UNIVERSITÉ SAVOIE MONT BLANC

# THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ SAVOIE MONT BLANC

Spécialité : **Sciences et techniques de l'information et de la communication, Traitement de l'information**

Arrêté ministériel : 25 Mai 2016

Présentée par


## Matthieu GALLET


Thèse dirigée par **Abdourrahmane ATTO** et
codirigée par **Fatima KARBOU et Emmanuel TROUVÉ**

préparée au sein du **Laboratoire LISTIC**
dans **l'École Doctorale SIE**

# Apprentissage automatique pour la caractérisation et l'analyse de la dynamique de la cryosphère par imagerie satellitaire radar

Thèse soutenue publiquement le **20 septembre 2024**,
devant le jury composé de :

**Mme Florence TUPIN**
Professeure, Telecom ParisTech, Paris (Rapportrice)
**M. Laurent FERRO-FAMIL**
Professeur, ISAE-Supaero,Toulouse (Rapporteur)
**M. Lionel BOMBRUN**
Maître de conférences, Université de Bordeaux, Bordeaux (Examinateur)
**Mme Nadia FOURRIÉ**
Directrice de recherche, CNRM GMAP/OBS, Toulouse (Présidente)
**M. Jordi INGLADA**
Expert scientifique senior, CESBIO, Toulouse (Examinateur)
**M. Abdourrahmane ATTO**
Professeur, Université Savoie-Mont-Blanc, Annecy (Directeur de thèse)
**Mme Fatima KARBOU**
Directrice de recherche, CNRM CEN, Grenoble (Co-directrice de thèse)
**M. Emmanuel TROUVÉ**
Professeur, Université Savoie-Mont-Blanc, Annecy (Co-directeur de thèse)

# Résumé

L'analyse de la cryosphère est essentielle en raison de son importance socio-économique et environnementale. Elle influence les ressources en eau douce, soutient le développement économique des régions dépendantes du tourisme hivernal, et joue un rôle significatif dans le bilan énergétique mondial grâce au phénomène de l'albédo. À l'échelle globale, la compréhension des dynamiques de la cryosphère est cruciale pour aborder les problématiques climatiques actuelles, comme souligné dans les rapports du Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC).

L'explosion des données satellitaires ces dernières années offre des perspectives de suivi à haute résolution spatiale et temporelle des régions montagneuses, notamment grâce aux données acquises par les radars à synthèse d'ouverture (RSO). Ces radars permettent d'observer les régions enneigées et glacées indépendamment des conditions météorologiques et de l'illumination solaire. Cependant, l'analyse de ces données reste complexe en raison de la présence de bruit multiplicatif (chatoiement) et de distorsions géométriques irréversibles, nécessitant des méthodes d'analyse adaptées. L'utilisation des méthodes d'apprentissage automatique pour l'analyse des données RSO offre des perspectives prometteuses pour la classification des surfaces enneigées et des glaciers, mais requiert des approches adaptées pour surmonter les défis liés à la complexité des données et à la quantité et la qualité des annotations.

Cette thèse vise à aborder ces défis en se concentrant sur plusieurs aspects clés. Dans un premier temps, elle identifie comment exploiter les données RSO en bande C et en bande X pour la classification de la neige et des glaciers, en analysant les comportements temporels ainsi que les corrélations entre les caractéristiques statistiques des images et les mesures in situ ou issues de modèles numériques.

À partir de cette analyse, elle propose l'élaboration de deux jeux de données annotés pour l'étude de la cryosphère et leur utilisation dans des tâches de classification. Ces jeux de données sont constitués d'images RSO acquises par les satellites Sentinel-1 et PAZ, centrées sur les Alpes françaises et la région du Mont-Blanc, et annotées automatiquement avec un modèle de simulation du manteau neigeux (CROCUS) pour la neige humide, et manuellement pour les types de glaciers. La création de ces jeux de données permet l'analyse de la détection de neige humide par un ensemble de modèles d'apprentissage automatique supervisés, et la compréhension des caractéristiques discriminantes pour cette tâche. Face à la variabilité des résultats et des performances des modèles, cette thèse propose une approche permettant l'agrégation de modèles de classification pour améliorer la robustesse des résultats face aux bruits inhérents au signal radar et aux incertitudes des annotations, en se basant sur la logique floue.

Elle propose enfin des solutions algorithmiques permettant une classification efficace et interprétable des données RSO, avec une approche frugale en termes de paramètres et interprétable, alliant estimation d'un ensemble de distributions de probabilité propres aux caractéristiques des images RSO et une architecture neuronale de combinaison de distances statistiques ou divergences paramétrées. En conséquence, ces travaux démontrent la capacité des méthodes d'apprentissage automatique à être adaptées de manière efficace et interprétable à l'analyse des données SAR pour l'étude de la cryosphère, en prenant en compte les spécificités des données et des annotations, et en proposant des solutions algorithmiques pour une classification efficace et interprétable.

# Abstract

The analysis of the cryosphere is essential due to its socio-economic and environmental importance. It influences freshwater resources, supports the economic development of regions dependent on winter tourism, and plays a significant role in the global energy balance through the albedo effect. Globally, understanding the dynamics of the cryosphere is crucial for addressing current climate issues, as highlighted in the reports of the Intergovernmental Panel on Climate Change (IPCC).

The explosion of satellite data in recent years offers opportunities for high spatial and temporal resolution monitoring of mountainous regions, particularly through data acquired by synthetic aperture radars (SAR). These radars allow for the observation of snowy and icy regions regardless of weather conditions and solar illumination. However, analyzing these data remains complex due to the presence of multiplicative noise (speckle) and irreversible geometric distortions, necessitating adapted analytical methods. The use of machine learning methods for SAR data analysis offers promising prospects for classifying snow-covered and glacier surfaces but requires approaches tailored to overcome the challenges related to data complexity and the quantity and quality of annotations.

This thesis aims to address these challenges by focusing on several key aspects. Firstly, it identifies how to exploit C-band and X-band SAR data for snow and glacier classification by analyzing temporal behaviors as well as correlations between statistical image features and in situ or model-derived measurements.

From this analysis, it proposes the creation of two annotated datasets for the study of the cryosphere and their use in classification tasks. These datasets consist of SAR images acquired by the Sentinel-1 and PAZ satellites, centered on the French Alps and the Mont-Blanc region, and annotated automatically with a snowpack simulation model (CROCUS) for wet snow and manually for glacier types. The creation of these datasets enables the analysis of wet snow detection by a set of supervised machine learning models and the understanding of discriminative features for this task. Given the variability in results and model performances, this thesis proposes an approach for aggregating classification models to improve the robustness of results in the face of inherent radar signal noise and annotation uncertainties, based on fuzzy logic.

Finally, it proposes algorithmic solutions for effective and interpretable classification of SAR data, with a frugal approach in terms of parameters and computational time, combining the estimation of a set of probability distributions specific to SAR image features with a neural architecture for combining statistical distances or parameterizable divergences. Consequently, this work demonstrates the ability of machine learning methods to be effectively and interpretable adapted for SAR data analysis for the study of the cryosphere, considering the specifics of the data and annotations, and proposing algorithmic solutions for effective and interpretable classification.

# Table des matières

**Table des matières** vii

**Notation** xi

**Liste des acronymes** xiii

**1 Introduction** 1
- 1.1 Problématiques . . . 3
- 1.2 Organisation du manuscrit . . . 4
- 1.3 Contributions . . . 5
  - 1.3.1 Articles de revue . . . 5
  - 1.3.2 Articles de conférence . . . 5
  - 1.3.3 Codes et données . . . 5

**2 Imagerie SAR et Cryosphère** 7
- 2.1 Cryosphère . . . 7
  - 2.1.1 Quelques éléments de la cryosphère: Objets d'étude . . . 8
  - 2.1.2 Mesures pour l'étude de la cryosphère . . . 11
- 2.2 Synthetic Aperture Radar (SAR) pour la cryosphère . . . 18
  - 2.2.1 Principes de la mesure SAR . . . 18
  - 2.2.2 Statistiques du signal . . . 25
  - 2.2.3 Analyse SAR pour la cryosphère . . . 29
- 2.3 Caractéristiques temporelles de la rétrodiffusion: comparaison entre les bandes X et C pour la cryosphère . . . 35
  - 2.3.1 Comportement temporel de la rétrodiffusion en bande X . . . 36
  - 2.3.2 Comportement temporel de la rétrodiffusion en bande C . . . 37
  - 2.3.3 Analyse comparative pour les glaciers . . . 38
  - 2.3.4 Analyse comparative pour la neige . . . 41
- 2.4 Conclusion . . . 44

