb{P}_X \circ q_{\mathcal{C}_J}^{-1}$.

□

La proposition précédente montre que l'erreur de quantification est égale à la valeur optimale de la redistribution avec la distance de Kantorovich (voir Théorème A.8). Le lien entre l'erreur de quantification (d'ordre 1) et la distance de Kantorovich a déjà été établie dans la littérature [Graf et Luschgy, 2009 ; Kreitmeier, 2011] pour le cas général où l'univers n'est pas nécessairement fini. En outre, le point (i) de la proposition précédente montre que l'erreur de quantification (et donc la valeur optimale de la redistribution) est indépendante de la partition de Voronoï $\mathcal{C}_J$ (et donc du choix de l'application α). Le point (iii) de la proposition précédente montre que la redistribution optimale pour l'ensemble J est complètement définie par une partition de Voronoï engendrée par $\Omega_{\overline{J}}$ par rapport à la distance c. Avec la définition 1.2 de [Villani, 2009], la relation $\mathbb{Q}^* = \mathbb{P}_X \circ q_{\mathcal{C}_J}^{-1}$ montre que $(\mathbb{P}_X, \mathbb{Q}^*)$ induit un couplage déterministe, résultat bien connu [Villani, 2009].

A.3.5 Choix optimal des scénarios à supprimer à nombre fixé

Dans cette sous-section, nous nous intéressons au choix optimal des scénarios à supprimer lorsque leur nombre est fixé.

Soit $n \in \{0, \dots, N-1\}$. On note $\mathcal{J}_n$ l'ensemble des parties de $\{1, \dots, N\}$ de cardinal n. Avec les résultats de la sous-section A.3.3, le problème que nous cherchons à résoudre est le suivant :

$$\min_{J \in \mathcal{J}_n} \left\{ \sum_{i \in J} p_i \min_{j \in \overline{J}} c_{ij} \right\} = \min_{J \in \mathcal{J}_n} \min_{\mathbb{Q} \in \mathcal{Q}_J} D_J(\mathbb{Q}). \tag{A.17}$$

Le problème de minimisation (A.17) est bien défini d'après le théorème A.8 et car l'ensemble $\mathcal{J}_n$ est fini. Il s'agit d'un problème de couverture par ensembles dont la résolution numérique n'est pas directement accessible en général [Gröwe-Kuska *et al.*, 2003 ; Alfandari et Monnot, 2014].

Les deux propositions suivantes, tirées de [Dupačová *et al.*, 2003], sont des cas particuliers importants où la résolution du problème de minimisation (A.17) est simple.

Proposition A.14 (Suppression d'un unique scénario)**.** *On suppose que* $n = 1$.

Le problème de minimisation (A.17) *s'écrit alors :*

$$\min_{1 \leq \ell \leq N} \left\{ p_\ell \min_{j \in \{1,\dots,N\} \setminus \{\ell\}} c_{\ell j} \right\}. \tag{A.18}$$

Soit $\ell_* \in \{1, \dots, N\}$ *un point de minimum de l'équation* (A.18) *(i.e.* ω_{ℓ_*} *est l'unique scénario supprimé), alors la redistribution optimale pour l'ensemble* $\{\ell_*\}$ *est donnée par :*

$$\mathbb{Q}^* = (p_{\ell_*} + p_{j_*})\, \delta_{\omega_{\ell_*}} + \sum_{\substack{1 \leq \ell \leq N \\ l \notin \{\ell_*, j_*\}}} p_\ell \delta_{\omega_\ell} \in \mathcal{Q}_{\{\ell_*\}},$$

avec $j_* \in \underset{j \in \{1,\dots,N\} \setminus \{\ell_*\}}{\arg\min} c_{\ell_* j}$.

Preuve. Conséquence du théorème A.8. □

Proposition A.15 (Conservation d'un unique scénario)**.** *On suppose que* $n = N - 1$.

Le problème de minimisation (A.17) *s'écrit alors :*

$$\min_{1 \leq u \leq N} \sum_{i=1}^{N} p_i c_{iu}. \tag{A.19}$$

Soit $u_* \in \{1, \dots, N\}$ *un point de minimum de l'équation* (A.19) *(i.e.* ω_{u^*} *est l'unique scénario conservé), alors la redistribution optimale pour l'ensemble* $\{1, \dots, N\} \setminus \{u_*\}$ *est donnée par :* $\mathbb{Q}^* = \delta_{\omega_{u_*}}$.

