



Université des Antilles



Ecole Doctorale de L'université des Antilles pôle Guadeloupe

Numéro d'ordre : n° 2015AGUY0872

THESE

Présentée en vue d'obtenir le titre de

Docteur de Physique

Mention : Energétique

Par

ANDRE Maïna

**PROPOSITION D'UN MODELE DE PREVISION
SPATIO –TEMPOREL A COURT TERME DE L'ENSOLEILLEMENT
GLOBAL A PARTIR DE TROIS SITES EN GUADELOUPE**

Dirigé par : Mr ZAHIBO Narcisse, Professeur, Université des Antilles
Mr SOUBDHAN Ted, Maître de Conférence, Université des Antilles

Rapporteurs :
Mr Philippe LAURET, Professeur, Université de La Réunion
Mr Philippe POGGI, Professeur, Université de Corse

Jury :
Mr Philippe LAURET, Professeur, Université de La Réunion
Mr Philippe POGGI, Professeur, Université de Corse
Mr Richard PEREZ, Professeur, Université d'Albanie Etats-unis
Mr Ted SOUBDHAN, Maître de Conférences, Université des Antilles
Mr Narcisse ZAHIBO, Professeur, Université des Antilles

Remerciements

Je tiens à remercier avant tout et pour tout L'Eternel.

Je remercie Mr Zahibo Narcisse de m'avoir accueilli au sein du Laboratoire de Recherches en Géosciences et Energies. Je tiens également à le remercier pour sa disponibilité, ses conseils et la sympathie qu'il a témoigné envers moi.

Merci à Mr Soubdhan Ted d'avoir accepté d'encadrer cette thèse jusqu'à son terme. Merci pour les conférences et les missions auxquelles j'ai pu participer. Merci pour cette richesse scientifique qu'il m'a apporté par ses conseils, ses suggestions et nos différents échanges scientifiques. Merci à lui pour cette confiance qu'il a témoigné envers moi.

Merci à Sophie Dabo-Niang, pour sa disponibilité, les échanges et l'apport scientifiques qui ont enrichi mes travaux de recherches. Merci pour tous ces bons moments passés ensemble.

Je remercie Mr Philippe Lauret, Mr Philippe Poggi, d'être mes rapporteurs et Mr Richard Pérez d'avoir accepté d'être membre du Jury.

Merci à tous les membres du laboratoire de m'avoir bien accueilli, tout particulièrement ceux du Groupe Energie, Mr Rudy Calif, Mr Ruddy Blonbou, Mr Jean-Louis Bernard, Mr Dupont Frédéric.

Je remercie également La Région Guadeloupe et L'ADEME d'avoir financé mes travaux de recherches.

Je remercie tous les jeunes docteurs et post-doc que je considère comme mes amis :

A toi Stéphanie, qui a été ma confidente avec qui j'ai pu échanger au bureau dans tous les domaines autant professionnels que personnels. Avec qui j'ai pu partager ma passion pour la cuisine.

A toi Hanany, pour toutes ses bonnes discussions scientifiques et ces connaissances supplémentaires que tu m'as apporté. Merci pour tous ces bons moments passés ensemble et cette foi et cette joie de vivre partagée. Pour ton soutien.

A toi Joseph, pour tes encouragements, ces rigolades et nos échanges scientifiques. Merci.

A Thomas, Raphaël, Jean-françois pour ces bons moments passés ensemble.

A Bernard Dudon, Sébastien Régis, Sandrine, Vanessa, Frédéric, Nadiège, Célia, Jimmy, Alexandra, Audrey, tous les membres de l'AJC,...

Je tiens à remercier en dernier mais non les moindres, Lisa, Valéry, Lynda, Barbara ma meilleure amie (merci pour tout Barbara), ma sœur Marielle, mon frère Samuel, Solise, Harry, Elisa, Jessica, Myriam, toute ma famille de Martinique et Guadeloupe particulièrement tonton Sébastien et tati Nadine, papi, mami. A toi mon Papa chéri, Marie-michèle. Rien ne remplacera la place que vous avez tous dans mon cœur.

Et enfin mon mari Yannick qui m'a soutenu et qui a su trouver les mots dans les moments de détresses, qui a su m'encourager. A mes enfants, Noah et Kenan, mes amours, je ne trouve pas de mots assez forts pour vous dire MERCI et combien vous m'avez apporté dans la vie. Je vous aime.

Résumé

En Guadeloupe, actuellement, 5,92% de la demande en énergie électrique sont couverts par la filière photovoltaïque et 3,14% par la filière éolienne soit 9,06% pour leur production cumulée selon le bilan 2015 de l'OREC (Observatoire Régional de l'Énergie et du Climat). Selon le plan énergétique régional de prospection, la production cumulée du photovoltaïque et de l'éolien devrait représenter 14% du mix électrique en 2020 et 18% en 2030. Pour atteindre les 14% du mix électrique d'ici les cinq prochaines années, il va donc falloir entre autres, améliorer la prédictibilité pour un développement à un rythme soutenu de ces énergies. Ces travaux de recherches ont consisté à apporter de nouveaux résultats de performance de prévision de l'ensoleillement global à court terme et à donner une connaissance plus fine de la ressource sur trois stations en Guadeloupe. L'étude est basée sur une analyse et un modèle de prévision de l'ensoleillement, faisant intervenir des paramètres spatiaux et temporels. La littérature montre qu'un important nombre de sites est en général utilisé pour une analyse spatio-temporelle, ce qui impliquerait pour nous, de poser de multiples capteurs sur l'ensemble du territoire. Les coûts d'un tel système seraient considérables. Notre approche ici consistera à effectuer une analyse spatio-temporelle sur trois stations. Avec peu de stations et des distances non uniformes nous avons donc cherché à développer un modèle de prévision de l'ensoleillement à court terme en dépit de ces contraintes qui ne répondent pas à une approche classique. Le modèle est basé sur une méthodologie VAR (Vecteur Autorégressif) incluant des paramètres spatiaux et temporels. Une stratégie de sélection des variables est développée afin de sélectionner les prédicteurs (stations) utiles pour la prévision sur une localisation. Cette stratégie itérative permettra d'une part d'être plus proche de la réalité, d'autre part d'un point de vue algorithmique, la tendance des calculs sera plus rapide. En amont du développement du modèle, une étude de la variabilité spatio-temporelle de l'ensoleillement a permis de quantifier et caractériser de manière fine, les interactions dynamiques entre ces trois stations. Par comparaison avec les modèles de la littérature, notre modèle de prévision montre une bonne performance avec des valeurs de RMSE relative allant de 17,48% à 23,79% pour des horizons de prévisions de 5 min à 1h. Les méthodologies développées pourraient à terme offrir une opportunité d'assurer des garanties au gestionnaire du réseau. Si d'avenir des solutions de prévision performantes se généralisaient, cette opportunité permettrait d'ouvrir le marché au-delà du seuil de 30% imposé actuellement.

Mots clés : classification, corrélation spatiale, corrélation temporelle, spatio-temporel VAR (Vecteur AutoRégressif), prévision à court terme

Abstract

Currently in Guadeloupe, there is 5,92 % of the electric power request covered by the photovoltaic sector and 3,14 % by the wind sector which represents 9,06 % for their accumulated production, according to the OREC report (Regional Monitoring center of Energy and Climate). According to the regional energy plan, the accumulated production of the photovoltaic and the wind energy should represent 14 % of the electric mix in 2020 and 18 % in 2030. To reach the 14 % of the electric mix within the next five years, we need, among other things, to improve forecast for a sustained development of these energies. These research works consisted in bringing new performance results of short-term forecast of the global solar radiation and in giving a finer knowledge of the resource onto three stations in Guadeloupe. The study is based on an analysis and a forecast model of global solar radiation, by including spatial and temporal parameters. The literature shows that an important number of sites is generally used for a spatio-temporal analysis, which would imply for us, to put multiple sensors on the whole territory. The costs of such a system would be considerable. Our approach here will consist in making a spatiotemporal analysis on three stations. With few stations and not uniform distances, we, thus, tried to define a short-term forecast model of global solar radiation, in spite of these constraints which do not answer to a classic approach. The model is based on a methodology the VAR (Autoregressive Vector) including spatial and temporal parameters. A strategy of selection of variables is developed to select useful predictors (stations) for the forecast on localization. This iterative strategy, on one hand will allow being closer to the reality, on the other hand to the point of algorithmic view, the trend of the calculations will be faster. Preliminarily, a study of the spatiotemporal variability of global solar radiation, allowed to quantify and to characterize in a fine way, the dynamic interactions between these three stations. Compared with the models of the literature, our forecast model shows a good performance with relative RMSE values going from 17.48 % to 23.79 % for horizons from 5 min to 1 hour. The developed methodologies could eventually offer an opportunity to assure guarantees to the network manager. If in the future the successful solutions of forecast became widespread, this opportunity would allow the opening of the market beyond the 30 % threshold imposed at present.

Key words : classification, spatial correlation, temporal correlation, spatio-temporal VAR (Autoregressive Vector), short term forecasting

Table des matières

Table des figures	13
Liste des tableaux	21
Introduction générale	23
I. L'enjeu de la prévision à court terme de l'ensoleillement	24
II. Problématique de la thèse	25
III. Organisation du manuscrit	26
Chapitre 1. Contexte de l'étude	28
I. Situation géographique et climatique de la Guadeloupe	29
II. Les consommations d'électricité en Guadeloupe	30
III. Un mix énergétique renouvelable diversifié en Guadeloupe	31
III.1. Historique de l'énergie solaire photovoltaïque en Guadeloupe	31
III.2. La production d'électricité à partir des énergies renouvelables en Guadeloupe.....	34
IV. Objectifs politiques concernant la filière photovoltaïque	35
V. Perspectives d'amélioration de l'intégration de la filière photovoltaïque	37
V.1. Le stockage.....	38
V.2. Smart grid.....	38
V.3. L'intérêt de la prévision.....	39
VI. Conclusion	41
Références	43
Chapitre 2. Etat de l'art pour la prédiction du rayonnement solaire	44
I. Horizon de prédiction	46
I.1. Horizon de prévision à très court terme	46
I.2. Horizon de prévision à court terme	47
I.3. Horizon de prévision à moyen terme	47
II. Prévisions météorologiques numériques (NWP)	48
II.1. GFS.....	49
II.2. NAM : Modèle Méso-échelle Nord-Américain (North American Mesoscale Model)	49
II.3. Centre européen de prévisions à moyen terme (ECMWF)	50
II.4. MOS	50
II.5. Aperçu des travaux NWP pour la prévision solaire	51
II.6. Limites des modèles NWP.....	52
III. Imagerie des nuages	53
III.1. Imagerie au sol des nuages	53
III.2. Imagerie par satellite des nuages	54
III.3. Aperçu des travaux d'imagerie des nuages	54
III.4. Limites des modèles d'imagerie des nuages.....	56
IV. Modèles statistiques	57
IV.1. Modèle de persistance	57
IV.2. Modèles linéaires	58

IV.3. Modèles non-linéaires	65
V. Modèles hybrides	70
V.1. Aperçu des travaux de modèles hybrides	70
V.2. Limites des modèles hybrides.....	73
VI. Modèles spatio-temporels et aperçu des travaux.....	74
VI.1. Modèle STARMA.....	74
VI.2. Modèle krigeage.....	77
VI.3. Modèle VARMA.....	80
VII. Synthèse et conclusion de l'état de l'art	81
Références.....	86

Chapitre 3. Classification de la variabilité de l'ensoleillement global..... 93

I. Les données de mesures	94
I.1. Présentation des données de mesures.....	94
I.2. Dépouillement des données.....	95
II. Classification des journées de flux solaire	96
II.1. Aperçu des travaux sur les classifications du flux solaire.....	96
II.2. La classification du flux solaire par méthode k-means	98
III. Analyse à composante principale fonctionnelle	99
III.1. Principe.....	99
III.2. Résultats	101
IV. Les classes de flux solaire.....	103
IV.1. Classe 1	103
IV.2. Classe 2	106
IV.3. Classe 3	108
IV.4. Classe 4	110
V. Analyse statistique des classes obtenues.....	112
V.1. Séquence de classes.....	112
V.2. Pourcentage des types de journées	113
V.3. Persistance des classes.....	114
V.4. Discussion et comparaison inter-site des classes	117
VI. Conclusion.....	119
Références.....	121

Chapitre 4. Corrélation spatio-temporelle de la variabilité de l'ensoleillement global à court-terme 124

I. Analyse sur trois stations de mesures	125
I.1. Localisation géographique	125
I.2. Contexte climatique des stations	127
II. Analyse statistique descriptive de la variabilité de flux solaire à court terme	127
II.1. Caractérisation de l'énergie reçue.....	128
II.2. Caractéristiques de la variabilité de l'ensoleillement global à court terme.....	130
III. Variabilité spatio-temporelle du flux solaire	138
III.1. Aperçu des travaux dans la littérature	138
III.2. Evaluation des corrélations : Outils statistiques utilisés	139
III.3. Analyse des corrélations inter sites spatiales.....	143
III.4. Analyse des corrélations inter sites temporelles	151
IV. Analyse dynamique des corrélations	162
IV.1. Décalage temporel des corrélations.....	162
IV.2. Corrélations intersites et paramètres physiques	164
IV.3. Description dynamique des corrélations	167
V. Conclusion	174
Références.....	176

Chapitre 5. Modèle spatio-temporel de prévision à court terme de l'ensoleillement global

.....	179
I. Modèle Vecteur AutoRégressif VAR spatio-temporel pour des données dispersées dans l'espace ..	180
I.1. Principe général du modèle VAR	180
I.2. Matrice des distances	182
I.3. Stationnarisation spatio-temporelle des données	182
I.4. Stratégie de sélection des prédicteurs.....	184
I.5. Estimation des paramètres.....	185
II. Le modèle spatio-temporel VAR pour la prévision de l'ensoleillement global.....	187
II.1. Un modèle multivarié pour notre contexte spatio-temporel	187
II.2. Matrice des données	188
II.3. Sélection des prédicteurs	199
II.4. Estimation des paramètres du modèle.....	200
III. Analyse de la performance prédictive	201
III.1. Test sur la sélection de l'ordre p.....	201
III.2. Analyse de la performance prédictive pour deux structures spatiales	202
III.3. Analyse de la performance prédictive en fonction de la variabilité du flux solaire.....	210
III.4. Analyse de la performance prédictive pour les saisons	216
III.5. Analyse de la performance prédictive pour différents horizons	217
IV. Amélioration des paramètres du modèle spatio-temporel VAR.....	225
IV.1. Tests des paramètres du modèle et amélioration de la performance de prévision	225
IV.1. Comparaison avec d'autres modèles	230
IV.2. Limites du modèle VAR et perspectives.....	235
V. Conclusion	238
V.1. Références	239
Conclusion générale	241
I. Bilan des résultats	242
II. Perspectives	244
Annexe	246
Analyse en composante principale fonctionnelle	246
I. Transformations des données brutes en données fonctionnelles	246
II. Extraction des harmoniques	247
Résultats du modèle pour plusieurs horizons de prévisions	249
Horizon à 5 minutes	249
Horizon à 15 minutes	254
Horizon à 30 minutes	258
Horizon à 1 heure	263
Liste des publications	268

Table des figures

Figure 1-1. Situation géographique de la Guadeloupe.....	29
Figure 1-2. Evolution de la part de l'électricité produite à partir d'énergie renouvelable dans le mix énergétique (source : Observatoire Régional de l'Energie et du Climat (OREC) en Guadeloupe Publication 2015).....	32
Figure 1-3. Source EDF : Production d'électricité à partir d'énergie solaire photovoltaïque de 2005 à 2012	33
Figure 1-4. Répartition des panneaux solaires photovoltaïques de l'archipel Guadeloupéen (carte réalisée à partir des données d'EDF.....	33
Figure 1-5. Schéma du système électrique prévisionnel. Source EDF SEI : Bilan prévisionnel de l'équilibre offre demande d'électricité juillet 2013	34
Figure 1-6. Evolution de la puissance du photovoltaïque par période d'échantillonnage de 2 minutes pendant 9 jours à partir du 2 septembre 2013. Mesures effectuées sur une installation PV de l'Université.	40
Figure 2-1 Classification des modèles basée sur une résolution spatiale et temporelle selon Diagne et al (2013) à laquelle nous avons ajouté les modèles spatio-temporels.	84
Figure 3-1 Courbes de flux solaire moyennes, représentatives des 4 classes pour le site de La Désirade, Petit-canal et Fouillole.....	99
Figure 3-2. Plan des scores des deux premiers scores et classification des groupements de points (La Desirade)	102
Figure 3-3. Plan des scores des deux premiers scores et classification des groupements de points (Petit-canal).....	102
Figure 3-4. Plan des scores des deux premiers scores et classification des groupements de points (Fouillole).....	103
Figure 3-5. Distributions du flux solaire pour la classe 1 pour toutes les stations des données moyennées par pas de 30 minutes.	104
Figure 3-6. Exemples de journées classe 1 pour les trois sites de mesures (La Désirade, Petit-canal, Fouillole)	105
Figure 3-7. Distributions du flux solaire pour la classe 2 pour toutes les stations des données moyennées par pas de 30 minutes.	106
Figure 3-8. Exemples de journées de classe 2 pour les trois sites de mesures (La Désirade, Petit-canal, Fouillole).....	107

Figure 3-9. Distributions du flux solaire pour la classe 3 pour toutes les stations des données moyennées par pas de 30 minutes.	108
Figure 3-10. Exemples de journées de classe 3 pour les trois sites de mesures (La Désirade, Petit-canal, Fouillole).....	109
Figure 3-11. Distributions du flux solaire pour la classe 4 pour toutes les stations, des données moyennées par pas de 30 minutes.	110
Figure 3-12 Exemple de journées de classe 4 pour les trois sites de mesures(La Désirade, Petit-canal, Fouillole).....	111
Figure 3-13. Séquence de classes {1, 2, 3, 4} sur 1096 jours (année 2010 à 2012) sur le site de La Désirade, Petit-canal et Fouillole.....	112
Figure 3-14. Nombre de jours d'apparition pour les classes {1, 2, 3, 4} au cours des trois années de mesures sur le site de La Désirade, Petit-canal et Fouillole.	113
Figure 3-15. Fréquence du nombre de jours d'apparition successifs pour les classes {1, 2, 3, 4} au cours des trois années de mesures sur les stations : La Désirade a), Petit-canal b), Fouillole c).....	115
Figure 3-16. Exemple de journée (31 janvier 2010) ayant la même classe 1 simultanément sur les trois sites de mesure : La Désirade, Petit-canal et Fouillole.....	119
Figure 4-1 Carte de l'Archipel de Guadeloupe et localisation géographique de nos trois stations de mesures: Petit-canal, Fouillole et La Désirade.....	126
Figure 4-2. Fraction d'énergie de l'ensoleillement sur l'ensemble de nos données de mesures (2010 à 2012) pour le site de La Désirade.....	130
Figure 4-3. Courbe de la moyenne, du mode, de la médiane journalière sur trois années pour le site de La Désirade	131
Figure 4-4. Courbe de la moyenne, du mode, de la médiane journalière sur trois années pour le site de Petit-canal.....	131
Figure 4-5. Courbe de la moyenne, du mode, de la médiane journalière sur trois années sur le site de Fouillole.....	132
Figure 4-6. Exemple d'une journée de flux solaire mesuré a) et de sa moyenne glissante correspondante b) (journée du 1er octobre 2010). c) est le signal de fluctuation obtenu.	133
Figure 4-7 . Fonction de répartition de l'amplitude des fluctuations pour le site de Fouillole	135
Figure 4-8. a) Fonction de répartition de la durée des fluctuations et b)densité de probabilité de l'échelle de temps Δt (durée des fluctuations en secondes) pour une gamme d'amplitude des fluctuations $\Delta G'$ pour le site de La Désirade. c) Zoom de la figure b).....	136

Figure 4-9. a) Fonction de répartition de la durée des fluctuations et b)densité de probabilité de l'échelle de temps Δt (durée des fluctuations en secondes) pour une gamme d'amplitude des fluctuations $\Delta G'$ pour le site de Petit-canal. c) Zoom de la figure b). 137

Figure 4-10. a) Fonction de répartition de la durée des fluctuations et b)densité de probabilité de l'échelle de temps Δt (durée des fluctuations en secondes) pour une gamme d'amplitude des fluctuations $\Delta G'$ pour le site de Fouillole. c) Zoom de la figure b). 137

Figure 4-11. Echantillon d'une journée à haute variabilité (1er janvier 2012) du signal Δkt en rouge (différentiation de l'indice de clarté) et du signal Δkt^* en bleu (différentiation de l'indice ciel clair) pour le site de La Désirade. 145

Figure 4-12. Echantillon d'une journée à haute variabilité (1er janvier 2012) montrant l'ensoleillement global mesuré (G) par pas de 20s, l'ensoleillement extraterrestre G_0 , G_c l'ensoleillement théoriquement mesuré par le capteur et l'écart-type sur 15 minutes. b) Signal Δkt^* correspondant..... 145

Figure 4-13 Corrélation de paires de stations en fonction des distances inter sites et des échelles de temps pour un échantillon d'une journée à haute variabilité. 146

Figure 4-14 Corrélation de paires de stations en fonction des distances inter sites et des échelles de temps pour un échantillon d'une année (366 jours). 148

Figure 4-15 Tendance moyenne des corrélations de paires de stations en fonction des distances inter sites et des échelles de temps pour un échantillon d'une année (366 jours). Les lignes en pointillé représentent l'intervalle de confiance à 95% et les points verts les C_{paire} calculés par l'équation (3.11). 148

Figure 4-16 Zoom de la Figure 4-15 149

Figure 4-17. Intercorrélation de paires de stations en fonction des distances inter sites et des échelles de temps pour un échantillon d'une année (366 jours). 150

Figure 4-18. Tendance moyenne des intercorrélations de paires de stations en fonction des distances inter sites et des échelles de temps pour un échantillon d'une année (366 jours). Les lignes en pointillé représentent l'intervalle de confiance à 95% 150

Figure 4-19 Fonction de cohérence entre les signaux de fluctuations pour chaque paire de sites comparés: Petit-canal/Fouillole (a), Petit-canal/La Desirade (b), La Desirade/ Fouillole (c). 152

Figure 4-20 Signaux de fluctuations du 10 octobre 2012 pour les intervalles du haut vers le bas [1 min ;15 min], [15 min ; 30 min], [30 min ;1 h], [2 h ;3 h] 154

Figure 4-21 Densité de probabilité des valeurs absolues des coefficients maximaux d'intercorrélacion des signaux de fluctuations journaliers dans l'intervalle [1 min ;15

min],[15min ;30min], [30min ;1h] calculés de 2010 à 2012 pour l'ensemble des paires de sites comparés.	155
Figure 4-22 Densité de probabilité des coefficients maximaux des signaux de fluctuations journaliers dans l'intervalle [1min ;15min],[15min ;30min],[30min ;1h],[1h ;2h],[2h ;3h], calculés de 2010 à 2012 pour la paire de site Petit-canal/Fouillole.	157
Figure 4-23 Densité de probabilité des coefficients maximaux des signaux de fluctuations journaliers dans l'intervalle [1min ;15min],[15min ;30min],[30min ;1h],[1h ;2h],[2h ;3h], calculés de 2010 à 2012 pour la paire de site Petit-canal/La Désirade.....	158
Figure 4-24 Densité de probabilité des coefficients maximaux des signaux de fluctuations journaliers dans l'intervalle [1min ;15min],[15min ;30min],[30min ;1h],[1h ;2h],[2h ;3h], calculés de 2010 à 2012 pour la paire de site Petit-canal/La Désirade.....	159
Figure 4-25. Densité de probabilité des coefficients de covariance des signaux de fluctuations journaliers dans l'intervalle [1h ;2h] calculés de 2010 à 2012 pour les paires de site comparés : Petit-canal/ Fouillole, Petit-canal/La Désirade, La Désirade/Fouillole.....	161
Figure 4-26. a) Journée de flux solaire mesuré sur deux stations avec b) leurs signaux de fluctuations correspondants obtenues par méthode de moyenne glissante. La fonction d'intercorrélacion calculée pour les signaux de fluctuations est présentée en c). Décalage temporel $\tau = 41$ min et coefficient maximal = 0,75 (1er février 2012 entre Petit-canal et La Désirade).....	162
Figure 4-27. Densité de probabilité des décalages temporels journaliers calculés pour RXY ($\tau \geq 0.6$) pour la comparaison intersite Petit-canal/Fouillole, La Désirade/Petit-canal, La Désirade/Fouillole.....	164
Figure 4-28. Densité de probabilité des décalages temporels journaliers calculés pour RXY ($\tau \geq 0.6$) pour l'ensemble des comparaisons intersites.	164
Figure 4-29. Fréquence d'apparition en nombres de jours des corrélacions RXY ($\tau \geq 0.6$) calculées pour chaque couple de sites par mois pour l'ensemble de l'année 2010.	165
Figure 4-30. Fréquence d'apparition en nombres de jours des corrélacions RXY ($\tau \geq 0.6$) calculées pour chaque couple de sites par mois pour l'ensemble de l'année 2011.	165
Figure 4-31. Fréquence d'apparition en nombres de jours des corrélacions RXY ($\tau \geq 0.6$) calculées pour chaque couple de sites par mois pour l'ensemble de l'année 2012.	166
Figure 4-32. Fréquence d'apparition en nombres de jours des corrélacions RXY ($\tau \geq 0.6$) calculées pour chaque couple de sites par mois pour l'ensemble des trois années. La ligne en pointillée délimite l'évolution de la moyenne journalière du flux solaire et mettant en exergue les quatre périodes saisonnières (voir chapitre 2).	166

Figure 4-33. Coefficients maximaux de fonction d'intercorrélation journalière en fonction des coefficients de covariance journaliers pour 3 années de mesures, pour l'ensemble des paires de sites comparés. 168

Figure 4-34. Figure 4-33 sectionnée en tranche de valeur de coefficients. Cadre jaune : coefficients de covariance compris dans l'intervalle $[-0.6 ; 0.6]$ et des coefficients d'intercorrélation compris entre $[0,2 ; 0,6]$. Cadre rouge : coefficients de covariance compris dans l'intervalle $[-0.6 ; 0.6]$ et coefficients d'intercorrélation supérieurs à 0,6. Cadre gris : coefficients de covariance égaux aux coefficients d'intercorrélation pour des valeurs supérieurs à 0,6. Cadre vert : Coefficients de covariance négatifs mais supérieurs à 0,6 en valeur absolue et coefficients d'intercorrélation supérieurs à 0,6. 169

Figure 4-35. a) Signaux de fluctuations, b) signaux lissés sur les séries de flux solaire mesurés par méthode de moyenne glissante, c) signaux de flux solaire mesurés. Exemple d'une journée (29 Mai 2010) pour laquelle coefficient de covariance et le coefficient d'intercorrélation sont égaux et de valeur supérieure à 0,6.(cadre gris Figure 4-34). 170

Figure 4-36. a) Signaux de fluctuations, b) signaux lissés sur les séries de flux solaire mesurés par méthode de moyenne glissante, c) signaux de flux solaire mesurés. Exemple d'une journée (14 juin 2010) pour laquelle le coefficient de covariance est compris dans l'intervalle $[-0.4 ; 0.4]$ et le coefficient d'intercorrélation dans l'intervalle $[0,4 ; 0,6]$ (cadre jaune Figure 4-34). 171

Figure 4-37. a) Signaux de fluctuations, b) signaux lissés sur les séries de flux solaire mesurés par méthode de moyenne glissante, c) signaux de flux solaire mesurés. Exemple d'une journée (25 juin 2010) pour laquelle le coefficient de covariance est compris dans l'intervalle $[-0.4 ; 0.4]$ et le coefficient d'intercorrélation supérieur à 0,6 (cadre rouge Figure 4-34). 172

Figure 4-38. a) Signaux de fluctuations, b) signaux lissés sur les séries de flux solaire mesurés par méthode de moyenne glissante, c) signaux de flux solaire mesurés. Exemple d'une journée (11 Mai 2010) pour laquelle le coefficient de covariance est compris dans l'intervalle $[-0.6 ; -0.9]$ et le coefficient d'intercorrélation supérieur à 0,6 (cadre vert Figure 4-34). 172

Figure 5-1 Moyenne journalière des résidus des modèles de régression linéaire simple (en vert) et de la régression linéaire multiple (en bleu) résultant de la combinaison linéaire des stations 188

Figure 5-2. a)Exemple d'une série temporelle correspondant à un bruit blanc. simulé b) Fonction d'autocorrélation partielle correspondante, caractéristique d'un processus stationnaire. 190

Figure 5-3. Exemple d'une série temporelle correspondant à une journée d'ensoleillement de 7h à 17h b) Fonction d'autocorrélation partielle correspondante à cette série temporelle. 190

Figure 5-4. Exemple d'une série temporelle correspondant à deux journées d'ensoleillement de 7h à 17h. b) Fonction d'autocorrélation partielle pour deux journées d'ensoleillement global. 191

Figure 5-5. a) Exemple d'une série temporelle correspondant à un mois d'ensoleillement de 7h à 17h. b) Fonction d'autocorrélation partielle pour un mois d'ensoleillement global. 192

Figure 5-6. Exemple du 2 mars 2012 a) Signal mesuré du flux solaire global et son signal de moyenne glissante correspondante en rouge. b) Signal sans tendance obtenu par différence entre le signal mesuré et la moyenne glissante. 195

Figure 5-5. a) Exemple d'une série temporelle correspondant à un mois d'ensoleillement de 7h à 17h. b) Fonction d'autocorrélation partielle pour un mois d'ensoleillement global. 192

Figure 5-6. Exemple du 2 mars 2012 a) Signal mesuré du flux solaire global et son signal de moyenne glissante correspondante en rouge. b) Signal sans tendance obtenu par différence entre le signal mesuré et la moyenne glissante. 195

Figure 5-7. Exemple du 2 mars 2012 a) Signal mesuré du flux solaire global b) Signal sans tendance obtenu par méthode des différenciations d'ordre 1. **Erreur ! Signet non défini.**

Figure 5-8. Fonction d'autocorrélation du signal sans tendance obtenu par méthode des moyennes glissantes à N=10points. **Erreur ! Signet non défini.**

Figure 5-9. Fonction d'autocorrélation du signal sans tendance obtenu par méthode des différenciations d'ordre1 **Erreur ! Signet non défini.**

Figure 5-10. Signal de flux solaire mesuré en fonction du temps a) et son signal correspondant sans tendance spatio-temporelle b) **Erreur ! Signet non défini.**

Figure 5-11. Fréquence des valeurs propres maximales en valeur absolue de la matrice des paramètres R_i calculées sur un mois soient 31 matrices **Erreur ! Signet non défini.**

Figure 5-12. Fréquence en pourcentage (%) des p-ordres pour $p = 1, \dots, 6$ sélectionnés par le critère a) AIC ,b) BIC et c) HQC **Erreur ! Signet non défini.**

Figure 5-13. Schéma simplifié des deux différentes structures spatiales (organisation spatiale) des stations : a) selon une information physique : la direction du vent b) selon les distances métriques ascendantes. **Erreur ! Signet non défini.**

Figure 5-14. Exemple de prévisions à 10 minutes pour une journée (1er novembre) incluant deux différentes structures spatiales dans le processus du modèle (ordre des localisations selon la direction prédominante Est-Ouest des Alizés en gris et l'ordre des localisations selon les distances ascendantes en pointillé)..... **Erreur ! Signet non défini.**

Figure 5-15. MAE normalisée des prédictions de flux solaire pour le site de la Désirade en fonction des mois **Erreur ! Signet non défini.**

Figure 5-16. MBE normalisée des prédictions de flux solaire pour le site de la Désirade en fonction des mois **Erreur ! Signet non défini.**

Figure 5-17. MAE normalisée des prédictions de flux solaire pour le site de Petit-canal en fonction des mois **Erreur ! Signet non défini.**

Figure 5-18. MBE normalisée des prédictions de flux solaire pour le site de Petit-canal en fonction des mois **Erreur ! Signet non défini.**

Figure 5-19. MAE normalisée des prédictions de flux solaire pour le site de Fouillole en fonction des mois **Erreur ! Signet non défini.**

Figure 5-20. MBE normalisée des prédictions de flux solaire pour le site de Fouillole en fonction des mois **Erreur ! Signet non défini.**

Figure 5-21. Exemple de journée de classe 1. Courbes des valeurs mesurées en bleue et des valeurs prédites en vert. b) Valeurs prédites en fonction des valeurs mesurées. Le coefficient de corrélation R entre les deux signaux est de 0,99. **Erreur ! Signet non défini.**

Figure 5-22. Exemple de journée de classe 2. Courbes des valeurs mesurées en bleue et des valeurs prédites en vert. b) Valeurs prédites en fonction des valeurs mesurées. Le coefficient de corrélation R entre les deux signaux est de 0,98. **Erreur ! Signet non défini.**

Figure 5-23. Exemple de journée de classe 3. Courbes des valeurs mesurées en bleue et des valeurs prédites en vert. b) Valeurs prédites en fonction des valeurs mesurées. Le coefficient de corrélation R entre les deux signaux est de 0,96. **Erreur ! Signet non défini.**

Figure 5-24. Exemple de journée de classe 4. Courbes des valeurs mesurées en bleue et des valeurs prédites en vert. b) Valeurs prédites en fonction des valeurs mesurées. Le coefficient de corrélation R entre les deux signaux est de 0,98. **Erreur ! Signet non défini.**

Figure 5-25. RMSE normalisée en fonction des saisons pour les années 2010, 2011, 2012 sur les trois stations (Des (La Désirade), Pc (Petit-canal), Fac (Fouillole)), par application du modèle spatio-temporel VAR(1) avec les stations ordonnées selon la direction du vent. **Erreur ! Signet non défini.**

Figure 5-26. a) Préviation à 10 minutes du 7 mars 2012 et son signal mesuré correspondant. b) Préviation à 5 minutes du même jour et son signal mesuré correspondant. **Erreur ! Signet non défini.**

Figure 5-27. a) Préviation à 10 minutes du 7 mars 2012 et son signal mesuré correspondant. b) Préviation à 15 minutes du même jour et son signal mesuré correspondant. **Erreur ! Signet non défini.**

Figure 5-28. a) Préviation à 10 minutes du 7 mars 2012 et son signal mesuré correspondant. b) Préviation à 30 minutes du même jour et son signal mesuré correspondant. **Erreur ! Signet non défini.**

Figure 5-29 Préviation à 10 minutes du 7 mars 2012 et son signal mesuré correspondant. b) Préviation à 30 minutes du même jour et son signal mesuré correspondant. **Erreur ! Signet non défini.**

Figure 5-30 Signal mesuré et signaux de prévisions correspondants pour différentes périodes d'initialisation : 15 points, 2 semaines de points d'initialisation soient 616 points, 2mois de points d'initialisation soient 1364 points, 2 mois d'initialisation soient 2728 points. **Erreur ! Signet non défini.**

Figure 5-31 Résultats des RMSE normalisées des prévisions avec transformation logarithmique des données (en bleue) et sans transformation logarithmique des données (en rouge)..... **Erreur ! Signet non défini.**

Figure 5-33 Résultats des RMSE normalisées des prévisions avec transformation des données en ΔKt (en bleue) et par différenciation d'ordre 1 des données (en rouge) **Erreur ! Signet non défini.**

Figure 5-34 Valeurs des prévisions obtenues avec le modèle appliquant ΔKt des données (en noir) et les valeurs de prévisions obtenues avec différenciation d'ordre 1 des données (en bleu cyan). La courbe des valeurs observées correspondante est en bleue (1er janvier 2012). **Erreur ! Signet non défini.**

Figure 5-36 RMSE relative en pourcentage du modèle de persistance et du modèle VAR spatio-temporel en fonction de l'horizon de prévision **Erreur ! Signet non défini.**

Figure 5-37 Séries de temps passées du flux solaire des sites de Petit-canal (Pc) et de Fouillole (Fac) au temps [t-10mn ; t-20mn ;t-30mn ;t-40mn] en fonction des séries de temporelles de flux solaire de La Désirade (Des) à l'instant t.(exemple pour le mois de mars 2012). 1a) pour t-10min ; 1b) pour t- 20min ; 1c) pour t-30min ; 1d) pour t-40min.... **Erreur ! Signet non défini.**

Liste des tableaux

Tableau 1-1 Ressources locales valorisées en Guadeloupe. Source : Observatoire Régional de l’Energie et du Climat (Publication 2015).....	35
Tableau 2-1 Erreurs statistiques de prévisions solaire pour des horizons de prédiction de 15 minutes à 5 minutes répertoriés dans la littérature (Reikard (2008), Wu et Chang (2011), Yang et al (2014), Dambreville (2014))	83
Tableau 3-1 Taux de variance des quatre harmoniques obtenues par ACPF sur les trois stations.....	101
Tableau 3-2. Statistiques de la variabilité journalière des types de journées de classes 1.....	105
Tableau 3-3 Statistiques de la variabilité journalière des types de journées de classes 2.....	107
Tableau 3-4 Statistiques de la variabilité journalière des types de journées de classe 3.....	109
Tableau 3-5 Statistiques de la variabilité journalière des types de journées de classe 4.....	111
Tableau 3-6 Probabilités de transition d’une classe i vers une autre classe j avec $i=\{1, 2, 3, 4\}$ et $j= \{1, 2, 3, 4\}$ calculées en en pourcentage (%) pour le site de La Désirade sur trois années de mesures.	116
Tableau 3-7 Probabilités de transition d’une classe i vers une autre classe j avec $i=\{1, 2, 3, 4\}$ et $j= \{1, 2, 3, 4\}$ calculées en en pourcentage (%) pour le site de Petit-canal sur trois années de mesures.	116
Tableau 3-8 Probabilités de transition d’une classe i vers une autre classe j avec $i=\{1, 2, 3, 4\}$ et $j= \{1, 2, 3, 4\}$ calculées en en pourcentage (%) pour le site de Fouillole sur trois années de mesures.	116
Tableau 3-9 Tableau des comparaisons journalières des classes $\{1, 2, 3, 4\}$ présentant les journées ayant la même classe entre les sites simultanément avec la fréquence des classes en nombre de jours sur chaque site individuel.	118
Tableau 4-1 Distance inter-site	126
Tableau 4-2 Fréquence d’apparition en pourcentage des journées pour chaque type de corrélation dynamique pour l’ensemble des 3 années de mesures sans distinction des paires de sites.	174
Tableau 5-1 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d’ordres des stations pour la prédiction du flux solaire sur la Désirade en 2010.....	204
Tableau 5-2 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d’ordres des stations pour la prédiction du flux solaire sur la Désirade en 2011.....	205
Tableau 5-3 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d’ordres des stations pour la prédiction du flux solaire sur la Désirade en 2012.....	205
Tableau 5-4 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d’ordres des stations pour la prédiction du flux solaire sur Petit-canal en 2010.....	206
Tableau 5-5 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d’ordres des stations pour la prédiction du flux solaire sur Petit-canal en 2011.....	206
Tableau 5-6 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d’ordres des stations pour la prédiction du flux solaire sur Petit-canal en 2012.....	207
Tableau 5-7 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d’ordres des stations pour la prédiction du flux solaire sur Fouillole en 2010.....	207

Tableau 5-8 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d'ordres des stations pour la prédiction du flux solaire sur Fouillole en 2011	208
Tableau 5-9 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d'ordres des stations pour la prédiction du flux solaire sur Fouillole en 2012	208
Tableau 5-10 Tableau récapitulatif des résultats sur les trois années de nRMSE à un temps de retard il moyennée sur trois années.....	209
Tableau 5-11 RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 5 min sur La Désirade	217
Tableau 5-12. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 5 min sur Petit-canal.....	218
Tableau 5-13 RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 5 min sur Fouillole	218
Tableau 5-14 RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 15 min sur La Désirade	219
Tableau 5-15. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 15 min sur Petit-canal.....	220
Tableau 5-16. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 15 min sur Fouillole	220
Tableau 5-17. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 30 min sur La Désirade	221
Tableau 5-18. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 30 min sur La Désirade	222
Tableau 5-19 RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 30 min sur La Désirade	222
Tableau 5-20 RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 1 heure sur La Désirade.....	223
Tableau 5-21. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 1 heure sur Petit-canal.....	224
Tableau 5-22. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 1 heure sur Fouillole.....	224
Tableau 5-23. Les erreurs RMSE normalisées en pourcentage pour chaque trimestre dans l'année et pour chaque site du modèle VAR spatio-temporel pour des prévisions à un horizon de 10 minutes.	232
Tableau 5-24. Comparaison des erreurs de prévisions de différents modèles.....	234
Tableau 5-25. Comparaison des erreurs de prévisions des modèles présentés dans Lauret et al (2015) et notre modèle VAR spatio-temporel.	234

Introduction générale

Sommaire

I. L'enjeu de la prévision à court terme de l'ensoleillement.....	Erreur ! Signet non défini.
II. Problématique de la thèse.....	Erreur ! Signet non défini.
III. Organisation du manuscrit	Erreur ! Signet non défini.

Cette thèse porte sur une analyse spatio-temporelle de l'ensoleillement global et l'élaboration d'un modèle de prévision spatio temporel à court terme à partir de trois stations, en Guadeloupe. Dans cette introduction générale, nous présentons l'enjeu de la prévision à court terme de l'ensoleillement. Nous discuterons également de l'objectif de ces travaux de recherches à travers la définition de la problématique enfin nous présenterons l'organisation du mémoire.

I. L'enjeu de la prévision à court terme de l'ensoleillement

La caractéristique aléatoire et très fluctuante dans le temps et dans l'espace de l'énergie solaire, particulièrement en milieu intertropical, entraîne la limitation de l'insertion de cette énergie renouvelable à 30% de la puissance active totale transitant sur le réseau. Une des solutions pour repousser cette limite, est la prévision de la ressource solaire à court terme.

Cette thèse s'inscrit dans un projet global d'intégration des énergies renouvelables dans un réseau insulaire non interconnecté.

L'utilisation croissante de l'énergie solaire pour la production d'électricité nécessite une connaissance de plus en plus fine de la dynamique du flux solaire disponible en un lieu donné. Les panneaux solaires photovoltaïques ayant un temps de réponse très court, le courant qu'ils fournissent en sortie va suivre presque instantanément les variations du rayonnement solaire incident. De nombreux chercheurs cités dans l'état de l'art de ce mémoire, ont ciblé la prévision de l'ensoleillement à court terme. Voyant (2011) présente une décomposition des horizons de prédiction en fonction des moyens que doit gérer le gestionnaire du réseau électrique corse. Il établit que la prévision à court terme concerne les moyens de production hydraulique et ceux liés à l'interconnexion qui sont très rapidement exploitables, puisqu'il suffit de quelques minutes pour que l'électricité soit disponible.

Selon les besoins du gestionnaire qui estime avoir besoin d'environ 15 minutes pour changer de système de production d'électricité, la prédiction à très court terme permettra de prévoir les réserves énergétiques qui prendront le relais pour préserver la sécurité du système, en cas de production photovoltaïque très faible. De plus au niveau des industriels, la mise à jour intra journalière des prévisions permettra une meilleure vente le jour même.

En outre, la prévision à très court terme peut améliorer la gestion de stockage. Les systèmes de stockage sont une solution apportée pour la sécurisation de l'alimentation des réseaux, via des dispositifs locaux ou globaux capables de pallier toute coupure temporaire. Ce stockage peut être infra journalier (à court terme). Monjoly (2013) propose une gestion dynamique du stockage à court terme. Dans cette application, elle démontre l'avantage d'utiliser la prévision de la puissance éolienne à court terme dans une gestion optimisée du stockage pour réduire l'incertitude sur le niveau de puissance injecté au réseau.

La prévision spatio-temporelle de l'ensoleillement global définie dans ce mémoire sous-entend le besoin futur d'une étape de conversion entre le rayonnement solaire reçu par les panneaux et la production en sortie en fonction des caractéristiques des panneaux installés.

Nos travaux de recherches présentés dans ce mémoire porteront ainsi sur l'étude spatio temporelle de la variabilité à court terme et l'amélioration de la prévision à court terme de l'ensoleillement global à partir de paramètres spatiaux et temporels. Le contexte spatio-temporel de cette recherche impose une problématique dont l'objectif est de définir un modèle spatio-temporel différent des modèles traditionnels proposés dans l'industrie de prévision de l'ensoleillement à court terme.

II. Problématique de la thèse

L'étude spatio-temporelle proposée ici s'appuiera sur une base de données mesurées à une fréquence de 1 Hz sur trois stations : Petit-canal (16°38N, 61°49W) sur la côte Est de la Guadeloupe dans le lieu-dit Gros Cap, Fouillole (16°26N, 61°24W) sur le Campus de l'Université des Antilles au Sud de la Grande-Terre, et le site le plus à l'Est est celui de La Désirade (16°31N, 61°55W), situé à l'extrémité Nord Est de La Désirade. Les distances entre les sites sont comprises entre 20km et 56km. L'étude spatio-temporelle est réalisée sur trois années de mesures 2010 à 2012. L'analyse spatio temporelle consiste à analyser une variable suivant les dimensions spatiales et temporelles. Récemment, quelques chercheurs, ont utilisé les statistiques spatio-temporelles pour modéliser des processus environnementaux. Les modèles spatio-temporels dans la littérature sont définis sur un espace telle une grille. Ils sont donc traitables pour plusieurs sites formant un réseau bidimensionnel. Les études sont basées sur la quantification de l'interaction de chaque point de la grille avec son voisin. On peut citer : Glasbey et al (2008) qui présente une analyse spatio-temporelle sur une grille carrée 10km*10 km pour 10 stations, Inoue et al (2012) pour une grille de 40km*40km composée de 25 stations, Yang et al (2014) pour une grille 22km*41km composée de 13 sites.

Ce qui impliquerait pour nous, de poser de multiples capteurs sur l'ensemble du territoire. Les coûts d'un tel système seraient considérables. Notre approche dans cette étude, se limitera à effectuer une analyse spatio-temporelle sur trois stations. Les Alizés, ayant une forte composante Est, la station de La Désirade est d'un intérêt particulier pour l'analyse spatio-temporelle, puisque cette station est située sur le premier relief îlien que les masses d'air océaniques atteignent. Ainsi, les masses nuageuses se déplacent, avec une forte probabilité, de l'île de La Désirade vers la Guadeloupe continentale. Nous émettons l'hypothèse, qu'il sera possible d'effectuer une prédiction de la ressource solaire à court terme sur un site à partir de la ressource solaire des autres sites.

Notre problématique de recherche serait donc : connaissant les données sur deux stations, pouvoir estimer la donnée sur la troisième station.

Ainsi, avec peu de stations et des distances non uniformes, nous devons définir un nouveau modèle de prévision spatio-temporel qui ne répond pas à une approche classique.

Pour répondre à cet objectif, plusieurs études ont été effectuées définissant la trame de ce mémoire.

III. Organisation du manuscrit

L'objectif principal de cette recherche est de définir un modèle spatio-temporel de prévisions de l'ensoleillement global à court terme à partir de trois stations. Néanmoins, une étude préliminaire a permis de caractériser la variabilité spatio-temporelle de l'ensoleillement et d'analyser les corrélations intersites.

Le mémoire de thèse sera organisé comme suit :

Le chapitre 1 présentera l'évolution du photovoltaïque dans le mix énergétique général diversifié de la Guadeloupe. L'enjeu du photovoltaïque est ce qui permettra de montrer l'apport de ces travaux pour ce contexte énergétique.

Le chapitre 2 donnera un état de l'art des modèles de prévisions de l'ensoleillement décrits dans la littérature. Ce chapitre permettra d'évaluer les limites et performances des techniques de prévisions déjà utilisées pour le solaire. Il justifiera également le choix de la méthodologie choisie pour notre modèle spatio-temporel pour des données d'ensoleillement rares en dimension spatiale (peu de localisation).

Préalablement à l'élaboration du modèle de prévision, nous avons effectué une étude spatio-temporelle de l'ensoleillement global et avons cherché à quantifier les interactions entre nos stations de mesures.

Le chapitre 3 portera sur une classification de l'ensoleillement global sur nos trois stations afin de caractériser et comparer la variabilité de l'ensoleillement à l'échelle journalière entre les stations. Une étude des pourcentages de types de journées, de l'évolution temporelle des classes obtenues, de leur persistance et de la dynamique d'évolution des classes a été effectuée. Les données de mesures sont préalablement présentées en première section de ce chapitre.

L'objectif du chapitre 4 est de quantifier les interactions de la variabilité de l'ensoleillement à très court terme entre les stations. Une première analyse étudiera les caractéristiques de la variabilité de l'ensoleillement à très court terme pour chaque station. Une comparaison d'un point de vue énergétique sera effectuée entre les stations. Une deuxième analyse statistique quantifiera et déterminera les corrélations inter sites temporelles et spatiales des fluctuations à court terme. On définira une méthode différente pour obtenir les signaux de fluctuations dans le cadre de l'analyse des corrélations spatiales d'une part et de l'analyse des corrélations temporelles d'autre part étudiée avec séparation d'échelle de temps.

Le chapitre 5 donnera une description détaillée de la procédure du modèle de prévision spatio temporelle appliquée à nos données. L'évaluation de sa performance prédictive sera étudiée en fonction de deux structures spatiales des stations, en fonction des classes obtenues au chapitre 3, en fonction des saisons et de l'horizon. Différents tests des paramètres seront montrés en vue d'une perspective d'amélioration du modèle.

Chapitre 1. Contexte de l'étude

Table des matières

I. Situation géographique et climatique de la Guadeloupe	29
II. Les consommations d'électricité en Guadeloupe	30
III. Un mix énergétique renouvelable diversifié en Guadeloupe	31
III.1. Historique de l'énergie solaire photovoltaïque en Guadeloupe	31
III.2. La production d'électricité à partir des énergies renouvelables en Guadeloupe	Erreur ! Signet non défini.
IV. Objectifs politiques concernant la filière photovoltaïque	Erreur ! Signet non défini.
V. Perspectives d'amélioration de l'intégration de la filière photovoltaïque.....	Erreur ! Signet non défini.7
V.1. Le stockage.....	Erreur ! Signet non défini.
V.2. Smart grid	Erreur ! Signet non défini.
V.3. L'intérêt de la prévision.....	Erreur ! Signet non défini.
VI. Conclusion.....	Erreur ! Signet non défini.
Références.....	Erreur ! Signet non défini.

Ce chapitre permet de situer les travaux de thèse dans le contexte énergétique spécifique à la Guadeloupe. On comprendra alors quelles sont les problématiques de la filière photovoltaïque sur cet archipel et par conséquent quel est l'apport de nos résultats de recherche pour cette filière en Guadeloupe.

I. Situation géographique et climatique de la Guadeloupe

La Guadeloupe est un département français d'Outre-mer de superficie de 1628 km² située dans l'arc antillais à 16,15°N 61,35°O, dans la zone intertropicale.



Figure 1-1. Situation géographique de la Guadeloupe

La Guadeloupe bénéficie d'un climat de type tropical maritime, chaud et humide. L'anticyclone des Açores dirige vers les îles un vent d'Est connu sous le nom d'Alizé. A l'équateur thermique, un fort réchauffement solaire reste présent et donne naissance à une importante convection, celle-ci est accompagnée d'une zone de basse pression permanente vers laquelle se déplacent des masses d'air en provenance des anticyclones subtropicaux situés au nord et au sud. La force de Coriolis dévie la trajectoire de ces masses d'air vers l'ouest dans les deux hémisphères. L'archipel bénéficie alors d'un fort ensoleillement, tout au long de l'année, pouvant atteindre plus de 1kW.m⁻².

La variabilité spatiale et temporelle des régimes climatiques, imposée par la morphologie des îles, l'échelle et la fréquence des perturbations atmosphériques, constitue la principale particularité du climat tropical, humide et insulaire de la Guadeloupe. Les températures, l'humidité et le rayonnement solaire varient peu tout au long de l'année. Cependant, le plateau calcaire de la Grande-

Terre, les dépendances et la côte sous le vent connaissent régulièrement des périodes de sécheresse. En Basse-Terre, le relief, perpendiculaire au flux des alizés, régule le régime des pluies. Des phénomènes d'échelle synoptique ou sous-synoptique (cyclones, lignes de grains, par exemple), ou à l'échelle locale (convection diurne favorisant le développement des nuages vecteurs d'averses parfois violentes et orageuses), provoquent parfois de fortes intempéries. Enfin la saison cyclonique type s'étend normalement de début juin à fin octobre, avec un paroxysme en septembre, mais peut perdurer jusqu'à fin novembre. (Brévignon C. 2005).

La durée moyenne d'ensoleillement en Grande -Terre sud (le Raizet) est d'environ 2400h/an, en Grande -Terre nord et Iles du Sud (Godet INRA) est de plus de 2500h/an et en relief de Basse-Terre est de 1900 à un peu plus de 2300h/an (données climatiques de Météo France).

La moyenne annuelle de l'énergie reçue sur une surface orientée au Sud et inclinée à la latitude est d'environ 5,2 kWh/m².jour.

II. Les consommations d'électricité en Guadeloupe

Dépendante à 93 % aux énergies fossiles, l'économie guadeloupéenne est particulièrement sensible à l'augmentation constante du prix des énergies.

En matière d'électricité le mix énergétique est très carboné : le tiers de l'électricité consommée en Guadeloupe est produit à base de charbon et plus de la moitié à partir de fuel lourd.

Ces évolutions ont entraîné une forte augmentation des émissions de gaz à effet de serre liées à l'énergie, à savoir 40% de 2006 à 2011.

Après une très forte progression entre 2000 et 2010, les consommations d'énergie finale semblent depuis 2011 se stabiliser selon l'Observatoire Régional de l'Energie et du Climat. Elles passent de plus de 40% à moins de 1%. Les transports demeurent toujours le premier secteur de consommation d'énergie en Guadeloupe avec 68 % des consommations finales. (OREC publication 2013).

Avec près de 700 000 tonnes de produits pétroliers importés chaque année, la Guadeloupe est hautement dépendante des approvisionnements extérieurs pour sa production énergétique.

La consommation d'électricité livrée au réseau est de 1726 GWh.

Les consommations mensuelles sont relativement constantes au cours de l'année 2012 avec une moyenne de consommation autour de 143 839 MWh. Le minimum de consommation est relevé au mois de février, avec 127 088 MWh consommés et le maximum au mois de juin avec une consommation de 155 300 MWh. Ces consommations semblent directement corrélées aux conditions

de température. La production de froid pour la climatisation influe donc directement sur la consommation d'électricité.

On compte une consommation primaire totale de 848 500 tep (tonne équivalent pétrole qui est la quantité de chaleur obtenue par la combustion parfaite d'une tonne de pétrole) en 2012 : 62 000 tep sont produits grâce aux ressources énergétiques locales et 775 700 tep sont issus de ressources énergétiques importées.

Dans un contexte de demande d'électricité en constante augmentation (2,8% d'augmentation attendue pour l'année 2012), les efforts à fournir sont conséquents pour développer et diversifier les moyens de production d'électricité renouvelable, et à terme parvenir aux objectifs d'autonomie énergétique. En 2011, 12% de la production d'électricité en Guadeloupe est d'origine renouvelable, soit environ 200 GWh. Les 88% restant proviennent d'énergies fossiles, principalement diesel et charbon selon le PRERURE de Guadeloupe publié en 2012 qui est le plan énergétique régional pluriannuel de prospection et d'exploitations des énergies renouvelables et d'utilisation rationnelle de l'énergie de la Guadeloupe.

Cependant, une évolution en 2012 est constatée selon l'Observatoire Régional de l'Energie et du Climat.

En effet, la Guadeloupe atteint 15,6 % de production d'électricité à partir d'énergie renouvelable en 2012 avec un nouveau record de 17,45% en 2013.

Malgré cette évolution et un important développement des énergies renouvelables, la Guadeloupe n'a pas encore atteint les objectifs d'autonomie énergétique imposés par le PRERURE dans lequel est stipulé entre autres que la Guadeloupe doit produire 25 % d'énergies renouvelables dans le mix énergétique en 2020 et 50 % en 2030, pour une autonomie énergétique en 2050.

III. Un mix énergétique renouvelable diversifié en

Guadeloupe

III.1. Historique de l'énergie solaire photovoltaïque en Guadeloupe

En 1998, la Guadeloupe ne compte dans son mix énergétique que trois ressources renouvelables : la géothermie, l'hydraulique et l'éolien. La géothermie fait son entrée avec près de 50% de la part d'électricité produite à partir d'énergie renouvelable. Ensuite vient la bagasse en 1999 avec près de trois cinquième de la part d'électricité produite à partir d'énergie renouvelable dans le mix énergétique. La production d'électricité à partir d'énergie solaire photovoltaïque n'a débuté qu'en

2005, elle représente à peine 1% de la part d'électricité produite à partir d'énergie renouvelable avec une production nette de 291 MWh.

En 2010, une diminution de la production géothermique a eu lieu, ce qui a entraîné une chute de la part de production d'électricité d'origine renouvelable avec un pourcentage de 9% au lieu de près de 12% en 2009. Néanmoins en 2012, la Guadeloupe va connaître un maximum historique de la part de production d'électricité à partir d'énergies renouvelables grâce au photovoltaïque avec une production nette pratiquement triplée de valeur 91 827 MWh (Figure 1-3). La part du photovoltaïque représente dorénavant en 2012, 5,3 % de la production totale d'électricité (91 827 MWh), ce qui la classe comme la première énergie renouvelable de l'île. L'essor du photovoltaïque représente désormais plus du tiers de la production énergétique d'origine renouvelable (Figure 1-2).

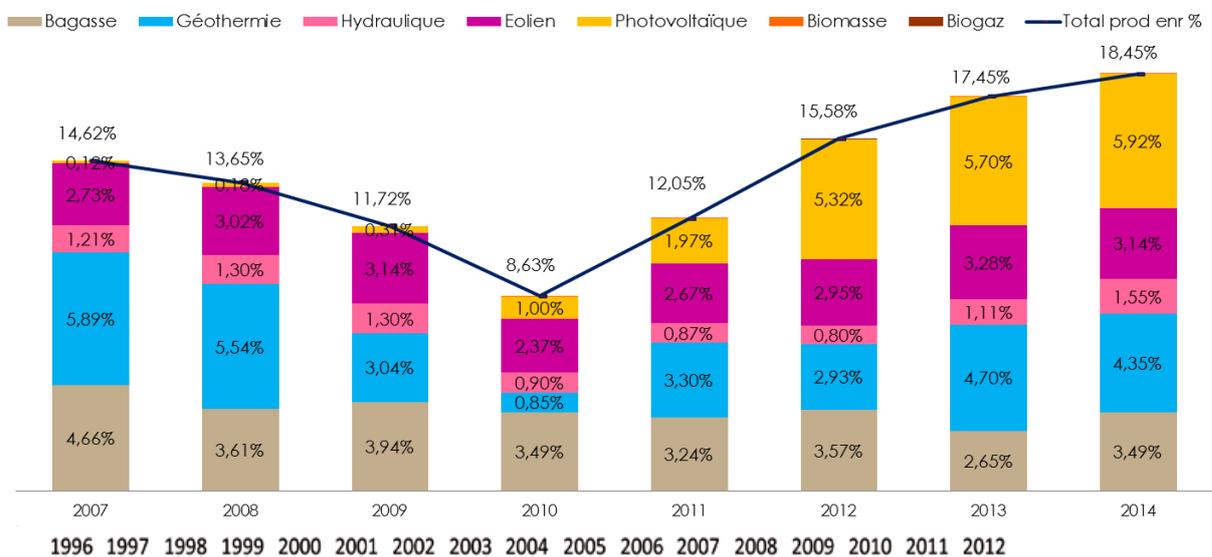


Figure 1-2. Evolution de la part de l'électricité produite à partir d'énergie renouvelable dans le mix énergétique (source : Observatoire Régional de l'Energie et du Climat (OREC) en Guadeloupe Publication 2015)

En effet, seuls 5MW photovoltaïques étaient installés fin 2008, la Figure 1-3 nous montre près de 2959 MWh de production nette pour cette même année. Néanmoins, une forte croissance de production nette de ce parc peut être observée dès 2010. En 2012, le parc photovoltaïque installé en Guadeloupe atteint une puissance de 91 827 MWh. En fin 2013, on compte en puissance installée maximale, une totalité de 66,85 MWc selon les données du gestionnaire de réseau EDF SEI. Actuellement, les installations de panneaux solaires photovoltaïques sont réparties sur toute la Guadeloupe (Figure1-4). La puissance installée la plus importante se trouve sur la commune de Baie-Mahault avec près de 17,34 MWc. Les trois premières puissances installées, respectivement Baie-

Mahault, Petit-canal et Le Moule représentent à elles seules près de 60% de la part de puissance installée de panneaux solaires photovoltaïques.

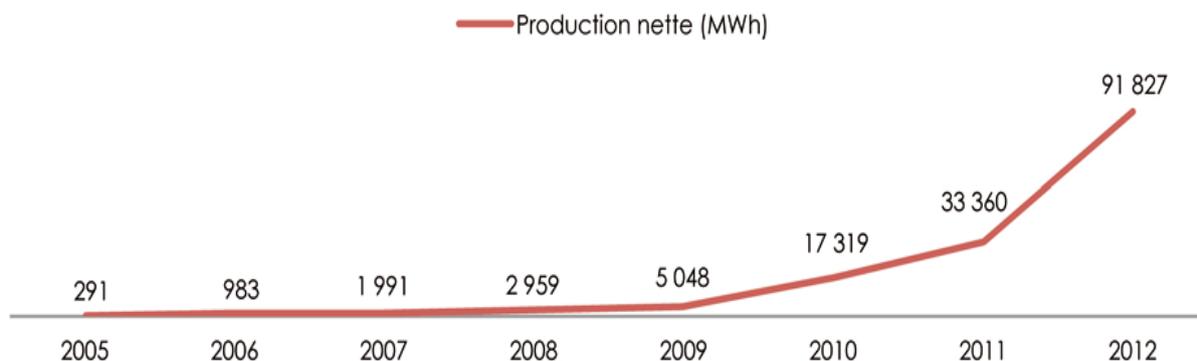


Figure 1-3. Source EDF : Production d'électricité à partir d'énergie solaire photovoltaïque de 2005 à 2012

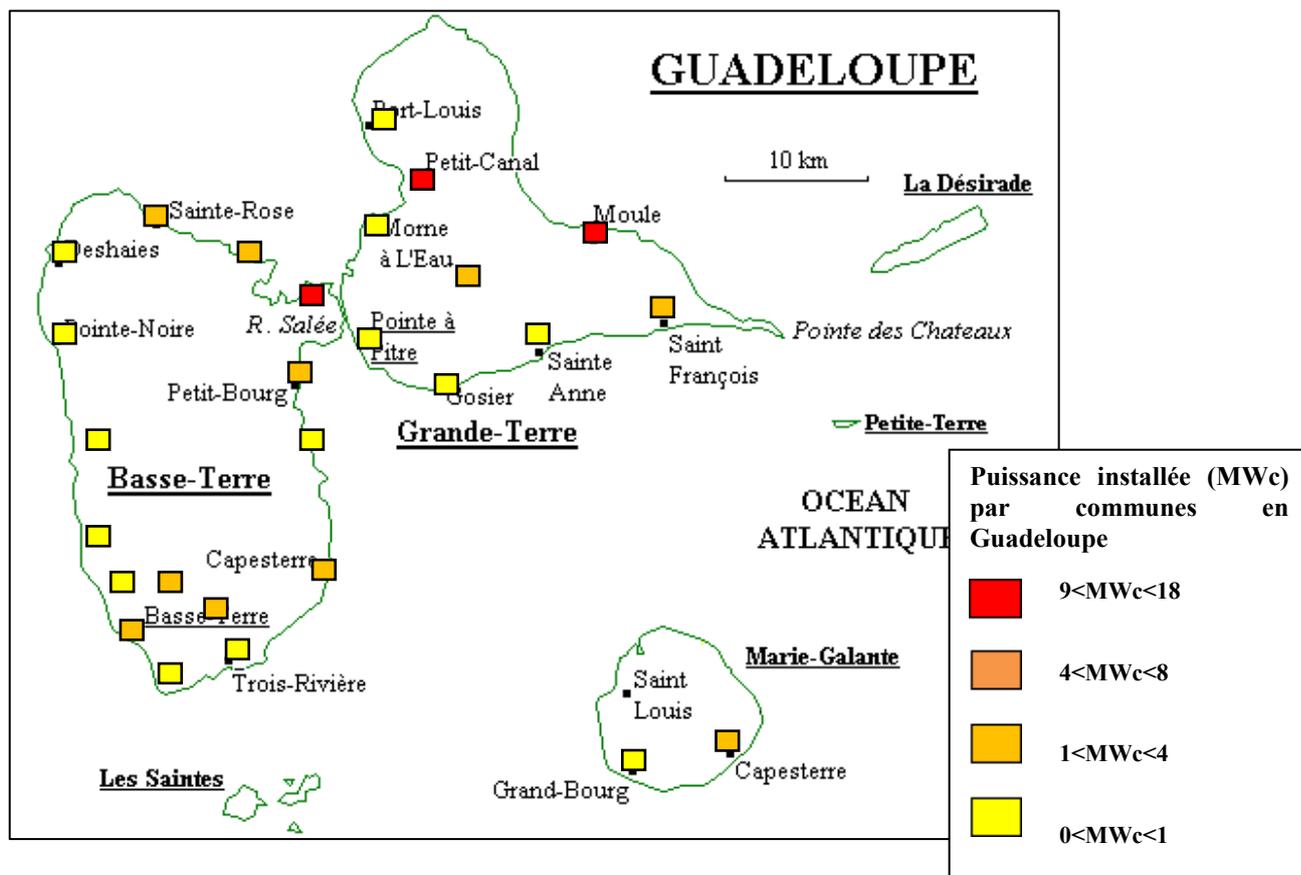


Figure 1-4. Répartition des panneaux solaires photovoltaïques de l'archipel Guadeloupéen (carte réalisée à partir des données d'EDF).

III.2. La production d'électricité à partir des énergies renouvelables en Guadeloupe

La Guadeloupe présente une diversité de sources d'énergie qui est une particularité forte de ce territoire. Le fuel et le charbon représentent l'essentiel de nos ressources énergétiques.

Néanmoins, la Guadeloupe a su développer de nombreuses sources d'énergies renouvelables : la géothermie, la biomasse, mais également l'hydraulique, l'éolien, le photovoltaïque.

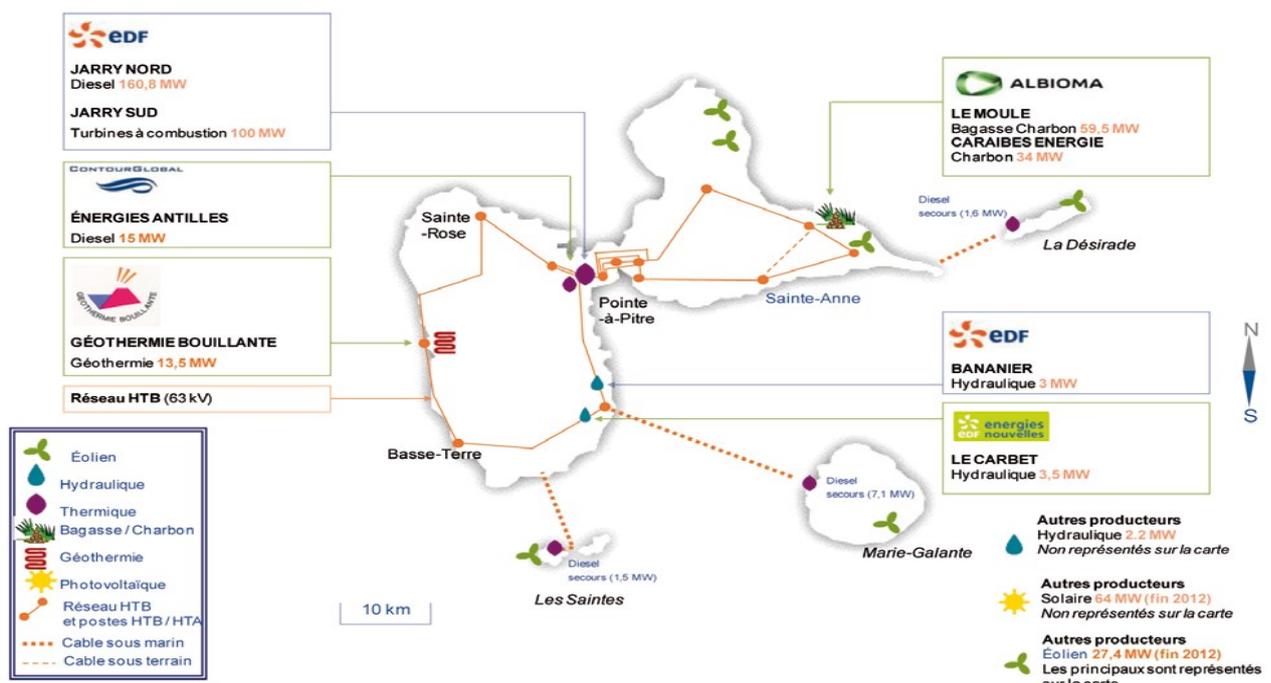


Figure 1-5. Schéma du système électrique prévisionnel. Source EDF SEI : Bilan prévisionnel de l'équilibre offre demande d'électricité juillet 2013

La localisation des exploitations des différentes sources d'énergie est représentée sur la carte Figure 1-5. Elles sont réparties sur l'ensemble du territoire. En 2011, l'usine géothermique de Bouillante ainsi que la centrale thermique du Moule (à travers sa consommation de biomasse) sont les deux principaux sites producteurs d'électricité d'origine renouvelable avec respectivement 28 % et 27 % de l'électricité d'origine renouvelable. La production d'électricité renouvelable est très dépendante de ces deux sites. Viennent ensuite les parcs éoliens et les parcs photovoltaïques. Par ailleurs, la Guadeloupe compte deux installations hydroélectriques qui contribuent à 7 % de la production totale.

Le tableau ci-dessous rend compte d'une production annuelle d'électricité à base d'énergies renouvelables conséquentes en 2012. Ces ressources locales entièrement valorisées sous la forme

d'énergies renouvelables ont permis de produire 269 006 MWh en 2012. La production d'électricité s'élève à 1726,1GWh en 2012 dont 269 GWh provenant d'énergies renouvelables. On peut constater dans le tableau ci-dessous, que la production d'électricité la plus importante est d'origine photovoltaïque avec 34% de la part de la production d'électricité d'origine renouvelable. Par rapport à l'année 2011, la part de production du photovoltaïque a nettement évolué. Cette part importante est due à une augmentation historique du photovoltaïque en 2012.

	2012		2013		2014	
	MWh	%	MWh	%	MWh	%
Biogasse	61691	23%	45853	15%	60456	19%
Hydraulique	13870	5%	19210	6%	26801	8%
Eolien	50901	19%	56794	19%	54482	17%
Photovoltaïque	31827	34%	98642	33%	102726	32%
Géothermie	50556	19%	81347	27	74877	23
Bologne-biomasse	18	<1%	12	<1%	9	<1%
Bologne-biogaz	83	<1%	12	<1%	9	<1%
Total	269006		301858		319903	

Tableau 1-1 Ressources locales valorisées en Guadeloupe. Source : Observatoire Régional de l'Energie et du Climat (Publication 2015)

En 2013, les énergies renouvelables permettent de produire 17,4% de la production totale d'électricité soit 301 858 MWh. De 2012 à 2013, la Guadeloupe a réalisé une évolution interannuelle de la production des énergies renouvelables de 12%. La Guadeloupe enregistre un nouveau record de 18,5% en 2014 selon l'Observatoire Régional de l'Energie et du Climat (publication 2015).

IV. Objectifs politiques concernant la filière

photovoltaïque

L'énergie est une problématique transversale qui se retrouve dans les différentes politiques sectorielles du territoire.

Les objectifs de développement du parc de production par source d'énergie primaire renouvelable sont fixés. Le schéma régional de raccordement au réseau des EnR (énergie renouvelable) est soumis à l'approbation conjointe du préfet et du président du Conseil Régional.

Le Président du Conseil Régional émet un avis conforme sur la programmation pluriannuelle des investissements de production d'énergie, sur la base des conclusions et objectifs du PRERURE. En effet, depuis 2008, la Guadeloupe dispose d'une planification énergétique. Il s'agit du plan énergétique régional pluriannuel de prospection et d'exploitation des énergies renouvelables et d'utilisation rationnelle de l'énergie - le PRERURE.

Ce plan énergétique définit la politique régionale de demande et d'offre énergétique centrée sur l'amélioration de l'efficacité énergétique et la valorisation des énergies renouvelables, dans une perspective de développement durable. Il constitue l'outil de référence pour la définition de la politique régionale de l'énergie. La réalisation et la mise en œuvre de ce plan doit ainsi permettre à la Région Guadeloupe de tirer parti des gisements d'économie d'énergie et des ressources énergétiques et de se fixer en particulier un objectif en termes d'autonomie énergétique pour l'horizon 2050. En 2012, le SRCAE, le schéma régional climat air énergie de la Guadeloupe est venu compléter cette planification énergétique en introduisant un volet air axé sur les polluants atmosphériques et un volet climat (atténuation et adaptation). La Région Guadeloupe, l'ADEME et la Commission européenne par le biais des fonds du FEDER, subventionnent les projets de maîtrise des consommations énergétiques, de limitation des gaz à effet de serre, et de développement des énergies renouvelables, dans le cadre de conventions annuelles annexées au contrat de projet Etat-Région 2007-2013 et du PO FEDER 2007-2013.

Les projets futurs de la filière énergie renouvelable, particulièrement les projets concernant les panneaux solaires photovoltaïques sont liés à la faisabilité de dépasser le seuil des 30 % ou de s'en affranchir. En effet, préserver la stabilité du réseau électrique dans les zones non interconnectées (ZNI) requiert une limite d'intégration des énergies intermittentes dans le système électrique fixée à 30% de la puissance maximale injectée sur le réseau. Le 20 juillet 2010 les quotas PV-éolien sont règlementés en fixant les règles de déconnexion du réseau des centrales PV et éoliennes.

Les solutions liées à la faisabilité de dépasser le seuil des 30% ou de s'en affranchir sont :

- soit le développement des solutions de stockage sous forme décentralisé de type batterie à l'échelle des installations PV ou de dispositif centralisé par exemple sous forme de Station de Transfert d'Énergie par Pompage (STEP) ;
- soit une amélioration de la gestion du réseau associée à la prédiction ou au développement de smart grids.

Dans le scénario PRERURE, la filière PV compte au total 124 MW installés en 2030, dont 36 MW avec stockage. La production PV en Guadeloupe doit s'élever à 173 GWh en 2030.

La production de chacune des filières PV et éolien permettra ainsi de couvrir 7% de la demande électrique en 2020 et 9% en 2030. Leur production cumulée représenterait 14% du mix électrique en 2020 et 18% en 2030. Les objectifs semblent ambitieux, mais la Guadeloupe compte actuellement en 2014, 5,92% de la demande électrique pour la filière PV et 3,14% pour la filière éolienne soit 9,06% pour leur production cumulée selon le bilan 2015 de l'OREC (Observatoire Régional de l'Energie et du Climat). Le développement de ces filières est fortement conditionné par les décisions politiques au niveau national : tarifs d'achat et évolution du cadre réglementaire. Les financeurs de projets de développement des énergies renouvelables sont principalement la Région, l'Etat, FEDER, l'ADEME et EDF. L'atteinte des objectifs définis par le PRERURE suppose une modification des niveaux de consommation d'énergie et des modes de production : à terme, la production d'énergie renouvelable devra couvrir 50 % de la consommation d'énergie finale. Cela implique une diminution des consommations d'énergie dans l'ensemble des secteurs (transports compris) et un développement à un rythme soutenu de la production d'énergie renouvelable. Pour atteindre les 14% du mix électrique d'ici les cinq années prochaines, il va donc falloir développer des solutions pour le développement de ces énergies.

V. Perspectives d'amélioration de l'intégration de la filière photovoltaïque

La sensibilité des systèmes électriques, constatée dans les petits systèmes insulaires, a conduit le ministère en charge de l'énergie à y limiter la puissance provenant de sources d'énergies intermittentes à 30% de la puissance globale injectée sur le réseau. En Guadeloupe, il arrive que ce seuil soit désormais dépassé.

Des solutions sont à l'étude dans les systèmes insulaires pour atténuer le caractère intermittent du solaire et de l'éolien et ainsi, permettre leur développement tout en préservant la stabilité des réseaux électriques :

- les technologies de stockage d'énergie
- Projet de Smart grid
- des modèles de prévisions de l'ensoleillement ou du vent .

V.1. Le stockage

Le stockage consiste à constituer un stock d'énergie potentielle à partir de flux d'énergie dont on n'a pas l'usage immédiat, afin de pouvoir en disposer plus tard, lorsque la demande sera plus importante.

L'indépendance énergétique est stratégique et économiquement essentielle pour La Guadeloupe. Pour les individus et les entreprises, l'énergie doit impérativement être disponible à la demande, sans coupure inopinée. Le stockage d'énergie répond à trois motivations principales :

- sécurisation de l'approvisionnement en énergie de la Guadeloupe ;
- ajustement de la production d'énergie en fonction de la demande ;
- compensation de l'irrégularité de la production des énergies dites intermittentes.

Grâce aux STEP, qui sont capables de stocker l'électricité, l'intermittence des énergies renouvelables ne serait plus un obstacle au développement du photovoltaïque et de l'éolien dans les départements d'outre-mer.

La chute d'eau nécessaire au fonctionnement des STEP marines serait créée entre un lac artificiel, creusé en haut d'une falaise et le niveau de la mer.

L'eau de mer serait pompée et stockée dans un lac. Le rejet de l'eau dans la mer par turbinage permettrait de produire de l'électricité au moment voulu.

Selon EDF (source : site EDF Guadeloupe archipel), on pourrait installer jusqu'à 5.000 MW de STEP marines en France, en incluant les falaises situées en métropoles.

Le relief des DOM n'étant pas assez marqué pour y envisager la construction de nombreuses STEP, EDF étudie la réalisation de deux projets de STEP marines : l'un en Guadeloupe (où se trouvent des falaises de 50 mètres), l'autre à la Réunion (falaises allant jusqu'à 100 mètres de hauteur).

V.2. Smart grid

Le smart grid est un réseau de distribution d'électricité « intelligent » qui utilise des technologies informatiques de manière à optimiser la production, la distribution, la consommation et qui a pour objectif d'optimiser l'ensemble des mailles du réseau d'électricité qui va de tous les producteurs à tous les consommateurs afin d'améliorer l'efficacité énergétique de l'ensemble.

L'intégration des nouvelles technologies de l'information et de la communication aux réseaux les rendra communicants et permettra de prendre en compte les actions des acteurs du système électrique, tout en assurant une livraison d'électricité plus efficace, économiquement viable et sûre.

Le système électrique sera ainsi piloté de manière plus flexible pour gérer les contraintes telles que l'intermittence des énergies renouvelables, ce qui aura également pour effet de faire évoluer le système actuel, où l'équilibre en temps réel est assuré en adaptant la production à la consommation, vers un système où l'ajustement se fera davantage par la demande. Ainsi, contrairement aux caractéristiques des réseaux actuels, les smart grid seront numériques, bidirectionnels avec une production décentralisée.

Un système de réseau intelligent, chez les consommateurs, est en cours d'expérimentation dans le cadre du projet Millener. En 2009, l'ADEME a lancé un appel à manifestation d'intérêt sur les réseaux intelligents « smart grid ». Millener est né de cet appel. Ce projet a trois terrains d'action : la Corse, la Réunion et la Guadeloupe. Millener se compose :

- de panneaux solaires photovoltaïques distribués aux clients avec des batteries qui permettent à celui-ci d'avoir une autosuffisance s'il a besoin d'une puissance importante ou s'il y a une baisse importante de la production d'électricité à partir de l'énergie solaire par rapport à la consommation. Ce qui permet de limiter la consommation des clients avec le réseau.
- une passerelle énergétique afin d'encourager la maîtrise de l'énergie
- un système d'agrégation et de pilotage géré depuis le dispatching, tour de contrôle du système.

Ce projet innovant pour les îles est en ce moment à l'étude avec près de 1500 clients.

V.3. L'intérêt de la prévision

L'intégration des énergies renouvelables intermittentes telles que l'éolien et le solaire au sein de réseaux électriques est une problématique importante comme il l'a été mentionné précédemment. A cause de leur caractère aléatoire et intermittent, ce type d'énergie ne peut être inséré que de manière limitée afin de mieux préserver la sécurité du réseau et une qualité de fourniture. En effet, la puissance globale de l'éolien et du photovoltaïque n'est pas garantie et leur sensibilité à la source primaire oblige les autres moyens de production à compenser leurs variations. L'énergie solaire est une énergie présentant un profil de production fluctuant. Les panneaux solaires photovoltaïques ayant un temps de réponse très court, le courant qu'ils fournissent en sortie va suivre presque instantanément les variations du rayonnement solaire incident (Woyte et al 2006). En effet, comme on peut l'observer sur la Fig.1-6 représentant l'évolution journalière de la production d'énergie photovoltaïque par pas de temps de 2 minutes, la puissance produite connaît des variations brutales et de forte amplitude, les signaux présentent de nombreuses fluctuations à court terme caractéristiques du flux solaire.

La Guadeloupe présente une problématique supplémentaire au caractère intermittent des énergies renouvelables : elle n'est pas interconnectée avec d'autres systèmes électriques et toute l'électricité

doit être produite sur le territoire. La non interconnexion à un grand réseau comme celui de l'Europe continentale est une fragilité structurelle : un court-circuit sur un ouvrage haute tension engendre un creux de tension sur toute l'île. La puissance de la grille connectée aux installations photovoltaïques augmente rapidement et peut interférer avec la stabilité réseau. De plus, la perte d'un groupe de production dans un système électrique insulaire engendre des perturbations importantes (Lew et al 2010). Ainsi le caractère insulaire du réseau électrique Guadeloupéen est une source intrinsèque d'instabilité du réseau électrique. Tant que la puissance d'origine solaire fournie par les panneaux photovoltaïques demeurait faible cela n'avait pas d'influence sur la stabilité du réseau. En 2012, l'Observatoire Régional de l'Energie et du Climat compte 64MW de puissance installée d'origine photovoltaïque raccordés au réseau, ce qui n'est plus négligeable.

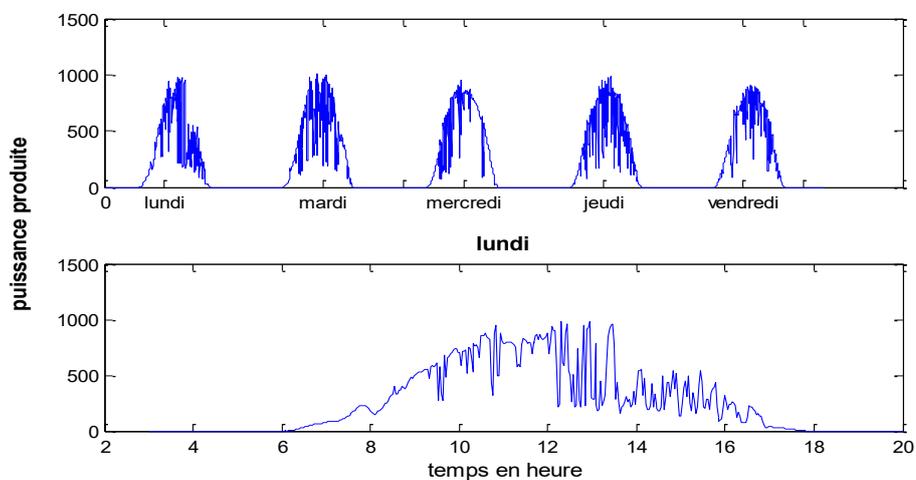


Figure 1-6. Evolution de la puissance du photovoltaïque par période d'échantillonnage de 2 minutes pendant 9 jours à partir du 2 septembre 2013. Mesures effectuées sur une installation PV de l'Université.

Les solutions pour palier à ces problèmes seraient alors :

- d'augmenter la réserve primaire et donc augmenter la production d'énergie, les coûts et la pollution
- le foisonnement mais à l'échelle d'un département c'est très limité.
- Stocker les énergies renouvelables aléatoires qui reste un projet très coûteux
- Adapter la charge : asservissement de certains usages à la production, utilisation de compteurs communicants
- la prévision est la solution la moins chère.

La gestion d'un système électrique s'effectue en temps réel. Le gestionnaire de réseau doit en permanence assurer l'équilibre offre-demande. Il a, en temps réel, la puissance réelle fournie par l'ensemble des fermes solaires photovoltaïques de la Guadeloupe. La prévision de l'ensoleillement global permet au gestionnaire d'avoir un aperçu de l'évolution future de la puissance fournie par des fermes de panneaux solaires photovoltaïques, en d'autres termes cette prévision lui permettra d'anticiper sur les prochains événements connaissant la disponibilité de la puissance à quelques heures.

VI. Conclusion

La production d'électricité à partir des énergies renouvelables en Guadeloupe est caractérisée par la diversité de ses sources d'énergies : la géothermie, la biomasse, mais aussi l'hydraulique, l'éolien et le solaire. Cette diversité énergétique est une particularité forte de ce territoire qui a permis d'atteindre un nouveau record en 2014 de 18,5% de production d'électricité à partir des énergies renouvelables. La part importante du photovoltaïque due à une augmentation historique en 2012 en est principalement la cause. En effet, plus d'1/3 de la part de production d'électricité produite par les énergies renouvelables est d'origine photovoltaïque en 2012. La production de PV a effectivement atteint une production de 91 827 MWh en 2012 comparé à 291 MWh en 2005. Néanmoins, l'énergie solaire est une énergie aléatoire caractérisée par d'importantes fluctuations à court terme particulièrement en milieu intertropical, ce qui aura pour conséquence une production d'électricité à partir des panneaux solaires photovoltaïques très variable.

Cette caractéristique aléatoire a entraîné la limitation de l'insertion de cette ENR non garanties à 30% de la puissance active totale transitant sur le réseau. Néanmoins, Le PRERURE a défini des objectifs impliquant un développement à un rythme soutenu de la production d'énergie renouvelable et entre autre, une diminution des consommations d'énergie dans l'ensemble des secteurs (transports compris). A terme, la production d'énergie renouvelable devra couvrir 50 % de la consommation d'énergie finale.

L'augmentation de l'utilisation de l'énergie solaire dans le contexte énergétique pour la production d'électricité présente donc un enjeu important pour atteindre ces objectifs et augmenter la compétitivité des Régions d'Outre-Mer dans les énergies renouvelables.

Un développement de la production PV requiert une connaissance statistique et dynamique plus fine de la ressource mais aussi des outils de prévision plus performants. La connaissance du comportement à court terme de fermes solaires permettra d'offrir une opportunité pour assurer des garanties au

gestionnaire du réseau et d'obtenir un gain industriel dans la maîtrise des potentialités du système de production.

Les travaux de recherche présentés dans cette thèse proposeront des outils de prévisions performants qui tiennent compte de la variabilité temporelle et spatiale de l'ensoleillement et tenteront d'apporter des connaissances fines de la dynamique de l'ensoleillement sur trois stations situés dans l'Archipel Guadeloupe.

Références

Brévignon C., 2005. L'environnement atmosphère de la Guadeloupe, de St Barthélémy et de St Martin, La direction Inter Régionale Antilles Guyane de Météo France.

Météo France de Guadeloupe : www.meteo.gp

Belfort Amélie, avec l'appui du comité de pilotage de l'Observatoire (Région Guadeloupe, ADEME, DEAL, EDF et Météo-France),

Observatoire Régional de l'Energie et du Climat, publication 2013

Observatoire Régional de l'Energie et du Climat, publication 2015

Périanin Jessy, Etude de la prédictibilité des énergies renouvelable en Guadeloupe, Rapport de stage Master 2 REMI-VERT, 2007-2008

PRERURE Guadeloupe, Plan énergétique régional pluriannuel de prospection et d'exploitation des énergies renouvelables et d'utilisation rationnelle de l'énergie de la Guadeloupe, publication septembre 2012

Kostylev APV, 2011. Solar power forecasting performance – towards industry standards. In: First international workshop on the integration of solar power into power systems Aarhus, Denmark,.

Lew, D., Piwko, R., 2010. Western wind and solar integration study. NREL Technical Report, NREL/SR-550-47434. Golden, Colorado: National Renewable Energy Laboratory.

Woyte, A., Thong, V.V., Belmans, R., Nijs, J.c., 2006. Voltage fluctuations on distribution level introduced by photovoltaic systems , IEEE Transactions on Energy Conversion 21, 202-209.