**3 Apprentissage automatique et jeux de données SAR** 47
- 3.1 Méthodes d'apprentissage statistique et automatique pour l'analyse de données SAR . . . 47
  - 3.1.1 Méthodes spatiales avec a priori de distribution . . . 48
  - 3.1.2 Caractéristiques et décompositions . . . 54
  - 3.1.3 Méthodes spatiales sans a priori de distribution . . . 55
  - 3.1.4 Méthodes d'apprentissage temporel . . . 65
- 3.2 Jeu de données SAR en bande C pour l'étude de la neige humide . . . 67
  - 3.2.1 Zone d'étude . . . 67
  - 3.2.2 Données satellitaire . . . 68
  - 3.2.3 LSD4WSD V1 . . . 70
  - 3.2.4 LSD4WSD VX . . . 76
- 3.3 Jeu de données SAR en bande X pour l'étude des glaciers . . . 77
  - 3.3.1 Zone d'étude . . . 77
  - 3.3.2 Données SAR . . . 77
  - 3.3.3 Labellisation des données . . . 78
- 3.4 Conclusion . . . 79

**4 Détection de la neige humide par apprentissage automatique** **81**
4.1 Modèles d'apprentissage automatique . . . 81
4.1.1 Modèles . . . 81
4.1.2 Prétraitement et représentation des données . . . 82
4.1.3 Gestion des données et scénarios d'apprentissage . . . 84
4.2 Résultats quantitatifs . . . 85
4.2.1 Évaluation des modèles . . . 85
4.2.2 Évaluation des données auxiliaires . . . 86
4.2.3 Ensemble de statistiques . . . 88
4.3 Résultats qualitatifs et analyse temporelle . . . 90
4.3.1 Cartographie en période de fonte . . . 90
4.3.2 Comparaison entre la période de fonte et de début de saison . . . 92
4.3.3 Répartition topographique de la neige humide . . . 94
4.3.4 Comparaison in-situ . . . 95
4.4 Conclusion . . . 98

**5 Méthodes d'ensemble floue et application pour la cryosphère** **101**
5.1 Contexte: logique floue et opérateurs d'agrégation . . . 101
5.1.1 Mesures floues . . . 101
5.1.2 Intégrale de Choquet . . . 102
5.1.3 T-normes et opérateur de pré-agrégation . . . 103
5.2 Apprentissage flou: intégrale de Choquet et pré-agrégation . . . 104
5.2.1 Apprentissage des paramètres de l'intégrale de Choquet . . . 105
5.2.2 Apprentissage des T-normes paramétriques . . . 106
5.3 Expérimentations . . . 107
5.3.1 Description du jeu de données . . . 107
5.3.2 Algorithmes et stratégie d'apprentissage . . . 108
5.4 Résultats quantitatifs sur l'apprentissage flou . . . 111
5.4.1 Apprentissage de la mesure floue . . . 111
5.4.2 Apprentissage des T-normes . . . 111
5.4.3 Robustesse aux perturbations . . . 112
5.5 Spatialisation des résultats . . . 114
5.6 Conclusion . . . 116

**6 Apprentissage d'ensemble de combinaisons de distances et divergences** **119**
6.1 Apprentissage d'ensemble de combinaisons de distances et divergences . . . 119
6.1.1 Descripteurs statistiques a priori . . . 120
6.1.2 Apprentissage de l'ensemble de divergence . . . 124
6.1.3 Résultats expérimentaux . . . 125
6.2 Apprentissage d'ensemble de divergences paramétriques . . . 127
6.2.1 Divergences de Rényi paramétriques . . . 129
6.2.2 Cadre d'apprentissage par divergence . . . 130
6.2.3 Expérimentations . . . 133
6.2.4 Résultats et discussion . . . 134
6.3 Conclusion . . . 137

**7 Conclusion & perspectives** **139**

**A Annexes** **145**
A.1 Compléments sur les fonctions spéciales . . . 145

A.2 Détails des régions d'étude pour la classification des glaciers . . . . . 145
A.2.1 Complément sur la comparaison des données SAR en bande C et en bande X pour les glaciers . . . . . 147

**Bibliographie** **155**

# Notation

Liste des symboles utilisés dans le document.

**Constantes**

| | | |
|---|---|---|
| $\varepsilon$ | Euler-Mascheroni | 0.577 21 |

**Ensembles**

| | |
|---|---|
| $\mathbb{C}$ | Ensemble des nombres complexes |
| $\mathbb{R}$ | Ensemble des nombres réels |
| $\mathbb{R}^+$ | Ensemble des nombres réels positifs |
| $\mathbb{R}^+{}_*$ | Ensemble des nombres réels positifs non nuls |

**Apprentissage automatique**

| | |
|---|---|
| $\ell_2$ | Norme euclidienne |
| $\hat{\mathbf{y}}$ | Vecteur de prédiction |
| $\mathcal{J}$ | Divergence de Jeffreys |
| $\mathcal{K}$ | Divergence de Kullback-Leibler |
| $\mathbf{w}$ | Vecteur de poids |
| $\mathbf{X}$ | Matrice ou tenseur de données |
| $\mathbf{x}$ | Vecteur de données |
| $\mathbf{y}$ | Vecteur de vérité terrain |
| $\mathbf{y}_i$ | Scalaire de vérité terrain de l'échantillon $i$ |
| $\mathcal{H}$ | Entropie croisée |
| $\mathcal{JS}$ | Divergence de Jensen-Shannon |
| $\mathcal{L}$ | Fonction de perte |
| $\nabla_\theta \mathcal{L}$ | Gradient de la fonction de perte $\mathcal{L}$ par rapport aux paramètres $\theta$ |
| $b$ | Quantité scalaire |
| $D_\alpha$ | Divergence de Renyi d'ordre $\alpha$ |

**Statistiques**

| | |
|---|---|
| $\gamma_1$ | Coefficient d'asymétrie (*Skewness*) |
| $\gamma_2$ | Coefficient d'aplatissement (*Kurtosis*) |

$\hat{\kappa}_i$ Estimateur du log-cumulant d'ordre $i$

$\mathscr{E}(x;\lambda)$ Fonction de densité de probabilité Exponentielle

$\mathscr{G}(x;\mu,L)$ Fonction de densité de probabilité Gamma

$\mathscr{L}(x;\mu,\sigma)$ Fonction de densité de probabilité Log-Normale

$\mathscr{N}(x;\mu,\sigma^2)$ Fonction de densité de probabilité Gaussienne

$\mathscr{R}(x;\sigma)$ Fonction de densité de probabilité Rayleigh

$\mathscr{W}(x;\mu,k)$ Fonction de densité de probabilité Weibull

$\mu$ Moyenne

$\sigma$ Écart-type

**Physique**

$\delta_e$ Épaisseur de peau

$\lambda$ Longueur d'onde

$\rho$ Densité

$\sigma_0$ Coefficient de rétrodiffusion (*Backscattering coefficient*)

$\tan(\delta)$ Tangente de pertes diélectriques

$\theta_l$ Angle d'incidence local

$\varepsilon_r$ Permittivité relative (complexe : $\varepsilon_r = \varepsilon_r' + j\varepsilon_r''$)

$dB$ décibel

**Autres symboles**

$\mathbf{A}^T$ Transposée de la matrice $\mathbf{A}$

$\odot$ Produit de Hadamard

$\oslash$ Division terme à terme

# Acronymes

**A**
**ABL** ablation area. 8
**ACC** accumulation area. 8
**ANFIS** Adaptive Neuro-Fuzzy Inference System. 59
**ANN** Artificial Neural Network. 12
**AUC** Area Under the ROC Curve. 62

**B**
**BAROC** Best Accuracy on the ROC. 82

**C**
**CNN** Convolutional Neural Network. 12
**CREA** Centre de Recherche sur les Écosystèmes d'Altitude. 41
**CV** coefficient de variation. 25

**D**
**DEM** Digital Elevation Model. 67
**DWT** Discrete Wavelet Transform. 69

**E**
**ELA** Equilibrium Line Altitude. 8
**ESA** European Space Agency. 19

**F**
**FCROC** False positive rate Constant on the ROC. 82
**FSC** Fractionnal Snow Cover. 68

**G**
**GIEC** Groupe d'experts intergouvernemental sur l'évolution du climat. 1
**GRD** Ground Range Detected. 21

**H**
**HAG** Hanging Glaciers. 8

**I**
**ICA** Ice Aprons. 8
**IGN** Information Géographique et Forestière français. 67
**InSAR** Interferometric Synthetic Aperture Radar. 22

**K**
**KNN** K-Nearest Neighbors. 54

**L**
**LR** Logistic Regression. 54
**LWC** Liquid Water Content. 9

**M**
**MBM** massif du Mont-Blanc. 75
**ML** Machine Learning. 12
**MLP** Multilayer Perceptron. 54
**MNS** Modèle Numérique de Surface. 67
**MNT** Modèle Numérique de Terrain. 67
**MoLC** Method of Log-Cumulant. 27

**N**
**NDSI** Normalized Difference Snow Index. 16
**NDVI** Normalized Difference Vegetation Index. 16

**P**
**PAC** Probabilities Approximately Correct. 116
**PCA** Principal Component Analysis. 131
**PolSAR** Polarimetric Synthetic Aperture Radar. 31

**R**
**RF** Random Forest. 17
**ROC** Receiver Operating Characteristic. 62
**ROI** Regions of Interest. 34

**S**
**SAR** Synthetic Aperture Radar. iii, 1, 18, 19, 21, 23, 25, 27, 29, 31, 33
**SCE** Snow Cover Extent. 15
**SLC** Single Look Complex. 21
**SMB** Surface Mass Balance. 13
**SPD** Symmetric Positive Definite. 31
**SVM** Support Vector Machine. 26, 54
**SWE** Snow Water Equivalent. 9
**SWIR** Short-Wave InfraRed. 16
**SWS** SAR Wet Snow. 73