Preuve. Conséquence du théorème A.8 et :

$$\mathbb{Q}^*[\{\omega_{u_*}\}] = p_{u_*} + \sum_{i \in \{1,\dots,N\} \setminus \{u_*\}} p_i = \sum_{i=1}^{N} p_i = 1.$$

□

Si $n = 0$, alors aucun scénario n'est supprimé ($\mathcal{J}_0 = \{\emptyset\}$) et une redistribution optimale pour l'ensemble $J = \emptyset$ est donnée par $\mathbb{Q}^* = \mathbb{P}_X$ (car d_K est une distance, voir Proposition A.5). Nous supposons à présent que $n \geq 1$.

Le théorème suivant donne une borne minimale et une borne maximale de la valeur optimale du problème d'optimisation (A.17). L'énoncé et la preuve sont très largement inspirés de ceux du théorème 4 de [Dupačová *et al.*, 2003].

Théorème A.16. *On définit une famille $(\ell_i)_{1\leq i\leq n}$ de n indices distincts dans $\{1,\dots,N\}$ en supprimant récursivement les scénarios un à un sans duplication :*

- *ℓ_1 est un point de minimum du problème* (A.18)*,*
- *pour tout $i \in \{2,\dots,n\}$, ℓ_i est un point de minimum de*

$$\min_{\ell\in\{1,\dots,N\}\setminus\{\ell_1,\dots,\ell_{i-1}\}} \left\{ p_\ell \min_{j\in\{1,\dots,N\}\setminus\{\ell\}} c_{\ell j} \right\}.$$

De façon analogue, on définit une famille $(u_j)_{1\leq j\leq N-n}$ de $N-n$ indices distincts dans $\{1,\dots,N\}$ en conservant récursivement les scénarios un à un sans duplication :

- *u_ℓ est un point de minimum du problème* (A.19)*,*
- *pour tout $j \in \{2,\dots,N-n\}$, u_j est un point de minimum de*

$$\min_{u\in\{1,\dots,N\}\setminus\{u_1,\dots,u_{j-1}\}} \sum_{i\in\{1,\dots,N\}\setminus\{u_1,\dots,u_{j-1},u\}} p_i \min_{\ell\in\{u_1,\dots,u_{j-1},u\}} c_{\ell i}.$$

Alors, on a les inégalités suivantes :

$$\sum_{i=1}^{n} p_{\ell_i} \min_{j\in\{1,\dots,N\}\setminus\{\ell_i\}} c_{\ell_i j} \leq \min_{J\in\mathcal{J}_n} \left\{ \sum_{i\in J} p_i \min_{j\in\overline{J}} c_{ij} \right\} \leq \sum_{\substack{i\in\{1,\dots,N\}\\ i\notin\{u_1,\dots,u_{N-n}\}}} p_i \min_{j\in\{u_1,\dots,u_{N-n}\}} c_{ij}. \tag{A.20}$$

Si, en outre, pour tout $i \in \{1,\dots,n\}$,

$$\operatorname*{arg\,min}_{j\in\{1,\dots,N\}\setminus\{\ell_i\}} c_{\ell_i j} \setminus \{\ell_1,\dots,\ell_{i-1},\ell_{i+1},\dots,\ell_n\} \neq \emptyset, \tag{A.21}$$

alors l'ensemble $\{\ell_1,\dots,\ell_n\} \in \mathcal{J}_n$ est un point minimum du problème (A.17)*.*

Preuve.

- La borne supérieure de l'équation (A.20) vient du fait que

$$\begin{cases} J_u = \{1,\dots,N\} \setminus \{u_1,\dots,u_{N-n}\} \in \mathcal{J}_n, \\ \overline{J_u} = \{u_1,\dots,u_{N-n}\}. \end{cases}$$

- Soit $J = \{j_1,\dots,j_n\} \in \mathcal{J}_n$. On considère une permutation σ de $\{1,\dots,n\}$ définie récursivement par :
  - $|J| = n > n-1 = |\{\ell_1,\dots,\ell_{n-1}\}|$, donc il existe $\sigma(n) \in \{1,\dots,n\}$ tel que $j_{\sigma(n)} \notin \{\ell_1,\dots,\ell_{n-1}\}$,
  - pour tout $i \in \{1,\dots,n-1\}$, $|J \setminus \{j_{\sigma(n)},\dots,j_{\sigma(i+1)}\}| = i > i-1 = |\{\ell_1,\dots,\ell_{i-1}\}|$, donc il existe $\sigma(i) \in \{1,\dots,n\} \setminus \{\sigma(n),\dots,\sigma(i+1)\}$ tel que $j_{\sigma(i)} \notin \{\ell_1,\dots,\ell_{i-1}\}$.