Chapitre 2. Etat de l'art pour la prédiction du rayonnement solaire

Table des matières

I. Horizon de prédiction	Erreur ! Signet non défini.
I.1. Horizon de prévision à très court terme	Erreur ! Signet non défini.
I.2. Horizon de prévision à court terme	Erreur ! Signet non défini.
I.3. Horizon de prévision à moyen terme	Erreur ! Signet non défini.
II. Prévisions météorologiques numériques (NWP)	Erreur ! Signet non défini.
II.1. GFS	Erreur ! Signet non défini.
II.2. NAM : Modèle Mésoéchelle Nord-Américain (North American Mesoscale Model)	Erreur ! Signet non défini.
Signet non défini.	
II.3. Centre européen de prévisions à moyen terme (ECMWF).....	Erreur ! Signet non défini.
II.4. MOS.....	Erreur ! Signet non défini.
II.5. Aperçu des travaux NWP pour la prévision solaire.....	Erreur ! Signet non défini.
II.6. Limites des modèles NWP.....	Erreur ! Signet non défini.
III. Imagerie des nuages	Erreur ! Signet non défini.
III.1. Imagerie au sol des nuages	Erreur ! Signet non défini.
III.2. Imagerie par satellite des nuages	Erreur ! Signet non défini.
III.3. Aperçu des travaux d'imagerie des nuages	Erreur ! Signet non défini.
III.4. Limites des modèles d'imagerie des nuages	Erreur ! Signet non défini.
IV. Modèles statistiques	Erreur ! Signet non défini.
IV.1. Modèle de persistance	Erreur ! Signet non défini.
IV.2. Modèles linéaires	Erreur ! Signet non défini.
IV.3. Modèles non-linéaires	Erreur ! Signet non défini.
V. Modèles hybrides	Erreur ! Signet non défini.
V.1. Aperçu des travaux de modèles hybrides	Erreur ! Signet non défini.
V.2. Limites des modèles hybrides.....	Erreur ! Signet non défini.

VI. Modèles spatio-temporels et aperçu des travaux	Erreur ! Signet non défini.
VI.1. Modèle STARMA	Erreur ! Signet non défini.
VI.2. Modèle krigeage.....	Erreur ! Signet non défini.
VI.3. Modèle VARMA.....	Erreur ! Signet non défini.
VII. Synthèse et conclusion de l'état de l'art	Erreur ! Signet non défini.
Références	Erreur ! Signet non défini.6

Dans ce chapitre, un état de l'art de la prévision du rayonnement solaire est présenté. Nous essayerons de recenser de manière globale les différentes méthodologies utilisées et les travaux de recherche correspondants, dans le domaine de la prédiction du rayonnement solaire.

Cette démarche est préliminaire à l'application d'un modèle de prédiction pour notre étude.

A partir de l'étude des différents modèles existants, nous préciserons dans la dernière section de ce chapitre, le choix du modèle de prédiction utilisé pour nos travaux de recherche correspondant à notre contexte d'étude. Soulignons que l'objectif final de notre recherche est d'établir un modèle de prédiction spatio-temporel de l'ensoleillement global à court terme.

I. Horizon de prédiction

L'horizon de prédiction envisagé, dépend des besoins du gestionnaire de réseau. On peut classer les horizons de prédiction en trois catégories (Voyant, 2011):

- Le moyen terme, qui prédit la ressource disponible les jours suivants ($j+1, j+2, j+3$).
- Le court terme, qui prédit la ressource au moins une heure à l'avance ($h+1/2, h+1$).
- Le très court terme, correspondant à l'horizon de quelques minutes apportant une prédiction utile pour les systèmes d'interconnexion (échelle de temps nécessaire pour que l'électricité soit disponible).

Dans la littérature, les modèles de prédiction sont basés sur différentes méthodologies. De manière générale, nous pourrions dire que deux approches sont envisagées :

- une approche météorologique basée sur la résolution de modèles numériques permettant en général la prédiction des variations temporelles et du potentiel du rayonnement solaire
- une approche locale basée sur le développement de modèles statistiques de prévision, en un point donné, à partir de mesures in situ. Il est plus simple d'envisager dans ce cadre une prévision sur les échelles de temps à court terme.

Les modèles de prévision du rayonnement global horizontal, peuvent être également classés selon les données d'entrée utilisées qui déterminent l'horizon de prévision (très court terme, court terme, long terme). Nous présenterons, de manière globale, l'intérêt de prédire l'ensoleillement à un horizon défini et la méthode de prévision la plus couramment utilisée à cet horizon.

I.1. Horizon de prévision à très court terme

La variabilité du rayonnement solaire a un impact sur le système photovoltaïque pour une échelle temporelle précise. Kostylev et al (2011), définissent la granularité du système photovoltaïque avec l'horizon de prédiction correspondant. Un horizon de 15 minutes à 2 heures avec 30 secondes à 5

minutes concerne les phénomènes dits de ramping qui correspondent entre autres à la montée progressive de la production photovoltaïque. L'échelle de temps de 15 minutes à 30 minutes correspond au temps nécessaire pour que le gestionnaire de réseau puisse optimiser au mieux l'arrêt ou le démarrage d'un moyen de production.

Ainsi, dans la littérature, plusieurs chercheurs ont ciblé l'horizon de prévision à très court terme. La deuxième approche (modèles statistiques) basée sur des mesures de l'ensoleillement, est celle souvent appliquée sur des prévisions à très court terme, de 5 minutes à 6 h dans Reikard (2009). Ces modèles utilisent les séries chronologiques du rayonnement. On peut citer les modèles Auto Régressifs (AR), les modèles autorégressifs à moyenne mobile (ARMA ou ARIMA), les fonctions de transfert, les réseaux de neurones et certains modèles hybrides. Les modèles utilisent un échantillon d'apprentissage des données passées avant le début de la période de prévision et les valeurs prédites sont comparées avec les valeurs réelles.

Néanmoins, la méthode récente basée sur l'imagerie du ciel a pu obtenir des résultats de prédiction pour cet horizon. Dans Chi Wai Chow et al (2011), des vecteurs de mouvement de nuages ont été générés par corrélation croisée de deux images de ciel consécutives. Des prévisions ont été effectuées jusqu'à 5 min par advection des deux dimensions de carte de nuage. Plus de détails seront donnés dans ce chapitre.

I.2. Horizon de prévision à court terme

Pour des prévisions à court terme de l'ensoleillement, des informations sur l'évolution temporelle des nuages, qui est la source dominante de la variabilité du flux solaire, sont souvent utilisés comme une base. L'application d'une méthode statistique pour détecter le mouvement des structures des nuages à partir d'images satellites, pour une plage temporelle de 30 min à 6 h, est souvent développée dans la littérature. Cette prédiction permet d'apporter des connaissances sur les moments critiques où le gestionnaire de réseau devra être vigilant (délai d'allumage des dispositifs thermiques, turbines à combustion, moteurs thermiques). Selon Kostylev et al (2011), un horizon de 1h à 6h est défini comme une prévision à court terme.

I.3. Horizon de prévision à moyen terme

Les horizons de prévisions d'environ 4 à 6h jusqu'à plusieurs jours sont essentiellement basées sur la prévision numérique du temps (NWP) qui sont des modèles généralement effectués pour des prévisions par satellite (Perez et al. (2010)). Les prévisions qui s'étendent à plus de 6 h sont basées sur le mouvement des nuages provenant d'images consécutives de satellites géostationnaires. Ces prévisions sont modélisées à partir des prévisions de la couverture nuageuse maillée.

L'utilité de ce type de prédiction se trouve essentiellement dans la gestion des stocks fossiles. Les modèles de charge prévus pour les 2 prochains jours constituent la base pour la planification des centrales et les opérations de planification dans le marché de l'électricité afin d'équilibrer l'offre et la demande d'énergie, et assurer un fonctionnement fiable du réseau (Lorenz et Heinemann (2012)). Ces prévisions sont utilisées par les entreprises de services publics, les opérateurs de réseaux de transport, les fournisseurs de services énergétiques, des négociants en énergie et producteurs d'énergie indépendantes dans leur programmation, l'envoi et le règlement de la puissance.

Ce chapitre présentera des sections définies selon trois types de modèles : la prédiction de temps météorologique numérique, l'imagerie des nuages et les modèles statistiques. Une section sur les modèles spatio-temporels sera ajoutée.

II. Prévisions météorologiques numériques (NWP)

Une connaissance précise de l'état météorologique de l'atmosphère et des lois physiques qui gouvernent les passages de transitions de l'atmosphère d'un état à un autre est liée à la réussite d'une bonne prévision du rayonnement solaire. Cette connaissance, sur la structure physique de l'atmosphère, constitue le principe de la méthodologie des NWP.

Un domaine physique est choisi, il est ensuite discrétisée à une résolution souhaitée. Les lois physiques thermodynamiques et du mouvement, la mécanique des fluides et la chimie sont alors numériquement résolues sur la grille spatiale. La Prévision météorologique numérique (NWP) est généralement l'outil le plus précis de prévision du rayonnement solaire pour plusieurs heures à l'avance. Le processus des modèles NWP peut être résumé en six équations physiques basiques :

- la loi des gaz qui donne la relation entre la pression, la densité et la température
- l'équation hydrostatique qui donne la relation entre la densité de l'air et la diminution de la pression avec l'altitude
- l'équation de continuité qui est une expression de la conservation de la masse et est utilisée pour déterminer la vitesse du vent verticale et les variations de pression en surface
- l'équation thermodynamique qui exprime l'évolution de la température d'un fluide (air) du au refroidissement adiabatique ou au réchauffement au cours du déplacement vertical. D'autres processus physiques comme la condensation, l'évaporation, le transport turbulent et des effets radiatifs y sont aussi inclus.
- l'équation du mouvement qui décrit l'accélération et la décélération de la vitesse d'un fluide (air) du à la force du gradient de pression, et la manière dont la force de Coriolis contribue à influencer sur la direction.
- l'équation de la conservation de l'humidité.

Des équations pronostiques spécifiques à la fraction nuageuse, l'eau, la teneur en glace et l'ozone, sont également utilisées. Les processus physiques comme les rayonnements, la turbulence, la friction et la formation des nuages régis par les équations de base, sont en raison de leurs petites échelles, décrits de manière statistique comme processus de paramétrisation.

Les modèles de prévisions numériques présentés dans les sections suivantes sont les plus couramment définis pour la prévision de l'ensoleillement global.

II.1. GFS

GFS est un modèle basé sur un système de prévision global. Il est calculé par le National Centers for Environmental Prediction (NCEP) un département du NOAA (Administration Nationale Océanique et Atmosphérique National), NWS (Service météorologique national). Depuis juillet 2010, ce modèle calcule des prévisions à 192 heures avec une résolution de 27km, puis des prévisions à 384 heures avec une résolution plus grossière. Les résultats du GFS actuellement disponibles sur le site du NOAA ont une résolution de 0.5 degré (environ 50km). Le modèle GFS est un modèle qui se compose de quatre modèles séparés (un modèle atmosphérique, un modèle océanique, un modèle terre/sol, un modèle de refroidissement de la mer) qui fonctionnent ensemble pour fournir une cartographie précise des conditions climatiques. Le modèle radiatif SW employé par GFS est le modèle rapide et précis de transfert radiatif RRTM-SW pour les applications climatiques. RRTM utilise la méthode de corrélation-k pour transformer la dépendance spectrale du coefficient d'absorption radiatif (k) en une fonction de distribution cumulative continue (Mlawer (1997)). La version actuelle de RRTM calcule les flux radiatifs atmosphériques et les taux de refroidissement pour la région spectrale de grande longueur d'onde (10 à 3000 cm^{-1}) pour une atmosphère arbitraire clair. Les espèces moléculaires traitées dans le modèle sont la vapeur d'eau, dioxyde de carbone, l'ozone, le méthane, l'oxyde nitreux et les hydrocarbures halogénés communs. En outre, la profondeur de nuage optique, l'albédo, le facteur d'asymétrie, et le rayon de particule contribuant aux propriétés radiatives de la couche sont traitées.

II.2. NAM : Modèle Méso-échelle Nord-Américain (North American Mesoscale Model)

Le NAM est un modèle régional de méso-échelle utilisant des processus avancés de paramétrisation des surfaces de terrain et des précipitations. Les prédictions NAM est une production du modèle nord-américain de recherches météorologiques et de prévision WRFNMM. Elles sont publiées par l'administration nationale océanique et atmosphérique (NOAA) qui est l'agence américaine responsable de l'étude de l'océan et de l'atmosphère sur 0.113° E-O par 0.111° N-S (approximativement 12 km x 12 km) sur l'entière grille continentale des Etats-Unis. Les données de sortie horaire, disponibles jusqu'à 36 heures à l'avance, sont publiées quatre fois dans la journée. La résolution étant bien plus haute que celle du GFS, donne en principe des prévisions à court terme plus

précises. Une douzaine de paramètres météorologiques sont disponibles sur les grilles NAM à partir de la température et de la précipitation ainsi que l'énergie cinétique turbulente. Dans ce modèle, ils prennent en compte entre autres, les effets de l'humidité dans l'atmosphère par les équations d'Euler, et ajoutent des équations en incluant quatre projections à la sphère : les équations de Lambert, stéréographique polaire, Mercator, et les projections latitude longitude (la force de Coriolis et les conditions de courbures interviennent pour les projections cartographiques). D'autres paramètres physiques sont inclus pour remanier les équations différentielles basées sur les mouvements de fluide et la chimie. Pour plus de détails voir Skamarock et al (2007). Le rayonnement est un paramètre de prévision qui est calculé à partir de plusieurs modèles de transfert radiatif : le modèle dynamique des fluides géophysiques (GFDL-SW) Lacis (1974), le modèle Dudhia utilisé également par MM5 ou le modèle Goddard.

II.3. Centre européen de prévisions à moyen terme (ECMWF)

L'ECMWF est un modèle de prévision globale. Les données d'ECMWF sont disponibles sur une grille de 0.25° E-O par 0.25° N-S. Le modèle ECMWF est constitué d'une composante dynamique, une composante physique et une composante couplée onde océan.

Le modèle ECMWF opérationnel contient le modèle radiatif McRad. La partie à ondes courtes de McRad est basée sur RRTM. Les constituants représentés dans McRad sont la vapeur d'eau, le dioxyde de carbone, l'ozone, le méthane, l'oxyde nitreux, les aérosols, et divers chlorofluorocarbures (CFC). Semblable au modèle GFDL-SW, les propriétés importantes de prévisions des nuages sont l'albédo et la profondeur optique

Les prévisions de ECMWF (centre européen de prévisions météorologiques à moyen terme) inclut le paramètre rayonnement comme une sortie direct du modèle et fournit des prévisions 10 jours à l'avance, mais dans une résolution spatiale maximale d'environ $60 \text{ km} \times 60 \text{ km}$, de sorte que les variations régionales, voire locales ne sont pas incluses. Pour résoudre cette pénurie, des modèles mésoéchelles peuvent par exemple être employés. Il y a également des services météorologiques qui prévoient le rayonnement solaire pour les prochains jours, comme le meteomedia AG, le plus grand service météorologique privé de l'Allemagne, qui y a inclus le MOS.

II.4. MOS

Les modèles de prévision basés sur les équations de la dynamique atmosphérique ne déterminent pas parfaitement les conditions météo proches de la terre. Des corrections statistiques ont été élaborées pour tenter de résoudre ce problème, ces modèles sont les MOS (statistiques de sortie de modèle), et ont été mis au point par le Service National Météorologique pour leur gamme de modèles de prévision météorologique en 1976. C'est une technique de post-traitement de sortie des modèles de prévisions

météo numériques qui utilise les statistiques de mesures météo locale historiques ou actuelles. Les modèles de statistiques pures sont excellents en "nowcasting" (prévisions à court-terme de météo locale) mais ils sont habituellement inutiles au-delà de six heures, car leur précision devient alors plus faible que celle des modèles physiques. La technique MOS combine la sortie de modèle et les statistiques, en utilisant les modèles numériques complexes se basant sur la physique de l'atmosphère afin de prévoir des situations météo à large échelle et ensuite en utilisant des équations de régression en post-traitement statistique pour clarifier des détails de météo de surface.

La précision est généralement meilleure qu'un modèle de statistique pur ou d'une sortie de modèle numérique pure. Une définition simplifiée du MOS peut être donnée par cette formule symbolique :

MOS = Modèle de prévisions + mesures locales+ statistiques locales

Les modèles statistiques ont été créés sur la base des champs tridimensionnels produits par les modèles numériques météorologiques, les observations de surface et les conditions climatologiques pour des emplacements spécifiques.

II.5. Aperçu des travaux NWP pour la prévision solaire

Heinemann et al. (2006) utilise des images satellite pour des horizons en dessous de 6 h; et montre une évaluation des prévisions par le modèle pour une période de 40 jours. Les images, dans la gamme du visible, du satellite géostationnaire METEOSAT sont utilisées comme des données en entrée pour la prévision. L'algorithme de prévision pour des durées à très court terme opère sur des images d'indice de nuage et est donc indépendant du modèle d'irradiance solaire journalier. Ceci permet de se concentrer sur le développement de structures de nuage. La composante déterministe journalière est ajoutée à l'aide d'un modèle ciel clair. La technique MOS basé pour l'irradiance solaire avec deux types de données en entrée (mesures au sol, mesures de Meteostat) a été introduite en utilisant le modèle ECMWF. Une comparaison des erreurs de ce modèle avec le modèle de persistance et le modèle Héliostat, a été effectuée. Ils trouvent une RMSE horaire de 32,1 % meilleure pour le modèle utilisant des données d'entrée de mesures au sol plutôt qu'avec les données de Météostat. Néanmoins, la meilleure performance est obtenue pour le modèle Héliostat avec une RMSE de 26%.

Bofinger and Heilscher (2006) ont utilisé le MOS localement avec ECMWF pour la prédiction du rayonnement. L'analyse de la RMSE présentent des valeurs de 58.8 Wh.m⁻² pour 24h de prévision et 76.8 Wh.m⁻² pour 120 h de prévision.

Perez et al. (2007) ont examiné l'exactitude de la base de données de prévisions numériques nationales (NDFD), un dérivé des modèles NWP publiés par le Centre National pour des prédictions environnementales (NCEP). Après l'application d'une fonction de correction locale Perez et al. (2007) ont constaté que, pour 8-26 heures d'horizons de prévision, le NDFD présentait une RMSE de 38%.

Remund et al. (2008) ont examiné les biais des NWP comparé à un seul site pour trouver que le modèle Centre Européen pour des prévisions de temps (ECMWF) et le modèle du Système de Prévision globale (GFS) obtenait des prévisions GHI (ensoleillement global) journalières avec une MBE de 19%. Cette MBE s'est révélée être environ constante pour des horizons de prévisions intra-journalières (une heure avant).

Breitkreuz et al. (2009) ont inclus les effets des aérosols sur le rayonnement dans le modèle ECMWF et montre ainsi une réduction de la RMSE de 11,5% à 7,2% dans le GHI horaire pour des horizons de prédictions de 2 à 3 jours sous des conditions de ciel clair. De même, la MBE a été améliorée de 9,8% à 5,1%.

Lorenz et al. (2009) ont relié angle solaire zénithal (SZA) et l'indice ciel clair k_c (clear sky index) à ECMWF, montrant une surestimation allant jusqu'à 100 W.m^{-2} pour des conditions de ciel nuageux intermédiaire. L'utilisation d'une fonction de correction MOS a éliminé le biais et réduit la RMSE pour des prévisions horaires de 5% pour 24h de prévisions. Pour un horizon d'une journée, ils trouvent une RMSE de 36% pour des prévisions du rayonnement sur des sites individuels en Allemagne. La précision des prévisions augmente avec la taille de la région (pour la zone complète de l'Allemagne, la RMSE s'élève à 13%).

Mathiesen et al (2011) ont utilisé les méthodes NWP à travers les modèles NAM, GFS et ECMWF avec une fonction de correction de biais par la technique MOS, pour la prédiction horaire de l'ensoleillement global. Les données horaires d'ensoleillement global provenant du réseau SURFRAD aux Etats-Unis ont été utilisées. Le modèle GFS corrigé par la méthode MOS fournit la meilleure valeur de RMSE d'environ 85 W.m^{-2} , suivi du modèle ECMWF et NAM. ECMWF effectue une prévision plus précise pour des journées présentant des conditions nuageuses, alors que pour des journées présentant un ciel clair c'est le modèle GFS.

La performance des modèles NWP comme source de données de prévision pour le rayonnement solaire est très variable, de 20% RMSE à 60% pour des prévisions à un jour à l'avance.

II.6. Limites des modèles NWP

Le modèle ECMWF fournit les prévisions jusqu'à 10 jours à venir, mais dans une résolution spatiale d'environ $60 \text{ km} \times 60 \text{ km}$. Le modèle de transfert radiatif dans la plupart des NWP produisent des données de sortie horaires (NAM) ou chaque 3heures (GFS et ECMWF) ou 4 heures (HRRR modèle : « High Resolution Rapid Refresh »). Les données de sorties ne sont pas réglées pour une prévision à une localisation particulière mais s'appliquent à de grandes cellules de la grille.

Les modèles NWP fournissent des prévisions pour un nombre de variables utiles pour la modélisation du rayonnement solaire avec des horizons de prévisions de 48 à 180 heures ou plus. La résolution spatiale de ces modèles est variable et en moyenne chaque cellule de la grille de prévisions est sur une échelle proche d'une centaine de kilomètre carré. Cependant, il y a une recherche active sur les

modèles à haute résolution avec des zones de cellules de moins de 10 km² (Benjamin et al, 2010; Lara-Fanego et al, 2011). L'imagerie satellite permet de fournir des prévisions précises jusqu'à 6 h (Perez et al, 2010; Schroedter-Homscheidt et al, 2009.) à une résolution de 1 km².

En raison du couplage spatio-temporel, l'utilisation des NWP est optimale pour des échelles de temps de prévision de 6 heures et plus longues. Les NWP à faible résolution, comme GEM, pourraient être utilisés pour produire des prévisions à des échelles de temps plus grandes que les NWP à haute résolution, par exemple HRRR ayant une méthode de résolution spatiale supérieure devrait être appliquées pour assurer une meilleure précision à des échelles temporelles plus fines (Kostylev 2011).

De plus, dans les modèles de prévision météorologique numérique existant, le rayonnement solaire est seulement utilisé pour l'équilibre d'énergie moyenne de surface, ce qui rend la variabilité du rayonnement global horizontal bien moins pertinente.

III. Imagerie des nuages

Les techniques de prévision numériques du temps n'ont pas la résolution spatiale et temporelle pour fournir des informations sur les fluctuations du rayonnement solaire à haute fréquence. Une autre solution est fournie grâce à l'imagerie au sol des conditions météorologiques locales. En appliquant les systèmes de satellite ou d'imageur au sol, les nuages peuvent être détectés, caractérisés, pour prédire avec précision le rayonnement solaire global à très court terme.

III.1. Imagerie au sol des nuages

Similairement à la prévision par satellite, l'approche de l'utilisation de modèles « sky images » (imageur du ciel) fournit la détection en temps réel des nuages et a la capacité de prédire le mouvement des nuages et leur position, dans un avenir proche. Dans Chi Wai Chow et al (2011), les images sont prises toutes les 30s lorsque le soleil est au-dessus d'une élévation d'angle de 3°. La caméra fournit des images qui ont une précision de 640 x 480 pixels et le miroir occupe 420 par 420 pixels. Pour déterminer la couverture du ciel par la caméra, des masques d'image sont générés, les bords des masques sont identifiés, et les valeurs de pixels de la région de bord sont utilisés pour interpoler l'intérieur de la région. Les imageurs du ciel fournissent une mesure visuelle de l'ensemble de la voûte céleste avec une haute résolution spatiale et temporelle.

Pour produire les séries de temps de rayonnement global horizontal à partir des stations de mesures au sol, une technique de rayonnement binaire (absence/présence) est utilisée. La présence de nuages à un pixel particulier est identifié en utilisant divers algorithmes de seuillage (Martins et al 2010, Ferreira et al 2010). Les vecteurs de mouvement des nuages sont générés par corrélation croisée d'images du ciel consécutives et utilisés pour prédire l'emplacement des nuages peu de temps avant, en fonction de la vitesse de déplacement des nuages. Une fois la structure nuageuse déterminée, et un modèle ciel clair

utilisé, la variation journalière du rayonnement est ajoutée. Plusieurs modèles ciel clair sont utilisés dans la littérature des méthodes d'imagerie du ciel, tel que le modèle développé par Ineichen et Perez (2002) utilisé par Chi Wai Chow et al (2011), le modèle European Solar Radiation Atlas (ESRA) utilisée par Rigollier et al.(2004) pour Heliostat 2, le modèle Solis développé par Mueller et al. (2004) pour le projet européen Heliosat-3.

III.2. Imagerie par satellite des nuages

Les satellites sont utilisés pour la détermination et la prévision du rayonnement solaire local. La base de cette approche est la même que celle de l'imagerie du ciel au sol qui consiste à déterminer la structure nuageuse par rapport à la structure enregistrée les temps précédents. L'extrapolation du mouvement des nuages mène à la prédiction de la position des nuages et donc à la situation locale du rayonnement solaire. De manière analogue à la méthode précédente, les données satellite détectent la structure des mouvements de nuages en déterminant un vecteur champ. Le rayonnement pour toutes les conditions du ciel, y compris le ciel nuageux, peut être obtenu en utilisant des modèles de transfert radiatif RTM, nécessitant en entrée la structure verticale des paramètres physiques des nuages, par exemple, les nuages et la teneur en eau de la glace ou le rayon des gouttelettes.

Par rapport aux données de satellite, les imageurs du ciel au sol offrent une résolution spatiale et temporelle beaucoup plus élevée, y compris la possibilité de capturer les changements brusques sur une échelle temporelle de moins d'1min.

III.3. Aperçu des travaux d'imagerie des nuages

Heinemann et al (1999) ont décrit l'application d'une méthode statistique pour détecter le mouvement des structures nuageuses à partir d'images satellite. En extrapolant le développement temporel de la situation nuageuse, le rayonnement solaire peut être prévue pour des échelles de temps de 30 min à 2 h. Les prévisions sont évaluées et un exemple d'algorithme de prévision pour prédire la puissance PV est présenté.

Hammer et al (1999) ont démontré que 30 minutes de prévision GHI à partir de la technique d'images par satellites donne des résultats de prévisions avec 17% de RMSE et pour 2h d'horizon de prédiction 30% RMSE. Pour des prévisions inférieures à la journée, ils constatent une réduction de la RMSE par 7 à 10 % comparé aux prévisions avec le modèle de persistance.

Rigollier et al (2004) présentent la méthode Heliosat-2 qui convertit les observations faites par les satellites météorologiques géostationnaires par des estimations du rayonnement global au niveau du sol. La méthode Heliosat-2 est appliquée à des images Météosat de l'Europe pour les mois de Janvier 1995, avril 1995 et juillet 1994 des mesures pyrométriques effectuées par trente-cinq stations

météorologiques sont utilisés pour évaluer les performances. Il trouve pour l'estimation du rayonnement global horaire une valeur minimale de RMSE de 18% pour le mois de juillet.

De plus dans Gueymard et al (2011) le rayonnement global mensuel résultant de la méthode d'images satellite montre une RMSE de 5,25%.

Perez et al. 2010 ont prédit l'ensoleillement global horaire jusqu'à un horizon de 6 h en utilisant les images satellites géostationnaires. Les auteurs ont utilisés 1 année de données horaires d'ensoleillement global provenant du réseau SURFRAD aux Etats-Unis. Les prévisions de l'ensoleillement à court terme sont produites en utilisant deux indices d'images d'ensoleillement global consécutives obtenues par satellite (Perez et al, 2002, 2004) à partir duquel le mouvement du nuage spécifique à un pixel est déterminé.

La méthodologie permettant de déterminer le mouvement nuageux localisé est basé par le calcul de la RMSE de la différence entre deux grilles consécutives d'indice de clarté K_t entourant le pixel considéré quand la seconde grille est advectée dans la direction d'un vecteur de mouvement. Le vecteur de mouvement sélectionné correspond à la RMSE la plus faible (méthode basée sur Lorenz et al. (2007)). Ce processus est répété pour chaque pixel d'image, et chaque pixel se voit attribuer un vecteur de mouvement individuel. Les images futures sont obtenues en déplaçant les pixels de l'image actuelle dans la direction de leur vecteur de mouvement. La méthode de prévision de l'ensoleillement est basée sur la relation expérimentale entre la couverture nuageuse et l'ensoleillement global mesuré. Les algorithmes de prévision sont validés en fonction des mesures au sol de 7 localisations aux Etats-Unis sur 1 année. Une comparaison de performance régionale est présentée. Une RMSE métrique annuelle entre 80 et 200 $W.m^{-2}$ ont été obtenus.

Chow et al. (2011) ont trouvé une réduction de 50 à 60% des erreurs de prévisions basée sur la méthode TSI (total sky imager) par rapport à la persistance pour une horizon de 30 secondes de prévisions. Les erreurs de prévisions des nuages augmentent avec l'horizon de prédiction à cause d'importante variabilité de la couverture nuageuse sur les sites côtiers. L'erreur moyenne de la "matching error" (erreur entre la carte de nuage prédite et celle observée) présente un taux de 6% à 30% pour 30 secondes à 5 minutes d'horizon.

Ferreira et al. (2012) ont créé leur propre imageur de ciel portable. Un ensemble de modèles de réseaux de neurones sont identifiés par une méthode d'optimisation, afin de fournir des fonctionnalités prédictives au capteur. Les ANNs sont identifiés par la méthode MOGA pour plus de détails voir Ferreira et al (2012). Le rayonnement global (GHI) est échantillonné à 5 minutes. L'objectif est de prédire le rayonnement global jusqu'à un horizon de prédiction de 4h. Pour un horizon de prédiction de 5 minutes ils trouvent 15,9 $W.m^{-2}$ de MAE et 8,8 $W.m^{-2}$ pour un horizon de 4 h pour leur modèle NAR.

Marquez et Coimbra (2013) décrit une méthode de traitement d'image (total Sky Imager :STI) pour générer des prévisions à court terme du rayonnement normal direct au niveau du sol. Ils utilisent des données de DNI moyennées à 1min des horizons de prévisions variant de 3 à 15 min. Ils trouvent une RMSE minimale de 213 W.m⁻² au mois de juin pour un horizon de prédiction de 4 minutes. Les erreurs augmentent avec les horizons de prévisions allant jusqu'à 496 W.m⁻² de RMSE pour 15 minutes.

Quesada et al (2014) utilise une méthode TSI (l'imagerie du ciel total) pour une prévision intra-horaire du DNI (rayonnement normal direct) jusqu'à un horizon de prédiction de 20 minutes. Les données utilisées sont les DNI moyennées à 1min mais mesurées à l'origine à 30 s. Les images ont été prises toutes les 20 secondes. Dans des conditions de ciel intermédiaire nuageux, et ciel nuageux, le modèle donne particulièrement de bons résultats par rapport à le modèle de persistance et la méthode PIV présentée par Marquez et Coimbra (2013). Cependant, pour les journées à ciel clair, le modèle de persistance présente une meilleure performance prédictive. Ils trouvent une RMSE en moyenne de 105,26 W.m⁻² pour tous types de journées confondus sur un ensemble de 15 jours.

Marquez et al (2013) utilise des indices de nuages obtenus à partir du TSI intégrant un algorithme de classification des nuages et des indices de nuages dérivés des mesures radiométriques infrarouges pour apporter des améliorations sur les résultats de prévisions du rayonnement global à un horizon de prédiction d'une heure. Un algorithme de réseau de neurones artificiels est employé pour améliorer la modélisation quantitative de l'indice de couverture céleste du rayonnement global. Ils trouvent une RMSE minimale de 35,9 W.m⁻² pour un horizon de prédiction horaire.

III.4. Limites des modèles d'imagerie des nuages

Le succès des imageurs du ciel dépendent de la précision de l'algorithme de détection des nuages et l'exactitude des prévisions bidimensionnelles des nuages. Néanmoins, cette méthode est contestée par le manque d'informations sur la structure à 3 dimensions des nuages et des dynamiques multicouches des nuages observés. La perspective de la couverture nuageuse résultant de la dimensionnalité des images peut conduire à une perception diminuée de la taille des nuages et peut également créer l'illusion qu'il existe un seul nuage tandis qu'il y a deux nuages à proximité l'un de l'autre, mais distincts. Un réseau d'imageurs du ciel avec différents points de vue pourrait réduire cette erreur de prévision (Chow et al 2011). La déformation des nuages, l'hétérogénéité de la vitesse des nuages à différentes altitudes sont également les principales sources d'erreur dans cette approche. La vitesse du nuage calculée à partir de la méthode CCM est représentative de la vitesse de tous les nuages dans l'image. Dans de rares cas, le cumulus à l'est de l'imageur de ciel peut se déplacer à l'est à une vitesse différente de celle des nuages stratiformes à l'ouest alors que la prévision suppose une vitesse de nuage homogène (environ la vitesse moyenne pondérée des deux couches de nuages), ce qui peut donner de grandes erreurs de prévision. La plupart des algorithmes de classification des nuages sont

basés sur la méthode binaire (présence/absence) ce qui ne différencie pas le type de nuage (opaque ou fin). Quelques travaux ont essayé d'améliorer ces problèmes tels que Ghonima et al. (2012) qui ont développé un algorithme basé sur la comparaison d'un ratio de pixels et Calbó et Sabburg (2010) qui ont utilisé des images de ciel comme entrée pour une classification automatique qui distingue différents types de conditions climatiques (clair, nuages stratiformes, nuages cumuliformes ...).

En définitive, la méthodologie de l'imageur total du ciel (TSI) est la plus réussie dans le suivi d'une seule couche de nuages se déplaçant à travers le ciel sans déformation rapide. Par conséquent, les conditions météorologiques intermédiaires sont plus difficiles à prédire avec les modèles basés sur TSI. Des améliorations telles que la prise en compte des informations sur le déplacement des nuages mais aussi sur la formation et la dissolution des nuages devraient être apportées (Hammer et al 1999). Le rayonnement solaire peut conduire à des convections et à l'évaporation des nuages pour de courte échelle de temps. Ce qui reste un paramètre complexe à introduire dans un modèle déterministe ou un algorithme. De plus, les limitations de cette approche sont également dues à la qualité des images (mauvais enregistrement et mauvaise performance du sky imager) Kostylev et al (2011).

IV. Modèles statistiques

IV.1. Modèle de persistance

Le modèle de persistance est l'un des modèles de prévision les plus simples à mettre en œuvre et est souvent utilisé comme modèle de référence pour la performance prédictive d'autres méthodes de prévision plus avancées. C'est un modèle de référence trivial particulièrement pour la prévision à court terme. En effet, le modèle de persistance est souvent une technique imprécise pour plus d'1h avant la prévision. La persistance suppose que le rayonnement solaire à chaque temps $t+1$ est mieux prédite par sa valeur au temps t :

$$\hat{X}_{t+1} = X_t \quad (2.1)$$

Un modèle de persistance à partir de l'indice de clarté est défini dans Inman et al (2013) (smart persistence) :

Soit :

$$k_{t+\Delta t} = k_t = \frac{I_t}{I_t^{clr}} \quad (2.2)$$

Avec I_t^{clr} le rayonnement solaire défini selon un modèle ciel clair (Kasten, Solis, Ineichen..).

$$K_{t+\Delta t} = K_t = \frac{I_t}{I_t^{EX}} \quad (2.3)$$

Avec I_t^{EX} le rayonnement solaire extraterrestre.

La prévision à partir du modèle de persistance devient :

$$I_{t+\Delta t}^{pers} = k_t I_{t+\Delta t}^{clr} \quad (2.4)$$

Il peut donc avoir différents modèles de persistance dépendant de comment I_t^{EX} (rayonnement solaire extraterrestre et I_t^{clr} (rayonnement solaire défini selon un modèle ciel clair) sont estimés.

L'exactitude des prévisions de la persistance diminue fortement avec l'horizon de prévision, particulièrement à cause des variations brutales et à court terme du rayonnement solaire due à la nébulosité ce qui deuxièmement établit un retard évident.

IV.2. Modèles linéaires

IV.2.1 Modèles auto régressifs

Les modèles AR, ARMA, ARIMA sont des catégories de modèles qui ont été popularisées et formalisées par Box et Jenkins (1976). Les processus autorégressifs supposent que chaque point peut être prédit par la somme pondérée d'un ensemble de points précédents, plus un terme aléatoire d'erreur.

IV.2.1.1 Modèles AR, MA

Modèle AR

Le terme AR est associé aux coefficients d'auto-régression. Les modèles autorégressifs supposent qu'une valeur à l'instant t est une fonction linéaire des valeurs précédentes. Chaque observation est constituée d'une composante aléatoire et d'une combinaison linéaire des observations précédentes (autrement dit chaque terme de la combinaison linéaire est composé du produit des observations linéaires avec leur coefficient d'auto-régression correspondant, plus une composante aléatoire, plus la constante du modèle). Les coefficients AR peuvent être estimés par une analyse en régression multiple.

Selon Tsay (2005) l'équation du modèle AR est définie par :

$$R_t = \phi_0 + \phi_1 R_{t-1} + \phi_2 R_{t-2} + \dots + \phi_p R_{t-p} + e_t \quad (2.5)$$

Avec R_t un processus linéaire avec $t \in Z$, p le nombre de termes auto-régressifs, ϕ_i sont des coefficients et e_t est supposé être un bruit blanc avec une moyenne nulle de variance σ^2 .

Modèle MA

Le terme MA fait référence à un modèle de moyenne mobile.

Un modèle de moyenne mobile correspond à des séries avec des fluctuations aléatoires autour d'une moyenne variant lentement. Au lieu de prendre la valeur précédente comme prédicteur, on utilise une moyenne de quelques observations précédentes, afin d'éliminer le bruit, et estimer plus précisément la moyenne locale. Cette méthode fait référence au lissage exponentiel simple, qui considère chaque observation comme la résultante d'une constante et d'un terme d'erreur.

$$R_t = \sum_{i=0}^q \theta_i \cdot e_t \quad \forall t \in Z \quad (q \text{ est le nombre de moyennes mobiles}) \quad (2.6)$$

La notation réduite faisant intervenir l'opérateur retard L transforme les équations précédentes :

$$L^i \cdot R_t = R_{t-i} \quad (2.7)$$

$$R_t = \theta(L)\epsilon(t) ; \quad \theta(L) = \sum_{i=0}^q \theta^i L^i \quad (2.8)$$

$$\phi(L) R_t = \epsilon(t) ;$$

$$\phi(L) = 1 - \sum_{i=1}^p \phi_i L^i \quad (2.9)$$

IV.2.1.2 Modèles ARMA, ARIMA

Modèle ARMA

Le modèle ARMA regroupe AR(p) et MA(q), c'est un processus stationnaire vérifiant la relation suivante :

$$\phi(L) R_t = \theta(L)\epsilon_t \quad (2.10)$$

Une notation plus explicite du modèle ARMA peut être définie tel que \tilde{z}_t est la série de temps pour laquelle la moyenne μ est retirée et $\phi(q)$ et $\theta(q)$ sont les poids :

$$\theta(q)\phi(q)\tilde{z}_t = \theta(q)\omega_t = \tilde{z}_t \quad (2.11)$$

avec la méthode MA comme somme finie de n chocs précédents $\omega_t, \omega_{t-1}, \omega_{t-2}, \dots, \omega_{t-n}$

L'équation précédente implique que :

$$\theta(q)\phi(q) = 1 \quad (2.12)$$

$$\phi^{-1}(q) = \theta(q) \quad (2.13)$$

Le processus MA(n), $\tilde{z}_t = \theta_n(q)\omega_t$, peut être écrit comme processus AR infini :

$$\tilde{z}_t = \omega_t - \theta_1\tilde{z}_{t-1} - \theta_1^2\tilde{z}_{t-2} - \dots \quad (2.14)$$

En pratique, pour réaliser une paramétrisation parcimonieuse avec les deux termes AR et MA, on obtient l'équation suivante :

$$\tilde{z}_t + \phi_1\tilde{z}_{t-1} + \dots + \phi_m\tilde{z}_{t-m} = \omega_t + \theta_1\omega_{t-1} + \dots + \theta_n\omega_{t-n} \quad (2.15)$$

$$\text{où } \phi_m(q)\tilde{z}_t = \theta_n(q)\omega_t \quad (2.16)$$

Cette dernière équation se réfère au modèle ARMA processus d'ordre (m,n) mixant les AR et MA modèles. Néanmoins le modèle Arma peut également se définir par l'équation suivante :

$$\tilde{z}_t = \frac{\theta_n(q)}{\phi_m(q)}\omega_t = \frac{1+\theta_1q^{-1}+\dots+\theta_nq^{-n}}{1+\phi_1q^{-1}+\dots+\phi_mq^{-m}} \quad (2.17)$$

\tilde{z}_t est donc la sortie d'un filtre linéaire soit une fonction de transfert qui est le rapport de deux polynômes $\theta_n(q)$ et $\phi_m(q)$ avec un bruit blanc ω_t comme entrée.

Un modèle ARMA typique se compose de trois étapes: l'identification de l'ordre, l'estimation des paramètres et la prévision. Parmi ces trois étapes, l'identification qui détermine l'ordre de la partie AR et MA est très importante. Cette étape requiert des informations statistiques telles que l'autocorrélation et l'autocorrélation partielle. La difficulté d'estimer l'ordre et les paramètres reste un domaine actif dans la recherche (Rojasa et al 2008). Les modèles ARIMA supposent que l'on travaille sur les séries stationnaires (moyenne de la série constante, variance constante). La meilleure méthode pour éliminer toute tendance est de différencier, c'est-à-dire de remplacer la série originale par la série des différences adjacentes. Une série temporelle qui a besoin d'être différenciée pour atteindre la stationnarité est considérée comme une version intégrée d'une série stationnaire (d'où le terme Integrated).

Le processus d'intégration suppose que chaque point présente une différence constante avec le point précédent. Les processus de moyenne mobile supposent que chaque point est fonction des erreurs entachant les points précédents, plus sa propre erreur.

Modèle ARIMA

Un modèle ARIMA est dit comme modèle ARIMA(p,d,q), dans lequel :

p est le nombre de termes auto-régressifs

d est le nombre de différences

q est le nombre de moyennes mobiles.

Une différenciation d'ordre 1 suppose que la différence entre deux valeurs successives de y est constante. Les modèles d'ordre 2 travaillent non plus sur les différences brutes, mais sur les différences de différence.

Soit le modèle ARMA :

$$\phi_m(q) \tilde{z}_t = \theta_n(q) \omega_t \quad (2.18)$$

On peut exprimer le modèle tel que :

$$\phi_m(q) \tilde{z}_t = \theta_n(q) (1 - q^{-1})^d \tilde{z}_t = \theta_n(q) \omega_t \quad (2.19)$$

Où $\phi_m(q)$ est un opérateur AR(m) stationnaire.

Puisque, $S^d \tilde{z}_t = S^d z_t$ quand $d \geq 1$ nous pouvons écrire :

$$\phi_m(q) S^d z_t = \theta_n(q) \omega_t \quad (2.20)$$

On définit $y_t = S^d z_t$, ainsi l'expression du modèle devient

$$\phi_m(q) y_t = \theta_n(q) \omega_t \quad (2.21)$$

Comme la fonction de transfert AR, $\phi_m(q)$, est d'ordre m , d le pas de différenciation pris et la fonction de transfert MA, $\theta_n(q)$, est d'ordre n , le processus est un modèle ARIMA(m,d,n). En pratique d est égale à 0, 1, au plus 2. Le modèle ARIMA(m, d, n) vérifie une relation récurrente avec le processus différencié d et équivaut à représenter z_t comme la sortie d'un filtre linéaire avec la fonction de transfert $\phi_n^{-1} S^d \theta_n$ et prend un bruit blanc ω_t comme entrée.

Les paramètres des modèles peuvent être estimés à partir de différentes méthodes telles que les équations de Yule Walker (estimation de la partie AR) et la méthode du maximum de vraisemblance, méthodologie des moindres carrés ordinaires MCO (pour la partie AM).

IV.2.2 Les modèles autorégressifs avec variables exogènes

IV.2.2.1 Modèle ARMAX

Les modèles autorégressifs précités sont univariés et n'utilisent que les valeurs passées de la série temporelle. Néanmoins, les modèles ARMA(m,n) peuvent être améliorés en incluant une information

extérieure aux séries de temps. En effet, en incluant dans le modèle des informations concernant l'évolution de température locale, d'humidité relative, de couverture nuageuse, la vitesse et la direction du vent... les erreurs de prévisions du rayonnement solaire peuvent être réduites. Ces variables indépendantes du modèle mais influençant les valeurs de rayonnement sont référées à des variables exogènes. Le modèle ARMA(m,n) avec p termes exogènes est intitulé le modèle ARMAX :

$$\tilde{z}_t + \phi_1 \tilde{z}_{t-1} + \dots + \phi_m \tilde{z}_{t-m} = \omega_t + \theta_1 \omega_{t-1} + \dots + \theta_n \omega_{t-n} + \lambda_1 e_{t-1} + \dots + \lambda_p e_{t-p} \quad (2.22)$$

Le modèle ci-dessus contient le modèle AR(m) et MA(n) mais également les p valeurs précédentes de séries de temps exogènes. L'opérateur d'entrée exogène est défini tel que :

$$\Lambda_p(q) = \sum_{k=1}^p \lambda_k q^{-k} = \lambda_1 q^{-1} + \lambda_2 q^{-2} + \dots + \lambda_p q^{-p} \quad (2.23)$$

Ainsi, le modèle ARMAX (m, n, p) peut être décrit conventionnellement par :

$$\phi(q) \tilde{z}_t = \frac{\theta(q)}{\psi(B)} \omega_t + \frac{\Lambda(q)}{\Xi(q)} \quad (2.24)$$

Où $\theta(q)$, $\phi(q)$, $\Lambda(q)$ et $\Xi(q)$ sont des polynômes dépendants de q .

IV.2.2.2 Modèle ARIMAX

Par méthode analogue, les p valeurs précédentes de séries de temps exogènes peuvent être aussi introduites dans le processus ARIMA(m,n,p) pour le modèle ARIMAX d'ordre (m,d, n, p) :

$$\tilde{z}_t = \phi_1 S^d z_{t-1} + \dots + \phi_m S^d z_{t-m} + \omega_t + \theta_1 \omega_{t-1} + \dots + \theta_n \omega_{t-n} + \lambda_1 e_{t-1} + \dots + \lambda_p e_{t-p} \quad (2.25)$$

Avec l'opérateur $y_t = S^d z_{t-1}$ terms of the backwards shift operator, on peut définir le modèle ARIMAX de manière plus compacte tel que :

$$\phi_m(q) y_t = \theta_n(q) \omega_t + \Lambda_p(q) e_t \quad (2.26)$$

D'autres modèles autorégressifs non cités dans cette section existent dans la littérature telle que le modèle ARARX (Auto Régressif-ARX), et le modèle ARARMAX (Auto-Régressif ARMAX).

IV.2.3 Aperçu des travaux pour les modèles autorégressifs

Ces techniques ont été utilisées depuis les années 1970, notamment Boileau et Guerrier (1979) ont évalué plusieurs modèles stochastiques dans le domaine de la météorologie solaire, Bernard et al (1978) ont utilisé des techniques stochastiques saisonnières et non-saisonnières pour la prédiction du rayonnement solaire.

Kärner et al (2002) ont effectué une analyse ARIMA pour les anomalies troposphériques et stratosphérique de température par satellite et des ensembles de données du rayonnement solaire. Les résultats ont souligné un rôle prépondérant de la variabilité du rayonnement solaire sur les variations de la température de la troposphère.

Dans Santos et al (2003) présente une méthodologie pour calculer les valeurs synthétiques du rayonnement solaire journalier et décrit comment des séquences de rayonnement global journalier peuvent être générées en utilisant le rayonnement moyen mensuel. Un modèle stochastique ARIMA (1,1,1) est présenté. Les séquences de l'indice de clarté journalier sont estimés par le modèle ARIMA(1,1,1).

Bacher et al. (2009) ont utilisé les modèles AR, ARX et RX (modèle autorégressif avec des variables non endogènes) pour prédire la production PV horaire en utilisant 1 an de données de PV à 15 minutes sur 21 toits. Les variables exogènes sont l'ensoleillement global prédit à partir du modèle HIRLAM NWP mésoéchelle provenant de l'Institut météorologique danois.

Reikard (2009) a comparé les modèles ARIMA avec d'autres méthodes (les fonctions de transfert, les réseaux de neurones et les modèles hybrides), pour six ensembles de données à des résolutions temporelles de 5, 15, 30 et 60 min en utilisant la composante horizontale du rayonnement solaire global. Reikard a constaté que, dans presque tous les tests, les meilleurs résultats sont obtenus en utilisant la méthode ARIMA avec une MAE minimale de 19,68% ; 18,65% ; 18,97% ; 13,21% pour respectivement des horizons de prévision de 1h, 30 min, 15 min et 5min. Les données sont mesurées par pas de temps de 1 minute.

Wu et Chan (2011) ont utilisé ARMA(1,1) pour prédire le rayonnement solaire horaire pour un horizon de 10 minutes. Nous pouvons voir que le modèle ARMA peut généralement prédire la tendance de la série solaire cependant il constate un certain «retard» dans la prédiction. Ils ont finalement utilisé le modèle hybride ARMA et TDNN que nous détaillerons dans la section des modèles hybrides.

Dans Voyant et al (2012) le modèle ARMA(1,0) a été utilisé pour la prédiction de rayonnement solaire global horaire pour plusieurs localisations sur la Méditerranée avec la meilleure moyenne annuelle de RMSE normalisée égale à 19.4 % sur le site de Ajaccio et de 16% sur Marseille dans un autre article de Voyant et al (2012). Dans ces articles, les chercheurs ont également préféré utilisé un modèle hybride qui obtenait de meilleurs résultats dont nous expliquerons brièvement la méthode dans la section des modèles hybrides.

Huang et al (2013) ont utilisé le modèle AR(2) pour prédire le rayonnement solaire horaire et a montré que ce modèle travaille particulièrement bien quand les résidus du rayonnement solaire est

décroissante. L'erreur statistique n'est pas précisée. Ils ont opté finalement pour un modèle hybride nommé CARDS que nous détaillerons dans la section des modèles hybrides.

Yang et al (2012) ont étudié trois approches basées sur le modèle ARIMA. La première méthode utilise directement des données d'ensoleillement global pour prédire à un horizon d'1h l'ensoleillement global en soustrayant la composante saisonnière. Les résidus de cette soustraction correspondant à une composante plus irrégulière, sont utilisés pour la prédiction à l'heure suivante en appliquant le modèle ARIMA. La seconde méthode décompose l'ensoleillement global en rayonnement diffus horizontal et rayonnement direct normalisé. Ces deux paramètres sont utilisés comme entrée dans l'application du modèle ARIMA séparément. L'ensoleillement global prédit est obtenu donc par la combinaison du rayonnement diffus prédits et du rayonnement direct normalisé prédit. La troisième méthode considère les effets de couverture nuageuse. Un modèle ARIMA est utilisé pour prédire les nuages transitoires. L'ensoleillement global à différents angles zénith et sous différentes conditions de couvertures nuageuses est construit en utilisant différents modèles de régression linéaire dépendante des différentes couvertures nuageuses. Ils obtiennent la prévision horaire de l'ensoleillement global en substituant l'angle zénith à cette heure précise, qui est un paramètre déterministe, dans la ligne de régression correspondante. La troisième méthode présente des résultats plus performants avec une RMSE moyenne de 32,80 W.m⁻² et 29,73 W.m⁻² comparé aux autres modèles approchant la centaine de W.m⁻² (102,69 à 116,85 W.m⁻²).

IV.2.4 Limites des modèles autorégressifs

Le succès du modèle ARIMA est principalement attribuable à sa capacité à capturer le cycle diurne de manière plus efficace que les autres méthodes. (Reikard 2009)

Comme mentionné précédemment, le modèle ARIMA est un processus stochastique couplant l'autorégression (AR) et la moyenne glissante (MA), après une différenciation appropriée pour retirer la tendance de la série temporelle. De cette manière, le modèle ARIMA a la capacité de traiter des séries non stationnaires. C'est un cas général tandis que le modèle ARMA est plus particulier car il ne traite que les séries stationnaires (Carlo gaetan et al 2011). Cependant, afin de traiter ce type de séries, il faut déterminer le niveau d'intégration c'est-à-dire le nombre de fois qu'il faut différencier la série avant de la rendre stationnaire.

De plus, le modèle ARIMA ou ARMA ne peut traiter plus d'une variable donc ne peut travailler sur des séries à plusieurs variables.

Certains de ces modèles autorégressifs fournissent de bons résultats dans des conditions climatiques normales, mais pas aussi efficaces que dans les situations météorologiques extrêmes. En effet, la série de temps doit être linéaire et stationnaire, mais les données de séries chronologiques de l'ensoleillement sont non linéaires et non stationnaires. Ceci peut conduire à de grandes valeurs d'erreurs de prévision particulièrement pour des situations météorologiques extrêmes. Par exemple, le

modèle ARMA qui est considéré comme un modèle stable peut produire de grandes erreurs pour des journées à ciel nuageux présentant une variabilité importante (Wu et Chan 2011). Huang et al (2013) ont utilisé le modèle AR (2) pour prédire le rayonnement solaire horaire et ont montré que quand le rayonnement désaisonnalisé a tendance à diminuer, AR(2) fournit des prédictions correctes mais pour les hauts pics AR(2) a tendance à sous-estimer les données.

Dans la littérature, différentes méthodes hybrides ARMA ont été proposées pour la prévision. Ainsi, de nombreuses combinaisons de modèles avec ARMA ont été étudiées telles que la combinaison du modèle ARMA et TDNN (modèle basé sur les réseaux de neurones) (Wu and Chan 2011), la combinaison AR(2) et Lucheroni modèle (Jing Huang et al 2012). Dans ces articles les modèles de combinaison ont présenté des résultats corrects pour une prédiction du rayonnement solaire à l'échelle d'une heure. L'idée de base de la combinaison de modèles de prévision est d'utiliser la fonction unique de chaque modèle pour capturer différentes caractéristiques dans les données.

IV.3. Modèles non-linéaires

IV.3.1 Réseau de Neurones Artificiels (ANNs)

D'autres modèles statistiques de prévision sont basés sur les réseaux de neurones artificiels). Les réseaux de neurones (NNs) ont été utilisés pour pallier aux limitations des approches classiques et linéaires et permettent de résoudre des problèmes complexes non linéaires.

Dans une structure NNs classique, chaque entrée (échantillon de données) est multipliée par un poids de connexion. Ensuite, les produits sont additionnés et passés à travers une fonction de transfert pour générer une sortie. La fonction de transfert est une équation algébrique qui peut être linéaire ou non linéaire, on peut citer les fonctions sigmoïdes (unipolaire et bipolaire), les fonctions logistiques (unipolaires et bipolaires), les fonctions à bases radiales gaussiennes. Chaque neurone est pondéré et interconnecté. Les neurones se développent en trois différentes couches : la couche d'entrée, la couche de sortie et une ou plusieurs couches cachées. Dans les modèles « feed-forward » (réseau de neurones non bouclé), chaque neurone de la couche cachée et de la sortie sont activés par une fonction non linéaire (fonction d'activation) qui dépend de la somme pondérée de ses entrées et un paramètre spécifique des neurones. La fonction d'activation modifie le signal en conséquence, et transmet le signal au neurone suivant. En général dans ce type de modèle, les fonctions d'activation sont de types logistiques ou tangentes hyperboliques (Sfetsos, 2000). La phase d'apprentissage correspond à la mise à jour et à la détermination des poids de connexion. Il existe différents types de connexion et algorithmes d'apprentissage. Un algorithme populaire et simple est l'algorithme de rétro-propagation qui est constitué de deux phases: la phase d'apprentissage et la phase de rappel. Avant la phase d'apprentissage, les poids du réseau sont initialisés de manière aléatoire. Ensuite, la sortie du réseau est calculée et comparée à la valeur désirée. A chaque étape durant l'apprentissage, l'erreur du réseau est calculée et utilisée pour ajuster les poids de la couche de sortie au moyen des méthodes de gradient

(Haykin, 1999). Si le réseau contient plus d'une couche de connexion, l'erreur du réseau est aussi propagée vers l'arrière et est utilisé pour la mise à jour des poids des couches précédentes. Une fois que les poids sont déterminés après plusieurs étapes d'apprentissage, la phase de rappel peut fonctionner, où les calculs de sortie du réseau sont effectués en utilisant les données d'entrée et les poids attribués lors de la phase d'apprentissage.

La capacité d'apprendre comment classifier des informations en adaptant des poids, est appelé perceptron. Le perceptron simple peut être défini par l'équation suivante :

$$\Delta w_{i,j}^{(n)} = \alpha (d_j^{(n)} - z_j^{(n)}) x_i^{(n)} - \alpha \varepsilon^{(n)} x_i^{(n)} \quad (2.27)$$

Où $d_j^{(n)}$ est la valeur de sortie désirée pour l'entrée $x_i^{(n)}$, $z_j^{(n)}$ est la sortie actuelle à partir du $j^{\text{ème}}$ neurone avec l'entrée $x_i^{(n)}$, $\varepsilon^{(n)}$ est l'erreur de sortie et α est appelé le coefficient du taux d'apprentissage.

La mise à jour du poids de la matrice \mathbf{W} se détermine par l'équation suivante :

$$W^{(n+1)} = W^{(n)} + \Delta W \quad (2.28)$$

Le résultat après n ensembles d'apprentissage est que chacun des poids w_{ij} auront été mises à jour selon la règle décrite ci-dessous :

$$w_{ij}^{(n+1)} = w_{ij}^{(n)} + \alpha x_i^{(n)} \quad \text{si la sortie est zéro et devrait être un} \quad (2.29)$$

$$w_{ij}^{(n+1)} = w_{ij}^{(n)} - \alpha x_i^{(n)} \quad \text{si la sortie est un et devrait être zéro} \quad (2.30)$$

$$w_{ij}^{(n+1)} = w_{ij}^{(n)} \quad \text{si la sortie était attendue} \quad (2.31)$$

La règle d'apprentissage du perceptron peut être optimisée grâce à la mise au point du coefficient du taux d'apprentissage. Un petit α correspond à un schéma d'apprentissage stable et lent (Inman et al 2013). Par conséquent, dans la pratique, on aimerait faire α aussi large que possible pour assurer un apprentissage rapide sans introduire des oscillations instables sur la valeur désirée qui résultent de sur-relaxation.

La détermination de la fonction de mise à jour n'est pas unique, et elle dépend de la méthode d'erreur de minimisation adoptée qui peut être choisie en fonction du problème et du jeu de données. A titre d'exemple, dans l'algorithme de minimisation basé sur les gradients conjugués, la mise à jour est conjuguée le long de la direction pour laquelle la fonction d'erreur diminue le plus rapidement. Cette méthode effectue généralement une convergence plus rapide que les célèbres algorithmes de « steepest descent » (décente forte). Il existe d'autres algorithmes qui arrivent à effectuer de rapides

convergences telles que la méthode de Newton ou l'algorithme Levenber-Marquard qui considère la seconde dérivée de la fonction d'erreur. La littérature couvre un vaste champ de réseaux de neurones. Néanmoins, la plupart des applications sont basés sur quatre types d'algorithmes de mise à jour basiques, à savoir la basique « back propagation (BBP), l'algorithme du gradient conjugué (CG), l'algorithme Quasi-Newton (QN), et l'algorithme Levenberg-Marquard (LM). Dans Fatih et al (2008) une brève description de chacune de ces méthodes est donnée ainsi qu'une application de celles-ci avec comparaison et discussion.

Les types de réseau plus simple sont les réseaux linéaires (LN). Ces réseaux n'ont pas de couche cachée et la fonction d'activation de la couche de sortie est linéaire. Les poids et les biais de ce réseau sont formés en utilisant la règle Widrow-Hoff (Lin et Lee, 1996).

IV.3.2 Réseau de neurones à structures de perceptrons multicouches

Le modèle de réseau de neurones le plus utilisé est le réseau de neurones à structure de perceptrons multicouches (MLP). La structure MLP est composée d'une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie. La couche cachée est caractérisée par plusieurs unités non linéaires (les neurones).

Le réseau de neurones avec d entrées, h neurones de la couche cachée et une seule sortie linéaire définissant un schéma paramétré non-linéaire à partir d'une entrée x à une sortie y donné par la relation suivante :

$$y = y(x; w) = \sum_{j=0}^h [w_j f(\sum_{i=0}^d w_{ji} x_i)] \quad (2.32)$$

La fonction d'activation f est une fonction tangente hyperbolique définie par exemple dans Diagne et al (2013). Les paramètres du réseau de neurone sont donnés par des poids et des biais qui connectent les couches entre elles dans l'équation (38). Les biais sont notés par i et j , les paramètres w gouvernent le schéma non linéaire.

Les perceptrons à multicouches avec apprentissage par rétropropagation sont très efficaces pour effectuer des schémas arbitraire non-linéaires. Les MLP sont également appelés MLFF (Multi-Layer Feed-Forward) qui souligne le fait que le flux vers l'avant de l'information par rapport à la direction vers l'arrière de réglage de poids utilisé dans l'algorithme BPL.

IV.3.3 Aperçu de l'application des modèles de réseaux de neurones

Mohandes et al (1998) ont estimé le rayonnement global avec le modèle de réseaux de neurones. Ils ont utilisé l'algorithme de rétro-propagation pour l'apprentissage de plusieurs réseaux de neurones

multicouches feed-forward pour estimer les valeurs mensuelles du rayonnement solaire global. Le meilleur réseau se compose de quatre entrées, 10 neurones dans une couche cachée et un neurone dans la couche de sortie. La performance prédictive montre une MAE de 6,5 à 19,1%.

Dorvlo et al (2002) ont utilisé 8 stations moyennées pour l'apprentissage et la validation du MLP réseaux de neurones et le modèle RBF (Fonction de base radiale) jugée pour sa rapidité de calculs par rapport au modèle MLP. Les paramètres d'entrée sont la latitude, la longitude, l'altitude et le ratio d'ensoleillement (S/S_0) et le mois de l'année. Les paramètres de sortie sont l'indice de clarté (H/H_0). Le rayonnement solaire mensuel a été obtenu en multipliant l'indice de clarté estimé par H_0 . Ils trouvent une RMSE minimale avec le RBF de $0.83 \text{ MJ/m}^2/\text{day}$.

Sfetsos et al (2000) ont introduit une approche simple pour la prévision du rayonnement solaire horaire en utilisant divers techniques basées sur l'intelligence artificielle (RNA et ANFIS). Ils ont également étudié d'autres variables météorologiques comme la température, la vitesse du vent et la pression. Une comparaison entre les différents modèles en termes d'erreur de prédiction a indiqué que le réseau formé avec l'algorithme Levenberg-Marquardt était la prédiction optimale du modèle. Il présente l'erreur RMSE la plus basse sur l'ensemble de la prédiction avec $27,58 \text{ W.m}^{-2}$ qui présente une amélioration de $2,3 \text{ W.m}^{-2}$ comparé à l'approche univariée.

Mihalakakou et al (2000) ont développé un modèle de simulation de série temporelle du rayonnement solaire global sur la base ANN et l'ont appliqué à Athènes. Le réseau logique Neural était identifié comme étant le modèle ayant le moins d'erreurs. Ils ont produit une erreur quadratique moyenne de 4,9% de moins que l'approche de persistance.

Les travaux de Tymvios et al (2005) ont démontré que le modèle de type ANN peut être une alternative à l'approche statistique traditionnelle d'estimation du rayonnement solaire sur un plan horizontal basé sur la méthodologie d'Angströms, particulièrement dans les cas où les mesures de rayonnements ne sont pas facilement accessibles. Les données utilisées sont les valeurs totales horaires du rayonnement sur une surface plane horizontale. Une comparaison a été effectuée entre 10 modèles statistiques basés sur la méthodologie d'Angströms et 7 modèles basés sur les ANN. En comparant le RMSE pour tous les modèles présentés dans cet article, il a été montré que le modèle ANN-5 avait les meilleurs résultats.

La méthode ANNs a été également utilisée dans les travaux de Cao et Lin (2008) utilisant un modèle DRWNN (réseau de neurone à ondelette diagonale récurrente) pour la prévision du rayonnement solaire global horaire. RMSE du modèle DRWNN est de 13.2 W m^{-2} et 19.05 W m^{-2} pour les prévisions horaires et journalières respectivement. Egalement Cao et Cao (2005) ont combiné les réseaux de neurones à l'analyse d'ondelettes. De même en 2008 Cao et al ont utilisé un modèle de

réseau de neurones diagonales pour l'estimation du rayonnement solaire à l'aide d'une technique floue. Ces modèles hybrides seront explicités dans la section IV.

Hocaoglu et al. (2008) ont utilisé une année de données d'ensoleillement global horaire de Iki Eylul, en Turquie pour effectuer une prédiction de l'ensoleillement global horaire en proposant un modèle de représentation (2D) bidimensionnel des données de radiation solaires horaires par des méthodes de traitement d'image. Pour tester l'efficacité de prédiction du modèle, neuf filtres linéaires différents sont optimisés et testés. Les résultats fournissent le modèle de corrélation nécessaire et les directions de prédiction pour obtenir le modèle de prédiction optimum pour la prédiction. Les filtres linéaires optimaux et le modèle de réseau de neurone sont comparés par les résultats de RMSE. Les structures de réseaux de neurones développées dans cette étude ont jusqu'à trois entrées et trois neurones dans la couche cachée, avec une seule sortie correspondant à la prédiction. 10 mois de données sont utilisées pour l'apprentissage et deux mois de données sont utilisés pour les tests. Il est observé que le modèle de réseaux de neurones le plus performant est celui ayant trois entrées avec une couche cachée à trois neurones avec l'application de la méthode Levenberg–Marquard (LM) pour le processus d'apprentissage. Il est observé que les NNs surpassent les filtres linéaires due à leur capacité à capturer la non linéarité qui est inhérente aux variations saisonnières des données. De plus, les meilleurs résultats sont obtenus pour les modèles NNs avec un prétraitement en 2-D avec une RMSE minimale de 39,37.

Mellit et al (2010) où un α -model adaptatif a été développé pour la prédiction du rayonnement solaire global, direct et diffus horaire. Les données utilisées sont cinq années par pas de 5 de rayonnement solaire global en Arabie Saoudite et les variables exogènes utilisées sont la durée d'ensoleillement, la température et l'humidité. Les résultats trouvés sont proches des données réelles. En effet, le coefficient de corrélation entre les données mesurées et prédites est de plus de 97%.

IV.3.4 Limites des modèles de réseaux de neurones

On peut citer quelques problèmes avec les réseaux de neurones notés dans la littérature.

Dans le processus itératif, le réseau identifie les minima locaux plutôt que les minima globaux. Dans un jeu de données complexes, de par sa trop grande capacité à stocker des informations, une structure dans une situation de sur apprentissage aura de la peine à généraliser les caractéristiques des données. Elle se comporte alors comme une table contenant tous les échantillons utilisés lors de l'apprentissage et perd ses pouvoirs de prédiction sur de nouveaux échantillons. Ceci est particulièrement un problème quand le réseau présente un grand nombre de nœuds dans la couche cachée, ou lorsque l'apprentissage implique un grand nombre d'itérations. Pour limiter ce type de problèmes dans le cas des réseaux de neurones, on doit veiller à utiliser un nombre adéquat de neurones et de couches cachées. Cependant, ces paramètres sont difficiles à déterminer à l'avance.

V. Modèles hybrides

V.1. Aperçu des travaux de modèles hybrides

L'idée d'une méthodologie hybride est partie du constat que les prédicteurs autonomes sont inefficaces dans certains cas et très efficace dans d'autres cas spécifiques. Une des solutions consisterait à combiner des prédicteurs pour comprendre et modéliser la dynamique du signal. Zhang (2003) confirme qu'un modèle hybride ayant à la fois des capacités de modélisation linéaires et non linéaires pourraient être une bonne alternative pour prédire des données de séries chronologiques. Bien souvent, les modèles hybrides pour la prédiction du rayonnement solaire utilise une combinaison du modèle ARMA pour les propriétés linéaires et les réseaux de neurones pour les propriétés non linéaires. La combinaison de différents modèles donne la possibilité d'augmenter la précision de prédiction, en récupérant les avantages et la performance de chaque méthodologie qui tendent à capturer différents aspects des tendances.

Zarzalejo et al. 2005 ont utilisé une technique basée sur la conjonction des méthodes ANN telles que la logique floue et les réseaux de neurones et l'indice de nuages dérivé à partir des techniques d'images satellites, pour l'estimation horaire du rayonnement global à partir d'images satellites. Les résultats de leur modèle hybride sont comparés avec les estimations du modèle basé sur la technique Héliostat I. Ils trouvent de meilleurs résultats pour leur modèle avec une RMSE minimale de 10,34% pour l'indice de clarté horaire du rayonnement global.

Mellit et al. (2005) présentent un modèle hybride qui combine les techniques de réseaux de neurones artificiels (ANNs) et une bibliothèque de matrices de transition de chaîne de Markov pour la prédiction du rayonnement global journalier. Ils trouvent une RMSE minimale de 5,25%.

Chaabene et Ammar (2008) ont introduit un estimateur employant une approche de neurones flous qui donne une estimation du comportement dynamique des paramètres météorologiques. Ces paramètres permettent d'obtenir des modèles de distribution qui serviront à la prévision à moyen terme, du rayonnement et de la température par un modèle ARMA. Les prévisions effectuées à moyen terme par le modèle ARMA sont utilisées pour effectuer des prévisions à court terme (5minutes) en utilisant la technique du filtre de Kalman. Pour la prévision à court terme (5min), ils trouvent une RMSE minimale de 2,1% (journée de flux solaire ne présentant pas de variations) et maximale de 8,22% (journée de flux solaire avec plus de variations mais de dynamique faible). Les données sont mesurées par pas de 5 minutes.

Cao et Lin (2008) combinent les caractéristiques dynamiques des réseaux de neurones récurrents(RNN) avec la capacité accrue des réseaux de neurones en ondelettes (WNN) dans les fonctions de cartographie non linéaires, un réseau de neurone récurrent à ondelettes diagonales (DRWNN) est créée pour des prévisions du rayonnement solaire horaire. Cette méthode est basée sur les travaux préliminaires de Cao et Cao (2005), ainsi que Cao et Cao (2006). Les erreurs de RMSE sont de $13,2\text{W.m}^{-2}$ pour les prévisions horaires pour des données mesurées à l'heure avec un vecteur d'entrée de 9 composants (pour plus de détails voir Cao J et Lin (2008)).

Reikard (2009) a effectué une comparaison des performances entre plusieurs modèles (ARIMA, régression, fonctions de transferts, réseaux de neurones) et des modèles hybrides. Pour la prédiction à court terme, le modèle hybride est basé sur une combinaison des réseaux de neurones et des techniques de régressions. Dans ces modèles hybrides, une régression eu ARIMA est estimée et les résidus sont ensuite traités à l'aide des réseaux de neurones. Ils trouvent une valeur minimale d'erreur MAE pour ces modèles hybrides de 12,64% à 5 minutes, 19,18% à 15 minutes et 18,73% à 30 minutes. Les données mesurées pour ces horizons de prédictions sont par pas de 1minute. Cette méthode de combinaison ARMA et réseaux de neurones est également utilisé dans Nelson et al.(1999); Hibbert et al. (2000); Zhang (2003); Zhang and Qi (2005); Hibon and Evgeniou (2005), Aburto and Weber (2007).

Mellit et al. (2010) ont adapté un α -modèle adaptatif, approprié au rayonnement solaire en tenant compte de l'influence de paramètres météorologiques, pour la prévision horaire du rayonnement global. Pour cela, ils tiennent compte dans leur méthode de la durée d'ensoleillement, la température de l'air, l'humidité relative, le rayonnement solaire direct et diffus et le rayonnement global mesuré par pas de 5 minutes. La performance du modèle montre une RMSE de $2,18\text{W.m}^{-2}$. La dynamique des variations des journées de flux solaire testées est assez faible.

Ji et Chee (2011) présentent une approche pour la prédiction horaire du rayonnement solaire en utilisant deux phases. La première phase consiste à retirer la tendance non-stationnaire des séries de temps du rayonnement solaire, pour laquelle plusieurs modèles sont appliqués. La deuxième phase. La deuxième phase consiste à prédire le rayonnement solaire par une combinaison de méthode ARMA et TDNN pour un horizon de prédiction de 1heure. Les données sont échantillonnées à 10minutes. Le modèle hybride de prédiction donne de bonnes performances pour des jours clairs avec une RMSE autour de 20%, mais quand les conditions météorologiques ne sont pas idéales, la RMSE augmente extrêmement pouvant atteindre 80%.

Marquez et Coimbra (2011) ont développé et validé une prévision du rayonnement solaire à moyen terme à la fois pour le rayonnement global et diffus basée sur des méthodes d'apprentissage stochastique, des expériences de terrain et la base de données NWS. Un algorithme génétique est

utilisé en tant que sélecteur d'entrée pour l'ANN afin de sélectionner l'entrée la plus pertinente. Pour les valeurs horaires prédites, ils trouvent une nRMSE minimale égale à 17,7 %.

Voyant et al (2011) ont proposé une étude de la contribution des données météorologiques exogènes comme des séries chronologiques à leur MLP (perceptron multi couches) optimisé. Pour ce faire, ils ont comparé différentes méthodes de prévision, y compris une prévision de persistance, indicateur de référence ARIMA, un ANN avec prétraitement en utilisant uniquement des entrées endogènes et un ANN avec prétraitement en utilisant deux entrées endogènes et exogènes. Il s'agit d'une prédiction journalière du rayonnement solaire. La NRMSE est de 25,43% pour le modèle multivarié qui donne les meilleurs résultats.

Dans Cyril Voyant (2012), trois modèles hybrides sont proposées: le premier propose d'utiliser ARMA pendant 6 mois au printemps et en été et d'utiliser un ANN optimisé pour l'autre partie de l'année; le second modèle est équivalent au premier mais avec un apprentissage saisonnier; le dernier modèle dépend de l'erreur qui s'est produite pendant l'heure précédente. Ces modèles ont été utilisés pour prévoir le rayonnement global horaire pour cinq localisations dans la région méditerranéenne. Ils ont pu constater qu'il n'y avait aucun intérêt à utiliser des méthodes comme ANN avec des variables endogènes et exogènes en été. En effet, il y a rarement des nuages au cours de cette période, un processus linéaire comme ARMA semble suffisant. L'utilisation du modèle ANN avec variables exogènes n'est intéressant que pendant de courtes périodes où beaucoup de nuages dans le ciel sont observés (essentiellement en automne et en hiver). Dans la meilleure configuration, la combinaison de ANN et ARMA permet une amélioration de plus de 1%, avec un maximum à l'automne (3,4%) et un minimum en hiver où seul ANN est la meilleure (0,9%).

De même, dans un autre article Voyant et al (2012) ont proposé une nouvelle technique pour prédire les séries de temps du rayonnement global horaire en utilisant les prédictions météorologiques à partir d'un modèle NWP. Ils ont optimisé le modèle de réseaux de neurones perceptrons multicouches (MLP) en incluant les données de prévisions du modèle ALADIN et les données endogènes précédemment rendues stationnaires. Ils ont utilisé une méthode de sélection de couche pré entrée (pre input layer) et ont ensuite combiné le MLP optimisé avec le modèle autorégressif ARMA. Les résultats trouvés ont montré que ce modèle hybride était très performant pour la prédiction horaire du rayonnement global horizontal pour cinq sites dans la zone Méditerranéenne. La nRMSE est de 14,9% pour le modèle hybride MLP/ARMA comparé au modèle de persistance qui est de 26,2%. D'autres travaux de Voyant et al (2013) montrent l'évolution de cette même approche. Dix années de données horaires de l'ensoleillement global mesuré sur cinq localisations en France méditerranéenne, sont exploitées. Ils comparent trois modèles. Le premier modèle étudié est déterministe: 6 mois sont prédits avec le modèle AR et les 6 derniers mois avec un modèle ANN. Le second modèle est équivalent au premier mais l'apprentissage est saisonnier: l'apprentissage du modèle ANN est effectué seulement en

automne et hiver alors que le modèle AR est appliqué pour les saisons été et printemps. Le troisième modèle est un modèle stochastique qui dépend de l'erreur de prédiction à l'heure précédente. Le troisième modèle montre une meilleure performance une RMSE normalisée minimale annuelle de 13,7%.

Marquez et al (2013) ont développé un modèle de prédiction basé les images satellites et les ANNs pour le rayonnement global à des horizons de prédiction de 30minutes à 2 heures. L'approche de prévision utilise les informations recueillies auprès de l'analyse d'images par satellite, y compris la vélocimétrie et l'indice des nuages comme entrées dans les modèles ANNs. Les données de rayonnement sont par pas de 30 minutes, ils trouvent une erreur RMSE de 56,79W.m⁻² pour un horizon de 30 min et de 80,21 W.m⁻² pour un horizon de 2h.

Chu et al (2013) propose un modèle de prédiction du rayonnement normal direct (DNI) qui combine la méthode du traitement d'image du ciel et les réseaux de neurones artificiels. Les données DNI sont par pas de 1 minute et les horizons de prévisions vont de 5 à 10 minutes. Ils trouvent une performance prédictive avec des erreurs RMSE minimales pour un DNI à faible variance de 44,8 W.m⁻² et de haute variance 109,3 W.m⁻² pour un horizon de prévisions à 5 minutes. Pour un horizon de prévisions à 10 minutes, la RMSE minimale est de 54,6 W.m⁻² et 132 W.m⁻² pour respectivement un DNI à faible variance et à haute variance.

Huang et al (2013) proposent le modèle CARDS qui combine deux approches, une approche autorégressive (modèle AR(2)) et une approche de système dynamique basée sur la méthode de Lucheroni (2007) pour la prévision horaire du rayonnement global. Les résultats trouvés par le modèle CARDS donne une meilleure performance pour les jours à ciel clair soit une RMSE de 17% et 32% pour les jours à ciel nuageux, bien qu'en moyenne pour tous les jours, ils trouvent une RMSE de 16,5%. Les données sont mesurées par pas horaire.

V.2. Limites des modèles hybrides

Les limites des modèles hybrides sont imposées par la résolution spatiale et temporelle des méthodes utilisées. En effet, la performance de chaque technique est une fonction de ces résolutions, particulièrement de l'horizon temporelle. Dans Inman et al (2013) un schéma représente la performance (RMSE) de plusieurs méthodes hybrides. Pour des horizons de prédictions de 30 à 120 minutes, le modèle hybride GA/ANN qui consiste à combiner un algorithme génétique dans le modèle ANN pour optimiser les paramètres (Pédro et al 2012), montre une meilleure performance métrique $s = 1 - \frac{RMSE}{RMSE_p}$ (ratio entre la RMSE du modèle considéré et la RMSE du modèle de persistance) comprise entre 30 et 40%.

Il est constaté également un autre type de limitation des modèles hybrides dans Inman et al (2013). En effet, les modèles hybrides sont souvent limités par le fait que les horizons temporels doivent être du même ordre de grandeur que la résolution temporelle des données. Par exemple, une fréquence d'échantillonnage de 2 Hz impliquerait la même limite sur la résolution temporelle, donc un horizon de prévision de l'ordre de quelques secondes serait approprié. Il est possible de prétraiter les données afin de produire, par exemple, la moyenne horaire ou journalière, ce qui réduit la fréquence d'échantillonnage et pour permettre une extension efficace de l'horizon temporel. Cela dit, l'argument de réciprocité ne tient pas: les données initialement enregistrées à basses fréquences d'échantillonnage, les moyennes journalières pour exemple, ne peuvent pas être pré-traités d'une manière qui rendrait appropriée les horizons à court terme.

VI. Modèles spatio-temporels et aperçu des travaux

Les comportements temporels et spatiaux du rayonnement solaire sont liés à des mécanismes atmosphériques complexes. Récemment des chercheurs se sont penchés sur la statistique spatio-temporelle, permettant d'analyser l'ensoleillement selon des processus non seulement dans des domaines temporels et spatiaux séparés, mais aussi dans l'ensemble (Finkenstädt et al (2006), Cressie et al (2011), Le ND et Zidek (2006)) tels que dans Gueymard et Wilcox où les auteurs fournissent une étude sur la variabilité spatio-temporelle (inter-annuelle) à long terme de l'ensoleillement global.

Dans cet état de l'art, seuls les modèles qui décrivent les processus spatio-temporels utilisés pour la prévision solaire sont cités. Dans la littérature nous citerons le modèle STARMA, le modèle VARMA et le krigeage.

VI.1. Modèle STARMA

Des modèles de prédiction utilisant l'approche statistique autorégressive mais basée sur des propriétés spatio-temporelles ont été utilisées pour la prédiction de séries temporelles caractéristiques d'une localisation spatiale. C.A Glasbey et D.J Allcroft en 2008 ont utilisé le modèle STARMA (spatio-temporel ARMA) pour évaluer la variabilité du potentiel de l'énergie solaire. Les modèles STARMA sont définis sur une grille carrée, et les paramètres ne sont pas liés à un site spécifique, mais dépendent simplement des vecteurs de distances entre les sites, ce qui conduit à des matrices de dépendance structurés (voir Bennett (1979)). Ils sont donc traitables pour plusieurs sites formant un réseau et n'ont pas besoin que les données soient mesurées sur la localisation pour laquelle la simulation doit être effectuée. Des hypothèses de spatio-stationnarité doivent être établies. Les données son recueillies à partir de 10 radiomètres dans le Pentland Hills au sud d'Edimbourg. Le rayonnement solaire a été enregistré en moyenne sur des intervalles de 10 minutes, sur chaque site pendant 2 ans. Les données utilisées pour l'application du modèle ne concernent qu'une séquence d'un mois. Les résultats de la

prédiction concerne une échelle de temps de 10 minutes pour 10 sites sur une séquence de 31 jours pour une grille de 10km*10 km. Ils constatent qu'un processus autorégressif du premier ordre (STAR (1)) et un processus de bruit Matérn fournit un ajustement adéquat des données. A Glasbey et D.J Allcroft nous démontre l'utilisation du modèle STAR (1) dans la simulation des réalisations de la production d'énergie.

Les processus STARMA sont un cas particulier des modèles VARMA (que nous détaillerons à la fin de cette section) où les éléments dans Y sont des observations sur une grille carrée. Soit une grille $n \times n$, un emplacement dans un quadrillage bidimensionnel i,j est représenté par l'élément $ni+j$ dans Y et $N = n^2$. Pour s'assurer que le processus est spatialement homogène, une entrée à $(ni+j, nk+l)$ dans une des matrices dépend seulement du vecteur de distance entre les localisations à deux dimensions (i,j) et (k,l) . :

Les modèles STARMA sont particulièrement adaptés à la simulation, en ce que, une fois que les processus d' Y et Z ont été initialisés, des points de temps ultérieurs peuvent être générés directement à partir du modèle :

$$Y_t = \sum_{r=1}^p \phi_r Y_{t-r} + \sum_{s=0}^q \theta_s Z_{t-s} \quad (2.33)$$

Y_t est le processus observé et Z_t le bruit, les deux N -vecteurs $Z_t \sim N(0, \Lambda)$ indépendants du temps t , et Λ est une $N \times N$ matrice de variance positive. ϕ_1, \dots, ϕ_p les coefficients autorégressifs et $\theta_0, \dots, \theta_q$ sont les coefficients de moyenne glissante, toutes les matrices $N \times N$ avec $\theta_0 \equiv I$, la matrice d'identité.

Dans Glasbey et Allcroft (2008), ils utilisent les distances euclidiennes plutôt que les vecteurs de distances.

Les auto-covariances de Y peuvent être calculées via l'équation de Yule-Walker :

$$\Gamma_j = \sum_{r=1}^p \phi_r \Gamma_{j-r} + \sum_{s=j}^q \theta_s \Lambda \Psi_{s-j}^T \quad (2.34)$$

Pour $j=0,1,\dots$ où $\Gamma_j = E(Y_t Y_{t-j}^T)$ est la $N \times N$ matrice d'auto-covariance au temps de retard j et les matrices auxiliaires sont obtenues par :

$$\Psi_j = \sum_{i=1}^j \phi_i \Psi_{j-i} + \theta_j \quad (2.35)$$

Pour $j=0,1,\dots$. Le premier $p+1$ de ces équations de matrice sont résolus pour $\Gamma_0, \dots, \Gamma_p$ en utilisant $\Gamma_{-j} \equiv \Gamma_j^T$ et ensuite $\Gamma_{p+1}, \Gamma_{p+2}, \dots$, sont obtenus récursivement. Pour un processus STAR(1) on a :

$$Y_t = \phi Y_{t-1} + Z_t \quad (2.36)$$

et les autocovariances sont données par :

$$\Gamma_0 = (I - \phi^2)^{-1} A \quad (2.37)$$

$$\Gamma_j = \phi \Gamma_{j-1} = \phi^j \Gamma_0 \quad \text{pour } j=1,2,\dots \quad (2.38)$$

Dans Glasbey et Allcroft (2008), l'espace est approximé par un tore, ainsi toutes les matrices deviennent des matrices TBC(Toeplitz bloc circulant) et la charge de calcul est considérablement réduite en utilisant des transformées de Fourier (pour plus de détails voir l'article). Ils définissent une fonction de distance circulaire et une distance euclidienne toroïdale entre les localisations à deux dimensions (i,j) et (k,l) . Les termes de la matrice dépendent seulement de la distance euclidienne toroïdale. En principe, les ordres des modèles STARMA peuvent être déterminées à partir des modèles de dégradation ou de coupure de corrélation ou les fonctions de corrélation partielle, comme avec les modèles ARMA univariés, pour des cas d'identification si $p>0$ et $q>0$ (Brockwell et Davis (1991). Toutefois, les critères d'information (AIC, BIC) sont plus faciles à utiliser.

Dans l'application du rayonnement solaire, les valeurs manquantes sont une complication supplémentaire, par conséquent des techniques simples ad hoc sont utilisées dans l'article. Les paramètres des données originales et les paramètres estimés sur 100 simulations sont comparés. Les résultats montrent des valeurs proches des valeurs réelles avec des différences au centième.

Dambreville et al (2014) ont également utilisé une base de modèle temporel autorégressif AR en y introduisant des paramètres spatiaux. L'intérêt majeur de cette étude se trouve dans la combinaison d'informations temporelles avec les informations spatiales des pixels HelioClim-3 entourant le site de mesure. Les cartes HelioClim-3 SSI présentent une grande quantité d'informations spatio-temporelles selon la région traitée dans l'étude. Les auteurs ont donc créé une stratégie de sélection de la région à étudier permettant d'obtenir les paramètres spatiaux à inclure dans les données d'entrée du modèle. Dans Marquez et al (2013), la sélection de la région pour laquelle les paramètres seront utilisés en données d'entrée, est développée par une approche locale basée sur le calcul de la vitesse des images satellites. Néanmoins, cette méthode exige une description précise du champ de vecteur de mouvement de nuage. Dans une tentative de profiter des informations spatiales en gardant un algorithme très simple, Dambreville et al (2014) ont extrait de la série de temps d'images seulement la tendance principale observée sur le site de mesure. Les auteurs se basent sur l'hypothèse que beaucoup de sites continentaux sont souvent affectés par les mêmes phénomènes météorologiques, et proposent ainsi de

déterminer ce comportement comme une première étape. Les données d'entrée spatiales sont ajustées pour chaque pas temporel selon un calcul de vecteur de mouvement des nuages Marquez et al (2013) tandis que pour le modèle de Dambreville et al (2014), les données d'entrée sont sélectionnées selon une direction moyenne calculée par traitement des données de l'année 2011. Ils créent donc le modèle Ast2, qui inclut différentes régions comme données d'entrée selon l'horizon de prévision. Les régions sont sélectionnées par analyse des cartes d'intercorrélation entre les séries de temps des deltas indices de clarté et chaque pixel HelioClim-3 pour différents décalages temporels. Pour chaque horizon de prévision ou chaque décalage temporel défini, une région sera alors sélectionnée en données d'entrées du modèle. Ce qui rend cette technique nouvelle par rapport au modèle Ast qui prend les données d'entrée spatiales de la grille de pixels centré sur la localisation du site. Ils ont appliqué leur modèle sur l'indice de clarté moyenné à 15 minutes pour correspondre à la résolution des bases de données de HelioClim-3 avec des horizons de prévisions à 15 min, 30 min, 45 min, 60 min. Pour certifier la stationnarité de ces tendances, ils ont examiné l'année 2011, mais ont exécuté l'algorithme pour l'année 2012. La performance prédictive du modèle Ast 2 donne les meilleurs résultats avec une RMSE de 75 W.m⁻² à 99 W.m⁻² pour respectivement les horizons de prévision de 15 min à 60 min. L'originalité de ce travail réside dans l'intégration d'informations spatio-temporelles à partir d'images satellites sans aucun calcul de vecteur de mouvement des nuages.