**T**
**TEL** Teneur en Eau Liquide. 9

**U**
**UTC** Coordinated Universal Time. 66

**V**
**VWC** volumetric water content. 9

# 1 Introduction

Le terme cryosphère désigne toutes les parties des mers et des terres émergées où l'eau est sous forme solide. Pour le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), la cryosphère inclut *"les volumes de neige, les glaciers, les inlandsis, les banquises, les icebergs, la glace de mer, de lac et de rivière, et de façon temporaire ou permanente les sols gelés (pergélisol/permafrosts, selon les saisons)"*. Une compréhension de la dynamique de la cryosphère est nécessaire à l'échelle locale, où les changements socio-économiques des régions dépendantes de la neige pointe le besoin d'une connaissance accrue sur les phénomènes liés à la cryosphère. Ces dépendances peuvent être liées aux ressources en eau douce découlant de la présence de neige et de glaciers, mais aussi à un développement économique dans des régions reposant sur un modèle touristique hivernal.
D'autre part, à l'échelle globale, cette compréhension fait échos aux problématiques climatiques actuelles. Nous retrouvons ainsi l'étude de la cryosphère dans le rapport spécial de 2019 du GIEC [1] et les problématiques en milieux montagneux et polaires dans le 6ème rapport [2]. Les risques liés à l'évolution rapide et inconnue de la cryosphère y sont analysés. Dans tous les scénarios de réchauffement étudiés, la diminution de la couverture neigeuse accélérerait le taux d'extinction des plantes dans certaines régions et la fonte des glaces favoriserait à l'échelle globale la libération de déchets et de polluants présents dans les glaces. La réduction de l'albédo[1] des zones enneigées quant à elle augmenterait le réchauffement du sol et l'évapo-transpiration (élément majeur dans le cycle de l'eau et dans le bilan énergétique mondial) [1].

[1]: Ipcc (2022), *The Ocean and Cryosphere in a Changing Climate*

[2]: Ipcc (2022), *Climate Change 2022*

**Définition (1)**
*Albédo:* L'albédo est la fraction de la lumière incidente réfléchie par une surface. Il est souvent utilisé pour caractériser la réflectivité d'une surface, et est un élément important dans le bilan radiatif de la Terre.

Dans ce contexte multi-échelles et multi-sectorielles, l'étude de la cryosphère, notamment les surfaces enneigées, est nécessaire et activement menée et les données satellitaires apparaissent comme un outil incontournable pour l'étude à large échelle de la cryosphère. L'intérêt est spatial : ces données permettent d'obtenir des informations sur des zones difficilement accessibles, sur de grandes étendues, ou encore, en fonction du capteur utilisé, sur des zones avec des conditions météorologiques défavorables. L'intérêt est également temporel, la répétitivité des acquisitions permet de suivre l'évolution des cibles d'intérêt, avec une fréquence pouvant descendre à quelques dizaines de minutes pour les satellites géostationnaires. Ainsi les données satellitaires ont connu un essor important ces dernières années, notamment avec l'arrivée de nouvelles missions ou l'accès libre à certaines données, tel que le programme Copernicus de l'Union Européenne [3].
Dès 1978, avec le lancement du premier satellite civil (SeaSat) RADAR à Synthèse d'Ouverture (RSO ou *Synthetic Aperture Radar* SAR), les données radar ont été utilisées pour l'étude de la cryosphère [4]. Ce dernier, durant sa brève période opérationnelle (~3.5 mois), a pu analyser la couverture de glace de mer et des glaciers grâce à son imageur radar (en bande L). Un des principaux buts de SeaSat était de démontrer la capacité de ce type de capteur à surveiller les phénomènes et caractéristiques océanographiques de la Terre depuis l'espace à une échelle globale. Le succès de cette mission a ouvert la voie à de nombreuses autres missions SAR[2] , caractérisées par des résolutions spatiales et temporelles variées, et des longueurs d'onde différentes, en fonction des objectifs scientifiques et des besoins opérationnels [5].

[3]: Aschbacher (2017), « ESA's earth observation strategy and Copernicus »

[4]: Kramer et al. (2002), *Observation of the Earth and its Environment : Survey of Missions and Sensors*

**Précision (2)**
Un descriptif détaillé des missions SAR est disponible sur le site `https://www.eoportal.org/satellite-missions`

[5]: Ustin et al. (2021), « Current and near-term advances in Earth observation for ecological applications »

Il convient de noter les avantages et les limites des capteurs de type SAR. Ces

derniers sont des capteurs actifs (cf. Figure 1.1), et permettent d'observer la surface terrestre de manière indépendante des conditions météorologiques et de l'illumination solaire grâce aux longueurs d'onde utilisées. Ils sont donc particulièrement adaptés pour l'observation des régions polaires et montagneuses. De plus, le SAR peut donner accès à des informations sur la structure des cibles considérées, et sur les propriétés de surface, en fonction de la longueur d'onde utilisée. Ceci s'avère être à la fois un avantage, permettant par exemple de renseigner sur la teneur en eau liquide au sein du manteau neigeux, mais aussi une limite, car la complexité des phénomènes d'interaction du signal radar dans la matière rend son interprétation difficile. De plus, les images SAR sont caractérisées par la présence de bruit multiplicatif appelé chatoiement (*speckle* en anglais) [6] et de phénomènes de distorsion géométrique, rendant leur analyse complexe.

[6]: Goodman (1976), « Some fundamental properties of speckle »

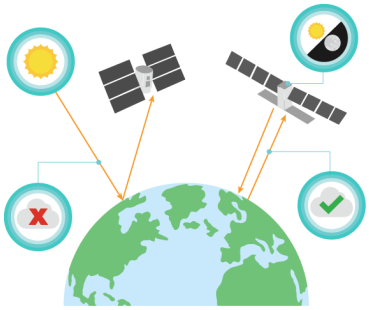


**Figure 1.1 –** Comparaison des satellites actifs et passifs. Les flèches orange représentent la propagation des ondes électromagnétiques. Extrait des ressources pédagogiques de ArcGIS

Pour répondre à ces problématiques, des méthodes d'analyse, de filtrage et de traitement des données SAR ont été développées [7]. Celles-ci permettent de réaliser différentes tâches répondant aux problématiques données dans la figure 2.1, telles que la classification de surfaces glaciaires, ou encore la caractérisation de la neige. Les outils méthodologiques sont variés, et peuvent être basés sur des analyses statistiques [8], des approches de segmentation [9], ou encore des méthodes d'apprentissage automatique [10]. Ces dernières années, l'utilisation de méthodes d'apprentissage automatique, et plus particulièrement de méthodes d'apprentissage profond [11], a connu un essor important, notamment pour l'analyse d'images SAR [12].

[7]: Passah et al. (2022), « SAR Image Classification »

[8]: Besic et al. (2015), « Stochastic Approach in Wet Snow Detection Using Multitemporal SAR Data »

[9]: Guiot et al. (2023), « Insights into Segmentation Methods Applied to Remote Sensing SAR Images for Wet Snow Detection »

[10]: Bouchayer et al. (2022), « A Machine Learning Framework to Automate the Classification of Surge-Type Glaciers in Svalbard »

[11]: Zhang et al. (2016), « Deep Learning for Remote Sensing Data »

[12]: Parikh et al. (2020), « Classification of SAR and PolSAR images using deep learning »

Dans un contexte où la frugalité est de plus en plus valorisée face au changement climatique, la compréhension croissante de son impact sur nos sociétés conduit à la mise en place de nouveaux outils de mesure et de méthodes d'analyse [13]. Cependant, il est notable que cette évolution s'accompagne d'une complexification des méthodes d'analyse.

Cette complexification se traduit également par une augmentation de la quantité de traitements, ce qui entraîne une croissance du nombre de paramètres et a un impact direct sur la mémoire et le temps de calcul. Ces facteurs ont deux conséquences principales : d'une part, un éventuel retard dans l'obtention des résultats, ce qui peut être problématique dans le contexte de la prise de décision, et d'autre part, une augmentation de la consommation énergétique. De plus, ces méthodes sont souvent considérées comme des boîtes noires, dépendant d'un grand nombre d'hyperparamètres[3] ajustés manuellement, ce qui rend leur interprétation difficile.

[13]: Zhang et al. (2022), « Progress and Challenges in Intelligent Remote Sensing Satellite Systems »

**Définition (3)**

*Hyperparamètres:* Les hyperparamètres en machine learning sont des paramètres ajustables qui définissent la structure d'un modèle et influent sur son apprentissage, mais qui ne sont pas directement appris à partir des données. Ils doivent être définis avant l'entraînement et comprennent des choix tels que la profondeur d'un réseau neuronal ou le taux d'apprentissage.

Cependant, ces méthodes nécessitent souvent une grande quantité de données correctement annotées pour être efficaces. Or, dans le cadre de l'étude de la cryosphère, les bases de données SAR sont souvent limitées, et les annotations manuelles sont complexes à réaliser de par les limitations précédemment évoquées.