Alors, par définition de la famille $(\ell_i)_{1\leq i\leq n}$,

$$\forall i \in \{1,\dots,n\}, \quad p_{\ell_i} \min_{j\in\{1,\dots,N\}\setminus\{\ell_i\}} c_{\ell_i j} \leq p_{j_{\sigma(i)}} \min_{j\in\{1,\dots,N\}\setminus\{j_i\}} c_{j_i j}.$$

Donc

$$\begin{aligned}\sum_{i=1}^{n} p_{\ell_i} \min_{j\in\{1,\dots,N\}\setminus\{\ell_i\}} c_{\ell_i j} \leq \sum_{i=1}^{n} p_{j_{\sigma(i)}} \min_{j\in\{1,\dots,N\}\setminus\{j_i\}} c_{j_i j} &= \sum_{i\in J} p_i \min_{j\in\{1,\dots,N\}\setminus\{i\}} c_{ij} \\ &\leq \sum_{i\in J} p_i \min_{j\in\overline{J}} c_{ij}.\end{aligned}$$

D'où

$$\forall J \in \mathcal{J}_n, \quad \sum_{i=1}^{n} p_{\ell_i} \min_{j\in\{1,\dots,N\}\setminus\{\ell_i\}} c_{\ell_i j} \leq \sum_{i\in J} p_i \min_{j\in\overline{J}} c_{ij}.$$

Comme seul le terme de droite dépend de $J \in \mathcal{J}_n$, il vient, en prenant le minimum sur $\mathcal{J}_n$,

$$\sum_{i=1}^{n} p_{\ell_i} \min_{j\in\{1,\dots,N\}\setminus\{\ell_i\}} c_{\ell_i j} \leq \min_{J\in\mathcal{J}_n} \left\{ \sum_{i\in J} p_i \min_{j\in\overline{J}} c_{ij} \right\}. \tag{A.22}$$

— On suppose que la condition (A.21) est vérifiée pour tout $i \in \{1,\dots,n\}$. Alors

$$\forall i \in \{1,\dots,n\}, \quad \min_{j\in\{1,\dots,N\}\setminus\{\ell_i\}} c_{\ell_i j} = \min_{j\in\{1,\dots,N\}\setminus\{\ell_1,\dots,\ell_n\}} c_{\ell_i j},$$

donc, en posant $J_* = \{\ell_1,\dots,\ell_n\} \in \mathcal{J}_n$, l'équation (A.22) s'écrit

$$\sum_{i\in J_*} p_i \min_{j\in\overline{J_*}} c_{ij} = \sum_{i=1}^{n} p_{\ell_i} \min_{j\in\{1,\dots,N\}\setminus\{\ell_i\}} c_{\ell_i j} \leq \min_{J\in\mathcal{J}_n} \left\{ \sum_{i\in J} p_i \min_{j\in\overline{J}} c_{ij} \right\} \leq \sum_{i\in J_*} p_i \min_{j\in\overline{J_*}} c_{ij}.$$

D'où J_* est un point minimum du problème (A.17), i.e.

$$\sum_{i\in J_*} p_i \min_{j\in\overline{J_*}} c_{ij} = \min_{J\in\mathcal{J}_n} \left\{ \sum_{i\in J} p_i \min_{j\in\overline{J}} c_{ij} \right\}.$$

□

A.3.6 Algorithmes numériques

Le théorème A.16 permet de définir les algorithmes itératifs *backward reduction* (en supprimant les scénarios un à un) et *forward selection* (en sélectionnant les scénarios un à un). Ces deux algorithmes sont présentés plus en détails dans [Dupačová *et al.*, 2000].

Pour améliorer les temps de calcul, ces algorithmes ont été améliorés par les algorithmes *simultaneaous backward reduction* et *fast forward selection* présentés dans [Gröwe-Kuska *et al.*, 2003].