VI.2. Modèle krigeage

Le modèle krigeage espace-temps est un prédicteur linéaire qui utilise la corrélation ou la covariance pour prédire le processus spatio-temporel $z(s,t)$ dans l'espace s et au temps t :

$$\hat{z}(s,t) = c_0' C^{-1} z \quad (2.39)$$

C est la matrice de covariance, c_0 est un vecteur de covariance et le terme $c_0' C^{-1}$ donnent les poids. C a la dimension $\mathbb{R}^{N \times M} \times \mathbb{R}^{N \times M}$, où N est le nombre de localisations spatiales et M est le maximum de temps de retard sous considération que :

$$C = \begin{pmatrix} C_0 & \cdots & C_{M-1} \\ \vdots & \ddots & \vdots \\ C_{M-1} & \cdots & C_0 \end{pmatrix}, \in \mathbb{R}^{N \times M} \times \mathbb{R}^{N \times M}, \quad (2.40)$$

C_τ est la matrice au temps de retard τ :

$$C_\tau = \begin{pmatrix} C_{11,\tau} & \cdots & C_{1N,\tau} \\ \vdots & \ddots & \vdots \\ C_{N1,\tau} & \cdots & C_{NN,\tau} \end{pmatrix}, \in \mathbb{R}^{N \times N}, \quad (2.41)$$

Où $C_{ij,\tau} = cov \{z(s_i, t), z(s_j, t - \tau)\}$

L'expression du vecteur c_0 est:

$$c_0 = (c_{0,1} \ c_{0,2} \ \dots \ c_{0,M}) \quad \in \mathbb{R}^{NM \times 1} \quad (2.42)$$

et

$$c_{0,\tau} = (c_{11,\tau} \ c_{12,\tau} \ \dots \ c_{1N,\tau}) \quad \in \mathbb{R}^{1 \times N} \quad (2.43)$$

$$\text{avec } c_{ij,\tau} = \text{cov} \{z(s_i, t), z(s_j, t - \tau)\}$$

z est un vecteur qui est construit en utilisant les valeurs observées à partir des M pas de temps passés et des N localisations spatiales ayant la dimension $\mathbb{R}^{NM \times 1}$. La méthode du modèle krigeage spatio temporel est la prévision du rayonnement par une combinaison linéaire pondérée des données passées de rayonnement. La matrice de covariance C et le vecteur de covariance c_0 peuvent être empiriques ou calculés à partir de fonctions de covariance. Plusieurs fonctions de covariance sont détaillées dans Yang et al 2013. En effet, le principe consiste à modéliser ou ajuster le variogramme expérimental qui représente la corrélation entre les sites en fonction des distances entre les sites. Le variogramme est un outil qui nécessite l'hypothèse de stationnarité. Les méthodes géostatistiques modélisent le variogramme expérimental par une fonction continue soumise à certaines contraintes. Les cinq variogrammes théoriques les plus utilisés associés à une covariance stationnaire permettant l'ajustement du variogramme expérimental, sont les variogrammes pépitique, exponentiel, sphérique, Matérn et puissance. Cette étape s'appelle la modélisation ou l'ajustement du variogramme. La modélisation est la partie essentielle du krigeage. Les poids $c_0' C^{-1}$ dépendent explicitement de la fonction de covariance qui dépend des données, à travers le variogramme, et entre autre de la géométrie de l'échantillon. Yang et al (2013) ont créé un modèle de prévision du rayonnement solaire basé sur une approche combinant les structures de covariance et la méthode de krigeage. Les données d'entrée sont l'indice de clarté horaire de 10 stations à Singapour pour la période de novembre 2012. Il est connu qu'aucun modèle de variogramme isotropique ne reflétera précisément des données de rayonnement solaire (Rehman et al (2000)). Ainsi, les auteurs ont développé une méthode de prédiction qui opère sur les structures de covariance spatiales qui ne supposent pas de stationnarité plutôt que des méthodes qui font faussement cette supposition. Le krigeage est utilisé dans les travaux de Yang et al (2013) pour évaluer la capacité de prédiction des structures spatio-temporelles et pour évaluer les poids des valeurs observées (M décalages temporels et N points spatiaux) en utilisant les structures de covariance connues. La tendance temporelle est retirée par une méthode basée sur les travaux de Sampson et Guttorp (2006) qui estime les structures de covariance spatiales utilisant une approche non paramétrique. La méthode construit une fonction entre les localisations dans l'espace géographique, où la stationnarité n'est pas supposée, à des localisations dans un nouvel espace où l'isotropie est supposée. Cette technique applique une transformation de coordonnées pour obtenir une stationnarité spatiale. L'espace géographique se réfère au plan G et l'espace de dispersion au plan D .

En appliquant la méthode MDS, la tendance spatiale est ainsi retirée. L'isotropie impliquera que les variations du processus spatial sont identiques quelque soient la direction, par conséquent la covariance spatiale est une fonction dépendante de la distance. Après transformation des données, la performance du modèle de prévision basé sur la méthode de krigeage est explorée. Deux structures de covariance sont présentées et adaptées. L'une des structures peut être séparée en une structure purement spatiale, et une structure purement temporelle. La deuxième structure de covariance entièrement symétrique fait intervenir un processus spatio-temporel. Le modèle plus général, soit entièrement symétrique, a une RMSE globale de 14,4%, tandis que le modèle séparable a une RMSE globale de 14,6%.

On peut également faire référence au modèle de krigeage spatio-temporel appliqué dans Yang et al (2014) pour la prédiction du rayonnement solaire global avec un horizon de prévision de 5 minutes. Les données sont préalablement mesurées à 1 seconde puis moyennée à 5 minutes collectées sur 13 stations météorologiques à Singapour. Les stations les plus proches sont distantes à environ 4 km sinon à une dizaine de kilomètres. Le modèle est testé pour plusieurs temps de retard (35 minutes et 20 minutes). Ils étudient au préalable les distributions de corrélations entre les sites à partir des fonctions d'inter-covariance et d'intercorrélation ainsi le seuil de temps de retard pour lequel les corrélations entre les sites sont significatives, est estimé. La distance seuil pour laquelle les corrélations entre les sites restent significatives, est estimée en utilisant L-méthode. De plus, ils ont étudié la méthode MDS (échelle non-métrique multidimensionnelle) afin d'améliorer l'hypothèse de spatio-stationnarité (isotropie spatio-temporelle) puisqu'ils considèrent que la variance de l'ensoleillement ne dépend pas strictement de la distance géographique entre les sites ce qui rend le contexte d'étude non isotropique. Les distances D-plan entre les stations en fonction de la dispersion spatiale sont ainsi analysés (voir Yang et al 2013 et Yang et al 2014).

Les résultats montrent une bonne performance prédictive du modèle krigeage spatio-temporel avec une RMSE minimale de 19,12 % pour un temps de retard de 20 minutes (4×5minutes).

Inoue et al (2012) ont utilisé l'information donnée par la vitesse et la direction du mouvement des nuages en utilisant des corrélations pour réaliser une carte du rayonnement solaire avec une résolution spatiale et temporelle de 500 mètres et 5 minutes par la méthode de krigeage spatio-temporel.

D'autres applications de la technique de krigeage ont été faites pour des données environnementales autres que le rayonnement solaire tel que dans Cellura et al (2008) qui a utilisé une technique combinant les méthodes de perceptron à multi couches et le krigeage pour obtenir une estimation de la vitesse spatiale du vent.

VI.3. Modèle VARMA

Le modèle spatio-temporel VARMA a été exploré dans les deux articles travaillant sur la prédiction spatio-temporelle du rayonnement solaire précédemment cités (Glasbey (2008) et Yang et al (2014)). Ici, nous détaillerons la méthodologie du modèle VARMA utilisée dans de Luna et Genton (2005).

Le modèle vectoriel autorégressif (VAR) est l'un des modèles les plus réussis, flexible et facile à utiliser pour l'analyse des séries temporelles multivariées. Il est une extension du modèle autorégressif univariée de séries temporelles multivariées dynamiques. Le modèle VAR s'est avéré particulièrement utile pour décrire le comportement dynamique des séries temporelles économiques et financières ainsi que pour la prévision. Il fournit souvent des prévisions supérieures à celles des modèles de séries temporelles univariées. En plus de la description des données et de la prévision, le modèle VAR est également utilisé pour l'inférence structurelle et l'analyse de la politique. Par ailleurs, les modèles VAR temporelles peuvent s'étendre à une méthodologie de prévision spatio-temporelle pour fournir des prévisions de temps, en tenant compte des dépendances spatio-temporelles.

de Luna et Genton (2005) ont utilisé le modèle VAR spatio-temporel pour la prévision de concentrations en monoxyde de carbone, enregistrées sur 4 stations pour des observations horaires. Les stations sont distantes de 400 m (distance minimale) à 1km (distance maximale). Ils ont comparés quatre stratégies de prévision, le modèle univarié AR testé avec le critère AIC et BIC et le modèle spatio-temporel VAR testé avec le critère AIC et BIC. Les erreurs de prévisions calculées (RMSE cumulative) montrent une meilleure performance pour le modèle univarié pour la Station 1, néanmoins avec des valeurs d'erreurs statistiques proches du modèle VAR (0,371 pour AR au lieu de 0,380 pour VAR). Pour la Station 3, le modèle VAR intégrant une structure spatiale montre de meilleurs résultats de prévisions avec des erreurs de 0,466 au lieu de 0,480 pour le modèle AR. De plus, les auteurs notent que le critère BIC offre les meilleures performances de prédiction. Le modèle VAR spatio-temporel a montré une bonne performance prédictive et a été d'un grand intérêt pour la prévision horaire de concentrations d'air polluantes. De plus, ils ont testé le modèle spatio-temporel VAR(3) pour la prédiction de vitesse de vent en analysant deux différents ordres des stations pour une grille de 11 stations. Un premier ordre basé sur les distances entre les sites et un autre ordonnant les sites en tenant compte de la direction prédominante du vent. Ils trouvent de meilleurs résultats pour l'ordre des localisations respectant cette information physique. Deux principales observations peuvent être données. Premièrement, le modèle VAR spatio-temporel surclasse les modèles de séries chronologiques univariés (modèle de persistance et AR avec le critère AIC et BIC). Deuxièmement, les résultats montrent une meilleure performance pour les sites ordonnés selon la direction du vent. Les détails de la méthodologie seront explicités dans le chapitre V.

Le modèle VAR avec une méthode plus ou moins similaire, a été également utilisé pour la prédiction du rayonnement solaire global à un horizon de prévision de 5 minutes dans Yang et al (2014). On peut

noter quelques distinctions entre ces deux méthodes. D'une part, celle utilisée dans Yang et al (2014) pour 13 localisations spatiales contrairement à de Luna et Genton (2005) qui applique leur méthode VAR pour des sites dispersés dans l'espace (méthode applicable même pour seulement quatre sites). D'autre part, différents paramètres de shrinkage (Lasso et ridge regression) sont testés et intégrés dans le modèle VAR créé par Yang et al (2014). Une RMSE minimale de 19,08% a été trouvée.

VII. Synthèse et conclusion de l'état de l'art

Ce chapitre fait l'état de l'art des techniques utilisées pour la prévision du rayonnement solaire, notamment les modèles NWP et les méthodes statistiques couramment utilisés par la recherche académique. L'analyse de la performance des prévisions solaires et des limites des différentes approches existantes est discutée. Ainsi, ce chapitre nous permet d'évaluer et de justifier le choix de notre modèle pour notre étude spatio-temporelle de l'ensoleillement global. Au regard des travaux de recherche cités, les meilleures valeurs de RMSE sont approximativement autour de 12% à 22% pour des jours à ciel clair et de 28% à 44% pour des jours de type nuageux.

Plusieurs études pour la prédiction du rayonnement solaire, sont basées sur l'observation d'un historique de données à savoir les modèles d'auto régression des séries de temps tels que les modèles de prévision ARIMA et ARMA caractérisée par une méthode linéaire. Une analyse publiée par Reikard (2009) montre qu'après une comparaison entre différents modèles de régression, ARIMA avec des coefficients variant dans le temps donne les meilleurs résultats, en raison de sa capacité à capturer le cycle diurne du rayonnement solaire plus efficacement que les autres méthodes.

Les modèles de réseaux de neurones démontrent par leurs résultats une bonne performance également due à leur méthode non-linéaire et semblent être une bonne alternative aux modélisations physiques pour la prédiction du rayonnement solaire quand il y a un historique de données disponibles suffisant. Ces modèles sont de bons prédicteurs pour des cas non-linéaires, stochastiques et multivariés et réussissent avec succès la prédiction du rayonnement solaire pour des horizons de prévisions à court terme (de quelques minutes à l'heure).

Les méthodes de prédiction basées sur les modèles numériques permettent une prédiction du rayonnement solaire, même aux localisations où les données sont indisponibles. Néanmoins, ces méthodes ont des limites spatiales et temporelles inévitables due à la fréquence d'échantillonnage et à la résolution spatiale des images satellite assez importantes. En effet, les techniques de prévision numériques du temps n'ont pas la résolution spatiale et temporelle pour fournir des informations sur les fluctuations du rayonnement solaire à haute fréquence.

Pour des horizons à court terme, l'imagerie basée au sol semble être une alternative très prometteuse. L'utilisation d'imageurs du ciel permet l'acquisition de couverture nuageuse locale qui fournit ainsi des informations sur les fluctuations à haute fréquence du rayonnement solaire pour des horizons de prédiction d'environ 5 à 20 minutes.

Les méthodes hybrides peuvent combiner deux ou plusieurs techniques différentes. Ces méthodes montrent beaucoup d'avantages par rapport aux méthodes traditionnelles, en ce qu'elles ont la capacité de combiner la prévision stochastique et déterministe. Les modèles hybrides combinant les modèles satellites avec des méthodes d'apprentissage stochastiques montrent des résultats de prédiction avec une bonne précision pour plusieurs horizons de prévisions.

Nous avons également effectué un état de l'art des modèles spatio-temporels développés pour la prévision de l'ensoleillement. Récemment, quelques chercheurs ont utilisé les statistiques spatio-temporelles pour modéliser des processus environnementaux. Les modèles spatio-temporels dans la littérature sont définis sur un espace régulier, partitionné avec des limites bien définies telle une grille. Ils sont donc traitables pour plusieurs sites formant un réseau bidimensionnel.

Le Tableau 2.1, résume les performances prédictives à très court terme (de 5minutes à 15 minutes) trouvées dans les articles cités préalablement dans notre état de l'art. Nous ne citerons que les performances prédictives évaluées par les erreurs statistiques telles que la RMSE, MBE ou MAE. Ainsi, bien que les travaux de Chi Wai et al (2011) et Glasbey et al (2008) discutent d'une méthode de prévision du rayonnement solaire basée respectivement sur l'imagerie des nuages au sol à un horizon de prédiction de 5minutes et sur le modèle STARMA pour des horizons de prévisions de 10 minutes, ils ne seront pas cités. En effet, les erreurs (matching error, écart-type et moyenne des paramètres estimés) évaluant la performance prédictive sont exprimées de telles sortes qu'elles ne peuvent être comparées avec les erreurs trouvées dans les articles nommés dans le tableau (2.1).

MAE		
Horizon de Modèles prévision	15 minutes	5 minutes
Régression	26,52%	12,63%
UCM	25,97%	12,66%
ARIMA	18,97%	13,21%
Fonction de transfert	19,26%	13,57%
Réseau de neurones	21,04%	12,78%
Hybride	19,18%	12,64%
RMSE ($W.m^{-2}$) ou RMSE relative (%)		
	15 minutes	5 minutes
krigeage		18,49%
VAR		18,09%
At	86 (en $W.m^{-2}$)	
Ast	78 (en $W.m^{-2}$)	
Ast2	75 (en $W.m^{-2}$)	

Tableau 2-1 Erreurs statistiques de prévisions solaire pour des horizons de prédiction de 15 minutes à 5 minutes répertoriés dans la littérature (Reikard (2008), Wu et Chang (2011), Yang et al (2014), Dambreville (2014))

Nous pourrions compléter ce tableau (Tab 2.1) par la Figure 2.1 représentant une analyse sur l'ensemble des modèles de prévision de l'ensoleillement. Cette figure issue de Diagne et al (2013), donne un résumé des caractéristiques des différentes techniques de modèles de prévision de l'ensoleillement de la littérature. Il s'agit plus précisément, d'une classification des modèles de prévisions effectuée par Diagne et al (2013) basée sur la résolution spatiale des données d'entrée et la résolution temporelle des données de prévisions. Compte tenu de notre étude sur les modèles spatio-temporels définis dans la littérature, nous avons tenu à ajouter ce type de modèles dans cette classification. Les modèles spatio-temporels présentent un intérêt particulier pour notre analyse.

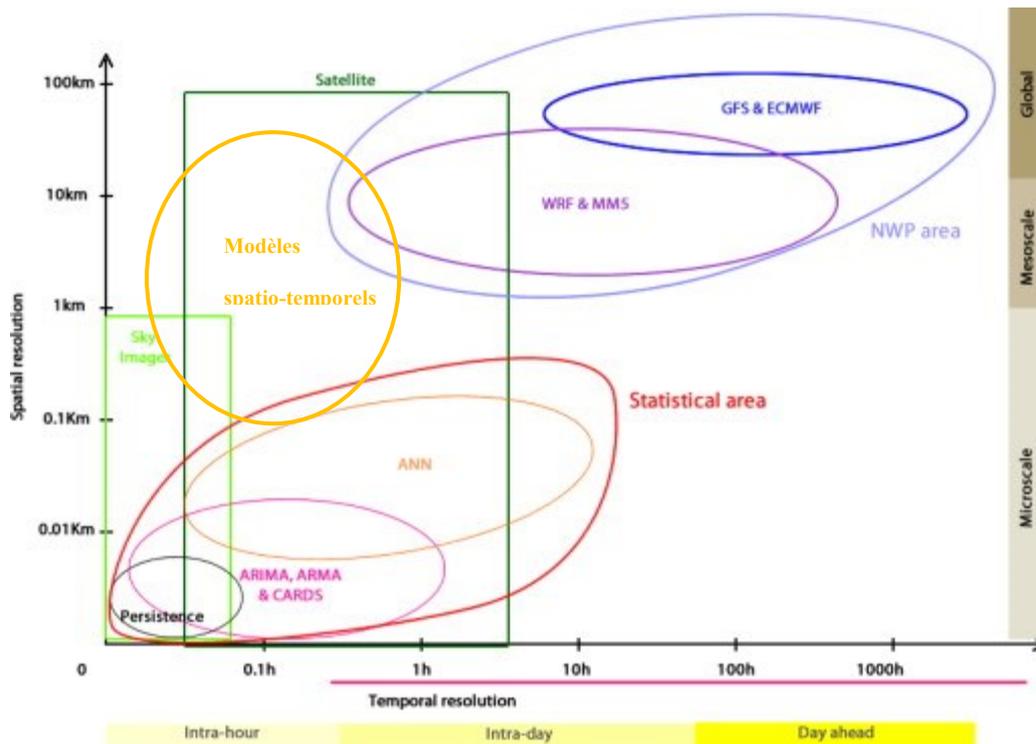


Figure 2-1 Classification des modèles basée sur une résolution spatiale et temporelle selon Diagne et al (2013) à laquelle nous avons ajouté les modèles spatio-temporels.

L'île de la Guadeloupe est un petit territoire avec des reliefs diversifiés, de nombreux microclimats et divers processus de formation de nuages. La sélection de modèle est basée sur des approches utilisant de petite résolution spatiale et temporelle. Diagne et al (2013) prennent en compte les besoins des gestionnaires sur l'île de La Réunion qui présentent les mêmes problématiques que La Guadeloupe. Diagne et al (2013) préconisent alors un modèle combinant les méthodes WRF (modèle de mésoéchelle), ARIMA et AR à des techniques d'imageries du ciel pour donner un produit de prévisions complet et plus précis avec différents horizons. Le filtrage de Kalman sera appliqué afin d'améliorer la précision des prévisions à partir WRF.

Nos travaux de recherche concernent spécifiquement l'étude spatio-temporelle de l'ensoleillement global pour trois sites de mesures. Nous pourrions dire que les modèles STARMA offrent un cadre idéal pour la modélisation spatio-temporelle des données sur une grille. Ce modèle étant un cas particulier des modèles VAR et VARMA, n'est défini que sur une grille donc pour une multitude de sites, et les paramètres ne sont pas liés à un site spécifique, mais plutôt dépendent uniquement des vecteurs de distances entre les sites. La méthode de krigeage réalise l'interpolation spatiale d'une variable régionalisée par calcul de l'espérance mathématique d'une variable aléatoire, utilisant l'interprétation et la modélisation du variogramme expérimental, ce qui contraint d'avoir une multitude de sites de mesures. Globalement, les analyses spatio-temporelles sont basées sur la quantification de l'interaction de chaque point de la grille avec son voisin. Ce qui impliquerait pour nous, de poser de

multiples capteurs sur l'ensemble du territoire. Les coûts d'un tel système seraient considérables. Nous devons créer une nouvelle approche qui sera contrainte à effectuer une analyse spatio-temporelle sur trois stations. Au vu des propriétés des modèles cités, la prédiction spatio-temporelle à court terme de l'ensoleillement global se basera sur l'approche de Luna et Genton (2005) utilisant la méthodologie du modèle VAR spatio-temporel. En effet, ce modèle est approprié pour des données riches dans la dimension temporelle mais dispersée et faible dans la dimension spatiale. L'hypothèse de spatio-stationnarité difficile à démontrer pour seulement trois sites de mesures n'est pas nécessaire pour ce modèle qui consiste à déduire la prévision solaire par une combinaison linéaire pondérée des séries temporelles passées à une localisation et des séries temporelles passées aux localisations corrélées avec un temps de retard défini. La suite des travaux se penchera sur l'adaptation et l'application de la méthode de prévision de ce modèle à nos données d'ensoleillement. Nous analyserons également sa performance prédictive. Toutefois, au préalable nous avons fait une étude spatio-temporelle de la variabilité de l'ensoleillement global entre nos trois stations.

Références

- Aburto, L., Weber, R., 2007. Improved supply chain management based on hybrid demand forecasts. *Applied Soft Computing*, 7, 136–144
- Anders Persson, Federico Grazzini, User Guide to ECMWF forecast products, Copyright 2005 ECMWF
- Armstrong, M., 2000, Comparison of MM5 forecast shortwave radiation with data obtained from the atmospheric radiation measurement program. In: Master of Science Scholarly Paper, University of Maryland, USA.
- Bacher P, Madsen H, Nielsen HA, 2009, Online short-term solar power forecasting. *Solar Energy*;83(10):1772e83.
- Benard C, Wirgin A, Body Y, Gobin D., 1978, Caractérisation de la stabilité de l'intensité solaire par l'analyse de séries aléatoires en France et au Pérou. *Météorologie*;VI:53e71.
- Bennett, R. J., . 1979, *Spatial Time Series: Analysis–Forecasting–Control*, London: Pion.
- Bofinger, S., Heilscher, G., , 2006, Solar Electricity Forecast – Approaches and first results. In: 21st PV Conference, Dresden, Germany.
- Boileau E, Guerrier B., 1979, Comparaison de modèles statistiques saisonniers et non saisonniers en météorologie solaire. *La Météorologie*;VI:115e30.
- Box, G.E.P., Jenkins, G.M., 1970, *Time Series Analysis: Forecasting and Control*. Holden-Day, , San Francisco.
- Breitkreuz, H., Schroedter-Homscheidt, M., Holzer-Popp, T., Dech, S., 2009, Short-range direct and diffuse irradiance forecasts for solar energy applications based on aerosol chemical transport and numerical weather modeling. *Journal of Applied Meteorology and Climatology*, 48 (9), 1766–1779.
- Brockwell, P. J. and Davis, R. A., , 1991, *Time Series: Theory and Methods*, 2nd edn. New York: Springer.
- Calbó J, Sabburg J., , 2010, Feature extraction from whole-sky ground-based images for cloud-type recognition. *Journal of Atmospheric and Oceanic Technology*;25:3e14.
- Cao J,LinX., 2008, Application of the diagonal recurrent wavelet neural network to solar irradiation forecast assisted with fuzzy technique. *Engineering Applications of Artificial Intelligence*,;21:1255–63.
- Cao J, CaoS., 2006, Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis.*Energy*;31:3435–45.
- Cao S,CaoJ., 2005, Forecast of solar irradiance using recurrent neural networks combined with wavelet analysis. *Applied Thermal Engineering*,;25: 161–72.
- Cellura M., Cirrincione G., Marvuglia A., Miraoui A., 2008, Wind speed spatial estimation for energy planning in Sicily:A neural kriging application, *Renewable Energy* 33, 1251–1266
- Chow C.W, Urquhart B, Lave M, Dominguez A, Kleissl J, Shields J, 2011, Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed, *Solar Energy*, 85: 2881-93

Chu, Y., Pedro, H.T.C., Coimbra, C.F.M., 2013, Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning. *Solar Energy*, 98, 592–603.

Cressie N, Wikle C., 2011, *Statistics for spatio-temporal data*. Wiley series in probability and statistics. Wiley.

Dambreville R., Blanc P., Chanussot J., Boldo D., 2014, Very short term forecasting of the Global Horizontal Irradiance using a spatio-temporal autoregressive model, *Renewable Energy*, 72, 291-300

de Luna Xavier et Marc G. Genton, 2005, Predictive spatio-temporal models for spatially sparse environmental data, *Statistical Sinica*, 15, 547-568

Deutsch, S. J. and Pfeifer, P. E., 1981, Space-time ARMA modeling with contemporaneously correlated innovations. *Technometrics*, 23, 401–409.

Diagne M., David M., Lauret P., Boland J., Schmutz N., 2013, Review of solar irradiance forecasting method and a proposition for small scale insular grids, *Renewable and Sustainable Energy Reviews*, 27:65–76

Dorvlo A.S.S., Joseph A.Jervase, Ali Al-Lawati, 2002, Solar radiation estimation using artificial neural networks, *Applied Energy*, Volume 71, Issue 4, 307-319

Fatih O., Gerek Hocaoglu, Omer N., 2008, Mehmet Kurba, Hourly solar radiation forecasting using optimal coefficient 2-D linear filters and feed-forward neural networks, *Solar Energy*, pp. 714–726.

Ferreira, P.M., Gomes, J.M., Martins, I.A., Ruano, A.E., 2012. A neural network based intelligent predictive sensor for cloudiness, solar radiation and air temperature. *Sensors* 12 (11), 15750–15777.

Ferreira P.M, Martins I.A.C, Ruano A.E, 2010, Cloud and Clear Sky Pixel Classification in Ground Based All-Sky Hemispherical Digital Images, “in Proceedings of CMTEE”, the IFAC Conference on Control Methodologies and Technology for Energy Efficiency, vol.1

Finkenstädt B, Held L, Isham V., 2006, *Statistical methods for spatio-temporal systems*. In *Monographs on statistics and applied probability*. Chapman & Hall/ CRC.

Flores J., Graff M., Rodriguez H., 2012 Evolutive design of ARMA and ANN models for time series forecasting, *Renewable Energy*, Volume 44, Pages 225–230

Gaetan C., Guyon X., *Modélisation et statistique spatiale*, Mathématiques et Applications, Comité de Lecture 2008-2011 Editorial Board 2008-2011

Ghonima MS, Urquhart B, Chow CW, Shields JE, Cazorla A, Kleissl J., , 2012, A method for cloud detection and opacity classification based on ground based sky imagery. *Atmospheric Measurement Techniques*;5(4):4535e69.

Glasbey C.A et Allcroft D.J, 2008, A spatiotemporal auto-regressive moving average model for solar radiation, *Biomathematics and Statistics Scotland*, Edinburgh, UK, *Appl. Statist.*, 57, Part 3, pp. 343–355

Gueymard C. A., Wilcox S. M., 2011, Assessment of Spatial and Temporal Variability in the U.S. Solar Resource from Radiometric Measurements and Predictions from Models Using Ground-Based or Satellite Data, *Solar Energy*, vol. 85, pp. 1068-1084

Hamill, T., Nehrkorn, T., 1993, A short-term cloud forecast scheme using cross correlation. *Weather Forecasting*, 8, 401–411.

Hammer, A., Heinemann, D., Lorenz, E., and Lückehe, B, 1999, Short-term forecasting of solar radiation: A statistical approach using satellite data. *Solar Energy*, 67:1-3, 139-150.

Hammer A., Heinemann D., Hoyer C., Lorenz E., 2001, Satellite Based Short-Term Forecasting of Solar Irradiance - Comparison of Methods and Error Analysis, in Proceedings of the EUMETSAT, Meteorological Satellite Data Users' Conference,

Haykin, S., 1999., Neural Networks: A Comprehensive Foundation. Prentice Hall, Upper Saddle River, NJ.

Heinemann D, Lorenz E, Girodo M., 2006., Forecasting of solar radiation. In: Solar Energy resource management for electricity generation from local level to global scale. Nova Science Publishers; p.223–33.

Hibon, M., Evgeniou, T., 2005, To combine or not to combine: selecting among forecasts and their combinations. International Journal of Forecasting 21, 15–24

Hibbert, H., Pedreira, C., Souza, R., 2000, Combining neural networks and ARIMA models for hourly temperature forecasts. In: Proceedings of the International Conference on Neural Networks (IJCNN 2000), Como, Italy, 414–419

Hocaoglu FO, Gerek ON, Kurban M., 2008, Hourly solar radiation forecasting using optimal coefficient 2-d linear filters and feed-forward neural networks. Solar Energy;82(8):714e26.

Huang J., Korolkiewicz M., Agrawal M., Boland J. , 2013, Forecasting solar radiation on an hourly time scale using a Coupled AutoRegressive and Dynamical System (CARDS) model, Solar Energy, Volume 87, Pages 136-149

Ineichen, P., Perez, R., 2002, A new airmass independent formulation for the Linke turbidity coefficient. Solar Energy, 73 (3), 151–157.

Inman Rich H., Hugo T.C. Pedro, Carlos F.M. Coimbra, 2013, Solar forecasting methods for renewable energy integration, Progress in Energy and Combustion Science 39, 535e576

Inoue, T., Sasaki, T., Washio, T., 2012, Spatio-temporal kriging of solar radiation incorporating direction and speed of cloud movement. In: The 26th Annual Conference of the Japanese Society for Artificial Intelligence, Yamaguchi City.

Ji W, Chee KC., 2011, Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN. Solar Energy, 85(5):808e17.

Kärner O., 2002, On nonstationarity and antipersistence in global temperature series, Journal of Geophysical Research;107(D20):4415

Konrad J. and Dubois E., 1992, Bayesian estimation of motion vector fields. IEE trans. Pattern Anal., Machine Intell.11, 910-927

Kostylev APV., 2011, Solar power forecasting performance – towards industry standards. In: First international workshop on the integration of solar power into power systems Aarhus, Denmark

Lacis, A., and Hansen, J., 1974, A parameterization for the absorption of solar radiation in the earth's atmosphere. J. Atmospheric Sciences, 31:1, 118-133.

Lauret P, Fock E, Randrianarivony R N, Manicom-Ramsamy J-F, 2008, Bayesian neural network approach to short time load forecasting. Energy Conversion and Management, 49:1156–66.

Lave M., Kleissl J., 2010, Solar variability of four sites across the state of Colorado, Renewable Energy, vol. 35 pp. 2867-2873

Le ND, Zidek JV., 2006, Statistical analysis of environmental space-time processes. Springer, ISBN 0-387-26209-1; 2006.

Lorenz E, Hammer A, Heinemann D., 2004, Short term forecasting of solar radiation based on satellite data. In: EUROSUN 2004 (ISES Europe Solar Congress), 841-8

Lorenz, E., Heinemann, D., Wickramaratne, H., Beyer, H.G., Bofinger, S., 2007, Forecast of ensemble power production by grid-connected PV systems. In: Proc. 20th European PV Conference, Milano, Italy.

Lorenz, E., Hurka, J., Heinemann, D., Beyer, H., 2009, Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, , 2 (1), 2–10.

Lorenz E., Remund J., Müller S.C, Traunmüller W., Steinmaurer G., Pozo D., Ruiz-Arias JA., Fanego VL, Ramirez L., Romeo M.G, Kurz C., Pomares L.M, Guerrero C.G, 2009, Benchmarking of Different Approaches to Forecast Solar Irradiance, 24th European Photovoltaic Solar Energy Conference, Hamburg, Germany, pp. 21-25

Lorenz E, Heinemann D., 2012, Prediction of solar irradiance and photovoltaic power. In Ali SAYigh E, editor. Comprehensive renewable energy. Oxford: Elsevier, p239-92

Lucheroni, C., 2007, Resonating models for the electric power market. Physical Review E , 76 (5), 9831.

Lütkepohl, H, 1991, Introduction to Multiple Time Series Analysis. Springer-Verlag, , Berlin.

Maher Chaabene, Mohsen Ben Ammara, 2008, Neuro-fuzzy dynamic model with Kalman filter to forecast irradiance and temperature for solar energy systems, Renewable Energy 33, 1435–1443

Marquez, R., Coimbra, C.F.M., 2013a. Intra-hour DNI forecasting methodology based on cloud tracking image analysis. Sol Energy, 91 (6), 327–336.

Marquez, R., Gueorguiev, V.G., Coimbra, C.F.M., 2013, Forecasting of global horizontal irradiance using sky cover indices. Journal of Solar Energy Engineering, 135 (011017), 5.

Marquez R, Pedro HTC, Coimbra CFM., 2013, Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. Solar Energy,;92:176e88.

Marquez R, Coimbra CFM., 2011, Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. Solar Energy;85(5):746e56.

Martins I.A.C, Ferreira P.M., Ruano A.E., 2010, Estimation and Prediction of Cloudiness from Ground-Based All-sky Hemispherical Digital Images, “in Proceedings of CONTROLO, 9th Portuguese Conference on Automatic control

Mathiesen P, Kleissl J., 2011, Evaluation of numerical weather prediction for intraday solar forecasting in the continental United States. Solar Energy;85(5):967e77.

Mellit A, Benghanem M, Arab AH, Guessoum A., 2005, A simplified model for generating sequences of global solar radiation data for isolated sites: using artificial neural network and a library of Markov transition matrices approach. Solar Energy,;79(5):469e82.

Mellit A, Eleuch H, Benghanem M, Elaoun C, Pavan AM., 2010, An adaptive model for predicting of global, direct and diffuse hourly solar irradiance. Energy Conversion and Management,;51(4):771e82.

Mlawer, E., Taubman, S., Brown, P., Iacono, M., and Clough, S., 1997, Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophysical Research*, 102:14, 16663-16682.

Mlawer, E. and Clough, S., 1997, Shortwave and longwave enhancements in the Rapid Radiative Transfer Model. Proceedings of the 7th Atmospheric Radiation Measurement (ARM) Science Team Meeting. U.S. Department of Energy, CONF-9603149.

Mohandes M., Rehman S., Halawani T.O, 1998, Estimation of global solar radiation using artificial neural networks, *Renewable Energy*, Volume 14, Issues 1–4, Pages 179-184

Moradi I., Mueller R., Alijani B., Kamali G.A, 2009, Evaluation of the Heliosat-II method using daily irradiation data for four stations in Iran, *Solar Energy*, vol. 83, pp. 150–156

Mori, N., Chang, K.-A., 2003. Introduction to MPIV <<http://www.oceanwave.jp/software/mpiv>

Mueller R, Dagestad K, Ineichen P, Schroedter-Homscheidt M, Cros S, Dumortier D, et al. 2004, Rethinking satellite-based solar irradiance modeling: the SOLIS clear-sky module. *Remote Sensing of Environment*, 91(2):160e74.

Nelson, M., Hill, T., Remus, W., O'Connor, M., 1999, Time series forecasting using neural networks: Should the data be deseasonalized first? *Journal of Forecasting*, 18, 359–367.

Pedro HTC, Coimbra CFM., 2012, Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, 86(7):2017e28.

Peña, D., Tiao, G. C. and Tsay, R. S, 2001, *A Course in Time Series Analysis*. Wiley, NewYork.

Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., Vignola, F., 2002, A new operational satellite-to-irradiance model. *Solar Energy*, 73 (5), 307–317.

Perez, R., Ineichen, P., Kmiecik, M., Moore, K., George, R., Renne, D., 2004, Producing satellite-derived irradiances in complex arid terrain. *Solar Energy*, 77 (4), 363–370.

Perez, R., Moore, K., Wilcox, S., Renne', D., Zelenka, A., 2007, Forecasting solar radiation—preliminary evaluation of an approach based upon the national forecast database. *Solar Energy*, 81 (6), 809–812.

Perez R, Kivalov S, Schlemmer J, Hemker Jr K, Renne D, Hoff TE, 2010, Validation of short and medium term operational solar radiation forecasts in the US, *Solar Energy*;84:2161–72.

Perez R., Kivalov S., Schlemmer J., Hemker Jr K., Hoff T.E, 2012, Short term irradiance variability: Preliminary estimation of station pair correlation as a function of distance, *Solar Energy*, Volume 86, Issue 8, Pages 2170-2176

Pfeifer, P. E. and Deutsch, S. J., 1980, A three-stage iterative procedure for space-time modeling. *Technometrics*, 22, 35–47.

Pfeifer, P. E. and Deutsch, S. J., 1980, Identification and interpretation of first order space-time ARMA models. *Technometrics*, 22, 397–408.

Pfeifer, P. E. and Deutsch, S. J., 1981, Variance of the sample space-time autocorrelation function. *J. R. Statist. Soc., B*, 43, 28–33.

Quesada-Ruiz S., Chu Y., Trovar-Pescador J., Pedro H.T.C, Coimbra C.F.M, 2014, Cloud-tracking methodology for intra-hour DNI forecasting, *Solar Energy*, 102:267-275

Rehman S, Ghori SG., 2000, Spatial estimation of global solar radiation using geostatistics. *Renewable Energy*, 21:583e605.

Reikard G., 2009, Predicting solar radiation at high resolutions: a comparison of time series forecasts. *Solar Energy*;83:342–9.

Remund, J., Perez, R., Lorenz, E., 2008, Comparison of solar radiation forecasts for the USA., European PV Conference, Valencia, Spain.

Rigollier C., Lefèvre M., Wald L., 2004, The method Heliostat-2 for deriving shortwave solar radiation from satellite images, *Solar Energy* 77, 159–169

Rojasa I., Valenzuela O., Rojasa F., Guillena A., Herrera L.J., Pomaresa H., Marquez L., Pasadas M., 2008, Soft-computing techniques and ARMA model for time series prediction, *Neurocomputing*, 71, pp. 519–537

Sampson PD, Guttorp P, 1992, Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417):108e19.

Santos JM, Pinazo JM, Canada J., 2003, Methodology for generating daily clearness index values K_t starting from the monthly average daily value K_t . Determining the daily sequence using stochastic models. *Renewable Energy*;28(10):1523e44.

Sfetsos A., Coonick A.H, 2000, Univariate and Multivariate forecasting of hourly solar radiation with artificial intelligence techniques, *Solar Energy*, 68, pp. 169–178.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Wang, W., and Powers, J., 2007, A description of the advanced research WRF version 2., NCAR Technical Note NCAR/ TN-468 +STR.

Stephens, G., 1984, The parameterization of radiation for numerical weather prediction and climate models.. *Monthly Weather Review*. 112, 826-867.

Tsay, R.S., 2005, *Analysis of Financial Time Series*, 2nd edn. A John Wiley & Sons, INC., Hoboken, New Jersey.

Voyant C, Muselli M, Paoli C, Nivet M-L., 2011, Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation. *Energy*, 36(1):348e59.

Voyant C., Muselli M., Paoli C., Nivet M-L, 2012, Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation, *Energy*, 39, pp. 341–355

Voyant C., Marc Muselli, Paoli C., Nivet M L, 2013, Hybrid methodology for hourly global radiation forecasting in Mediterranean area, *Renewable Energy*, Volume 53, 2013, Pages 1-11

Wu Ji, Keong Chan Chee, , 2011, Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN, *Solar Energy*, Volume 85, Issue 5, Pages 808-817

Yang D, Jirutitijaroen P, Walsh WM., 2012, Hourly solar irradiance time series forecasting using cloud cover index. *Solar Energy* 86(12):3531e43.

Yang, D., Dong, Z., Nobre, A., Khoo, Y.S., Jirutitijaroen, P. Walsh, W.M., 2013, Evaluation of transposition and decomposition models for converting global solar irradiance from tilted surface to horizontal in tropical regions. *Sol. Energy*, 97, 369–387

Yang D., Gu C., Dong Z., Jirutitijaroen P., Chen N., 2013, Walsh W.M, Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging, *Renewable Energy* 60 (2013) 235e245

Yang D., Dong Z., Reindl T., Jirutitijaroen P., Walsh W.M., 2014, Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage, *Solar Energy*, Volume 103, Pages 550-562

Yousefizadeh H., Zilouchian A., 2001., Neural Network Architectures, Chap. 3, Intelligent Control Systems Using Soft Computing Methodologies, Edited By A. Zilouchian, Mo Jamshidi, CRC Press.

Zarzalejo LF, Ramirez L, Polo J., 2005, Artificial intelligence techniques applied to hourly global irradiance estimation from satellite-derived cloud index. *Energy*,;30(9):1685e97.

Zhang GP., 2003, Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* ,50:159e75.

Zhang, G.P., Qi, M., 2005, Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160, 501–514.

Chapitre 3. Classification de la variabilité de l'ensoleillement global

Table des matières

I. Les données de mesures	Erreur ! Signet non défini.
I.1. Présentation des données de mesures.....	Erreur ! Signet non défini.
I.2. Dépouillement des données.....	Erreur ! Signet non défini.
II. Classification des journées de flux solaire	Erreur ! Signet non défini.
II.1. Aperçu des travaux sur les classifications du flux solaire.....	Erreur ! Signet non défini.
II.2. La classification du flux solaire par méthode k-means	Erreur ! Signet non défini.
III. Analyse à composante principale fonctionnelle	Erreur ! Signet non défini.
III.1. Principe.....	Erreur ! Signet non défini.
III.2. Résultats	Erreur ! Signet non défini.
IV. Les classes de flux solaire	Erreur ! Signet non défini.
IV.1. Classe 1	Erreur ! Signet non défini.
IV.2. Classe 2	Erreur ! Signet non défini.
IV.3. Classe 3	Erreur ! Signet non défini.
IV.4. Classe 4	Erreur ! Signet non défini.
V. Analyse statistique des classes obtenues	Erreur ! Signet non défini.
V.1. Séquence de classes	Erreur ! Signet non défini.
V.2. Pourcentage des types de journées	Erreur ! Signet non défini.
V.3. Persistance des classes.....	Erreur ! Signet non défini.
V.4. Discussion et comparaison inter-site des classes	Erreur ! Signet non défini.
VI. Conclusion	Erreur ! Signet non défini.
Références	Erreur ! Signet non défini.

Préliminairement à l'élaboration d'un modèle spatio-temporel de prévision de l'ensoleillement, nous avons cherché à effectuer une étude spatio-temporelle de la variabilité de l'ensoleillement global sur nos trois stations. Bien que les saisons soient des échelles temporelles clés dirigeant la météorologie tropicale, la plus grande variabilité du rayonnement solaire se produit à l'échelle journalière. Une démarche permettant de décrire la variabilité du flux solaire est l'identification de différents types de régimes. Pour cela, une classification des journées de flux solaire est effectuée, basée sur la méthodologie k-means. Deux méthodes de classification existent : la classification supervisée et la classification non supervisée à laquelle appartient la méthode k-means. La différence entre les deux situations est la connaissance des classes. Dans le cas non supervisé, on fait une recherche à l'aveugle. La question du nombre de groupe est donc centrale. Selon les travaux de recherches cités dans la littérature, un certain nombre de classes est déterminé. Afin de justifier et valider cette hypothèse, nous effectuons une analyse à composante principale fonctionnelle pour notre base de données de flux solaire. Les techniques basées sur l'ACP fonctionnelle permettent d'avoir une visualisation, une quantification, une détection de la variabilité et une analyse des harmoniques des données solaires. A partir de l'analyse à composante principale fonctionnelle, on peut également effectuer une analyse des regroupements de points autour des axes fonctionnels permettant de distinguer et de déterminer le nombre de classes. Cette méthode nous permettra de valider la classification effectuée a priori par la méthode k-means. L'hypothèse validée, une comparaison entre les sites sera alors étudiée afin de caractériser la fréquence d'apparition des classes propres à chaque site et de repérer les éventuelles corrélations de variabilité journalière.

I. Les données de mesures

I.1. Présentation des données de mesures

Le laboratoire dispose d'une base de données d'ensoleillement mesuré pour une période allant de 2010 à 2012. L'ensoleillement global est mesuré à partir d'un pyranomètre Kipp et Zonen de type SP Lite dont le temps de réponse est inférieure à la seconde. Il capte les spectres de rayonnement allant de 400 à 1100 nm, sous un angle d'incidence allant de 0° à 90°. Une parfaite réponse montre une sensibilité maximale à un angle d'incidence de 0° (perpendiculaire à la surface du capteur) et une sensibilité nulle à un angle d'incidence de 90° (ensoleillement passant au-delà de la surface du capteur). Pour un angle compris entre 0° et 90° la sensibilité devrait être proportionnelle au cosinus de l'angle d'incidence. La sensibilité est de $10 \frac{\mu V}{W.m^{-2}}$. La sensibilité électrique de la photodiode dépend de la température. La valeur nominale de cette variation est de 0,2% variation par degré celsius. Le SP-LITE a été calibré à une température de 20°C pour des rayonnements naturels. Le SP-LITE est constitué d'une diode qui va capter l'énergie des photons incidents et générer une tension à la sortie, proportionnelle à la quantité de radiation reçue. Le circuit inclut une résistance de mesure pour la photodiode afin de

générer une tension de sortie. La radiation globale en $W.m^{-2}$ incident est calculée en divisant le signal de sortie (tension) par la sensibilité de l'appareil. Le temps de réponse est inférieur à 1 seconde et sa fréquence de mesure est de 1 Hz soit 86400 données par jour.

Les données sont stockées sur une carte mémoire de type compact flash industrielle d'une capacité de 1 Go et la centrale est alimentée par un jeu de batteries. Ce dispositif a une autonomie vérifiée de deux mois. Les capteurs sont connectés à une centrale d'acquisition Campbell Scientific de type CR23X. Les données sont stockées dans la mémoire tampon de la centrale puis sont transférées vers un ordinateur portable. De plus, une liaison téléphonique permet de contrôler et de piloter la chaîne de mesures depuis le laboratoire.

Les mesures sont effectuées sur les sites de Petit-canal Gros-Cap, sur la côte Est de la Guadeloupe ($16^{\circ}38N$, $61^{\circ}49W$) en bord de falaise, et de l'Université à Fouillole ($16^{\circ}26N$, $61^{\circ}24W$). D'après la classification de C. Brévignon (2003) le site de Petit-Canal se trouve dans le régime des côtes au vent, le vent y souffle en permanence. Dans cette zone la formation nuageuse est majoritairement due à l'advection des masses d'air marin. La station de Fouillole est soumise au régime continental insulaire, dans cette zone le vent souffle moins fort que sur le régime des côtes au vent. La formation nuageuse est principalement due à la convection des masses d'air. Cette base de données existante a été complétée en introduisant un nouveau site de mesures dont l'implantation a fait l'objet d'un choix judicieux, vis-à-vis de la dynamique des nuages. Les vents dominants étant l'Alizé (vent d'Est), une nouvelle station de mesure a été installée à l'extrémité Est de l'île de la Désirade ($16^{\circ}31N$, $61^{\circ}55W$). Ce site étant le plus à l'Est de l'archipel Guadeloupéen, il présente un intérêt particulier vis-à-vis d'une analyse spatiotemporelle. Ce site étant installé sur la côte Est, se trouve également dans le régime des côtes au vent. Dans cette zone la formation nuageuse est majoritairement due à l'advection des masses d'air marin. L'approche envisagée dans ce travail, est d'établir à partir d'un site pilote, des corrélations spatio temporelles qui seront utilisées ensuite dans la prédictibilité de la ressource solaire.

I.2. Dépouillement des données

Les données étant mesurées à une fréquence de 1Hz soit à chaque seconde, sont au nombre de 86400 données par jour soit dans un mois 2 592 000 données à traiter ce qui est équivalent à 31 536 000 de données à traiter par site pour une année. Avant l'analyse proprement dite des données, il a été nécessaire de procéder à une détection d'erreur de mesure et de corrections de ces dernières. Par ailleurs, des séquences de données peuvent être manquantes dues à des défaillances du système de mesure. Un dépouillement fin des données a permis de détecter et d'identifier les périodes de mesures avec des données inexploitable. Suite à cette opération les données sont archivées suivant une nomenclature permettant de dissocier le site, l'année et le mois correspondant. Pour le site de Fouillole nous pouvons compter 98,90% des données enregistrées sur la totalité de l'année 2010, 94,52% des données pour l'année 2011, et 100% des données en 2012. Par conséquent, nous possédons une totalité

de 97,81% des données de mesures sur nos trois années pour le site de Fouillole. Concernant le site de Petit-canal, nous avons 100% des données enregistrées sauf en 2010 avec 87,12% soit une totalité de 95,71% des données de mesures sur nos trois années. Pour La Désirade, nous avons 100% des données pour 2010, 2011 et 2012. Ces séquences de données sont donc bien exploitables pour les années 2010 à 2012 sur nos trois sites de mesures.

II. Classification des journées de flux solaire

II.1. Aperçu des travaux sur les classifications du flux solaire

Plusieurs études se sont penchées sur la classification de jours typiques du rayonnement solaire, généralement en utilisant une approche paramétrique.

Boullier et Le Chapellier(1984) présente une classification basée sur la méthode de classification hiérarchique Ward en utilisant la distance Euclidienne, afin de classer les données climatiques journalières. Cette étude se base sur douze paramètres (le rayonnement horizontal, température de l'air, vitesse du vent, humidité, nébulosité, la pression...) qui permet une analyse des transitions entre les périodes diurnes et nocturnes. Dans ce travail, 4 classes dont 9 sous-classes sont obtenues. La méthodologie proposée est difficile à utiliser car elle a besoin de plusieurs paramètres enregistrés qui sont rarement mesurés.

Aranovitch et al (1985) ont élaboré une classification composée de 3 groupes de jours: ciel clair, partiellement obscurci et ciel nuageux (opaque) afin de synthétiser l'ensoleillement globale et d'étudier les performances du système PV à Ispra (Italie). Les paramètres utilisés étaient la moyenne mensuelle de l'ensoleillement global horizontal, la température et la vitesse du vent.

Louche et al (1991) présente une méthode de classification des du rayonnement solaire directs en utilisant quatre paramètres journaliers caractéristiques : rayonnement normal direct, dimension fractale, et deux paramètres statistiques basés sur le coefficient de variation. Quatre classes ont été caractérisées.

Fabero et al (1997), ont distingués neuf jours « typiques » en décomposant chaque jour étudié en trois périodes. Ils ont utilisé le rayonnement horizontal global enregistré avec un pas de temps d'acquisition de 10 min, et ont étudié les transitions entre les périodes pour un jour donné afin d'évaluer le potentiel solaire.

Muselli et al (2000) ont défini des paramètres élaborés à partir de l'indice de clarté horaire calculé à partir du rayonnement horizontale et enregistré sur deux sites côtiers méditerranéens en Corse. La classification utilise la méthode de Ward, et les résultats sont confirmé par analyses discriminantes . Deux études basées sur différents pas de temps (mensuel et annuel) ont été menées. Les résultats

montrent que les jours enregistrés sont groupés en trois, quatre ou cinq pour un pas de temps mensuel, et pour un pas de temps annuel sur environ 4 années de données, 3 jours typiques météorologiques sont obtenus.

Maafi et Harrouni (2003) traite de la modélisation fractale du rayonnement solaire journalier mesuré avec un temps d'échantillonnage de 10 min sur deux sites en Algérie. Les fractales journaliers et l'indice de clarté ont été utilisés pour proposer un modèle de classification qui conduit à trois classes typiques. Ces classes correspondent à des types de journées à ciel clair, ciel partiellement nuageux et ciel complètement nuageux.

Harrouni et al (2005) utilise également une méthode basée sur les dimensions fractales et trouve également trois jours typiques ayant les mêmes caractéristiques.

Soubdhan et al (2009) classe les distributions journalières de l'indice de clarté k_t en utilisant une méthode basée sur les distributions de Dirichlet sans aucune hypothèse paramétrique. La méthode est appliquée aux mesures du rayonnement solaire réalisées en Guadeloupe sur le site de Fouillole. Les résultats montrent quatre classes distinctes de distributions correspondant à différents types de journée (ciel clair, ciel clair intermédiaire, ciel nuageux et ciel nuageux intermédiaire).

Badosa et al (2013) ont identifié des régimes de rayonnement solaire en utilisant la méthode de classification k-means basées sur les paramètres d'indice de clarté, variation relative cumulative totale (TRC), variation relative absolue (RARC). Ils ont trouvé cinq régimes distincts : journée à ciel clair, journée à ciel nuageux, journée avec des périodes nuageuses dans la matinée, journée avec des périodes nuageuses dans l'après-midi, journée avec variabilité aléatoire.

Soubdhan et al (2014) ont appliqué un algorithme de classification basé sur la méthode k-means sur les densités de probabilité de l'indice de clarté. Ils ont obtenu quatre types classes distinctes définies et déterminées pour chaque tranche horaire : 7h-9h, 9h-11h, 11h-13h, 13h-15h, 15h-17h. Il apparaît que pendant la tranche de 9h à 13h; les classes obtenues correspondent particulièrement à des journées avec des conditions de ciel clair.

D'autres auteurs ont abordé la classification de régime pour des applications d'énergie : (Adelard et al., 2000; David et al., 2010).

Pour classer nos journées de flux solaire, nous utiliserons la méthode k-means. Dans cette partie, nous définirons succinctement ce qu'est une classification k-means. Ensuite, nous présenterons nos résultats.

II.2. La classification du flux solaire par méthode k-means

II.2.1 Principe

Les algorithmes de classification peuvent être classés comme indiqué ci-dessous:

- Classification exclusive
- Classification avec chevauchement
- Classification hiérarchique
- Classification probabiliste

Dans le premier cas, les données sont regroupées de manière exclusive, de sorte que si un certain nombre de données appartient à un groupe précis, alors il ne peut être inclus dans une autre classe. Au contraire, le second type, le regroupement se chevauche, de sorte que chaque point peut appartenir à deux ou plusieurs groupes.

La classification k-means appartient au premier cas (classification exclusive).

La méthode des "K-means" reste actuellement la méthode la plus utilisée surtout pour les grands fichiers de données. Le principe de la méthode des "k-means" c'est que la classification se fait sur la base du critère des plus proches voisins. Celui-ci signifie que chaque individu est affecté à une classe s'il est très proche de son centre de gravité.

Ayant initialisé k centres de classes par tirage aléatoire, tous les individus sont affectés à la classe dont le centre est le plus proche au sens de la distance choisie (en principe, euclidienne pour cette méthode). Dans une deuxième étape, l'algorithme calcule des barycentres de ces classes qui deviennent les nouveaux centres. Le procédé (affectation de chaque individu à un centre, détermination des centres) est itéré jusqu'à convergence vers un minimum (local) ou un nombre d'itérations maximum fixé. (Hartigan and Wong (1979)).

Il n'y a pas de critère absolu pour décider de ce qui constitue un bon regroupement. Par conséquent, c'est l'utilisateur qui doit fournir ce critère, de telle sorte que le résultat du regroupement sera fonction de ses besoins. Nous choisissons alors de mettre un terme à l'itération dès que par observations, nous trouvons que les types de journées correspondants à chaque classe sont bien distinctes entre elles et soient caractéristiques des 4 classes décrites dans la littérature par Soubdhan et al. (2009) : journées à ciel clair, journées à ciel clair intermédiaire, journées à ciel nuageux, journées à ciel nuageux intermédiaire.

II.2.2 Résultats

Les figures ci-dessous (Figure 3-1) représentent les courbes de flux solaire moyennes des 4 classes résultant de l'algorithme appliquant la fonction k-means pour nos trois sites de mesures sur 3 années

de données (2010, 2011, 2012). Nous avons également testé l'algorithme pour 5 classes, cependant on constatait que deux classes parmi les 5 se chevauchaient. Par conséquent, les 5 classes n'étaient pas distinctes. Nous validerons l'hypothèse du nombre de classe plus précisément par méthode d'analyse à composante principale fonctionnelle.

Nous avons pu constater que nous retrouvons les mêmes caractéristiques de classes de journée décrites dans Soubdhan et al. (2009). La classe 3 citée dans cet article correspond pour nous à la classe 4 et donc la classe 4 correspond à la classe 3, les classes 1 et 2 restent similaires.

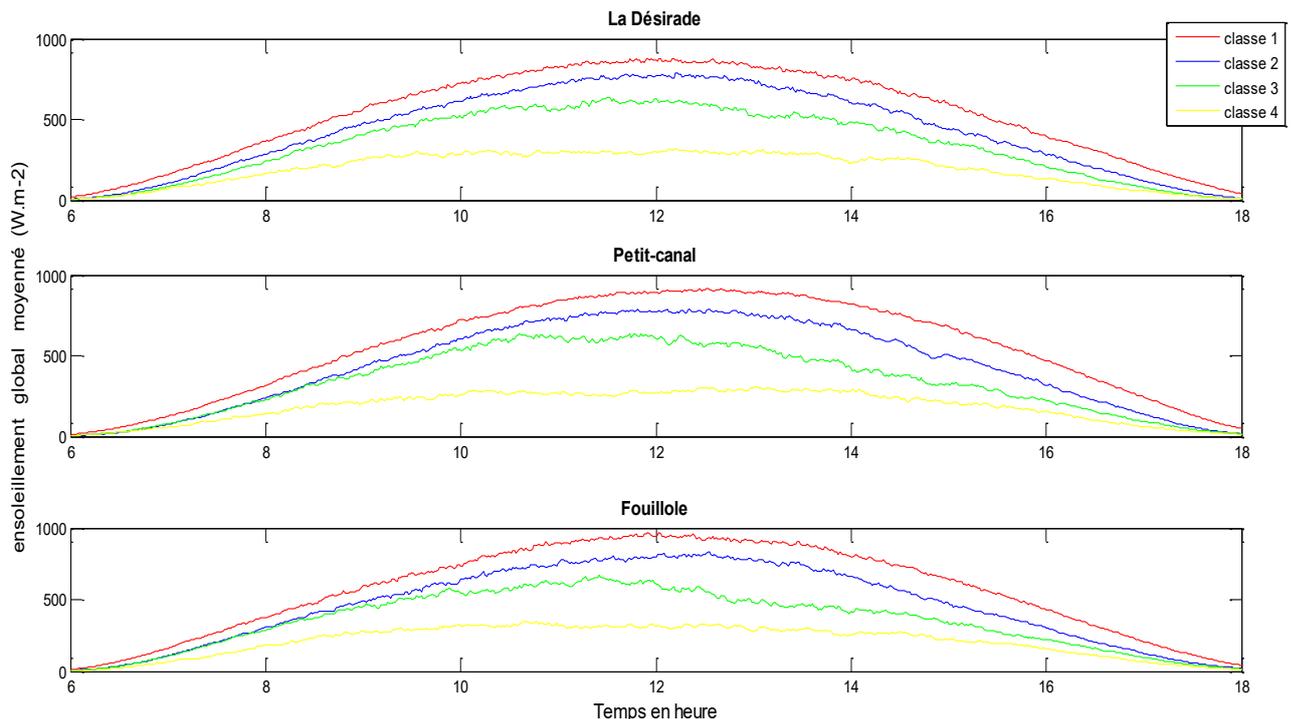


Figure 3-1 Courbes de flux solaire moyennes, représentatives des 4 classes pour le site de La Désirade, Petit-canal et Fouillole

III. Analyse à composante principale fonctionnelle

III.1. Principe

L'objectif de l'analyse à composante principale fonctionnelle dans ce chapitre est d'appuyer l'hypothèse du nombre de classe choisi pour la classification de l'ensoleillement global journalier basée sur l'algorithme kmeans, effectuée précédemment. La classification k-means étant un cas non supervisé, la question du nombre de classes à déterminer est donc centrale.

Dans le cadre de l'ACP (analyse à composante principale) fonctionnelle à contrario de l'ACP classique, les caractéristiques de la variabilité du flux solaire sont considérées comme des courbes continues. La principale idée de cette extension est de remplacer les vecteurs par des fonctions, des

matrices par des opérateurs linéaires compacts, des matrices de covariance par des opérateurs de covariance et des produits scalaires dans l'espace des vecteurs par des produits scalaires dans l'espace L^2 (Shang 2014). Pour calculer les harmoniques (composantes principales) et leurs scores, il existe au moins trois approches :

- discrétisation : l'ACP fonctionnelle est effectuée d'une façon semblable à l'ACP classique, sauf qu'il est nécessaire de renormaliser les vecteurs propres et les interpoler avec un lisseur approprié.
- expansion de fonctions de base : la deuxième approche implique l'expression d'un processus stochastique comme une combinaison linéaire de fonctions de base.
- approximation numérique: cette approche consiste dans l'approximation de composants principaux fonctionnels par des règles de quadrature.

La deuxième approche sera appliquée pour le calcul des harmoniques basée sur la méthodologie de Chebana et al (2012). On peut sectionner le processus d'analyse en deux étapes :

- La transformation des données discrètes en données fonctionnelles. L'objectif de cette étape consiste à préparer les données pour être utilisée dans le contexte de l'AFD (analyse de données fonctionnelles) en utilisant des techniques de lissage de données par combinaison linéaires de fonctions de bases (Fourier, Spline, polynomiale...).
- L'application de l'analyse à composante principale fonctionnelle pour réduire la dimensionnalité des variables d'origine et capturer le maximum des variances des données. Après conversion des données en fonctions, l'ACP fonctionnelle (ACPF) permet de trouver de nouvelles fonctions qui révèle le type de variations le plus important dans la courbe des données. La méthode d'ACPF maximise la variance des scores sujets à des contraintes orthonormales. Elle décompose les données fonctionnelles centrées en termes de base orthogonale. Chaque composante principale est alors représentée par une fonction continue.

Les avantages de cette méthode résident dans le fait que, les données fonctionnelles sont un ensemble de courbes continues. Dans le cas d'analyses temporelles, souvent les expériences sont des mesures d'un même phénomène à des temps différents et peuvent présenter une irrégularité, l'ACPF permet d'analyser des données même si les mesures ne sont pas effectuées aux mêmes instants pour chaque variable (interpolation par les fonctions de base). De plus, dans le cas temporel, traiter les données manquantes est possible lorsque celles-ci sont peu nombreuses (interpolation à l'aide des fonctions de base). L'analyse donne une formulation analytique de chaque composante principale fonctionnelle, sous la forme d'une combinaison linéaire des fonctions de base, par conséquent la continuité des courbes rend plus simple par analyse harmonique, l'extraction des caractéristiques principales (amplitude, période). Notre choix pour l'ACP fonctionnelle s'explique aussi par le fait qu'elle utilise

une méthode de lissage qui consiste en une combinaison linéaire de fonctions de bases apportant ainsi une information supplémentaire sur les données par rapport à l'ACP classique.

III.2. Résultats

Les harmoniques sont les principales composantes obtenues par maximisation de la variance des scores correspondants (voir Chebana et al (2012) et Annexe). Ainsi la variance expliquée diminue en proportion au fur et à mesure que les harmoniques sont extraites. La proportion de variance totale cumulée à travers les différentes composantes atteindra les 100% si le processus d'extraction est mené à terme, c'est à dire si le nombre de composantes extraites équivaut au nombre de variables initialement soumises à l'analyse. Néanmoins, l'analyse en composantes principales est utilisée dans une perspective de réduction des données. Les résultats de taux de variance obtenus pour les quatre premières harmoniques sont présentés dans le tableau (Tab.1).

	Harmonique 1	Harmonique 2	Harmonique 3	Harmonique 4
La Désirade	53%	8%	5%	3%
Petit-canal	54%	9%	6%	4%
Fouillole	49%	9%	6%	4%

Tableau 3-1 Taux de variance des quatre harmoniques obtenues par ACPF sur les trois stations

Selon le tableau Tab.3.1, pour le site de La Désirade, le taux de variance total obtenu à partir de quatre harmoniques s'élève à 69%. Pour le site de Fouillole, le taux de variance total obtenu est de 68%. On obtient 73% de variance totale à partir de quatre harmoniques pour le site de Petit-canal. Les quatre harmoniques obtenues à partir des scores résument en quatre fonctions les informations les plus importantes de la matrice correspondant à 3 années de flux solaires par pas de 1 minute. Les caractéristiques des données solaires sont représentées par une simplification maximale du phénomène réel. En d'autres termes, elles représentent les types de journées ayant les plus fortes probabilités d'apparition dans la séquence totale de mesures (3ans). On peut constater qu'avec quatre harmoniques, nous obtenons une totalité jusqu'à 73 % de variance du flux solaire. La cinquième harmonique présente un taux de variance très faible pour les trois stations, soit inférieure à 3% .

Une fois les données discrètes sont transformées en courbe, elles peuvent être analysées selon une structure fonctionnelle. Par projection des données fonctionnelles, de nouvelles variables appelées scores vont résumer l'information contenue dans les données de mesures. En traçant les scores 2 en fonction des scores 1 obtenus, on peut effectuer une analyse des regroupements de points autour des

axes fonctionnels. En effet, on considère les scores de composante principale comme les substituts de courbes fonctionnelles (Jones et Rice 1992), ce qui permet d'appliquer un algorithme de classification aux deux premiers scores, pour révéler le sous-groupe homogène dans un ensemble de données fonctionnelles (Shang 2014). L'algorithme de classification des groupements de points est basé sur la méthode k-means.

Nous vérifions ensuite, si chacun de ces groupements de points correspond à une classe.

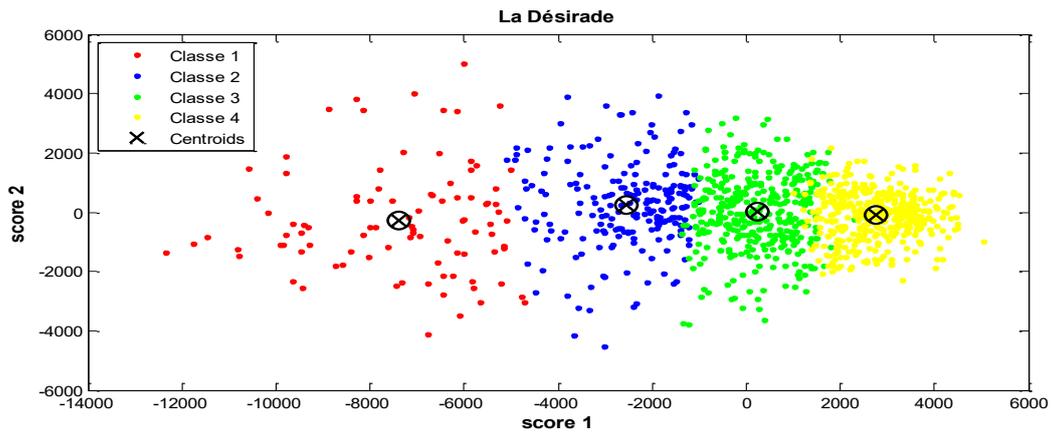


Figure 3-2. Plan des scores des deux premiers scores et classification des groupements de points (La Désirade)

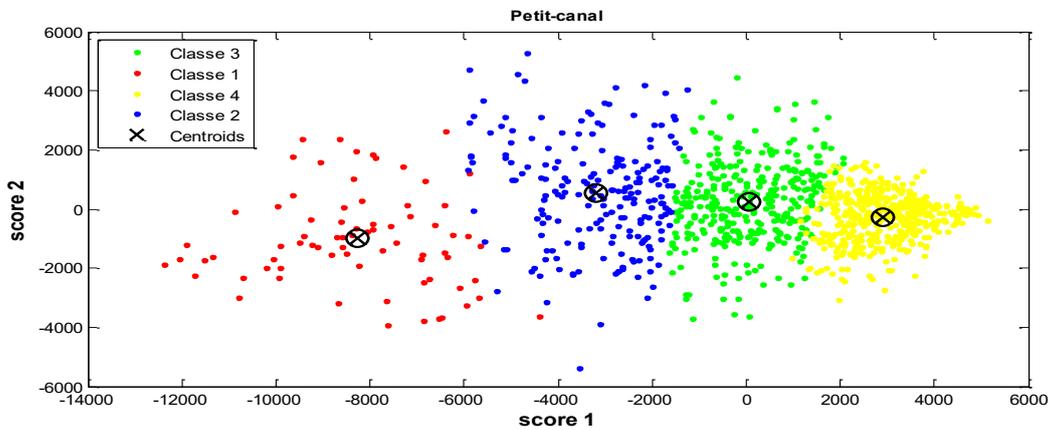


Figure 3-3. Plan des scores des deux premiers scores et classification des groupements de points (Petit-canal)

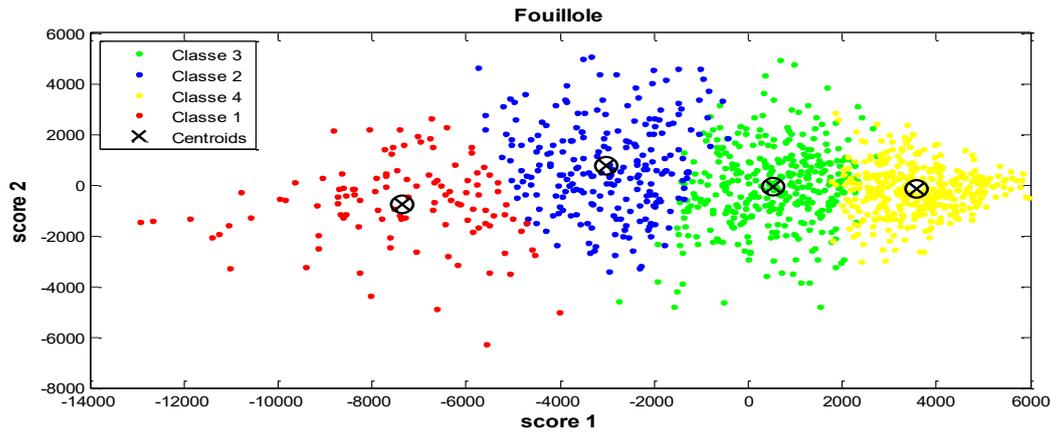


Figure 3-4. Plan des scores des deux premiers scores et classification des groupements de points (Fouillole)

Les plans scores 1/scores2 pour chacun des sites ont permis de constater quatre regroupements de points bien nets, classés par la méthode k-means (Figure 3-2 à 3-4). Nous vérifions à quoi correspondent chaque groupement de points. On retrouve effectivement les classes de journées identifiées en section I.2.2, lorsque nous vérifions à quelles journées correspondent chaque groupe de points.

Les résultats de l'analyse à composante principale fonctionnelle, nous ont permis de constater qu'à partir de quatre harmoniques on peut extraire 73% de la variance du flux solaire. Ainsi, sachant que les harmoniques suivantes (cinquième harmonique et plus) présentent un taux de variance très faible soit inférieur à 3%, on peut estimer que quatre harmoniques suffisent à extraire les informations sur la variabilité de l'ensoleillement journalier. L'analyse des regroupements de points placés dans le plan formé par les deux premiers scores et le taux de variance total important des quatre premières harmoniques, appuient l'hypothèse du nombre de classes déterminé en section I.

IV. Les classes de flux solaire

L'hypothèse du nombre de classes étant vérifiée par l'analyse à composante principale fonctionnelle, nous validons la classification par méthode k-means effectuée précédemment distinguant quatre classes. Une description des caractéristiques des classes et leur fréquence d'apparition est discutée dans les sections suivantes.

IV.1. Classe 1

La première classe est composée de courbes de flux solaire ayant une moyenne journalière entre 470 $W.m^{-2}$ et 665 $W.m^{-2}$. Les valeurs maximales de flux solaire peuvent atteindre environ 1100 $W.m^{-2}$. Ses courbes sont représentatives de jours ayant des conditions météorologiques à ciel clair, les moyennes

de flux solaire sont assez importantes. En effet, les passages nuageux ayant pour conséquence la diminution du flux solaire mesurée au sol, ces valeurs moyennes importantes de flux solaire caractérisent une faible fréquence de passage nuageux au cours de la journée. Ainsi, nous pouvons observer sur les courbes représentatives de cette classe (Figure 3-6), une dynamique très lente du flux solaire, que l'on peut également constater en comparant les valeurs de coefficient de variation par rapport aux autres classes (Tab3.2).

$$c_v = \frac{\sigma}{\mu} \quad (3.1)$$

Avec c_v coefficient de variation, σ écart-type et μ la moyenne.

La boîte à moustache est représentée pour chaque 30 minutes, pour des journées correspondant la classe 1. Ces graphiques de boîtes à moustaches permettent non seulement d'avoir une vue synthétique, globale mais aussi une vue locale sur les données (valeurs atypiques). La Figure 3-5 représente les boîtes à moustache du flux solaire pour chaque 30 minutes calculée sur l'ensemble des journées appartenant à la classe 1 pour toutes les stations. Cette figure permet de rendre compte de la médiane, du quartile Q1 qui correspond à 25 % des observations et du quartile Q3 qui correspond à 75 % des observations. Les valeurs dites extrêmes, atypiques, situées au-delà des valeurs adjacentes sont individualisées. Elles sont représentées par des marqueurs (étoile rouge). Les marqueurs étant peu nombreux signifient que la variabilité des valeurs est faible. L'étendue des boîtes à moustache n'étant pas grande montre que les valeurs sont assez homogènes (Le Guen (2002)).

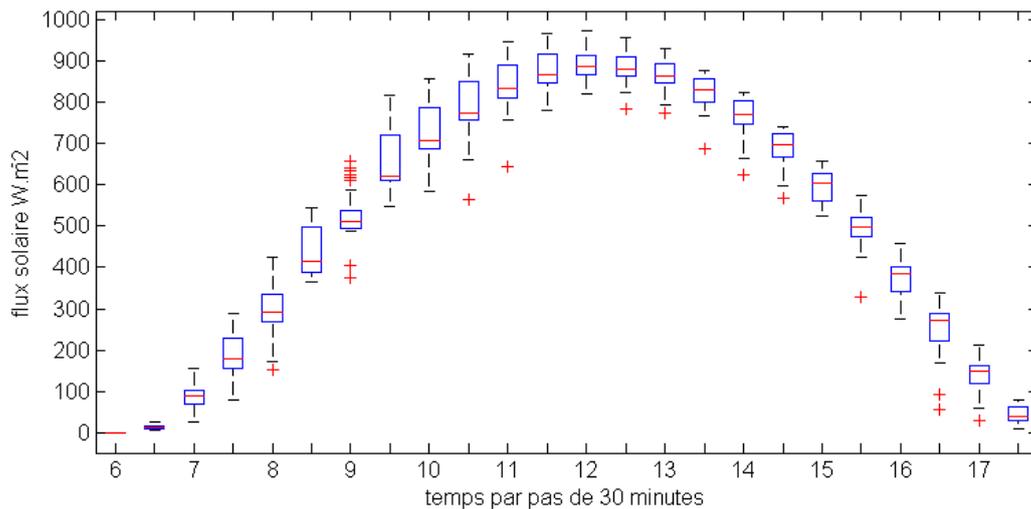


Figure 3-5. Distributions du flux solaire pour la classe 1 pour toutes les stations des données moyennées par pas de 30 minutes.

Classe 1			
	Moyenne journalière (W.m ⁻²)	Ecart-type journalier (W.m ⁻²)	Coefficient de variation journalier
Désirade	[469,3 ; 601,6]	[250,5 ; 365,1]	[0,53 ; 0,61]
Petit-canal	[478,4 ; 629,8]	[266,7 ; 387,7]	[0,56 ; 0,61]
Fouillole	[480 ; 664,7]	[279,9 ; 410,8]	[0,58 ; 0,62]

Tableau 3-2. Statistiques de la variabilité journalière des types de journées de classes 1

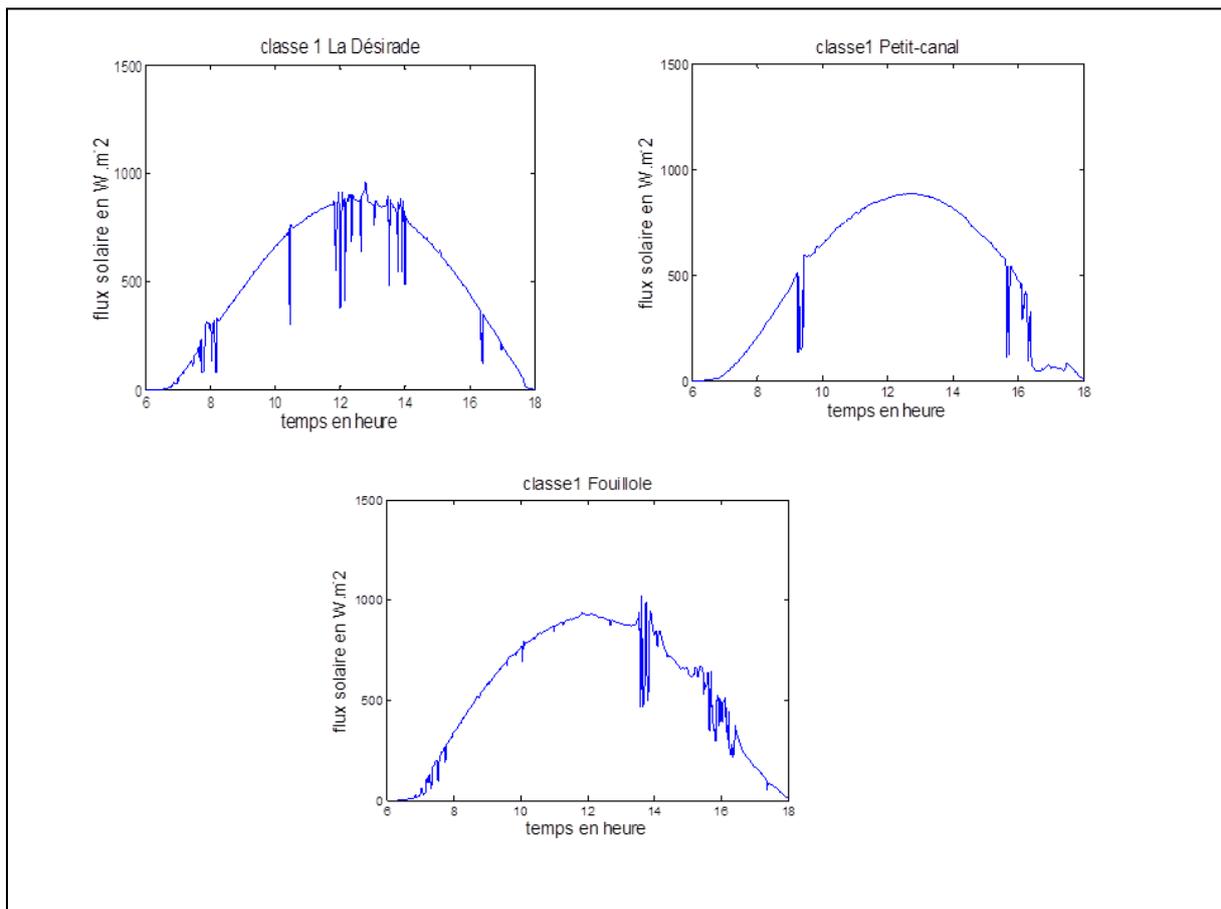


Figure 3-6. Exemples de journées classe 1 pour les trois sites de mesures (La Désirade, Petit-canal, Fouillole)

IV.2. Classe 2

La deuxième classe est composée de courbes de flux solaire ayant en moyenne entre 300 W.m^{-2} et 531 W.m^{-2} . Les valeurs maximales de flux solaire peuvent atteindre environ 900 W.m^{-2} . Ce qui est représentatif de jours ayant un important rayonnement solaire avec quelques passages nuageux. Les fluctuations de flux solaire sont effectivement plus importantes que celles de la classe 1 que l'on peut vérifier grâce au coefficient de variation journalier (comparaison entre Tab3-3. et Tab3.2).

En comparaison avec la classe 1, les courbes représentatives de la classe 2 ont effectivement une dynamique à niveau moyen que l'on peut observer sur la Figure 3-8. Sur la (Figure 3-7), représentant la distribution de flux solaire journalier pour des journées appartenant à la classe 2, les valeurs dites extrêmes, atypiques, situées au-delà des valeurs adjacentes sont plus nombreuses que celles observées pour la classe 1, ce qui signifie que la variabilité est plus importante. Néanmoins, les valeurs restent plus ou moins homogènes en observant la petite taille des boîtes à moustaches (Le Guen (2002)).

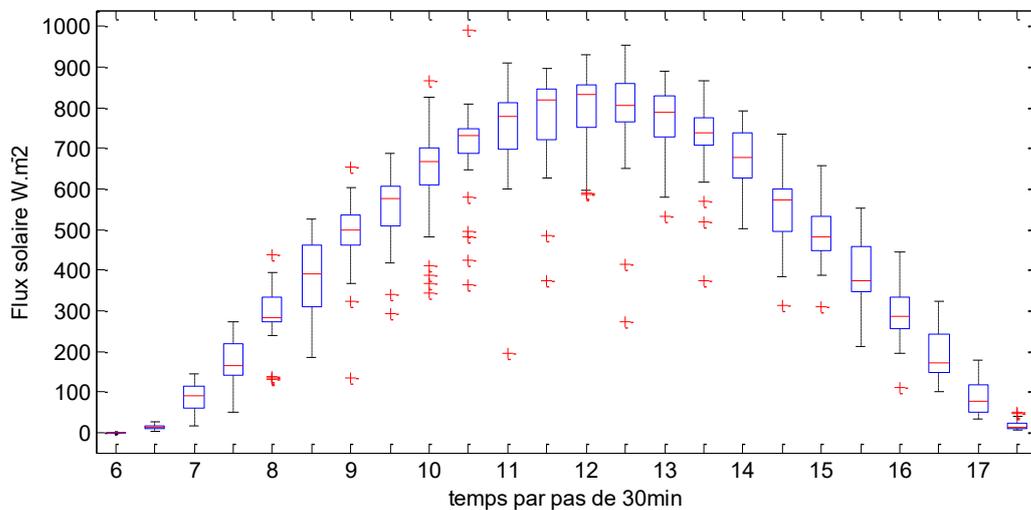


Figure 3-7. Distributions du flux solaire pour la classe 2 pour toutes les stations des données moyennées par pas de 30 minutes.

Classe 2			
	Moyenne journalière (W.m ⁻²)	Ecart-type journalier (W.m ⁻²)	Coefficient de variation journalier
Désirade	[354,3 ; 491,5]	[239,3 ; 402,3]	[0,67 ; 0,82]
Petit-canal	[313,0 ; 506,3]	[241,4 ; 387,2]	[0,77 ; 0,76]
Fouillole	[354,1 ; 531,0]	[229,1 ; 419,6]	[0,65 ; 0,79]

Tableau 3-3 Statistiques de la variabilité journalière des types de journées de classes 2

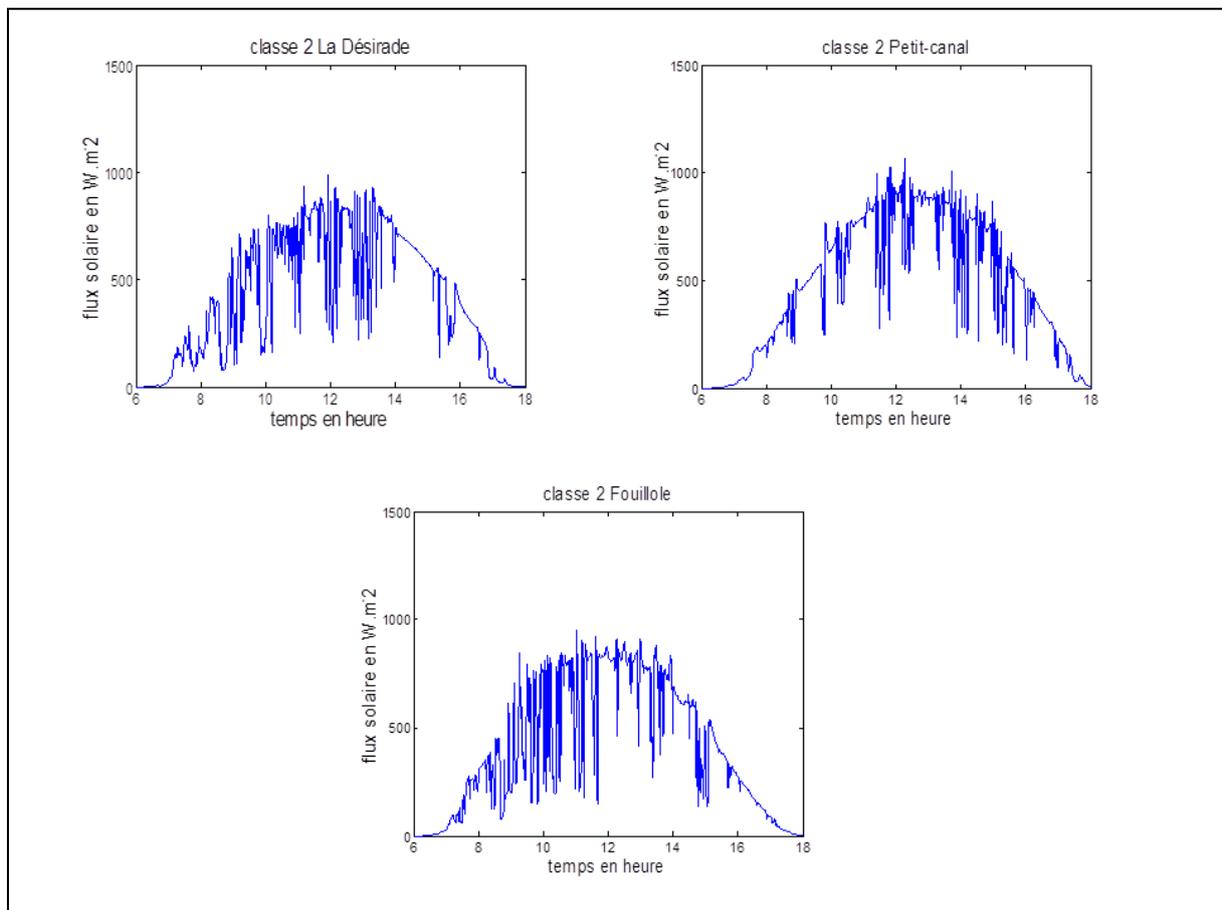


Figure 3-8. Exemples de journées de classe 2 pour les trois sites de mesures (La Désirade, Petit-canal, Fouillole)

IV.3. Classe 3

La troisième classe est composée de courbes de flux solaire ayant une moyenne journalière entre 246 $W.m^{-2}$ et 463 $W.m^{-2}$. Les valeurs maximales de flux solaire peuvent atteindre environ 600 $W.m^{-2}$. Ses courbes de flux solaire sont représentatives d'un ensoleillement significatif, toutefois avec un nombre important de passages nuageux (en général de petites masses nuageuses avec une vitesse de passage rapide). Les jours correspondant à cette classe présente une courbe de flux solaire avec une dynamique très importante comme on peut l'observer sur la (Figure 3-10), Les coefficients de variations compris entre 0,74 et 0,78 montrent une variabilité importante de ces types de journées.

La Figure 3-9 représente les boîtes à moustache du flux solaire pour chaque 30 minutes pour toutes les journées appartenant à la classe 3. L'étendue des boîtes à moustache est assez grande par rapport aux distributions des journées appartenant à la classe 1 et 2 ce qui indique que les valeurs sont très hétérogènes (Le Guen (2002)). Cette figure indique également une variabilité importante pour ce type de journée et ce, durant toute la journée particulièrement de 8h30 à 15h30.

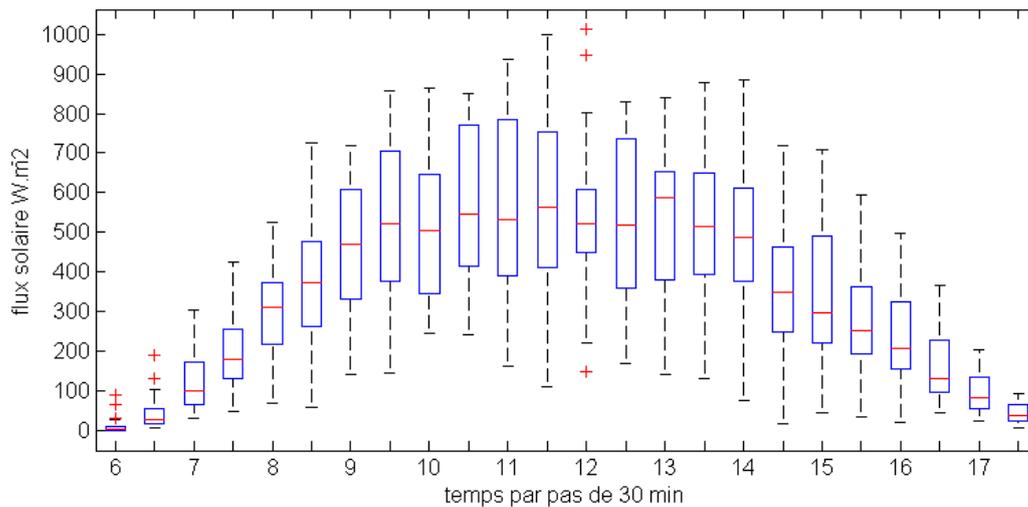


Figure 3-9. Distributions du flux solaire pour la classe 3 pour toutes les stations des données moyennées par pas de 30 minutes.