L'utilisation de l'apprentissage automatique pour le spatial en général et plus particulièrement pour l'étude de la cryosphère fait face à de nombreux défis. Le premier est la faible disponibilité et la qualité des données d'apprentissage, souvent exacerbées par des problèmes de variabilité des acquisitions en temps et en espace. À cela s'ajoute la labellisation des données, nécessaire dans un cadre d'apprentissage supervisé, qui est complexe à obtenir, et est sujette à des problèmes de validité temporelle des annotations, de sur ou sous-représentation des classes, ou plus simplement de qualité des annotations [14]. En présence de jeux de données construits sur des données spatio-temporelles,

[14]: Hansch et al. (2019), « The Truth About Ground Truth »

les pièges sont nombreux pour l'évaluation des modèles, par la sélection des données utilisées pour l'entraînement et la validation [15]. Dans un second temps, l'évaluation des modèles, notamment pour la comparaison des méthodes, pose la question de la généralisation des résultats à d'autres jeux de données, mais aussi de la pertinence de la vérification des modèles sur des données de validation indépendantes (mesures terrain, données d'autres capteurs). Cette validation se heurte souvent à des problèmes de disponibilité des données (complexité de la mesure, coût de la mesure, spatialisation des mesures). Enfin, l'interprétabilité des résultats est un enjeu majeur pour l'acceptation des méthodes d'apprentissage automatique par les utilisateurs de produits spatiaux. Cette interprétabilité est d'autant plus cruciale dans le cadre de l'étude de la cryosphère, où les utilisateurs sont souvent des experts du domaine, et où la compréhension de la décision du modèle est nécessaire pour établir des liens avec la physique du signal.

[15]: Hänsch et al. (2017), « Correct and Still Wrong »

Malgré ces défis, l'accès libre à certaines données satellitaires, telles que celles du programme Copernicus, offre des opportunités pour la création de telles bases de données, et pour le développement de méthodes d'apprentissage automatique, notamment via l'utilisation de codes communautaires, ouverts et reproductibles.

## 1.1 Problématiques

Dans le cadre de ces travaux de thèse, nous nous sommes intéressés à répondre à plusieurs problématiques :

- Quelles sont les caractéristiques des données SAR (en bande C et en bande X) pour l'étude de la cryosphère, et comment les exploiter pour la classification de la neige et des glaciers ?
- Comment construire des jeux de données labellisés pour l'étude de la cryosphère, et comment les exploiter pour la classification de la neige et des glaciers ?
- Comment introduire des méthodes d'apprentissage automatique pour la classification de neige humide à partir de données SAR, et quelles sont leurs validités au vu de l'état de l'art ?
- À partir des nombreuses méthodes d'apprentissage automatique existantes, comment choisir ou agréger des méthodes, en prenant en compte la variabilité des données et des annotations ?
- Comment rehausser l'interprétabilité des résultats, mais aussi la frugalité des solutions algorithmiques dans le cadre de la classification de données SAR ?

**Précision (4)**
Le terme frugalité est ici utilisé pour désigner des méthodes de classification nécessitant un minimum de paramètres, et un temps de calcul réduit, tout en conservant une bonne performance dans la tâche considérée.

Dans ce cadre, cette thèse cherche à explorer dans un premier temps la capacité des approches d'apprentissage automatique à réaliser des tâches de classification de données SAR dans le cadre de l'étude de la cryosphère. Dans un second temps, nous cherchons à proposer de nouvelles méthodes pour la classification de données SAR. Ces méthodes sont expérimentées à travers deux cas applicatifs :

- la caractérisation de la neige humide dans les Alpes françaises du Nord,
- la classification de différents types de glaciers dans la région du Mont-Blanc.

**Précision (5)**
Le terme hybride est ici utilisé pour désigner des méthodes de classification combinant des approches paramétriques et non-paramétriques, soit d'estimation statistique et d'apprentissage par la donnée.

À travers ces deux cas d'études, nous cherchons à étudier l'apport de méthodes de classification frugales[4] ou hybrides,[5] motivée par le besoin de garder un

temps d'inférence et des quantités de données plus mesurés, mais aussi par le besoin de garder une compréhension extensive du modèle, permettant une connexion simple avec des experts propres au domaine applicatif.

## 1.2 Organisation du manuscrit

De ces questions découlent la structure de ce manuscrit, qui est organisé en 5 chapitres principaux :

Le Chapitre 2 s'articule autour de la présentation de la cryosphère, des acquisitions SAR et d'une première contribution sur la comparaison des données SAR en bande C et en bande X pour le suivi de la neige et des glaciers. Il vise à donner un aperçu global et synthétique des éléments nécessaires à la compréhension des travaux menés dans cette thèse mais aussi à dégager certaines problématiques liées à la classification de données SAR dans le cadre de l'étude de la cryosphère.

Le Chapitre 3 présente les aspects méthodologiques nécessaires pour répondre aux problématiques posées. Il introduit les méthodes d'apprentissage statistique et automatique utilisées pour l'analyse de données SAR, et présente la deuxième contribution de cette thèse, à savoir la construction de jeux de données labellisés pour l'étude de la cryosphère, en visant à donner un aperçu des données utilisées, et des spécificités des zones d'étude.

Le Chapitre 4 présente notre troisième contribution, les travaux menés pour la caractérisation de la neige humide dans les Alpes françaises du Nord. Il présente l'application d'un ensemble de méthodes de classification, sur différentes caractéristiques des données SAR, en exploitant un jeu de données SAR en bande C labellisées. Nous y comparons les performances de ces méthodes au vu des produits existants.

Nous proposons dans le Chapitre 5 une méthode ensembliste, basée sur l'agrégation floue de modèles d'apprentissage automatique supervisé. Nous présentons le contexte théorique et notre quatrième contribution, qui vise à montrer l'intérêt de l'agrégation de modèles pour la classification de données SAR, avec une application sur la classification de différents types de glaciers dans la région du Mont-Blanc, à partir de données SAR en bande X extraites du dataset décrit dans le chapitre Chapitre 3.

Le Chapitre 6 présente l'utilisation de méthodes hybrides, alliant des approches basées sur des modèles avec des a priori de distribution, et des approches d'apprentissage profond. Nous présentons notre cinquième contribution, à travers deux scénarios. Le premier est basé sur des divergences et mesures sur une variété de caractéristiques, le second basé sur un seul type de divergence, mais paramétrique. Dans ce chapitre, nous proposons une évaluation de ces méthodes sur des données SAR en bande X, extraites du dataset décrit dans le chapitre Chapitre 3.

Enfin, le chapitre Chapitre 7 présente les conclusions de ce travail, et les perspectives pour la suite de ces travaux.

## 1.3 Contributions

Les travaux présentés dans ce manuscrit ont donné lieu à plusieurs contributions, qui sont résumées ci-dessous :

### 1.3.1 Articles de revue

**[J1]** **M. Gallet**, A. Atto, F. Karbou and E. Trouvé, "Wet Snow Detection From Satellite SAR Images by Machine Learning With Physical Snowpack Model Labeling," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 2901-2917, 2024, doi : 10.1109/JSTARS.2023.3342990.

**[J2]** **M. Gallet**, A. Atto, F. Karbou and E. Trouvé, "Aggregation of Forest of Classifiers with Fuzzy Learning for Glacier Classification in SAR Images" soumis à IEEE Transactions on Geoscience and Remote Sensing.

### 1.3.2 Articles de conférence

**[C1]** **M. Gallet**, A. Mian and A. Atto, "Renyi Divergences Learning for explainable classification of SAR Image Pairs," ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, Republic of, 2024, pp. 7445-7449, doi : 10.1109/ICASSP48485.2024.10448227.

**[C2]** C. Lin-Kwong-Chon, **M. Gallet**, S. Kaushik and E. Trouvé, "Supervised classification for analysis of cryospheric zones using SAR statistical timeseries," IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium

**[C3]** **M. Gallet**, A. Atto, E. Trouvé and F. Karbou. Apprentissage explicable d'un ensemble de divergences pour la similarité inter-classe de données SAR. GRETSI, XXIXème Colloque Francophone de Traitement du Signal et des Images, GRETSI, Aug 2023, Grenoble, France.

**[C4]** **M. Gallet**, A. Atto, E. Trouvé and F. Karbou, "CNN Classification of Wet Snow by Physical Snowpack Model Labeling," IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 2023, pp. 5351-5354, doi : 10.1109/IGARSS52108.2023.10281833.

**[C5]** S. Kaushik, **M. Gallet**, Y. Yan, A. Atto, L. Ravanel and E. Trouvé, "Temporal Evolution of X and C Band Sar Backscattering In The Mont-Blanc Massif," IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 2023, pp. 207-210, doi : 10.1109/IGARSS52108.2023.10282799.

### 1.3.3 Codes et données

Les codes développés pour ces travaux sont disponibles sur le dépôt GitHub de l'auteur : `https://github.com/Matthieu-Gallet?tab=repositories`.