Un algorithme *random search* a également été introduit dans [Armstrong *et al.*, 2013] dont le principe est d'appliquer la redistribution optimale de la sous-section A.3.3 sur un grand nombre d'ensembles de $\mathcal{J}_n$ tirés aléatoirement. Le problème d'optimisation (A.17) est alors résolu sur les ensembles tirés aléatoirement au lieu de $\mathcal{J}_n$.

Ces algorithmes sont itératifs et ils nécessitent un critère d'arrêt. En général, ce critère d'arrêt porte sur au moins un des points suivants :

- la proportion de scénarios conservés par rapport au scénarios supprimés,
- la précision définie par la distance de Kantorovich entre la distribution initiale et la distribution réduite des scénarios,
- les écarts entre les premiers moments (par exemple, espérance et variance) de la distribution initiale et ceux de la distribution réduite des scénarios.

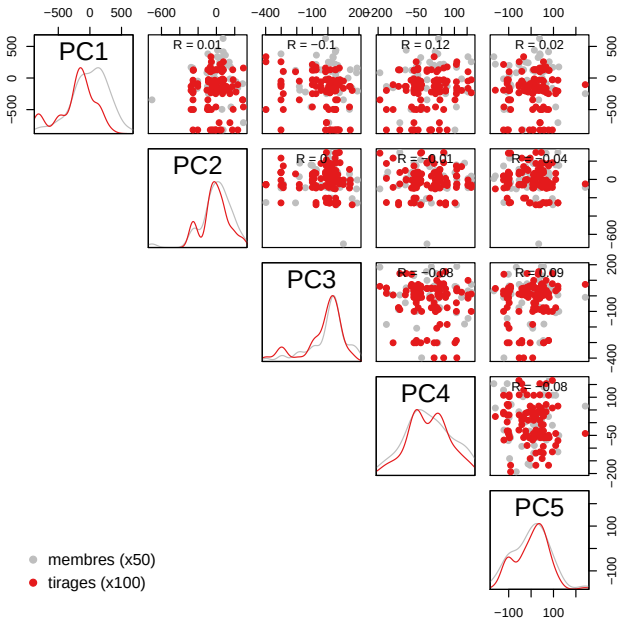


FIGURE A.4 – Nuages de points des cinq premières composantes principales initiales (en gris) et celles obtenues par tirage avec rejet (en rouge) pour la prévision d'ensemble brute des apports en eau du Haut-Rhône émise le 26/04/2014 à 0 h UTC.

A.4 Génération de nouveaux scénarios

Dans cette section, nous présentons une approche simple pour générer de nouveaux scénarios. Cela pourrait servir, par exemple, à augmenter la taille du jeu de données de la section 5.3 pour l'apprentissage du métamodèle dans la phase de pré-traitement. Nous présentons l'approche uniquement pour les apports en eau mais la même approche pourrait être appliquée pour les prix et les productions variables.

L'approche consiste à tirer aléatoirement les coefficients de la projection sur les directions principales selon les lois empiriques des composantes principales supposées indépendantes. Le nombre de composantes principales retenues est défini par rapport à un seuil minimal sur la variance expliquée égale à 99% (le seuil est plus important que celui utilisé pour l'analyse en composante principale de la sous-section 4.2.2). La difficulté est que l'on peut générer de cette façon des courbes d'apports en eau avec des débits négatifs. Pour pallier ce problème, nous choisissons d'effectuer un tirage avec rejet dès que l'on rencontre un débit négatif.

Cette approche est testée sur la prévision d'ensemble brute émise le 26/04/2014 à 0 h UTC (voir Sous-section 2.4.4) pour générer 100 nouvelles courbes d'apports en eau. Le nombre d'itérations nécessaires pour rejeter les courbes ayant au moins un débit négatif est assez faible (10 lots de retirages). Les nuages de points associés aux cinq premières composantes principales (sur les 21 pour expliquer au moins 99% de la variance expliquée) sont présentés à la figure A.4. On remarque que les lois des coefficients de la projection sur les directions principales des points tirés (en rouge) sont légèrement déformées par rapport aux lois des composantes principales (en gris). Une partie du support supérieur de la première composante principale est par exemple sous-représentée par les nouvelles courbes. La figure A.5 présente la reconstitution des courbes tirées des apports en eau. Même si visuellement les allures des courbes semblent acceptables, des études supplémentaires permettraient de quantifier les erreurs d'approximation sur les corrélations spatio-temporelles, une caractéristique importante des apports en eau.