Classe 3			
	Moyenne journalière (W.m ⁻²)	Ecart-type journalier (W.m ⁻²)	Coefficient de variation journalier
Désirade	[251,3 ; 416,8]	[195,7 ; 326,8]	[0,77; 0,78]
Petit-canal	[248,2 ; 455,19]	[190,2 ; 339,9]	[0,74 ; 0,77]
Fouillole	[246,4 ; 463,0]	[187,8 ; 362,9]	[0,76 ; 0,78]

Tableau 3-4 Statistiques de la variabilité journalière des types de journées de classe 3

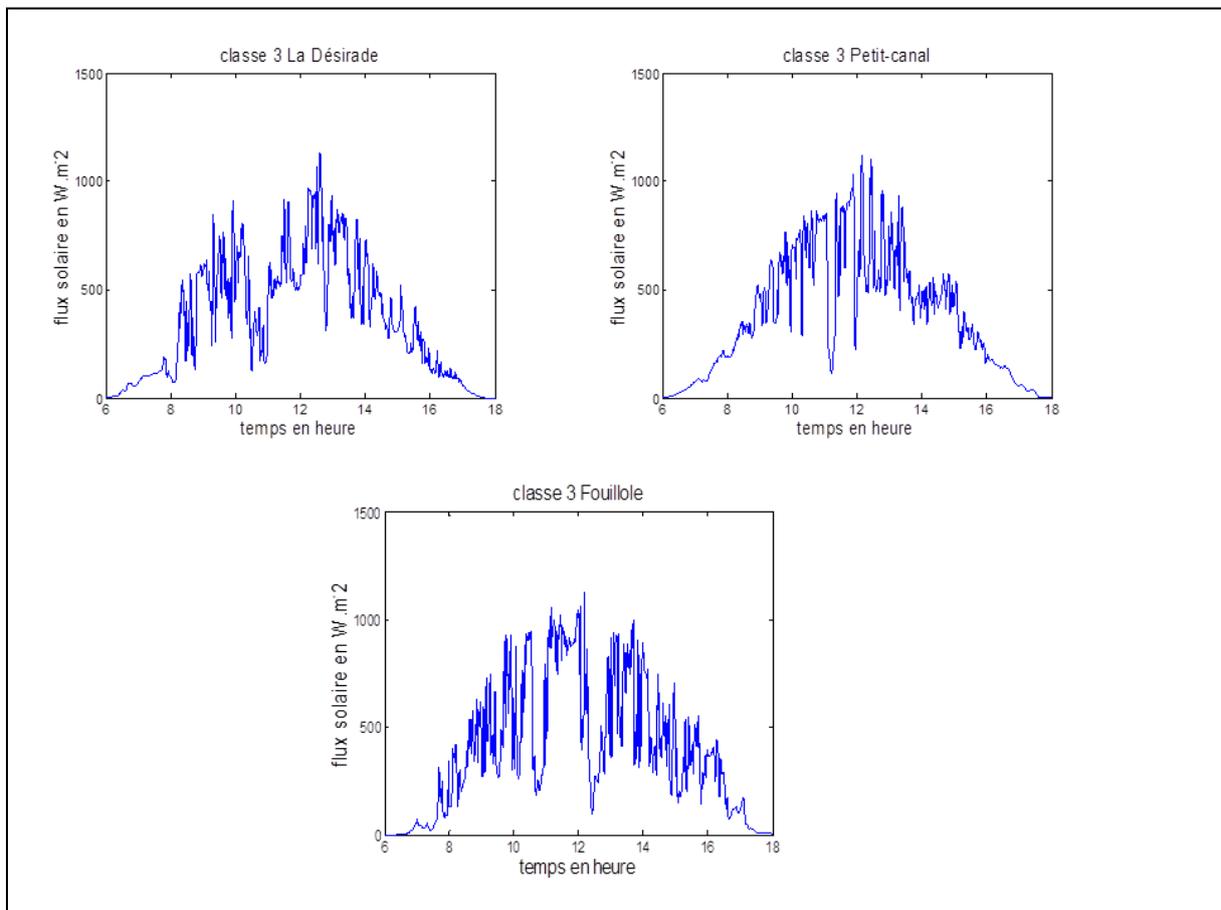


Figure 3-10. Exemples de journées de classe 3 pour les trois sites de mesures (La Désirade, Petit-canal, Fouillole)

IV.4. Classe 4

La quatrième classe est composée de courbes de flux solaire ayant une moyenne journalière entre 32 $W.m^{-2}$ et 331 $W.m^{-2}$. Ses courbes de flux solaire sont représentatives d'un ensoleillement très faible dû à un ciel très nuageux. Les conditions météorologiques des journées de cette classe, présentent des masses nuageuses de grandes tailles ayant une vitesse de passage assez lente. En effet, comme on peut l'observer sur les figures ci-dessous, les variations correspondant aux diminutions de flux solaire ont tendance à durer dans le temps et les courbes de flux solaire sont très basses comparées aux journées appartenant aux autres classes. Les coefficients de variation ayant des valeurs très proches de 1 indiquent une variabilité importante de ces types de journées. Néanmoins, ces journées peuvent présenter une dynamique très lente avec un rayonnement solaire faible, dus à des événements météorologiques caractéristiques de passages de masses nuageuses très épaisses et de taille importante.

La Figure 3-11, représente les boîtes à moustache du flux solaire pour chaque 30 minutes pour toutes les journées appartenant à la classe 3. L'étendue des boîtes à moustache est assez petite en début (7h à 9h) et en fin de journée (16h à 18h), ce qui indique que les valeurs sont homogènes dans ces tranches horaires. Toutefois, les valeurs sont particulièrement hétérogènes entre 10h30 et 15h30, la variabilité de l'ensoleillement est particulièrement importante pour ce type de journée durant cette tranche horaire.

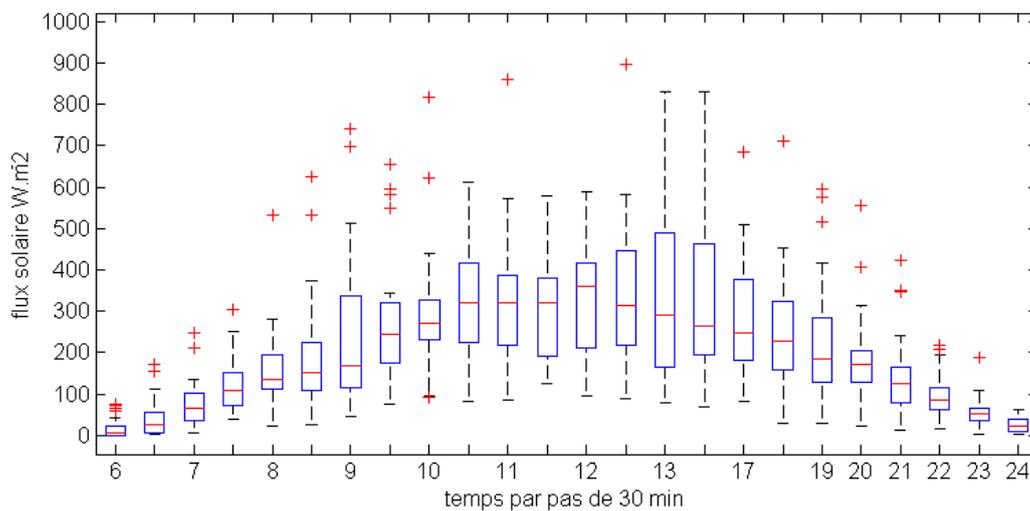


Figure 3-11. Distributions du flux solaire pour la classe 4 pour toutes les stations, des données moyennées par pas de 30 minutes.

Classe 4			
	Moyenne journalière (W.m ⁻²)	Ecart-type journalier (W.m ⁻²)	Coefficient de variation journalier
Désirade	[32,7 ; 308,8]	[31,31 ; 267,65]	[0,96; 0,87]
Petit-canal	[41,37 ; 314,1]	[39,23 ; 261,7]	[0,83; 0,94]
Fouillole	[35,4 ; 331,1]	[23,1 ; 272,9]	[0,65 ; 0,82]

Tableau 3-5 Statistiques de la variabilité journalière des types de journées de classe 4

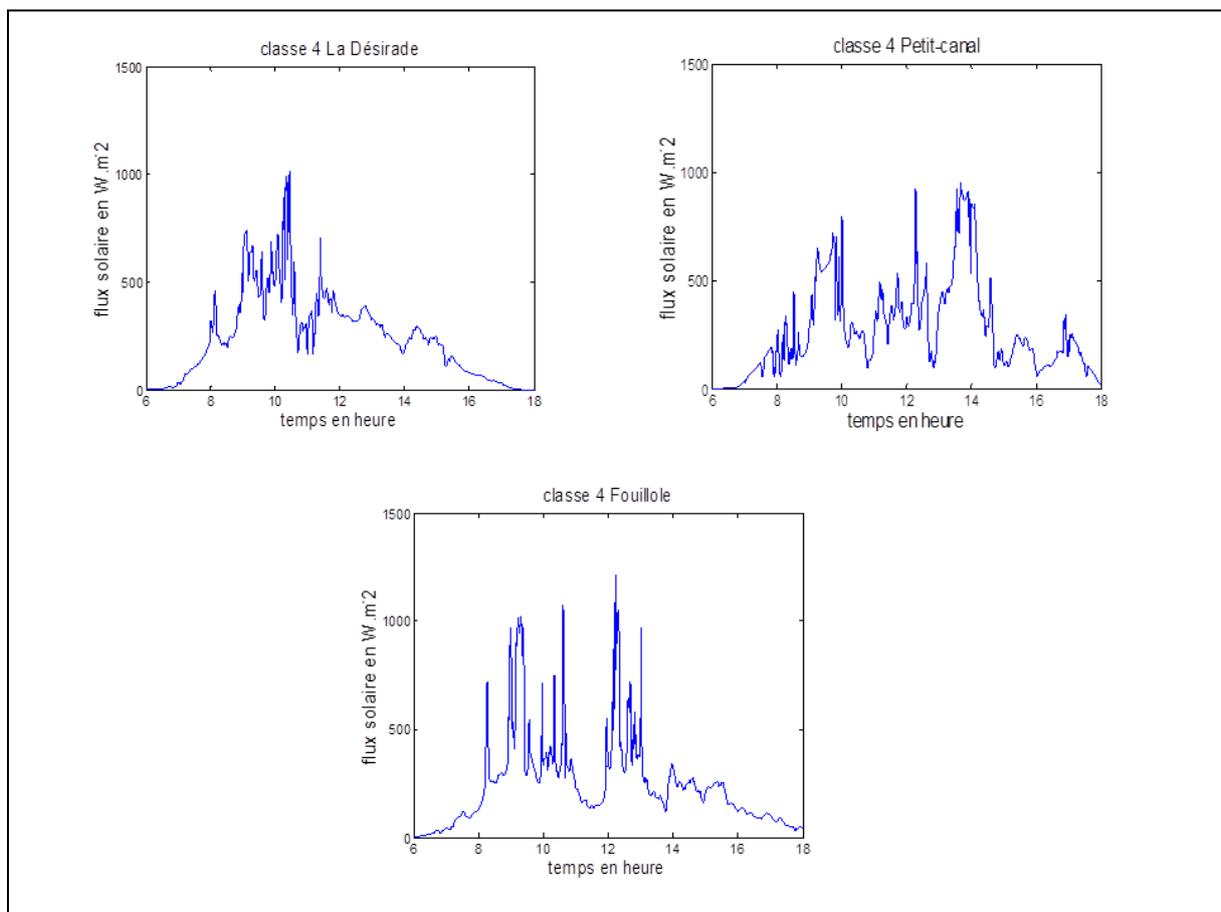


Figure 3-12 Exemple de journées de classe 4 pour les trois sites de mesures(La Désirade, Petit-canal, Fouillole)

Dans les sections suivantes nous chercherons à caractériser les pourcentages de types de journées, la persistance des classes et à analyser les corrélations des classes journalières entre les stations.

V. Analyse statistique des classes obtenues

Les journées étant classifiées en quatre classes, nous pouvons alors remplacer les numéros de jours par le numéro de la classe correspondante. On obtient ainsi, une séquence composée des numéros 1, 2, 3 et 4. Cette séquence apporte de nombreuses informations notamment sur le pourcentage de chacune des classes, le type de jour prédominant, leur évolution durant nos trois années de mesure. De plus, il est intéressant d'observer la transition d'une classe à une autre.

Premièrement, nous chercherons à décrire l'apparition des classes journalières de l'ensoleillement tout au long des années de mesures.

V.1. Séquence de classes

La Figure 3-13, représente les séquences de classes journalières $\{1,2,3,4\}$ obtenues par la classification basée sur la méthode k-means de l'ensoleillement global, sur l'ensemble des trois années de mesures soit 1096 jours.

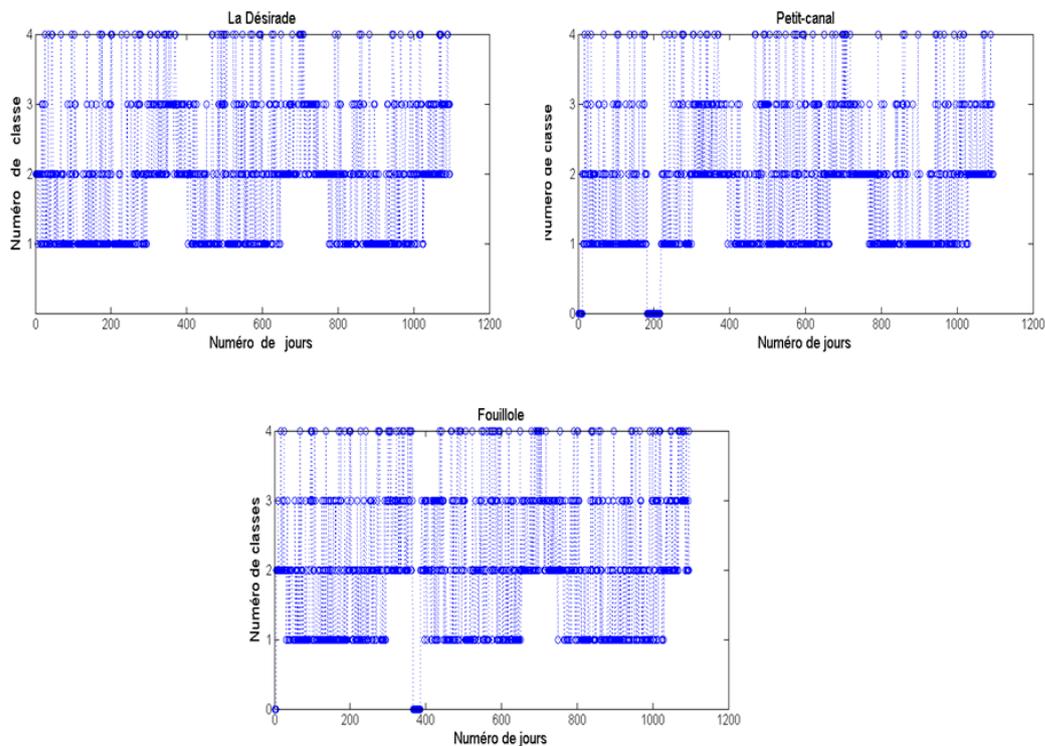


Figure 3-13. Séquence de classes $\{1, 2, 3, 4\}$ sur 1096 jours (année 2010 à 2012) sur le site de La Désirade, Petit-canal et Fouilliole.

On peut observer sur la Figure 3-13 ; en identifiant les numéros de jours au mois correspondant, que la classe 1 n'apparaît pas du mois d'octobre à janvier sur l'ensemble des trois années de mesures. Ce phénomène d'apparition périodique des classes 1 est observable pour les trois stations. La classe 1 n'étant pas obtenue du mois d'octobre à janvier, peut s'expliquer par le fait que la période d'octobre à décembre correspond à la saison d'hivernage aux Antilles, où l'ensoleillement est moins important, avec des conditions météorologiques assez nuageuses. Le mois de janvier est une période de transition entre l'hivernage et le carême, qui se caractérise par un ciel nuageux intermédiaire. On peut également constater que la classe 4 bien que peu fréquente au cours de l'année, n'apparaît pas de février à mars au cours des trois années qui correspond à une période de l'année bien ensoleillée. Les classes 2 et 3 sont présentes tout au long de l'année. Toutefois, la classe 3 se produit principalement de mars à juin. A l'échelle annuelle, des similitudes d'évolution des classes entre les trois stations est constatée.

V.2. Pourcentage des types de journées

L'étude de la séquence des classes de journées sur trois années, nous a permis d'obtenir les pourcentages de journées type . La Figure 3-14 représente le nombre de journées d'apparition en fonction d'une classe donnée pour chaque station de mesures.

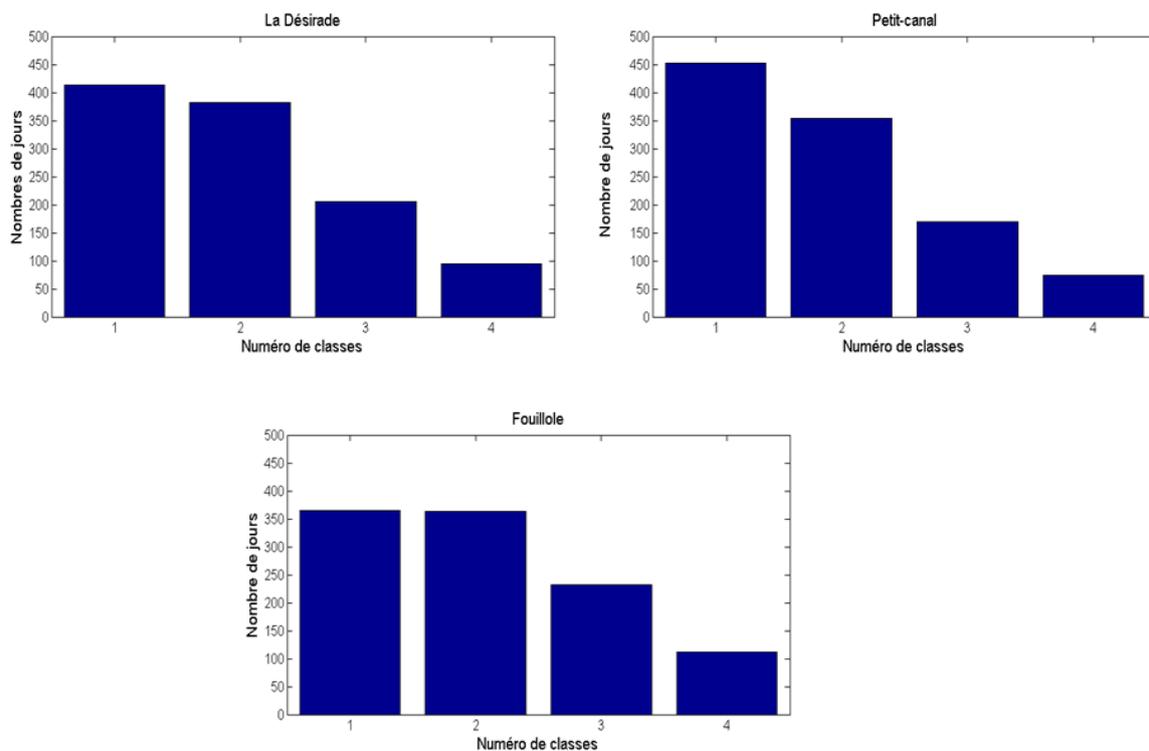


Figure 3-14. Nombre de jours d'apparition pour les classes {1, 2, 3, 4} au cours des trois années de mesures sur le site de La Désirade, Petit-canal et Fouillole.

Sur le site de La Désirade, on trouve 413 jours pour la classe 1 sur l'ensemble des trois années 2010, 2011 et 2012 ce qui correspond à un pourcentage de 38% de journées appartenant à cette classe. On obtient 382 jours soit 35% pour la classe 2, 206 jours soit 19% pour la classe 3 et 95 jours soit 9% pour la classe 4. On peut donc conclure que la classe 1 et 2 sont les classes prédominantes pour les données de mesure de flux solaire sur La Désirade. Les journées types les plus fréquentes pour cette localisation sont donc des journées à ciel clair et des journées intermédiaires à ciel clair représentatives de passages nuageux moyennement fréquents.

Sur le site de Petit-canal, on trouve 452 jours pour la classe 1 sur l'ensemble des trois années de mesures, ce qui correspond à un pourcentage de 43% de journées correspondant à ce type de classe. On obtient 354 jours soit 34% pour la classe 2, 169 jours soit 16% pour la classe 3 et 74 jours soit 7% pour la classe 4. On en déduit que les classes 1 et 2 sont les classes prédominantes pour les données de mesure de flux solaire sur Petit-canal. Les journées types à ciel clair et intermédiaires ciel clair sont également les plus fréquentes pour cette station.

Sur le site de Fouillole, on trouve 365 jours pour la classe 1 ce qui correspond à un pourcentage de 34% de journées correspondant à ce type de classe. On obtient 363 jours soit 34% pour la classe 2, 232 jours soit 22% pour la classe 3 et 112 jours soit 10,4% pour la classe 4. Les classe 1 et 2 sont également les classes prédominantes pour les données de mesure de flux solaire sur le site de Fouillole.

Bien que les classes 3 et 4 sont plus fréquentes sur le site de Fouillole comparé aux autres stations, on retrouve environ les mêmes proportions d'apparition pour une classe donnée, d'une station à l'autre.

V.3. Persistance des classes

Une autre caractéristique pertinente de la séquence de classe est la persistance. Une étude de la durée, en terme de jours, pour laquelle une classe persiste est présentée dans cette section. La Figure 3-15 représente le nombre de jours consécutifs pour laquelle une classe donnée persiste. La fréquence (nombre de fois au cours des trois années de mesures) correspondant à nombre de jours successifs identifié est également décrite en ordonnée. On constate qu'il s'agit de la classe 1 qui dure plus longtemps en termes de jours soit de 13 à 24 jours consécutifs avec une persistance maximale sur le site de Petit-canal. La classe 2 peut durer de 8 à 18 journées successives avec une durée maximale également sur le site de Petit-canal. La classe 3 peut durer au plus, 4 à 10 journées successives avec une durée maximale sur le site de La Désirade. Concernant la classe 4, le nombre de jours successifs est maximal pour le site de Fouillole, soit 5 journée successives. Ainsi, la classe 1 peut durer jusqu'à 24 jours consécutifs tandis que la classe 4 peut durer jusqu'à 5 jours consécutifs. C'est sur le site de Fouillole que les journées nuageuses persistent plus longtemps comparé aux autres stations. Néanmoins, les journées ensoleillées persistent plus longtemps sur Petit-canal. La localisation

géographique de la station de Petit-canal étant sur la côte Ouest de la Guadeloupe est caractéristique d'un régime des côtes au vent, le vent y souffle en permanence. Dans cette zone, la formation nuageuse est majoritairement due à l'advection des masses d'air marin. Cette configuration impose des paramètres climatiques avec des formations nuageuses moins fréquentes que le régime continental insulaire, où se situe la station de Fouillole pour laquelle la formation nuageuse est principalement due à la convection des masses d'air.

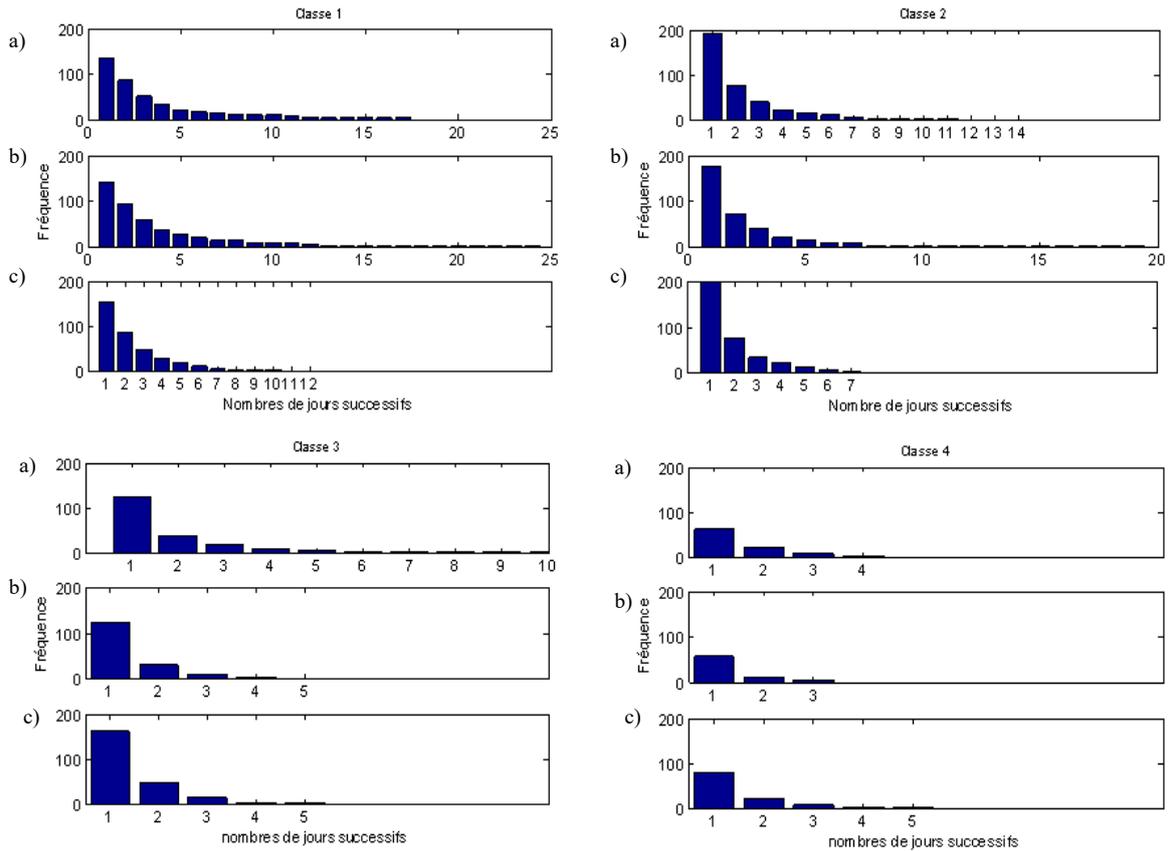


Figure 3-15. Fréquence du nombre de jours d'apparition successifs pour les classes {1, 2, 3, 4} au cours des trois années de mesures sur les stations : La Désirade a), Petit-canal b), Fouillole c)

Hormis la persistance de chaque classe, la séquence des classes présente des propriétés intéressantes et permet de rendre compte d'une transition d'une classe à une autre. Nous nous sommes donc intéressés à calculer la probabilité de transition d'une classe i vers une autre classe j . Nous avons ainsi assimilé nos séquences de classes par une suite de variables aléatoires $\{X_n\}$ $n \geq 0$ à valeurs dans l'espace dénombrable E , appelé processus stochastique à temps discret à valeurs dans E . L'ensemble E est l'espace d'état, dont les éléments seront notés 1, 2, 3, 4. Lorsque $X_n = 1$, le processus est dit être dans, ou visiter, l'état 1 au temps n . L'évolution peut se modéliser à l'aide d'une matrice de transition.

	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	62,3	25,1	5,6	7,0
Classe 2	15,2	65,1	16,3	3,3
Classe 3	3,0	12,3	81,2	3,5
Classe 4	1,7	3,5	3,5	91,3

Tableau 3-6 Probabilités de transition d'une classe i vers une autre classe j avec $i=\{1, 2, 3, 4\}$ et $j=\{1, 2, 3, 4\}$ calculées en en pourcentage (%) pour le site de La Désirade sur trois années de mesures.

	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	56,9	27,6	10,9	4,6
Classe 2	15,1	66,2	15,3	3,3
Classe 3	5,1	8,6	83,4	2,4
Classe 4	1,4	2,3	3,3	92,9

Tableau 3-7 Probabilités de transition d'une classe i vers une autre classe j avec $i=\{1, 2, 3, 4\}$ et $j=\{1, 2, 3, 4\}$ calculées en en pourcentage (%) pour le site de Petit-canal sur trois années de mesures.

	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	65,9	23,2	7,8	3,0
Classe 2	12,7	66,1	16,2	4,9
Classe 3	5,9	11,1	78,4	4,7
Classe 4	1,9	4,1	4,4	89,5

Tableau 3-8 Probabilités de transition d'une classe i vers une autre classe j avec $i=\{1, 2, 3, 4\}$ et $j=\{1, 2, 3, 4\}$ calculées en en pourcentage (%) pour le site de Fouillole sur trois années de mesures.

Les tableaux (Tab3.6, 3-7 et 3-8) représentent les probabilités calculées de transition de i vers j , qu'on peut assimiler à une matrice de transition de la chaîne. Ces probabilités sont calculées pour chacun des sites de mesures. La probabilité que la journée suivante puisse appartenir à la même classe que la journée précédente présente une forte probabilité nettement supérieure à celle obtenue quand elle transite à un autre type de journée. Les dynamiques de transitions d'une classe i à une classe j présentent des probabilités ayant des valeurs proches par comparaison intersites. Bien que les transitions d'une classe 1 à une classe 2, d'une classe 2 à une classe 3, ainsi que le phénomène inverse, présentent des probabilités moyennes, on observe une très faible probabilité de transition de la classe 3 à classe 4 (inférieures à 5%). Un commentaire principal peut être donné sur l'ensemble des trois

tableaux : sur les trois stations, la dynamique de transition des classes journalières d'ensoleillement est constatée similaire.

L'analyse des séquences de classes a permis de mettre en exergue le pourcentage des types de journées, de caractériser une dynamique de variabilité journalière pouvant persister sur une période maximale calculée. Cette analyse a également permis d'identifier un ordre des états de variabilité de flux solaire et une dynamique de transition. L'identification d'un tel ordre peut être assimilée à une chaîne de Markov discrète ayant quatre états {1, 2, 3, 4}. Ceci peut être intéressant pour des recherches plus approfondies sur la prévision du rayonnement solaire à l'échelle journalière.

V.4. Discussion et comparaison inter-site des classes

En comparant les séquences de classes sur nos trois sites, on peut constater des corrélations inter sites de la variabilité de l'ensoleillement à l'échelle journalière. En effet, les classes se produisent dans les mêmes périodes de l'année. De plus, les classes prédominantes obtenues sont identiques sur nos trois sites à savoir la classe 1 et 2. Seuls les pourcentages diffèrent quelque peu, avec pour le site de Petit-canal un pourcentage supérieur pour la classe 1 comparé aux autres sites. Pour la classe 2, c'est le site de La Désirade qui obtient un plus fort pourcentage, la classe 3 est plus fréquente sur le site de Fouillole et pour la classe 4, il s'agit également du site de Fouillole. On peut alors en déduire que le site de Petit-canal et de La Désirade ont une énergie solaire reçue supérieure à celle de Fouillole. A l'échelle annuelle, les phénomènes périodiques d'apparition des classes sont similaires d'une station à l'autre.

Afin de repérer les similitudes inter-sites entre les journées, nous effectuons une comparaison des classes inter journalières entre les stations. Nous étudions alors, la fréquence des journées qui ont la même classe simultanément sur deux stations (entre La Désirade et Petit-canal, entre La Désirade et Fouillole, et entre Petit-canal et Fouillole) mais aussi simultanément sur les trois sites. Nos résultats sont montrés dans le tableau Tableau 3-9 ci-dessous.

	Désirade / Petit-canal	Désirade/ Fouillole	Petit-canal /Fouillole	Désirade/ Petit-canal /Fouillole	Désirade	Petit-canal	Fouillole
Classe 1	348 jours soit 31,7%	287 jours soit 26,2%	297 jours soit 27,1%	249 jours soit 27,7%	413 jours	452 jours	365 jours
Classe 2	225 jours soit 20,5%	190 jours soit 17,3%	194 jours soit 17,7%	132 jours soit 12,04%	382 jours	354 jours	363 jours
Classe 3	96 jours soit 8,76%	89 jours soit 8,12%	93 jours soit 8,49%	55 jours soit 5,02%	206 jours	169 jours	232 jours
Classe 4	58 jours soit 5,25%	62 jours soit 5,66%	61 jours soit 5,57%	50 jours soit 4,56%	95 jours	74 jours	112 jours

Tableau 3-9 Tableau des comparaisons journalières des classes {1, 2, 3, 4} présentant les journées ayant la même classe entre les sites simultanément avec la fréquence des classes en nombre de jours sur chaque site individuel.

Le pourcentage des corrélations est calculé par le rapport entre le nombre de journées appartenant à la même classe simultanément sur deux stations (ou trois stations) et le nombre total de journées sur l'ensemble des années de mesures (1096 journées). Les trois dernières colonnes représentent le nombre de journées appartenant à une classe donnée sur un site. Ces résultats sont tirés de la Figure 3-14. Ces résultats sont rappelés dans ce tableau, pour avoir une idée de la proportion des journées simultanées appartenant à la même classe entre une paire de stations. On peut constater, à partir de ce tableau, que les corrélations sont moins importantes entre Fouillole et La Désirade. Ce qui peut s'expliquer par le passage des masses nuageuses se déplaçant selon la direction des Alizés prédominante des Alizés Nord-Est et Sud-Est. Cette direction prédominante entraîne donc une probabilité faible pour que les masses nuageuses se déplacent dans la direction partant de La Désirade vers Fouillole. La distance étant plus grande entre Fouillole et La Désirade peut aussi en être une cause. A contrario, les corrélations sont plus importantes entre La Désirade et Petit-canal par rapport aux autres comparaisons inter-sites. Pourtant, la distance entre Petit-canal et Fouillole est plus petite qu'entre La Désirade et Petit-canal. Ceci peut s'expliquer par le fait que le passage des masses nuageuses se déplaçant selon la direction des vents Nord-Est, Sud-Est, vont de La Désirade vers Petit-canal. On peut en déduire que les corrélations du flux solaire entre les sites sont plus influencées par la direction du vent que par les distances entre les sites. Ce qui est une information intéressante pour la prédiction spatio-temporelle du flux solaire.

Toutefois, les corrélations entre les classes signifient seulement que le même type de dynamique à l'échelle journalière se produit entre deux stations mais pas forcément les mêmes variations flux solaire à court terme (échelle intra journalière). On peut observer un exemple dans la Figure 3-16 où les fluctuations ne se produisent pas toujours exactement à la même période de la journée. Une étude alors plus fine sur les fluctuations du flux solaire à court terme sera effectuée à partir d'outils statistiques dans le chapitre suivant.

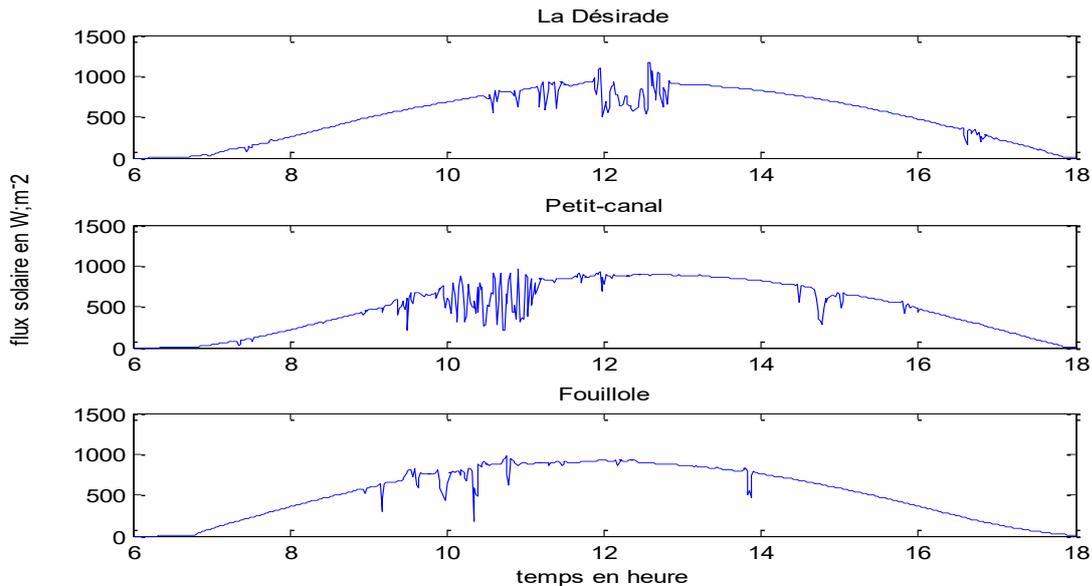


Figure 3-16. Exemple de journée (31 janvier 2010) ayant la même classe 1 simultanément sur les trois sites de mesure : La Désirade, Petit-canal et Fouilllole.

VI. Conclusion

Dans ce chapitre, la variabilité de l'ensoleillement a été caractérisée et quantifiée à l'échelle journalière par une analyse des classifications journalières de l'ensoleillement.

Nous avons utilisé une méthode de classification basée sur l'algorithme k-means. Cette classification étant non-supervisée, la question du nombre de classes à déterminer était donc centrale. L'application de l'ACP fonctionnelle à nos données de flux solaire mesurées sur trois sites (La Désirade, Petit-canal et Fouilllole) a fait ressortir quatre harmoniques présentant une totalité jusqu'à 73% de la variance du flux solaire. Ces résultats ont validé le nombre de classe déterminé préalablement par une hypothèse fondée sur les résultats de classification de l'ensoleillement dans la littérature.

Les quatre classes journalières identifiées, sont caractérisées par quatre types de dynamiques différentes : dynamique lente avec peu de fluctuations (jour à ciel clair), dynamique moyenne (jour ciel clair avec passages nuageux), dynamique rapide (jour à ciel nuageux avec une petite période à ciel clair), dynamique lente mais variabilités fréquentes avec un ensoleillement faible (jour à ciel nuageux).

Cette classification a permis de montrer une forte corrélation des classes de journées entre les trois sites. En effet, l'étude de l'évolution des séquences de classes $\{1, 2, 3, 4\}$ a montré des classes de journées apparaissant à des périodes identiques dans l'année sur les trois sites avec des corrélations très significatives de classes apparaissant simultanément sur les stations deux à deux.

Des régimes de classes constatées identiques entre les sites et la probabilité de transition d'une classe à une autre, apportent des informations intéressantes pour une modélisation. Dans les articles Muselli (2000) et Soubdhan et al (2009), une étude plus approfondie est suggérée en utilisant des processus de Markov pour synthétiser des séquences de jours typiques. La caractérisation de la variabilité du flux solaire par identification des classes de journées permettra d'évaluer la performance du modèle de prédiction spatio-temporel en fonction de la variabilité du flux solaire dans le chapitre 5. De plus, l'analyse des corrélations entre les sites à partir des séquences de classes a permis de constater que les corrélations étaient plus fortes entre les stations de Désirade et Petit-canal comparé aux autres paires de sites. Nous pouvons alors en déduire que les corrélations du flux solaire entre les sites sont plus influencées par un phénomène physique que par les distances entre les sites. C'est une information intéressante qui nous mène à faire dans le chapitre suivant, une étude statistique et dynamique plus fine des corrélations de la variabilité de flux solaire entre les stations, c'est-à-dire pour des fluctuations de l'ensoleillement à très court terme (quelques minutes à quelques heures).

Références

- Aranovitch E., Gandino C, Gillaert D, 1984, Colloque Météorologie et Energies Renouvelables. Valbonne, p.315±338.
- Badosa J., Haeffelin M., Chepfer H., Scales of spatial and temporal variation of solar irradiance on Reunion tropical island, *Solar Energy*, Volume 88, February 2013, Pages 42-56
- Bardos M., *Analyse Discriminante - Application au risque et scoring financier*, Dunod, 2001.
- Benhenni K., Hedli-Griche S., Rachdi M., Vieu P., 2008, Consistency of the regression estimator with functional data under long memory conditions , *Statistics and Probability Letters*, Volume 78, Issue 8, 1043-1049
- Berkes I., Gabrys, R., Horvath, L. and Kokoszka, P. (2009). Detecting changes in the mean of functional observations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 71 927–946.
- Biernacki C., Celeux G., Govaert G., Langrognet F., (2006) Model-based cluster and discriminant analysis with the MIXMOD software, *Computational Statistics & Data Analysis* 51 (2), 587-600
- Brévignon C., 2005. L'environnement atmosphère de la Guadeloupe, de St Barthélémy et de St Martin, La direction Inter Régionale Antilles Guyane de Météo France.
- Boente, G. and Fraiman, R. (2000) Kernel-based functional principal components. *Statist. Probab. Lett.*, 48, 335-345.
- Boullier, P., Le Chapellier M., 1984. *Meteorology and Renewable Energy symposium*. Valbonne, 632-653.
- Brumback, B. A. and Rice, J. A. (1998) Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Am. Statist. Ass.*, 93, 961-976.
- Cardot, H. (2000) Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *J Nonparam. Statist.*, 12, 503-538
- Chebana, F., S. Dabo-Niang, and T. Ouarda, 2012, Exploratory functional flood frequency analysis and outlier detection, *Water Resources Research*, 48 (4), W04, 514.
- Chen Mu-Yen, 2011, Predicting corporate financial distress based on integration of decision tree classification and logistic regression, *Expert Systems with Applications*, Volume 38, Issue 9, September 2011, Pages 11261-11272
- Dabo-Niang S., Ferraty F. and Vieu P., 2007, On the using of modal curves for radar wave curves classification. *Computational Statistics and Data Analysis*, Volume 51, 4878-4891
- Dauxois, J., Pousse, A. & Romain, Y. (1982), 'Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference', *Journal of Multivariate Analysis* 12(1), 136–154.
- Delaigle, A., Hall, P., and Bathia, N. (2012). Componentwise classification and clustering of functional data. *Biometrika*, page ass003.

Fabero F., Alonso-Abella M., Chenlo F., 1997. Proc. of the 14th European PV Solar Energy Conference, Barcelone, 2299-2302.

Febrero, M., Galeano, P., and González-Manteiga, W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal NO_x levels. *Environmetrics*, 19(4) :331–345.

Ferraty F., Laksaci A., Tadj A., Vieu P., 2010, Rate of uniform consistency for nonparametric estimates with functional variables, *Journal of Statistical Planning and Inference*, Volume 140, Issue 2, 335-352

Hadd M., mémoire juin 1999, Classification de la population en catégories socio-économiques: méthodologie et application pratique, Institut national statistiques et d'économie appliquée

Harrouni, S, Guessoum, A., Maafi, A., 2005. Classification of daily solar irradiation by fractional analysis of 10-min-means of solar irradiance, *Theoretical and Applied Climatology* 80, 27–36.

Hartigan J.A, Wong M.A,(1979), *Applied Statistics*, *Journal of the Royal Statistical Society*, Vol.28, No1, pp.100-108

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, 38(3) :1845–1884.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 68(1) :109–126.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psy.* 417-441.

Hyndman, R. J. and Shang, H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1).

James, G. M., Hastie, T. J. and Sugar, C. A. (2000) Principal component models for sparse functional data. *Biometrika*, 87, 587-602.

Jones, M. C. & Rice, J. A. (1992), Displaying the important features of large collections of similar curves, *The American Statistician* 46(2), 140–145.

Le Guen.M ,La boîte à moustaches pour sensibiliser à la statistique, CNRS-Matisse, publié dans "Bulletin de méthodologie sociologique, 73 (2002) 43-64

Loh WY, 2002, Regression trees with unbiased variable selection and interaction detection, *Statistica Sinica* 12 (2), 361-386

Louche, A., Notton, G., Poggi, P., Simonot, G., 1991, Classification of direct irradiation days in view of energetic applications, *Solar Energy* 46 (4), 255. posed *Solar Energy* 46 (4), 255.

Maafi, S., Harrouni, 2003, Preliminary results of the fractal classification of daily solar irradiances. *Solar Energy* 75, 53–61.

Muselli, M., Poggi, P., Notton, G., Louche, A., 2000, Classification of typical meteorological days from global irradiation records and comparison between two Mediterranean coastal sites in Corsica Island. *Energy Conversion and Management* 41, 1043–1063.

Ramsay J., and B. Silverman (2005), *Functional data analysis*, Springer, New York.

- Saporta G., Probabilités, Analyse des données et Statistiques, Paris, Éditions Technip, 2006, 622p
- Shang, H. L. (2014). A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98(2) :121–142.
- Shao Y-H, Deng N-Y, Yang Z-M, 2012, Least squares recursive projection twin support vector machine for classification, *Pattern Recognition*, Volume 45, Issue 6, June 2012, Pages 2299-2307
- Soubdhan T, Richard Emilion, Calif.R , février 2009, Classification of daily solar radiation distributions using a mixture of Dirichlet distributions, *Solar Energy* 83 (2009) 1056–1063
- Soubdhan T, Abadi M., Richard Emilion, Time Dependent Classification of Solar Radiation Sequences Using Best Information Criterion, *Energy Procedia*, Volume 57, 2014, Pages 1309-1316
- Tomson T., Tamm G., Short-term variability of solar radiation, 2005, *Solar Energy* 80 p600–606
- Voyant C., November 2011, Time series prediction of global solar radiation and photovoltaic energy production using artificial neural networks, thèse, University of Corse-Pascal Paoli, UMR CNRS 6134(SPE)
- Yang X, Tan L., He L., A robust least squares support vector machine for regression and classification with noise, *Neurocomputing*, Volume 140, 22 September 2014, Pages 41-52
- Yao, F., Müller, H.-G. and Wang., J.-L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* 33 2873–2903. MR2253106
- Wu D., Liang L., Yang Z., 2008, Analyzing the financial distress of Chinese public companies using probabilistic neural networks and multivariate discriminate analysis, *Socio-Economic Planning Sciences*, Volume 42, Issue 3, Pages 206-220
- I. Jolliffe, 2002, *Principal Component Analysis*, 2nd edition éd., Springer-Verlag
- Barra V, 2004 Analysis of gene expression data using functional principal components., *Comput Methods Programs Biomed.* 2004 Jul;75(1):1-9
- Kaiser, H. F., 1960,. The application of electronic computers to factor analysis, *Educational and Psychological Measurement*, 20, 141-151.
- Cattell, R. B., 1966, The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-27

Chapitre 4. Corrélation spatio-temporelle de la variabilité de l'ensoleillement global à court-terme

Table des matières

I. Analyse sur trois stations de mesures.....	Erreur ! Signet non défini.
I.1. Localisation géographique	Erreur ! Signet non défini.
I.2. Contexte climatique des stations.....	Erreur ! Signet non défini.
II. Analyse statistique descriptive de la variabilité de flux solaire à court terme	Erreur ! Signet non défini.
II.1. Caractérisation de l'énergie reçue.....	Erreur ! Signet non défini.
II.2. Caractéristiques de la variabilité de l'ensoleillement global à court terme	Erreur ! Signet non défini.
III. Variabilité spatio-temporelle du flux solaire.....	Erreur ! Signet non défini.
III.1. Aperçu des travaux dans la littérature	Erreur ! Signet non défini.
III.2. Evaluation des corrélations : Outils statistiques utilisés.....	Erreur ! Signet non défini.
III.3. Analyse des corrélations inter sites spatiales.....	Erreur ! Signet non défini.
III.4. Analyse des corrélations inter sites temporelles.....	Erreur ! Signet non défini.
IV. Analyse dynamique des corrélations	Erreur ! Signet non défini.
IV.1. Décalage temporel des corrélations.....	Erreur ! Signet non défini.
IV.2. Corrélations intersites et paramètres physiques	Erreur ! Signet non défini.
IV.3. Description dynamique des corrélations	Erreur ! Signet non défini.
V. Conclusion	Erreur ! Signet non défini.
Références	Erreur ! Signet non défini.

L'énergie solaire est une ressource qui varie à la fois dans l'espace et dans le temps et fait l'objet d'initiatives de recherches majeures comme aux Etats-Unis cités dans Pérez et al (2012). Les îles tropicales, telles que La Guadeloupe possèdent une ressource solaire très importante cependant fortement variable à la fois dans l'espace et dans le temps à cause du climat chaud et humide qui favorise la formation de nombreuses masses nuageuses hétérogènes, sources principales des variations rapides de l'ensoleillement global. Dans ce chapitre, nous avons mené une analyse statistique et dynamique spatio-temporelle de la variabilité de l'ensoleillement global à court terme. Nous nous intéresserons aux corrélations spatiales et temporelles de la variabilité du flux solaire existantes entre nos trois sites. La connaissance fine du comportement à court-terme du flux solaire et des corrélations inter-sites est une étape indicative et informative pour la réalisation de notre modèle de prévision spatio-temporelle. Premièrement, nous détaillerons le contexte météorologique de la localisation de chaque station mettant en jeu des paramètres physiques liés aux corrélations spatio-temporelles. Cette description sera suivie d'une caractérisation de la dynamique des variations de l'ensoleillement global à court terme pour chacune de nos stations. Pour notre étude spatio-temporelle, nous effectuerons une analyse statistique des corrélations intersites spatiales et temporelles (fréquence, probabilité). Nous terminerons par une étude des différents paramètres liée aux résultats des corrélations intersites spatio-temporelles (décalage temporel, interaction de la variabilité intersite).

I. Analyse sur trois stations de mesures

I.1. Localisation géographique

Bien que les données de mesures ont déjà été décrites dans le chapitre précédent, il est nécessaire de rappeler la localisation géographique des stations afin de mieux comprendre la physique des fluctuations de l'ensoleillement.

Les données sont enregistrées sur trois stations de mesures en Guadeloupe : Petit-canal (16°38N, 61°49W) sur la côte Ouest de la Guadeloupe dans le lieu dit Gros Cap, La Fouillole (16°26N, 61°24W) sur le Campus de l'Université des Antilles au Sud de la Grande-Terre, et le site le plus à l'Est est celui de La Désirade (16°31N, 61°55W), situé à l'extrémité Nord Est de La Désirade. La Localisation géographique de chaque site est représentée sur la carte de la Figure 4.1.

Nous disposons de trois sites de mesures dont les distances entre les sites sont comprises environ entre 26km et 56km. Celles-ci sont résumées dans le Tableau 4.1. L'approche classique des travaux de recherches sur les statistiques spatio-temporelles trouvées dans la littérature est d'utiliser de nombreux sites proches de la dizaine voir de la vingtaine pour des espaces avec des limites bien définies et de

quantifier les interactions intersites (Glasbey et al (2008) grille carrée 10km*10 km pour 10 station, Inoue et al (2012) grille 40km*40km de 25 stations, Yang et al (2014) grille 22km*41km de 13 sites). Nous ne disposons que de trois stations de mesure ce qui rend impossible la réalisation de variogramme expérimental effectué entre autres dans Hoff et Perez (2011), Pérez et al (2011), Pérez et al (2012), Yang et al (2014), Inoue et al (2012). De plus, la configuration géographique des stations de mesures séparées par un bras de mer n'a pas encore été étudiée dans la littérature. Cette configuration impose des paramètres climatiques (formation de masses nuageuses, direction du vent...) influençant les corrélations spatio-temporelles entre nos stations. Notamment, les Alizés ayant une forte composante Est, le site de La Désirade est d'un intérêt particulier puisqu'il est situé sur le premier relief îlien que les masses d'air océaniques vont rencontrer. Il y a donc une forte probabilité que les nuages sur la Désirade atteignent La Guadeloupe continentale avec un décalage temporel. La Figure 4-3 représente les localisations géographiques des trois stations en Guadeloupe. Nous identifions les différents régimes climatiques de chacune de nos stations imposés par leur situation géographique au sein de l'Archipel Guadeloupe dans la section suivante.

De	à	Distance	Distance	Distance
			Nord-Sud	Est-Ouest
Petit-canal	Fouillole	26272 m	22264	46383
Petit-canal	La Désirade	41818 m	12987	11132
Fouillole	La Désirade	55819 m	9276	57515

Tableau 4-1 Distance inter-site

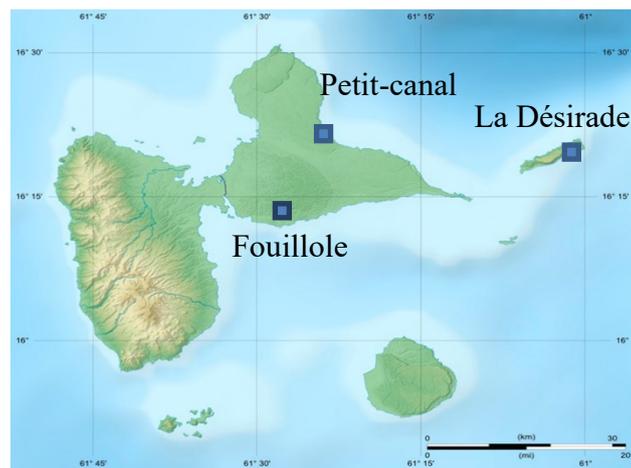


Figure 4-1 Carte de l'Archipel de Guadeloupe et localisation géographique de nos trois stations de mesures: Petit-canal, Fouillole et La Désirade.

I.2. Contexte climatique des stations

Cette section permet d'avoir une connaissance générale des phénomènes climatiques spécifiques à chaque site. En effet, elle apporte les informations physiques nécessaires à l'étude spatio-temporelle de l'ensoleillement global.

Selon la classification proposée par Brévignon C. (2005) à l'aide d'observation de terrain et d'analyse de données, le site de Petit-Canal se trouve dans le régime des côtes au vent, le vent y souffle en permanence. Dans cette zone la formation nuageuse est majoritairement due à l'advection des masses d'air marin. La station de Fouillole est soumise à un régime continental insulaire, dans cette zone le vent souffle moins fort que sur le régime des côtes au vent. La formation nuageuse est principalement due à la convection des masses d'air. La station de La Désirade se situe dans le même régime climatique que celle située à Petit-canal. Ces trois stations se situent donc dans des régimes climatiques distincts : les côtes au vent (La Désirade et Petit-canal) et une localisation située à l'intérieur des terres parallèle à la direction générale du vent des Alizés (Fouillole).

Si on considère les stations deux à deux, chaque paire de site montre un système physique avec une diversité de conditions météorologiques possibles. Néanmoins, comme mentionné préalablement, il y a une forte probabilité que les masses nuageuses se déplacent de l'île de La Désirade à la Guadeloupe continentale selon les données de Météo France (direction du vent, images satellites et radar). Par hypothèse, il sera donc possible d'effectuer une prédiction de la ressource solaire à court terme sur les autres sites. Cette hypothèse peut notamment être justifiée par notre analyse statistique et dynamique des corrélations spatio-temporelles de la variabilité du flux solaire à court terme.

II. Analyse statistique descriptive de la variabilité de flux solaire à court terme

Le chapitre 3 montrant une analyse de la variabilité de l'ensoleillement à l'échelle journalière, dans cette section, nous avons mené une analyse statistique et dynamique spatio-temporelle de la variabilité de l'ensoleillement global à court terme. Cette analyse statistique a pour but, d'une part d'effectuer une caractérisation du point de vue de l'énergie reçue et d'autre part d'effectuer une caractérisation de la variabilité du flux solaire à court-terme (de quelques minutes à quelques heures).

II.1. Caractérisation de l'énergie reçue

II.1.1 Définition de la fraction d'énergie

Des modèles déterministes permettent de calculer une courbe dite « idéale de l'ensoleillement » qui correspond à une journée d'ensoleillement sans passage nuageux. Ces modèles tiennent compte des effets de diffusion et d'absorption que subit le rayonnement solaire lors de sa traversée de l'atmosphère. Ils sont basés sur la détermination des coefficients de transmission des différents constituants atmosphériques. Ces coefficients nécessitent la disponibilité des paramètres météorologiques courants (l'humidité relative, la température ambiante, la pression atmosphérique, ...) et des paramètres géographiques du site (latitude, longitude et altitude). De nombreux auteurs se sont intéressés à l'étude théorique et expérimentale du rayonnement solaire. Sans être exhaustif, nous pouvons citer : Zidani C. et al. (2003) qui ont présenté une simulation de la position apparente du soleil à chaque instant de la journée et de l'année. Ce travail est enrichi par un calcul d'estimation des énergies instantanée, journalière et annuelle sur un capteur d'orientation quelconque placé sur le site Tlemcen. A. Baig et al. (1991) ont proposé un modèle basé sur la version modifiée de la fonction de distribution gaussienne pour estimer l'éclairement global journalier. Un modèle physique basé sur l'équation générale de transfert radiatif du rayonnement solaire dans le système sol atmosphère a été utilisé par Mechaqrane A. et al. (1993) pour l'estimation des irradiances globales horaires et journalières d'une surface horizontale au sol à partir des mesures dans la bande 0,4-1,1 μm du satellite Météosat-2. Plusieurs modèles théoriques sont définis tels que le modèle de Lacis & Hansen, le modèle de Davies & Hay, le modèle de Bird & Hulstrom et celui de Atwater & Ball, dont des travaux sont présentés dans la littérature (Koussa.M 2006).

Nous utilisons le modèle d'estimation du rayonnement théorique défini par Kasten. Ce modèle est basé sur la formule pyréliométrique développée dans Kasten (1980). Comme le modèle d'Ineichen, il exige seulement la turbidité de Linke. Une description du modèle peut être trouvée dans l'article de Kasten (1984). Ce modèle a été déjà utilisé dans la littérature par plusieurs scientifiques tels que Soubdhan et al (2007), Boubekri (2009), Soubdhan et al (2009).

La fraction d'énergie journalière du flux solaire φ^* , est le ratio entre l'énergie disponible et l'énergie attendue en l'absence de nuages calculée par le modèle de Kasten, selon l'expression suivante :

$$\varphi^* = \frac{\bar{\varphi}_m}{\bar{\varphi}_c} \quad (4.1)$$

$\bar{\varphi}_m$ est la moyenne du flux solaire mesuré entre le lever et le coucher du soleil et $\bar{\varphi}_c$ est la valeur moyenne du flux solaire théorique calculé selon le modèle de Kasten, en fonction de la position géographique, du jour Julien, de l'heure et du facteur de trouble de l'atmosphère. Une valeur de φ^* proche de 1 correspond à une journée de ciel clair tandis qu'une faible valeur de ce paramètre correspond à une journée ayant une forte couverture nuageuse. L'étude de la fraction d'énergie pour chaque site permettra donc d'analyser l'énergie solaire reçue propre à chaque site. Une analyse sur deux de nos sites a déjà été étudiée dans la littérature par Soubdhan et al (2007). Nous effectuons une étude analogue mais sur trois sites. Les résultats de densité de probabilité de ce paramètre au cours des trois années de mesure pour chaque site sont présentés.

II.1.2 Analyse intersite de la fraction d'énergie

Nous avons défini dans le chapitre 2, quatre classes de journées distinctes par méthode kmeans: journées à ciel clair, journées à ciel intermédiaire, journée à ciel nuageux intermédiaire et journées à ciel nuageux. Parallèlement aux valeurs de fraction d'énergie, nous trouvons $0.7 \leq \varphi^* \leq 1$ correspondant aux journées à ciel clair, $0.5 \leq \varphi^* \leq 0.7$ correspondant aux journées à ciel clair intermédiaire, $0.2 \leq \varphi^* \leq 0.5$ correspondant aux journées à ciel nuageux intermédiaire, et $0 \leq \varphi^* \leq 0.2$ correspondant aux journées à ciel nuageux.

La densité de probabilité de la fraction d'énergie est montrée pour chaque site de mesure dans la Figure 4-32. Le site de La Désirade, de Petit-canal et de Fouillole montrent des occurrences plus importantes pour des valeurs de fraction d'énergie φ^* entre 0.7 et 1, la densité de probabilité de cette gamme est entre 0.05 et 0.23. Néanmoins, le site de Fouillole se détache nettement des autres stations par une plus faible densité de probabilité dans cet intervalle de fraction d'énergie. La Désirade et Petit-canal ont une densité de probabilité totale similaire pour $0.7 \leq \varphi^* \leq 1$, égale à environ 0.75. Cependant, on peut remarquer que pour une fraction d'énergie φ^* comprise entre 0.95 et 1, Petit-canal montre une nette différence par une densité plus importante que celle de La Désirade. Cette gamme de fraction d'énergie, $0.7 < \varphi^* < 1$, est caractéristique d'une énergie solaire reçue et rend compte que les journées à ciel clair sont plus probables sur les trois stations de mesure.

Néanmoins, le site de Fouillole présente une plus forte probabilité de $\varphi^* < 0.7$ comparé aux autres sites. En effet, on peut observer que la fraction d'énergie est distribuée plus vers les grandes valeurs pour Petit-canal et La Désirade alors que pour Fouillole, la densité est un peu plus étalée avec une densité plus importante par rapport aux deux autres stations vers les valeurs de fraction d'énergie

inférieures à 0.7. Ces valeurs sont caractéristiques d'une plus faible énergie reçue pour le site de Fouillole qui peut se traduire par un plus grand nombre de passages nuageux.

Ces premiers résultats montrent une différence entre les 3 stations d'un point de vue énergétique.

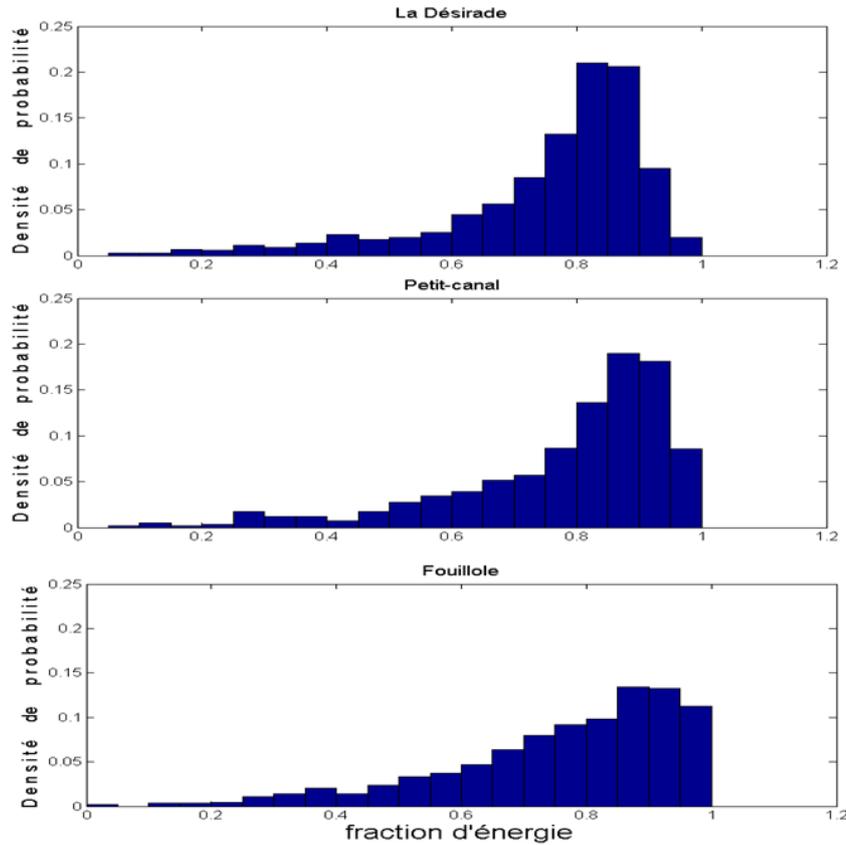


Figure 4-2. Fraction d'énergie de l'ensoleillement sur l'ensemble de nos données de mesures (2010 à 2012) pour le site de La Désirade

II.2. Caractéristiques de la variabilité de l'ensoleillement global à court terme

Nous mettons en évidence dans cette section, essentiellement les variations de l'ensoleillement à l'échelle intra-journalière (quelques heures) et à très court-terme (quelques secondes à quelques minutes), qui sont les échelles de temps les plus pertinentes du point de vue de la variabilité de l'ensoleillement.

II.2.1 Variabilité à l'échelle intra-journalière

La variabilité de l'ensoleillement à l'échelle intra-journalière est particulièrement notable. Nous présentons les propriétés de centralité et de dispersion des données. Pour cela, nous avons effectué la moyenne, le mode et la médiane au même instant t entre les journées de mesures. Nous obtenons alors des courbes représentant une journée moyennée sur trois années (ainsi que le mode et la médiane) [Figure 4-3 ; Figure 4-34 ; Figure 4-35]. Nous avons choisi de montrer les résultats pour les données de flux solaire par pas de 10 minutes. Ce rééchantillonnage met plus nettement en exergue les trois régimes de fluctuations dans la journée puisque les signaux par pas de 1 secondes ou 1 minute sont très bruités. Les conclusions et observations sont les mêmes quel que soit la période de rééchantillonnage utilisée. Entre autre, les signaux par pas de 10 minutes étant moyennés sur un plus grand nombre de points par rapport à 1 minute, ceux-ci ne gardent alors que les fluctuations les plus importantes et donnent une perception plus globale des résultats concernant les courbes de moyennes, modes et médianes.

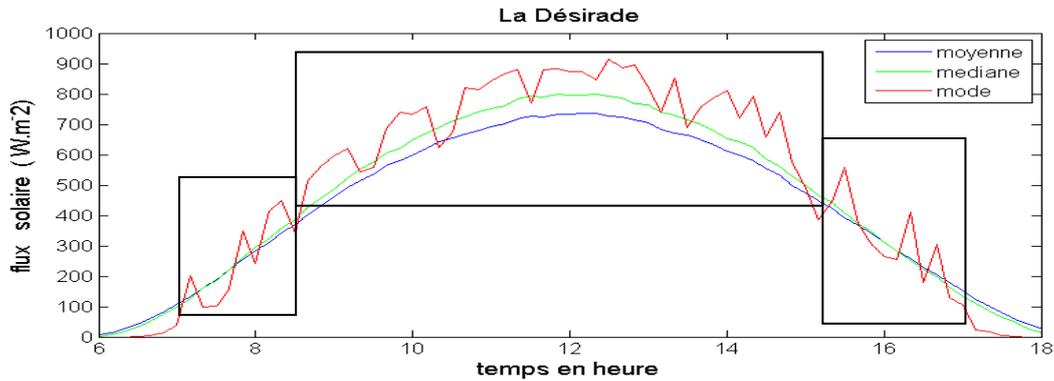


Figure 4-3. Courbe de la moyenne, du mode, de la médiane journalière sur trois années pour le site de La Désirade

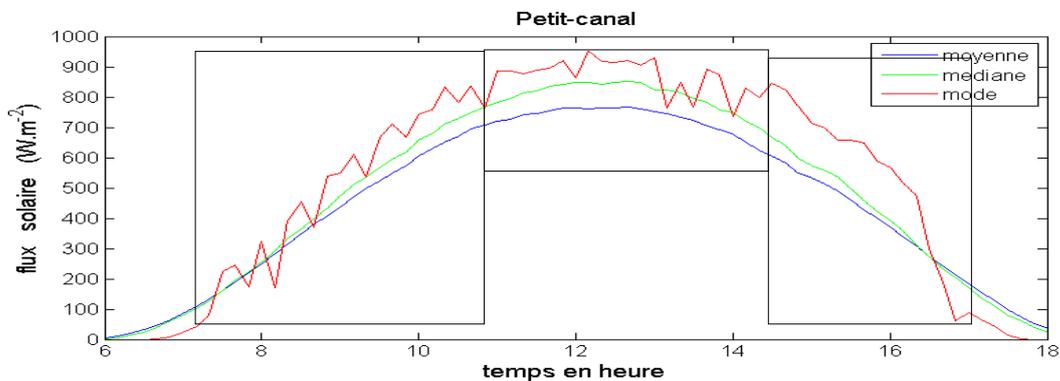


Figure 4-4. Courbe de la moyenne, du mode, de la médiane journalière sur trois années pour le site de Petit-canal

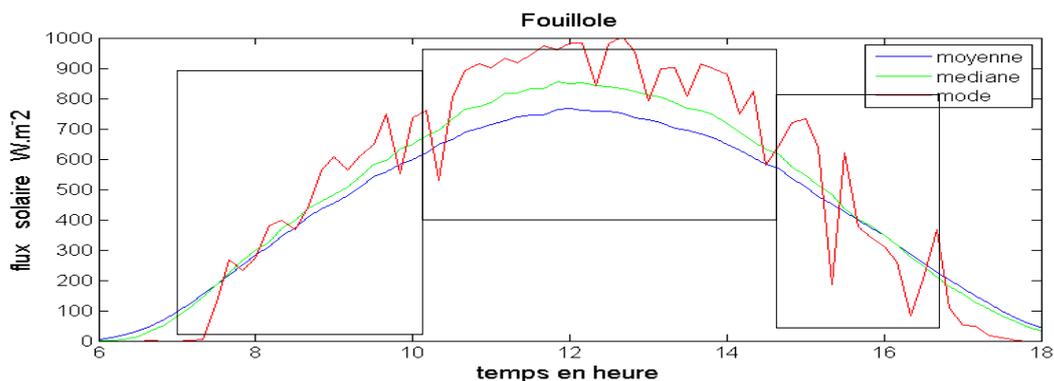


Figure 4-5. Courbe de la moyenne, du mode, de la médiane journalière sur trois années sur le site de Fouillole

Les courbes moyennes sont sensiblement égales sur nos trois sites. La médiane étant une fonction statistique qui donne une meilleure mesure de la tendance centrale en cas de valeurs extrêmes, montrent que les courbes médianes sont aussi sensiblement égales entre les trois sites de mesures. Le mode permet également de capturer d'éventuelle hétérogénéité qui ne se verrait pas en ne considérant que la moyenne. En effet, le mode représente la valeur la plus fréquente localement. En comparant les trois figures on peut constater que l'amplitude des fluctuations est plus importante sur le site de Fouillole.

De plus, on peut remarquer trois régimes de variations du flux solaire au cours de la journée. En effet, sur le site de la Désirade, on pourrait observer que le premier régime se situe dans une tranche horaire de 7h00 à 8h30, le deuxième régime de 8h30 à 15h et le troisième régime de 15h à 17h30 (Figure 4-3). Concernant le site de Petit-canal, on peut situer le premier régime entre 7h et 11h, le deuxième entre 11h et 13h30, le troisième entre 13h30 et 17h (Figure 4-34). Pour Fouillole, le premier régime serait entre 7h30 et 10h15, le second entre 10h15 et 15h40 et le troisième entre 15h40 et 17h40 (Figure 4-5).

On peut constater que le second régime des trois sites, qui se situe en milieu de journée présente le plus de variabilité comparée aux deux autres régimes.

Nous avons cherché à obtenir dans la suite de notre étude une connaissance plus fine des variations à court-terme constatées.

II.2.2 Variabilité à très court-terme de l'ensoleillement global

L'ensoleillement global présente une variabilité à très court-terme (inférieures à l'heure), dont une composante est distribuée de manière aléatoire, avec des journées pouvant présenter de fortes fluctuations. Ces fluctuations peuvent être rapides et de grande amplitude. Leur fréquence est liée non seulement à la vitesse de passage des nuages, à leur taille mais aussi à leur nombre. La Guadeloupe,

étant une île située en milieu intertropicale, elle présente des conditions météorologiques caractérisées par de très fréquents passages nuageux.

Dans cette section nous décrivons une méthode permettant d'obtenir les signaux de fluctuations et présenterons une analyse statistique descriptive de ces fluctuations à court terme.

II.2.2.1 Signaux des fluctuations

Afin d'étudier la variabilité spatio-temporelle de l'ensoleillement global, les séries de temps sans tendance doivent être utilisées. Dans plusieurs travaux de recherche telles que dans Hoff et Perez 2010, Pérez et al 2012, Badosa et al 2013, Yang et al 2014), c'est l'indice de clarté qui est utilisé pour évaluer la variabilité de l'ensoleillement et les corrélations entre les sites. Dans ce chapitre, nous analyserons la variabilité du flux solaire obtenue par les signaux de fluctuations calculés à partir d'une méthode simple basée sur la moyenne glissante.

Les signaux de fluctuations de l'ensoleillement global sont obtenus en retirant la tendance du signal mesuré. Pour cela, nous considérons le signal mesuré G décrit par :

$$G = G' + \overline{G_N} \quad (4.2)$$

où G est le signal mesuré, G' le signal de fluctuation et $\overline{G_N}$ le signal de moyenne glissante à N points. Ce dernier est obtenu selon l'expression suivante :

$$\overline{G_N}(t) = \frac{1}{N} \sum_{i=t-\frac{N-1}{2}}^{t+\frac{N-1}{2}} G(i) \quad (4.3)$$

Avec N fixé.

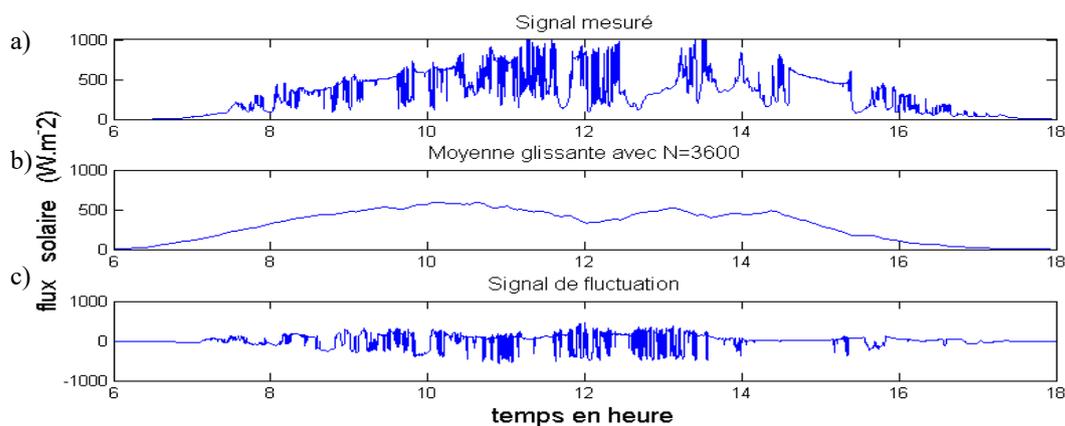


Figure 4-6. Exemple d'une journée de flux solaire mesuré a) et de sa moyenne glissante correspondante b) (journée du 1^{er} octobre 2010). c) est le signal de fluctuation obtenu.

L'utilisation d'une moyenne glissante peut être considérée comme une méthode de lissage des données. La moyenne glissante $\overline{G_N}$, qui est un filtre passe bas, rend compte principalement des variations du flux solaire sur des échelles de temps supérieures à N. La Figure 4-5 un exemple d'ensoleillement global mesuré pour une journée, le signal de moyenne glissante correspondant avec N=3600 et le signal de fluctuations obtenu. Ainsi, nous avons obtenu le signal de fluctuations filtré pour une échelle de temps de coupure égale à 1 h. Cette transformation a pour fonction d'atténuer les fréquences supérieures à sa fréquence de coupure (ici les fréquences correspondantes à une période inférieure à 1h) et ce, dans le but de conserver uniquement les basses fréquences. Nous obtenons donc un signal donnant des informations sur toutes les fluctuations supérieures à 1h.

Les variations irrégulières de l'ensoleillement global dues à la couverture nuageuse, quantifié par les fluctuations, sont alors étudiées pour caractériser la variabilité à court-terme du rayonnement solaire.

II.2.2.2 Statistique descriptive des signaux de fluctuations

Les fluctuations sont caractérisées par leur amplitude, leur fréquence et leur temps caractéristiques. Une connaissance fine de la dynamique de l'ensoleillement disponible en un lieu donné est nécessaire et indicative pour l'étude spatio-temporelle de la variabilité du flux solaire à court-terme. Dans cette étude, l'analyse statistique descriptive sera effectuée sur les signaux de fluctuations durant la période de l'ensoleillement c'est-à-dire de 7h à 17h (Tomson et Tamm (2005), Voyant (2011)), pour l'ensemble des journées de mesure (soit 36000 points par jour) sur trois années (2010, 2011, 2012)..

Amplitude des fluctuations

Afin de mettre en évidence les valeurs d'amplitudes des fluctuations et leur probabilité, la Figure 4-7 présente la fonction de répartition de l'amplitude des fluctuations pour chaque station.

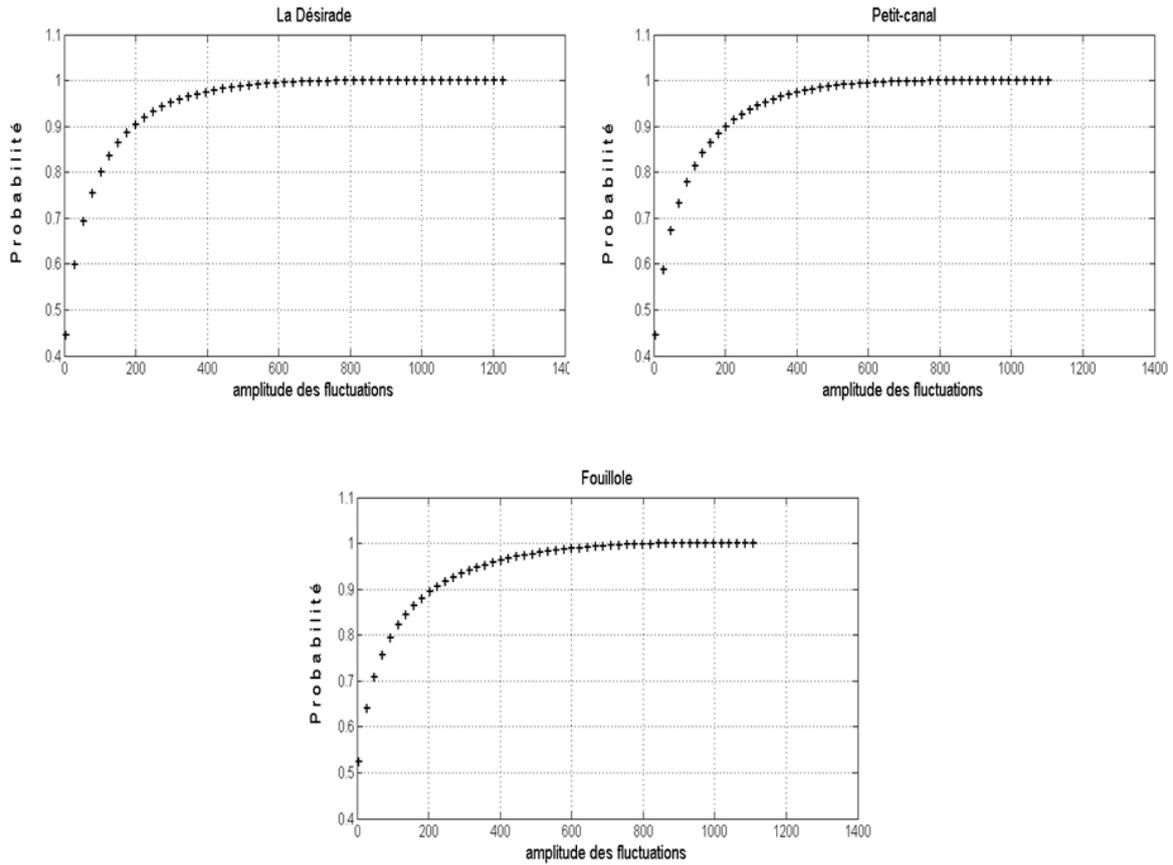


Figure 4-7 . Fonction de répartition de l'amplitude des fluctuations pour le site de Fouillole

On observe que 90% des fluctuations ont une amplitude inférieure à 200 W.m^{-2} sur les trois stations. On remarquera que l'amplitude des fluctuations peut atteindre des valeurs proches de la valeur d'ensoleillement mesurée (800 à 1000 W.m^{-2}). Néanmoins, la probabilité d'obtenir des amplitudes supérieures à 600 W.m^{-2} est très faible (environ 4%). Bien que les valeurs extrêmes d'amplitude (supérieures à 1100 W.m^{-2}) soient fortement liées aux erreurs de mesures, on peut constater que les signaux de fluctuations pour le site de La Désirade présente un intervalle de valeurs d'amplitude des fluctuations plus grand que pour le site de Petit-canal et de Fouillole. On peut noter toutefois les mêmes résultats de probabilité d'amplitude des fluctuations entre les trois stations.

Durée des fluctuations (intervalle de temps stable ou instable)

Pour compléter notre analyse statistique descriptive de la variabilité à court-terme de l'ensoleillement global, nous étudierons la durée caractéristique des fluctuations. L'étude de la probabilité des amplitudes a montré que les fluctuations probables à environ 97% ont des valeurs inférieures ou égales à 600 W.m^{-2} . Ainsi, nous ciblerons les amplitudes de fluctuations inférieures ou égales à 600 W.m^{-2} pour la suite de l'analyse statistique.

Les fluctuations sont évaluées par augmentation de la variation du signal. Ces différences $\Delta G'$ peuvent être positives ou négatives. On considère comme fluctuations uniquement les valeurs $\Delta G' > 5 \text{ W.m}^{-2}$.

Les durées caractéristiques des fluctuations sont assimilées aux calculs des temps de « montée » des fluctuations.

Afin de mettre en évidence les échelles de temps caractérisant les fluctuations, nous avons étudié la probabilité de la durée (en pourcentage) sur laquelle une variation de flux s'étend dont l'amplitude appartient à un intervalle donné. Ces probabilités ont été obtenues en considérant l'ensemble des journées de mesure pour les trois stations. Les figures suivantes: Figure 4-8 ; Figure 4-9 ; Figure 4-10, représentent ainsi les durées caractéristiques des fluctuations correspondant à un intervalle de valeurs d'amplitudes pour chaque localisation. Par analyse des figures Figure 4-58 b) et c); Figure 4-59 b) et c); figure 4-10 b) et c) on remarque que la durée des fluctuations est une fonction croissante de l'amplitude des fluctuations. Les figures : Figure 4-58 a); Figure 4-59 a); Figure 4-510 a), permettent d'établir que quelque soient les amplitudes considérées, 90% des fluctuations ont une durée caractéristique inférieure à 200 secondes soit environ 4 min. Toutefois, les fluctuations peuvent atteindre une durée de 2300 secondes soit près de 38 minutes. Un dernier commentaire peut être donnée à partir des figures Figure 4-58 c); Figure 4-59 c); Figure 4-510 c): quelle que soit l'amplitude des fluctuations, la plus forte probabilité de leur durée correspond à un temps caractéristique de 8 à 20 secondes (densité comprise entre 0.10 et 0.20).

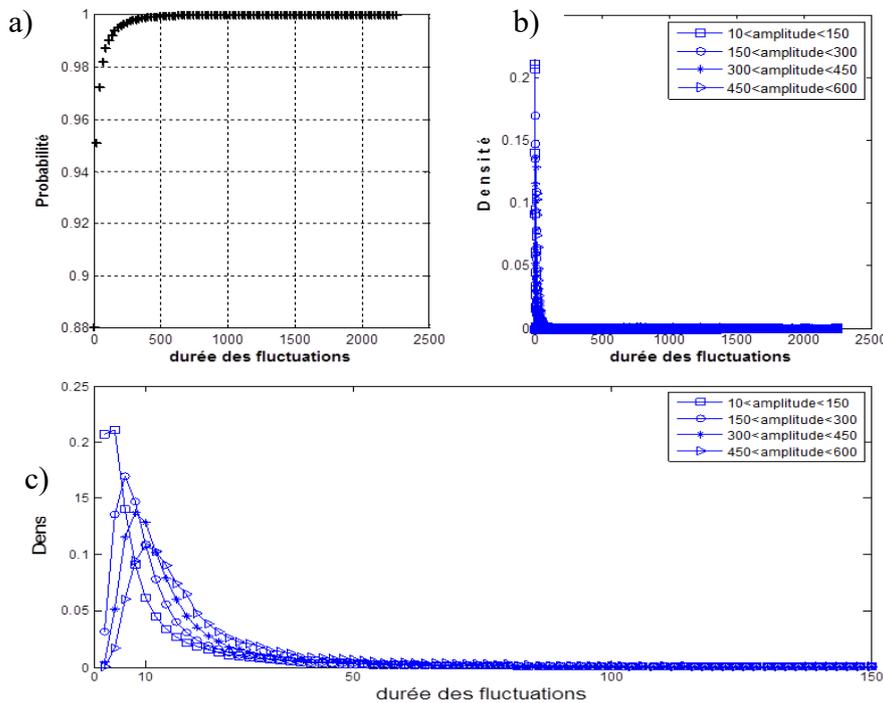


Figure 4-8. a) Fonction de répartition de la durée des fluctuations et b) densité de probabilité de l'échelle de temps Δt (durée des fluctuations en secondes) pour une gamme d'amplitude des fluctuations $\Delta G'$ pour le site de La Désirade. c) Zoom de la figure b).

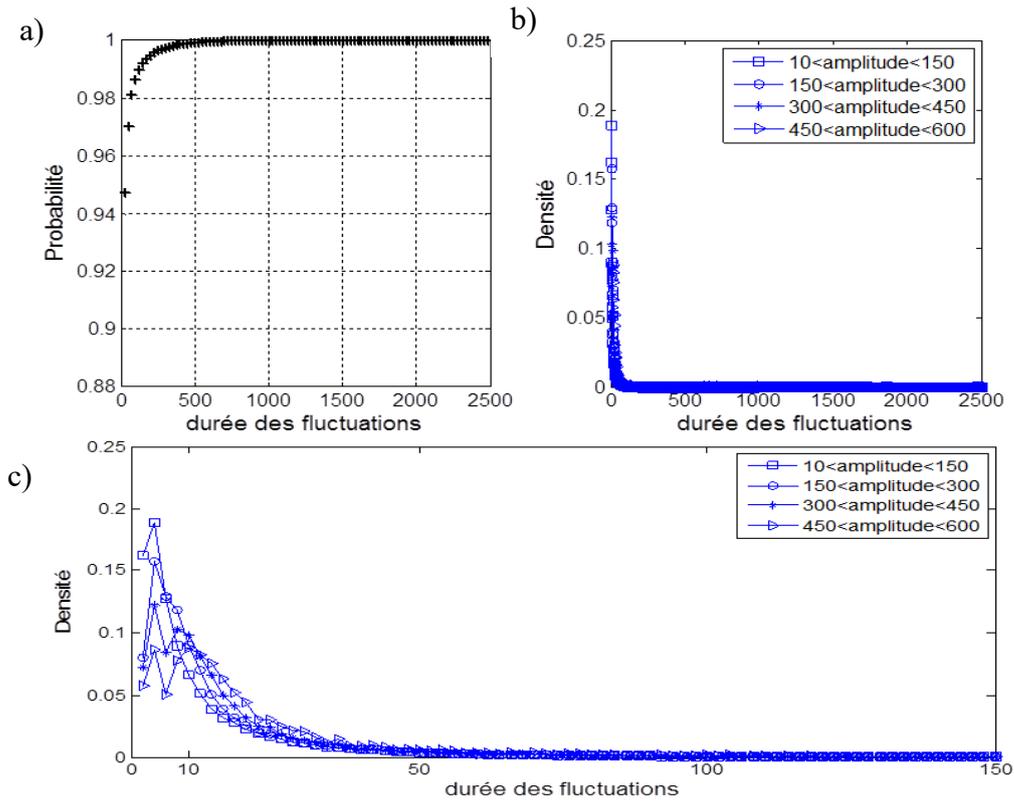


Figure 4-9. a) Fonction de répartition de la durée des fluctuations et b) densité de probabilité de l'échelle de temps Δt (durée des fluctuations en secondes) pour une gamme d'amplitude des fluctuations $\Delta G'$ pour le site de Petit-canal. c) Zoom de la figure b).