Les jeux de données construits pour ces travaux :

**[D1]** **M. Gallet**, A. Atto, F. Karbou and E. Trouvé, "LSD4WSD : An Open Dataset for Wet Snow Detection with SAR Data and Physical Labelling (VX)". Zenodo, Oct. 30, 2023. doi : `https://doi.org/10.5281/zenodo.10046730`.

**[D2]** **M. Gallet**, A. Atto, F. Karbou and E. Trouvé, "LSD4WSD : An Open Dataset for Wet Snow Detection with SAR Data and Physical Labelling (V1)". Zenodo, Oct. 30, 2023. doi : https://doi.org/10.5281/zenodo.8111485.

**[D3]** **M. Gallet**, S. Kaushik, C. LIN KWONG CHONand E. Trouvé, "CPAZ-MAL : Cryosphere PAZ satellite MAchine Learning". Zenodo, Dec. 18, 2023. doi : https://doi.org/10.5281/zenodo.10401360.

# Imagerie SAR et Cryosphère 2


**2.1 Cryosphère** . . . . . 7
**2.2 Synthetic-Aperture Radar** . . . . . 18
**2.3 Comparaison des données SAR en bande C et en bande X** . . . . . 35
**2.4 Conclusion** . . . . . 44


Le champ d'étude de la cryosphère s'avère être large comme l'illustre la figure 2.1. Nous proposons dans ce chapitre de présenter un état de l'art sur les domaines abordés tout au long de ce manuscrit. Nous commencerons par une présentation de la cryosphère, puis nous aborderons les principes de l'imagerie radar et les spécificités de l'imagerie SAR pour la cryosphère. Enfin, nous proposons une analyse du comportement temporel (évolution saisonnière) et des corrélations entre les données SAR en bande C et en bande X pour le suivi de la neige et des glaciers.

## 2.1 Cryosphère

Sur Terre, l'eau sous forme solide peut être trouvée dans les calottes glaciaires, les glaciers, la neige (permanente ou saisonnière), les glaces de sol et pergélisol. Plus de 69% de l'eau douce de la planète est stockée sous ces différentes formes [17]. La cryosphère est ainsi un élément clé du système climatique terrestre, jouant un rôle important dans la régulation du climat et de l'hydrologie. Plus précisément, l'étude des glaciers et de la neige revêt une importance capitale dans divers domaines en raison de ses implications multiples. Ces études sont essentielles pour appréhender les changements climatiques [18],

[17]: Gleick (1993), « Water in crisis : a guide to the world's fresh water resources »

[18]: Beniston (2003), « Climatic Change in Mountain Regions : A Review of Possible Impacts »

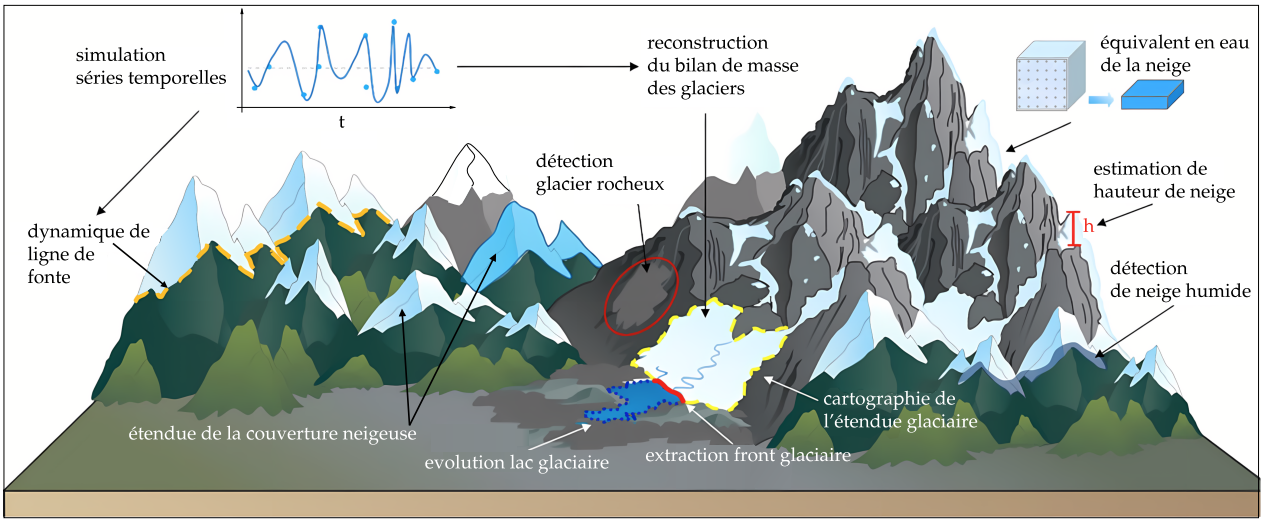



**Figure 2.1** – Schéma des différentes études de la cryosphère adapté de [16]

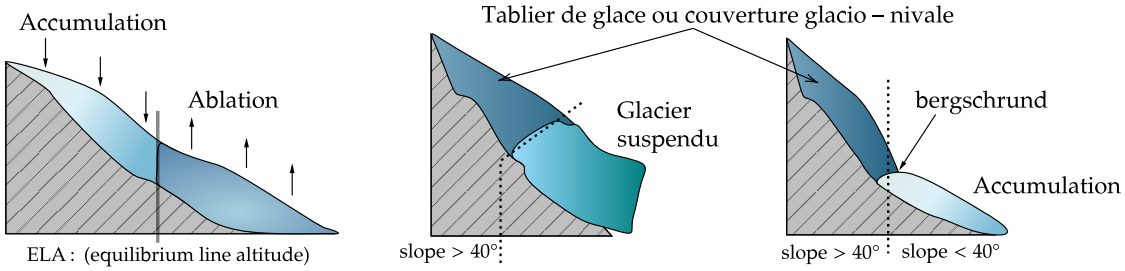



**Figure 2.2 –** Illustration des 4 types de surfaces glaciaires étudiées dans ce travail.

ou comprendre les dynamiques socio-économiques [19] [20]. Elles permettent également d'analyser et quantifier les flux élémentaires, de la disponibilité des ressources en eau douce à la production hydroélectrique [21] [22] [23]. Enfin, le suivi de la stabilité de ces zones permet d'anticiper et de gérer les risques naturels tels que les avalanches [24] ou les inondations [25].

### 2.1.1 Quelques éléments de la cryosphère : Objets d'étude

Nous présentons dans cette partie les deux principaux objets d'études que nous considérons par la suite dans notre travail : la distinction entre la neige humide et la neige sèche, et la classification de quatre types de surfaces glaciaires.

#### Glaciers

Nous proposons de nous concentrer sur quatre types de glaciers, à savoir les glaciers de vallée, répartis en zone d'accumulation (*accumulation area* ACC) et zone d'ablation (*ablation area* ABL), les glaciers suspendus (*Hanging Glaciers* HAG) et les tabliers de glace ou couvertures glacio-nivales (*Ice Aprons* ICA). Ces glaciers sont illustrés dans la figure 2.2. Ce choix se justifie par la diversité des caractéristiques de ces glaciers, leur présence dans le massif d'étude du Mont-Blanc et leur large utilisation pour des applications variées, de la prévention des risques naturels à la dynamique climatique[6] [26].

- Les tabliers de glace ou couvertures glacio-nivales (ICA), sont des zones glaciaires de faible superficie (généralement < 0,009 km$^2$ pour le massif du Mont-Blanc) formées de plaques de glaces minces de forme irrégulière, existant au-dessus de l'Altitude de la Ligne d'Équilibre (*Equilibrium Line Altitude* ELA) sur des pentes très raides (généralement entre 40 et 65°) [27] ;
- Les glaciers suspendus (HAG), qui sont des glaciers dont le front ne repose pas sur un lit rocheux et se termine de manière abrupte sur une falaise de roche [28] ;
- Les glaciers de vallée avec leurs zones d'accumulation (ACC) où la neige se transforme en firn[7] et en glace, et leurs zones d'ablation (ABL) avec de la glace nue après la fonte des neiges ;

Les tabliers de glace, en raison de leur faible volume et de leur localisation, offrent une perspective intéressante pour surveiller la santé des glaciers de haute montagne, notamment en ce qui concerne l'impact du changement climatique [29]. Les glaciers suspendus, en revanche, sont des sujets d'étude sensibles en raison des risques potentiels qu'ils représentent, la rupture d'un

[19]: François et al. (2014), « Crossing numerical simulations of snow conditions with a spatially-resolved socio-economic database of ski resorts : A proof of concept in the French Alps »

[20]: Morrison et al. (2013), « Perceptions of climate change impacts, adaptation and limits to adaption in the Australian Alps : the ski-tourism industry and key stakeholders »

[21]: Viel et al. (2016), « How seasonal forecast could help a decision maker : an example of climate service for water resource management »

[22]: Rondeau-Genesse et al. (2016), « Monitoring snow wetness in an Alpine Basin using combined C-band SAR and MODIS data »

[23]: Bahrami et al. (2020), « Analysing the contribution of snow water equivalent to the terrestrial water storage over Canada »

[24]: Baggi et al. (2008), « Characteristics of wet-snow avalanche activity : 20 years of observations from a high alpine valley (Dischma, Switzerland) »

[25]: Clague et al. (2021), « Chapter 14 - Glacier-related outburst floods »

**Précision (6)**
La diminution du volume des glaciers est responsable à hauteur de 17.3% de l'élévation du niveau de la mer[26]. La diminution de la surface des glaciers impacte directement la température de surface par la réduction de l'albédo et donc une augmentation de l'absorption de la chaleur solaire.