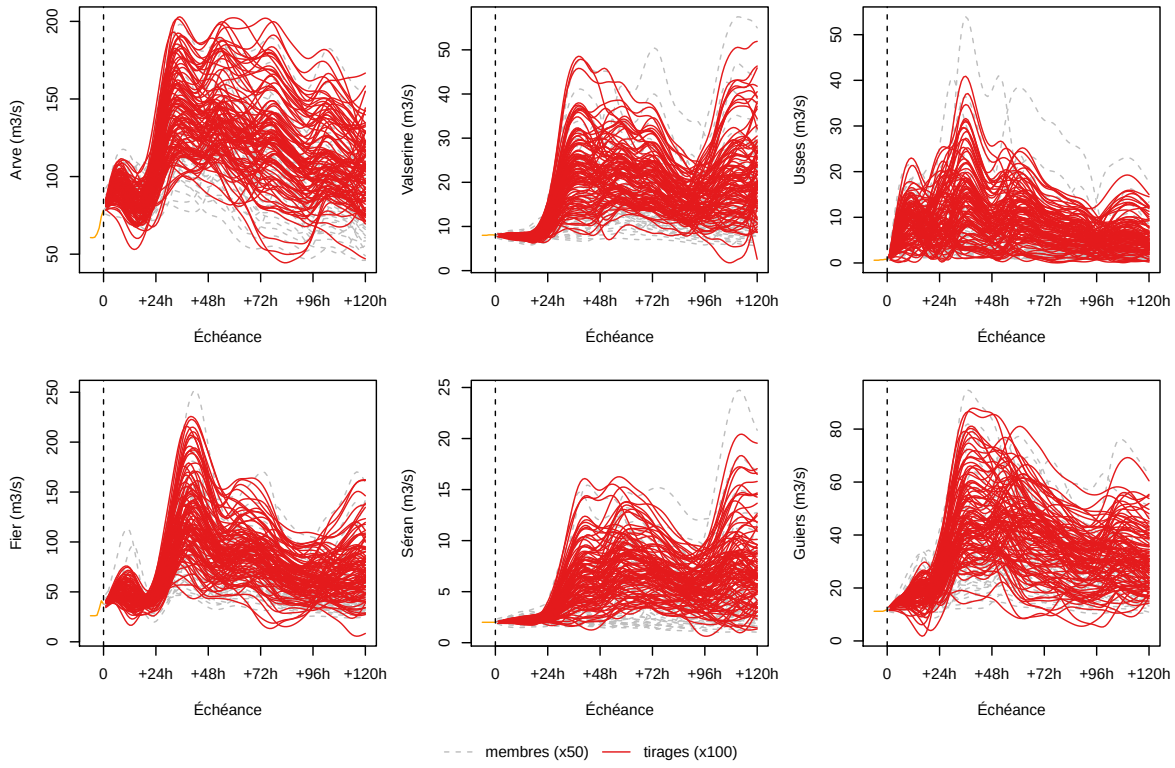


FIGURE A.5 – Membres de la prévision d'ensemble brute des apports en eau du Haut-Rhône émise le 26/04/2014 à 0 h UTC (en gris) et 100 nouvelles courbes obtenues par tirage avec rejet des composantes principales (en rouge).

dernière page de la thèse

AUTORISATION DE SOUTENANCE

Vu les dispositions de l'arrêté du 25 mai 2016,

Vu la demande du directeur de thèse

Monsieur G. VIAL

et les rapports de

Mme M. MOUGEOT
Professeur - Centre Borelli
Ecole Nationale Supérieure d'Informatique pour l'Industrie et l'Entreprise
1 rue de la résistance - 91000 Evry-Courcouronnes

et de

Mme M. CLAUSEL
Professeure - Université de Lorraine - IECL - Site de Nancy
Faculté des sciences et Technologies - Boulevard des Aiguillettes
54506 Vandoeuvre-lès-Nancy

Monsieur PIGUET Antoine

est autorisé à soutenir une thèse pour l'obtention du grade de **DOCTEUR**

Ecole doctorale Informatique et Mathématiques

Fait à Ecully, le 25 avril 2022

Le directeur de l'Ecole centrale de Lyon

Pascal RAY

Optimisation sous incertitudes par métamodèle pour la planification à court terme de la production conjointe d'une chaîne hydroélectrique et d'actifs de production variable.