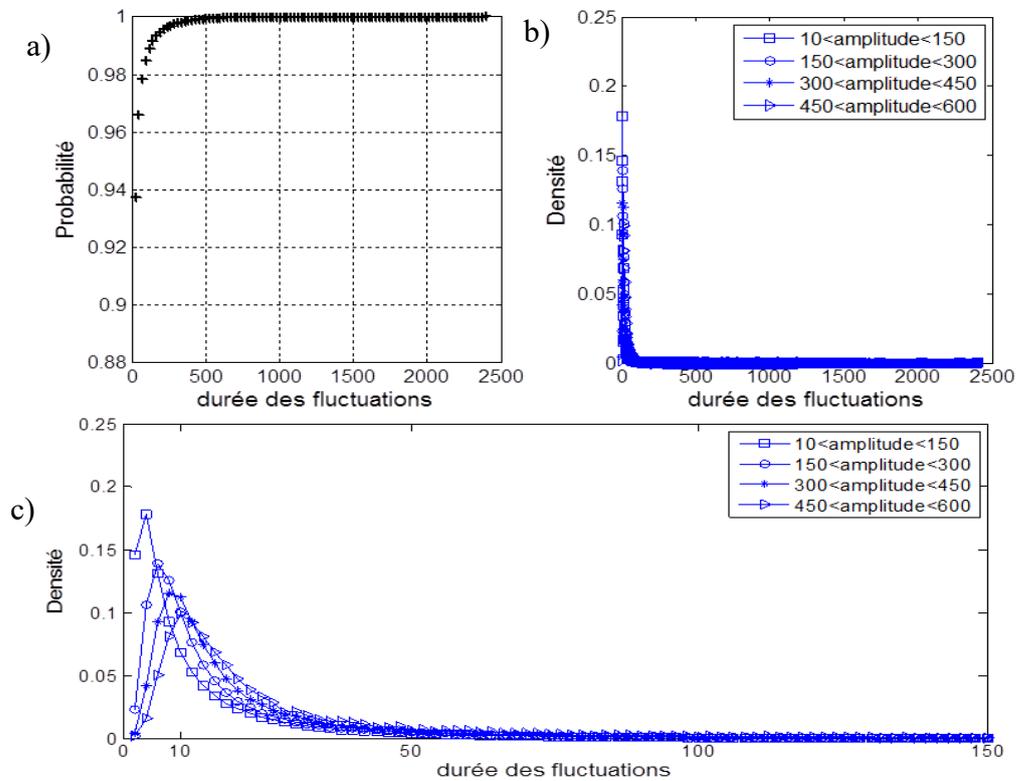


Figure 4-10. a) Fonction de répartition de la durée des fluctuations et b) densité de probabilité de l'échelle de temps Δt (durée des fluctuations en secondes) pour une gamme d'amplitude des fluctuations $\Delta G'$ pour le site de Fouillole. c) Zoom de la figure b).

Les résultats de l'analyse statistique descriptive de la variabilité de l'ensoleillement à très court terme ont montré une même dynamique d'évolution dans le temps entre les 3 stations. Néanmoins, une différence d'un point de vue énergétique entre les stations de mesures est constatée. L'étude de la fraction d'énergie solaire sur chacune des stations a révélé une plus grande variabilité de l'ensoleillement global soit un plus grand nombre de passages nuageux sur la station de Fouillole comparé à Petit-canal et La Désirade.

III. Variabilité spatio-temporelle du flux solaire

III.1. Aperçu des travaux dans la littérature

L'énergie solaire est une ressource qui varie à la fois dans l'espace et dans le temps particulièrement à échelle de temps courte. La variabilité spatio-temporelle du rayonnement solaire fait l'objet d'initiatives de recherches majeures comme aux Etats-Unis : CSI, 2010; USDOE, 2009; BNL, 2010; SMUD, 2010; et en France : IEA, 2010 (Pérez et al (2012)). De nombreux travaux de recherches mettent en évidence cette variabilité spatio-temporelle par l'étude de l'ensoleillement sur plusieurs stations de mesures tels que Barnet et al (1997), Soubdhan et al (2007), Murata et al (2009), Mills et Wiser (2010), Gueymard et Wilcox (2011), Lave et Kleissl (2010), Lave et Kleissl (2012), Hoff et Perez (2011), Pérez et al (2011), Pérez et al (2012), Badosa et al (2013) . Lave et Kleissl (2010) ont analysé le rayonnement solaire global à 5 minutes de résolution pour quatre sites de l'état du Colorado. Le but était d'examiner l'effet de lissage des sites de PV distribués géographiquement. Ils ont trouvé des corrélations entre les sites pour des échelles de temps supérieures à 3 h. Ils ont également montré que la station virtuelle composée de la moyenne des quatre sites n'avait pas d'extrême rampe et a eu une réduction de manière significative des taux de rampe. Ces travaux de recherches ont permis d'apporter une avancée pour l'exploitation distribuée des sites PV.

Hoff et Pérez (2011) ont étudié la corrélation des variations de l'indice de clarté en utilisant les données de rayonnement solaire dérivées des satellites pour un réseau virtuel de 24 localisations aux Etats-Unis. Les résultats ont montré une corrélation décroissante avec la distance et augmentant avec les intervalles de temps. Un modèle a été présenté simulant l'évolution du coefficient de corrélation en fonction de la distance de paire de station et également en fonction de l'intervalle de temps.

Badosa et al (2013) présente une étude faite à la Réunion, une île tropicale au Sud Ouest de l'Océan Indien. Ces travaux étudient la nébulosité et le rayonnement solaire à différentes échelles (grande échelle, méso-échelle et échelle locale). Sept stations de mesures sont utilisées à travers l'île correspondant à trois régimes climatiques différents (la côte au vent, côte sous le vent et côte parallèle à la direction générale du vent). Ils ont décrit la variabilité par une classification des journées d'ensoleillement avec leur fréquence d'occurrence et une comparaison des différentes classes entre les

stations. Les coefficients de corrélation de la variabilité de l'indice de clarté journalière et horaire entre toutes les paires de stations sont étudiées en fonction de la distance de station à station. Les résultats d'été et d'hiver ont été traités séparément et une fonction simulant l'évolution des coefficients est déterminée par une expression proposée dans Hoff et Perez (2011). Les distances correspondant à une décorrélation pour l'indice de clarté journalier entre deux stations sont trouvées assez grandes pour la période d'été (62 km) due à l'occurrence de systèmes de nuages méso-échelles affectant l'île entière. En hiver, cette distance atteint 29 km. A l'échelle de temps d'une heure, la distance est bien plus faible (4 km en été et 3 km en hiver) montrant la forte variabilité spatiale et les avantages de l'agrégation des stations. Ces résultats sont en accord avec ceux trouvés pour l'île tropicale d'Hawaï par Hoff et Perez (2011). Notre étude spatio-temporelle de la variabilité de l'ensoleillement global entre trois stations sur l'île de la Guadeloupe viendra compléter ces travaux de recherche traités en milieu insulaire. Dans les sections suivantes, les outils statistiques utilisés pour l'analyse spatio-temporelle sont premièrement présentés. Les résultats de corrélation intersite obtenus sont ensuite discutés. Puis, nous élargissons notre étude à une interprétation des phénomènes physiques liés à ces corrélations.

III.2. Evaluation des corrélations : Outils statistiques utilisés

Cette section présente les outils statistiques utilisés pour comparer les séries temporelles de fluctuations entre les sites étudiés, en particulier les valeurs de corrélation de Pearson, de cohérence, covariance et intercorrélation. Les études des corrélations permettront de mesurer à quel seuil une station peut affecter une autre en terme de variabilité de la ressource solaire.

III.2.1 Corrélation de Pearson

Le coefficient de corrélation de Pearson est le plus fréquemment utilisé en statistique pour mesurer la covariabilité linéaire entre deux variables. Dans la littérature l'idée essentielle de corrélation a été conçue par Francis Galton et a été formellement développée par Karl Pearson d'où le nom de corrélation de Pearson. Bien que le coefficient de corrélation ne mesure pas la relation causale entre deux variables, il joue un rôle important dans plusieurs domaines scientifiques (finance, économie...) puisque qu'il permet d'évaluer 'la co-relation' entre deux variables.

Considérons deux processus stochastiques $\{x_t\}_{t=1,\dots,T}$ et $\{y_t\}_{t=1,\dots,T}$ où x_t est supposé être identiquement distribué avec la fonction de répartition F_x et y_t est supposé être identiquement distribué avec la fonction de répartition F_y . La mesure conventionnelle de corrélation dénotée C , appelé coefficient de Pearson est donné par :

$$C = \frac{E[(x_t - \mu_x)(y_t - \mu_y)]}{\sqrt{\sigma_x^2} \sqrt{\sigma_y^2}} \quad (4.4)$$

Où $\mu_x = E(x_t)$, $\mu_y = E(y_t)$, $\sigma_x^2 = E[(x_t - \mu_x)^2]$, $\sigma_y^2 = E[(y_t - \mu_y)^2]$ et l'espérance E est calculée en accord avec la fonction de répartition de x_t et y_t . Usuellement, l'estimation de C remplace l'espérance de l'échantillon E par la moyenne de chaque processus stochastique :

$$\hat{C} = \frac{\frac{1}{T} \sum_{t=1}^T [(x_t - \hat{\mu}_x)(y_t - \hat{\mu}_y)]}{\sqrt{\hat{\sigma}_x^2} \sqrt{\sigma_y^2}} \quad (4.5)$$

Où $\hat{\mu}_x = \frac{1}{T} \sum_{t=1}^T x_t$, $\hat{\mu}_y = \frac{1}{T} \sum_{t=1}^T y_t$, $\hat{\sigma}_x^2 = \frac{1}{T} \sum_{t=1}^T (x_t - \hat{\mu}_x)^2$ et $\hat{\sigma}_y^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu}_y)^2$

Le coefficient de corrélation est compris entre -1 et 1. Il est égal à 1 dans le cas où l'une des variables est une fonction affine croissante de l'autre variable, à -1 dans le cas où une variable est une fonction affine et décroissante. Les valeurs intermédiaires renseignent sur le degré de dépendance linéaire entre les deux variables. Plus le coefficient est proche des valeurs extrêmes -1 et 1, plus la corrélation entre les variables est forte ; on emploie simplement l'expression « fortement corrélées » pour qualifier les deux variables. Une corrélation égale à 0 signifie que les variables ne sont pas corrélées.

III.2.2 Cohérence

Le concept de cohérence des séries chronologiques a été introduit par Wiener (1949) et largement développé par Rosenberg et al. (1989), pour ses applications aux données d'imagerie fonctionnelle. Bien que la corrélation soit définie dans le domaine temporel, les mesures de cohérence sont effectuées dans le domaine fréquentiel. La cohérence dans le domaine des énergies solaires a été démontrée comme étant une statistique utile pour étudier l'échelle de temps seuil pour lequel les corrélations entre les sites sont faibles comme indiqué par Lave et al (2012). Les mesures de cohérence représentent implicitement les retards des effets d'une région sur une autre. Si une série temporelle mesurée à une localisation serait globalement similaire à celle de l'autre, mais avec un décalage temporel, alors la corrélation d'ordre zéro ordinaire entre les deux sera modérée ou faible; la cohérence, en revanche, sera élevée dans la bande passante de la courbe puisqu'elle tient compte du décalage temporel. La fonction de cohérence est définie par :

$$c_{XY}(f) = \frac{|P_{XY}(f)|^2}{P_{XX}(f)P_{YY}(f)} \quad (4.6)$$

où $P_{XX}(f)$ et $P_{YY}(f)$ sont respectivement la densité de puissance spectrale des signaux X et Y, et $P_{XY}(f)$ est l'inter densité de puissance spectrale de X et Y.

Une valeur $c_{XY}(f)$ de proche de 1 signifie que X est corrélée à Y pour une fréquence donnée. Le coefficient au carré dans la cohérence peut être interprété comme la proportion d'énergie dans une des séries de temps (à une fréquence donnée), qui peut s'expliquer par une régression linéaire de l'autre série au cours de l'évolution du temps. Comme une mesure de la corrélation spatiale des signaux de fluctuations, nous avons calculé le spectre de cohérence par paire de sites. Le spectre de cohérence fournit la covariance normalisée à chaque fréquence, permettant une analyse de la corrélation à des échelles de temps courtes.

III.2.3 Covariance

La fonction de covariance est définie telle que :

$$C_{XY}(t1, t2) = E|(X(t1) - \mu_X)(Y(t2) - \mu_Y)| \quad (4.7)$$

où X et Y sont les signaux de fluctuations journaliers pour deux, μ_X et μ_Y sont respectivement les moyennes statistiques de X et Y, $t2 = t1 - \tau$ avec $\tau=0$. La matrice de covariance R utilisée pour notre étude est calculée avec la fonction de covariance normalisée par la variance de X et Y. C'est une matrice carrée composée en première ligne des coefficients de covariance pour les sites comparés à Petit-Canal, la seconde ligne pour les sites comparés à La Désirade et la troisième pour les sites comparés à Fouillole. Nous obtenons une matrice carrée où M (1,2) est le coefficient de covariance pour Petit-Canal/La Désirade, M (1,3) pour Petit-Canal/Fouillole et M(2,3) pour La Désirade/Fouillole. Cette matrice est symétrique.

Petit-canal DesiradeFouillole

$$M = \begin{matrix} & \begin{matrix} \text{Petit-canal} \\ \text{Desirade} \end{matrix} & \begin{pmatrix} 1 & 0.4228 & 0.3162 \\ 0.4228 & 1 & -0.0664 \\ 0.3162 & -0.0664 & 1 \end{pmatrix} \end{matrix}$$

$$P = \begin{pmatrix} 1 & 5.9 \cdot 10^{-11} & 2.810^{-11} \\ 5.9 \cdot 10^{-11} & 1 & 4.3 \cdot 10^{-11} \\ 2.810^{-11} & 4.3 \cdot 10^{-11} & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 5.9 \cdot 10^{-11} & 2.810^{-11} \\ 5.9 \cdot 10^{-11} & 1 & 4.3 \cdot 10^{-11} \\ 2.810^{-11} & 4.3 \cdot 10^{-11} & 1 \end{pmatrix}$$

Tableau 2: Matrice de covariance et de P-value pour un jour donnée (9 Mai 2010) entre les trois sites

Lorsque l'hypothèse nulle H_0 est acceptée alors $H_0 = 0$ à contrario lorsqu'elle est rejetée on trouve $H_0 = 1$. Pour ces coefficients de corrélation, on se fixe un seuil de confiance $\alpha = 0.05$. Cela signifie que l'on autorise que l'hypothèse soit rejetée à tort (alors qu'elle est vraie) dans 5 % des cas. De fait la valeur de p doit être supérieure au seuil de confiance α pour ne pas rejeter l'hypothèse H_0 . Les coefficients de covariance sont significatifs à 95% de confiance.

Un exemple Tableau 2 présente la M et P matrice (matrice des p-value) pour une journée (9 mai 2010). Une limite de cette mesure est qu'elle ne se soucie que de la tendance générale. Par exemple si $Y=2X$, alors X et Y sont corrélés parfaitement positives. Néanmoins, si X et Y étaient corrélés avec un décalage temporel, la corrélation semblerait faible. Une méthode pour mesurer la corrélation avec un décalage temporel entre deux séries est l'utilisation de la fonction d'intercorrélation.

III.2.4 Intercorrélation

La fonction d'intercorrélation $R_{XY}(\tau)$ est calculée selon l'expression suivante (Zhao et al., 2011) :

$$R_{XY}(\tau) = \begin{cases} \sum_{n=0}^{N-t-1} \frac{X_{n+\tau} Y_n}{\sigma_X \sigma_Y} & \tau \geq 0 \\ R_{YX}(-\tau) & \tau < 0 \end{cases} \quad (4.8)$$

où X et Y sont les signaux de fluctuations pour deux sites donnés, τ est le temps de retard entre les signaux. Si X et Y sont des vecteurs de longueur N. Les coefficients sont calculées pour tous les décalages temporels $\tau = 1 - N, 2 - N, \dots, 0, 1, \dots, N - 1$. Les résultats d'intercorrélation résultent en une séquence $2N-1$, chaque élément étant le coefficient de corrélation maximal entre X et un décalage de Y lorsque nous faisons glisser Y le long de l'axe des x.

Par conséquent, si Y ne diffère par un décalage de k pas en avant, nous devrions trouver le coefficient maximale de corrélation croisée à $\text{corr}(k)$. Comparée à la cohérence cette fonction nous permettra d'analyser les corrélations spatio-temporelles à court terme dans le domaine temporel et non fréquentielle. Les Corrélations croisées ou intercorrélation permettront alors d'évaluer les relations linéaires retardés ou différés des variabilités de flux solaire entre les stations de mesures. La corrélation entre la station X et la station Y à des décalages positifs indiquerait une relation entre la variabilité de la région X et la variabilité ultérieure de la région Y, ou vice versa. L'intercorrélation et le coefficient de Pearson sont des méthodes souvent utilisées pour définir une distance seuil qui spécifie la distance maximale entre les stations pour lesquelles les valeurs d'ensoleillement devrait être corrélées (Barnet et al (1997), Murata et al (2009), Mills et Wiser (2010), Lave et Kleissl (2010), Hoff et Perez (2011), Pérez et al (2011), Pérez et al (2012), Badosa et al (2013), Yang et al 2014). La fonction d'intercorrélation a été particulièrement utilisée dans Yang et al 2014, pour quantifier le

retard temporel entre deux stations en utilisant le coefficient maximal d'intercorrélation de l'indice de clarté entre deux stations.

On s'intéressera d'une part à l'évolution de la corrélation spatiale pour nos trois paires de stations avec un échantillonnage fixé. D'autre part, l'analyse de la corrélation temporelle sera effectuée en étudiant l'intercorrélation avec séparation d'échelle. Le but étant de déterminer à quelle échelle de temps seuil les corrélations intersites deviennent significatives. D'autre part, la probabilité des coefficients maximaux d'intercorrélation est analysée pour chaque paire de sites (Petit-Canal/Fouillole, Petit-Canal/ Désirade, et Désirade/Fouillole). Ce qui permettra d'identifier la décroissance des corrélations, à savoir si elle est liée aux distances isotropiques. Badosa et al (2013) ont précisé que la décorrélation entre deux séries de temps est définie pour des coefficients de corrélation inférieurs à 0,5. Par observation des signaux, leur coefficient d'intercorrélation et en accord avec cet article, nous définirons une corrélation significative entre les signaux de fluctuations pour des coefficients supérieurs ou égaux à 0,6. Le décalage temporel des coefficients maximaux d'intercorrélation sera également évalué pour l'analyse spatio-temporelle de la variabilité du flux solaire d'un point de vue dynamique.

III.3. Analyse des corrélations inter sites spatiales

Plusieurs approches existent pour étudier la variabilité spatiale dans la littérature comme mentionné en section I.3. L'étude préliminaire présentée dans Hoff et Pérez (2011) a montré l'évolution de la corrélation des variations de l'indice de clarté en utilisant les données de rayonnement solaire (dérivées des satellites) pour 24 stations pour des distances intersites allant de 100m à 100km. Les résultats des coefficients ont montré une corrélation décroissante avec la distance et augmentant avec les intervalles de temps. Une étude plus récente dans la suite de ces premiers travaux, discutée dans Pérez et al (2012), a démontré des corrélations de paires de stations comme une fonction de distance station-à-station donnant ainsi l'évolution spatiale du rayonnement solaire aux Etats-Unis. Nous utiliserons ces travaux de recherche comme référence en effectuant la même analyse de l'évolution des corrélations spatiales sur nos stations.

III.3.1 Stationnarisation des données

En première phase, nous avons cherché à rendre stationnaire les données de mesures par la même méthode présentée dans Pérez et al (2012). En effet, il est fondamental d'utiliser les séries temporelles de flux solaire sans tendance pour étudier les éventuelles corrélations inter-sites. Dans plusieurs travaux de recherche telles que dans Hoff et Perez 2010, Badosa et al 2013, Yang et al 2014, c'est l'indice de clarté qui est utilisé, notamment dans Pérez et al (2012), le Δk_t^* est la différenciation à l'ordre 1 de l'indice ciel clair. L'indice de clarté a pour équation :

$$k_t = \frac{G}{G_0} \quad (4.9)$$

Le flux solaire hors atmosphère G_0 est une courbe "idéale", qui n'est sujette à aucunes fluctuations rapides mais rend compte de la variation du flux solaire due à la position du soleil par rapport au point de mesure et G les données de flux solaire mesurées. Le rapport $\frac{G}{G_0}$, rend compte d'une part de l'atténuation du flux solaire due à la traversée de l'atmosphère mais également des fluctuations dues aux nuages. Pour notre cas d'étude le flux solaire hors atmosphère est calculé par le modèle de Kasten.

L'indice ciel clair a pour équation :

$$k_c = \frac{G}{G_c} \quad (4.10)$$

Le flux solaire G_c est une courbe "idéale", qui n'est sujette à aucunes fluctuations rapides mais rend compte de la variation du flux solaire théoriquement mesurée par le capteur.

D'où Δk_t^* :

$$\Delta k_t^* = k_c(t) - k_c(t - 1) \quad (4.11)$$

Pour notre étude, nous utiliserons le Δk_t qui est la différenciation à l'ordre 1 de l'indice de clarté :

$$\Delta k_t = k_t - k_{t-1} \quad (4.12)$$

Nous avons constaté effectivement que la normalisation avec le Δk_t^* était plus importante qu'avec le Δk_t , mais les valeurs de corrélations n'ont pas changé. En effet, les variations des deux signaux sont les mêmes (voir Figure 4-11), il n'y a qu'une différence d'amplitude ce qui n'influence pas les valeurs de corrélations. Notre étude des corrélations spatiales sera effectuée avec le paramètre Δk_t .

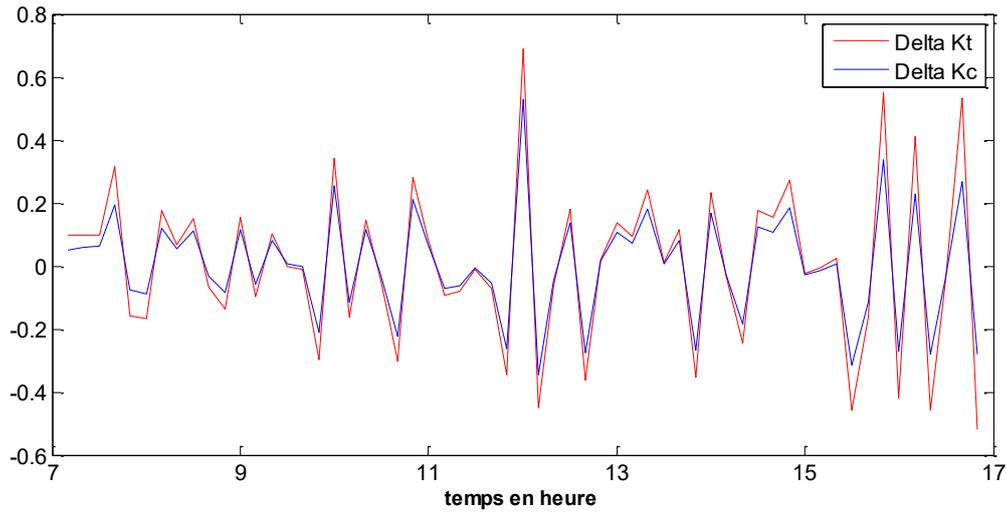


Figure 4-11. Echantillon d'une journée à haute variabilité (1^{er} janvier 2012) du signal Δk_t en rouge (différentiation de l'indice de clarté) et du signal Δk_t^* en bleu (différentiation de l'indice ciel clair) pour le site de La Désirade.

La Figure 4-12 représente un exemple de signal Δk_t obtenu pour journée présentant une importante variabilité.

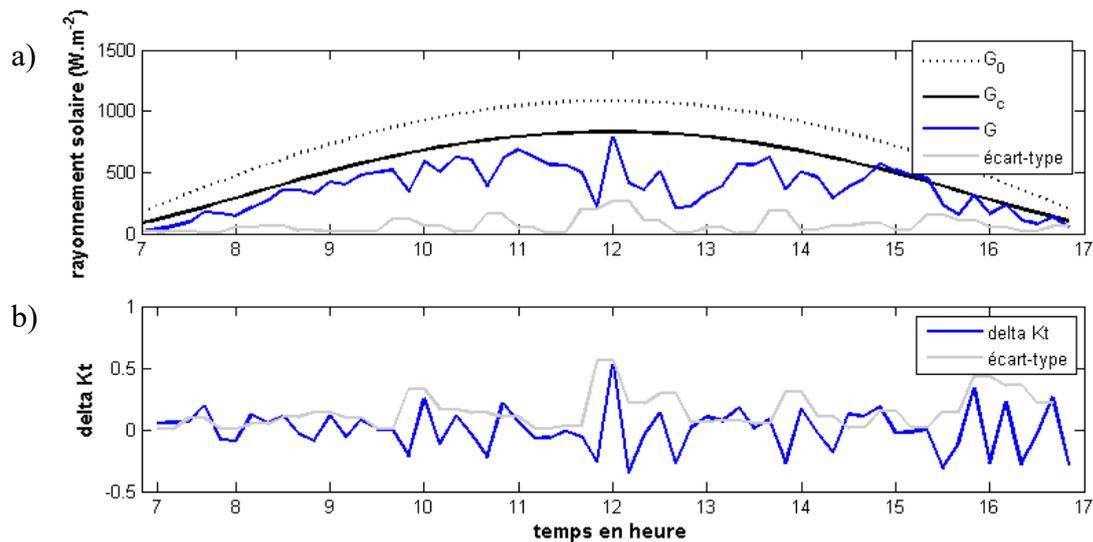


Figure 4-12. Echantillon d'une journée à haute variabilité (1^{er} janvier 2012) montrant l'ensoleillement global mesuré (G) par pas de 20s, l'ensoleillement extraterrestre G_0 , G_c l'ensoleillement théoriquement mesuré par le capteur et l'écart-type sur 15 minutes. b) Signal Δk_t^* correspondant.

Comme on peut l'observer sur la Figure 4-12 b), ce paramètre offre deux avantages : la normalisation de la variabilité à l'unité, et la suppression des effets de géométrie solaire qui sont communs à toutes les stations, due à la tendance journalière. Il évite alors de fausser les résultats de corrélation intersites.

III.3.2 Corrélation spatiale intersite

L'objectif est de comprendre comment la corrélation de paire de station varie en fonction des distances pour un intervalle d'échantillonnage fixé. En parallèle avec l'étude de Pérez et al (2012), les corrélations intersites sont calculées par le coefficient de Pearson entre les stations sur les signaux de fluctuations Δk_t .

Pour l'exemple de l'évolution de la corrélation spatiale d'un seul jour, nous avons choisi un jour fortement variable (1^{er} janvier 2012) illustrée dans la Figure 4-5. La corrélation de paire de station pendant ce jour est tracée dans la figure Figure 4-5, comme une fonction de distance pour chaque intervalle d'échantillonnage. Selon Pérez et al (2012), la corrélation atteint zéro pour les signaux de fluctuations à respectivement 500 m, 1km, 4 km et 10 km pour les intervalles d'échantillonnage de 20 s, 1 min, 5 min et 15 min. Au-delà du passage à zéro, les paires de stations considérées ne présentent pas de corrélation positive significative. Il est intéressant de noter que dans chaque cas, la corrélation devient légèrement négative au-delà du point de passage à zéro. La corrélation négative indique qu'à une certaine distance ce jour-ci, il se produit un phénomène de succession de périodes nuageuses et ensoleillées qui a tendance à être en opposition de phase pour toutes les échelles de temps sur deux stations. Dans notre cas, pour cette journée les coefficients sont positifs mais quasi nulles (0,01 à 0,2).

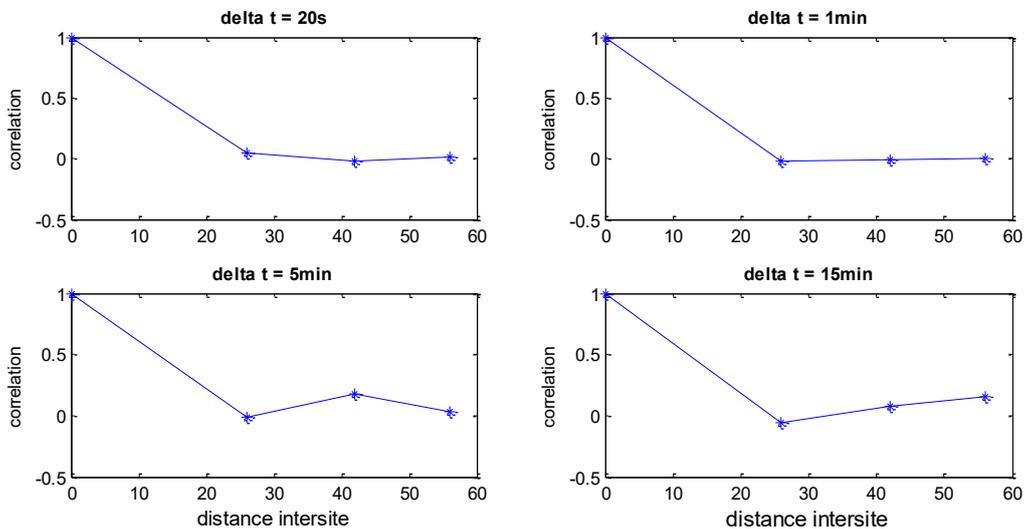


Figure 4-13 Corrélation de paires de stations en fonction des distances inter sites et des échelles de temps pour un échantillon d'une journée à haute variabilité.

La Figure 4-5 présente les résultats obtenus à partir de la même méthode que précédemment, mais sur tous les 366 jours analysés, pour trois paires de stations. La tendance moyenne de l'évolution des corrélations spatiales de cet échantillon est montrée sur la Figure 4-55 (ligne en rouge). La ligne en

pointillé est la moyenne plus ou moins deux fois l'écart-type qui représente un intervalle de confiance à 95%.

L'étroitesse de l'intervalle de confiance, particulièrement pour des intervalles d'échantillonnage courts (20 s et 1 minute), indique que l'évolution de la corrélation spatiale intersite pour un jour individuel considéré ne sort pas significativement de la tendance, autrement dit la corrélation spatiale intersite évolue de manière similaire quelque soient les jours considérés, pour un échantillonnage très court.

Les points verts (Figure 4-15 et Figure 4-56) sont les coefficients calculés à partir de la formule dans Pérez et al (2012) :

$$C_{pair} = \exp\left(\frac{x \ln(0.2)}{1.5 * \Delta t * CS}\right) \quad (4.13)$$

x est la distance, Δt est le pas de temps et CS la vitesse des nuages que nous avons associé à la vitesse du vent. N'ayant pas accès aux données de vent, nous avons estimé la vitesse de vent en calculant la moyenne annuelle des vitesses mensuelles fournit dans le bilan climatique de l'année 2012 par Météo France, soit 25 km/h.

Les coefficients C_{paire} calculés pour nos échantillons ont des valeurs très proches des coefficients de corrélation de pearson moyennés comme le montre la Figure 4-55 et la Figure 4-516. L'absence de corrélation se produisant régulièrement dans tous les sites à des distances au-delà du pic de corrélation négative (500 m, 1 km, 4 km et 10 km pour les intervalles d'échantillonnage de 20 s, 1 min, 5 min et 15 min) présenté dans Pérez et al (2012), est tenu dans notre système spatio-temporel composé de données expérimentales provenant de différents environnements climatiques pour trois stations (à des distances de 26 km, 42 km et 52 km). L'analyse de Pérez reste donc valide dans notre cas d'étude.

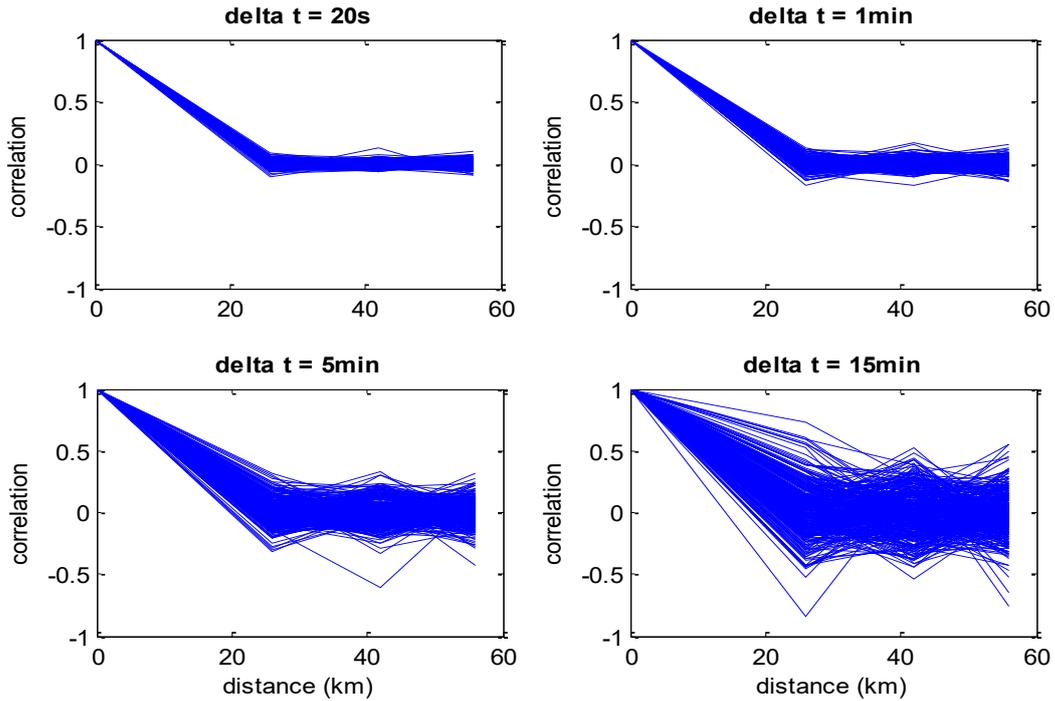


Figure 4-14 Corrélation de paires de stations en fonction des distances inter sites et des échelles de temps pour un échantillon d'une année (366 jours).

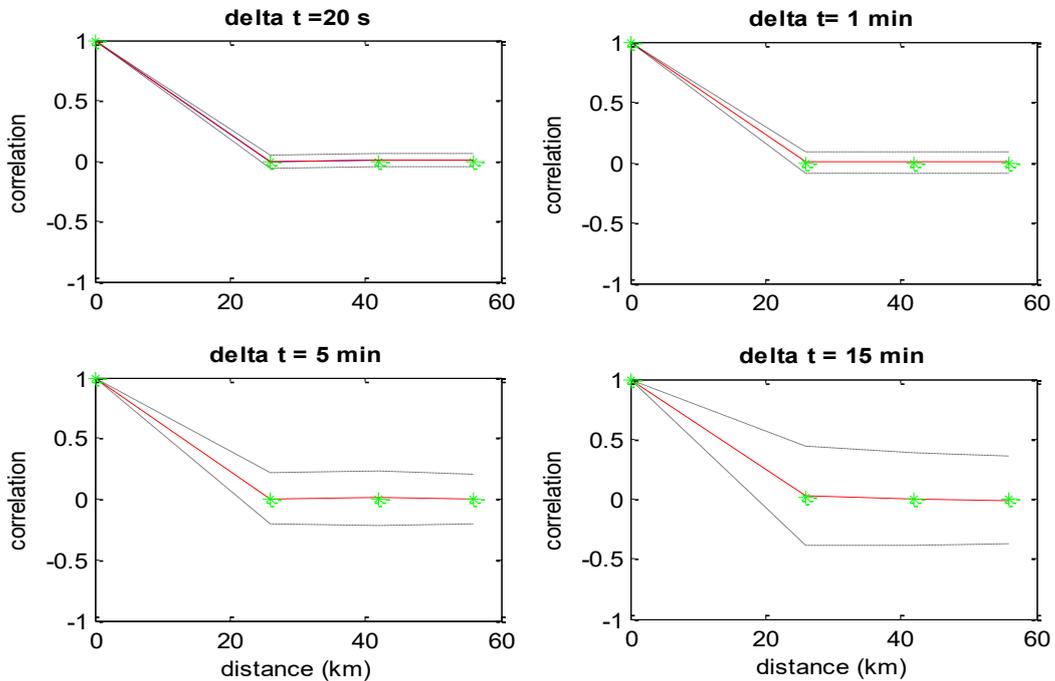


Figure 4-15 Tendence moyenne des corrélations de paires de stations en fonction des distances inter sites et des échelles de temps pour un échantillon d'une année (366 jours). Les lignes en pointillé représentent l'intervalle de confiance à 95% et les points verts les C_{paire} calculés par l'équation (3.11).

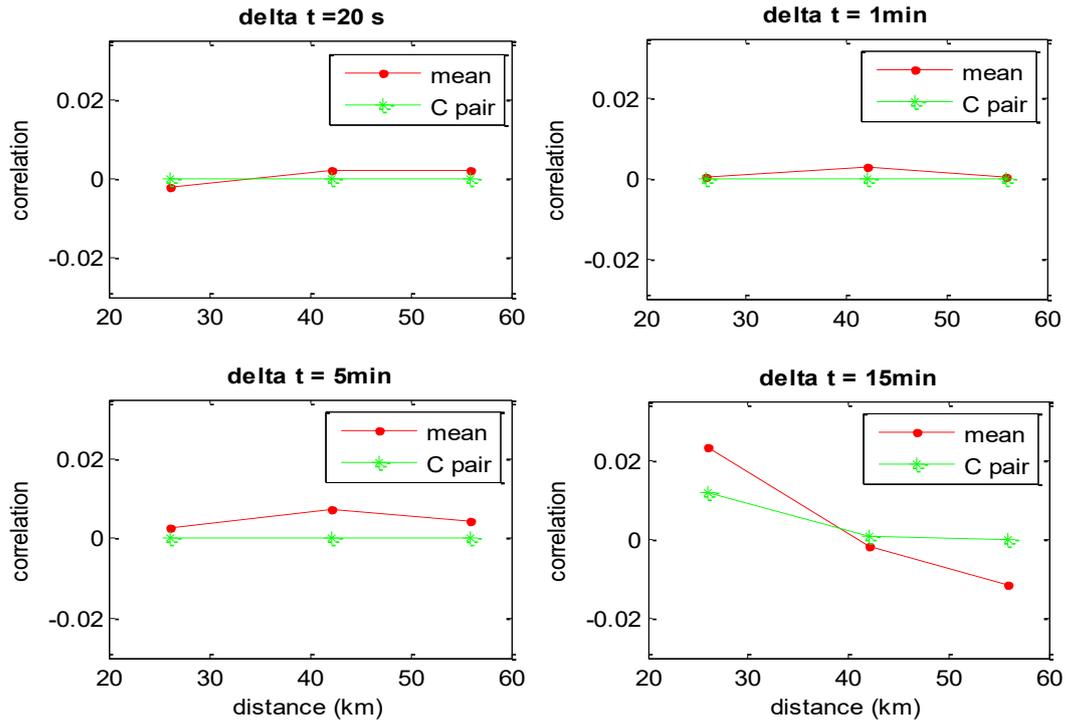


Figure 4-16 Zoom de la Figure 4-15

Nous avons effectué une étude analogue avec les calculs des coefficients d'intercorrélation au lieu des coefficients de Pearson (Figure 4-17). Les valeurs de corrélation deviennent bien plus importantes (0,15 à 0,8). De plus, on observe un pic à une distance intersite de 42 km qui correspond à la paire de stations Désirade/Petit-canal. L'augmentation des valeurs de corrélations lorsque le décalage temporel est pris en compte et l'évolution des corrélations spatiales n'étant pas fonction des distances (pic à 42 km), indique qu'un phénomène physique est prépondérant au phénomène spatial (distances intersites). Il s'agit ici d'un phénomène physique lié à la direction prédominante du vent ayant une forte composante Est (les masses nuageuses se déplacent fréquemment de la station de La Désirade vers le site de Petit-canal). Il y a donc absence de corrélation spatiale instantanée mais présence de corrélation spatiale en tenant compte du décalage temporel.

Les valeurs des coefficients C_{paire} de l'équation (4.13) ne sont plus valables dans ce cas, néanmoins l'évolution des corrélations varient bien en fonction de l'échantillonnage (Figure 4-17).

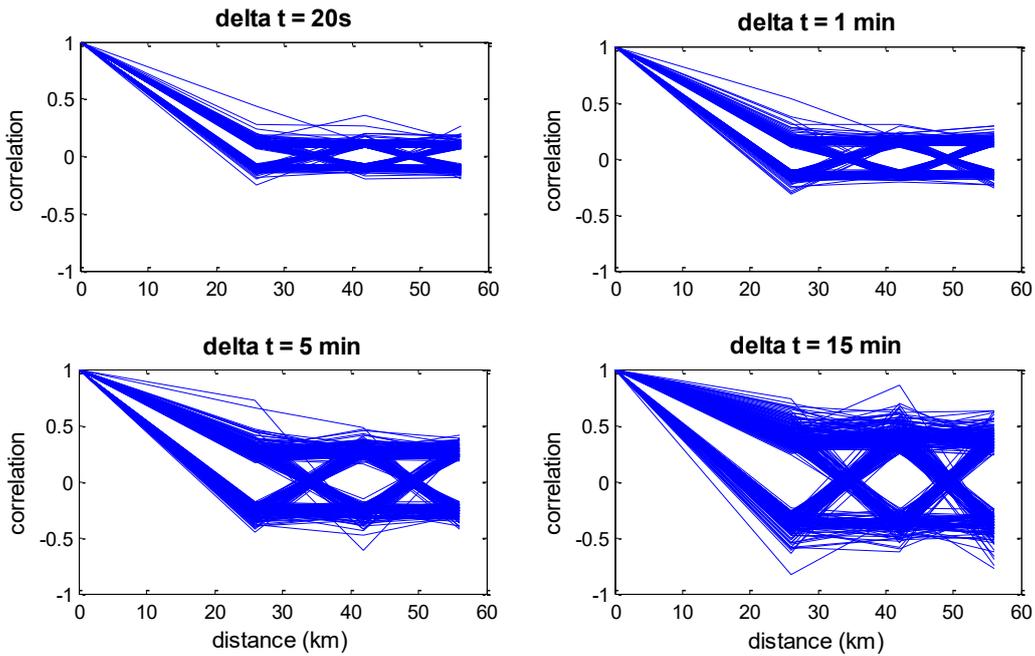


Figure 4-17. Intercorrélation de paires de stations en fonction des distances inter sites et des échelles de temps pour un échantillon d'une année (366 jours).

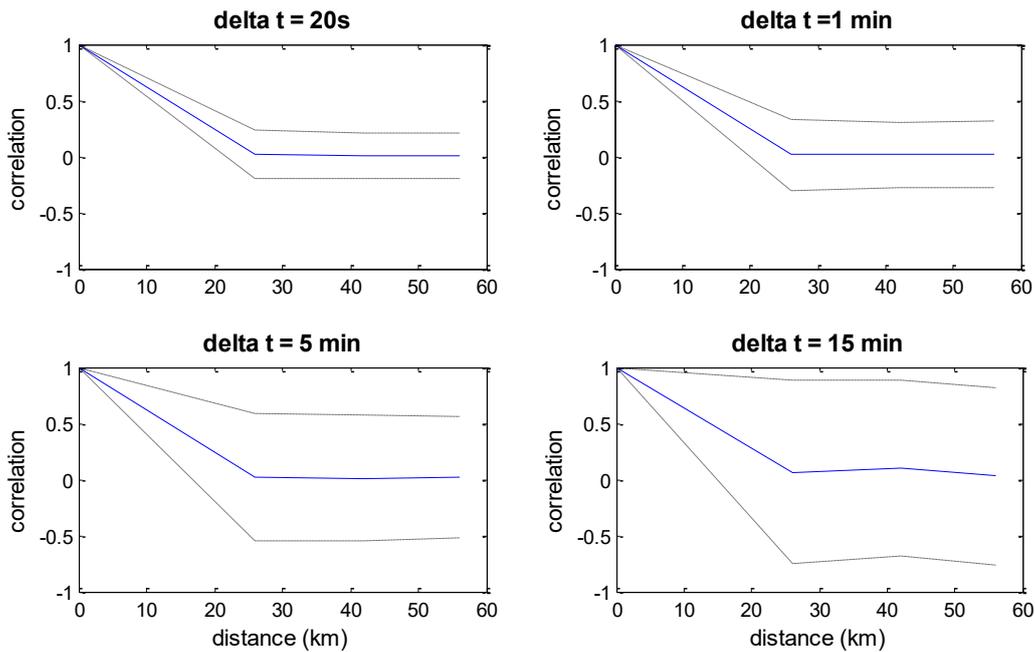


Figure 4-18. Tendence moyenne des intercorrélations de paires de stations en fonction des distances inter sites et des échelles de temps pour un échantillon d'une année (366 jours). Les lignes en pointillé représentent l'intervalle de confiance à 95% .

Cette première analyse, a permis d'évaluer les corrélations spatiales intersites. Pour cela, un parallèle avec les travaux de Pérez décrit dans la littérature, qui utilise le coefficient de Pearson, est présenté. L'objectif a été de définir l'évolution des corrélations de paires de stations en fonction des distances et

de l'intervalle d'échantillonnage des mesures considéré. Cette même étude a été effectuée avec la fonction d'inter corrélation, qui a permis d'évaluer les corrélations, pour une distance et un échantillonnage donnés, en faisant varier un décalage temporel entre les signaux. Cette première étude nous permet d'établir qu'il y a absence de corrélation spatiale instantanée mais présence de corrélation spatiale en tenant compte du décalage temporel.

III.4. Analyse des corrélations inter sites temporelles

III.4.1 Cohérence des signaux de fluctuations

Pour examiner la relation entre les densités de puissance spectrale des signaux de fluctuations entre les sites, nous utilisons la fonction de cohérence. La cohérence estimée est une fonction de fréquence avec des valeurs comprises entre 0 et 1. Elles indiquent dans quelle mesure le signal de fluctuations du site 1, X , est corrélé au signal de fluctuation du site 2, Y , à chaque fréquence donnée, où $P_{XX}(f)$ et $P_{YY}(f)$ sont respectivement la densité de puissance spectrale de X et Y et la densité de puissance spectrale croisée $P_{XY}(f)$.

Une valeur $C_{XY}(f)$ proche de 1 signifie que X et Y sont corrélés pour une fréquence donnée. Nous avons calculé la fonction de cohérence pour chaque paire de sites: Petit-Canal/Désirade, Petit-Canal/Fouillole, Fouillole/Désirade et les résultats sont tracés sur la Figure 4-519. Les signaux de fluctuations utilisés sont des séquences de 7h à 17h afin d'effectuer une analyse sur la période d'ensoleillement.

Les valeurs de cohérence ne sont pas proches de 1 et sont assez faibles contrairement dans le cas où on utiliserait les signaux de flux solaire mesurés plutôt que les signaux de fluctuations, puisque le cycle journalier a été retiré. Les sites sont décorrélés à des fréquences inférieures à 10^{-4} Hz soit des échelles de temps plus courtes que 2h45. La valeur de cohérence la plus élevée est observée pour la paire de sites Petit-canal et Désirade. Entre Désirade et Fouillole les valeurs de cohérence sont très faibles même pour de longues échelles de temps avec une valeur de cohérence maximale de 0,12.

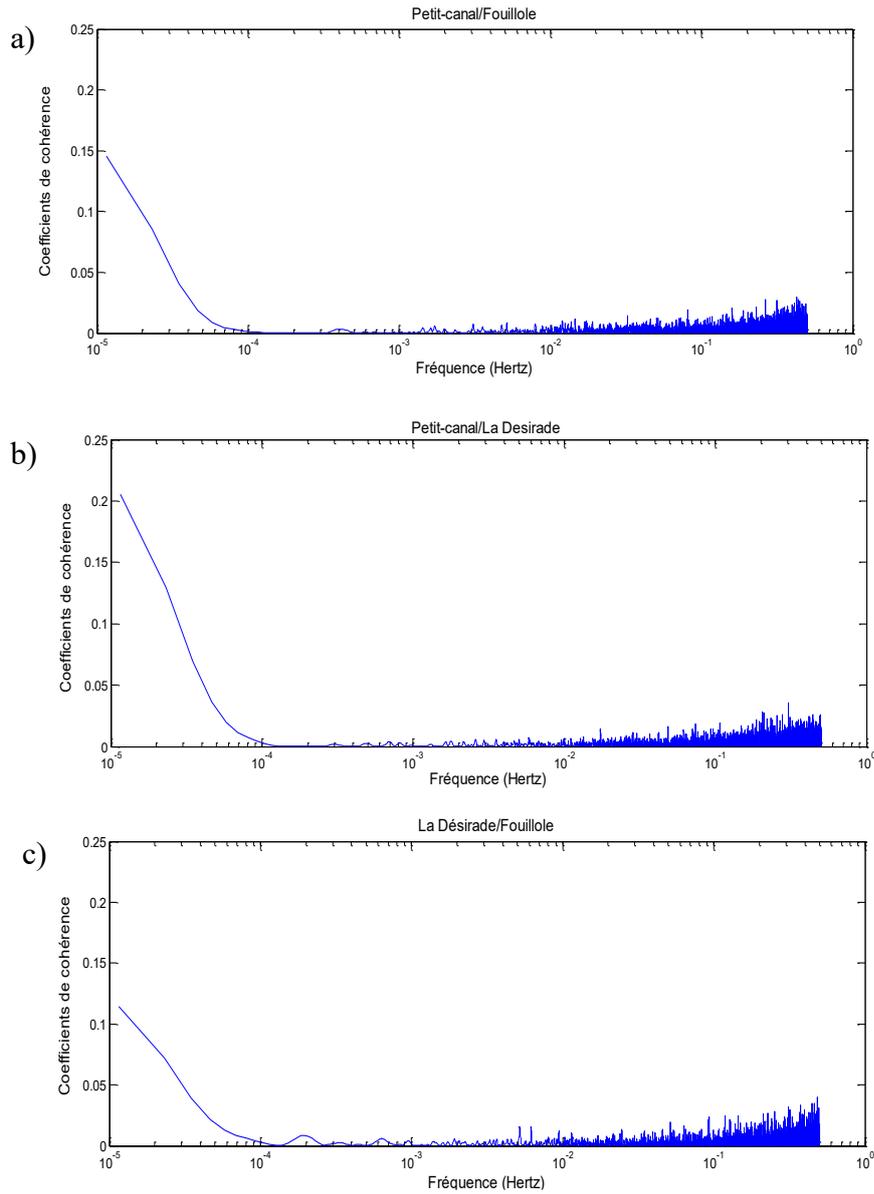


Figure 4-19 Fonction de cohérence entre les signaux de fluctuations pour chaque paire de sites comparés: Petit-canal/Fouillole (a), Petit-canal/La Desirade (b), La Desirade/ Fouillole (c).

Cette analyse permet de constater que la variabilité du flux solaire entre les sites n'est pas corrélée à des échelles de temps inférieures à 2h45. Lave and Keissl (2010) ont trouvé que deux sites (NWTC and SRRI) avec une distance minimale de 19 km deviennent décorrélés pour des échelles de temps inférieures à 3h. Dans Lave et Kleissl (2012), un calcul des cohérences entre EBU2 et cinq autres sites, est également effectué et a montré l'influence de paramètres météorologiques sur les résultats de cohérences. En effet, leurs résultats ont permis de déduire différentes conditions météorologiques à l'ouest de EBU2 comme à l'est. Leurs observations du

ciel anecdotiques ont confirmé que les nuages s'évaporent souvent lorsqu'ils se déplaçaient vers l'est, qui se traduirait par une cohérence plus petite.

Nous avons cherché à approfondir notre étude par une mesure des corrélations sur des échelles de temps plus courtes des signaux de fluctuations.

III.4.2 Séparation des échelles de temps

III.4.2.1 Méthode

Nous nous sommes intéressés à la séparation d'échelle de temps, premièrement indépendamment des sites comparés. Le but est de déterminer à quelle échelle de temps, les corrélations deviennent significatives. Pour cette analyse nous avons utilisé la méthodologie suivante. Premièrement, nous avons appliqué, pour chaque paire de sites et pour chaque jour, la fonction d'intercorrélation sur nos années de mesures 2010, 2011 et 2012 soit 1029 journées. Cette fonction a été appliquée sur les signaux de fluctuations obtenues à partir d'une différence de deux moyennes glissantes calculés sur N points différents (N_1 et N_2) qui définiront l'intervalle d'échelle de temps $[N_1, N_2]$. Rappelons l'équation 4.3 de la moyenne glissante :

$$\bar{G}_N(t) = \frac{1}{N} \sum_{i=t-\frac{N-1}{2}}^{t+\frac{N-1}{2}} G(i) \quad (4.3)$$

Avec N fixé.

La procédure suivante décrira comment nous calculons les signaux de fluctuations avec un intervalle d'échelle de temps $[N_1, N_2]$ nommé $G'_{[N_1;N_2]}$:

$$G = G'_{N_1} + \bar{G}_{N_1} \quad (4.14)$$

$$G = G'_{N_2} + \bar{G}_{N_2} \quad (4.15)$$

Avec $N_1 < N_2$

On calcule la différence entre la moyenne glissante à N_2 points et celles à N_1 points pour obtenir les fluctuations à l'échelle de temps $[N_1; N_2]$. Soit l'équation :

$$\bar{G}_{N_2} - \bar{G}_{N_1} = G - G + G'_{N_1} - G'_{N_2} \quad (4.16)$$

ainsi on obtient:

$$\bar{G}_{N_2} - \bar{G}_{N_1} = 0 + G'_{N_1} - G'_{N_2} \quad (4.17)$$

$$G'_{N_1} - G'_{N_2} = G'_{[N_1;N_2]} \quad (4.18)$$

Avec $G'_{[N_1;N_2]}$ les signaux de fluctuations à l'échelle temps $[N_1; N_2]$.

Nous avons calculé la densité de probabilité des coefficients maximaux d'intercorrélations journaliers des signaux de fluctuations $G'_{[N_1;N_2]}$ entre les paires de stations. L'analyse permet d'obtenir un histogramme des fréquences relatives des différentes valeurs de coefficients maximaux. Les échelles de temps choisies sont définies par les intervalles suivants: [1 min ;15 min], [15 min ; 30 min], [30 min ;1 h], [2 h ;3 h]. Les densités de probabilité des coefficients d'intercorrélations sont représentées par la Figure 4-21. De telles échelles de temps sont choisies car elle marque un changement significatif des densités. La Figure 4-20, donne un exemple de signaux de fluctuations de la même journée obtenues pour chaque intervalle de temps.

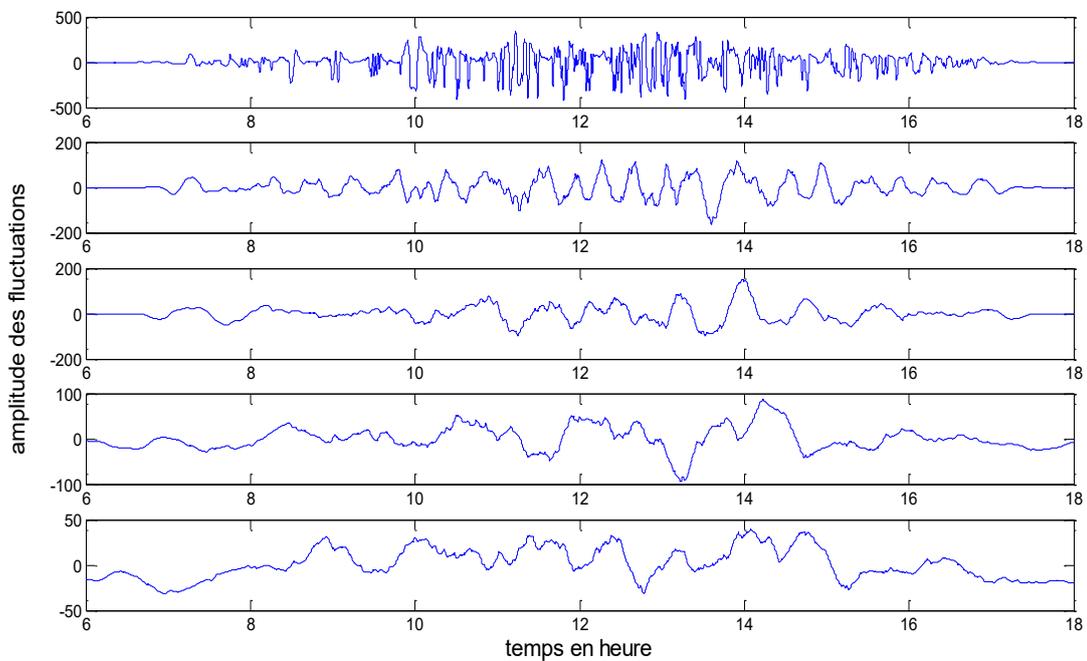


Figure 4-20 Signaux de fluctuations du 10 octobre 2012 pour les intervalles du haut vers le bas [1 min ;15 min], [15 min ; 30 min], [30 min ;1 h], [2 h ;3 h] .

III.4.2.2 Intercorrélation intersite avec séparation des échelles de temps

Nous présentons les résultats obtenus pour l'ensemble des paires de stations (Petit-canal/ La Désirade, Petit-canal/Fouillole, Fouillole/ La Désirade), de densité de probabilité en fonction des coefficients maximaux d'intercorrélations pour les trois années de mesures. L'objectif est d'évaluer les

corrélations temporelles par séparation d'échelle de manière globale (sans distinction des paires de stations).

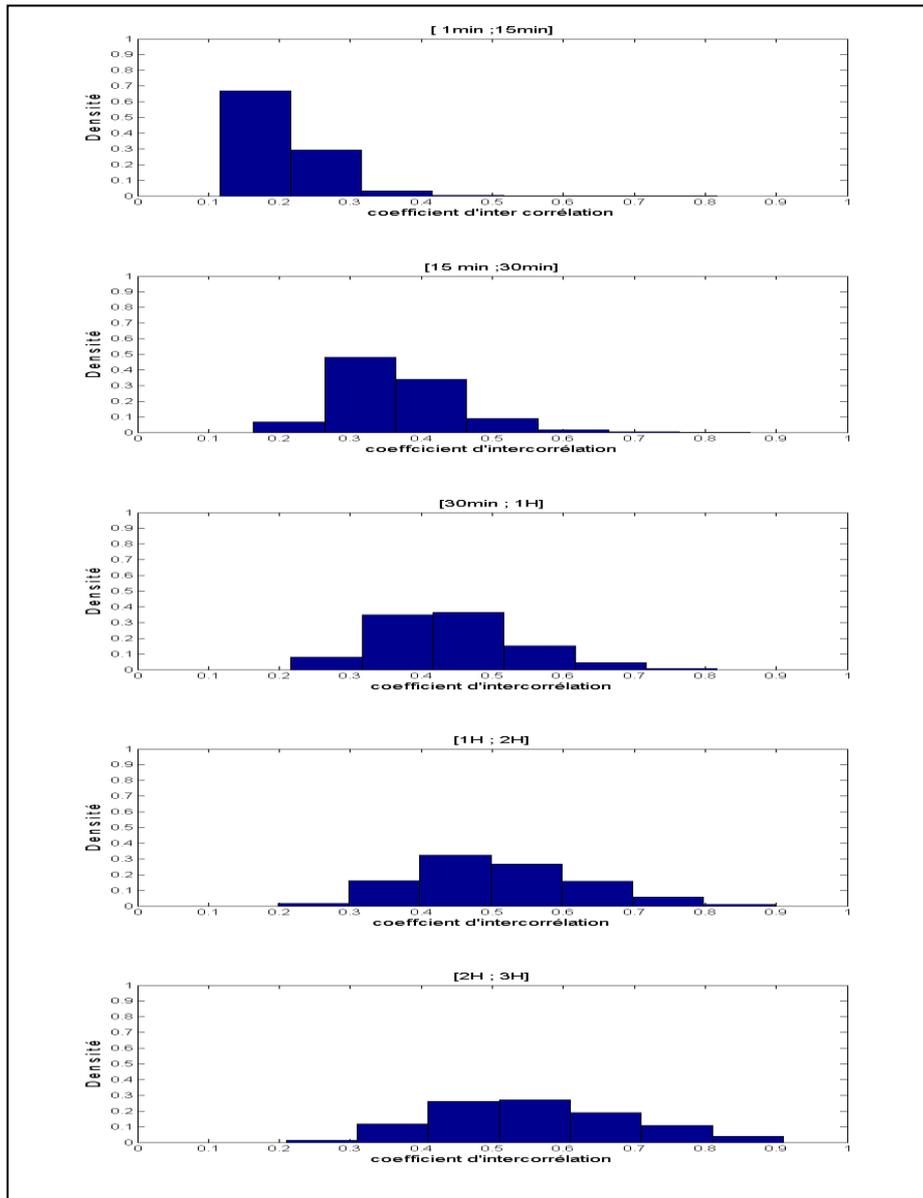


Figure 4-21 Densité de probabilité des valeurs absolues des coefficients maximaux d'intercorrélation des signaux de fluctuations journaliers dans l'intervalle [1 min ;15 min],[15min ;30min], [30min ;1h] calculés de 2010 à 2012 pour l'ensemble des paires de sites comparés.

Nous pouvons remarquer que plus l'échelle de temps augmente, plus la valeur maximale des coefficients maximaux d'intercorrélation augmente. Les corrélations significatives entre les signaux de fluctuations comparés entre deux stations ont une probabilité faible pour des échelles de temps inférieures à 1h et n'atteignent pas des coefficients de valeur supérieure à 0,8. Pour une échelle de temps entre [1min ; 15min], les corrélations significatives (supérieures à 0,6) entre les signaux de fluctuations ont une densité proche de 0, entre [15 min ; 30 min] une densité égale à 0,03, entre [30 min ; 1 h] une densité de 0,07. Les corrélations significatives qui existent à ces échelles de temps

correspondent principalement à des conditions climatiques uniformes (ciel clair, ou ciel totalement nuageux) dans la zone limitée par nos trois sites de mesure. Entre [1h ; 2 h] une densité environ deux fois plus grande de valeur 0,18 avec des coefficients pouvant atteindre 0,9. Pour les échelles de temps supérieures à 2 h, les signaux de fluctuations sont alors très lissés et les valeurs de coefficients d'intercorrélation $\geq 0,6$ ont une densité significative approximant 0,3.

L'échelle de temps [1h ; 2 h] correspondant à une densité significative et environ deux fois plus grande par rapport aux densités pour des échelles de temps inférieures, marque par conséquent l'échelle de temps seuil pour obtenir une probabilité non faible des corrélations significatives intersite de la variabilité du flux solaire. Ce résultat signifie que les phénomènes liés aux fluctuations qui surgissent et durent sur de longues échelles de temps ont plus de chance de se déplacer d'un site à l'autre. L'échelle de temps au cours de laquelle les sites deviennent corrélés est également une indication de l'échelle de temps seuil à laquelle la variabilité globale est atténuée.

En considérant cette séparation d'échelle par paires de sites, on peut analyser l'évolution des corrélations temporelles de variabilité de flux solaire dans l'espace. Nous évaluons donc le maximum d'intercorrélation pour différentes échelles de temps avec distinction des paires de sites comparés : Petit-Canal/Fouillole (Figure 4-22) ; Petit-Canal/La Désirade (Figure 4-23) ; et La Désirade/Fouillole. (Figure 4-24). La densité de probabilité est effectuée afin d'analyser l'occurrence des valeurs à chaque échelle de temps.

Sites comparés : Petit-canal et Fouillole

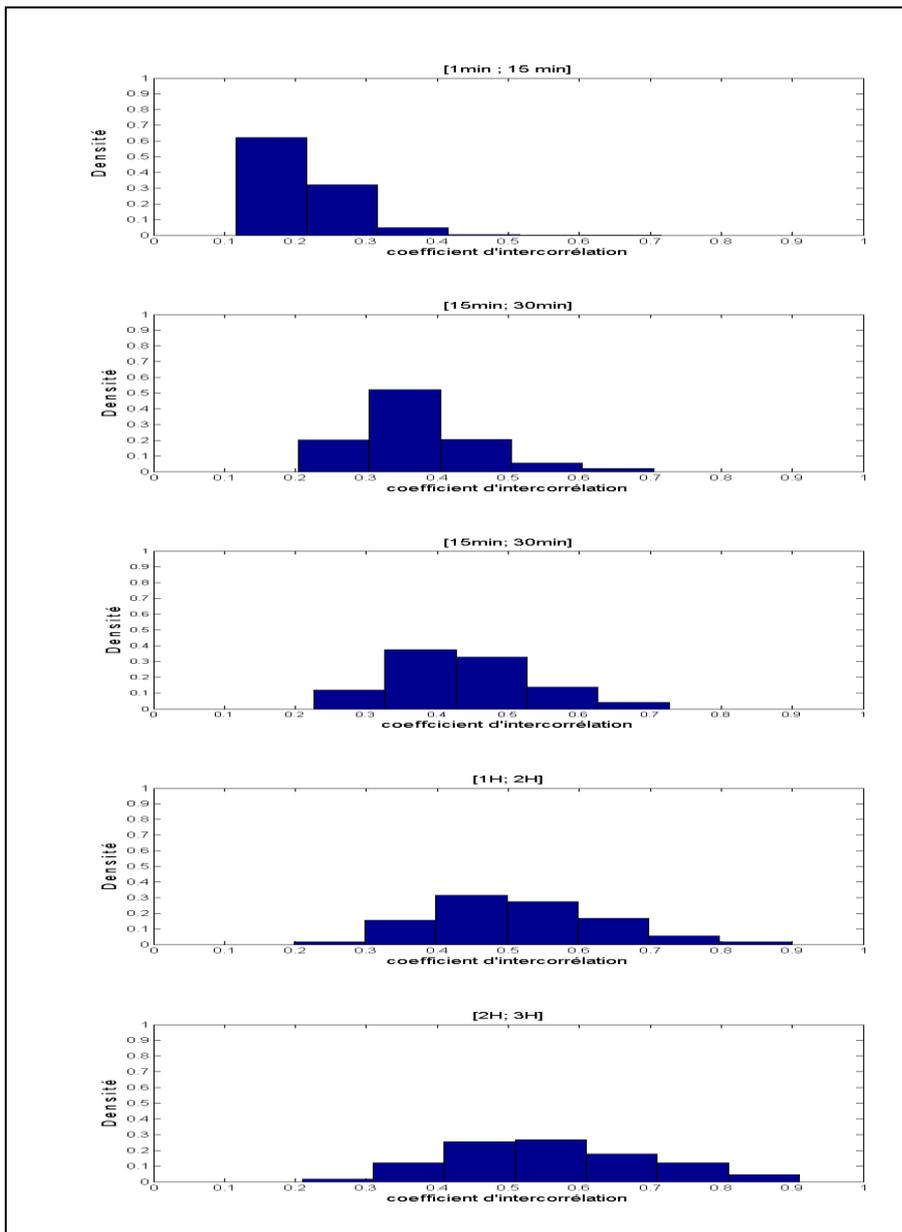


Figure 4-22 Densité de probabilité des coefficients maximaux des signaux de fluctuations journaliers dans l'intervalle [1min ;15min],[15min ;30min],[30min ;1h],[1h ;2h],[2h ;3h], calculés de 2010 à 2012 pour la paire de site Petit-canal/Fouillole.

Sites comparés : Petit-canal et La Désirade

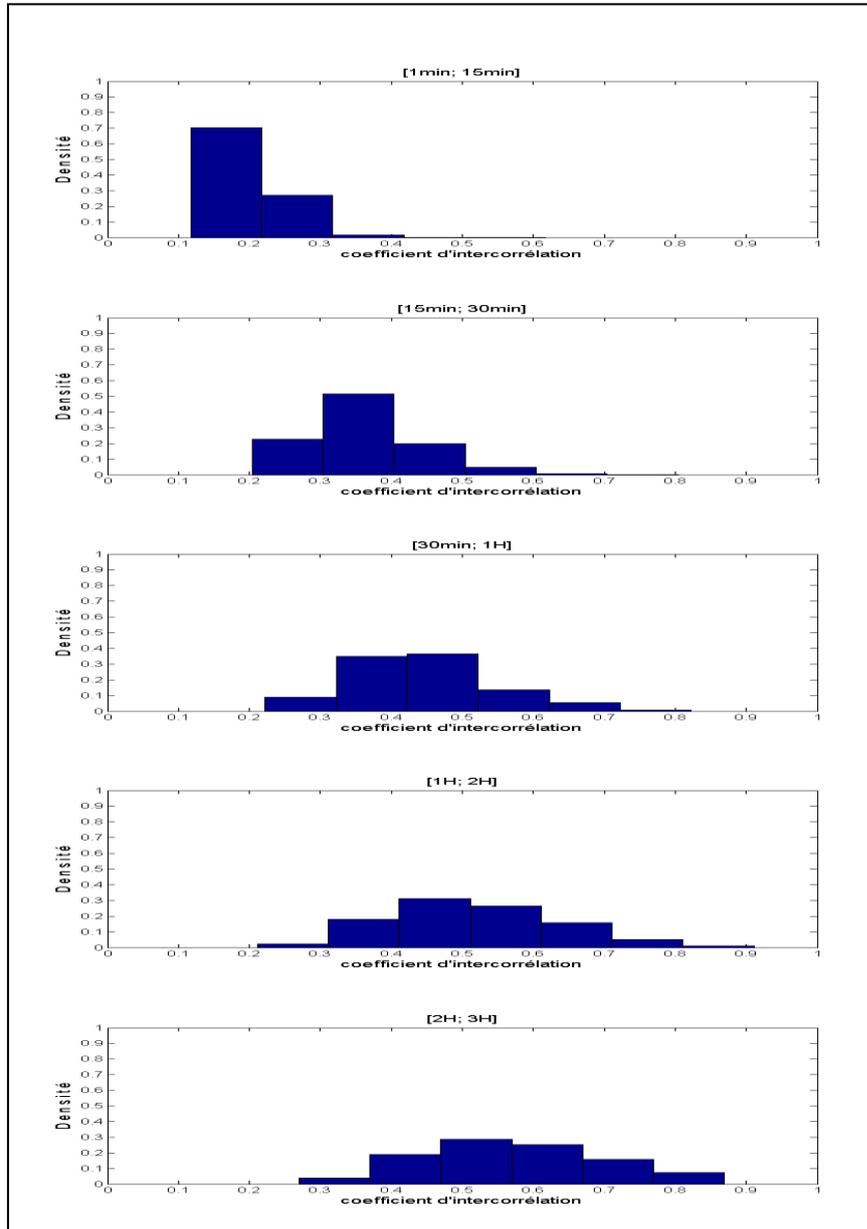


Figure 4-23 Densité de probabilité des coefficients maximaux des signaux de fluctuations journaliers dans l'intervalle [1min ;15min],[15min ;30min],[30min ;1h],[1h ;2h],[2h ;3h], calculés de 2010 à 2012 pour la paire de site Petit-canal/La Désirade.

Sites comparés : La Désirade et Fouillole

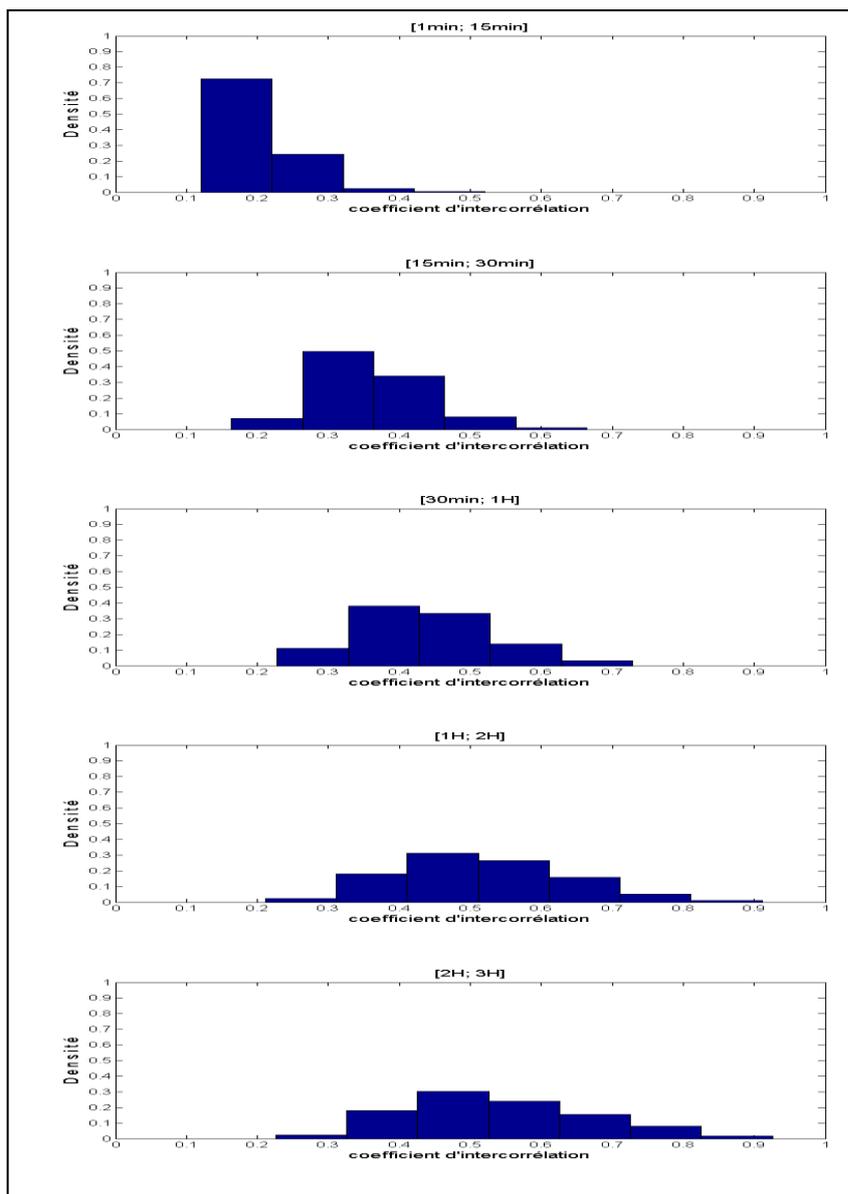


Figure 4-24 Densité de probabilité des coefficients maximaux des signaux de fluctuations journaliers dans l'intervalle [1min ;15min],[15min ;30min],[30min ;1h],[1h ;2h],[2h ;3h], calculés de 2010 à 2012 pour la paire de site Petit-canal/La Désirade.

On peut remarquer qu'à l'échelle de temps comprise entre 1min et 15min les coefficients $\geq 0,6$ ont une occurrence nulle pour l'intercorrélation des signaux de fluctuations entre La Désirade et Fouillole, tandis que les paires de sites comparés correspondant à La Désirade/ Petit-canal et Petit-canal/ Fouillole présentent bien une corrélation des signaux de fluctuations $\geq 0,6$ mais avec une densité très faible à peine visible sur l'histogramme. Pour l'échelle de temps entre 30min et 1h, la probabilité non nulle est valable seulement pour les coefficients ne dépassant pas une valeur de 0,7 concernant la paire de sites La Désirade/ Fouillole contrairement aux autres paires

de stations. Pour les coefficients $\geq 0,6$, la densité montre des valeurs environ égale à 0,04 pour la paire de sites Petit-canal/ Fouillole et de 0,15 pour la paire de sites La Désirade/ Petit-canal.

L'échelle de temps comprise entre 1h et 2h, montre une augmentation brutale de l'occurrence des valeurs de coefficients d'intercorrélation $\geq 0,7$ pour les trois paires de sites comparés. Néanmoins, seule la paire de site La Désirade/ Petit-canal présente une probabilité d'obtenir des coefficients d'intercorrélation de variabilité de flux solaire supérieurs à 0,9. Les résultats permettent de confirmer que l'échelle de temps seuil pour laquelle les signaux de variabilité deviennent corrélés est [1h ;2h], soit les signaux calculé par la suppression de fluctuations transitoires inférieures à l'heure soulignant les tendances de variabilité à plus long terme que 1h.

Ainsi, selon les résultats obtenus, les corrélations de variabilité de flux solaire significatives ont une plus forte occurrence entre Petit-Canal/La Désirade, et Petit-Canal/Fouillole. Ce résultat peut être expliqué par la localisation géographique de nos stations de mesures. Petit-Canal et Fouillole sont les sites les plus proches du point de vue des distances métriques, leur proximité expliquerait la probabilité d'obtenir une forte corrélation de la variabilité de flux solaire. Concernant Petit-Canal/La Désirade, les vents prédominants se déplaçant selon une direction à forte composante Est (Les Alizés) aux Antilles provoqueraient le déplacement de masse nuageuse du site de La Désirade au site de Petit-canal. Nous chercherons à analyser plus finement dans la section IV, l'influence des paramètres météorologiques sur les corrélations de la dynamique du flux solaire entre nos trois sites. La corrélation ne diminue donc pas de manière isotropique (la paire de site La Désirade/ Petit-canal présente une plus grande occurrence de valeurs fortes de coefficients d'intercorrélation que Petit-canal/Fouillole bien que la Désirade soit plus éloignée géographiquement de Petit-canal que Fouillole). Il y a donc une plus forte probabilité de journées où les corrélations intersites dépendent des conditions météorologiques plutôt que des distances. Pérez et al (2012) montrent que les corrélations décroissent en fonction de la distance entre les sites pour des distances métriques entre les sites maximales de 100km et minimale de 100 m. La localisation climatique et géographique différentes pour nos sites de mesures localisés en milieu îlien avec le site de La Désirade séparé par les autres sites par un bras de mer est à prendre en compte pour la différence des résultats.

III.4.3 Covariance

La covariance pouvant être assimilée à la fonction d'intercorrélation centrée, l'application de cette fonction sur nos signaux de fluctuations permet d'analyser les corrélations de la variabilité de flux solaire qui existent entre les sites sans faire intervenir de décalage temporel. Pour cette analyse, nous n'utiliserons que les signaux de fluctuations pour l'échelle de temps [1h ;2h], puisque cette échelle est le temps seuil pour que les corrélations soient significatives. Nous avons représenté les résultats de densité de probabilité des coefficients de covariance pour chaque paire de sites sur la Figure 4-25.

L'occurrence très faible cependant non nulle, des coefficients significatifs de covariance signifient qu'il y a une très faible probabilité que le même type de variabilité de flux solaire, c'est-à-dire que le même évènement météorologique, se produise au même instant sur nos trois stations de mesures. Nous notons que les fortes valeurs de coefficients de covariance, ayant une probabilité non nulle, peuvent atteindre la valeur de 0,8 pour la paire de sites Petit-canal / Fouillole tandis que pour la paire de sites Petit-canal/La Désirade, la plus grande valeur de coefficients de covariance est de 0,7 et pour La Désirade/Fouillole les valeurs n'atteignent pas 0,7. La situation est donc inversée entre les paires de sites Petit-canal/Fouillole et La Désirade/ Petit-canal par rapport aux résultats avec l'application de la fonction d'intercorrélation. Ainsi, c'est entre le site de Petit-canal et Fouillole que les corrélations de variabilité de flux solaire sont plus significatives. L'analyse des coefficients de covariance donne de toute évidence la conclusion que la corrélation diminue de manière isotropique (plus les distances métriques sont grandes plus les corrélations sont faibles). Ce résultat est en accord avec bien d'autres recherches dans la littérature tels que dans Gueymard et Wilcox (2011), Lave et Kleissl (2010), Hoff et Perez (2011), Pérez et al (2011), Pérez et al (2012), Badosa et al (2013) .

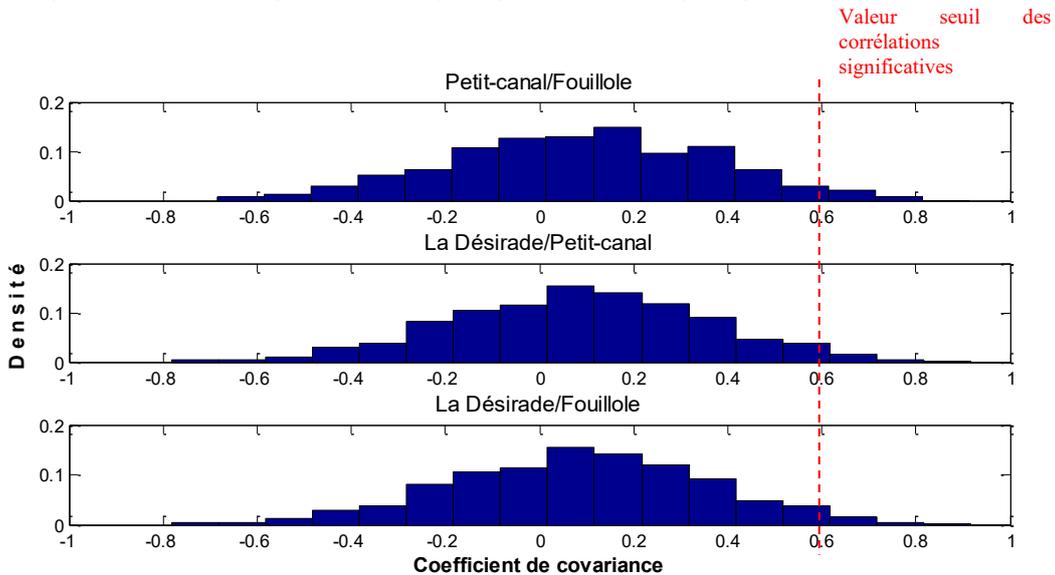


Figure 4-25. Densité de probabilité des coefficients de covariance des signaux de fluctuations journaliers dans l'intervalle [1h;2h] calculés de 2010 à 2012 pour les paires de site comparés : Petit-canal/ Fouillole, Petit-canal/La Désirade, La Désirade/Fouillole.

L'analyse des corrélations temporelles avec séparation d'échelles de temps a permis de définir un intervalle de temps seuil pour que les corrélations entre les sites soient significatives. Cette échelle de temps est [1h ; 2h]. L'étude de la fonction de covariance pour les signaux de fluctuations obtenus à cette échelle de temps montre que les corrélations sont prédominantes pour la paire de sites Petit-canal /Fouillole tandis que les résultats d'intercorrélation montrent qu'il s'agit de la paire de sites La Désirade/Petit-canal. La prise en compte du décalage temporelle montre une dynamique d'évolution des corrélations spatio-temporelles différentes. Notamment, l'étude des coefficients maximum de la

fonction d'intercorrélation montre que le processus spatio-temporel aléatoire de l'ensoleillement n'est pas isotropique dans notre cas d'étude, dans le sens où la corrélation n'est pas une fonction des séparations des distances.

Les résultats de corrélations prédominantes pour la paire de stations La Désirade /Petit-canal sont en accord avec le fait que les conditions météorologiques (masses nuageuse) qui tendent à se déplacer en direction prédominante Est-Ouest. L'impact de la direction du vent prédominant est donc un paramètre à considérer pour l'élaboration du modèle spatio-temporel.

IV. Analyse dynamique des corrélations

IV.1. Décalage temporel des corrélations

Le décalage temporel dépendant tant sur la distance considérée que la vitesse de nuage observée, nous avons cherché à quantifier ce paramètre évaluée entre deux stations en utilisant le maximum d'intercorrélation de deux signaux de fluctuations. Cette étude donnera une évaluation de la durée pour laquelle un évènement météorologique (généralement une masse nuageuse) se déplacerait d'un site à l'autre. La Figure 4-26 montre un exemple d'intercorrélation entre la station de Petit-canal et La Désirade(41818 m de distance) pour l'exemple d'une journée. Un maximum d'intercorrélation de 0,75 est observé à un temps de retard de $\tau = 41$ min. Ceci indique que la variabilité causée par un phénomène (masse nuageuse) se produisant sur le site de La Désirade devrait se déplacer sur le site de Petit-canal, 41 minutes plus tard pour ce même jour.

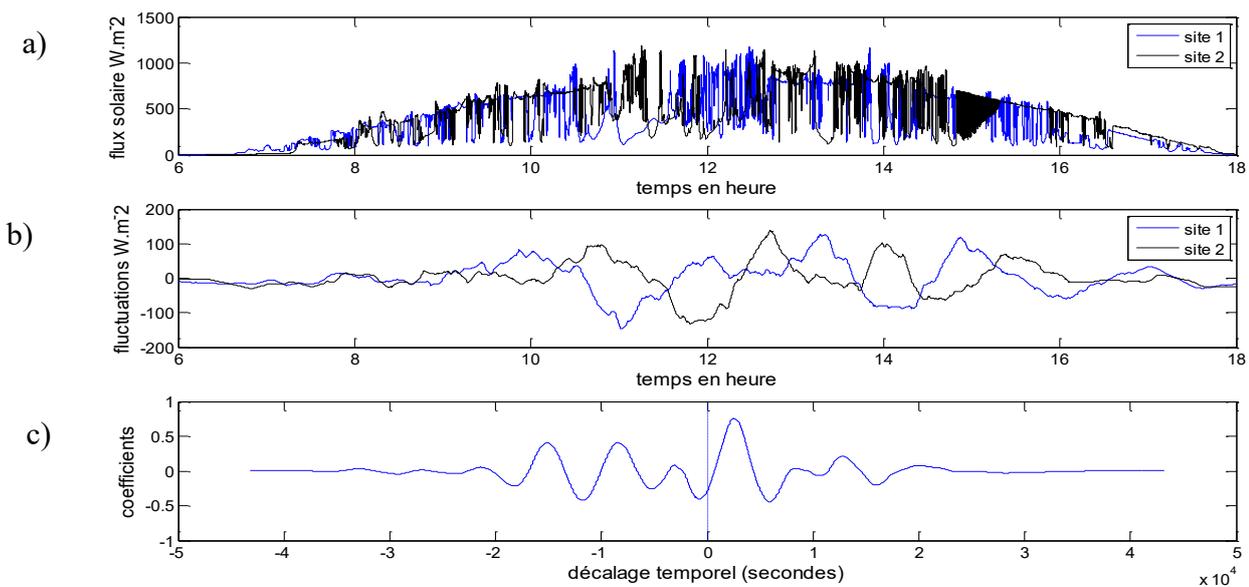


Figure 4-26. a) Journée de flux solaire mesuré sur deux stations avec b) leurs signaux de fluctuations correspondants obtenues par méthode de moyenne glissante. La fonction d'intercorrélation calculée pour les signaux de fluctuations est présentée en c). *Décalage temporel $\tau = 41$ min et coefficient maximal = 0,75 (1^{er} février 2012 entre Petit-canal et La Désirade)*

Comme mentionné en section III, les coefficients d'intercorrélation dont les valeurs sont comprises entre 0,6 et 1 sont des coefficients correspondants à des similarités significatives entre les signaux. Nous avons donc évalué la densité de probabilité de toutes les valeurs de décalage temporel τ calculé pour chaque paire de sites pour lequel $R_{XY}(\tau) \geq 0.6$. Les résultats sont montrés sur les figures 4-26. On constate que les temps de retard les plus probables sont entre [0 ; 30min] plus précisément avec une moyenne de 10min pour les sites de Petit-canal/Fouillole. Les décalages temporels pour la paire de site Petit-canal / Fouillole sont essentiellement dues à la proximité de ces sites (près de 26 km de distance entre ces deux stations).

Entre Petit-canal et La Désirade, on remarque une probabilité prédominante à 30% d'un décalage temporel compris entre [1h ; 1h 30min]. Pour les sites de La Désirade/Fouillole on trouve un décalage temporel plus probable entre [30min ; 1h]. Bien que La Désirade est plus distante de Fouillole que de Petit-canal, on trouve le décalage temporel ayant la plus grande probabilité, inférieur à celui de Petit-canal/ La Désirade. Puisqu'il est impossible qu'un quelconque phénomène prenne moins de temps pour se déplacer du site de La Désirade à Fouillole, comparé à un phénomène se déplaçant de La Désirade à Petit-canal, on peut en déduire que les corrélations entre La Désirade et Fouillole peuvent correspondre à un évènement météorologique qui se produit de manière opposée (absence/présence de nuages) avec un décalage temporel. Le passage nuageux se déplaçant de la Désirade vers Fouillole peut correspondre à un décalage temporel compris entre [1h ; 2h] qui présente une probabilité totale de 22% (somme des occurrences correspondant aux intervalles [1h ; 1h30], [1h30 ; 2h]) (Figure 4-27 c)). La plus forte occurrence se produisant pour des phénomènes météorologiques opposés entre La Désirade et Fouillole plutôt qu'un déplacement nuageux entre ces stations, est due au fait que les directions prédominantes du vent sont Nord-Est, Sud-Est.

Sur l'ensemble des corrélations de la variabilité de flux solaire sans distinction de paire de site (Figure 4-28), la dynamique du flux solaire présente des corrélations entre les sites avec une plus forte occurrence de décalage temporel compris entre [0 ; 30 min] avec une probabilité d'environ 24%. Néanmoins les intervalles compris entre [30min ; 1h30] présente une occurrence assez proche avec un pourcentage d'environ 20% (en sommant les occurrences des intervalles [30min ; 1h], [1h ; 1h30]). Considérant que les décalages temporels correspondent à la durée de déplacement d'une masse nuageuse d'un site à l'autre, on trouve alors une vitesse de vent probable comprise entre 24 km/h et 43 km/h (rapport entre l'intervalle de décalage temporel prédominant et l'intervalle de distances métriques entre les stations). Ce résultat conforte la validité de notre étude statistique en ce sens où l'intervalle des valeurs de vitesse de vent trouvées correspond approximativement aux données de vitesse de vent de METEO France Guadeloupe hors des périodes de météo à bas vent (www.meteofrance.gp).