[26]: Climate Change (IPCC) (2023), « Ocean, Cryosphere and Sea Level Change »

[27]: Guillet et al. (2020), « Variations in surface area of six ice aprons in the Mont-Blanc massif since the Little Ice Age »

[28]: Benn et al. (2010), *Glaciers and Glaciation, 2nd Edition*

glacier suspendu menant à de potentielles conséquences graves [30]. Les glaciers de vallée[8] , répartis ici en zone d'accumulation et d'ablation, sont des glaciers de plus grande taille et plus souvent étudiés. Ces deux zones représentent les principaux composants du bilan massique du glacier, la zone d'accumulation recevant plus de neige qu'elle ne perd de glace par son mouvement vers l'aval, tandis que la zone d'ablation perd plus de glace par fonte et sublimation qu'elle ne reçoit de neige et de glace reçu de l'amont [31].

**Définition (7)**
*Firn:* Névé partiellement compacté, neige laissée par les saisons précédentes et recristallisée en une substance plus dense que le névé. Il s'agit d'une glace qui se trouve à un stade intermédiaire entre la neige et la glace glaciaire.

[29]: Kaushik et al. (2022), « Analysis of the Temporal Evolution of Ice Aprons in the Mont-Blanc Massif Using X and C-Band SAR Images »

[30]: Faillettaz et al. (2016), « Time forecast of a break-off event from a hanging glacier »

**Précision (8)**
Le terme *bergschrund*, donné dans la figure 2.2, indique une zone constituées de profondes crevasses transversales près des têtes des glaciers de vallée à l'interface de la zone d'accumulation et des tabliers de glace.

[31]: Engeset et al. (2002), « Change Detection and Monitoring of Glacier Mass Balance and Facies Using ERS SAR Winter Images over Svalbard »

### Neige

Les chutes de neige proviennent de la condensation solide de vapeur d'eau dans les nuages. La transition entre neige et pluie se produit à une altitude de 200 à 400 mètres en dessous de l' isotherme 0°C[9] [32]. Durant la saison hivernale, le manteau neigeux se forme par l'accumulation de différentes couches, résultantes de précipitations neigeuses, chacune possédant des caractéristiques mécaniques, physiques et thermiques distinctes. En plus de cette variabilité verticale, la composition du manteau neigeux varie considérablement en fonction de l'orientation des pentes, de l'altitude et de la zone géographique. Les cristaux de neige subissent des changements tout au long de l'hiver sous l'influence des flux thermodynamiques, entraînant des modifications dans les couches de neige jusqu'à leur fonte. La fonte du manteau neigeux peut se décomposer en trois phases distinctes [33] : (a) l'humidification, (b) la maturation, et (c) l'écoulement ou le ruissellement. L'humidification initie le processus de fonte lorsque l'élévation de la température de l'air et l'intensification du rayonnement solaire, combinées aux échanges thermiques avec le sol ou aux précipitations, entraînent la fonte des couches superficielles du manteau neigeux. La phase de maturation commence dès que la capacité de rétention maximale des pores est dépassée, provoquant l'infiltration du front humide à travers le manteau neigeux. Ce phénomène, résultant de cycles de gel et de dégel répétés, n'aboutit pas immédiatement à l'écoulement de l'eau de fonte. Pendant cette phase, le manteau neigeux atteint un état isotherme, et quand il ne peut plus contenir aucune eau liquide, la phase de ruissellement débute. Le processus de fonte est complexe et non linéaire, soumis à de nombreuses variables de forçages externes (conditions météorologiques, topographiques) et internes (type de cristaux, densité du manteau, température). Ce processus de fonte est particulièrement important puisque son étude permet, entre autres, de quantifier le stockage d'eau douce potentiel dans le manteau neigeux, de connaître à quel moment la phase de ruissellement commence. Pour son étude les variables les plus significatives sont l'équivalent en eau de la neige (*Snow Water Equivalent* SWE) [10] et la Teneur en Eau Liquide (TEL ou *Liquid Water Content* LWC) [11] . Par simplification de ce processus d'humidification continue, il est courant de parler de neige sèche et de neige humide. La neige humide est définie comme étant caractérisée par une teneur en eau liquide supérieure à 3% en poids, tandis que la neige sèche est définie par une teneur en eau liquide inférieure à 1.5% en poids [34] [35]. Le détail de cette classification est donné dans le tableau 2.1. Il est important de noter que ce découpage est sujet à débat et [36] donne deux autres définitions pour la distinction entre neige sèche et humide. La première est basée sur le contenu volumique en eau (*volumetric water content* VWC) [12] , où l'on considère une neige sèche quand cet indicateur est en dessous de 1%. La deuxième définition repose sur la température du manteau neigeux. Si la température est inférieure à 0°C, le manteau est considéré comme sec, et humide dans le cas contraire. L'information de présence de neige humide est importante

**Définition (9)**
*Isotherme 0°C:* L'isotherme zéro degré marque l'altitude la plus basse où la température atteint 0 degré Celsius dans une région donnée de la surface terrestre et sur une période de temps déterminée et à atmosphère libre.

[32]: Ancey (1996), *Guide neige et avalanches : connaissances, pratiques, sécurité*

[33]: Dingman (2002), *Physical Hydrology*

**Définition (10)**
*SWE:* Snow Water Equivalent ou équivalent en eau de la neige en français. Représente le contenu en eau liquide résultant de la fonte immédiate de toute la couche de neige à un emplacement donné. L'équivalent en eau de la neige est inférieur à la hauteur de la neige au sol, de par la composition mixte de la neige (eau et air). Cette mesure est donnée en millimètres (mm).

**Définition (11)**
*TEL:* Teneur en Eau Liquide ou Liquid Water Content (LWC) en anglais. Indique la quantité d'eau liquide présente à l'intérieur du manteau neigeux (en % ou en kg/m$^2$).

[34]: Techel et al. (2010), « Point observations of liquid water content in wet snow - investigating methodical, spatial and temporal aspects »

**Table 2.1 –** Classification de la neige en fonction de sa teneur en eau liquide (TEL). Adapté de [35]

| Type de neige | Description |
|---|---|
| Sèche | Généralement, la température de surface du sol (Ts) est en dessous de 0°C, mais de la neige sèche peut se former à n'importe quelle température jusqu' Les grains de neige désagrégés ont peu tendance à adhérer les uns aux autr lorsqu'ils sont pressés ensemble, comme lors de la formation d'une boule de |
| Humide | Ts = 0°C. L'eau n'est pas visible même à un grossissement de 10x. Lorsqu'elle est légèrement écrasée, la neige a une tendance distincte à coller |
| Mouillée | Ts = 0°C. L'eau peut être reconnue à un grossissement de 10x par son ménis entre les grains de neige adjacents, mais l'eau ne peut pas être extraite en la serrant modérément dans les mains (régime pendulaire). |
| Très mouillée | Ts = 0°C. L'eau peut être extraite en serrant modérément la neige dans les m mais une quantité appréciable d'air est confinée dans les pores (régime funi |
| Imbibée | Ts = 0°C. La neige est imprégnée d'eau et contient une fraction volumique d'air de 20 à 40% (régime funiculaire). |

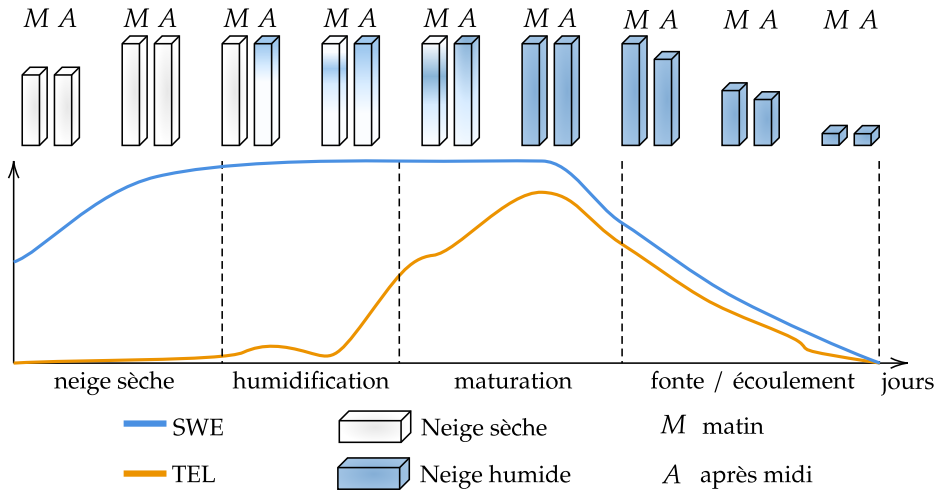



**Figure 2.3 –** Schéma de l'évolution générale de deux indicateurs du manteau neigeux (SWE et TEL) au cours d'une saison hivernale. Adapté de [41].