La planification à court terme de la production conjointe d'une chaîne hydroélectrique et d'actifs de production variable est un problème d'optimisation. Le programme de production hydroélectrique est construit à un horizon de quelques jours pour respecter les contraintes d'exploitation et pour maximiser le chiffre d'affaires obtenu par la vente sur le marché *day-ahead* et par la pénalisation des écarts, en considérant la production totale de tous les actifs.

Le travail de la thèse consiste à prendre en compte les incertitudes qui entachent les prévisions des apports en eau, des prix de l'électricité et des productions variables dans le problème d'optimisation. À partir d'une modélisation des incertitudes par un ensemble fini de scénarios multivariés, le problème d'optimisation en univers probabiliste proposé consiste en un problème de programmation dynamique stochastique à deux niveaux linéaires mixtes.

La problématique des temps de calcul étant cruciale pour un usage opérationnel, le problème d'optimisation probabiliste est résolu numériquement en remplaçant la valeur optimale du second niveau par un métamodèle calibré par apprentissage supervisé durant une phase de pré-traitement. Pour calibrer ce métamodèle, les données d'entrée sont simplifiées par des approches spécifiques. Un échantillonnage par analogie permet d'abord d'approcher le domaine de la donnée d'entrée associée aux variables de décision du premier niveau. Les données d'entrée fonctionnelles sont ensuite réduites par analyse en composantes principales. Un plan d'expérience peut ainsi être construit par échantillonnage par hypercube latin pour former le jeu de données nécessaire à l'apprentissage du métamodèle. Plusieurs métamodèles linéaires sont ensuite proposés.

La méthodologie proposée est testée sur un cas d'étude réel simplifié. Le métamodèle obtenu par régression linéaire sur les données d'entrée réduites donne des performances acceptables et il permet d'obtenir une solution du problème d'optimisation en univers probabiliste. Néanmoins, en l'absence de productions variables dans le cas d'étude, le problème d'optimisation en univers probabiliste ne permet pas d'apporter des gains significatifs par rapport à celui en univers déterministe. En outre, les temps de calcul ne permettent pas une utilisation en opérationnel sans calcul distribué. Plusieurs pistes de recherche sont toutefois proposées pour améliorer la méthodologie. Une validation sur plusieurs cas d'étude, voire sur un horizon roulant, permettrait d'estimer les performances de la méthodologie de manière plus robuste.

Mots-clés : programmation linéaire mixte, programmation stochastique à deux niveaux, programmation dynamique stochastique, prévision d'ensemble, prévision par analogie, réduction de données fonctionnelles, réduction de scénarios, plan d'expérience, régression linéaire.

Optimization under uncertainty with a surrogate model for the short-term production planning of a hydropower cascade combined with variable energy sources.

Short-term production planning of a hydropower cascade combined with variable energy sources is an optimization problem. The hydropower production plan is constructed within a time horizon of a few days to comply with operational requirements and to maximize revenues obtained by selling the total production from all energy sources on the day-ahead market and by penalizing imbalances.

This PhD thesis focuses on the optimization problem considering data uncertainty on water inflows, electricity prices and variable generation, modeled by a finite set of multivariate scenarios. The proposed optimization problem in the stochastic framework is then written with a two-stage stochastic dynamic mixed-integer linear programming formulation.

Since computation time is a crucial issue for an operational use, the optimization problem is solved by replacing the value function of the second stage with a surrogate model fitted by supervised learning during a pre-processing step. The surrogate model is fitted on input data that are simplified by specific approaches. The domain of the input data related to the decision variables of the first stage is first estimated by analogue-based sampling, and the functional inputs are reduced by principal components analysis. The learning data set can then be obtained from a design of experiments using latin hypercube sampling. Several linear models are finally proposed for the surrogate model.

The proposed methodology is tested on a real but simplified case study. The surrogate model obtained by linear regression on the reduced inputs provides acceptable performance, and it can be used to get a solution to the optimization problem in the stochastic framework. Since variable generation is not considered in the case study, the stochastic framework provides, however, minor profit compared to the deterministic framework. Besides, computation time is not compatible with an operational use without distributed computing. Several research tracks are then proposed to improve the methodology. A validation procedure on several case studies or on a rolling horizon could be used to assess the performance of the methodology in a more robust way.

Keywords: mixed-integer linear programming, two-stage stochastic programming, stochastic dynamic programming, ensemble forecasting, analog forecasting, functional reduction, scenario reduction, design of experiments, linear regression.