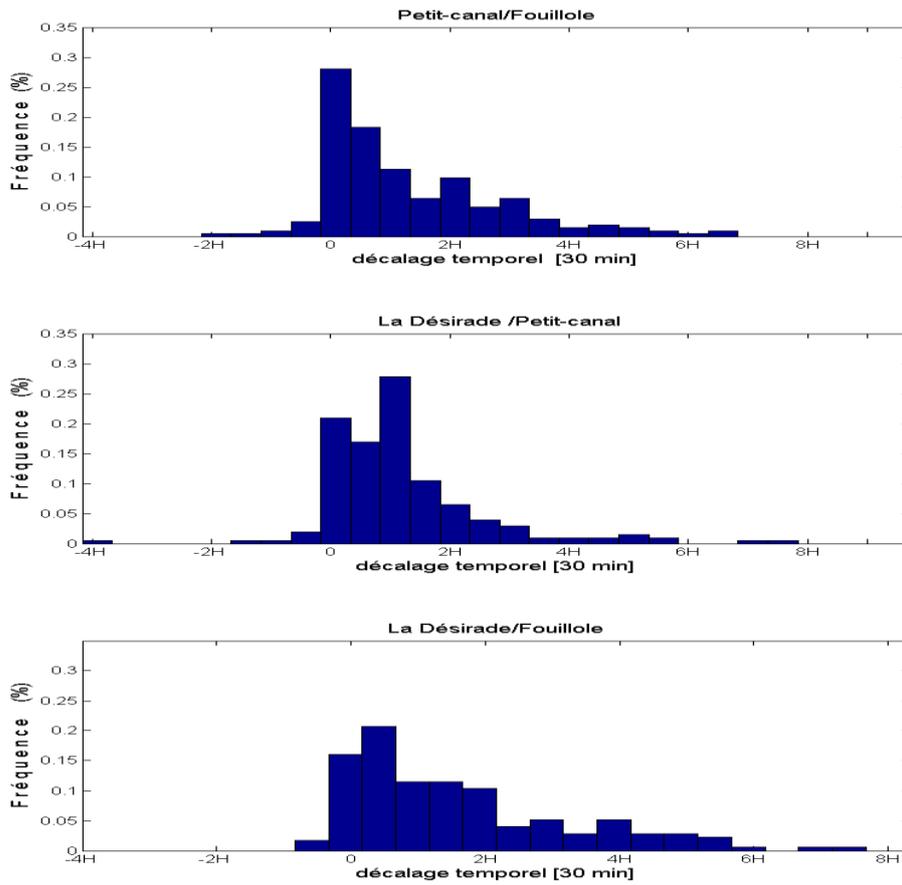


Figure 4-27. Densité de probabilité des décalages temporels journaliers calculés pour $R_{XY}(\tau) \geq 0.6$ pour la comparaison intersite Petit-canal/Fouillole, La Désirade/Petit-canal, La Désirade/Fouillole

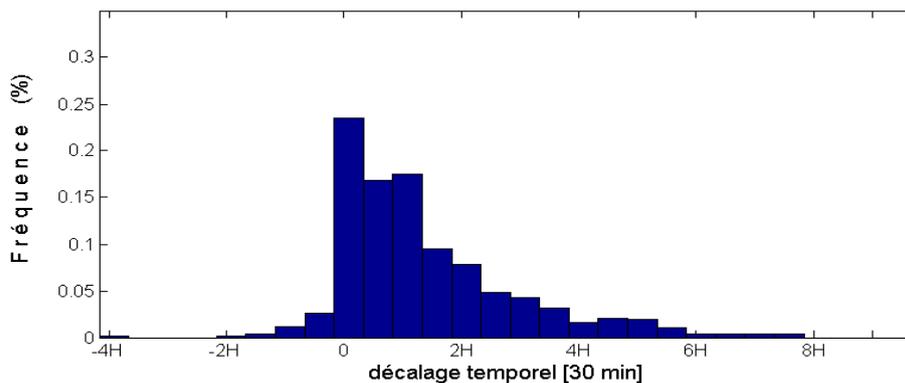


Figure 4-28. Densité de probabilité des décalages temporels journaliers calculés pour $R_{XY}(\tau) \geq 0.6$ pour l'ensemble des comparaisons intersites.

IV.2. Corrélations intersites et paramètres physiques

Nous nous sommes intéressés aux paramètres physiques qui sont sources des corrélations de la variabilité du flux solaire observées statistiquement. Seules les journées présentant des corrélations

significatives (coefficients d'intercorrélation supérieures à 0,6) sont considérées. Les corrélations spatio-temporelles sont regroupées par mois et par paires de sites afin d'étudier de manière générale l'évolution des corrélations de la variabilité spatio-temporelle au cours de l'année. Les histogrammes, Figure 4-29, à Figure 4-31 présentent le nombre de jours, associés aux corrélations significatives, en fonction du mois et de la paire de sites considérée pour respectivement l'année 2010, 2011, 2012 et l'ensemble des trois années Figure 4-32. La case vide pour l'historgramme Figure 4-59 est due à un trou de mesure pour le mois de juillet 2010.

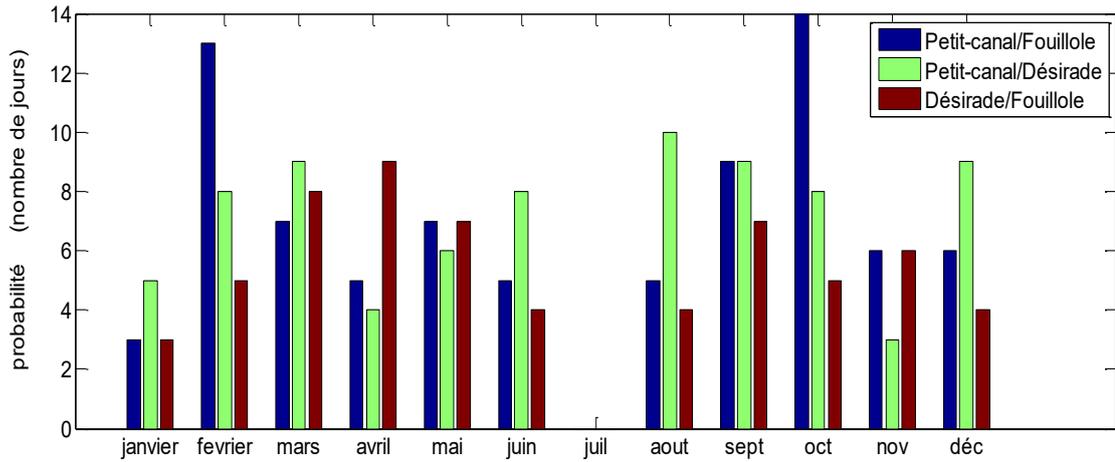


Figure 4-29. Fréquence d'apparition en nombres de jours des corrélations $R_{XY}(\tau) \geq 0.6$ calculées pour chaque couple de sites par mois pour l'ensemble de l'année 2010.

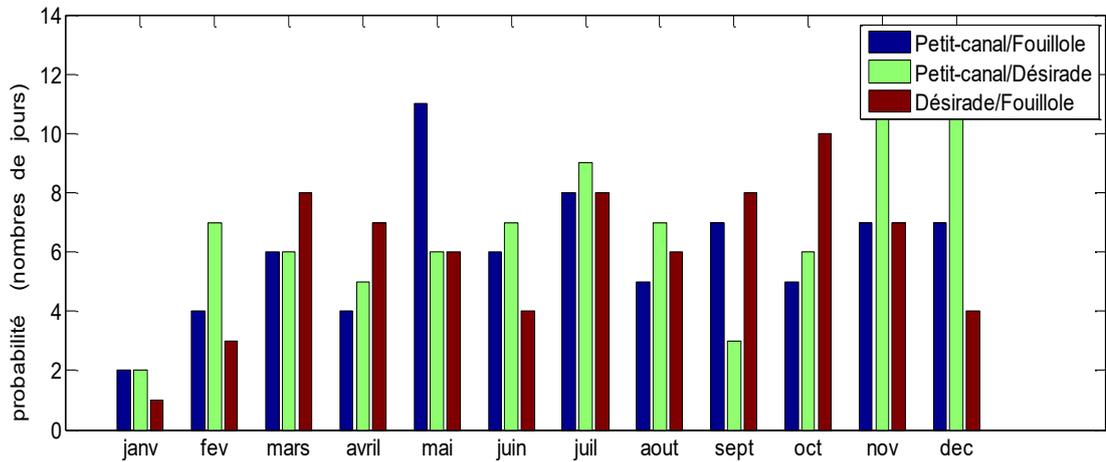


Figure 4-30. Fréquence d'apparition en nombres de jours des corrélations $R_{XY}(\tau) \geq 0.6$ calculées pour chaque couple de sites par mois pour l'ensemble de l'année 2011.

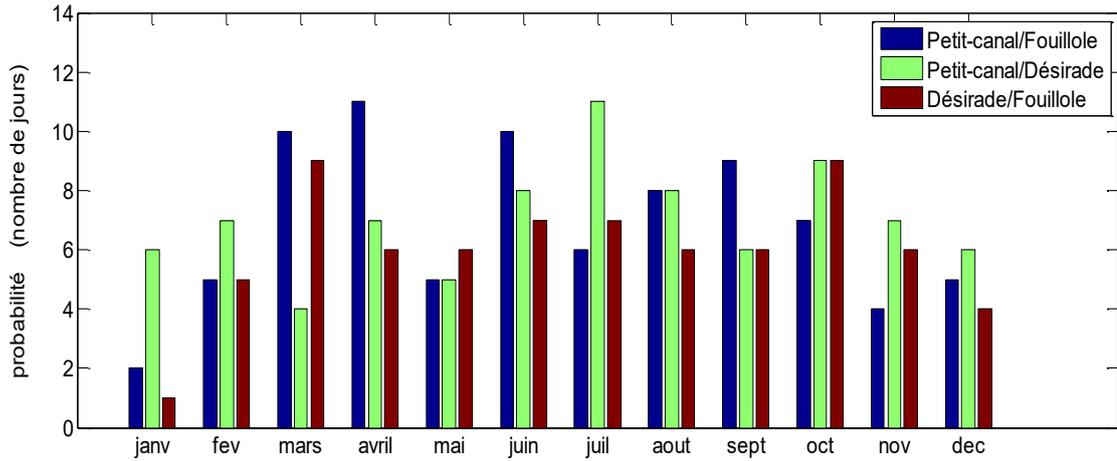


Figure 4-31. Fréquence d'apparition en nombres de jours des corrélations $R_{XY}(\tau) \geq 0.6$ calculées pour chaque couple de sites par mois pour l'ensemble de l'année 2012.

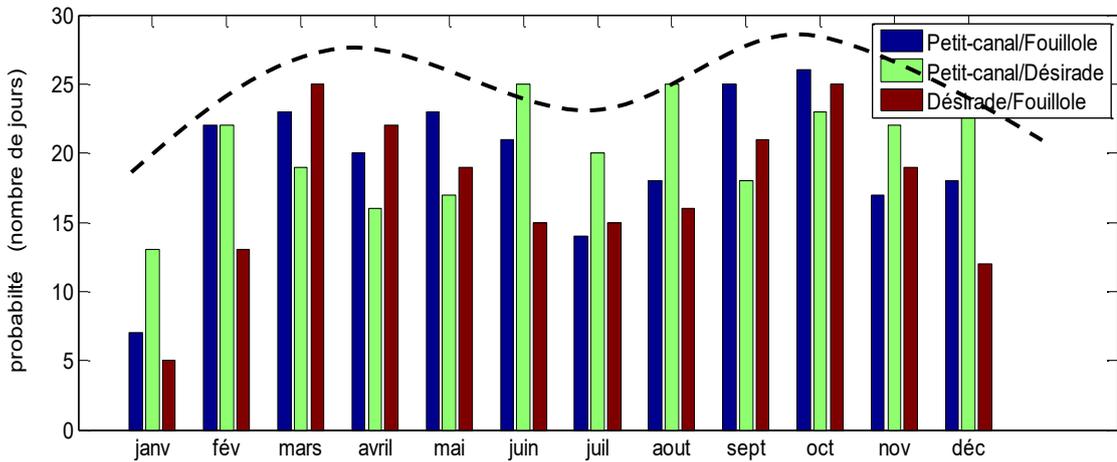


Figure 4-32. Fréquence d'apparition en nombres de jours des corrélations $R_{XY}(\tau) \geq 0.6$ calculées pour chaque couple de sites par mois pour l'ensemble des trois années. La ligne en pointillée délimite l'évolution de la moyenne journalière du flux solaire et mettant en exergue les quatre périodes saisonnières (voir chapitre 2).

On reconnaît sur la Figure 4-32, qui généralise les résultats sur nos trois années de mesures, que les fréquences des corrélations spatio-temporelles suivent la même évolution que le flux solaire théorique moyenné au cours de l'année mettant en valeur quatre saisons. Ceci indique bien que les corrélations de la variabilité du flux solaire évoluent selon les saisons. Nous pouvons noter que les probabilités les plus importantes d'obtenir des corrélations significatives sont de février à juin correspondant à la saison du carême, une période très ensoleillée et de septembre à novembre correspondant à la saison cyclonique. Janvier et juillet pourraient être décrits comme des mois transitoires. De plus sur l'ensemble des trois années, on constate que 60% des corrélations sont prédominantes pour la paire de sites Petit-canal/La Désirade. Le même type de régime climatique entre La Désirade et Petit-canal (section I.2) peut expliquer les corrélations de variabilité de flux solaire. De

plus, la raison pour laquelle les corrélations sont plus fréquentes entre La Désirade et Petit-canal (d'août à décembre) est également due à la même masse nuageuse provenant de l'Atlantique se déplaçant de La Désirade à Petit-canal sous l'influence du vent de forte composante Est. Ce phénomène se produit souvent dans l'année, particulièrement durant la saison des pluies (juillet à décembre).

Petit-canal et Fouillole sont les sites de mesures les plus proches. Ainsi, leur proximité métrique expliquerait les corrélations fréquentes observées après la paire de site Petit-canal/ La Désirade, qui sont dues principalement aux mêmes conditions météorologiques dans cette zone. Nous avons pu observer des corrélations particulièrement importantes pour cette paire de station, lorsque le mois est très sec et ensoleillé comme en février 2010 et septembre 2012 qui ont été des mois particulièrement secs selon l'historique météorologique (Météo France). Ce qui est valable aussi pour les mois très pluvieux comme observé en octobre 2010 et mai 2011 sur les histogrammes (Figure 4-29 et Figure 4-30) qui correspondaient à des mois particulièrement pluvieux.

Concernant Fouillole et La Désirade, les corrélations sont dues à une masse nuageuse, se déplaçant de La Désirade au site de Fouillole, dans la même journée. Néanmoins, à cause de la direction prédominante du vent Sud-Est Nord-Est, ce phénomène se produit à une probabilité faible comme l'indique le décalage temporelle observé (section IV.1). D'autres phénomènes physiques pourront être la source des corrélations de la variabilité du flux solaire entre ces deux stations, décrits par les deux hypothèses suivantes. La première est que, les résultats de corrélations significatives pour cette paire de sites sont dus à une zone climatique uniforme couvrant nos trois sites. La deuxième hypothèse est qu'une ligne de nuages se formant sur la Grande-Terre principalement entre le centre et le Sud de la Grande Terre, atteigne également l'île de La Désirade. Cette ligne de nuage est décrite dans Cécé et al (2013). Elle est générée principalement par la convergence de flux d'air marin due au contraste thermique terre-mer. Les conditions météorologiques (passages de petites masses nuageuses) peuvent entraîner un phénomène météorologique opposé entre les deux stations préalablement mentionné dans la section IV.1).

Pour cette analyse, un commentaire peut être donné. Trois paramètres physiques influencent les corrélations de la variabilité solaire, analysées par des outils statistiques, entre les couples de stations : la saisonnalité du flux solaire sur l'ensemble de l'année influe sur la fréquence d'apparition des corrélations significatives, la direction du vent en faveur de la localisation géographique de deux stations et les régimes climatiques propres à chaque station.

IV.3. Description dynamique des corrélations

Nous traçons le maximum d'intercorrélation en fonction des coefficients de covariance des signaux de fluctuations pour l'échelle de temps entre 1h et 2h (Figure 4 -33). En effet, nous l'avons constaté dans

la section III, la probabilité d'obtenir des coefficients supérieurs ou égaux à 0,6 pour des échelles de temps inférieures à 1h est très faible.

L'analyse de cette figure apporte des connaissances sur l'aspect dynamique et spatio-temporel des corrélations des signaux de fluctuations entre les sites de mesures. La covariance et l'intercorrélation donne deux informations différentes. La première, l'évaluation de la corrélation centrée de la variabilité (signaux de fluctuations) entre deux sites, la deuxième l'évaluation de la corrélation avec décalage temporel entre deux sites. Par conséquent la Figure 4-33, permet de mettre en évidence et de quantifier les jours où les corrélations entre les signaux de fluctuations sont sans décalage temporel. Nous avons observé que chaque tranche de valeurs des points correspond à une dynamique précise des corrélations de variabilité de flux solaire entre les sites. Nous sectionnons donc la Figure 4-33 par des tranches définies par un groupe de points situés dans un intervalle de valeurs et décrivant un aspect dynamique des corrélations.

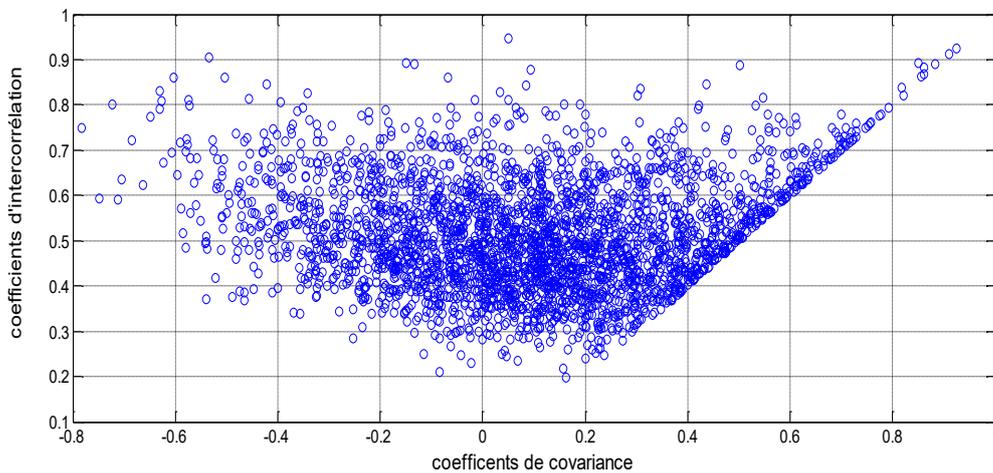


Figure 4-33. Coefficients maximaux de fonction d'intercorrélation journalière en fonction des coefficients de covariance journaliers pour 3 années de mesures, pour l'ensemble des paires de sites comparés.

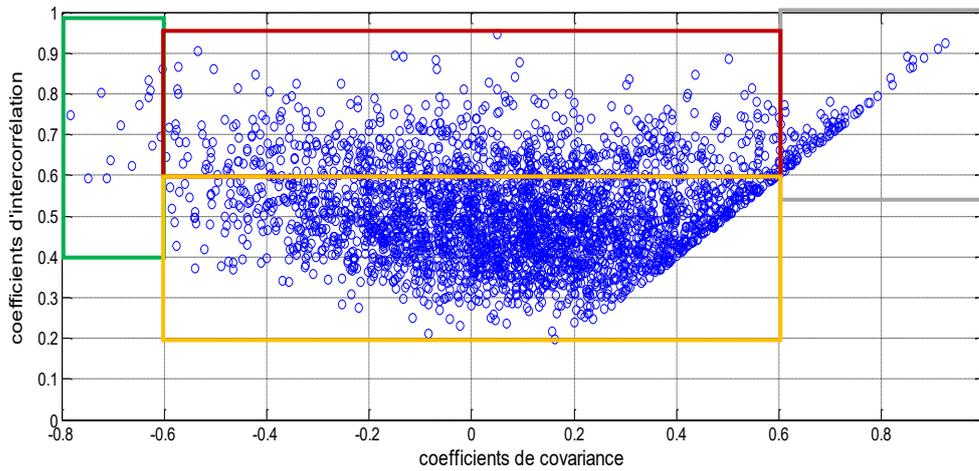


Figure 4-34. Figure 4-33 sectionnée en tranche de valeur de coefficients. *Cadre jaune* : coefficients de covariance compris dans l'intervalle $[-0,6 ; 0,6]$ et des coefficients d'intercorrélation compris entre $[0,2 ; 0,6]$. *Cadre rouge* : coefficients de covariance compris dans l'intervalle $[-0,6 ; 0,6]$ et coefficients d'intercorrélation supérieurs à 0,6. *Cadre gris* : coefficients de covariance égaux aux coefficients d'intercorrélation pour des valeurs supérieurs à 0,6. *Cadre vert* : Coefficients de covariance négatifs mais supérieurs à 0,6 en valeur absolue et coefficients d'intercorrélation supérieurs à 0,6.

Premièrement, nous pouvons noter que les coefficients d'intercorrélation ne sont jamais nulles, dues à une tendance retrouvée dans tous les signaux (amplitudes très faibles de fluctuations ou quasi nulles en début et fin de journée, et maximales en milieu de journée). Si la valeur de coefficient d'intercorrélation est la même que celle de la covariance, les deux signaux correspondants sont corrélés sans décalage temporel. Nous avons obtenus 2% de journées (soit 56 jours) où des évènements similaires se produisent sur une paire de sites au même instant t . La Figure 4-35, illustre ce cas de figure. En effet, sur cette figure, on peut observer que les signaux de fluctuations présentent la même dynamique (variations des signaux similaires) sur le site 1 et le site 2. Les fluctuations pour les deux sites, diminuent et augmentent avec des valeurs d'amplitudes similaires et ceci approximativement au même instant. Il n'y a pas de déphasage temporel entre les deux signaux. Nous pouvons observer le même phénomène sur les signaux de moyenne glissante et les signaux de flux solaire mesuré.

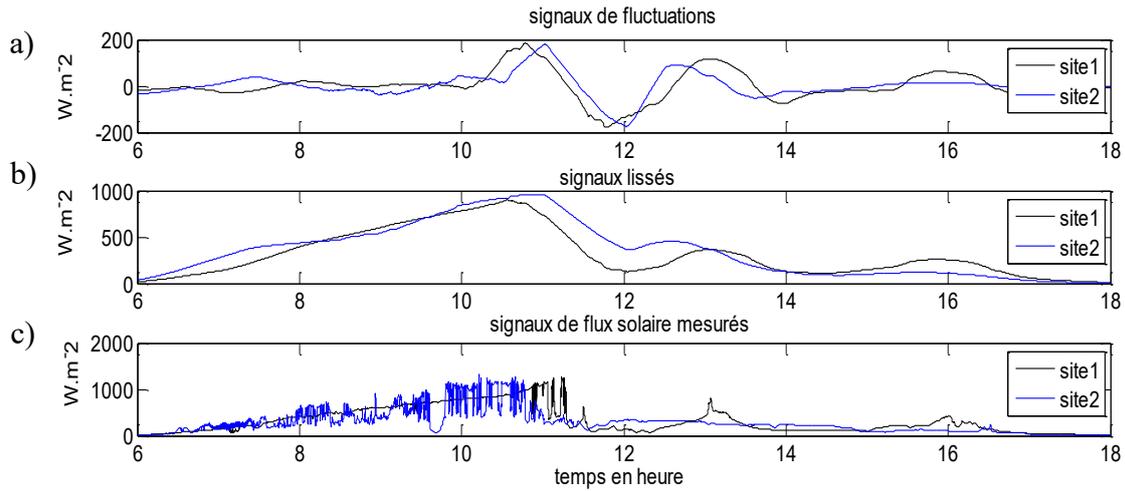


Figure 4-35. a) Signaux de fluctuations, b) signaux lissés sur les séries de flux solaire mesurés par méthode de moyenne glissante, c) signaux de flux solaire mesurés. Exemple d'une journée (29 Mai 2010) pour laquelle coefficient de covariance et le coefficient d'intercorrélation sont égaux et de valeur supérieure à 0,6. (cadre gris Figure 4-34).

Nous pouvons observer sur la Figure 4-34, une tranche de valeurs où la majorité des points se regroupent. Ces points correspondent à des coefficients de covariance compris dans l'intervalle $[-0.4 ; 0.4]$ et des coefficients d'intercorrélation compris entre $[0,4 ; 0,6]$ (cadre jaune). Ce groupe de points correspond à des journées qui ne présentent pas de corrélation de variabilité de flux solaire entre les sites. La Figure 4-36 illustre un exemple de ce cas. Cette figure montre effectivement qu'il n'y a pas de corrélations entre les signaux. Sur le site 1, la variabilité du signal de fluctuation est plus faible que le signal de fluctuation sur le site 2. Nous pouvons observer que les signaux de mesures (Figure 4-36 a)) correspondent à deux types de journées différentes : sur le site 2, il s'agit d'une journée à ciel clair intermédiaire et sur le site 1, il s'agit d'une journée à ciel clair selon les critères de classification du chapitre 3.

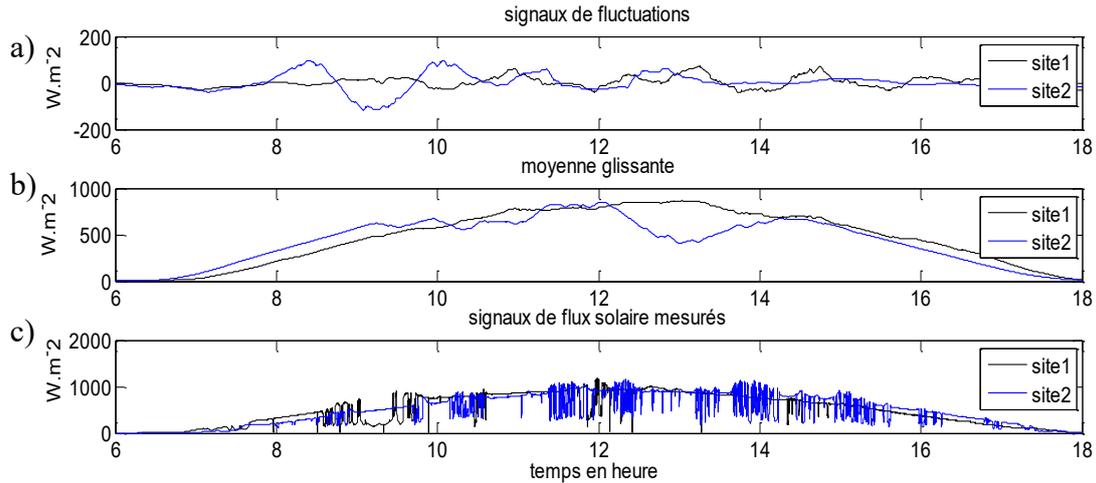


Figure 4-36. a) Signaux de fluctuations, b) signaux lissés sur les séries de flux solaire mesurés par méthode de moyenne glissante, c) signaux de flux solaire mesurés. Exemple d'une journée (14 juin 2010) pour laquelle coefficient de covariance est compris dans l'intervalle $[-0.4 ; 0.4]$ et le coefficient d'intercorrélation dans l'intervalle $[0,4 ; 0,6]$ (cadre jaune Figure 4-34).

Un autre groupe de points correspond à une faible valeur de coefficients de covariance (inférieure à 0,6) et une grande valeur de coefficients d'intercorrélation (supérieure à 0,6). Ces cas correspondent à une corrélation significative avec néanmoins un grand déphasage temporel qui explique la faible valeur obtenue par la fonction de covariance. La Figure 4-37 illustre ce cas. Nous pouvons observer sur cette figure que la variabilité des signaux est similaire sur les deux sites pour la même journée mais avec un décalage temporel égal à 1h. Sur le site 1, à 10h05 l'amplitude des fluctuations arrive à son maximum alors que sur le site 2, l'amplitude des fluctuations arrive à son maximum à 11h05. Ce même évènement se produit deux fois dans la journée comme on peut le constater également sur les signaux de moyenne glissante et les signaux mesurés. Malgré ces similitudes, on obtient pour la covariance une valeur faible à cause du décalage temporel qui n'est pas pris en compte dans le calcul. Nous pouvons noter que plus la valeur de covariance est faible avec un coefficient d'intercorrélation significatif, plus le déphasage temporel est important.

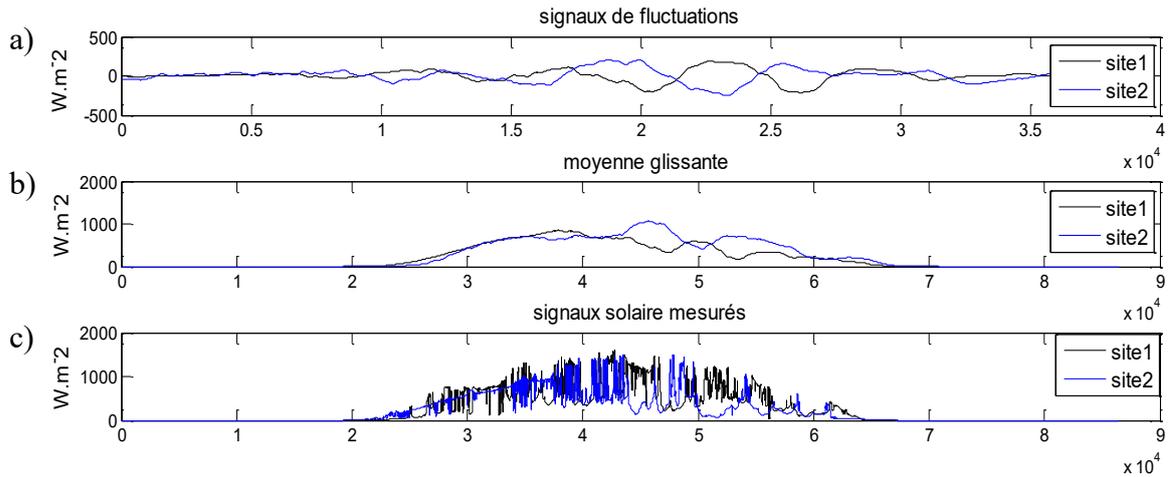


Figure 4-37. a) Signaux de fluctuations, b) signaux lissés sur les séries de flux solaire mesurés par méthode de moyenne glissante, c) signaux de flux solaire mesurés. Exemple d'une journée (25 juin 2010) pour laquelle le coefficient de covariance est compris dans l'intervalle $[-0.4 ; 0.4]$ et le coefficient d'intercorrélacion supérieur à 0,6 (cadre rouge Figure 4-34).

En ce qui concerne le cas des valeurs de coefficients de covariance négatives et supérieures en valeur absolue à 0,6 (cadre vert Figure 4-34), on observe qu'il y a bien des similitudes significatives de la variabilité du flux solaire. Néanmoins, le signal de fluctuation journalier sur le site 1 a un comportement dynamique opposé, comparé au signal de fluctuation journalier du site 2. La Figure 4-38 illustre ce cas.

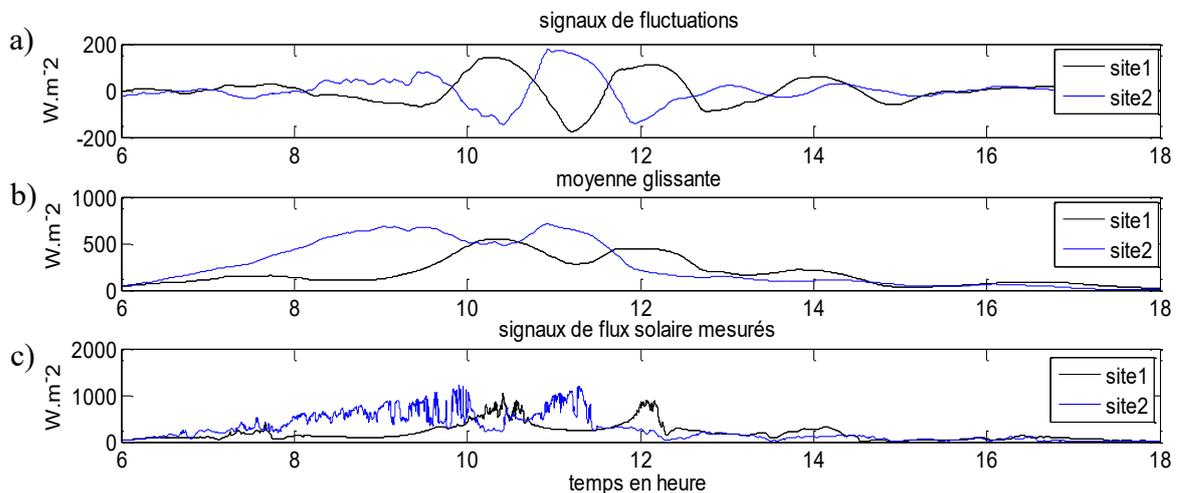


Figure 4-38. a) Signaux de fluctuations, b) signaux lissés sur les séries de flux solaire mesurés par méthode de moyenne glissante, c) signaux de flux solaire mesurés. Exemple d'une journée (11 Mai 2010) pour laquelle le coefficient de covariance est compris dans l'intervalle $[-0.6 ; -0.9]$ et le coefficient d'intercorrélacion supérieur à 0,6 (cadre vert Figure 4-34).

La Figure 4-38 montre que le signal de fluctuation sur les deux sites est bien corrélé. L'amplitude du signal de fluctuation pour le site 2 varie dans la même gamme mais en opposition de phase. Cette figure illustre un comportement dynamique : quand il y a une masse nuageuse au-dessus du

site 1, au-dessus du site 2 il y a un ciel clair. Environ 45 minutes après, l'évènement opposé se produit : un ciel clair au-dessus du site 1 et une masse nuageuse (il ne s'agit pas de la même masse) au-dessus du site 2.

A partir des intervalles de valeur de coefficient de covariance et d'intercorrélation, nous avons pu obtenir des groupes de points correspondant à une description de comportement dynamique se produisant entre deux sites. Cette analyse nous permet de quantifier sur nos trois années de mesures le type de corrélation dynamique existant entre deux sites. Le tableau ci-dessous reporte les résultats statistiques correspondant à chaque type d'évènement dynamique spatio-temporel du flux solaire entre deux sites.

Pour la paire de sites La Désirade / Fouillole on trouve des corrélations significatives quel que soit l'évènement dynamique considéré entre les deux sites à 20,6% sur trois années de mesures, pour Petit-canal/Fouillole 23,7% et Petit-canal/ La Désirade 24,4%.

Le tableau ci-dessous permet de conclure que la corrélation de la variabilité de l'ensoleillement entre deux sites est majoritairement décrite par des évènements dynamiques (passage nuageux) qui se produisent le même jour avec un déphasage temporel important d'un site à l'autre. Des conditions météorologiques uniformes sur la zone délimitée par la paire de site n'arrive que très rarement (soit 2% des cas que le même évènement dynamique se produit le même jour d'un site à l'autre sans décalage temporel : coefficient d'intercorrélation égale au coefficient de covariance). De même, un phénomène météorologique (passage de masse nuageuse) opposé à un phénomène météorologique se produisant sur un autre site (ciel clair) dans la même journée au même instant t (sans décalage temporel) est également peu probable soit 0,05%.

Ces analyses apportent des connaissances statistiques et dynamiques sur la variabilité spatio-temporelle de l'ensoleillement global à très court terme entre nos trois stations de mesures. Ce sont des informations temporelles et spatiales indicatives pour notre modèle de prédiction spatio-temporelle. Pour rappel, cette analyse a été effectuée pour les signaux de fluctuations à l'échelle de temps [1h ; 2h]. Ce qui consiste à dire que les phénomènes liés aux fluctuations qui peuvent survenir et durer sur une grande échelle de temps (grande masse nuageuse) ont de forte chance d'être retrouvé sur une autre station. Pour des fluctuations à petites échelles, il y a une très faible probabilité voir inexistante qu'il y ait des corrélations entre les sites.

Dynamique entre les deux sites	Probabilité d'apparition de l'évènement sur 3 années de mesures (2010, 2011,2012)
Pas de corrélation entre deux sites	2170 jours soit 70%
Corrélations avec des évènements de variabilité de flux solaire se produisant au même instant dans la même journée entre deux sites	56 jours soit 2%
Même dynamique de variation se produisant sur un autre site avec un déphasage temporel important dans la même journée	614 jours soit 20%
Variation du flux solaire d'un site se produisant sur un autre site avec un déphasage temporel faible dans la même journée	93 jours soit 3%
Comportement dynamique de flux solaire opposé entre deux sites	14 jours soit 0,05%
Corrélation significative indépendamment du type d'évènement dynamique rencontré entre les deux sites (tous les cas où le coefficient est supérieur à 0,6)	707 jours soit 30%

Tableau 4-2 Fréquence d'apparition en pourcentage des journées pour chaque type de corrélation dynamique pour l'ensemble des 3 années de mesures sans distinction des paires de sites.

V. Conclusion

Le but de cette étude est d'évaluer les corrélations de la variabilité de l'ensoleillement à court terme entre les stations de mesures. Les résultats obtenus nous permettent de déduire que les corrélations varient dans l'espace et dans le temps avec absence de corrélation instantanée et présence de corrélation avec décalage temporel. La séparation d'échelle de temps a révélé des corrélations significatives à faible probabilité entre nos sites de mesure pour les échelles de temps inférieures à 1 heure qui dorénavant est défini comme temps seuil. Nous avons donc pu établir qu'il y a absence de corrélation entre les fluctuations du flux solaire global pour des tendances de variations inférieures à 1 heure. Ce résultat montre que les fluctuations sur de grandes échelles de

temps correspondant à une grande masse nuageuse ont de forte chance d'être retrouvé sur une autre station, contrairement aux fluctuations à petite échelle correspondant à de petits nuages. Conformément à ce résultat, nous nous sommes focalisés sur les tendances de fluctuations comprises entre 1 heure et 2 heures. L'étude de la fonction de covariance pour les signaux de fluctuations obtenus à cette échelle de temps montre que les corrélations sont prédominantes pour la paire de sites Petit-canal /Fouillole tandis que les résultats d'intercorrélation montrent qu'il s'agit de la paire de sites La Désirade/Petit-canal. La prise en compte du décalage temporel montre une dynamique d'évolution des corrélations spatio-temporelles différentes. Notamment, l'étude des coefficients maximum de la fonction d'intercorrélation montre que le processus spatio-temporel aléatoire de l'ensoleillement n'est pas isotropique pour notre cas d'étude (la corrélation n'est pas une fonction des séparations des distances). La Désirade et Petit-canal, néanmoins séparés par un bras de mer, présentent effectivement davantage de corrélations en raison de la direction prédominante du vent Nord Est et Sud-est qui tend à déplacer la masse nuageuse, source des fluctuations à court terme, de la Désirade vers Petit-Canal.

Par analyse statistique, la dynamique de flux solaire présentent des similitudes significatives avec un décalage temporel compris entre environ [10min ;1h30min]. Cela devrait correspondre à la durée de déplacement d'une masse nuageuse d'un site à l'autre.

L'étude des paramètres météorologiques montrent une évolution des corrélations en fonction des saisons et mais aussi marque l'influence de la direction du vent sur les interactions entre les stations.

En analysant les coefficients d'intercorrélation en fonction des coefficients de covariance, nous avons pu obtenir les caractéristiques dynamiques de la corrélation spatio-temporelle entre les sites (comportement dynamique, important ou faible décalage temporel, probabilité d'un évènement de corrélation intersite). Nous avons trouvé que 30% des signaux de fluctuation présentent une corrélation significative entre les sites et 2% pour une corrélation sans déphasage temporel (évènements se produisant sur deux stations le même jour au même instant). Il serait intéressant d'étudier la distance seuil entre les sites à partir de laquelle les conditions météorologiques ont plus d'influence sur les corrélations intersites plutôt que le paramètre de distance. Plus de stations de mesures du flux solaire seraient alors nécessaires.

Les résultats obtenus à partir de notre analyse statistique et dynamique spatio-temporelle de la variabilité de l'ensoleillement à court terme, seront utilisés comme information indicative pour l'élaboration des paramètres du modèle.

Références

- Badosa J., Martial Haeffelin, Hélène Chepfer, 2013, Scales of spatial and temporal variation of solar irradiance on Reunion tropical island, *Solar Energy* 88, 42–56
- Baig A., Akhter P and A. Mufti, 1991, A novel approach to estimate the clear day global radiation“, *Renewable Energy* Vol.1, No.1, pp. 119-123.
- Barnett TP, Ritchie J, Stokes G., 1998, On the space-time scales of the surface solar radiation field. *Journal of Climate*; 11:88e96
- BNL, Brookhaven National Laboratory, 2010. Personal Communication Overview of BNL's Research Agenda on Solar Energy. BNL, Upton, NY
- Boland, J., Ridley, B., 2002. Time series modeling daily and hourly solar radiation. In: *Proceedings of EuroSun, Conference on CD-ROM*, paper 01_11, 6p.
- Brévignon C., 2005. L'environnement atmosphère de la Guadeloupe, de St Barthélémy et de St Martin, La direction Inter Régionale Antilles Guyane de Météo France.
- Cécé R., Bernard D., Dorville J.F., D'Alexis C., 2013, Numerical simulations of island-induced circulations and windward katabatic flow over the Guadeloupe archipelago, *American meteorological society, Mon. Wea. Rev.*, 142, 850–867.
- CSI, California Solar Initiative R&D Program, 2010. California Public Utility Commission, San Francisco, CA.
- Gueymard Christian A., Wilcox Stephen M., 2011, Assessment of spatial and temporal variability in the US solar resource from radiometric measurements and predictions from models using ground-based or satellite data, *Solar Energy*, Volume 85, Issue 5, Pages 1068-1084
- Hoff, T., Perez, R., 2010. Quantifying PV power output variability, *Solar Energy* 84 (10), 1782–1793.
- Horvatic, D., Stanley, H.E., Podobnik, B., 2011. Detrended crosscorrelation analysis for non-stationary time series with periodic trends. *Eur. Lett.(EPL)* 94
- IEA Solar Heating and Cooling Programme, 2010. Task 46 – Solar Resource Assessment and Forecasting, Subtask A – Resource Applications for High Penetration of Solar Technologies. International Energy Agency, Paris, France.
- Jenkins, N., 2004. Integrating PV with the power system. In: *Keynote Lecture on the Conference EuroSun, Freiburg, Germany*.
- Kasten F., 1980, A simple parametrization of two pyrheliometric formulae for determining the Linke turbidity factor, *Meteorologische Rundschau*, 33:124e7
- Kasten, 1997, 'The Linke Turbidity Factor Based on Improved Values of the Integral Rayleigh Optical Thickness', *Solar Energy*, Vol. 60, pp. 17 – 24
- Koussa M., Malek A. et Haddadi M., 2006, 'Validation de Quelques Modèles de Reconstitution des Eclaircissements dus au Rayonnement Solaire Direct, Diffus et Global par Ciel Clair', *Revue des Energies Renouvelables* Vol. 9 N°4 pp. 307 – 332

Kyriakidis, P. C. et Journel, A. G., 1999, Geostatistical space-time models: a review. *Math.Geol.* 31, 651-684.

Lave M. , Kleissl J., 2010, Solar variability of four sites across the state of Colorado, *Renewable Energy* 35, 2867e2873

Lave M., Jan Kleissl, Ery Arias-Castro, 2012, High-frequency irradiance fluctuations and geographic smoothing, *Solar Energy* 86, 2190–2199

Mechagrane A., Chaoui-Roquai M., Buret-Bahraoui J.,1993, Modèle d'estimation de l'irradiation solaire globale d'une surface horizontale au sol à partir des images satellitaires, *Journal de Physique III*, Vol.3, Issue 5, , pp 1005-1019

Mesri-Merad M., Rougab I, Cheknane A, Bachari N.I, 2012, Département de Génie Electrique, Faculté des Sciences et des Sciences de l'Ingénieur, Université Amar Tilidji, Route de Ghardaïa, Laghouat, Algérie

Mills, A., Wiser, R., 2010. Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power. LBNL Report No. 3884E.

Molineaux B, Ineichen P, O'Neill N., 1998, Equivalence of pyrhelimetric and monochromatic aerosol optical depths at a single key wavelength. *Applied Optics* ;37(30):7008e18.

Morf, H., 1998. The stochastic two-state solar irradiancemodel. *Solar Energy* 60, 257–270.

Moumami A., Hamani N., Moumami N. et.Mokhtari Z., 2006, "Estimation du rayonnement solaire par deux approches semi empiriques dans le site de Biskra", 8ème Séminaire International sur la Physique Énergétique, Centre Universitaire de Béchar-Algérie, 11,12

Mueller R, Dagestad K, Ineichen P, Schroedter-Homscheidt M, Cros S, Dumortier D, et al. , 2004, Rethinking satellite-based solar irradiance modeling: the SOLIS clear-sky module. *Remote Sensing of Environment*, 91(2):160e74.

Murata, A., Yamaguchi, H., Otani, K., 2009. A method of estimating the output fluctuation of many photovoltaic power generation systems dispersed in a wide-area. *Electrical Engineering in Japan* 166 (4), 9–19.

Mustacchi, C., Cena, V., Rocchi, M., 1979. Stochastic simulation of hourly global radiation sequences. *Solar Energy* 23,47–51.

Perez Richard, Kivalov Sergey, Schlemmer James, Karl Hemker Jr, Tom Hoff, 2011, Parameterization of site-specific short-term irradiance variability , *Solar Energy* 85, 1343–1353

Perez R., Sergey Kivalov, Jim Schlemmer, Karl Hemker Jr, Thomas E. Hoff, 2012, Short-term irradiance variability: Preliminary estimation of station paire correlation as a function of distance, *Solar Energy* 86, 2170–2176

Régis S., T. Soubdhan, R. Calif, M. Abadi, R. Blonbou E. Grandchand, F. Dupont, J. Desachy, 2007, Singularity detection in Solar Radiation Signal using wavelet transform, IEEE PES General Meeting, Tampa

Rigollier C, Bauer O, Wald L., 2000, On the clear sky model of the ESRA e European Solar Radiation Atlas e with respect to the Heliosat method. *Solar Energy*, 68(1):33e48.

Rosenberg et al. (1989), <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2728136/>

SMUD (ask Obadiah)—alternatively, use Mark Rawson, 2010. Solar and the smart grid session SPI

Soubdhan Ted, Calif Rudy, Dupont Frédéric, Blonbou Ruddy, SFT 2007, Solar flux variability in Guadeloupe: Comparative study in two geographical points, French Congress of Thermal

Teolan Tomson , Gunnar Tamm, 2006, Short-term variability of solar radiation, Solar Energy 80, 600–606

USDOE, US Department of Energy, 2009. Recovery Act: High Penetration Solar Deployment R&D Program. <http://www1.eere.energy.gov/solar/financial_opps_detail.html?sol_id=286>.

Voyant C., 2011, Time series prediction of global solar radiation and photovoltaic energy production using artificial neural networks, thèse, University of Corse-Pascal Paoli, UMR CNRS 6134(SPE)

Wiener (1949) ,<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2728136/>

Yang D., Dong Z., Reindl T., Jirutitijaroen P., Walsh W.M., 2014, Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage, Solar Energy, Volume 103, Pages 550-562

Zidani C., Benyarou F.et Benyoucef B., 2003,“Simulation de la position apparente du soleil et estimation des énergie solaires incidentes sur un capteur plan pour la ville de Tlemcen en Algérie“, Rev. Energ. Ren, Vol.6, pp. 69-76.

Zhao Xiaojun, Pengjian Shang, Aijing Lin, Gang Chen, 2011, Multifractal Fourier detrended cross correlation analysis of traffic signals, Physica A: Statistical Mechanics and its Applications, Volume 390, Issues, Pages 3670-3678

Chapitre 5. Modèle spatio-temporel de prévision à court terme de l'ensoleillement global

Table des matières

I. Modèle Vecteur AutoRégressif VAR spatio-temporel pour des données dispersées dans l'espace	Erreur ! Signet non défini.
I.1. Principe général du modèle VAR	Erreur ! Signet non défini.
I.2. Matrice des distances	Erreur ! Signet non défini.
I.3. Stationnarisation spatio-temporelle des données	Erreur ! Signet non défini.
I.4. Stratégie de sélection des prédicteurs	Erreur ! Signet non défini.
I.5. Estimation des paramètres	Erreur ! Signet non défini.
II. Le modèle spatio-temporel VAR pour la prévision de l'ensoleillement global	Erreur ! Signet non défini.
II.1. Un modèle multivarié pour notre contexte spatio-temporel	Erreur ! Signet non défini.
II.2. Matrice des données	Erreur ! Signet non défini.
II.3. Sélection des prédicteurs	Erreur ! Signet non défini.
II.4. Estimation des paramètres du modèle	Erreur ! Signet non défini.
III. Analyse de la performance prédictive	Erreur ! Signet non défini.
III.1. Test sur la sélection de l'ordre p	Erreur ! Signet non défini.
III.2. Analyse de la performance prédictive pour deux structures spatiales	Erreur ! Signet non défini.
III.3. Analyse de la performance prédictive en fonction de la variabilité du flux solaire	Erreur ! Signet non défini.
III.4. Analyse de la performance prédictive pour les saisons	Erreur ! Signet non défini.
III.5. Analyse de la performance prédictive pour différents horizons	Erreur ! Signet non défini.
IV. Amélioration des paramètres du modèle spatio-temporel VAR	Erreur ! Signet non défini.
IV.1. Tests des paramètres du modèle et amélioration de la performance	Erreur ! Signet non défini.
IV.2. Comparaison avec d'autres modèles	Erreur ! Signet non défini.
IV.3. Limites du modèle VAR et perspectives	Erreur ! Signet non défini.
V. Conclusion	Erreur ! Signet non défini.
Références	Erreur ! Signet non défini.

Suite à l'état de l'art des travaux effectués sur la prédiction spatio-temporelle de l'ensoleillement global, et tenant compte de notre contexte d'étude (peu de sites de mesures mais des données riches en dimension temporelle), nous décidons d'appliquer un modèle de prédiction basé sur la méthodologie du modèle VAR spatio-temporel de Luna et Genton (2005). Ce modèle diffère des modèles linéaires traditionnels en incorporant des paramètres spatiaux et temporels, déduisant la prédiction de l'ensoleillement par combinaison linéaire des séries de temps passées de la localisation et des séries de temps passées des localisations corrélées. Ce modèle peut s'appliquer pour des données dispersées dans l'espace et riches en dimension temporelle. Nous ciblons la prévision à un horizon de 10 min selon les besoins du gestionnaire de réseau, toutefois le modèle sera également évalué pour des horizons de prédictions inférieures ou égale à l'heure (5min à 1h).

En première partie, le principe statistique et la méthodologie générale du modèle de prédiction sont explicités. Dans la seconde partie, nous appliquons cette méthodologie pour la prévision de l'ensoleillement global à court terme. Une sélection des paramètres spatiaux et temporels est présentée pour l'optimisation du modèle. Cette analyse sera suivie d'une évaluation de la performance prédictive du modèle pour différentes dynamiques de flux solaire, pour plusieurs horizons, et sera comparé à d'autres modèles de la littérature.

Finalement, nous discuterons des perspectives d'amélioration de la performance du modèle.

I. Modèle Vecteur AutoRégressif VAR spatio-temporel pour des données dispersées dans l'espace

Déjà cité et explicité dans le chapitre de l'état de l'art, le modèle de prédiction présenté dans ce chapitre est un modèle spatio-temporel qui est adapté pour fournir des prévisions de temps dans les applications environnementales où les données sont rares dans la dimension spatiale, mais riches dans la dimension temporelle. Une autre particularité du modèle réside dans le fait que les hypothèses de stationnarité spatio-temporelle couramment utilisées en modélisation spatiale ne sont pas nécessaires. Ce modèle consiste en une spécification du vecteur autorégressif (VAR) pour chaque localisation.

I.1. Principe général du modèle VAR

Le modèle VAR est particulièrement populaire pour modéliser les séries temporelles multivariées à cause de sa simplicité et de sa richesse.

Le processus du modèle se base sur le modèle défini dans Luna et Genton (2005). Ce modèle est bien adapté à notre contexte d'étude car déjà validé pour un système spatio-temporel composé de

quatre stations (voir le chapitre 2 : état de l'art) et permet de modéliser les interactions existantes entre les variables. Ainsi, nous proposons dans cette recherche un modèle de prévision basé sur une méthode de modèle multivarié. A contrario d'un modèle autorégressif univarié dans lequel la valeur d'une variable est obtenue par ses propres valeurs retardées, le modèle VAR contient un ensemble de n variables dans lequel chacune d'elle est exprimée par une combinaison linéaire des p retards de ses propres valeurs et de toutes les $n-1$ variables, plus une erreur ou un terme appelé innovations. De plus, des paramètres spatio-temporels γ sont intégrés. L'apport de ce modèle est le développement d'une stratégie de sélection des prédicteurs (stations). De Luna et Genton (2005) montrent que leur méthodologie peut s'appliquer sans faire d'hypothèse de spatio-stationnarité, pour un contexte d'étude composé de très peu de stations, souvent supposée dans la littérature. Pour cela, ils appliquent leur modèle sur les mêmes données spatio-temporelles analysées par Tonellato (2001) qui suppose une hypothèse d'isotropie en utilisant une fonction spatiale isotrope de corrélation exponentielle. Le modèle spatio temporel VAR de de Luna et Genton (2005), montre une bonne performance de prévision (détaillée dans l'état de l'art) sans hypothèse de restriction de stationnarité spatio-temporelle.

Le modèle spatio temporel vecteur autorégressif est décrit par le système dynamique suivant dans de Luna et Genton (2005) :

Soient les données spatio-temporelles $Z(s_i, t)$, les observations d'un processus stochastique dans $\mathbb{R}^2 \times \mathbb{R}$. Chaque observation est localisée à une station $s_i = 1, \dots, n$ au temps $t = 1, \dots, T$. Le modèle de prédiction VAR (vecteur autorégressif) est tel que :

$$Z_t - \beta = \sum_{i=1}^p R_i (Z_{t-i} - \beta) + \varepsilon_t \quad (5.1)$$

où $Z_t = (Z(s_1, t), Z(s_2, t), \dots, Z(s_n, t))'$, $\varepsilon_t = (\varepsilon_t(s_1), \dots, \varepsilon_t(s_n))'$ est un bruit blanc avec $E(\varepsilon_t) = 0$, $E(\varepsilon_t, \varepsilon_u) = 0$ pour $u \neq t$, $E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon$, $\beta = (\beta(s_1), \dots, \beta(s_n))'$ est la tendance spatiale, R_i sont les matrices $n \times n$ de paramètre inconnu. Les n lignes de ces matrices correspondent aux n localisations où les séries de temps sont observées.

La particularité de ce modèle spatio-temporel VAR réside dans le fait qu'il traite, dans le processus du modèle, chaque localisation (sites) séparément. La méthodologie générale du modèle proposé est décrite dans la section suivante.

Il pourrait se résumer en quatre étapes principales qui détermineront les quatre sections suivantes:

- 1) Ordonner les stations selon une structure spatiale définie, déterminant une matrice des distances
- 2) Paramétrer la matrice des données et retirer les tendances temporelles et spatiales
- 3) Définir une stratégie itérative de sélection des variables explicatives et du décalage temporel

- 4) Prédire à court terme les données en calculant les paramètres de la régression linéaire gaussienne composée des variables explicatives sélectionnées.

Ce modèle est développé à l'aide du logiciel R project.

I.2. Matrice des distances

La paramétrisation du modèle commence par la matrice des distances qui doit être définie selon une structure spatiale choisie, bien souvent ordonnée selon l'ordre croissant des distances géographiques entre les sites. L'ordre choisi n'implique aucune hypothèse de stationnarité ou même d'isotropie en contraste avec Tonellato (2001). Ces hypothèses sont arbitraires puisque, avec seulement quatre stations dans le cas d'étude de Tonellato (2001), de Luna et Genton (2005) affirme qu'il n'est pas possible d'évaluer leur validité. La structure spatiale définie (ordre des sites) aide à l'amélioration de la sélection du modèle. Un ordre inapproprié implique seulement moins de coefficients des paramètres R_i de l'équation (1) qui seraient identifiés sous forme de zéros, ce qui impliquerait de plus grandes erreurs de prédiction (de Luna et Genton, 2005). Il est donc nécessaire d'évaluer l'ordre spatial qui donnera la meilleure performance du modèle

I.3. Stationnarisation spatio-temporelle des données

Le modèle n'est applicable que sur les données spatio-temporellement stationnaires. Les transformations sont nécessaires pour stabiliser une série avant d'estimer des modèles. Ceci est particulièrement important pour les modèles ARIMA, VARMA, VAR qui requièrent que les séries soient stationnaires avant l'estimation des modèles. Cette étape consiste à supprimer les tendances temporelles et spatiales des données de mesures. Bien que les séries environnementales ne soient pas stationnaires, les modèles autorégressifs sont efficaces tant que la série peut être rendue stationnaire par l'application de transformations. Pour rappel une série est stationnaire, si le niveau global (moyenne) et l'écart moyen à partir du niveau (variance) sont constants dans l'ensemble de la série.

Soit un processus temporel à valeurs réelles et en temps discret Z_1, Z_2, \dots, Z_t . Il est dit stationnaire au sens faible (ou de « second ordre », ou « en covariance ») si :

- 1) $E[Z_i] = \mu \quad \forall i=1 \dots t$
- 2) $Var[Z_i] = \sigma^2 \neq \infty \quad \forall i=1 \dots t$
- 3) $Cov[Z_i, Z_{i-k}] = f(k) = \rho_k \quad \forall i=1 \dots t, \forall k=1 \dots t$

La stationnarité faible est dite stationnarité du second ordre, car sa définition se base exclusivement sur les deux premiers moments de la variable aléatoire de Z_t .

Les données sans tendance temporelle sont définies à l'aide d'une méthode de différenciation telles que :

$$\nabla Z(s_i, t) = \log Z(s_i, t) - \log Z(s_i, t - 1) \quad (5.3)$$

Où $\nabla Z(s_i, t)$ est le processus stationnaire temporel par définition. Une fois la tendance temporelle retirée, nous estimons la tendance spatiale sur les données stationnaires temporellement. Dans la dimension spatiale, la différenciation a été également suggérée pour par exemple les fonctions aléatoires intrinsèques (Matheron 1973). La tendance spatiale supposée dépendante seulement de s_i est alors estimée par :

$$\beta(s_i) = E(\nabla Z(s_i, t) - \nabla Z(s_i, t - 1)) \quad (5.4)$$

Par conséquent, les données sans tendance spatio temporelle sont telles que :

$$\tilde{Z}(s_i, t) = \nabla Z(s_i, t) - \beta(s_i) \quad (5.5)$$

La tendance spatiale est réinjectée dans le résultat final de prédiction.

La propriété de stabilité du modèle assure que l'itération du système dynamique décrit par l'équation (5.1) converge vers une constante autrement dit que la méthode de stationnarisation des séries est valide. Cette stabilité peut être vérifiée par la méthode suivante.

Le modèle est dit stable quand le polynôme caractéristique défini par le déterminant a ses racines à l'extérieur du cercle unité c'est-à-dire (Hamilton (1994)) .:

$$\det(I_N - R_1 z - \dots - R_p z^p) \neq 0 \quad \text{pour } z \in \mathbb{C}, |z| \geq 1 \quad (5.6)$$

Où I_N est la matrice identité $N \times N$.

Une autre méthode pour caractériser notre système dynamique est de calculer les valeurs propres de la matrice des paramètres R_i . Si les valeurs propres de la matrice des paramètres R_i sont inférieures à 1, le système dynamique de l'équation (3) est dit stable (Hamilton 1994) :

Propriétés : Les valeurs propres de la matrice des paramètres satisfont :

$$|I_n \lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \dots - \phi_p| = 0 \quad (5.7)$$

ϕ matrice des coefficients autorégressif pour $j=1,2,\dots,p$, où I_N est la matrice identité $N \times N$

Par conséquent un processus VAR(p) est covariance-stationnaire aussi longtemps que $\|\lambda\| < 1$ (Hamilton 1994 page 259).

Nous chercherons à utiliser cette propriété plus loin, pour démontrer que cette méthode de stationnarisation appliquée à l'ensoleillement impliquera la stationnarité temporelle du processus stochastique.

I.4. Stratégie de sélection des prédicteurs

Dans un modèle multivarié, il est fondamental de choisir les variables explicatives. En d'autres termes, le modèle devra évaluer quelles seront les séries temporelles des stations qui pourront être utilisées pour prédire la série temporelle d'une localisation. La stratégie se basera sur une introduction successive de variables de telle sorte qu'avant toute introduction d'une variable supplémentaire, la signification des variables explicatives déjà présentes dans l'équation soit testée. Les variables qui n'apportent pas de contribution significatives sont éliminées. Plusieurs techniques statistiques destinées à valider et améliorer les résultats fournis par des techniques basées sur la régression linéaire multiple existent, telles que les sélections par optimisation des critères : R^2 , critère du PRESS, ou des techniques basées sur le F partiel de Fisher (Rakotomalala, 2011). Les techniques les plus appliquées pour des modèles basés sur des régressions linéaires multiples sont : la sélection de variable basée sur la corrélation partielle, la sélection de variables basée sur les régressions partielles et celle basée sur les critères d'information (Rakotomalala, 2011).

La méthodologie développée pour le modèle VAR spatio-temporel dans de Luna et Genton (2005) est basée sur la procédure de sélection de variables par corrélation partielle.

Supposons que nous voulons prédire la valeur du système pour une station s_j au temps t ; $Z(s_j, t)$.

Dans ce cas nous considérons les prédicteurs dans l'ordre suivant :

$$Z(s_j, t - 1), Z(s(1), t - 1), Z(s(2), t - 1), \dots, Z(s(n - 1), t - 1), Z(s_j, t - 2), \\ Z(s(1), t - 2), Z(s(2), t - 2), \dots, Z(s(n - 1), t - 2)$$

où $s(1), \dots, s(n-1)$ est un ordre des $n-1$ stations. Regardons les autocorrélations partielles en renommant la séquence précédente comme :

$$X_1 = Z(s_j, t - 1), X_2 = Z(s(1), t - 1), \dots, X_n = Z(s(n - 1), t - 1);$$

$$X_{N+1} = Z(s_j, t - 2); X_{N+2} = Z(s(1), t - 2), \dots, X_{2N} = Z(s(n - 1), t - 2), \dots$$

La fonction de corrélation partielle (PCF) pour une station s_j est définie par :

$$\rho_{Z_{S_j}(h)} = \text{Corr}(Z(s_j, t), X_h | X_1, \dots, X_{h-1}) \quad (5.8)$$

Soit h_1 définit tel que $\rho_{Z_{S_j}(h_1)} \neq 0$ et $\rho_{Z_{S_j}(h)} = 0$ pour $h_1 < h \leq n$. De même, une valeur h_i peut être définie pour chaque temps de retard i , tel que $\rho_{Z_{S_j}(h_i)} \neq 0$, $\rho_{Z_{S_j}(h)} = 0$ pour $h_i < h \leq in$. Les ordres de h_i peuvent être identifiés en observant la fonction de corrélation partielle $\hat{\rho}_{Z_{S_j}(h)} = \text{Corr}(Z(s_j, t) - \hat{P}(Z(s_j, t) | X_1, \dots, X_{h-1}), X_h - \hat{P}(X_h | X_1, \dots, X_{h-1}))$ où $\hat{P}(Z(s_j, t) | X_1, \dots, X_{h-1})$ est le meilleur prédicteur linéaire de $Z(s_j, t)$ donné par X_1, \dots, X_{h-1} .

Identification de l'algorithme

Etape 0 : Choisir un des sites observés s_j

Etape 1 : Identifier h_1 en observant l'échantillon $\hat{\rho}_{Z_{S_j}(h)}$, $h=1, \dots, n$

Etape 2 : Identifier h_2 en observant l'échantillon $\hat{\rho}_{Z_{S_j}(h)}$, $h=n+1, \dots, 2n$ quand X_{h+1}, \dots, X_n a été rejeté comme étant inutile en expliquant $Z(s_j, t)$ dans l'étape précédente

Etape 3 : Identifier h_3 en observant l'échantillon $\hat{\rho}_{Z_{S_j}(h)}$, $h=2n+1, \dots, 3n$ quand X_{h+1}, \dots, X_n et X_{2h+1}, \dots, X_{2n} ont été rejetés comme étant inutiles en expliquant $Z(s_j, t)$ dans l'étape précédente

Etape 4 : L'étape 3 est répétée de manière similaire pour tous les décalages temporels nécessaires afin d'identifier h_4, h_5, \dots

Etape 5 : Répéter les étapes précédentes pour tous les sites observés.

La suppression de prédicteurs inintéressants à chaque décalage temporel améliore l'efficacité en évitant l'estimation des coefficients nuls. Cette stratégie sélectionne alors les variables explicatives et le décalage temporel maximal des séries temporelles pour la combinaison linéaire des séries de temps pouvant prédire une localisation donnée. Cette procédure est également utilisée pour l'identification de l'ordre de modèles AR linéaires. Pour de tels modèles, tous les paramètres sont évalués jusqu'à un décalage donné par convention.

Une alternative à l'utilisation de cette technique est l'utilisation d'un critère de sélection automatique du modèle à chaque étape de la stratégie d'identification : AIC (information d'Akaike critère, Akaike (1974)) ou BIC (critère d'information bayésien, Schwarz (1978)). Le premier critère étant généralement préféré à des fins prédictives.

I.5. Estimation des paramètres

L'estimation des paramètres peut être évaluée par différentes méthodes telles que le maximum de vraisemblance (si des hypothèses distributionnelles sont faites), la méthode des moindres carrés ou des estimateurs de moment (Yule Walker). On peut trouver la théorie liée à de tels estimateurs dans Lütkepohl (1991, Chap.3) and Pena et al. (2001, Chap.14). Pour notre modèle l'estimation des paramètres se fait par la méthode des moindres carrés. Notre système peut s'écrire sous la forme de régression linéaire multiple telle que :

$$Z_{i,t} = R_i + \sum_{k=1}^p (R_{i,1}^k \dots R_{i,n}^k) \begin{pmatrix} Z_{1,t-k} \\ \vdots \\ Z_{n,t-k} \end{pmatrix} + \varepsilon_{i,t} \quad i = 1 \dots p \quad (5.9)$$

Soit :

$$Z_{i,t} = \underbrace{(R_i \ R_{i,1}^1 \dots R_{i,n}^1 \ R_{i,1}^2 \dots R_{i,n}^2 \ R_{i,1}^p \dots R_{i,n}^p)}_{=\theta_i'} \begin{pmatrix} 1 \\ Z_{1,t-1} \\ \vdots \\ Z_{n,t-1} \\ Z_{1,t-2} \\ \vdots \\ Z_{n,t-2} \\ \vdots \\ Z_{n,t-p} \end{pmatrix} + \varepsilon_{i,t} \quad (5.10)$$

$= Z$

L'équation 5.10 peut se résumer à l'équation de régression :

$$Z_{i,t} = \theta_i' z + \varepsilon_{i,t} = z' \theta_i + \varepsilon_{i,t} \quad (5.11)$$

La méthode des moindres carrés consiste à chercher une valeur $\theta_i = \hat{\theta}_i$ qui minimise la somme des carrés des déviations. $\hat{\theta}_i$ est la solution de ce problème de minimisation appelé estimateur des moindres carrés. En pratique z est une matrice de taille $((np+1) \times (T-p))$ et $Z_{i,t}$ est un vecteur de taille $((T-p) \times 1)$ de forme $(Z_{i,p+1} \dots Z_{i,T})'$ avec T la longueur des observations et p l'ordre du modèle VAR(p). La meilleure estimation de θ_i est donnée par :

$$\hat{\theta}_i = (zz')^{-1} z \begin{pmatrix} Z_{1,p+1} \\ \vdots \\ Z_{i,T} \end{pmatrix} \quad (5.12)$$

La méthode générale du modèle étant décrite, nous chercherons à l'appliquer pour la prévision solaire à court terme

II. Le modèle spatio-temporel VAR pour la prévision de l'ensoleillement global

Dans cette section nous développons un modèle spatio temporel de l'ensoleillement global pour la prévision à court terme à partir de la méthodologie décrite par de Luna et Genton (2005). Nous avons cherché à définir les paramètres spatiaux et temporels adaptés à notre contexte d'étude. Une optimisation du modèle a donc été effectuée. Préalablement, nous discuterons de l'amélioration possible de la prédictibilité par un modèle multivarié tel que le modèle spatio-temporel VAR proposé ici, pour notre contexte d'étude spatio-temporel.

II.1. Un modèle multivarié pour notre contexte spatio-temporel

La variabilité temporelle et spatiale de l'ensoleillement est liée à différents mécanismes atmosphériques complexes. Les stations de mesures sont vraisemblablement sujets à différents microclimats (côte au vent, intérieur des terres, séparation du site de La Désirade par la mer,...). Néanmoins, nous avons constaté grâce à une étude préliminaire présentée dans le chapitre III, que les corrélations linéaires entre les trois sites calculées à partir de la fonction de covariance et de la fonction d'inter corrélation sont bien existantes.

Des tests ont été effectués en faisant une prédiction de l'ensoleillement par pas de 10 minutes pour un horizon de 10 minutes pour une localisation.

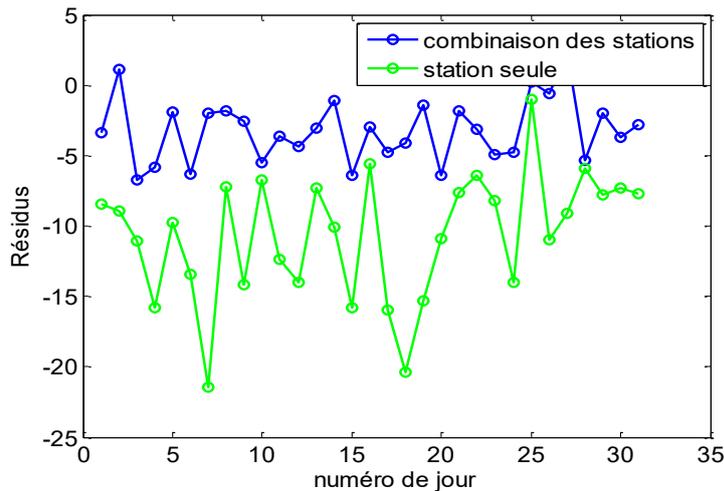


Figure 5-1 Moyenne journalière des résidus des modèles de régression linéaire simple (en vert) et de la régression linéaire multiple (en bleu) résultant de la combinaison linéaire des stations

La première méthode de prévision est basée sur une régression linéaire des données passées de ce site tel un modèle AR et la deuxième méthode est basée sur une régression linéaire multiple déduit de la combinaison linéaire des séries de temps passées du site de référence et des séries de temps passées des localisations spatiales corrélées. La Figure 5-1 représente les résultats des moyennes journalières des résidus obtenus par les deux modèles pour 31 jours pour l'exemple de la station de Petit-canal.

On constate que la courbe en bleue correspondant aux résultats de résidus pour une combinaison linéaire entre les séries de temps passées du site de référence et les sites corrélés présente de meilleurs résultats avec une erreur maximale de $4,5 \text{ W.m}^{-2}$ au lieu de 24 W.m^{-2} en valeur absolue pour le modèle ne considérant que la série de temps passée du site de référence. Par conséquent, la connaissance des séries temporelles passées de flux solaire mesurées sur les trois stations apporte une information supplémentaire et améliore la prédictibilité du flux solaire localisé. De nombreux chercheurs ont démontré l'amélioration de la qualité de la prévision par des techniques de modèles multivariés, notamment en incluant une information spatiale. On peut citer Glasbey et Allcroft (2008), Yang et al (2014), Bessa et al (2015).

Un modèle multivarié tel que le modèle spatio-temporel VAR proposé, améliorera donc la prévision de l'ensoleillement.

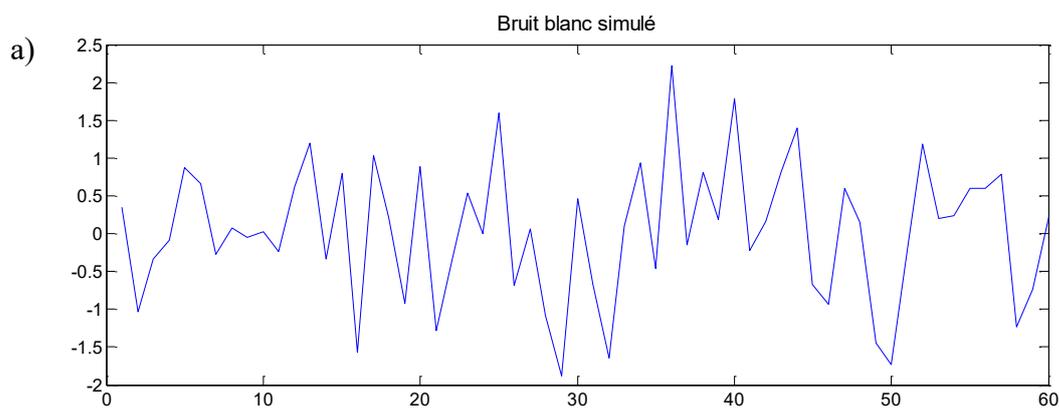
Les sections décrivent les paramètres du modèle pour la prévision à court terme de l'ensoleillement.

II.2. Matrice des données

II.2.1 Données journalières du flux solaire

Les données d'origine sont mesurées à une fréquence de 1Hz mais nous les moyennons à 10 minutes pour la suite de l'étude du modèle, puisque le gestionnaire du réseau nécessite 15 minutes pour changer de système d'exploitation d'énergie électrique sur le réseau. Une prévision à un horizon de 10 minutes est alors ciblée, bien que nous testerons notre modèle pour d'autres échelles de temps afin d'évaluer sa performance. L'analyse des résultats est effectuée sur trois années de mesures 2010, 2011 et 2012. L'ensemble des données disponibles est de 600 minutes d'observations de l'ensoleillement global pour une journée (soit 60 points par pas de 10minutes). Notons que le modèle est approprié une fois que la tendance temporelle et spatiale ont été retirées.

Le modèle est applicable que sur des séries stationnaires. Le modèle est exécuté pour un vecteur de données correspondant à une période d'une journée. Ce choix est justifié par l'analyse d'un outil statistique : la fonction d'autocorrélation partielle. Cette analyse statistique a montré que pour une série temporelle correspondant à une journée d'ensoleillement, on retrouve les mêmes caractéristiques qu'un processus stationnaire autorégressif d'ordre 1. En effet, la Figure 5-3 indique des coefficients inférieurs à 0,4 restant dans l'intervalle $[-0,2 ; 0,2]$ à partir du décalage temporel d'ordre 2. Tandis que lorsqu'on analyse la fonction d'autocorrélation partielle pour deux journées (Figure 5-4) ou un mois (Figure 5-5), on retrouve les caractéristiques d'un processus autorégressif supérieur ou égale à l'ordre 2. Ce résultat est dû aux tendances cycliques et tendance saisonnière qui composent une série temporelle d'ensoleillement global pour une période supérieure à une journée. On peut comparer ces résultats d'autocorrélation partielle à celui d'un bruit blanc qui constitue la référence d'un processus stationnaire (Figure 5-2).



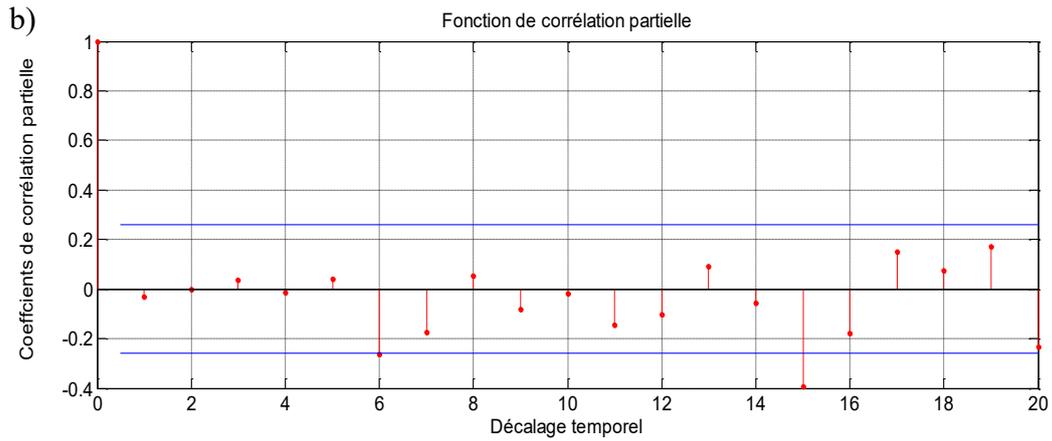


Figure 5-2. a)Exemple d'une série temporelle correspondant à un bruit blanc. simulé b) Fonction d'autocorrélation partielle correspondante, caractéristique d'un processus stationnaire.

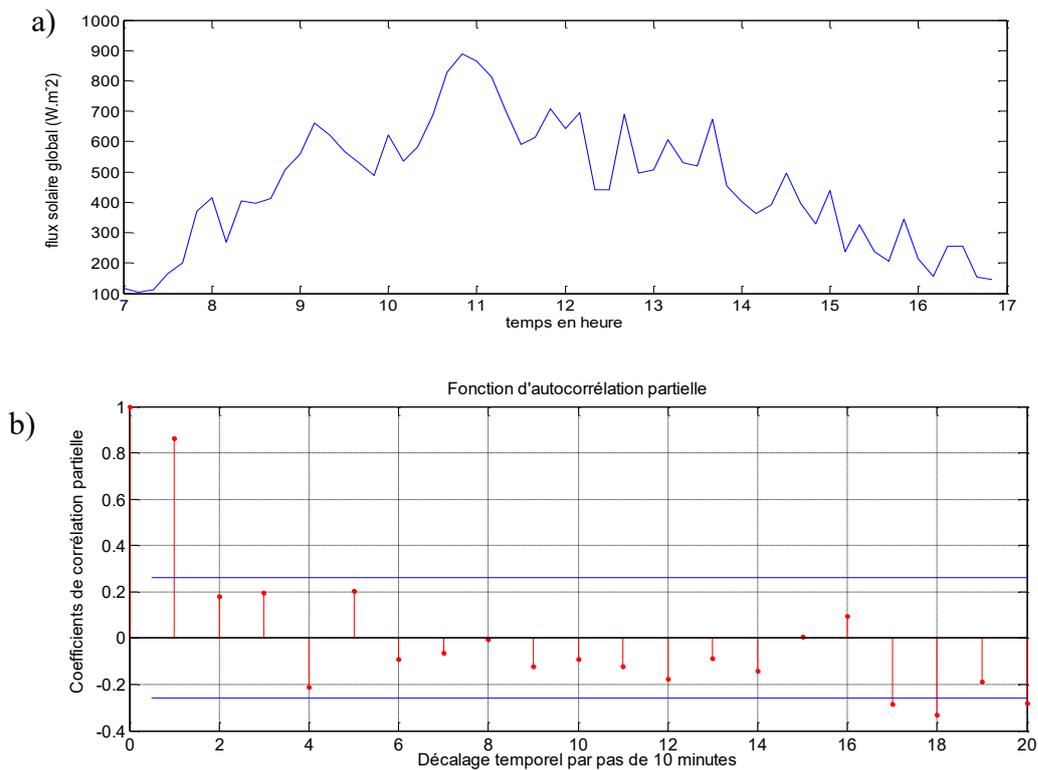


Figure 5-3. Exemple d'une série temporelle correspondant à une journée d'ensoleillement de 7h à 17h b) Fonction d'autocorrélation partielle correspondante à cette série temporelle.

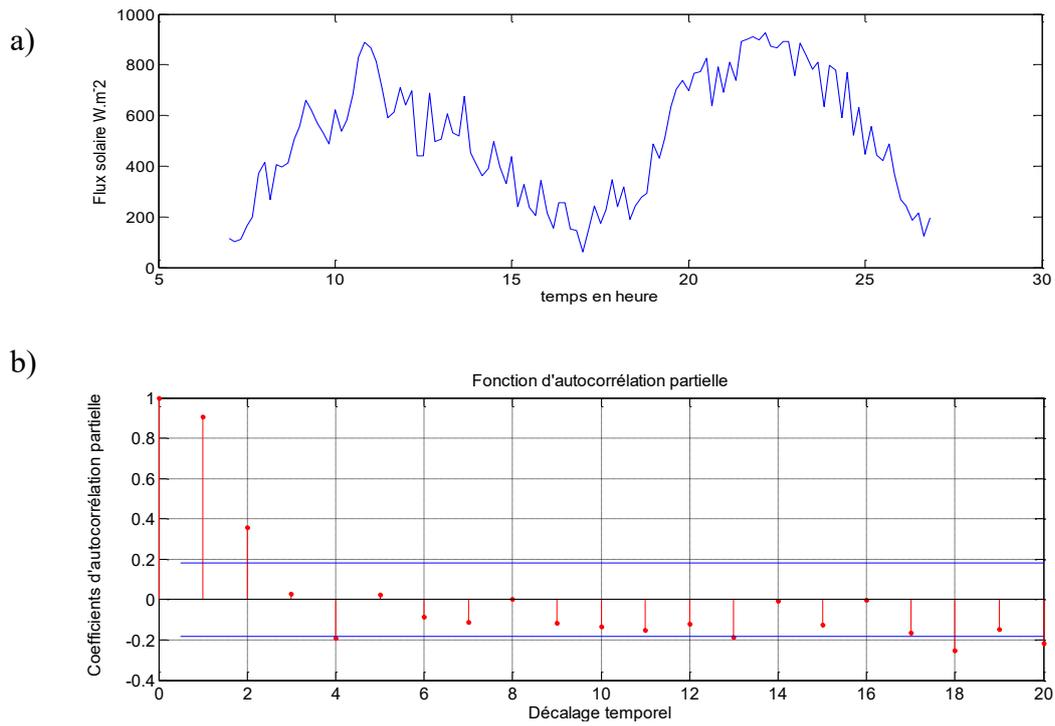
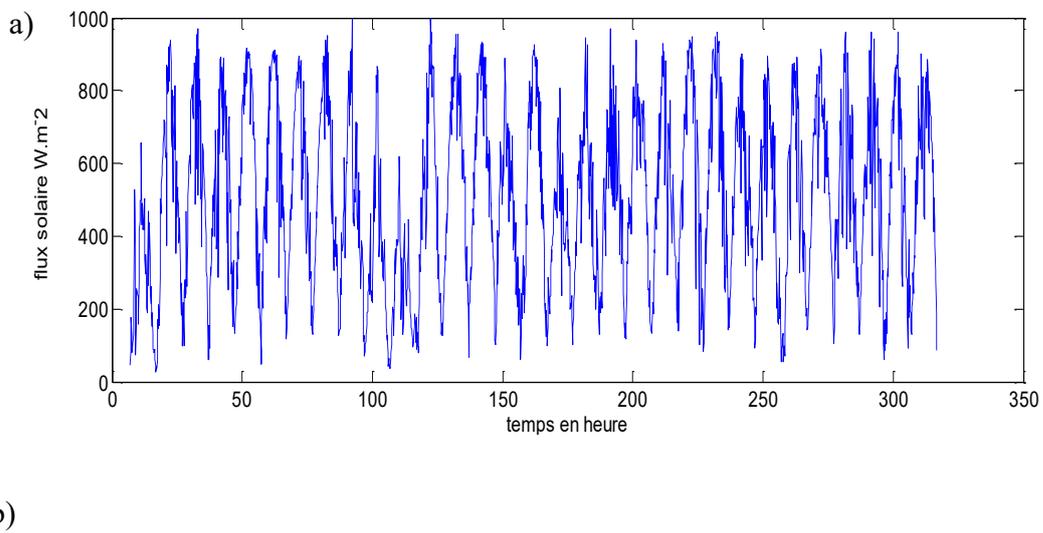


Figure 5-4. Exemple d'une série temporelle correspondant à deux journées d'ensoleillement de 7h à 17h. b) Fonction d'autocorrélation partielle pour deux journées d'ensoleillement global.



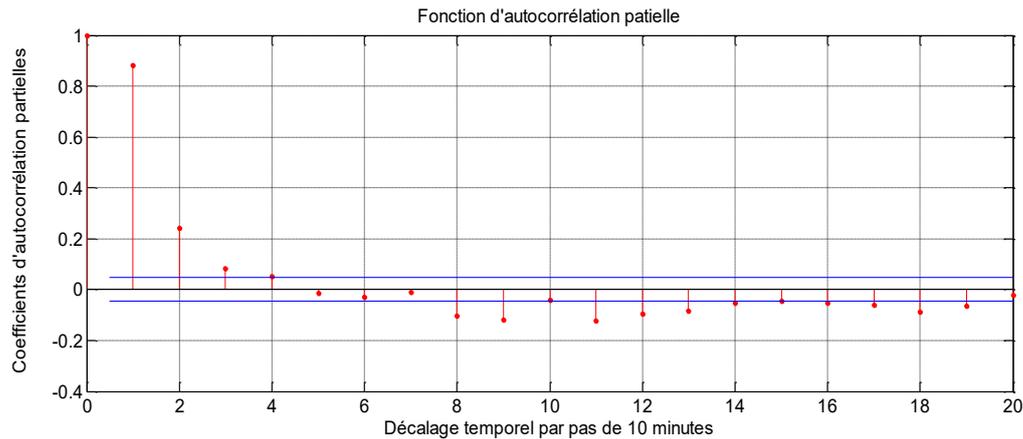


Figure 5-5. a) Exemple d'une série temporelle correspondant à un mois d'ensoleillement de 7h à 17h. b) Fonction d'autocorrélation partielle pour un mois d'ensoleillement global.

Il est donc plus judicieux d'appliquer le modèle prédictif pour une série de temps correspondant à une seule journée d'ensoleillement global. De plus, en exécutant le modèle de prédiction pour une journée, on évite d'ajouter la saisonnalité qui est un des éléments qui constituent la série temporelle de flux solaire global pour des périodes supérieures à la journée. Ce qui simplifiera la transformation de nos séries temporelles en séries stationnaires.

D'autre part, nous nous intéressons à la prévision pour la période d'ensoleillement du flux solaire global ce qui correspond aux séries de temps de flux solaire comprises entre 7h et 17h. Tomson et Tamm (2005) affirme que cette tranche horaire correspond à une période durant la quelle l'angle d'incidence est inférieur à 75° . Au-delà de 75° , l'efficacité optique du capteur approchera rapidement de la valeur nulle et le rayonnement solaire ne peut être converti efficacement.

Les autres tranches horaires ne sont pas intéressantes, d'une part d'un point de vue énergétique (les valeurs de flux solaires sont très faibles). D'autre part, leur prédiction est compliquée car les données au lever et coucher du soleil sont difficiles à rendre stationnaires (Voyant 2011). La matrice des données sur laquelle le modèle sera exécuté, se composera donc de trois vecteurs correspondant aux séries temporelles journalières de flux solaire de 7h à 17h des trois sites.

II.2.2 Stationnarisation spatio-temporelle

II.2.2.1 Analyse de la tendance temporelle

La série temporelle du flux solaire global comme toute série temporelle environnementale est généralement constituée de plusieurs éléments à savoir :

- une tendance : caractéristique d'une courbe en cloche due à l'évolution du flux solaire entre le lever et le coucher du soleil.

- une saisonnalité : variabilité saisonnière du flux solaire particulièrement caractérisée par des périodes trimestrielles dans l'année
- une composante stationnaire (ou résiduelle) : ce qui reste lorsque l'on a enlevé les autres composantes. Elle décrit l'évolution à court terme de la série.

Certaines séries temporelles se comportent selon un modèle multiplicatif. Par définition, une série temporelle (ou chronologique) est une suite d'observations :

x_1, x_2, \dots, x_n indexée par le temps. On supposera qu'il s'agit d'une réalisation d'un processus X , c'est à dire d'une suite $\{X_i\}$ de variables aléatoires.

On a alors :

$$X_t = T_t * S_t * Y_t \quad \text{pour tout } t \in \{1, \dots, n\} \quad (5.13)$$

avec T_t la tendance, S_t la composante saisonnière et Y_t la composante stationnaire.

Quand le modèle s'écrit comme une somme de plusieurs composantes on parle de modèle additif.

$$X_t = T_t + S_t + Y_t \quad \text{pour tout } t \in \{1, \dots, n\} \quad (5.14)$$

avec T_t la tendance, S_t la composante saisonnière et Y_t la composante stationnaire.

Le modèle le plus usuel consiste à supposer que la série initiale s'écrit sous la forme d'un modèle additif. L'application de notre modèle de prévision se basera effectivement, sur le fait que la série temporelle de l'ensoleillement global s'écrit sous la forme d'un modèle additif.

La série temporelle utilisée correspond à une journée de flux solaire ce qui nous autorise à retirer le terme correspondant à la saisonnalité.

On peut alors décrire la série temporelle de flux solaire comme variable étudiée du modèle de prédiction, sous forme :

$$X_t = T_t + Y_t \quad \text{pour tout } t \in \{1, \dots, n\} \quad (5.15)$$

avec T_t la tendance et Y_t la composante stationnaire.