[35]: CRYOSPHÉRIQUES et al. (2009), « The International classification for seasonal snow on the ground »

[36]: TSAI et al. (2019), « Remote Sensing of Snow Cover Using Spaceborne SAR »

**Définition (12)**

*Contenu volumique en eau:* volume water content VWC en anglais, est définie comme le rapport entre le volume d'eau et le volume total du matériau. Se réfère à la quantité totale d'eau, qu'elle soit sous forme liquide ou solide, présente dans un matériau, comme le sol ou la neige, par unité de volume.

[37]: HAPUARACHCHI et al. (2011), « A review of advances in flash flood forecasting »

[39]: BELLAIRE et al. (2017), « On forecasting wet-snow avalanche activity using simulated snow cover data »

pour de nombreuses applications, notamment pour la prévision des crues printanières [37][38], la prédiction [39] ou la simulation [40] des avalanches de neige humide.

La figure 2.3 illustre l'humidification du manteau neigeux au cours d'une saison hivernale. Les barres présentent au dessus du graphe schématisent une coupe verticale du manteau. Nous pouvons distinguer la différence entre le manteau du matin (M) et celui de l'après-midi (A). L'humidification commence d'abord l'après midi, quand le manteau a chauffé toute la journée, puis progressivement s'étend au reste de la journée pour rendre un manteau complètement humide. Une fois ce stade atteint, la phase de ruissellement commence et le manteau par sa fonte diminue en hauteur. Parallèlement, les deux indicateurs décrits précédemment (SWE et TEL) sont illustrés. Une augmentation de la TEL au fil du temps révèle un processus d'humidification en cours, avec une infiltration progressive de l'eau à travers le manteau neigeux. Le SWE, quant à lui, est lié à la hauteur de la neige par :

$$SWE = \frac{\hat{\rho}_s}{\rho_w} \cdot h,$$

où $\rho_w$ est la densité de l'eau, $\hat{\rho}_s$ et $h$ respectivement la densité moyenne et la

hauteur de la colonne de neige[13] . On observe une augmentation graduelle du SWE au cours de la saison hivernale (aussi appelée saison d'accumulation), correspondant aux chutes de neige successives. L'humidification est initiée à la surface [34] par l'intermédiaire de pluie sur la surface ou de la fonte due au rayonnement solaire. L'eau liquide qui se libère ou qui est absorbée par les couches superficielles gèle en entrant en contact avec la neige des couches inférieures. Ce processus dégage de la chaleur latente, ce qui entraîne le réchauffement du manteau neigeux, amorçant ainsi le processus de maturation de la neige. Les cycles répétés de fonte partielle pendant la journée et de regel pendant la nuit favorisent la formation du front humide dans la neige. Cette évolution n'est généralement pas uniforme, car les infiltrations débutent souvent par des "doigts d'écoulement"[14] isolés qui se transforment en canaux d'eau de fonte plus importants au fil du temps. Ainsi cette transformation diffère selon la topographie, mais aussi d'une année à l'autre. Une fois que le manteau neigeux atteint l'isothermie et une saturation totale en eau, il cesse de retenir l'eau liquide. Un surplus d'énergie peut entraîner un écoulement d'eau qui, selon les caractéristiques du sol telles que sa composition en glace et en eau, peut soit s'infiltrer dans le sol soit se transformer en ruissellement de surface [42]. À ce stade, l'équivalent en eau de la neige (SWE) diminue rapidement à mesure que l'eau s'écoule du manteau neigeux, entraînant également une diminution de la hauteur et de la teneur en eau liquide du manteau. À partir de ces indicateurs, il convient de pouvoir les estimer de manière précise et continue pour suivre l'évolution de la cryosphère.

**Précision (13)**
La densité du manteau neigeux est comprise entre 50 et 800 kg/m$^3$. Au dessus de 800 kg/m$^3$, la neige peut être considérée comme de la glace. Une neige fraîche a une densité de l'ordre de ~ 100 kg/m$^3$, ce qui correspond à un SWE de 100 mm pour une hauteur de neige de 1 mètre.

[34]: Techel et al. (2010), « Point observations of liquid water content in wet snow - investigating methodical, spatial and temporal aspects »

**Précision (14)**
Ces doigts d'écoulement sont des zones de la neige où l'eau liquide s'infiltre en premier. Ce phénomène de fonte localisée, couplé à des événements de pluie sur neige, favorise une augmentation de la rugosité de la surface.

[42]: DeWalle et al. (2008), « Snowmelt-runoff processes »

### 2.1.2 Mesures pour l'étude de la cryosphère

La surveillance continue des glaciers ou du manteau neigeux dans des environnements montagneux escarpés et accidentés par des observations sur le terrain se révèle être difficile [43, 44] par manque d'accessibilité. En conséquence, les mesures pour la cryosphère dans la plupart des régions montagneuses sont souvent rares et discontinues. Nous proposons un bref aperçu des différents moyens de surveillance et de mesure pour l'étude de la cryosphère.

[43]: Gultepe (2015), « Chapter Three - Mountain Weather : Observation and Modeling »
[44]: Kaushik et al. (2022), « Monitoring Hanging Glacier Dynamics From Sar Images Using Corner Reflectors And Field Measurements In The Mont-blanc Massif »

**Mesures In-Situ :**

Des mesures continues de SWE et de TEL sont essentielles pour suivre la dynamique de fonte et caractériser le manteau neigeux. Ces mesures peuvent être complétées par des observations de la hauteur de neige et de la température de la neige. Pour ce faire, les méthodes courantes consistent à mesurer manuellement le SWE à l'aide de tubes d'échantillonnage de neige ou de manière automatique avec des coussins de neige [45]. Cependant, l'installation et la maintenance de ce type de mesures sont coûteuses, et un nombre relativement limité de mesures continues de SWE sont disponibles dans les Alpes.

[45]: Kinar et al. (2015), « Measurement of the physical properties of the snowpack »

En ce qui concerne les mesures directes de TEL, elles peuvent être réalisées en se basant sur les propriétés diélectriques de la neige par l'intermédiaire de sondes à micro-ondes. Alternativement, elles peuvent être obtenues de manière indirecte en exploitant une information de température de neige ou encore en utilisant les signaux GPS [46]. Des méthodes similaires sont également utilisées pour l'estimation du SWE [47]. Ces paramètres physiques nécessitent des instruments spécifiques, tels que les stations Nivose de Météo-France (cf figure 2.4), qui sont moins courants que les instruments plus standards

[46]: Koch et al. (2014), « Measuring Snow Liquid Water Content with Low-Cost GPS Receivers »
[47]: Steiner et al. (2022), « (Near) Real-Time Snow Water Equivalent Observation Using GNSS Refractometry and RTKLIB »

utilisés pour mesurer la température ou les précipitations [48]. Ces appareils, en raison de leur spécificité, peuvent être plus coûteux, gourmands en énergie et nécessitent d'être suffisamment robustes pour résister aux conditions climatiques difficiles, ce qui réduit encore la densité de ces mesures.

Cependant, un grand nombre de stations météorologiques sont équipés de capteurs plus standard tels que des capteurs de température, d'humidité, de précipitations, de vitesse de vent, ou encore de hauteur de neige. Ces stations peuvent être utilisées pour estimer indirectement le SWE ou la TEL, en utilisant des modèles empiriques ou des modèles de régression [48] [49]. Le lecteur pourra se référer à [50] pour une description détaillée des différentes méthodes de mesure.

Deux tendances émergent de ce type de mesures. Tout d'abord, la qualité de l'estimation : les stations ou appareils spécialisés permettent d'obtenir des mesures précises et fiables, mais peuvent nécessiter des interventions ou des étalonnages. Ensuite, la densité de ces mesures est souvent très locale, limitant ainsi la possibilité de surveiller et de comprendre le manteau neigeux et les phénomènes associés à la fonte, en tenant compte de la variabilité géographique à plus grande échelle.

Pour surmonter le problème de spatialisation des mesures in-situ, des stratégies d'interpolation spatiale peuvent être utilisées pour estimer les valeurs en des points non mesurés à partir des mesures existantes. Ces méthodes peuvent être basées sur des techniques de géostatistique, telles que la krigeage[15] [51], ou sur des méthodes d'apprentissage automatique (*Machine Learning* ML), telles que les Réseaux de Neurones Artificiels (*Artificial Neural Network* ANN) [52]. Cette approche reste cependant locale à l'étendue spatiale du réseau de mesures et nécessite une densité de mesures suffisante, mais possède une bonne résolution temporelle. En effet, les relevés manuels peuvent être quotidiens et les relevés automatiques (températures, vitesses de vent, humidité) peuvent avoir une période d'acquisition courte (de l'ordre de la dizaine de minutes). De manière similaire [53] et [54] ont exploité une information dérivée d'images de type webcam pour respectivement inférer la vitesse de déplacement des glaciers ou la hauteur de neige. L'un utilise un recalage d'image pour déterminer les déplacements des pixels tandis que l'autre utilise un Réseau de Neurones Convolutifs (*Convolutional Neural Network* CNN) pour remonter à l'épaisseur du manteau. Ces travaux ont l'avantage de proposer une étude locale d'un phénomène, moins ponctuelle qu'une mesure physique, de par le champ couvert par la caméra, tout en gardant une bonne résolution temporelle. Néanmoins, l'information obtenue est liée à des indicateurs de présence (détection de neige) ou linéiques (hauteur de neige) et ne permet pas de caractériser les propriétés internes de la cible. Ces propriétés nécessitent des mesures faisant appel à des capteurs travaillant dans des domaines spectraux spécifiques différents de la lumière visible, comme les micro-ondes [55].