L'analyse de la tendance pour un modèle additif peut utiliser deux méthodes : la moyenne glissante et la différenciation. Une autre approche populaire consiste à modéliser une tendance déterministe par

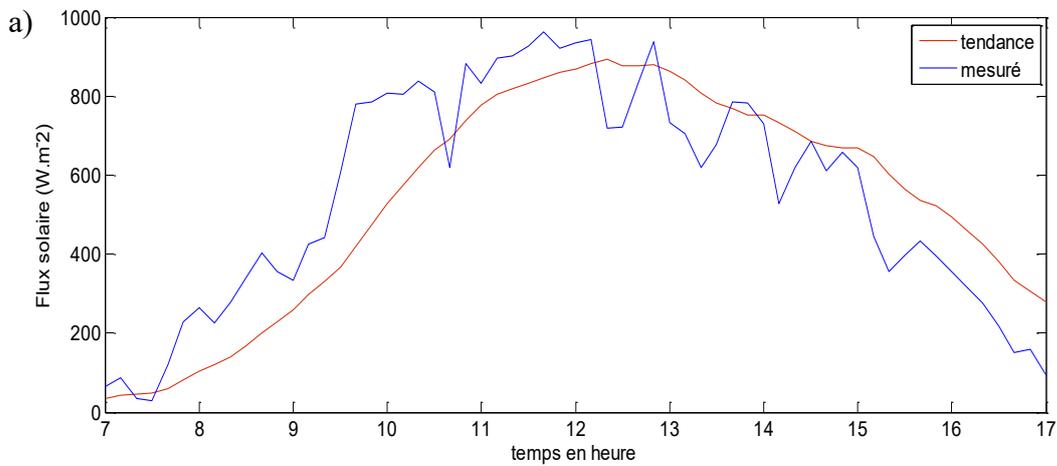
une somme pondérée de fonctions de base connues, où les pondérations sont estimées généralement par régression. Par exemple, des fonctions périodiques peuvent être utilisées pour rendre compte de la saisonnalité, et les polynômes peuvent être utilisés pour modéliser les variations lisses dans l'espace. D'autres discussions peuvent être trouvées dans l'article Kyriakidis et Journal (1999).

Méthode par moyenne glissante

La moyenne glissante a été définie préalablement dans le chapitre 4 tel que :

$$\bar{G}_N(t) = \frac{1}{N} \sum_{i=t-\frac{N-1}{2}}^{t+\frac{N-1}{2}} G(i) \quad (5.16)$$

Avec N fixé.



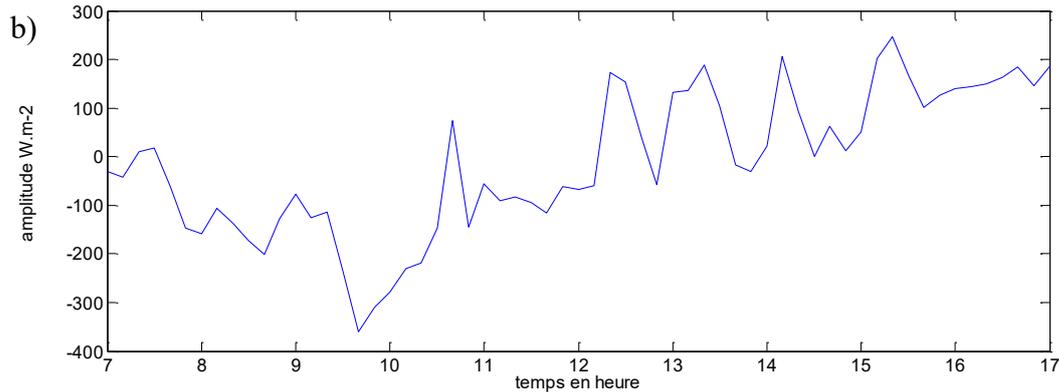


Figure 5-6. Exemple du 2 mars 2012 a) Signal mesuré du flux solaire global et son signal de moyenne glissante correspondante en rouge. b) Signal sans tendance obtenu par différence entre le signal mesuré et la moyenne glissante.

La limite de cette méthode est qu'on obtient souvent un décalage entre la courbe mesurée et la courbe de moyenne glissante, à l'image du graphique ci-dessus Fig.5-6; le lissage, en rouge, n'est pas centrée sur la courbe bleue des données observées. La moyenne glissante a été effectuée pour $N=10$ points sur les signaux de mesures par pas de 10 minutes.

Méthode par différenciation

Par définition, ∇ l'opérateur de différenciation, est tel que :

$$\nabla x_t = x_t - x_{t-1} \text{ pour tout } t \geq 2. \quad (5.17)$$

Propriétés

1. Soient u_t et v_t deux séries temporelles, $\nabla(u_t + v_t) = \nabla u_t + \nabla v_t$
2. Soit x_t une série temporelle et $\lambda \in \mathbb{R}$, $\nabla(\lambda x_t) = \lambda \nabla x_t$
3. Soit $y_t = a_0 + a_1 t + \dots + a_k t^k$, alors $\nabla^{(k)} y_t = k! a_k$ et $\nabla^{(k+1)} y_t = 0$.

La 3ème propriété implique que si $X_t = T_t + z_t$ avec T_t une tendance polynomiale, alors on peut supprimer la tendance en appliquant successivement plusieurs fois l'opérateur ∇ .

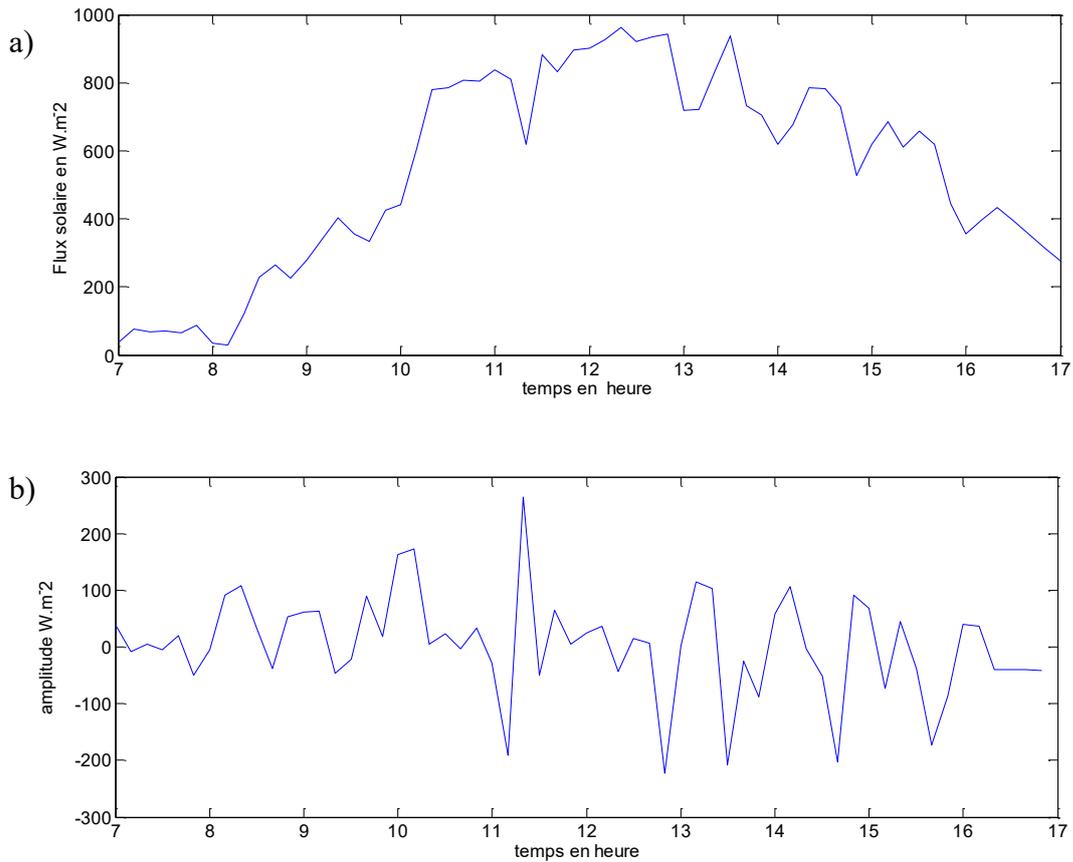


Figure 5- 7 Exemple du 2 mars 2012 a) Signal mesuré du flux solaire global b) Signal sans tendance obtenue par méthode des différenciations d'ordre 1.

Les régimes de variations observées sur la Figure 5-7 a) et b) sont sensiblement identiques. La variabilité du signal mesuré est respectée même après transformation des données. La limite de cette méthode est que les différences entre chaque valeur et la valeur précédente de la série sont calculées, à l'exception, de la valeur la plus ancienne. En d'autres termes, la série différenciée inclura une valeur de moins que la série d'origine.

Les fonctions d'autocorrélation des signaux obtenus sans tendance par les deux différentes méthodes montrent que seul le signal obtenu par différenciation d'ordre 1 est caractéristique d'un processus stationnaire (Figure 5-8 et Figure 5-9).

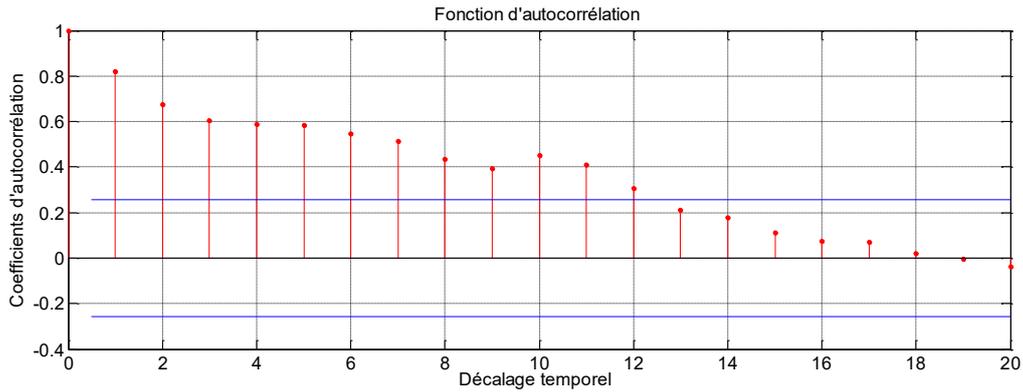


Figure 5-8. Fonction d'autocorrélation du signal sans tendance obtenu par méthode des moyennes glissantes à N=10 points.

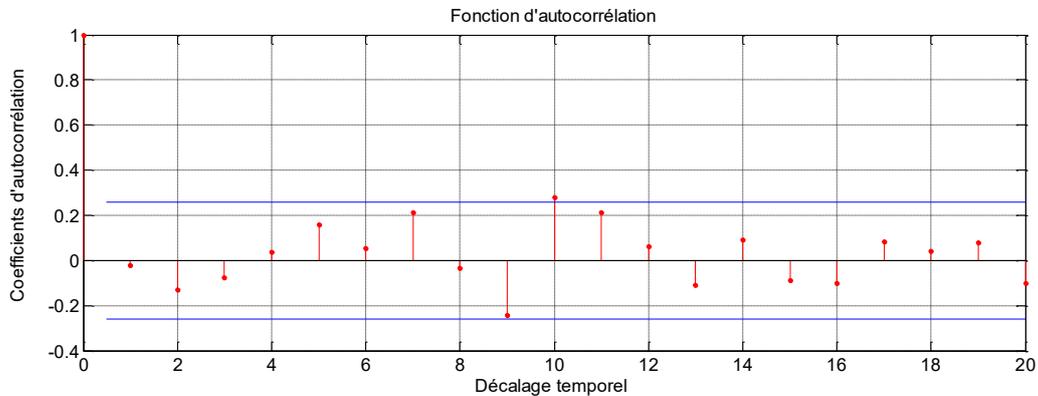


Figure 5-9. Fonction d'autocorrélation du signal sans tendance obtenu par méthode des différenciations d'ordre 1

Les tendances temporelles présentes dans les données peuvent être estimées et / ou enlevées de manière classique. Comme mentionné précédemment, dans les séries de temps univariées, il y a deux types de tendances communes : les tendances déterministes et les tendances stochastiques. On suppose que notre série initiale est assimilée à un modèle additif. Par conséquent, enlever la tendance déterministe par méthode de différenciation semble le plus approprié.

Un test de non stationnarité largement utilisé et répandu est le test de racine unitaire proposé par Dickey et Fuller en 1979. L'hypothèse nulle du test est la présence de racine unitaire, soit la non stationnarité de type stochastique. Le test consiste à tester :

$H_0 : \varphi = 1$ contre $H_1 : \varphi < 1$ dans le modèle :

$$Y_t = \varphi Y_{t-1} + \varepsilon_t \text{ avec } \varepsilon_t \text{ bruit blanc } (0, \sigma^2). \tag{5.18}$$

L'hypothèse nulle correspond au cas de marche aléatoire pure et l'hypothèse alternative correspond au cas d'un modèle AR(1) stationnaire (Lubrano M. 2008).

Nous avons effectué ce test à l'aide du code « adftest » sur Matlab. Les résultats montrent que le signal obtenu par méthode de différenciation est stationnaire alors que celui obtenu par moyenne glissante est non- stationnaire ce qui renforce le résultat obtenu par la fonction d'autocorrélation. La méthode de différenciation pour rendre notre variable étudiée stationnaire en dimension temporelle est alors vérifiée et justifiée.

II.2.2.2 Analyse de la tendance spatiale

Il est souvent possible de stabiliser les séries dans lesquelles la variance évolue dans le temps à l'aide d'une transformation par le logarithme naturel ou par la racine carrée. Ces transformations sont également appelées transformations fonctionnelles.

Etant donné la présence de fluctuations de forte amplitude et présentant des variations à très court terme et assez brutales, nous prenons le logarithme des observations, ce qui stabilise la variance.

Les données sans tendance temporelle sont alors définies telles que :

$$\nabla Z(s_i, t) = \log Z(s_i, t) - \log Z(s_i, t - 1) \quad (5.19)$$

Où $\nabla Z(s_i, t)$ est le processus stationnaire temporel par définition. Une fois la tendance temporelle retirée, nous estimons la tendance spatiale sur les données stationnaires temporellement. Dans la dimension spatiale, la différenciation a été également suggérée pour par exemple les fonctions aléatoires intrinsèques (Matheron 1973). La tendance spatiale supposée dépendante seulement de s_i est alors estimée par :

$$\beta(s_i) = E(\nabla Z(s_i, t) - \nabla Z(s_i, t - 1)) \quad (5.20)$$

Par conséquent, les données sans tendance spatio temporelle sont telles que :

$$\tilde{Z}(s_i, t) = \nabla Z(s_i, t) - \beta(s_i) \quad (5.21)$$

La tendance spatiale est réinjectée dans le résultat final de prédiction. La figure ci-dessous montre un exemple de journées (ciel nuageux) et sa transformation sans tendance spatio-temporelle.

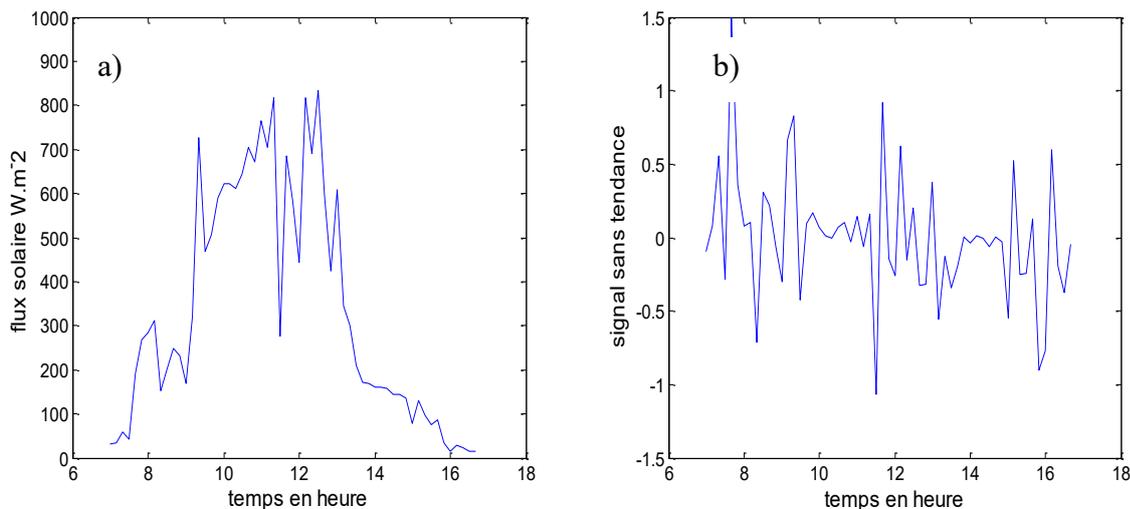


Figure 5-10. Signal de flux solaire mesuré en fonction du temps a) et son signal correspondant sans tendance spatio-temporelle b)

II.3. Sélection des prédicteurs

La sélection des variables explicatives ou des stations est basée sur la même stratégie algorithmique décrite en section I.3. Cette stratégie de sélection sera utilisée en alternative par les critères d'information dans notre modèle. La finalité de ce processus est de trouver le modèle optimal qui minimise le critère AIC. Nous adoptons donc une démarche backward, qui consiste, à partir du modèle complet comportant toutes les variables, à éliminer unes à unes les variables qui permettent de diminuer l'AIC, et de continuer ainsi tant que la suppression d'une variable améliore le critère.

Avec le critère Akaike (AIC), nous cherchons la régression qui minimise la quantité suivante :

$$AIC(p) = \log[\det(\widehat{\Omega})] + \frac{2n^2p}{T} \quad (5.22)$$

Avec le critère BIC de Schwartz, on cherche à optimiser :

$$BIC = \log[\det(\widehat{\Omega})] + \frac{\log T}{T} n^2p \quad (5.23)$$

p est le nombre de retards et $\widehat{\Omega}$ l'estimateur de la matrice de covariance des résidus dans le cadre des moindres carrés, n est le nombre de variables (nombre de stations), T est le nombre d'observations dont dispose la variable.

Dans le cadre vectoriel (adapté à notre cas d'étude puisque le processus du modèle traite chaque site séparément) les équations 15 et 16 deviennent :

$$AIC = 2k - 2\ln(L) \quad (5.24)$$

Avec le critère BIC de Schwartz, on cherche à optimiser :

$$BIC = -2\ln(L) + \ln(T)k \quad (5.25)$$

avec T le nombre d'observations dans l'échantillon étudié et k le nombre de paramètres, L est le maximum de la fonction de vraisemblance du modèle.

Nous introduisant par conséquent un modèle construisant une stratégie pour identifier les valeurs de la matrice des paramètres R_i de l'équation 5.1. Pour chaque prévision sur une station s , les prédicteurs disponibles sont les valeurs retardées de toutes les stations. Ceci pourrait être un modèle complexe si toutes les valeurs retardées de toutes les stations doivent être considérés comme prédicteurs potentiels. Le fait que les prédicteurs peuvent être entrés séquentiellement simplifie les étapes de construction du modèle. De plus, la suppression de prédicteurs inintéressants à chaque décalage temporel améliore l'efficacité de prévision en évitant l'estimation des coefficients nuls.

II.4. Estimation des paramètres du modèle

Rappelons l'équation du système du modèle spatio-temporel VAR déjà mentionné en section I.

Soient les données spatio-temporelles $Z(s_i, t)$, les observations d'un processus stochastique dans $\mathbb{R}^2 \times \mathbb{R}$. Chaque observation est localisée à une station $s_i = 1, \dots, n$ au temps $t = 1, \dots, T$. Le modèle de prédiction VAR (vecteur autorégressif) est tel que :

$$Z_t - \beta = \sum_{i=1}^p R_i (Z_{t-i} - \beta) + \varepsilon_t \quad (5.1)$$

où $Z_t = (Z(s_1, t), Z(s_2, t), \dots, Z(s_n, t))'$, $\varepsilon_t = (\varepsilon_t(s_1), \dots, \varepsilon_t(s_n))'$ est un bruit blanc avec $E(\varepsilon_t) = 0$, $E(\varepsilon_t, \varepsilon_u) = 0$ pour $u \neq t$, $E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon$, $\beta = (\beta(s_1), \dots, \beta(s_n))'$ est la tendance spatiale,

R_i sont les matrices $n \times n$ de paramètres inconnus avec $i = 1, \dots, p$. Les n lignes de ces matrices correspondent aux n localisations où les séries de temps sont observées. Les paramètres des matrices R_i auront une structure spécifique dépendant de la structure spatiale des données imposée. Si la structure spatiale des stations dépend des distances dans l'ordre croissant, par exemple pour un p -ordre du modèle VAR (p) égale à 1, la matrice sera telle que :

$$R_1 = \begin{matrix} Pc_t \\ F_t \\ Des \end{matrix} \begin{bmatrix} Pc_{t-1} & F_{t-1} & Des_{t-1} \\ F_{t-1} & Pc_{t-1} & Des_{t-1} \\ Des_{t-1} & Pc_{t-1} & Des_{t-1} \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}$$

Pc_{t-1} ; F_{t-1} ; Des_{t-1} sont les coefficients des paramètres R_{ij} associés aux séries temporelles respectivement de Petit-canal, Fouillole et La Désirade au temps $t-1$ composant la combinaison linéaire pour la prédiction de la série temporelle du site de référence (nommé à gauche de la matrice : Pc_t ; F_t ; Des_t , respectivement les séries temporelles de Petit-canal, Fouillole et La Désirade à l'instant t .

L'évaluation des paramètres peut être effectuée pour toutes les stations simultanément. Néanmoins, notre modèle est construit de manière à effectuer la prédiction pour chaque station séparément afin d'éviter toute hypothèse de spatio-stationnarité (de Luna et Genton 2005).

II.4.1 Propriétés des paramètres de régression linéaire multiple

Le système dynamique déterministe défini par l'équation (1) peut être caractérisé comme mentionné préalablement en section I.2.

Nous vérifions alors la stabilité de notre système en calculant les valeurs propres des paramètres R_i . Nous retenons les valeurs propres maximales des paramètres en valeur absolue. Les résultats obtenus pour 31 prédictions soit 31 matrices R_i (3×3) sont représentés par la densité de probabilité afin d'analyser la dispersion des résultats de valeurs propres maximales et leur fréquence (Figure 5-11).

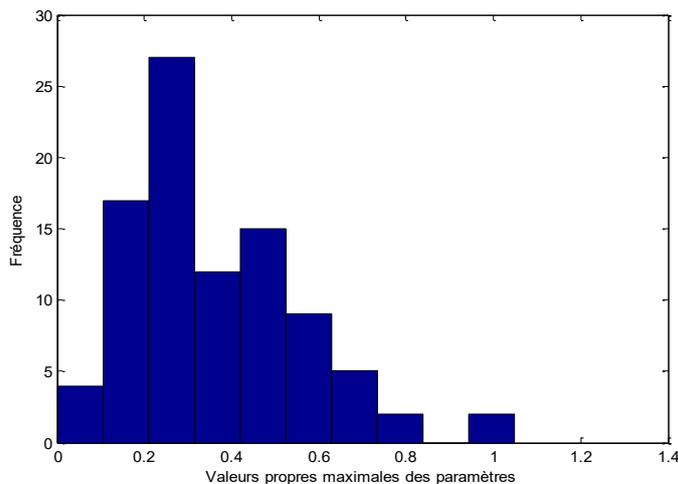


Figure 5-11. Fréquence des valeurs propres maximales en valeur absolue de la matrice des paramètres R_i calculées sur un mois soient 31 matrices

On constate que toutes les valeurs propres maximales en valeur absolue sont inférieures à 1 seules deux valeurs sont égales à 1. On peut estimer que notre système dynamique régit par l'équation (1) est stable selon la propriété de stabilité.

La stabilité est importante car elle implique une stationnarité temporelle du processus stochastique (Lütkepohl (1991), Hamilton (1994), Pena et al. (2001)).

III. Analyse de la performance prédictive

III.1. Test sur la sélection de l'ordre p

Tel que dans les modèles $AR(p)$, la longueur du décalage temporel peut être déterminé par les critères d'informations notamment : le critère d'information Akaike (AIC) (Akaike 1974), le critère Hannan et Quinn (HQC) (Hannan et Quinn 1979) ou le critère d'information Bayésien (Scharwz et al (1978)).

Les critères AIC et BIC sont décrits respectivement par les équations 5.24 et 5.25 préalablement définies en section II.3.3. Le critère HQC est défini par l'équation suivante :

$$HQC = \log[\det(\hat{\Omega})] + \frac{\log(\log(T))}{T} 2n^2p \quad (5.26)$$

p est le nombre de retards, $\hat{\Omega}$ l'estimateur de la variance des résidus dans le cadre des moindres carrés, T est le nombre d'observations dont dispose la variable et n le nombre de variables.

Nous avons effectués des tests pour ces trois critères sur 366 jours pour des décalages temporels de 10 min à 1h, soient pour $p=1, \dots, 6$. Les résidus sont calculés par la différence entre le flux solaire estimé (obtenu à partir des 15 points d'initialisation et des coefficients des paramètres estimés) et le flux solaire mesuré. Les critères d'informations préalablement cités sont ainsi calculés à partir des résultats des résidus journaliers suivant les équations 5.24, 5.25, et 5.26. Chaque critère d'information est minimisé à un p -ordre pour chaque journée et sélectionne ainsi l'ordre p du modèle VAR. La Figure 5-12 représente les histogrammes des fréquences en pourcentage de l'ordre p qui a été sélectionné pour chaque jour par les critères d'informations sur un ensemble de 366 journées. On constate que c'est plus souvent pour l'ordre $p=1$ (environ 34%) que les critères d'informations sont minimisés. La fréquence à laquelle l'ordre $p=5$ et $p=6$ sont sélectionnés par les critères d'informations est très faible (moins de 7%). Nous évaluerons donc la performance du modèle VAR que pour l'ordre p égale de 1 à 4 en section III.2.

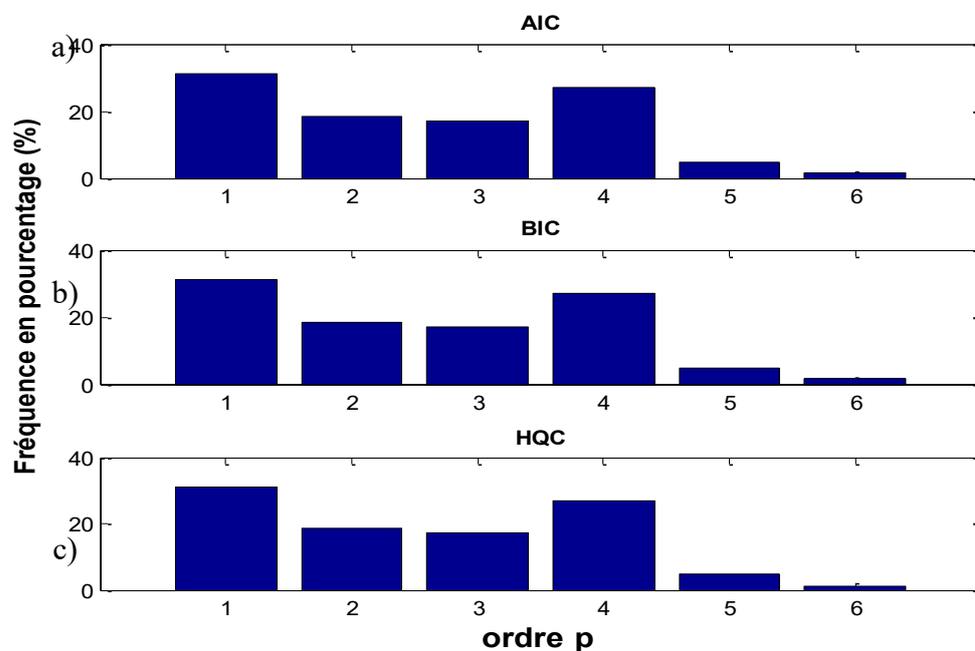


Figure 5-12. Fréquence en pourcentage (%) des p -ordres pour $p = 1, \dots, 6$ sélectionnés par le critère a) AIC ,b) BIC et c) HQC

III.2. Analyse de la performance prédictive pour deux structures spatiales

Nous avons appliqué le modèle VAR pour différents ordres p allant de 1 à 4 soit un temps de retard égal à 10 minutes jusqu'à 40 minutes pour nos trois années 2010, 2011 et 2012. Cette analyse permet d'évaluer la performance prédictive du modèle VAR à l'aide de l'erreur statistique RMSE normalisée (nRMSE) pour chaque p -ordre et de déterminer ainsi l'ordre p approprié du modèle VAR(p) qui minimisera le mieux la nRMSE. Une évaluation de deux schémas structurels spatiaux des localisations a été effectuée par comparaison des nRMSE. La première, pour les stations ordonnées selon la direction prédominante du vent, la deuxième ordonnant les stations selon les distances métriques dans l'ordre croissant. Pour la première structure spatiale, nous prenons en compte que les vents prédominants, les Alizés, viennent de l'Est de l'île de la Guadeloupe, on a ainsi l'ordre des stations suivant : La Désirade (site 1 [s1]), Petit-canal(site 2 :[s2]), et Fouillole (site 3 :[s3]). La deuxième structure spatiale selon les distances géographiques ascendantes ordonne les stations tel que :Petit-canal (site1 : [s1]), Fouillole (site 2 :[s2]), La Désirade (site 3 : [s3]). La Figure 5-13, schématise l'ordre des stations imposé par les deux structures spatiales tenant compte d'une information physique (direction du vent) et l'autre d'une information géographique (distance entre les sites). Ces structures spatiales détermineront la paramétrisation de la matrice des distances et celle de la matrice des vecteurs de données de flux solaire.

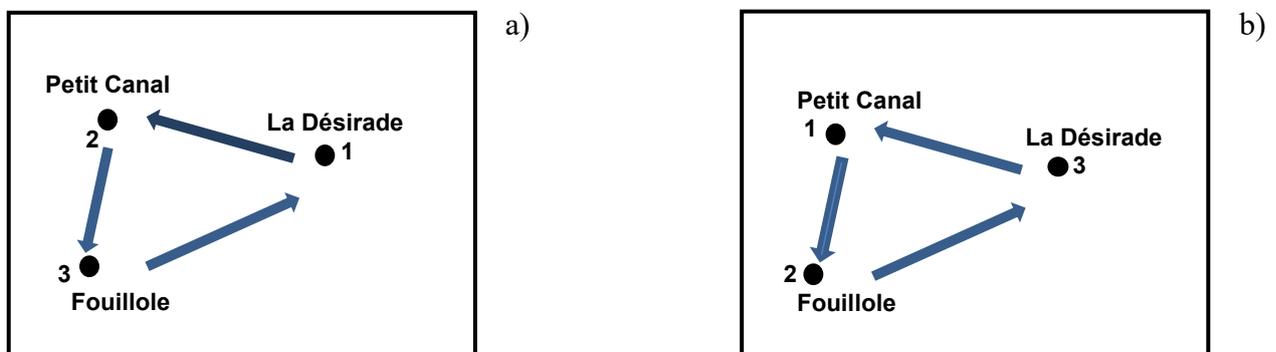


Figure 5-13. Schéma simplifié des deux différentes structures spatiales (organisation spatiale) des stations : a) selon une information physique : la direction du vent b) selon les distances métriques ascendantes.

Les prédictions ont été appliquées sur les données de mesures par pas de 10 minutes sur une journée de 7h à 17h pour les années 2010 à 2012 avec les 15 premiers points étant utilisés pour l'initialisation du modèle (sachant que le nombre d'observations doit être supérieur au nombre de variables). Ce qui revient à 45232 points prédits au total. Les RMSE mensuelles sont normalisées par la différence entre la valeur maximale et la valeur minimale des données observées du mois donné:

$$nRMSE = \frac{\sqrt{\frac{1}{n} \sum (\hat{G} - G)^2}}{(\max(G) - \min(G))} \quad (5.27)$$

Avec nRMSE la RMSE normalisée, \hat{G} les valeurs prédites et G les valeurs observées où n est le nombre de points des données.

Les résultats trouvés sont présentés par les tableaux suivants (Tableau 5-1 à Tableau 5-9) pour la prédiction du flux solaire pour chaque site.

	Direction Est-Ouest				Distance ascendante			
	p=1	p=2	p=3	p=4	p=1	p=2	p=3	p=4
Jan	0.13	0.15	0.16	0,20	0.13	0.23	0.23	0.20
Fev	0.11	0.13	0.15	0.18	0.11	0.13	0.15	0.18
Mar	0.12	0.16	0.18	0.20	0.12	0.16	0.18	0.20
Avr	0.13	0.14	0.16	0.47	0.12	0.14	0.16	0.47
Mai	0.22	0.33	0.36	0.49	0.22	0.34	0.36	0.49
Juin	0.12	0.16	0.19	0.22	0.12	0.16	0.19	0.22
Juil	–	–	–	–	–	–	–	–
Août	0.10	0.19	0.25	0.45	0.10	0.19	0.26	0.44
Sept	0.12	0.16	0.20	0.22	0.12	0.18	0.20	0.22
Oct	0.13	0.17	0.20	0.22	0.13	0.18	0.20	0.22
Nov	0.13	0.18	0.19	0.27	0.13	0.18	0.19	0.27
Dec	0.29	0.29	0.29	0.34	0.29	0.29	0.29	0.34
Moyenne annuelle	0.14	0.21	0.21	0.32	0.14	0.21	0.22	0.32

Tableau 5-1 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d'ordres des stations pour la prédiction du flux solaire sur la Désirade en 2010

	Direction Est-Ouest				Distance ascendante			
	p=1	p=2	p=3	p=4	p=1	p=2	p=3	p=4
Jan	0.15	0.24	0.20	0,25	0.15	0.20	0.22	0.25
Fev	0.13	0.18	0.21	0.28	0.14	0.18	0.21	0.28
Mar	0.12	0.24	0.29	0.20	0.13	0.14	0.17	0.37
Avr	0.09	0.17	0.23	0.24	0.09	0.12	0.17	0.36
Mai	0.13	0.15	0.20	0.39	0.13	0.14	0.20	0.39
Juin	0.11	0.17	0.19	0.22	0.11	0.15	0.19	0.28
Juil	0.17	0.28	0.19	0.23	0.13	0.17	0.19	0.30
Août	0.09	0.20	0.23	0.27	0.10	0.17	0.23	0.34
Sept	0.14	0.17	0.19	0.23	0.14	0.17	0.19	0.23
Oct	0.13	0.17	0.21	0.31	0.13	0.17	0.21	0.41
Nov	0.17	0.20	0.23	0.25	0.14	0.18	0.20	0.25
Dec	0.17	0.20	0.29	0.35	0.17	0.20	0.29	0.35
Moyenne annuelle	0.13	0.20	0.22	0.27	0.13	0.17	0.21	0.32

Tableau 5-2 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d'ordres des stations pour la prédiction du flux solaire sur la Désirade en 2011

	Direction Est-Ouest				Distance ascendante			
	p=1	p=2	p=3	p=4	p=1	p=2	p=3	p=4
Jan	0.17	0.23	0.26	0,30	0.17	0.23	0.33	0.30
Fev	0.15	0.19	0.22	0.41	0.15	0.19	0.22	0.25
Mar	0.12	0.15	0.17	0.18	0.12	0.15	0.17	0.19
Avr	0.11	0.15	0.17	0.22	0.11	0.15	0.17	0.22
Mai	0.12	0.15	0.20	0.40	0.12	0.15	0.20	0.47
Juin	0.08	0.11	0.14	0.15	0.09	0.11	0.14	0.15
Juil	0.10	0.12	0.16	0.16	0.10	0.12	0.14	0.16
Août	0.11	0.15	0.17	0.23	0.11	0.15	0.17	0.23
Sept	0.13	0.13	0.16	0.21	0.13	0.15	0.16	0.21
Oct	0.12	0.16	0.40	0.30	0.12	0.16	0.60	0.30
Nov	0.13	0.16	0.20	0.25	0.34	0.16	0.64	0.26
Dec	0.16	0.22	0.30	0.46	0.10	0.22	0.30	0.46
Moyenne annuelle	0.15	0.20	0.26	0.33	0.18	0.19	0.30	0.34

Tableau 5-3 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d'ordres des stations pour la prédiction du flux solaire sur la Désirade en 2012

	Direction Est-Ouest				Distance ascendante			
	p=1	p=2	p=3	p=4	p=1	p=2	p=3	p=4
Jan	0.13	0.13	0.17	0.19	0.11	0.19	0.23	0.23
Fev	0.17	0.23	0.21	0.21	0.17	0.23	0.21	0.21
Mar	0.12	0.15	0.17	0.19	0.12	0.15	0.17	0.22
Avr	0.11	0.15	0.17	0.21	0.11	0.15	0.17	0.21
Mai	0.10	0.14	0.16	0.48	0.10	0.14	0.17	0.48
Juin	0.11	0.31	0.35	0.24	0.11	0.31	0.35	0.24
Juil	-	-	-	-	-	-	-	-
Août	0.15	0.20	0.24	0.40	0.15	0.20	0.24	0.40
Sept	0.13	0.16	0.18	0.20	0.13	0.16	0.18	0.20
Oct	0.11	0.14	0.18	0.20	0.11	0.14	0.17	0.20
Nov	0.13	0.18	0.23	0.29	0.13	0.18	0.23	0.30
Dec	0.15	0.18	0.22	0.26	0.15	0.18	0.22	0.26
Moyenne annuelle	0.13	0.18	0.21	0.26	0.13	0.18	0.21	0.27

Tableau 5-4 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d'ordres des stations pour la prédiction du flux solaire sur Petit-canal en 2010

	Direction Est-Ouest				Distance ascendante			
	p=1	p=2	p=3	p=4	p=1	p=2	p=3	p=4
Jan	0.14	0.19	0.21	0.25	0.15	0.19	0.20	0.20
Fev	0.14	0.17	0.20	0.24	0.14	0.16	0.21	0.24
Mar	0.11	0.13	0.14	0.18	0.11	0.13	0.15	0.32
Avr	0.10	0.13	0.15	0.20	0.10	0.13	0.14	0.40
Mai	0.11	0.15	0.17	0.48	0.11	0.13	0.17	0.41
Juin	0.09	0.12	0.19	0.20	0.09	0.12	0.18	0.47
Juil	0.10	0.14	0.17	0.19	0.11	0.14	0.19	0.48
Août	0.11	0.30	0.20	0.33	0.12	0.15	0.19	0.60
Sept	0.12	0.17	0.19	0.29	0.12	0.16	0.24	0.41
Oct	0.14	0.16	0.17	0.43	0.13	0.16	0.17	0.40
Nov	0.20	0.21	0.35	0.30	0.15	0.17	0.43	0.49
Dec	0.16	0.20	0.25	0.45	0.16	0.20	0.27	0.36
Moyenne annuelle	0.13	0.17	0.20	0.30	0.15	0.15	0.23	0.40

Tableau 5-5 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d'ordres des stations pour la prédiction du flux solaire sur Petit-canal en 2011

	Direction Est-Ouest				Distance ascendante			
	p=1	p=2	p=3	p=4	p=1	p=2	p=3	p=4
Jan	0.14	0.18	0.20	0.27	0.14	0.18	0.21	0.28
Fev	0.13	0.17	0.19	0.49	0.13	0.17	0.19	0.24
Mar	0.13	0.16	0.19	0.26	0.13	0.16	0.19	0.25
Avr	0.11	0.15	0.18	0.22	0.11	0.15	0.18	0.22
Mai	0.11	0.14	0.17	0.21	0.11	0.16	0.17	0.21
Juin	0.07	0.09	0.10	0.14	0.07	0.09	0.10	0.14
Juil	0.10	0.14	0.15	0.16	0.10	0.14	0.15	0.17
Août	0.11	0.14	0.17	0.23	0.12	0.18	0.27	0.29
Sept	0.09	0.20	0.24	0.39	0.09	0.19	0.22	0.39
Oct	0.16	0.23	0.27	0.35	0.17	0.21	0.26	0.42
Nov	0.13	0.17	0.20	0.25	0.48	0.17	0.44	0.27
Dec	0.16	0.19	0.21	0.31	0.15	0.19	0.23	0.32
Moyenne annuelle	0.12	0.16	0.19	0.27	0.15	0.19	0.22	0.26

Tableau 5-6 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d'ordres des stations pour la prédiction du flux solaire sur Petit-canal en 2012

	Direction Est Ouest				Distance ascendante			
	p=1	p=2	p=3	p=4	p=1	p=2	p=3	p=4
Jan	0.13	0.19	0.25	0.25	0.13	0.31	0.33	0.39
Fev	0.12	0.14	0.39	0.45	0.12	0.14	0.49	0.50
Mar	0.13	0.16	0.20	0.27	0.13	0.15	0.20	0.29
Avr	0.15	0.19	0.23	0.28	0.15	0.19	0.23	0.28
Mai	0.13	0.17	0.22	0.29	0.13	0.17	0.22	0.28
Juin	0.12	0.16	0.22	0.29	0.12	0.16	0.22	0.29
Juil	-	-	-	-	-	-	-	-
Août	0.14	0.20	0.27	0.34	0.14	0.20	0.27	0.34
Sept	0.14	0.20	0.34	0.39	0.14	0.20	0.44	0.39
Oct	0.15	0.24	0.30	0.45	0.15	0.24	0.37	0.60
Nov	0.14	0.18	0.31	0.48	0.14	0.18	0.31	0.39
Dec	0.31	0.33	0.35	0.49	0.31	0.33	0.35	0.57
Moyenne annuelle	0.15	0.20	0.28	0.36	0.15	0.21	0.31	0.39

Tableau 5-7 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d'ordres des stations pour la prédiction du flux solaire sur Fouillole en 2010

	Direction Est-Ouest				Distance ascendante			
	p=1	p=2	p=3	p=4	p=1	p=2	p=3	p=4
Jan	0.16	0.21	0.24	0.47	0.16	0.20	0.23	0.45
Fev	0.18	0.22	0.28	0.36	0.18	0.23	0.29	0.37
Mar	0.18	0.21	0.32	0.46	0.20	0.25	0.35	0.47
Avr	0.13	0.20	0.23	0.29	0.13	0.19	0.23	0.39
Mai	0.11	0.14	0.19	0.23	0.12	0.12	0.21	0.33
Juin	0.13	0.17	0.23	0.26	0.13	0.17	0.22	0.33
Juil	0.12	0.16	0.23	0.32	0.13	0.17	0.25	0.49
Août	0.13	0.34	0.24	0.31	0.13	0.18	0.23	0.39
Sept	0.11	0.15	0.22	0.36	0.11	0.15	0.20	0.30
Oct	0.17	0.24	0.29	0.43	0.17	0.23	0.29	0.40
Nov	0.22	0.25	0.28	0.53	0.16	0.20	0.23	0.30
Dec	0.21	0.31	0.40	0.56	0.21	0.30	0.40	0.50
Moyenne annuelle	0.15	0.22	0.26	0.38	0.15	0.20	0.26	0.39

Tableau 5-8 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d'ordres des stations pour la prédiction du flux solaire sur Fouillole en 2011

	Direction Est-Ouest				Distance ascendante			
	p=1	p=2	p=3	p=4	p=1	p=2	p=3	p=4
Jan	0.19	0.26	0.34	0.43	0.20	0.26	0.35	0.45
Fev	0.19	0.27	0.35	0.50	0.19	0.27	0.34	0.46
Mar	0.15	0.20	0.23	0.29	0.14	0.19	0.23	0.39
Avr	0.14	0.18	0.22	0.26	0.14	0.18	0.22	0.26
Mai	0.11	0.17	0.20	0.30	0.11	0.15	0.18	0.28
Juin	0.09	0.11	0.14	0.22	0.09	0.11	0.13	0.24
Juil	0.14	0.17	0.21	0.26	0.13	0.17	0.21	0.30
Août	0.17	0.21	0.24	0.30	0.16	0.20	0.39	0.34
Sept	0.13	0.17	0.22	0.30	0.14	0.17	0.22	0.28
Oct	0.15	0.26	0.36	0.41	0.15	0.19	0.36	0.29
Nov	0.14	0.21	0.32	0.36	0.41	0.21	0.38	0.38
Dec	0.18	0.25	0.31	0.40	0.17	0.25	0.31	0.39
Moyenne annuelle	0.15	0.20	0.26	0.33	0.17	0.19	0.28	0.34

Tableau 5-9 RMSE normalisée (nRMSE) pour différents ordres p et deux différents schémas d'ordres des stations pour la prédiction du flux solaire sur Fouillole en 2012

Nous pouvons noter que les valeurs de nRMSE sont minimales (entre 13% et 16% Tableau 5-1 à Tableau 5-9) pour un p-ordre égale à 1 soit un décalage temporel de 10 minutes des séries de temps passées. Sélectionner un modèle VAR approprié pour une séquence de données requiert la sélection du p-ordre du modèle. Selon les résultats trouvés, il s'agit de p=1 soit le modèle spatio temporel VAR(1). Par ailleurs, nous pouvons noter que les valeurs de nRMSE pour les différents types d'ordres spatiaux des stations semblent sensiblement égales mais une légère différence est remarquable : la performance prédictive du modèle est plus efficace pour les stations ordonnées selon la direction prédominante du vent (direction Est-Ouest). Le tableau récapitulatif moyennant les nRMSE sur les trois années pour le modèle VAR(1), nous le montre (Tableau 5-10). Les nRMSE montrent une très bonne performance prédictive avec en moyenne un pourcentage de 14%. Cette performance est particulièrement valable pour les prédictions de flux solaire pour le site de Petit-canal qui indique de meilleurs résultats que les autres sites (Tableau.5-4 à Tableau.5-6). Ce site situé sur la côte Est est très ensoleillé et assez proche des deux autres sites, comparé à La Désirade et Fouillole qui sont plus éloignés entre eux. Suite à l'analyse de ces tableaux de résultats, un principal commentaire peut être donné : le modèle adéquat est le modèle spatio-temporel VAR(1), qui prend en compte la direction du vent prédominant dans le processus de structure spatiale des stations (matrice des distances et des vecteurs de données), donnant de bonne voire une meilleure performance prédictive que le modèle ignorant cette information physique (ordre selon les distances). On peut le constater sur la Figure 5-14 qui montrent des valeurs prédites du modèle prenant en compte la direction du vent (courbe en gris), plus proches des valeurs observées que celles obtenues par le modèle ignorant ce paramètre physique (courbe en pointillé).

Tableau récapitulatif (nRMSE)			
	La Désirade	Petit-canal	Fouillole
distance	0,15	0,14	0,16
Vent(direction Est Ouest)	0,14	0,13	0.15

Tableau 5-10 Tableau récapitulatif des résultats sur les trois années de nRMSE à un temps de retard *il* moyennée sur trois années.

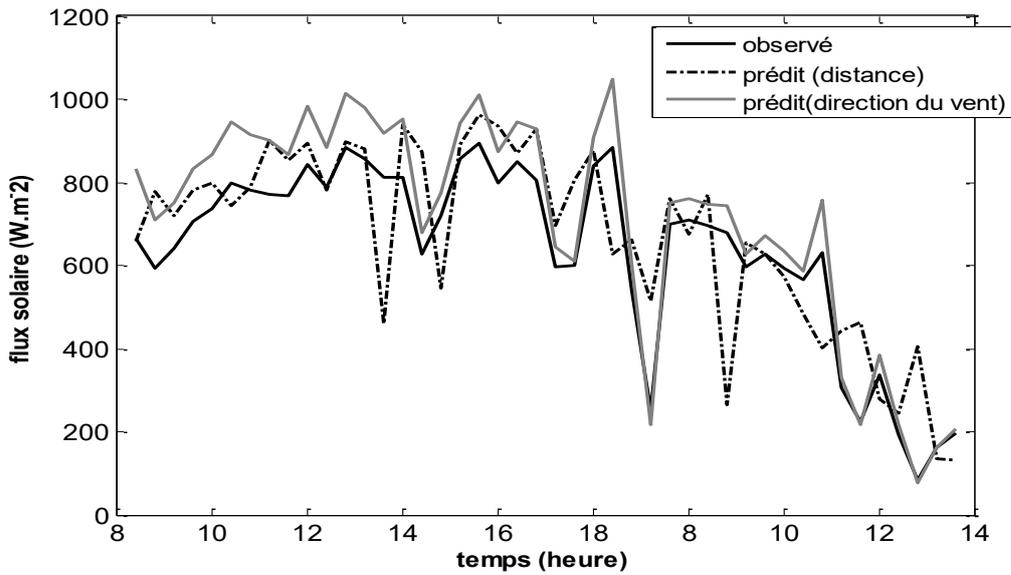


Figure 5-14. Exemple de prévisions à 10 minutes pour une journée (1er novembre) incluant deux différentes structures spatiales dans le processus du modèle (ordre des localisations selon la direction prédominante Est-Ouest des Alizés en gris et l'ordre des localisations selon les distances ascendantes en pointillé). Nous pouvons observer que les valeurs prédites en gris sont plus proches des valeurs mesurées que le signal en pointillé. De plus, le signal de couleur gris respecte bien la dynamique du signal mesuré. Les 15 premiers points sont utilisés pour l'initialisation du modèle soit 1h30 de données.

III.3. Analyse de la performance prédictive en fonction de la variabilité du flux solaire

Afin de mettre en évidence l'influence de la variabilité de l'ensoleillement sur les résultats de prévision nous avons appliqué le modèle sur une classe donnée de journées. Cette classification de l'ensoleillement global journalier est effectuée par un algorithme basée sur la méthode k-means (voir chapitre 3). Les classes obtenues ont les mêmes caractéristiques que celles présentées dans Soubdhan et al (2009) qui ont utilisé une méthode basée sur les distributions de Dirichlet : journées à ciel clair à journées à ciel nuageux. Les valeurs de MAE et MBE mensuelles sont normalisées par la différence entre la valeur maximale et la valeur minimale des données observées du mois donné ((respectivement nMAE et nMBE équation 5.28 et 5.29). Les résultats des erreurs statistiques pour les trois années de mesures, nMAE et nMBE sont présentés par des histogrammes des valeurs de ces erreurs statistiques pour chaque classe (Figures 5-15 à 5-20).

$$nMAE = \frac{\frac{1}{n} \sum |\hat{G} - G|}{(\max(G) - \min(G))} \quad (5.28)$$

$$nMBE = \frac{\frac{1}{n} \sum \hat{G} - G}{(\max(G) - \min(G))} \quad (5.29)$$

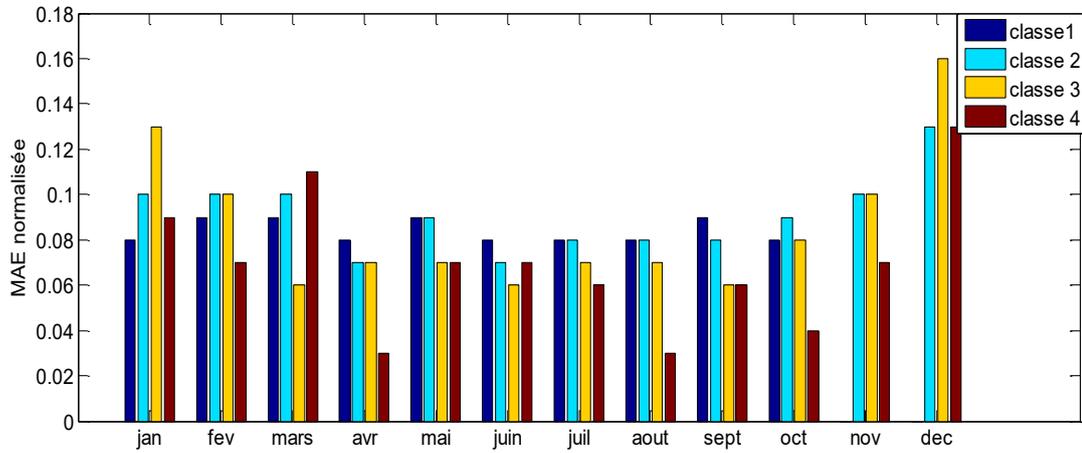


Figure 5-15. MAE normalisée des prédictions de flux solaire pour le site de la Désirade en fonction des mois

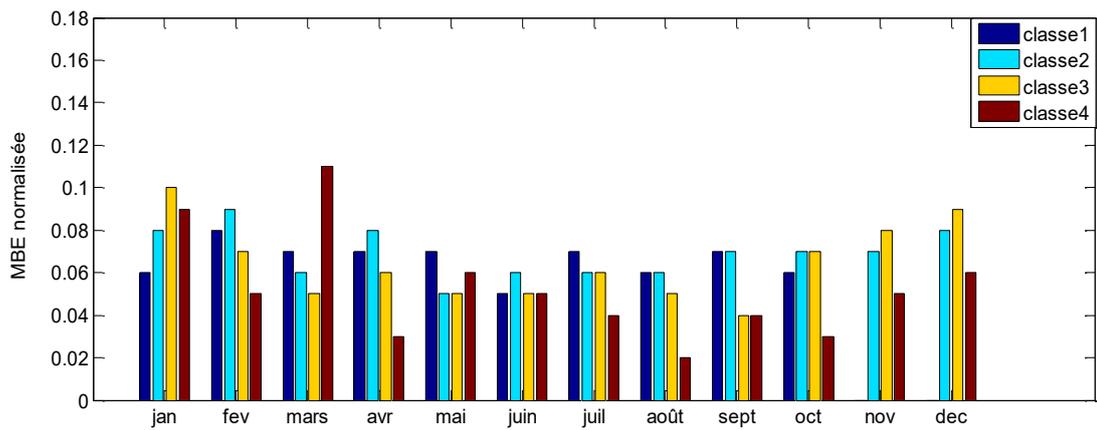


Figure 5-16. MBE normalisée des prédictions de flux solaire pour le site de la Désirade en fonction des mois

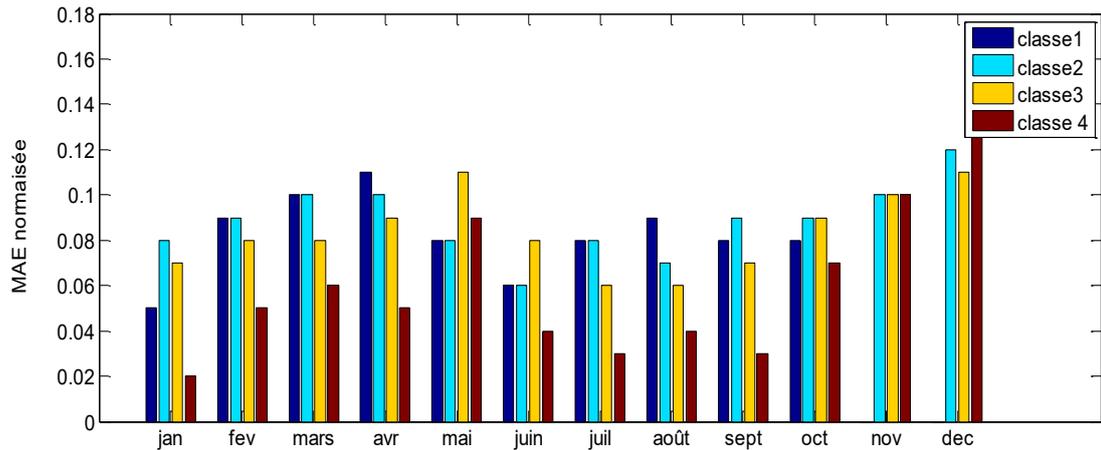


Figure 5-17. MAE normalisée des prédictions de flux solaire pour le site de Petit-canal en fonction des mois

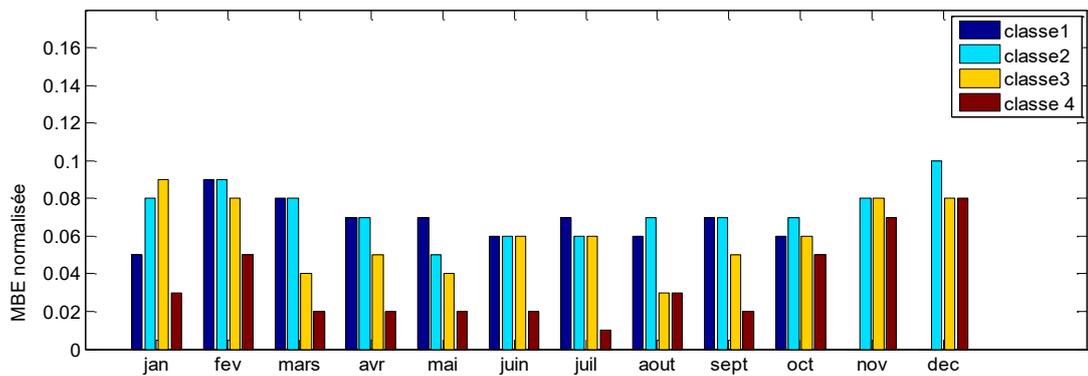


Figure 5-18. MBE normalisée des prédictions de flux solaire pour le site de Petit-canal en fonction des mois

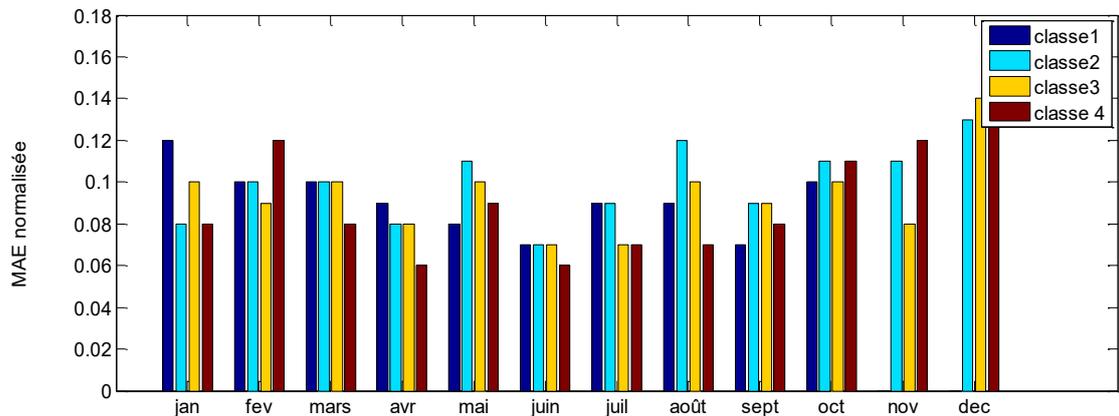


Figure 5-19. MAE normalisée des prédictions de flux solaire pour le site de Fouillole en fonction des mois

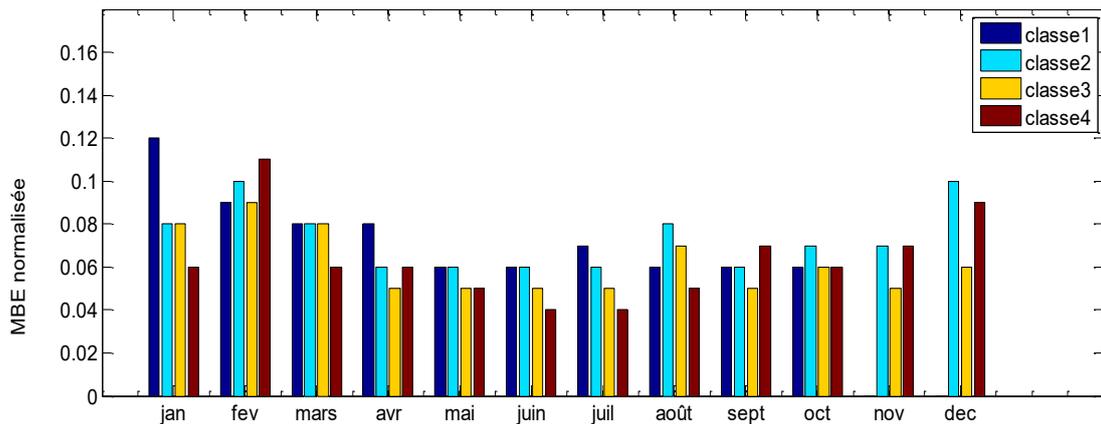


Figure 5-20. MBE normalisée des prédictions de flux solaire pour le site de Fouillole en fonction des mois

Cette analyse est effectuée pour chaque station et permet de vérifier si le modèle tend à sous-estimer ou surestimer les valeurs de flux solaire pour différentes dynamiques des signaux de rayonnement solaire global.

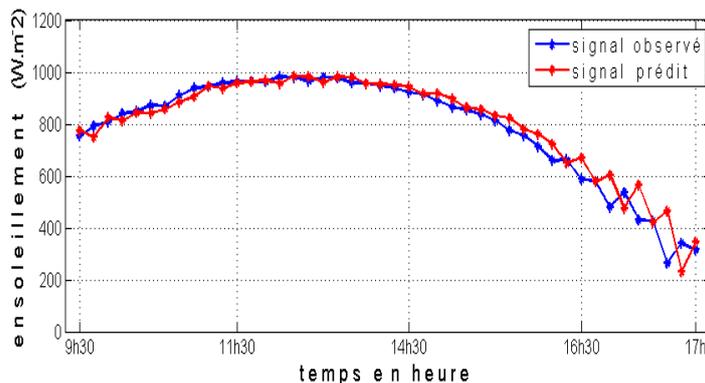
Nous avons pu observer des valeurs positives et négatives de MBE journalier. Néanmoins, en effectuant une normalisation mensuelle qui tend à généraliser les résultats, nous pouvons noter que les valeurs de MBE ne sont pas négatives ce qui signifie que le modèle a souvent tendance à surestimer les valeurs mesurées. Les valeurs de MAE sont inférieures à 15% excepté pour La Désirade en décembre et peut atteindre les 3% ce qui montre une bonne performance prédictive (Figure5-15 ; Figure 5-17 et Figure 5-19). Les classes qui présentent le plus souvent les plus faibles erreurs statistiques sont les journées de classe 4 particulièrement visible sur toute l'année pour Petit-canal (Fig.5-17 et Fig 5-18). Néanmoins, elles peuvent présenter (assez rarement) la plus forte valeur d'erreur statistique comme en Mars sur la Désirade ou en Février sur Fouillole. Cette classe est représentative de journées à ciel nuageux avec des masses nuageuses de grandes tailles. Dans ce cas où les valeurs de nRMSE sont faibles, le flux solaire est principalement diminué par les nuages et présente de faibles valeurs de flux solaire. Quand la masse nuageuse diminuant le rayonnement solaire a une vitesse lente, le niveau dynamique est faible et lent, ce qui peut expliquer la possibilité de faibles valeurs d'erreurs de prédiction.

Si les conditions météorologiques des journées de cette classe, présentent des masses nuageuses de grandes tailles ayant une vitesse de passage un peu plus rapide, les variations correspondant aux diminutions de flux solaire ont tendance à durer moins dans le temps et peuvent parfois présenter une forte dynamique à certaines périodes dans la journée, pour des valeurs de flux solaire très faibles ce qui peut donner de grandes valeurs d'erreurs. Les journées de classe 2 et 3 présentent les plus grandes valeurs d'erreurs de prévisions. La classe 2 est représentative de jours ayant un important rayonnement solaire avec quelques passages nuageux. En comparaison avec la classe 1, les courbes représentatives de la classe 2 ont une dynamique à niveau moyen comme on peut le constater dans le chapitre 3. Les

fluctuations de flux solaire sont plus importantes. Les jours de classe 3 sont représentatives d'une courbe de flux solaire avec une dynamique très importante. Ces journées présentent un nombre important de passages nuageux, en général de petites masses nuageuses avec une vitesse de passage rapide. Les erreurs statistiques ayant des valeurs moyennes sont obtenues pour la classe 1 qui est représentative de jours à ciel clair avec un important rayonnement solaire. Bien que ces journées présentent une dynamique très lente, il peut y avoir des passages nuageux épisodiques et rares au cours de la journée qui surviennent brutalement et donc qui suffisent parfois à augmenter les erreurs de prévisions tel qu'en janvier sur Fouillole (Figure 5-19 et Figure 5-20).

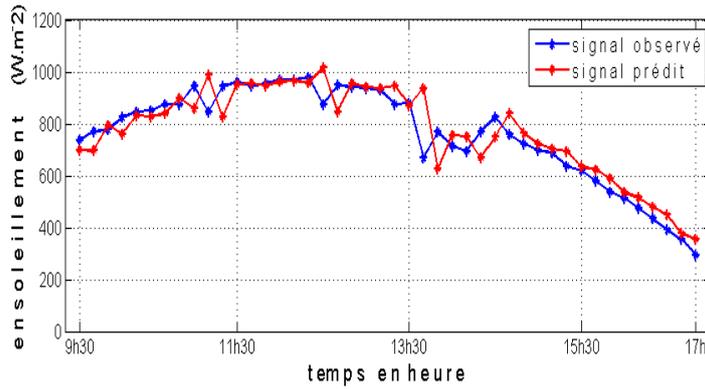
Selon les résultats, nous pouvons dire que la performance du modèle décroît quand les signaux de rayonnement solaire présentent un niveau de dynamique important. Le modèle a des difficultés à bien prédire les valeurs quand les variations sont rapides et brutales. Néanmoins, les valeurs des erreurs statistiques montrent une bonne performance du modèle VAR (1) spatio-temporel pour l'ensemble des classes de flux solaire journalier et n'excède pas en général les 15% pour la MAE et la MBE, quelque soit la dynamique (rapide ou lente) des fluctuations des signaux mesurés. Quelques exemples des signaux bien prédits par le modèle spatio-temporel VAR(1) pour chaque classe de journée sont présentés ci-dessous (Figure 5-21 à Figure.5-24).

Cette analyse est effectuée pour chaque station et permet de vérifier si le modèle tend à sous-estimer ou surestimer les valeurs de flux solaire pour différentes dynamiques des signaux de rayonnement solaire global.



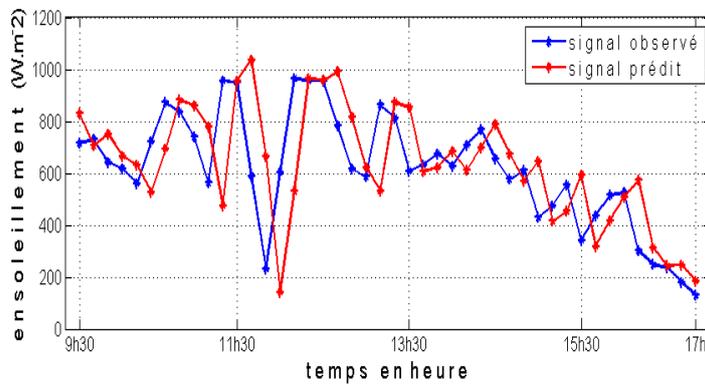
NRMSE (%)	NMAE(%)	NMBE(%)
7,12%	4,17%	-1,61%

Figure 5-21. Exemple de journée de classe 1 avec le tableau d'erreurs statistiques de prévision pour cette journée : courbes des valeurs mesurées en bleue et des valeurs prédites en rouge. Le coefficient de corrélation R entre les deux signaux est de 0,99.



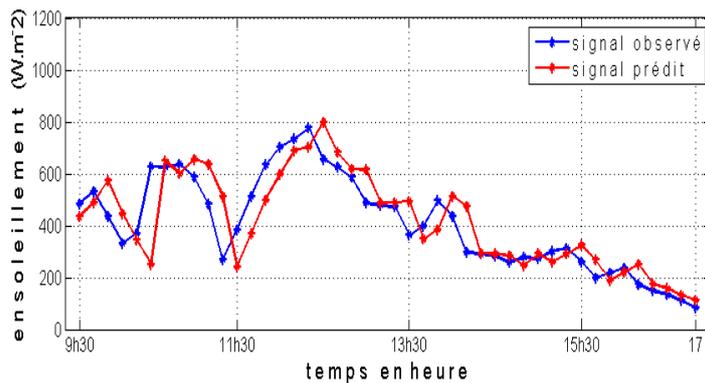
NRMSE (%)	NMAE(%)	NMBE(%)
10,61%	6,86%	-1,32%

Figure 5-22. Exemple de journée de classe 2 avec le tableau d'erreurs statistiques de prévision pour cette journée : courbes des valeurs mesurées en bleu et des valeurs prédites en rouge. Le coefficient de corrélation R entre les deux signaux est de 0,98.



NRMSE (%)	NMAE(%)	NMBE(%)
21,27%	17,40%	-2,19%

Figure 5-23. Exemple de journée de classe 3 avec le tableau d'erreurs statistiques de prévision pour cette journée : courbes des valeurs mesurées en bleu et des valeurs prédites en rouge. Le coefficient de corrélation R entre les deux signaux est de 0,96.



NRMSE (%)	NMAE(%)	NMBE(%)
15,20%	11,01%	-1,31%

Figure 5-24. Exemple de journée de classe 4 avec le tableau d'erreurs statistiques de prévision pour cette journée : courbes des valeurs mesurées en bleu et des valeurs prédites en rouge. Le coefficient de corrélation R entre les deux signaux est de 0,98.

III.4. Analyse de la performance prédictive pour les saisons

Météo France de Guadeloupe décrit une classification des temps au cours de l'année sur l'île de la Guadeloupe dans Brévignon C.(2005) qui correspond entre autre à la distinction et la caractérisation des saisons. Une saison sèche commence début janvier pour se terminer mi-avril. Le régime de temps est sec en général. L'Alizé est constant et soutenu. De mi-avril à mi-juin, les Antilles connaissent une saison intermédiaire assez longue. Le temps alterne souvent entre averses et embellies. Mais parfois de copieuses pluies, souvent orageuses, se manifestent entre fin avril et début mai. A partir de mi-juin et jusqu'à mi-novembre vient la saison des pluies, appelée « la saison cyclonique ». Les précipitations y sont plus abondantes. Mais occasionnellement, la faiblesse du vent ou panne d'alizé se traduit par un temps lourd qui débouche sur des orages. L'ensoleillement est donc altéré par un ciel souvent chargé de nuages. On retrouve une nouvelle saison de transition extrêmement rapide de la mi-novembre à décembre.

Le but de cette analyse est d'évaluer la performance du modèle spatio temporel en fonction des saisons au cours de l'année. Pour cela, nous avons calculé une moyenne des résultats des Tableaux 5-1 à 5-9 pour une saison donnée correspondant une période mensuelle définie par la classification de Brévignon C. (2005) mentionnée précédemment.

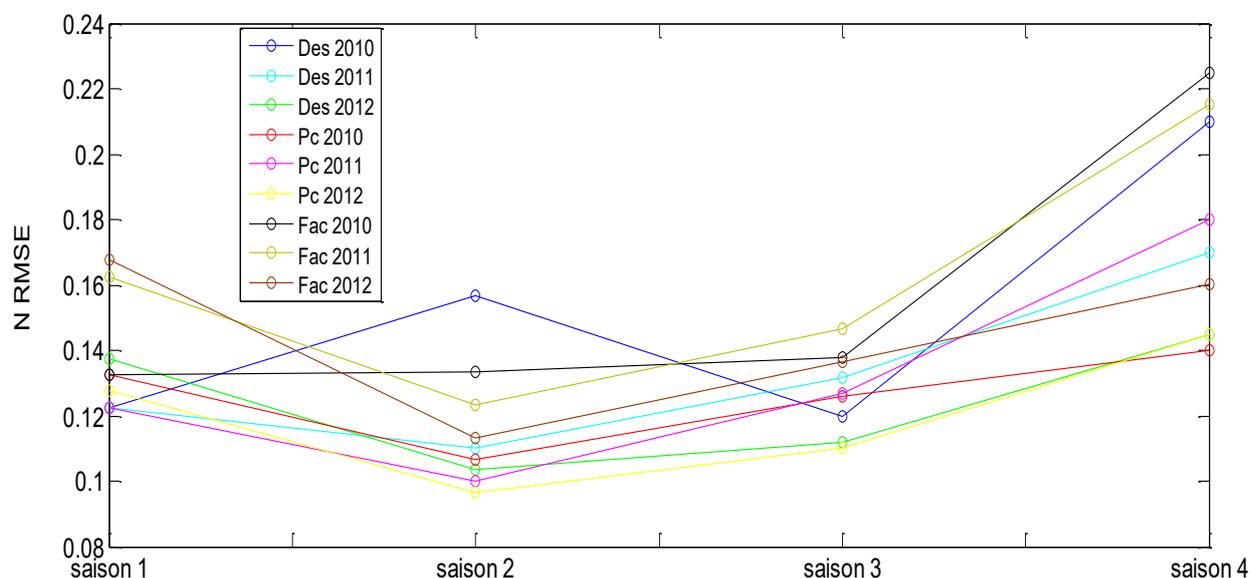


Figure 5-25. RMSE normalisée en fonction des saisons pour les années 2010, 2011, 2012 sur les trois stations (Des (La Désirade), Pc (Petit-canal), Fac (Fouillole)), par application du modèle spatio-temporel VAR(1) avec les stations ordonnées selon la direction du vent.

Selon la Figure 5-25, la saison 1 présente une gamme d'erreur de nRMSE de 12% à 17%, la saison 2 de 10% à 16%, la saison 3 de 11% à 15% et la saison 4 de 11% à 22%.

Seule la saison 3 présente des résultats très homogènes d'une année à l'autre. Les autres saisons, particulièrement la saison 4 présente une large gamme d'erreurs de prévisions statistiques certainement dues aux conditions météorologiques instables en cette période. En effet, durant cette saison 4 qui est une saison de transition extrêmement rapide, bien que les alizés, appelés les « avents » à cette période, deviennent plus réguliers, les averses alternent avec de belles éclaircies, ce qui rend très irréguliers les résultats de prévisions d'une année à l'autre.

III.5. Analyse de la performance prédictive pour différents horizons

Les résultats pour les horizons de prévisions de 10 minutes, présentés préalablement, sont les plus ciblés selon la requête du gestionnaire de réseau puisqu'il lui faut 15 minutes pour changer de système d'exploitation d'énergie électrique en cas de problèmes techniques. Néanmoins, nous testons le modèle pour plusieurs horizons de prévisions afin d'analyser l'évolution de la performance prédictive du modèle spatio-temporelle VAR en fonction des horizons. Dans cette section nous présentons les résultats des erreurs statistiques (RMSE normalisée) pour respectivement les horizons de prédiction de 5 min, 15min, 30 min et 1h.

III.5.1 Prévisions à 5 minutes

Le modèle VAR a été appliqué pour plusieurs ordres p à un horizon de 5 minutes pour chaque station (La Désirade, Petit-canal, Fouillole), et chaque année, soit des tests sur 1029 jours. Afin d'évaluer la performance des prévisions à cet horizon, les résultats de RMSE normalisés moyennés annuellement ont été calculés et sont présentés dans les Tableaux 5-11 à 5-13 pour chaque station. Pour une évaluation plus détaillée des RMSE, les résultats mensuels sont présentés en Annexe.

	p=1	p=2	p=3	p=4
2010	0,09	0,19	0,19	0,27
2011	0,13	0,17	0,26	0,29
2012	0,10	0,15	0,18	0,21

Tableau 5-11 RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 5 min sur La Désirade

	p=1	p=2	p=3	p=4
2010	0,09	0,16	0,21	0,27
2011	0,12	0,16	0,20	0,24
2012	0,09	0,14	0,17	0,20

Tableau 5-12. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 5 min sur Petit-canal

	p=1	p=2	p=3	p=4
2010	0,10	0,22	0,26	0,36
2011	0,12	0,18	0,24	0,29
2012	0,10	0,17	0,32	0,27

Tableau 5-13 RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 5 min sur Fouillole

Suite à l'analyse de ces tableaux de résultats, deux principaux commentaires peuvent être donnés. Le modèle adéquate est le modèle spatio-temporel à p -ordre égale à 1 c'est-à-dire VAR(1) pour un horizon de prévision de 5 minutes et les valeurs des RMSE annuelles présentent un pourcentage compris entre 9% et 13 % à l'ordre 1 sur l'ensemble des sites.

Les erreurs statistiques sont par conséquent inférieures à celles des prévisions pour un horizon de 10 minutes (de 12% à 15%). Les vecteurs de données présentent plus de variabilités à 5 minutes qu'à 10 minutes, néanmoins, l'information contenue dans les données passées (5 minutes avant) est plus récente comparée à celle de 10 minutes antérieures. Ce qui pourrait expliquer cette amélioration de la prédictibilité pour un horizon à 5 minutes comparé à 10 minutes. Un exemple d'une journée bien prédite pour un horizon à 5 minutes est présentée sur la Figure 5-26. En Annexe dans les Tableaux Tab .1 à Tab.9, les nRMSE peuvent atteindre des valeurs autour de 50% et 60% pour l'ordre p égale à 4 correspondant à des combinaisons linéaires des séries temporelles passées de la station de référence et des stations corrélées jusqu'à 20 minutes avant soit quatre points avant la donnée à prédire. Ce sont des erreurs très grandes comparées à celles de 10 minutes à l'ordre p égale 4. En effet, la dynamique des signaux à 5 minutes est plus rapide (le flux solaire à $t-20$ minutes a peu de probabilité de fournir des informations sur les données de flux solaire à l'instant t).

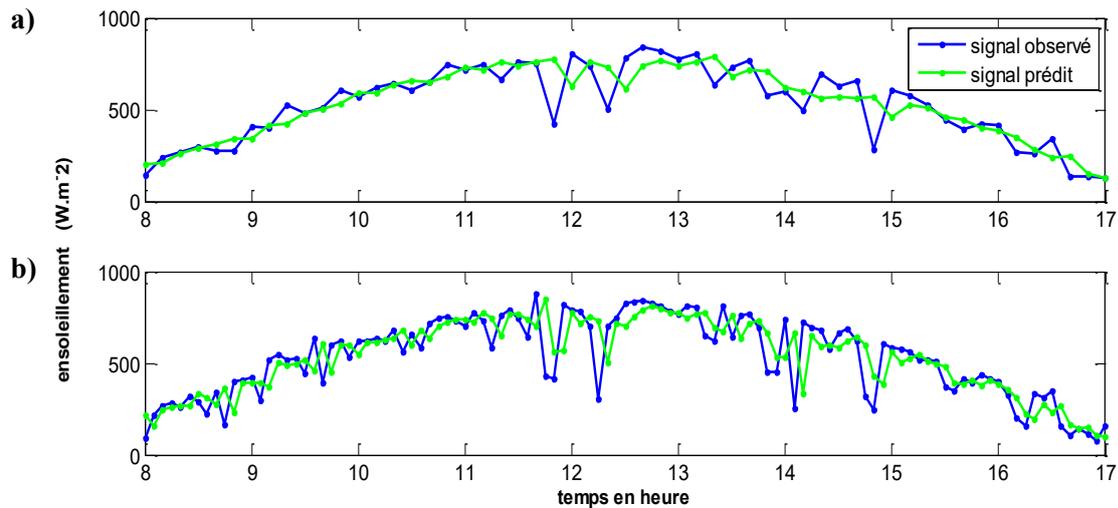


Figure 5-26. a) Prévision à 10 minutes du 7 mars 2012 et son signal mesuré correspondant . b) Prévision à 5 minutes du même jour et son signal mesuré correspondant.

On constate bien que malgré la forte variabilité du signal à 5minutes, le modèle spatio-temporel VAR arrive à donner une prévision très proche des valeurs observées (Figure 5-26). Pour cet exemple de journée, on trouve une erreur statistique RMSE de $34,8 \text{ W.m}^{-2}$ pour un horizon de prévision à 5 minutes au lieu de $58,48 \text{ W.m}^{-2}$ pour un horizon de prévision à 10 minutes.

III.5.2 Prévisions à 15 minutes

Nous avons également testé le modèle VAR pour un horizon de prévision de 15 minutes pour plusieurs ordres p pour chaque station et chaque année. Les résultats de RMSE normalisés moyennés annuellement sont présentés dans les Tableaux 5-14 à 5-16. Pour une évaluation plus détaillée des RMSE, les résultats mensuels sont présentés en Annexe.

	p=1	p=2	p=3	p=4
2010	0,20	0,21	0,25	0,31
2011	0,17	0,21	0,31	0,31
2012	0,15	0,18	0,22	0,30

Tableau 5-14 RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 15 min sur La Désirade

	p=1	p=2	p=3	p=4
2010	0,16	0,22	0,25	0,34
2011	0,16	0,20	0,31	0,35
2012	0,17	0,18	0,22	0,27

Tableau 5-15. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 15 min sur Petit-canal

	p=1	p=2	p=3	p=4
2010	0,19	0,18	0,26	0,33
2011	0,21	0,27	0,42	0,40
2012	0,19	0,23	0,36	0,35

Tableau 5-16. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 15 min sur Fouillole

Les résultats sont analogues concernant l'évaluation des erreurs statistiques en fonction des différents ordres. Le modèle adéquate reste le modèle spatio-temporel VAR(1) pour également un horizon de prévision de 15 minutes. Les valeurs des RMSE annuelles présentent un pourcentage compris entre 15% et 21 % à l'ordre 1 sur l'ensemble des sites. Les erreurs statistiques sont par conséquent supérieures à celles des prévisions pour un horizon de 10 minutes (de 12% à 15%) et 5 minutes (de 12% à 15%). Les vecteurs de données présentent moins de variabilités à 15 minutes qu'à 10 minutes, néanmoins, l'information contenue dans les données passées (15 minutes avant) sont moins récentes, comparé à 10 minutes. Ce qui pourrait expliquer cette diminution plus ou moins légère de la prédictibilité pour un horizon à 15 minutes comparé à 10 minutes (avec des valeurs d'erreurs environ 4% supérieures). Un exemple de prévision à 15 minutes pour une journée est présenté sur la Figure 5-27.

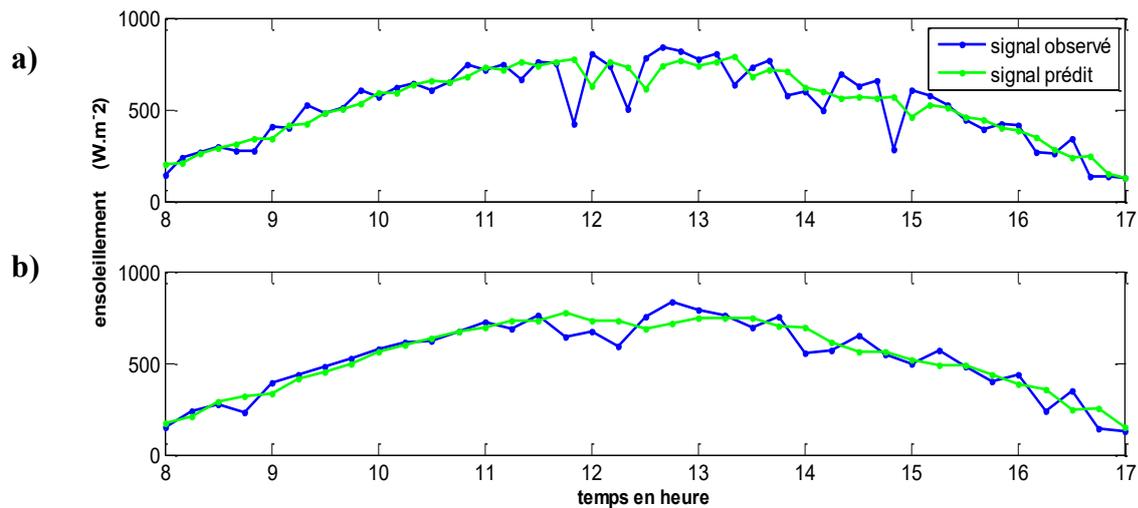


Figure 5-27. a) Prévision à 10 minutes du 7 mars 2012 et son signal mesuré correspondant. b) Prévision à 15 minutes du même jour et son signal mesuré correspondant.

Le modèle spatio-temporel VAR(1) arrive à donner une prévision très proche des valeurs observées (Figure 5-27). On trouve une erreur statistique RMSE de $60,93 \text{ W.m}^{-2}$ pour un horizon de prévision à 15 minutes au lieu de $58,48 \text{ W.m}^{-2}$ pour un horizon de prévision à 10 minutes.

III.5.4 Prévisions à 30 minutes

Nous avons testé le modèle VAR pour un horizon de prévision de 30 minutes pour plusieurs ordres p , pour chaque station (La Désirade, petit-canal, Fouillole), et chaque année. Les résultats de RMSE normalisés annuellement sont présentés dans les Tableaux 5-17 à 5-19. Pour une évaluation plus détaillée des RMSE, les résultats mensuels sont présentés en Annexe.

	p=1	p=2	p=3	p=4
2010	0,11	0,15	0,21	0,41
2011	0,13	0,17	0,22	0,37
2012	0,11	0,14	0,19	0,37

Tableau 5-17. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 30 min sur La Désirade

	p=1	p=2	p=3	p=4
2010	0,12	0,15	0,23	0,41
2011	0,15	0,20	0,20	0,33
2012	0,12	0,16	0,22	0,38

Tableau 5-18. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 30 min sur La Désirade

	p=1	p=2	p=3	p=4
2010	0,12	0,16	0,24	0,47
2011	0,13	0,19	0,24	0,44
2012	0,11	0,15	0,20	0,37

Tableau 5-19 RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 30 min sur La Désirade

Le modèle approprié reste le modèle spatio-temporel VAR(1) pour également un horizon de prévision de 30 minutes. Les valeurs des RMSE annuelles présentent un pourcentage compris entre 11% et 15 % à l'ordre 1 sur l'ensemble des sites. Les erreurs statistiques sont par conséquent inférieures à celles des prévisions pour un horizon de 15 minutes (15 à 21%), 10 minutes (de 12% à 15%) et 5 minutes (de 12% à 15%). Les vecteurs de données présentent bien moins de variabilités qu'à 15 minutes. Bien que, l'information contenue dans les données passées (30 minutes avant) sont moins récentes que l'information contenue dans les données 15 minutes avant, 10 minutes avant et 5 minutes avant, la prévision à un horizon de 30 minutes est plus proche des valeurs mesurées. Ce qui justifie ces résultats est le fait que le signal est très lissé à 30 minutes, les signaux présentent une variabilité plus ou moins faible. Ce qui pourrait expliquer cette amélioration de la prédictibilité pour un horizon à 30 minutes comparé à 15 minutes, 10 minutes et 5 minutes. Un exemple de prévision à 30 minutes pour une journée est présenté sur la Figure 5-28. Néanmoins, si on observe les résultats en Annexe pour un horizon de prévision de 30 minutes avec $p=4$, les valeurs des RMSE normalisées mensuelles peuvent atteindre 50 à 60 %, ce qui montre des valeurs de prévision très écartées des valeurs observées. Le modèle avec $p=4$ soit VAR (4), utilise des séries temporelles passées à un temps de retard maximal de 2h des stations corrélées pour expliquer et prédire la future valeur de la station de référence. 2 heures est une échelle trop grande pour expliquer qu'un évènement soit à l'origine de la variabilité du flux solaire sur une autre station causant une diminution ou augmentation du flux solaire 2 heures après.

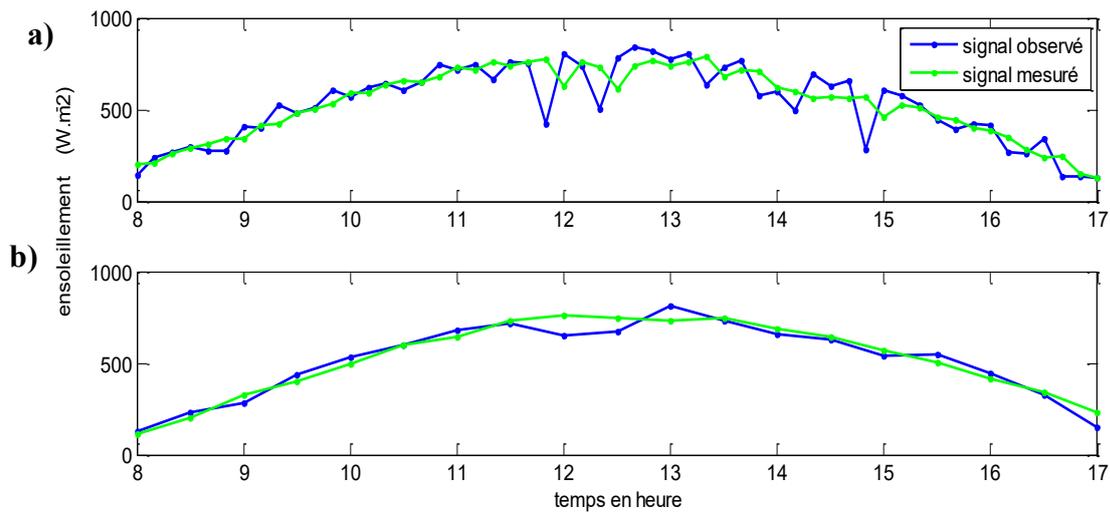


Figure 5-28 a) Prédiction à 10 minutes du 7 mars 2012 et son signal mesuré correspondant. b) Prédiction à 30 minutes du même jour et son signal mesuré correspondant.

La Figure 5-28 montre que le modèle spatio-temporel VAR(1) à un horizon de prévision de 30 minutes donne des valeurs très proches des valeurs observées. On obtient effectivement une performance de prédiction supérieure à celle de 10 minutes puisque le signal présente une faible variabilité comparé au signal de 10 minutes. Pour cet exemple de journée, on trouve une erreur statistique RMSE de $33,91 \text{ W.m}^{-2}$ pour un horizon de prévision à 30 minutes au lieu de $58,48 \text{ W.m}^{-2}$ pour un horizon de prévision à 10 minutes et $34,8 \text{ W.m}^{-2}$ pour un horizon de prévision à 5 minutes.

III.5.4 Prévisions à 1 heure

Le modèle VAR spatio-temporel a été appliqué et testé pour une prévision à 1 h pour les stations, chaque année. Les résultats de RMSE normalisés moyennés annuellement sont présentés dans les Tableaux 5-20 à 5-22. Pour une évaluation plus détaillée des RMSE, les résultats mensuels sont présentés en Annexe.

	p=1	p=2	p=3	p=4
2010	0,14	0,24	0,35	0,53
2011	0,18	0,20	0,35	0,53
2012	0,12	0,17	0,29	0,30

Tableau 5-20 RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 1 heure sur La Désirade

	p=1	p=2	p=3	p=4
2010	0,14	0,22	0,35	0,54
2011	0,16	0,20	0,41	0,35
2012	0,13	0,18	0,26	0,31

Tableau 5-21. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 1 heure sur Petit-canal

	p=1	p=2	p=3	p=4
2010	0,14	0,25	0,38	0,56
2011	0,17	0,19	0,36	0,36
2012	0,12	0,18	0,25	0,34

Tableau 5-22. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à un horizon de 1 heure sur Fouillole

Le modèle approprié reste le modèle spatio-temporel à p -ordre égale à 1 c'est-à-dire VAR(1) pour également un horizon de prévision de 1heure. Les valeurs des RMSE annuelles présentent un pourcentage compris entre 12% et 18 % à l'ordre 1 sur l'ensemble des sites. Les erreurs statistiques sont par conséquent inférieures à celles des prévisions pour un horizon de 15minutes (15 à 21%), mais plus ou moins proches des erreurs statistiques des prévisions à 10 minutes (de 12% à 15%) et des prévisions à 5 minutes (de 12% à 15%). Les vecteurs de données à cette échelle de temps sont très lissés et expliqueraient les bons résultats trouvés. Néanmoins, comme pour les prévisions à 30 minutes, si on observe les résultats en Annexe pour un horizon de prévision de 1 heure avec $p=4$, les valeurs des RMSE normalisées mensuelles peuvent atteindre 60 à 70 %, ce qui montre des valeurs de prévision très écartées des valeurs observées. Le modèle avec $p=4$ soit VAR (4), utilise des séries temporelles passées à un temps de retard maximal de 4h des stations corrélées pour expliquer et prédire la future valeur de la station de référence. Comme mentionné précédemment, 2 heures voir 4 heures est une échelle de temps trop grande pour expliquer qu'un évènement soit à l'origine de la variabilité du flux solaire sur une autre station causant une diminution ou augmentation du flux solaire 4 heures après.

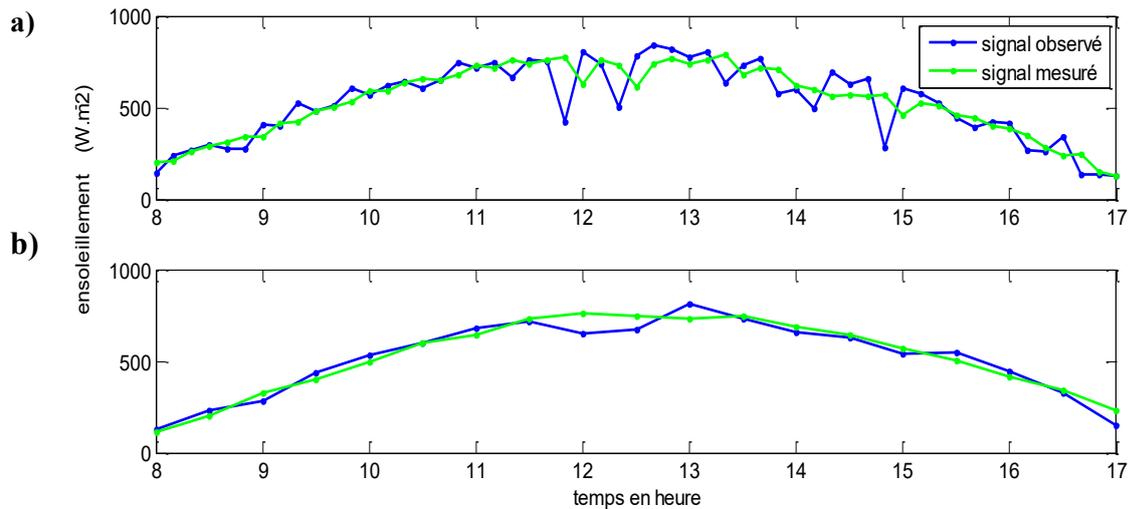


Figure 5-29 Prévision à 10 minutes du 7 mars 2012 et son signal mesuré correspondant. b) Prévision à un horizon d' 1h du même jour et son signal mesuré correspondant.

La Figure 5-29 montre que le modèle spatio-temporel VAR(1) à un horizon de prévision de 1 heure donne des valeurs très proches des valeurs observées. On obtient effectivement une performance de prédiction supérieure à celle de 10 minutes puisque le signal présente une très faible variabilité comparé au signal de 10 minutes. Le signal est effectivement très lissé. Pour cet exemple de journée, on trouve une erreur statistique RMSE de $32,27 \text{ W.m}^{-2}$ pour un horizon de prévision à 1 heure au lieu de $58,48 \text{ W.m}^{-2}$ pour un horizon de prévision à 10 minutes.

IV. Amélioration des paramètres du modèle spatio-temporel VAR

Cette section met en exergue différents tests qui ouvrent des perspectives d'amélioration du modèle VAR spatio-temporel. Nous notons deuxièmement les limites du modèle VAR.

IV.1. Tests des paramètres du modèle et amélioration de la performance de prévision

Plusieurs tests ont été effectués et ont montré une amélioration des résultats de prévision.

Tests sur les paramètres d'initialisation

Il est nécessaire d'avoir au minimum 15 points d'initialisation pour que le modèle effectue une prévision. Même si on récupère plus de points de la journée à prédire pour initialiser le modèle, les erreurs de prévision ne sont pas modifiées puisque le modèle fonctionne en prenant par itération les

valeurs passées de la série temporelle mesurée. En effet, pour prédire le point suivant, le modèle récupère tous les points passés de la série temporelle mesurée. Néanmoins, nous avons testé les paramètres d'initialisation pour un ensemble de points d'initialisation correspondants à des séries de temps passées d'une semaine, un mois, deux mois.

Nous avons constaté que lorsqu'on applique le modèle avec une période de 2 semaines pour l'initialisation de la prévision, les erreurs sont plus importantes comparés à 15 points d'initialisation. Par contre, pour des points d'initialisation correspondant à une période de 1 mois (soit le double de 2 semaines) ou 2 mois, les résultats sont meilleurs, il y a une amélioration de la prévision comparé à 15 points d'initialisation. La Figure 5-30 nous permet en effet de le constater. Les valeurs prédites les plus proches des valeurs observées correspondent aux résultats d'une prévision obtenus avec 1 à 2 mois de points d'initialisation. On trouve une RMSE de 55% pour 2 semaines de points d'initialisation, 12% pour 15 points d'initialisation, 5,8% pour un mois de points d'initialisation et 8,3% pour 2 mois de points d'initialisation, pour cette journée.

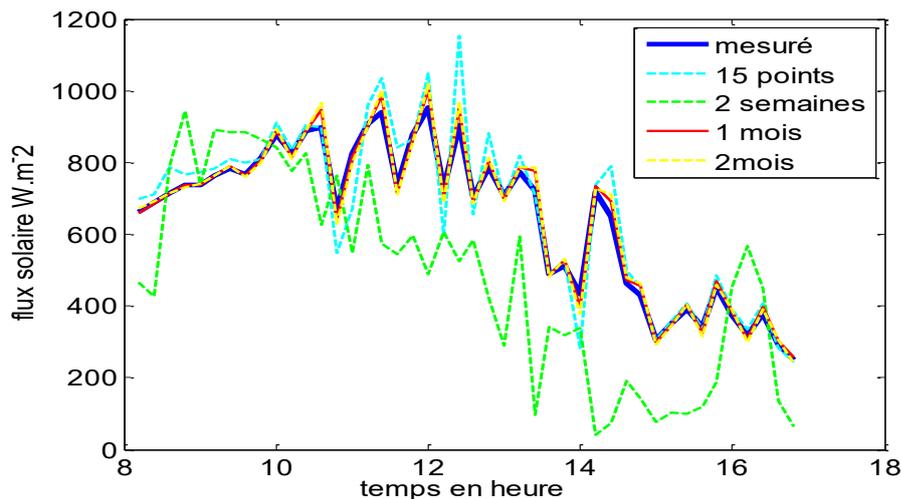


Figure 5-30 Signal mesuré et signaux de prévisions correspondants pour différentes périodes d'initialisation : 15 points, 2 semaines de points d'initialisation soient 616 points, 2mois de points d'initialisation soient 1364 points, 2 mois d'initialisation soient 2728 points.

Par conséquent, on peut améliorer les résultats de prévisions obtenus dans les sections précédentes en utilisant une période de 1 mois soit 1364 points pour initialiser le modèle.

Tests sans transformation logarithmique

Nous avons évalué la performance du modèle sans transformation logarithmique sur les séries temporelles. Bien que cette transformation ait été initialement utilisée pour stabiliser la variance des

séries temporelles, on trouve de meilleurs résultats sans son application. On peut le constater sur la Figure 5-31 qui représente l'histogramme des RMSE journaliers moyennés mensuellement normalisées par la moyenne mensuelle des valeurs observées. En effet, les erreurs statistiques montrent une meilleure performance de la prévision sur les séries temporelles sans transformation logarithmique quelque soit le mois considéré avec une valeur maximale de 11% RMSE au lieu de près de 18% RMSE avec transformation logarithmique. Un exemple de journée sur la Figure 5-32 montre que la courbe de prévision sans transformation logarithmique sur les données a des valeurs plus proches du flux solaire mesuré que la courbe de prévision avec transformation logarithmique du vecteur des données.

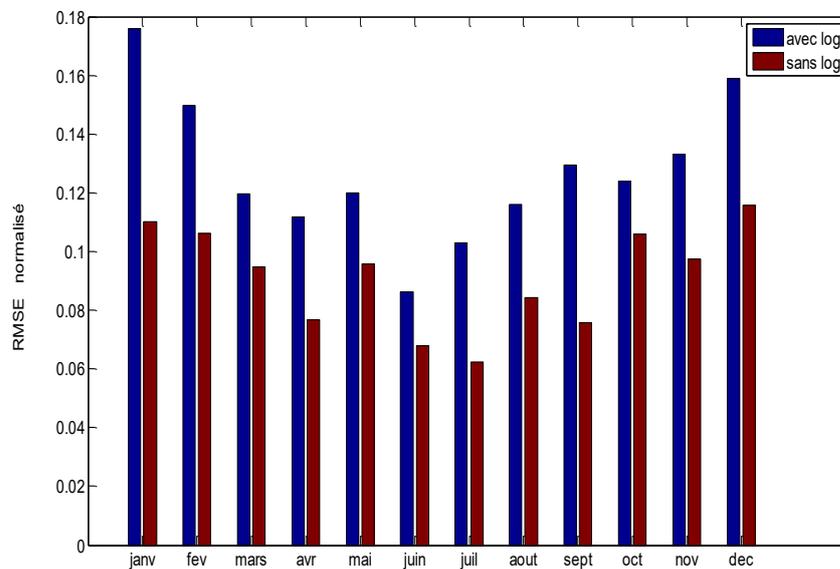


Figure 5-31 Résultats des RMSE normalisées des prévisions avec transformation logarithmique des données (en bleu) et sans transformation logarithmique des données (en rouge)

Ces tests apportent une nouvelle amélioration pour la prédictibilité du modèle. La transformation logarithmique des données de flux solaire pour stabiliser la variance n'est donc pas nécessaire pour notre cas d'étude.

Tests avec l'indice ciel clair

Pour l'application du modèle spatio-temporel VAR, nous avons considéré que la série temporelle de l'ensoleillement global se comportait selon un modèle additif, ce qui a permis de justifier la méthode de différenciation pour stationnariser la série en dimension temporelle. Nous avons également testé le modèle spatio-temporel VAR, en considérant que la série temporelle pouvait suivre un modèle multiplicatif. On a alors assimilé la série d'origine à cette équation:

$$X_t = T_t * S_t * Y_t \quad \text{pour tout } t \in \{1, \dots, n\}$$

avec T_t la tendance, S_t la composante saisonnière et Y_t la composante stationnaire.

Etant donné qu'il s'agit de série temporelle d'une seule journée on peut supprimer le terme correspondant à la saisonnalité. On obtient alors un modèle multiplicatif qui décrit la série temporelle telle que :

$$X_t = T_t * Y_t \quad \text{pour tout } t \in \{1, \dots, n\}$$

avec T_t la tendance et Y_t la composante stationnaire.

La composante stationnaire est donc sous la forme :

$$Y_t = \frac{X_t}{T_t} \tag{5.30}$$

Ce qui peut être assimilé à l'indice ciel clair :

$$k_c = \frac{G}{G_c} \tag{5.31}$$

G_c est le flux solaire théorique reçu par le capteur, qui n'est sujet à aucunes fluctuations rapides mais rend compte de la tendance du flux solaire due à la position du soleil par rapport au point de mesure. Pour notre cas d'étude le flux solaire théorique est calculé par le modèle de Kasten. G correspondant aux séries temporelles de flux solaire mesuré. L'indice ciel clair représente donc le processus stochastique du flux solaire sans tendance.

Dans plusieurs travaux de recherche telles que dans Hoff et Perez 2010, Pérez et al 2012, Badosa et al 2013, Yang et al 2014, c'est l'indice de clarté et l'indice ciel clair qui sont utilisés pour évaluer la variabilité de l'ensoleillement. L'indice ciel clair offre deux avantages : la normalisation de la variabilité à l'unité, et la suppression des effets de géométrie solaire qui est une source de la tendance de l'ensoleillement.

Nous avons donc testé le modèle considérant la série temporelle de flux solaire initiale suivant un modèle multiplicatif, et le processus stochastique étant associé à l'indice de clarté. La courbe de flux solaire est retrouvée en multipliant les résultats de prévision par le rayonnement théorique G_c . Les erreurs statistiques de prévision obtenues sont montrées dans la Figure 5-33 représentant l'histogramme des RMSE journaliers moyennés mensuellement normalisée par la moyenne mensuelle des valeurs observées. On peut constater que les erreurs statistiques montrent une meilleure performance de la prévision sur les séries temporelles avec transformation en K_c quelque soit le mois considéré avec une valeur maximale de 12% nRMSE au lieu de près de 18% nRMSE avec transformation juste par différenciation d'ordre 1.

Un exemple de journée sur la Figure 5-34 permet d'observer que la courbe de prévision utilisant dans le processus du modèle les données d'indice ciel clair (Kc), a des valeurs plus proches du flux solaire mesuré que la courbe de prévision avec méthode de différenciation du vecteur des données.

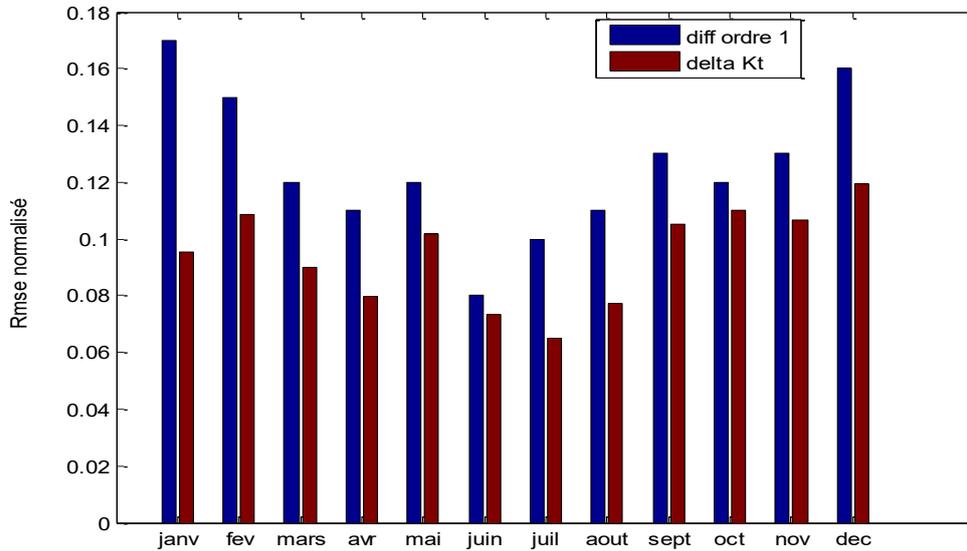


Figure 5-33 Résultats des RMSE normalisées des prévisions avec transformation des données en ΔKc (en bleu) et par différenciation d'ordre 1 des données (en rouge)

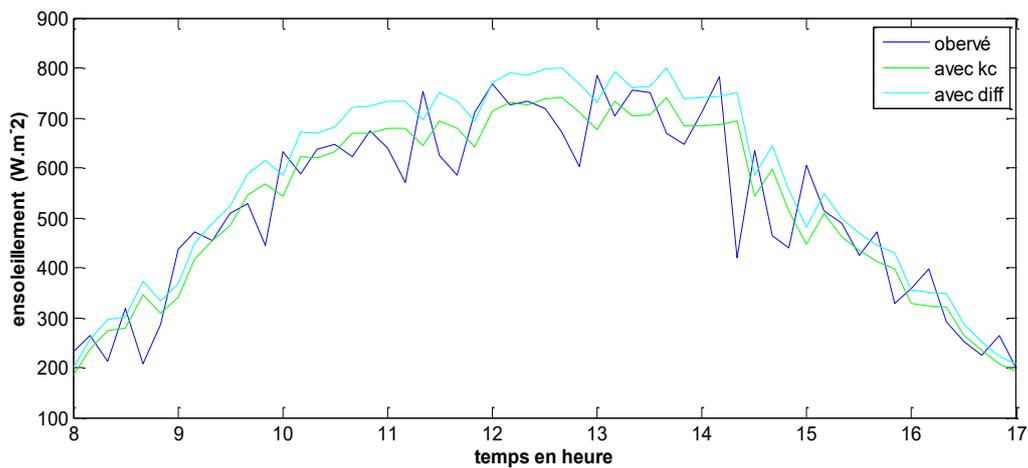


Figure 5-34 Valeurs des prévisions obtenues avec le modèle appliquant ΔKc des données (en noir) et les valeurs de prévisions obtenues avec différenciation d'ordre 1 des données (en bleu cyan). La courbe des valeurs observées correspondante est en bleu (1^{er} janvier 2012).

On peut observer que la prédiction obtenue par différenciation d'ordre 1 des données observées montre la même performance du point de vue de la dynamique que celle obtenue avec le Kc . La vraie

différence de performance prédictive se trouve au niveau des écarts de valeurs. Ce phénomène peut se justifier par le fait que le signal Kc est le même signal (même variabilité) que celui obtenu par différenciation mais avec une différence d'échelle. Les signaux Kc ont des valeurs environ 100 fois plus petite, due à la normalisation par l'indice ciel clair mais le plus pertinent est au niveau de la stationnarisation temporelle plus efficace par le paramètre Kc qui montre une moyenne constante au cours du temps notoire par rapport à la méthode de différenciation. Ces tests apportent une nouvelle amélioration pour la prédictibilité du modèle.

IV.2. Comparaison avec d'autres modèles

Les tests des paramètres ont montré une meilleure performance prédictive du modèle pour des points d'initialisation correspondant à un mois de donnée, une stationnarisation temporelle par la différenciation d'ordre 1 de l'indice ciel clair et la suppression d'une transformation logarithmique sur les données. Après application de ces améliorations aux processus du modèle spatio-temporel VAR, nous avons comparé les résultats de prévisions du modèle avec d'autres modèles de la littérature. Pour des horizons à 10min, Chu et al (2013) dont le modèle est discuté dans le chapitre état de l'art, utilisant des données de rayonnement direct moyennées à 1 min trouve une meilleure performance de $54,6 \text{ W.m}^{-2}$ RMSE pour un rayonnement direct faible et 132 W.m^{-2} pour un rayonnement direct important. Pour notre modèle spatio-temporel VAR, on trouve une RMSE minimale de 38 W.m^{-2} pour une journée de flux solaire à dynamique faible et pour une journée à forte dynamique, une RMSE minimale de 98 W.m^{-2} . Dans Glasbey et al (2007) la performance du modèle STAR(1) à une échelle de 10 min de prévision est décrit par une comparaison de moyennes des paramètres estimés et des valeurs moyennes des paramètres réels. Dans Chow (2011), la prévision à 10 minutes effectuée par un imageur du ciel est décrite par la moyenne et l'écart-type des "matchings errors" calculés entre deux cartes de nuages. Ces deux dernières méthodes d'évaluation ne peuvent être comparées à l'évaluation de la performance de notre modèle.

Une première comparaison entre le modèle de persistance et le modèle spatio-temporel VAR est effectuée pour chaque horizon de prévision sur l'ensemble des données de mesures (Tableau 5-23). Le modèle de persistance est l'un des modèles les plus faciles à mettre en œuvre et souvent utilisé comme une base de référence pour l'évaluation de la performance des autres modèles de prévisions tel que dans Perez et al (2010). Les prévisions effectuées par le modèle de persistance, comme son nom l'indique, suppose que le flux solaire à l'instant t persiste à l'instant $t+1$. Les détails du processus du modèle de persistance sont présentés dans le chapitre état de l'art. Ce modèle a été appliqué sur l'indice ciel clair. La RMSE normalisée calculée pour les modèles de la littérature se réfère souvent à la RMSE relative décrite par cette équation :

$$rRMSE = \frac{\sqrt{\frac{1}{n} \sum (\hat{G} - G)^2}}{\frac{1}{n} \sum G} \quad (5.27)$$

Avec rRMSE la RMSE relative normalisée par la moyenne des données observées pour une période d'une année, \hat{G} les valeurs prédites et G les valeurs observées avec n est le nombre de points des données.

Les résultats des RMSE relative pour chaque localisation sont présentés dans le tableau et ces résultats moyennés sur les trois sites sont représentées par la Figure 5-35. Ces résultats montrent une meilleure performance du modèle spatio-temporel VAR comparé au modèle de persistance. La RMSE maximale trouvée pour les horizons à 5 min, 10 min, 15 min, 30 min et 1h est respectivement de 26,24%, 26,12%, 26,06%, 27,18%, 35,94% pour le modèle de persistance comparé à 24,54%, 23,79%, 23,63%, 22,48% et 22,95%, pour le modèle spatio-temporel VAR. Une valeur minimale de 19,21% est trouvée pour le modèle de persistance à un horizon de 5 min, alors que pour le modèle spatio-temporel la rRMSE minimale calculée est de 17,41% à un horizon de 30 min (voir Figure 5-35). Le modèle spatio-temporel VAR prend en compte des paramètres spatiaux et temporels qui sont influencés par la configuration spatiale et temporelle des localisations. Ainsi, les résultats du modèle montre que par rapport à la structure spatiale de notre zone de mesures composé de trois stations, la prévision de l'ensoleillement sur une station est plus performante en utilisant les données passées à t-30 min des autres stations. Pour le processus du modèle de persistance, plus l'information fournie par les séries temporelles passées pour obtenir la prévision est proche (en dimension temporelle) de la donnée future plus la performance de prévision augmentera (en l'occurrence des données passées à t-5min). Il était alors prévisible que la performance prédictive du modèle de persistance soit meilleure à t-5min. On constate une mauvaise performance de ce modèle à partir d'un horizon de 15 min certainement due à plusieurs phénomènes météorologiques variant dans l'espace et le temps influençant la variabilité de l'ensoleillement. La mesure d'ensoleillement global à un instant t a une très faible probabilité de persister 15 min plus tard.

Station de La Désirade					
	5 min	10 min	15 min	30 min	1H
Modèle VAR spatio-temporel	20,35	19,65	19,15	18,31	18,37
Modèle de persistance	22,00	21,55	21,80	24,02	32,82
Station de Petit-canal					
	5 min	10 min	15 min	30 min	1 H
Modèle VAR spatio-temporel	17,82	17,67	17,48	17,41	18,42
Modèle de persistance	19,21	19,72	20,38	23,66	33,97
Station de Fouillole					
	5 min	10 min	15 min	30 min	1 H
Modèle VAR spatio-temporel	24,53	23,79	23,63	22,48	22,95
Modèle de persistance	26,24	26,12	26,06	27,18	35,94

Tableau 5-23. Les erreurs RMSE relatives en pourcentage pour chaque site du modèle VAR spatio-temporel pour des prévisions à un horizon de 5 min, 10 min, 15min, 30 min, 1H.

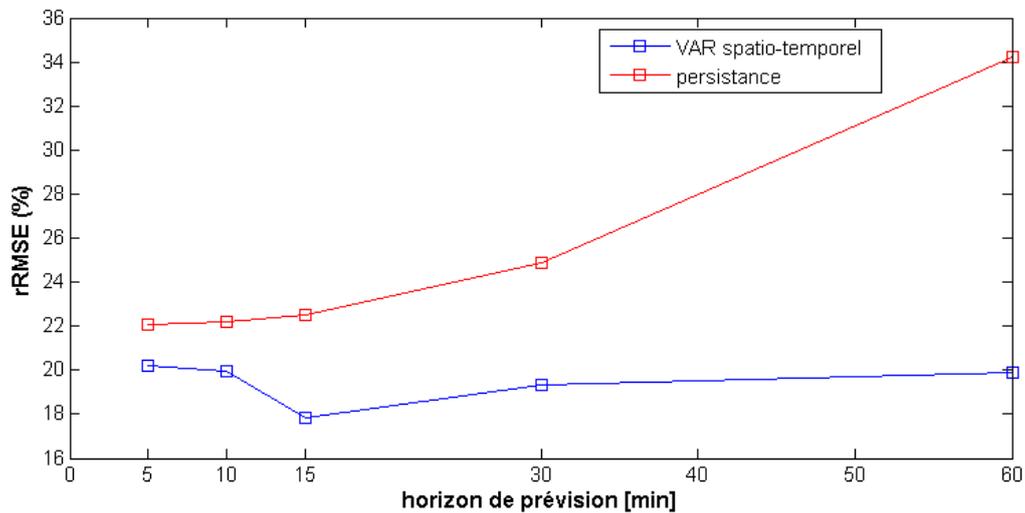


Figure 5-36 RMSE relative en pourcentage du modèle de persistance et du modèle VAR spatio-temporel en fonction de l'horizon de prévision.

Les résultats des erreurs statistiques de différents modèles trouvés dans la littérature sont bien souvent valables pour des horizons de prévision de 5 minutes et 15 minutes. Nous comparons ces évaluations aux résultats d'erreurs statistiques obtenus pour des prévisions du modèle VAR spatio-temporel et du modèle de persistance pour des horizons de 5 minutes et 15 minutes. Les résultats de prévisions de notre modèle spatio-temporel VAR(1) sont donnés par une moyenne des meilleures nMAE et nRMSE de tous les sites. Le ratio entre la RMSE du modèle et la RMSE de la persistance décrit dans Marquez et al (2013) et Dambreville et al (2014) est également présenté. La comparaison entre les performances prédictives (meilleure erreur) des modèles trouvés dans la littérature et de notre modèle est exposée dans le Tableau 5-25 pour différents horizons.

Dans l'ensemble, nous pouvons noter que le modèle spatio-temporel VAR donne de meilleurs résultats que les autres modèles validés dans la littérature (Tableau 5-24). Entre autre, comme nos résultats sont tirés de données mesurées en Guadeloupe, qui a un climat tropical fortement variable dans le temps, nous nous attendons à ce que la performance de notre méthode soit encore plus élevée à d'autres emplacements géographiques.

Sachant que les erreurs statistiques des prévisions dépendent fortement des conditions météorologiques et de la variabilité des journées d'ensoleillement prédites, cette comparaison n'est seulement qu'indicative pour l'évaluation de la performance de notre modèle de prévision. En perspective, nous appliquerons chacun de ces modèles à nos données d'ensoleillement afin d'avoir une comparaison plus valide.

Horizon de prévision	Modèle de prévision dans la littérature	Meilleure erreur statistique (%)
5 min	Modèle Krigeage	nRMSE :18,49
5 min	« shrinkage » dans le modèle VAR	nRMSE :18,09
5min	Regression	nMAE :12,63
5 min	ARIMA	nMAE : 13,21
5 min	Réseau de neurone	nMAE : 12,78
15 min	Regression	nMAE :26,52
15 min	ARIMA	nMAE :18,97
15 min	Réseau de neurone	nMAE :21,04
15 min	Modèle At	1-RMSE/RMSEp = 0.08
15 min	Modèle Ast	1-RMSE/RMSEp = 0.17
15 min	Modèle Ast2	1-RMSE/RMSEp = 0.20
Horizon de prévision	Modèle de prévision appliqué	Erreur statistique
5 min	Modèle de persistance	nRMSE :11,26
5 min	Modèle de persistance	nMAE : 9,54
5 min	Modèle VAR spatio-temporel	nRMSE : 7.33
5 min	Modèle VAR spatio-temporel	nMAE : 5,30
15 min	Modèle de persistance	nRMSE :19,4
15 min	Modèle de persistance	nMAE : 12,9
15 min	Modèle VAR spatio-temporel	nRMSE : 12,33
15 min	Modèle VAR spatio-temporel	1-RMSE/RMSEp = 0.36
15 min	Modèle VAR spatio-temporel	nMAE : 11,33

Tableau 5-25. Comparaison des erreurs de prévisions de différents modèles.

Les résultats de performance prédictive de plusieurs modèles de la littérature appliqués aux données de Guadeloupe sont présentés dans Lauret et al (2015). Ils ont testé plusieurs techniques prédictives tels que les modèles naïfs (modèle de persistance), les processus autorégressifs (AR), un modèle à réseau de neurones, les processus gaussiens, le modèle SVR. Nous avons donc comparé ces modèles à notre modèle VAR spatio-temporel en utilisant la RMSE relative (rRMSE). Cette comparaison est présentée dans le tableau suivant pour un horizon d'1h.

Modèles	rRMSE(%)	S(%)
V A R spatio-temporel	19,91	42,00
SVR	25,69	7,86
GP	25,95	6,92
Réseau de neurones	25,99	6,78
AR	26,74	4,09
persistance	34,24	0

Tableau 5-26. Comparaison des erreurs de prévisions des modèles présentés dans Lauret et al (2015) et de notre modèle VAR spatio-temporel.

Le modèle présente les meilleures performances avec 42% de skill score.

Cette performance constatée est principalement due au fait que l'utilisation de paramètres spatio-temporels apportent des informations supplémentaires et par conséquent améliorent la prédictibilité d'une station localisée par rapport aux modèles temporels n'utilisant que l'information sur une seule station. Ces résultats révèlent l'amélioration de la prévision de l'ensoleillement global par un processus multivarié et spatio-temporel.

IV.3. Limites du modèle VAR et perspectives

Les tests de prédiction du modèle VAR à différents horizons ont permis d'obtenir des résultats corrects avec une valeur maximale annuelle de RMSE de 21% obtenue pour une prévision à un horizon de 15 minutes. Néanmoins, le modèle nécessite un certain nombre de points d'initialisation pour prédire la journée. Ces points utilisés pour fournir la prédiction posent problème lorsqu'il s'agit de données horaires ou de données par pas de 30 minutes. En effet, les signaux par pas de 30 minutes et 1h correspondent respectivement à des signaux de 20 points et de 10 points pour une journée. En prenant 15 points d'initialisation, il y a une grande perte de données en début de journée.

La solution utilisée pour nos résultats présentés en section III.5.3 et section III.5.4, a été de récupérer 10 points d'initialisation sur la journée précédente et 5 points de la journée à prédire pour des séries temporelles par pas de 30min. De même, pour les séries temporelles par pas d'1h, 12 points de la journée précédentes et 3 points de la journée à prédire sont utilisés pour l'initialisation. On perd donc une journée de données pour initialiser le modèle. Une alternative serait de combiner le modèle VAR à d'autres modèles ne nécessitant que très peu de points d'initialisation afin d'obtenir une prévision sans trop de perte de données et permettre au gestionnaire du réseau d'avoir une prédiction de l'ensoleillement dès le début de la journée. Nous avons testé la combinaison des modèles AR et VAR

comme alternative. Un exemple est donné dans la Figure 5-36 pour une série de temps par pas de 30 min. Seuls les trois premiers points sont utilisés pour initialiser le modèle AR(1). Le modèle AR(1) prédit donc trois autres points puis le modèle VAR prend le relais pour la prévision des autres points.

On peut observer effectivement une mauvaise performance en début de journée, avec une NRMSE de 20,87% pour AR et VAR spatio-temporel combinés, au lieu de 19,13% pour le modèle VAR seul. Néanmoins, cette solution reste une alternative pour que le gestionnaire du réseau ait une prévision sur plus ou moins l'ensemble de la journée avec très peu de perte de données.

Une autre possibilité serait d'utiliser un nombre de points d'initialisation avec un historique de données conséquent pour commencer à prédire dès le premier point de la journée, même si les points d'initialisation n'appartiennent pas à la journée à prédire. L'étude précédemment faite sur les points d'initialisation montre qu'un mois donné peut suffire pour obtenir une bonne prédictibilité.

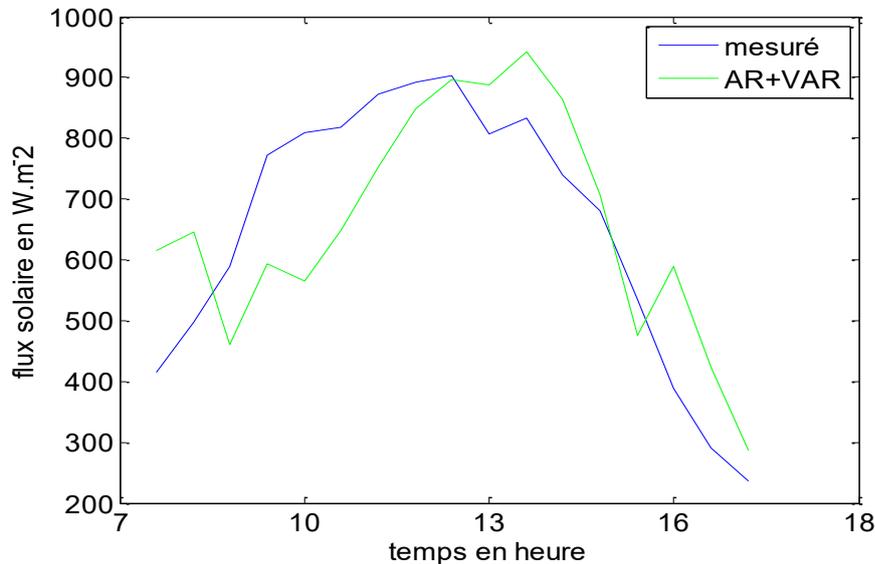


Figure 5-37 Signal observé et signal des valeurs prédites obtenues par combinaison linéaire des modèles AR et VAR

Une autre limite du modèle réside dans le fait que les séries temporelles ne montrent pas de relation linéaire forte entre les stations quelque soit le décalage temporel entre les stations (10 minutes, 20 minutes, 30 minutes et 40 minutes). On peut notamment l'observer dans un plan trivarié des données d'ensoleillement par pas de 10 minutes pour différents décalages temporels (Figure 5-37). Dans l'idée d'obtenir un modèle de prévision adaptatif, une solution serait d'utiliser des méthodes de prévision semi-paramétriques, des méthodes non linéaires. Récemment des méthodes semi-paramétriques d'estimation d'une régression qui améliorent l'approche non-paramétrique pure ont été proposées dans divers domaines. L'une d'elles, basée sur la notion de "directions révélatrices" appelée MAVE

(Moving Average -conditional- Variance Estimation), peut s'appliquer aux séries temporelles. Lefieux V. (2007) a étudié empiriquement son efficacité pour prédire les valeurs futures d'une série temporelle autorégressive. Il adopte ensuite cette méthode, d'un point de vue pratique, pour prédire la consommation électrique (Lefieux V. 2007, thèse) . La procédure d'estimation proposée se révèle efficace en pratique. Mehrota et Sharma (2007) expose un modèle semi-paramétrique pour une génération stochastique d'averse journalière sur plusieurs stations, pour une variabilité de basse fréquence. Dans la première étape, les occurrences des averses sont générées par un modèle Markov du premier ordre. La deuxième étape implique la génération de quantités d'averse journalières utilisant une approche basée sur l'estimation noyau qui est une méthode non paramétrique simulant la densité des averses, tout en assumant la dépendance Markovienne du premier ordre. Del Brio et al (2014) présentent un modèle VAR Semi-non paramétrique qui prédit pour des fonds spéculatifs pendant la crise. Des méthodes semi-non paramétriques capturent le risque précisément quand la volatilité est haute. La méthode de Gramme-Charlier incorpore des dépendances non-linéaires dans l'évaluation de risque de portefeuille. Il serait donc pertinent d'améliorer le modèle VAR spatio-temporel en étendant sa méthode à une approche non linéaire.

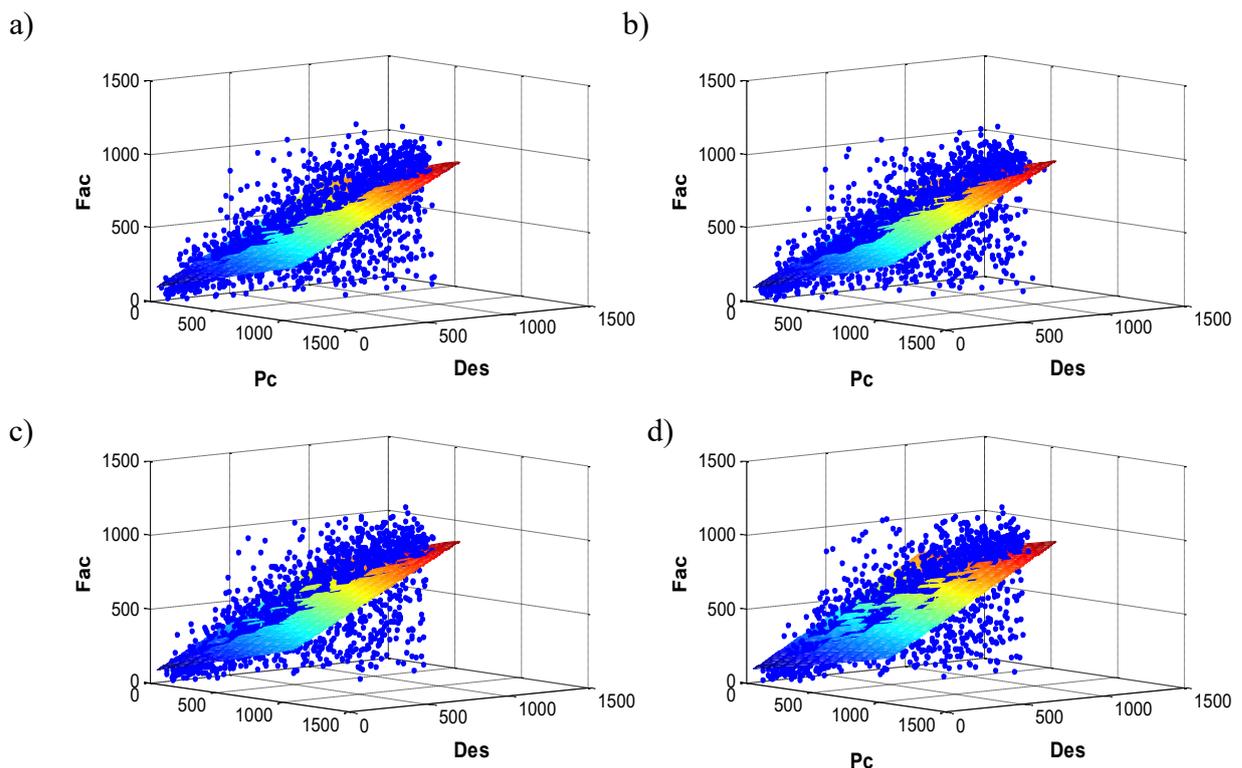


Figure 5-37 Séries de temps passées du flux solaire des sites de Petit-canal (Pc) et de Fouillole (Fac) au temps $[t-10mn ; t-20mn ; t-30mn ; t-40mn]$ en fonction des séries de temporelles de flux solaire de La Désirade (Des) à l'instant t . (exemple pour le mois de mars 2012). 1a) pour $t-10min$; 1b) pour $t-20min$; 1c) pour $t-30min$; 1d) pour $t-40min$.

V. Conclusion

Ces travaux de recherche présentent une nouvelle approche de prédiction lorsque les données spatio-temporelles disponibles sont riches en dimension temporelle et clairsemées dans la dimension spatiale. Le modèle VAR proposé est spécifiquement conçu pour un contexte spatio-temporel dont le processus consiste à fournir une prédiction à une localisation spatiale donnée à partir d'une somme pondérée de ses séries temporelles passées et des séries temporelles passées des localisations corrélées en utilisant des paramètres spatio-temporels. Les prédictions sont basées sur un minimum d'hypothèses puisqu'il traite chaque localisation spatiale séparément, ce qui permet d'appliquer le modèle sans hypothèse de stationnarité spatio-temporelle souvent difficile à démontrer pour des données environnementales et impossible à valider pour trois sites. Des études comparatives pour la prévision de flux solaire à plusieurs décalages temporels p et différentes structures spatiales sont présentées et discutées. Ces résultats ont permis l'optimisation du modèle en validant que le modèle VAR, à l'ordre 1 et utilisant une information physique respectant la direction prédominante du vent, est le plus performant et le plus approprié pour le système spatio-temporel composé de nos trois stations. L'étude de la performance prédictive en fonction des classes typiques du rayonnement solaire journalier ayant une variabilité spécifique a montré que malgré un haut niveau de dynamique des signaux de flux solaire, les erreurs statistiques (MAE et MBE) de prédictions sont principalement inférieures à 15%. Une comparaison avec le simple modèle de persistance et d'autres modèles trouvés dans la littérature a été également présentée. Notre modèle montre de meilleurs résultats, entre autre de plus faibles erreurs statistiques que celles du modèle VAR shrinkage avec une nRMSE de 18,09% au lieu de 10,44% pour notre modèle. Ainsi, l'intégration de notre modèle de prévision dans le processus du système de panneau solaire photovoltaïque offre une opportunité de fournir des garanties au gestionnaire de réseau d'énergie solaire. En perspective, nous comparerons ce modèle avec d'autres modèles de prévisions temporels appliqués à nos données afin de mettre en valeur l'influence de l'inclusion de paramètres spatiaux et temporels sur la précision de prédiction de l'ensoleillement global à une localisation. Des pistes d'amélioration du modèle ont également débutées.

Références

- Bessa, R., Trindade, A., Miranda, V., 2015, Spatial-temporal solar power forecasting for smart grids. *Industrial Informatics, IEEE Transactions on* 11 (1), 232–241.
- Brévignon C., 2005. L'environnement atmosphère de la Guadeloupe, de St Barthélémy et de St Martin, La direction Inter Régionale Antilles Guyane de Météo France.
- Charpentier A., 2002, Séries temporelles : Théorie et applications, Volume II, Université de Rennes
- Chow C. W., Urquhart B., Lave M., Dominguez A., Kleissl J., Shields J., Washom B., Intra-hour forecasting with a total sky imager at the uc san diego solar energy testbed, *Solar Energy* 85 (2011) 2881–2893
- Chu Y., Pedro H., Coimbra C., 2013, Hybrid intra-hour dni forecasts with sky image processing enhanced by stochastic learning, *Solar Energy*, p 592–603.
- Del Brio Esther B. , Mora-Valencia Andrés, 2014, Semi-non parametric VAR forecasts for hedge fund during the recent crisis, *Physica A* 401, 330–343
- Dickey, D., Fuller W. (1979): “Distribution of the Estimator for the autoregressive Time Series with a Unit Root,” *Journal of the American Statistical Association*, 74, 427–431.
- Faraway, J., *Practical Regression and ANOVA using R*, July 2002, <http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>
- Glasbey C., Allcroft D., A spatiotemporal auto-regressive moving average model for solar radiation, *Applied Statistics* 57 (2007) 343–355
- Hamilton J. D. *Time series analysis*, 1994, published by Princeton University Press
- Kyriakidis, P. C. and Journel, A. G. (1999). Geostatistical space-time models: a review. *Math. Geol.* 31, 651–684.
- Lefieux V., *Modèles semi-paramétriques appliqués à la prévision des séries temporelles. Cas de la consommation d'électricité*, 2007, Thèse de Mathématique, Université Rennes 2
- Lubrano M. , 2008, Test racine unitaire, Cours Chapitre III, Université de Marseille
- Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.
- Ma, Y. and Genton, M. G. (2000). Highly robust estimation of the autocovariance function. *J. Time Ser. Anal.* 21, 663–684

Marquez R., Coimbra M., Intra-hour dni forecast based on cloud tracking image analysis, *Solar Energy* 91 (2013) 327–336.

Mehrotra R, Sharma A., A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability, *Journal of Hydrology*, Volume 335, Issues 1–2, 8 March 2007, Pages 180-193

Pena, D., Tiao, G. C. and Tsay, R. S. (2001). *A Course in Time Series Analysis*. Wiley, New York.

Perez R., Kivalov J., Schlemmer, J. K. Hemker, D. Renne, T. Hoff, Validation of short and medium term operational solar radiation forecasts in the us, *Solar Energy* 84 (2010) 2161–2172.

Rakotomalala R., *Pratique de la Régression linéaire multiple*, Université Lyon 2, 2011, http://eric.univ-lyon2.fr/~ricco/cours/cours/La_regression_dans_la_pratique.pdf

Soubdhan T, Richard Emilion, Calif.R , février 2009, Classification of daily solar radiation distributions using a mixture of Dirichlet distributions, *Solar Energy* 83 p1056–1063

Tomson T., Tamm G., Short-term variability of solar radiation, 2005, *Solar Energy* 80 p600–606

Tonellato, S. F. (2001). A multivariate time series model for the analysis and prediction of carbon monoxide atmospheric concentrations. *Appl. Statist.* 50, 187-200.

Voyant C., November 2011, Time series prediction of global solar radiation and photovoltaic energy production using artificial neural networks, thèse, University of Corse-Pascal Paoli, UMR CNRS 6134(SPE)

Yang, D., Dong, Z., Reindl, T., Jirutitijaroen, P., Walsh, W. M., 2014. Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage. *Solar Energy* 103 (0), 550 – 562.

Conclusion générale

Sommaire

I. Bilan des résultats	241
II. Perspectives	243

I. Bilan des résultats

En dépit du faible nombre de stations de mesures disponible, nous avons développé une approche non classique par rapport à la littérature, en introduisant un nouveau modèle améliorant la prédictibilité de l'ensoleillement global à court-terme.

Le modèle spatio temporel a été développé à partir de données mesurées sur trois stations : Petit-canal (16°38N, 61°49W) sur la côte Est de la Guadeloupe dans le lieu-dit Gros Cap, Fouillole (16°26N, 61°24W) sur le Campus de l'Université des Antilles au Sud de la Grande-Terre, et le site le plus à l'Est est celui de La Désirade (16°31N, 61°55W), situé à l'extrémité Nord Est de La Désirade. Les distances entre les sites sont comprises entre 20km et 56km.

Dans le chapitre 2, l'état de l'art des techniques utilisées pour la prévision du rayonnement solaire a été étudié et décrit, ce qui nous a permis d'évaluer et de justifier le choix de notre modèle. Au regard des propriétés des modèles étudiés et présentés, notre prédiction spatio-temporelle à court terme de l'ensoleillement global se basera sur l'approche de Luna et Genton (2005) utilisant la méthodologie du modèle VAR (Vecteur Auto Régressif) dans le contexte de la prévision de concentration de particules.

Préliminairement, nous avons cherché à caractériser la variabilité de l'ensoleillement sur chacun des sites. Ainsi, une première analyse statistique descriptive a montré une variabilité du flux solaire sur des échelles différentes, allant de quelques secondes pour les fluctuations rapides, jusqu'aux variations annuelles.

Dans le chapitre 3, une première analyse de la variabilité du flux solaire a été effectuée par identification des classes de journées. Cette classification a été basée sur un algorithme k-means . Celle-ci étant non-supervisée, l'analyse à composante principale fonctionnelle a permis d'appuyer l'hypothèse du nombre de classes déterminé au préalable par rapport aux travaux de classification présentés dans la littérature. Les quatre classes journalières identifiées, sont caractérisées par quatre types de dynamiques différentes : dynamique lente avec peu de fluctuations (jour à ciel clair), dynamique moyenne (jour ciel clair avec passages nuageux), dynamique rapide (jour à ciel nuageux avec une petite période à ciel clair), dynamique lente mais variabilités fréquentes et de fortes amplitudes avec un ensoleillement faible (jour à ciel nuageux). Cette classification a permis de montrer une forte corrélation de l'évolution temporelle des classes de journées entre les trois sites.

L'analyse des signaux de fluctuations sur des courtes échelles de temps, dans le chapitre 4, a permis de caractériser la dynamique de l'ensoleillement. Les résultats de cette analyse ont établi que 90% des fluctuations ont une amplitude inférieure à 200 W.m^{-2} . Le temps caractéristique le plus probable de ces fluctuations est entre 8 et 20 secondes et peut atteindre une durée de 2300 secondes soit près de 38 minutes. On observe ces mêmes caractéristiques dynamiques sur les trois sites de l'étude.

L'analyse de la fraction d'énergie (rapport entre le flux moyen journalier mesuré et le flux moyen journalier théorique), qui rend compte de la fraction d'énergie disponible par rapport à l'énergie attendue en absence de nuage, a révélé une plus grande variabilité de flux solaire soit un plus grand nombre de passages nuageux sur la station de Fouillole comparé à Petit-canal et La Désirade. Ces premiers résultats ont montré une même dynamique d'évolution dans le temps entre les 3 stations mais une différence d'un point de vue énergétique.

Nous avons par la suite, effectué une analyse statistique des corrélations spatio-temporelles entre les stations. La première analyse, a permis d'évaluer les corrélations spatiales intersites. Pour cela, un parallèle avec les travaux de Pérez décrit dans la littérature, qui utilise le coefficient de Pearson, est présenté. L'objectif a été de définir l'évolution des corrélations de paires de stations en fonction des distances et de l'intervalle d'échantillonnage des mesures considéré. Cette même étude a été effectuée avec la fonction d'inter corrélation, qui permet d'évaluer les corrélations, pour une distance et un échantillonnage donnés, en faisant varier un décalage temporel entre les signaux. Trois principaux commentaires peuvent être donnés. Premièrement, la corrélation spatiale est plus significative si on tient compte du décalage temporel. Deuxièmement, la corrélation spatiale n'est pas fonction des distances inter sites quelque soit l'intervalle d'échantillonnage. Troisièmement, la corrélation prédominante pour la paire de station Petit-canal/la Désirade suggère que dans notre contexte spatio-temporel, un phénomène physique est prédominant par rapport aux paramètres des distances. Ce phénomène est lié à la direction prédominante du vent. Ce qui est une information indicative pour l'élaboration du modèle de prévision spatio-temporel.

Une autre analyse a consisté à évaluer la corrélation temporelle intersite. Une analyse statistique des corrélations avec séparation des échelles de temps a donc été décrite. Les résultats des densités de probabilités des coefficients maximaux d'intercorrélation avec séparation d'échelles de temps [1 min ;15 min], [15 min ; 30 min], [30 min ;1 h], [1h ;2 h], [2 h ;3 h] , a permis de déterminer une échelle de temps seuil où les corrélations temporelles entre les sites sont significatives. Cette échelle de temps seuil est comprise entre 1h et 2h. Ce qui consiste à dire que les phénomènes liés aux fluctuations qui peuvent survenir et durer sur une grande échelle de temps ont de forte chance d'être retrouvé sur une autre station. Ces analyses spatio-temporelles statistiques montrent une évolution des corrélations entre les sites qui va dépendre du temps et de l'espace.

Ces études spatio-temporelles ont démontré qu'il y a absence de corrélation instantanée mais présence de corrélation avec décalage temporel.

Avec peu de stations (soit 3 stations) et des distances non uniformes, imposant un contenu informatif faible, nous avons cherché à définir un nouveau modèle pour la prévision de l'ensoleillement global à court terme, qui diffère des approches classiques présentées dans la littérature.

Dans le chapitre 5, le modèle VAR proposé consiste à fournir une prédiction, à une localisation spatiale donnée, à partir d'une combinaison linéaire pondérée de la série temporelle passée du site considéré, et des séries temporelles passées des localisations corrélées en utilisant des paramètres spatiaux et temporels. Chaque localisation spatiale est traitée séparément, ce qui permet d'appliquer le modèle sans hypothèse de stationnarité spatio-temporelle souvent difficile à démontrer pour des données environnementales et impossible à valider pour trois sites.

Nous nous sommes ensuite intéressés à l'optimisation de la structure spatiale et temporelle du modèle. Nous avons ainsi testé un arrangement spatial fonction des distances inter sites et fonction de la direction prédominante du vent. Nous avons également testé des décalages temporels (ordre du modèle) allant de 10 à 40 minutes. Il en ressort que le modèle VAR, à l'ordre 1 utilisant un arrangement spatial respectant la direction prédominante du vent, est le plus performant et le plus approprié pour le système composé de nos trois stations. Les erreurs statistiques (MAE et MBE) de prédictions sont principalement inférieures à 15%.

Lorsque le modèle est appliqué aux classes de journées d'ensoleillement identifiées, les résultats sont compris entre 8% et 24% RMSE normalisé pour les classes ciel clair à ciel nuageux respectivement.

Les résultats obtenus ont été comparés aux modèles de prévisions classique de la littérature, entre autres le modèle de persistance, ARIMA, réseaux de neurones, krigeage, modèles Ast et Ast 2 , a été également présentée. Notre modèle a montré de meilleurs résultats.

II.Perspectives

Au terme de cette étude menée, nous prévoyons d'appliquer ce modèle à des données de production d'électricité d'origine photovoltaïque. L'intérêt de ce modèle est qu'il permet de faire une prévision à court terme du productible photovoltaïque agrégé à partir d'un ensemble de sites distincts.

Néanmoins, malgré une performance prédictive notoire, il serait judicieux d'appliquer préalablement les modèles de la littérature à nos données (réseaux de neurones, modèle ARIMA, modèle Ast2,...) afin d'effectuer un benchmarking qui validera la comparaison des performances. De plus, cette comparaison vérifiera l'influence de l'inclusion de paramètres spatiaux et temporels sur la précision de prédiction de l'ensoleillement global à une localisation.

Nous nous intéresserons également à évaluer notre modèle pour des stations plus nombreuses. L'idée serait d'analyser sa performance en fonction de la quantité d'informations données. On pourrait se demander si le modèle diminue en performance, pour la prévision de l'ensoleillement à court terme sur une localisation, avec plus de stations.

En tenant compte de toutes les améliorations qu'on peut apporter aux paramètres du modèle, discutées dans le chapitre V, il serait également intéressant d'étendre le modèle à une hypothèse plus complexe tel qu'un modèle non linéaire (bilinéaire, polynomial).

Pour finir, appliquer le modèle au foisonnement de la production PV reste une perspective notable. L'objectif à terme est d'améliorer la prévision de production d'électricité à partir du photovoltaïque pour repousser les limites d'intégration de cette énergie dite fatale.

Annexe

Analyse en composante principale fonctionnelle

I. Transformations des données brutes en données fonctionnelles

Afin d'analyser les composantes principales du flux solaire, nous transformerons premièrement les données discrètes en données fonctionnelles. Pour cela, nous utilisons une méthode de lissage qui s'effectue par un modèle de régression basée sur une combinaison linéaire de fonctions de bases : polynomiales, ondelettes, Fourier, Spline... (Chebana et al (2012)).

Soit $Y_i = (y_i(t_1), \dots, y_i(t_T))$, $i=1, \dots, n$ étant un ensemble de n observations discrètes où chaque $t_j \in C \subset \mathbb{R}^+$, $j=1, \dots, T$ est le $j^{\text{ème}}$ point temporel enregistré à partir d'un sous-ensemble de temps donné C . Pour une observation fixée i , chaque ensemble de mesures $(y_i(t_1), \dots, y_i(t_T))$ est convertie en données fonctionnelles notées $y_i(t)$ où l'indice t couvre le sous-ensemble continu C . On suppose que l'observation discrète $(y_i(t_j))_{j=1, \dots, T}$ est le modèle de régression :

$$y_i(t_j) = x_i(t_j) + \epsilon_{ij} \quad i=1, \dots, n \quad j=1, \dots, T \quad (1)$$

où ϵ_{ij} sont les erreurs et les fonctions $x_i(\cdot)$ sont des combinaisons linéaires de fonctions de bases $\phi_k(\cdot)$, ce qui permet d'expliquer la variation contenue dans les observations fonctionnelles :

$$x_i(t) = \sum_{k=0}^p c_k^i \phi_k(t), \quad t \in C \quad (2)$$

L'ensemble des données fonctionnelles $(y_i(t))_{i=1, \dots, n}$ sont alors données par :

$$y_i(t) = \hat{x}_i(t) = \sum_{k=0}^p \hat{c}_k^i \phi_k(t), \quad t \in C \quad (3)$$

où les coefficients estimés \hat{c}_k^i sont obtenus par minimisation de la somme suivante :

$$SSE(i) = \sum_{j=1}^T (y_i(t_j) - x_i(t_j))^2, \quad i = 1, \dots, n \quad (4)$$

Pour notre cas d'étude, nous utiliserons la fonction de Fourier qui est recommandée pour les fonctions périodiques puisqu'elle capture la saisonnalité des données (les données de flux solaire présentent une variabilité saisonnière et une périodicité sur le cycle journalier, mensuel et annuel). Les fonctions de base de Fourier $(\phi_k(\cdot))_{k=0, \dots, p}$ sont définies par :

$$\phi_0(t) = 1, \phi_{2j-1}(t) = \sin(jwt), \quad \phi_{2j}(t) = \cos(jwt), \quad w = 2\pi/T. \quad (5)$$

L'inspection des courbes résultant de la méthode de lissage est recommandée pour s'assurer que celles-ci capturent la majorité des points et les points spéciaux tels que les pics.

Une fois les données discrètes sont transformées en courbe, elles peuvent être analysées selon une structure fonctionnelle. Le domaine fonctionnel est plus général et plus flexible et peut représenter une large variété de courbes de flux solaire. Par projection des données fonctionnelles, de nouvelles variables appelées scores vont résumer au mieux l'information contenue.

II. Extraction des harmoniques

Les harmoniques sont les principales composantes obtenues par maximisation de la variance des scores correspondants. La méthode extrait des composantes fonctionnelles représentatives des mesures, et évalue l'adéquation de chaque donnée à ces composantes sous la forme de scores. Les coordonnées (en anglais "scores") des observations et des variables sont les projections sur les vecteurs propres:

Soit $y_i(t), i = 1, \dots, n$ les données fonctionnelles obtenues par lissages des données discrètes $(y_i(t_1), \dots, y_i(t_T)), i = 1, \dots, n$

Les données fonctionnelles centrées sont alors telles que : $(y_i^*(t) = y_i(t) - \bar{y}(t))_{i=1, \dots, n}$ avec $\bar{y}(t)$ la fonction moyenne de l'échantillon $y_1(t), \dots, y_n(t)$.

La fonction moyenne de l'échantillon étant :

$$\bar{y}(t) = \sum_{i=1}^n \frac{1}{n} y_i(t) \quad t \in C \quad (6)$$

Une analyse à composante principale fonctionnelle est alors appliquée à $(y_i^*(t))$ pour créer des harmoniques qui révèlent le plus important type de variations contenues dans les données.

La première principale composante fonctionnelle de $(y_i^*(t))_{i=1,\dots,n}$ notée par $w_1(t)$ est une fonction telle que la variance des scores correspondants est :

$$z_{i,1} = \int_C w_1(s) y_i^*(s) ds, i = 1, \dots, n \quad (7)$$

est maximisée sous la contrainte que : $\int_C w_1(s)^2 ds = 1$.

Les composantes principales suivantes $w_k(t)$ sont obtenues par maximisation de la variance des scores correspondants $z_{i,k}$:

$$z_{i,k} = \int_C w_k(s) y_i^*(s) ds, i = 1, \dots, n \quad (8)$$

Les quatre harmoniques obtenues à partir des scores résument en quatre fonctions les informations les plus importantes de la matrice correspondant à 3 années de flux solaires par pas de 1 minute. Les caractéristiques des données solaires sont représentées par une simplification maximale du phénomène réel. En d'autres termes, elles représentent les quatre types de journées ayant les quatre plus fortes probabilités d'apparition dans la séquence totale de mesures (3ans).

Résultats du modèle pour plusieurs horizons de prévisions

Horizon à 5 minutes

	p=1	p=2	p=3	p=4
Jan	0,08	0,10	0,12	0,14
Fev	0,08	0,17	0,18	0,22
Mar	0,07	0,21	0,23	0,26
Avr	0,08	0,11	0,14	0,17
Mai	0,08	0,12	0,14	0,17
Juin	0,09	0,37	0,34	0,37
Juil				
Août	0,10	0,12	0,14	0,21
Sept	0,07	0,10	0,13	0,32
Oct	0,10	0,20	0,24	0,44
Nov	0,09	0,24	0,28	0,55
Dec	0,10	0,13	0,15	0,18
Moyenne annuelle	0,09	0,19	0,19	0,27

Tab1. RMSE normalisée (nRMSE) pour différents temps de retard p pour la prédiction du flux solaire à 5 min sur la Désirade 2010

	p=1	p=2	p=3	p=4
Jan	0,13	0,17	0,26	0,36
Fev	0,18	0,26	0,38	0,45
Mar	0,26	0,33	0,45	0,51
Avr	0,08	0,10	0,12	0,14
Mai	0,09	0,12	0,14	0,17
Juin	0,10	0,15	0,17	0,20
Juil	0,12	0,17	0,19	0,22
Août	0,08	0,11	0,13	0,41
Sept	0,10	0,14	0,51	0,19
Oct	0,13	0,16	0,19	0,25
Nov	0,19	0,23	0,27	0,34
Dec	0,11	0,17	0,25	0,21
Moyenne annuelle	0,13	0,17	0,26	0,29

Tab2. RMSE normalisée (nRMSE) pour différents temps de retard p pour la prédiction du flux solaire à 5 min sur la Désirade 2011

	p=1	p=2	p=3	p=4
Jan	0,11	0,14	0,17	0,20
Fev	0,10	0,19	0,21	0,26
Mar	0,10	0,21	0,23	0,28
Avr	0,10	0,13	0,17	0,19
Mai	0,09	0,13	0,14	0,17
Juin	0,08	0,15	0,17	0,19
Juil	0,07	0,07	0,22	0,23
Août	0,08	0,11	0,13	0,15
Sept	0,09	0,12	0,14	0,15
Oct	0,11	0,18	0,21	0,23
Nov	0,12	0,22	0,25	0,27
Dec	0,11	0,14	0,17	0,21
Moyenne annuelle	0,10	0,15	0,18	0,21

Tab3. RMSE normalisée (nRMSE) pour différents temps de retard p pour la prédiction du flux solaire à 5 min sur la Désirade 2012

	p=1	p=2	p=3	p=4
Jan	0,07	0,10	0,11	0,14
Fev	0,08	0,15	0,17	0,20
Mar	0,10	0,20	0,22	0,25
Avr	0,08	0,11	0,14	0,15
Mai	0,07	0,10	0,11	0,14
Juin	0,14	0,36	0,57	0,36
Juil				
Août	0,08	0,10	0,15	0,42
Sept	0,11	0,13	0,15	0,17
Oct	0,08	0,18	0,22	0,43
Nov	0,08	0,23	0,27	0,50
Dec	0,10	0,13	0,16	0,18
Moyenne annuelle	0,09	0,16	0,21	0,27

Tab4. RMSE normalisée (nRMSE) pour différents temps de retard p pour la prédiction du flux solaire à 5 min sur Petit-canal 2010

	p=1	p=2	p=3	p=4
Jan	0,15	0,13	0,15	0,17
Fev	0,16	0,24	0,36	0,42
Mar	0,23	0,30	0,41	0,46
Avr	0,07	0,15	0,12	0,14
Mai	0,08	0,11	0,13	0,16
Juin	0,10	0,14	0,16	0,19
Juil	0,12	0,17	0,19	0,22
Août	0,09	0,12	0,14	0,15
Sept	0,09	0,12	0,13	0,18
Oct	0,11	0,14	0,18	0,23
Nov	0,18	0,21	0,25	0,32
Dec	0,10	0,13	0,16	0,21
Moyenne annuelle	0,12	0,16	0,20	0,24

Tab5. RMSE normalisée (nRMSE) pour différents temps de retard p pour la prédiction du flux solaire à 5 min sur Petit-canal 2011

	p=1	p=2	p=3	p=4
Jan	0,10	0,13	0,16	0,20
Fev	0,10	0,18	0,20	0,25
Mar	0,08	0,19	0,21	0,26
Avr	0,08	0,11	0,13	0,15
Mai	0,06	0,08	0,10	0,18
Juin	0,06	0,14	0,16	0,18
Juil	0,07	0,15	0,20	0,22
Août	0,07	0,10	0,11	0,13
Sept	0,08	0,10	0,12	0,14
Oct	0,09	0,17	0,19	0,22
Nov	0,10	0,20	0,23	0,26
Dec	0,12	0,15	0,18	0,25
Moyenne annuelle	0,09	0,14	0,17	0,20

Tab6. RMSE normalisée (nRMSE) pour différents temps de retard p pour la prédiction du flux solaire à 5 min sur Petit-canal 2012

	p=1	p=2	p=3	p=4
Jan	0,09	0,12	0,14	0,18
Fev	0,10	0,15	0,16	0,20
Mar	0,09	0,20	0,21	0,24
Avr	0,08	0,13	0,16	0,20
Mai	0,11	0,40	0,48	0,53
Juin	0,10	0,45	0,52	0,52
Juil				
Août	0,11	0,15	0,17	0,32
Sept	0,10	0,15	0,19	0,61
Oct	0,08	0,20	0,23	0,23
Nov	0,09	0,24	0,28	0,65
Dec	0,11	0,14	0,20	0,27
Moyenne annuelle	0,10	0,22	0,26	0,36

Tab7. RMSE normalisée (nRMSE) pour différents temps de retard p pour la prédiction du flux solaire à 5 min sur Fouillole 2010

	p=1	p=2	p=3	p=4
Jan	0,11	0,18	0,31	0,38
Fev	0,17	0,25	0,37	0,44
Mar	0,18	0,37	0,49	0,56
Avr	0,09	0,14	0,22	0,22
Mai	0,08	0,12	0,15	0,18
Juin	0,10	0,14	0,16	0,19
Juil	0,13	0,17	0,19	0,22
Août	0,11	0,15	0,18	0,23
Sept	0,08	0,11	0,15	0,20
Oct	0,12	0,15	0,19	0,25
Nov	0,17	0,21	0,25	0,31
Dec	0,12	0,17	0,20	0,25
Moyenne annuelle	0,12	0,18	0,24	0,29

Tab8. RMSE normalisée (nRMSE) pour différents ordre p pour la prédiction du flux solaire à 5 min sur Fouillole 2011

	p=1	p=2	p=3	p=4
Jan	0,14	0,19	0,22	0,28
Fev	0,11	0,19	0,21	0,26
Mar	0,12	0,20	0,22	0,27
Avr	0,11	0,16	0,53	0,65
Mai	0,09	0,13	0,15	0,17
Juin	0,08	0,14	0,16	0,18
Juil	0,10	0,16	0,22	0,24
Août	0,10	0,23	0,23	0,32
Sept	0,08	0,12	0,15	0,17
Oct	0,10	0,17	0,19	0,22
Nov	0,11	0,21	0,24	0,27
Dec	0,12	0,17	0,20	0,27
Moyenne annuelle	0,10	0,17	0,32	0,27

Tab9. RMSE normalisée (nRMSE) pour différents temps de retard p pour la prédiction du flux solaire à 5 min sur Fouillole 2012

Horizon à 15 minutes

	p=1	p=2	p=3	p=4
Jan	0,17	0,17	0,17	0,19
Fev	0,16	0,21	0,23	0,27
Mar	0,16	0,25	0,27	0,30
Avr	0,15	0,14	0,15	0,19
Mai	0,15	0,17	0,22	0,27
Juin	0,15	0,22	0,23	0,28
Juil				
Août	0,15	0,18	0,25	0,42
Sept	0,25	0,14	0,16	0,21
Oct	0,21	0,30	0,36	0,41
Nov	0,19	0,33	0,40	0,45
Dec	0,22	0,20	0,27	0,30
Moyenne annuelle	0,20	0,21	0,25	0,31

Tab10. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 15 min sur La Désirade 2010

	p=1	p=2	p=3	p=4
Jan	0,22	0,20	0,42	0,30
Fev	0,19	0,30	0,50	0,44
Mar	0,16	0,38	0,56	0,48
Avr	0,12	0,11	0,40	0,16
Mai	0,14	0,14	0,18	0,22
Juin	0,14	0,18	0,22	0,25
Juil	0,15	0,16	0,18	0,21
Août	0,15	0,14	0,16	0,22
Sept	0,16	0,16	0,19	0,22
Oct	0,15	0,21	0,27	0,33
Nov	0,20	0,29	0,36	0,44
Dec	0,22	0,25	0,32	0,41
Moyenne annuelle	0,17	0,21	0,31	0,31

Tab11. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 15 min sur La Désirade 2011

	p=1	p=2	p=3	p=4
Jan	0,20	0,28	0,33	0,28
Fev	0,20	0,18	0,20	0,43
Mar	0,15	0,14	0,17	0,46
Avr	0,13	0,15	0,17	0,20
Mai	0,15	0,17	0,21	0,27
Juin	0,12	0,17	0,23	0,25
Juil	0,12	0,18	0,20	0,22
Août	0,13	0,14	0,17	0,24
Sept	0,15	0,14	0,19	0,26
Oct	0,15	0,21	0,27	0,32
Nov	0,19	0,25	0,30	0,36
Dec	0,16	0,18	0,25	0,30
Moyenne annuelle	0,15	0,18	0,22	0,30

Tab12. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 15 min sur La Désirade 2012

	p=1	p=2	p=3	p=4
Jan	0,17	0,18	0,21	0,26
Fev	0,15	0,19	0,21	0,24
Mar	0,14	0,24	0,27	0,29
Avr	0,16	0,18	0,18	0,22
Mai	0,16	0,30	0,40	0,59
Juin	0,14	0,21	0,23	0,27
Juil				
Août	0,16	0,19	0,21	0,25
Sept	0,16	0,17	0,20	0,30
Oct	0,19	0,27	0,32	0,37
Nov	0,17	0,32	0,38	0,44
Dec	0,18	0,18	0,20	0,46
Moyenne annuelle	0,16	0,22	0,25	0,34

Tab13. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 15 min sur Petit-canal 2010

	p=1	p=2	p=3	p=4
Jan	0,15	0,17	0,34	0,50
Fev	0,18	0,28	0,46	0,40
Mar	0,14	0,34	0,50	0,43
Avr	0,14	0,14	0,48	0,19
Mai	0,15	0,15	0,18	0,25
Juin	0,13	0,17	0,20	0,24
Juil	0,13	0,19	0,23	0,26
Août	0,16	0,15	0,17	0,20
Sept	0,15	0,13	0,18	0,25
Oct	0,16	0,20	0,25	0,31
Nov	0,22	0,27	0,33	0,40
Dec	0,24	0,23	0,36	0,60
Moyenne annuelle	0,16	0,20	0,31	0,35

Tab14. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 15 min sur Petit-canal 2011

	p=1	p=2	p=3	p=4
Jan	0,22	0,26	0,25	0,26
Fev	0,19	0,16	0,23	0,41
Mar	0,16	0,13	0,17	0,42
Avr	0,14	0,13	0,16	0,18
Mai	0,15	0,16	0,18	0,20
Juin	0,11	0,20	0,21	0,23
Juil	0,12	0,15	0,25	0,27
Août	0,14	0,13	0,16	0,21
Sept	0,12	0,20	0,17	0,20
Oct	0,25	0,23	0,25	0,30
Nov	0,17	0,18	0,28	0,33
Dec	0,20	0,18	0,22	0,23
Moyenne annuelle	0,17	0,18	0,21	0,27

Tab15. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 15 min sur Petit-canal 2012

	p=1	p=2	p=3	p=4
Jan	0,17	0,17	0,19	0,23
Fev	0,18	0,18	0,21	0,23
Mar	0,15	0,16	0,26	0,29
Avr	0,17	0,19	0,21	0,24
Mai	0,17	0,21	0,20	0,25
Juin	0,15	0,22	0,23	0,27
Juil				
Août	0,20	0,24	0,24	0,56
Sept	0,21	0,29	0,29	0,34
Oct	0,26	0,20	0,34	0,40
Nov	0,20	0,34	0,40	0,46
Dec	0,26	0,23	0,29	0,38
Moyenne annuelle	0,19	0,18	0,26	0,33

Tab16. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 15 min sur Fouillole 2010

	p=1	p=2	p=3	p=4
Jan	0,28	0,28	0,47	0,39
Fev	0,21	0,30	0,49	0,43
Mar	0,20	0,42	0,62	0,54
Avr	0,14	0,18	0,20	0,26
Mai	0,15	0,16	0,20	0,24
Juin	0,19	0,16	0,25	0,23
Juil	0,15	0,21	0,19	0,29
Août	0,18	0,17	0,21	0,27
Sept	0,17	0,16	0,26	0,27
Oct	0,20	0,21	0,33	0,33
Nov	0,21	0,27	0,42	0,40
Dec	0,24	0,29	0,24	0,51
Moyenne annuelle	0,19	0,23	0,36	0,35

Tab24. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 15 min sur Fouillole 2011

	p=1	p=2	p=3	p=4
Jan	0,27	0,51	0,36	0,47
Fev	0,24	0,23	0,28	0,43
Mar	0,19	0,19	0,24	0,45
Avr	0,16	0,16	0,21	0,24
Mai	0,15	0,16	0,22	0,25
Juin	0,13	0,16	0,21	0,23
Juil	0,14	0,22	0,27	0,30
Août	0,14	0,14	0,20	0,42
Sept	0,14	0,15	0,20	0,24
Oct	0,18	0,20	0,25	0,30
Nov	0,20	0,25	0,30	0,35
Dec	0,26	0,25	0,32	0,48
Moyenne annuelle	0,18	0,24	0,25	0,35

Tab25. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 15 min sur Fouillole 2012

Horizon à 30 minutes

	p=1	p=2	p=3	p=4
Jan	0,13	0,15	0,21	0,67
Fev	0,15	0,14	0,29	0,59
Mar	0,19	0,17	0,32	0,56
Avr	0,06	0,13	0,17	0,26
Mai	0,07	0,09	0,12	0,23
Juin	0,10	0,15	0,20	0,70
Juil				
Août	0,07	0,12	0,14	0,16
Sept	0,09	0,13	0,17	0,24
Oct	0,13	0,20	0,29	0,31
Nov	0,17	0,24	0,34	0,36
Dec	0,06	0,10	0,13	0,42
Moyenne annuelle	0,11	0,15	0,21	0,41

Tab26. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur La Désirade 2010

	p=1	p=2	p=3	p=4
Jan	0,13	0,18	0,23	0,35
Fev	0,20	0,27	0,22	0,34
Mar	0,26	0,33	0,23	0,38
Avr	0,06	0,11	0,16	0,21
Mai	0,07	0,10	0,11	0,17
Juin	0,10	0,14	0,22	0,25
Juil	0,12	0,16	0,24	0,28
Août	0,06	0,10	0,14	0,67
Sept	0,09	0,12	0,15	0,25
Oct	0,13	0,18	0,32	0,60
Nov	0,20	0,26	0,43	0,76
Dec	0,11	0,13	0,24	0,22
Moyenne annuelle	0,13	0,17	0,22	0,37

Tab27. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur La Désirade 2011

	p=1	p=2	p=3	p=4
Jan	0,07	0,12	0,16	0,27
Fev	0,12	0,17	0,24	0,37
Mar	0,16	0,21	0,28	0,42
Avr	0,07	0,10	0,16	0,35
Mai	0,08	0,10	0,13	0,22
Juin	0,09	0,13	0,16	0,20
Juil	0,13	0,17	0,20	0,25
Août	0,06	0,09	0,25	0,35
Sept	0,08	0,11	0,13	0,20
Oct	0,14	0,16	0,20	0,55
Nov	0,18	0,19	0,24	0,57
Dec	0,09	0,12	0,17	0,67
Moyenne annuelle	0,11	0,14	0,19	0,37

Tab28. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur La Désirade 2012

	p=1	p=2	p=3	p=4
Jan	0,16	0,17	0,28	0,38
Fev	0,14	0,13	0,27	0,52
Mar	0,19	0,17	0,32	0,53
Avr	0,09	0,12	0,15	0,21
Mai	0,10	0,11	0,16	0,21
Juin	0,09	0,15	0,20	0,70
Juil				
Août	0,09	0,15	0,19	0,78
Sept	0,09	0,12	0,16	0,27
Oct	0,12	0,18	0,26	0,28
Nov	0,16	0,23	0,33	0,34
Dec	0,12	0,14	0,18	0,33
Moyenne annuelle	0,12	0,15	0,23	0,41

Tab29. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur Petit-canal 2010

	p=1	p=2	p=3	p=4
Jan	0,17	0,31	0,25	0,59
Fev	0,19	0,25	0,20	0,32
Mar	0,24	0,30	0,20	0,34
Avr	0,10	0,13	0,19	0,21
Mai	0,09	0,13	0,19	0,23
Juin	0,09	0,13	0,21	0,23
Juil	0,11	0,16	0,24	0,27
Août	0,12	0,15	0,24	0,29
Sept	0,24	0,29	0,40	0,38
Oct	0,13	0,17	0,31	0,38
Nov	0,18	0,23	0,39	0,40
Dec	0,14	0,18	0,31	0,40
Moyenne annuelle	0,15	0,20	0,20	0,33

Tab30. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur Petit-canal 2011

	p=1	p=2	p=3	p=4
Jan	0,12	0,23	0,30	0,61
Fev	0,12	0,16	0,23	0,35
Mar	0,14	0,19	0,25	0,37
Avr	0,11	0,19	0,19	0,26
Mai	0,09	0,13	0,16	0,22
Juin	0,08	0,12	0,15	0,18
Juil	0,12	0,16	0,19	0,23
Août	0,11	0,17	0,22	0,33
Sept	0,09	0,14	0,29	0,70
Oct	0,13	0,14	0,18	0,50
Nov	0,17	0,18	0,22	0,53
Dec	0,11	0,15	0,23	0,24
Moyenne annuelle	0,12	0,16	0,22	0,38

Tab31. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur Petit-canal 2012

	p=1	p=2	p=3	p=4
Jan	0,12	0,12	0,27	0,49
Fev	0,13	0,13	0,26	0,51
Mar	0,18	0,16	0,31	0,44
Avr	0,08	0,12	0,20	0,75
Mai	0,06	0,11	0,16	0,78
Juin	0,10	0,15	0,20	0,23
Juil				
Août	0,09	0,13	0,17	0,23
Sept	0,10	0,15	0,24	0,25
Oct	0,13	0,19	0,30	0,30
Nov	0,18	0,25	0,35	0,37
Dec	0,15	0,21	0,23	0,55
Moyenne annuelle	0,12	0,16	0,24	0,47

Tab32. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur Fouillole 2010

	p=1	p=2	p=3	p=4
Jan	0,11	0,18	0,22	0,52
Fev	0,20	0,27	0,22	0,34
Mar	0,29	0,36	0,25	0,42
Avr	0,10	0,17	0,22	0,39
Mai	0,06	0,10	0,18	0,22
Juin	0,09	0,13	0,20	0,23
Juil	0,12	0,16	0,25	0,28
Août	0,10	0,13	0,18	0,78
Sept	0,09	0,15	0,29	0,52
Oct	0,13	0,17	0,31	0,58
Nov	0,18	0,23	0,39	0,69
Dec	0,12	0,19	0,21	0,28
Moyenne annuelle	0,13	0,19	0,24	0,44

Tab33. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur Fouillole 2011

	p=1	p=2	p=3	p=4
Jan	0,11	0,16	0,24	0,39
Fev	0,13	0,17	0,24	0,37
Mar	0,16	0,21	0,27	0,40
Avr	0,11	0,15	0,22	0,33
Mai	0,06	0,10	0,14	0,18
Juin	0,08	0,12	0,15	0,19
Juil	0,13	0,17	0,20	0,25
Août	0,07	0,11	0,13	0,59
Sept	0,09	0,10	0,14	0,41
Oct	0,14	0,15	0,19	0,52
Nov	0,18	0,19	0,23	0,56
Dec	0,08	0,15	0,26	0,30
Moyenne annuelle	0,11	0,15	0,20	0,37

Tab34. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur Fouillole 2012

Horizon à 1 heure

	p=1	p=2	p=3	p=4
Jan	0,14	0,41	0,39	0,50
Fev	0,14	0,20	0,31	0,61
Mar	0,19	0,24	0,31	0,67
Avr	0,08	0,16	0,16	0,10
Mai	0,07	0,11	0,23	0,14
Juin	0,13	0,14	0,15	0,70
Juil				
Août	0,11	0,18	0,49	0,72
Sept	0,10	0,43	0,73	0,73
Oct	0,18	0,49	0,68	0,76
Nov	0,23	0,12	0,17	0,13
Dec	0,09	0,41	0,39	0,50
Moyenne annuelle	0,14	0,24	0,35	0,53

Tab35. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur La Désirade 2010

	p=1	p=2	p=3	p=4
Jan	0,29	0,19	0,27	0,36
Fev	0,33	0,20	0,25	0,45
Mar	0,37	0,21	0,25	0,50
Avr	0,08	0,12	0,15	0,75
Mai	0,11	0,14	0,36	0,14
Juin	0,14	0,17	0,40	0,17
Juil	0,13	0,23	0,50	0,62
Août	0,08	0,13	0,21	0,56
Sept	0,16	0,23	0,42	0,29
Oct	0,15	0,21	0,47	0,39
Nov	0,23	0,30	0,58	0,32
Dec	0,11	0,22	0,35	0,39
Moyenne annuelle	0,18	0,20	0,35	0,53

Tab36. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur La Désirade 2011

	p=1	p=2	p=3	p=4
Jan	0,09	0,18	0,46	0,15
Fev	0,14	0,25	0,27	0,22
Mar	0,18	0,29	0,31	0,28
Avr	0,08	0,10	0,10	0,13
Mai	0,08	0,09	0,21	0,15
Juin	0,11	0,16	0,26	0,18
Juil	0,10	0,11	0,18	0,19
Août	0,09	0,11	0,18	0,12
Sept	0,12	0,13	0,56	0,65
Oct	0,16	0,18	0,30	0,66
Nov	0,19	0,21	0,33	0,66
Dec	0,13	0,21	0,31	0,24
Moyenne annuelle	0,12	0,17	0,29	0,30

Tab37. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur La Désirade 2012

	p=1	p=2	p=3	p=4
Jan	0,11	0,15	0,27	0,58
Fev	0,13	0,18	0,47	0,63
Mar	0,18	0,23	0,49	0,68
Avr	0,10	0,22	0,38	0,37
Mai	0,09	0,12	0,18	0,66
Juin	0,13	0,14	0,15	0,63
Juil				
Août	0,12	0,17	0,28	0,64
Sept	0,10	0,13	0,23	0,15
Oct	0,16	0,39	0,54	0,65
Nov	0,22	0,48	0,58	0,70
Dec	0,13	0,15	0,30	0,26
Moyenne annuelle	0,14	0,22	0,35	0,54

Tab38. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur Petit-canal 2010

	p=1	p=2	p=3	p=4
Jan	0,21	0,29	0,69	0,65
Fev	0,21	0,19	0,24	0,41
Mar	0,24	0,20	0,23	0,44
Avr	0,09	0,14	0,18	0,17
Mai	0,14	0,18	0,25	0,18
Juin	0,13	0,16	0,38	0,16
Juil	0,13	0,18	0,28	0,55
Août	0,14	0,17	0,28	0,20
Sept	0,20	0,28	0,36	0,37
Oct	0,14	0,19	0,71	0,35
Nov	0,21	0,27	0,59	0,45
Dec	0,14	0,18	0,72	0,29
Moyenne annuelle	0,16	0,20	0,41	0,35

Tab39. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur Petit-canal 2011

	p=1	p=2	p=3	p=4
Jan	0,15	0,32	0,43	0,26
Fev	0,14	0,24	0,25	0,20
Mar	0,15	0,24	0,26	0,21
Avr	0,10	0,12	0,13	0,20
Mai	0,12	0,14	0,20	0,19
Juin	0,10	0,15	0,24	0,16
Juil	0,16	0,17	0,22	0,68
Août	0,10	0,15	0,21	0,20
Sept	0,12	0,13	0,16	0,15
Oct	0,15	0,17	0,27	0,65
Nov	0,18	0,20	0,31	0,67
Dec	0,13	0,16	0,48	0,17
Moyenne annuelle	0,13	0,18	0,26	0,31

Tab40. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur Petit-canal 2012

	p=1	p=2	p=3	p=4
Jan	0,10	0,16	0,29	0,69
Fev	0,12	0,17	0,25	0,61
Mar	0,18	0,22	0,28	0,59
Avr	0,10	0,21	0,38	0,11
Mai	0,09	0,10	0,12	0,63
Juin	0,13	0,13	0,15	0,63
Juil	0,14	0,22	0,28	0,64
Août	0,14	0,37	0,51	0,61
Sept	0,18	0,42	0,69	0,67
Oct	0,23	0,50	0,78	0,65
Nov	0,13	0,27	0,43	0,32
Dec	0,10	0,16	0,29	0,69
Moyenne annuelle	0,14	0,25	0,38	0,56

Tab41. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 30 min sur Fouillole 2010

	p=1	p=2	p=3	p=4
Jan	0,27	0,19	0,25	0,55
Fev	0,32	0,20	0,25	0,22
Mar	0,31	0,24	0,28	0,58
Avr	0,09	0,11	0,15	0,14
Mai	0,11	0,14	0,36	0,14
Juin	0,13	0,16	0,37	0,16
Juil	0,13	0,18	0,28	0,55
Août	0,11	0,16	0,24	0,32
Sept	0,11	0,17	0,48	0,34
Oct	0,14	0,19	0,51	0,35
Nov	0,20	0,27	0,58	0,46
Dec	0,14	0,32	0,53	0,47
Moyenne annuelle	0,17	0,19	0,36	0,36

Tab42. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 1h sur Fouillole 2011

	p=1	p=2	p=3	p=4
Jan	0,13	0,26	0,28	0,24
Fev	0,14	0,25	0,27	0,23
Mar	0,18	0,28	0,30	0,26
Avr	0,08	0,13	0,30	0,12
Mai	0,08	0,14	0,24	0,17
Juin	0,10	0,15	0,24	0,16
Juil	0,16	0,17	0,22	0,62
Août	0,08	0,11	0,19	0,14
Sept	0,10	0,12	0,21	0,69
Oct	0,15	0,17	0,27	0,65
Nov	0,19	0,21	0,32	0,69
Dec	0,09	0,12	0,18	0,18
Moyenne annuelle	0,12	0,18	0,25	0,34

Tab42. RMSE normalisée (nRMSE) pour différents ordres p pour la prédiction du flux solaire à 1h sur Fouillole 2012

Liste des publications

Journaux

Andre Maïna, Sophie Dabo-Niang, Ted Soubdhan, " Predictive spatio-temporal model for spatially sparse global solar radiation data, *Energy*, 2015, *soumis*

Soubdhan Ted, Ould-Baba Hanany, Minh-Thang Do, Ndong Joseph, Andre Maïna, "Granger Causality Analysis and a Dynamic Parameter estimation to set up a vector Autoregressive Model: Application to photovoltaic Power Forecast, *Energy*, 2015, *soumis*

Conférences

Andre Maïna, Ted Soubdhan, "Time and space dependent correlation of global solar radiation time series using empirical mode decomposition", *ISES International Congress XIII*, 8-12 Novembre 2015, Daegu, South Corea

Andre Maïna, Sophie Dabo-Niang, Ted Soubdhan, "Spatio temporal correlations and intra-hour forecasting of global solar radiation with VAR model in Guadeloupe", *WREC International Congress XIII*, 3-8 August 2014, London-Uk

Andre Maïna, Sophie Dabo-Niang, Ted Soubdhan, "Forecasting of different class of global solar radiation with spatio-temporal VAR model", *CM2E international Conference*, 5-9 May 2014, Martinique.

Andre Maïna, Soubdhan Ted, Bernard Jean-Louis, "Variability of global sunshine in Guadeloupe", *WREF international Conference*, 13-17 May 2012, Denver, Colorado.

Andre Maïna, Bernard Jean-Louis, "Analysis of fluctuations of global sunshine in Guadeloupe", *IASTED international Conference*, 2-4 April 2012, Phuket, Thailand