Un compromis entre résolution spatiale et résolution temporelle peut être trouvé en s'attachant à une description physique des phénomènes observés par l'utilisation de modèles numériques.

### Modèle numérique

Pour atténuer ces limitations, des modèles du manteau physique multicouches basés sur les échanges d'énergies et sur la modélisation physique des interactions entre la neige et l'environnement peuvent être utilisés [56] [57].

[48]: Lundberg et al. (2010), « Towards automated 'Ground truth' snow measurements–a review of operational and new measurement methods for Sweden, Norway, and Finland »



**Figure 2.4 –** Station Nivose de Météo-France du camp des Rochilles. Crédit photo : hermes05100.free.fr

[48]: Lundberg et al. (2010), « Towards automated 'Ground truth' snow measurements–a review of operational and new measurement methods for Sweden, Norway, and Finland »

[49]: Fontrodona-Bach et al. (2023), « NH-SWE : Northern Hemisphere Snow Water Equivalent dataset based on in situ snow depth time series »

[50]: Carmagnola (2014), « Mesure, analyse et modélisation des processus physiques du manteau neigeux sec »

**Définition (15)**

*Krigeage:* Technique d'interpolation spatiale provenant de la géostatistique, visant à minimiser la variance en se basant sur l'interprétation et la modélisation du variogramme expérimental. Elle prend en considération non seulement la distance entre les données et le point d'estimation, mais également les distances entre toutes les paires de données.

[51]: Tapsoba et al. (2005), « Apport de la technique du krigeage avec dérive externe pour une cartographie raisonnée de l'équivalent en eau de la neige : Application aux bassins de la rivière Gatineau »

[52]: Odry et al. (2020), « Using artificial neural networks to estimate snow water equivalent from snow depth »

[53]: Marsy et al. (2020), « Monitoring Mountain Cryosphere Dynamics by Time Lapse Stereo Photogrammetry »

[54]: Kopp et al. (2019), « Fully automated snow depth measurements from time-lapse images applying a convolutional neural network »

Ces modèles permettent de simuler les paramètres physiques du manteau neigeux, tels que la température, la densité, le contenu en eau liquide [58], la nature de la neige par couche et la hauteur de neige, à des échelles spatiales et temporelles plus grandes que les mesures in-situ. Les modèles peuvent être forcés par des données météorologiques obtenues localement par des stations météorologiques fiables.

Dans le cas de la neige, de nombreux modèles ont pu être développés pour simuler le manteau neigeux : des modèles plus simples se concentrant sur l'interaction entre la neige et la canopée [59], à des modèles plus complexes comme le modèle SNOWPACK [60], ou encore le modèle CROCUS [61] [62].

Un atout majeur de ces modèles est leur capacité à simuler des variables non mesurées et de le faire a priori quelle que soit la localisation spatiale et temporelle. Cependant il peut êter numériquement coûteux de réaliser des simulations sur des zones à l'échelle du massif Alpin avec des résolutions spatio-temporelles fines (1 m, 1 heure). Il est important de noter que ces modèles nécessitent une calibration et une validation régulières pour garantir la fiabilité des résultats. De plus, cette stratégie, par la simplification de certains processus lors de la modélisation des systèmes physiques complexes en jeu, introduit une incertitude importante sur les paramètres modélisés, qui peuvent être difficiles à quantifier [63]. Ces incertitudes peuvent provenir de diverses sources : du forçage météorologique [64], de la structure [65] ou du paramétrages du modèle [66].

La modélisation ne se limite pas à la seule simulation du manteau neigeux. D'autres stratégies ont vu le jour avec l'explosion des techniques d'intelligence artificielle. [67] propose ainsi une simulation du bilan de masse de surface (*Surface Mass Balance* SMB) à partir de variables climatiques, avec l'utilisation d'un réseau de neurones, afin de suivre l'évolution des glaciers. Ici, l'utilisation de réseaux de neurones permet de gagner en vitesse de simulation une fois le modèle entraîné, mais aussi de s'affranchir de la complexité des modèles physiques et de découvrir des relations non modélisées entre les variables. Cependant, ces modèles nécessitent une quantité de données importante pour être entraînés et ne permettent pas de remonter aux processus physiques sous-jacents.

Nous proposons de détailler le modèle CROCUS, qui sera utilisé dans la suite de ce travail.

**CROCUS :** Développé par le Centre d'Études de la Neige (CEN) de Météo-France, le modèle unidimensionnel CROCUS [61] vise à décrire l'état et l'évolution des propriétés du manteau neigeux. Le modèle CROCUS suit les changements dans les propriétés physiques du manteau neigeux, y compris la stratigraphie (jusqu'à 50 couches) et le sol sous-jacent. CROCUS décrit l'évolution des transferts thermiques et radiatifs internes du manteau neigeux en se basant sur une description semi-quantitative des propriétés morphologiques des grains de neige. CROCUS est contraint par des variables ou des mesures externes (température de l'air, l'humidité relative de l'air, la vitesse du vent, les précipitations pluvieuses et neigeuses, flux radiatifs solaires et atmosphériques), notamment par l'utilisation de données météorologiques réanalysées à intervalles horaires à partir d'un autre modèle numérique SAFRAN [68], mais sans aucune information sur le manteau neigeux observé. Les couches de neige numériques sont décrites par leur épaisseur, la température au milieu de la couche, la masse volumique de la neige, la TEL, et le type de

[55]: Stefko et al. (2022), « Coherent backscatter enhancement in bistatic Ku- and X-band radar observations of dry snow »

[56]: Armstrong et al. (2008), « Snow and Climate : Physical Processes, Surface Energy Exchange and Modeling »

[57]: Essery et al. (2013), « A comparison of 1701 snow models using observations from an alpine site »

[58]: Yamaguchi et al. (2004), « Application of the numerical snowpack model (SNOWPACK) to the wet-snow region in Japan »

[59]: Strasser et al. (2011), « Modeling Snow–Canopy Processes on an Idealized Mountain »

[60]: Lehning et al. (1999), « SNOWPACK model calculations for avalanche warning based upon a new network of weather and snow stations »

[61]: Vionnet et al. (2012), « The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2 »

[62]: Brun et al. (1992), « A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting »

[63]: Günther et al. (2019), « Uncertainties in Snowpack Simulations—Assessing the Impact of Model Structure, Parameter Choice, and Forcing Data Error on Point-Scale Energy Balance Snow Model Performance »

[64]: Raleigh et al. (2015), « Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework »

[65]: Avanzi et al. (2016), « Model complexity and data requirements in snow hydrology : seeking a balance in practical applications »

[66]: Engel et al. (2017), « Snow model sensitivity analysis to understand spatial and temporal snow dynamics in a high-elevation catchment »

[67]: Bolibar et al. (2020), « Deep learning applied to glacier evolution modelling »

[68]: Durand et al. (1993), « A meteorological estimation of relevant parameters for snow models »

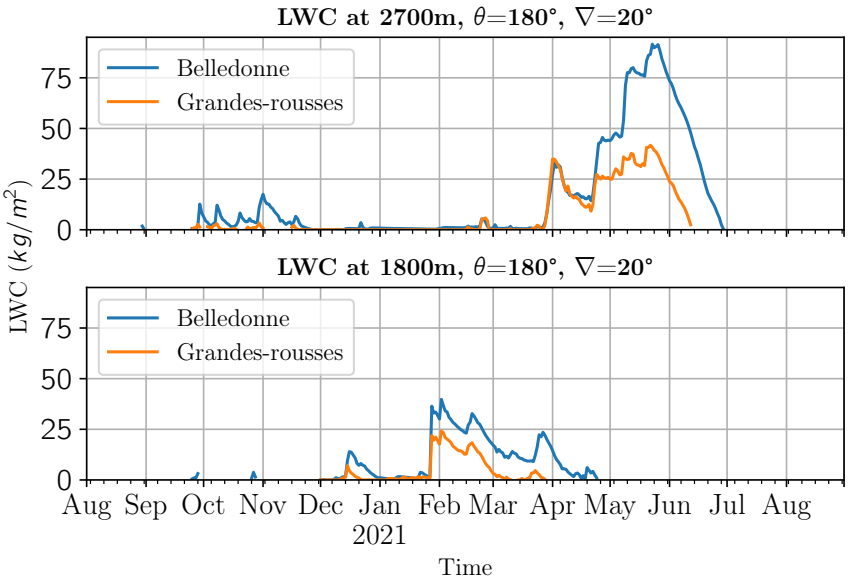



**Figure 2.5 –** Variation temporelle de teneur en eau liquide (TEL) pour le modèle CROCUS simulé quotidiennement à 18h pour deux massifs alpins (Belledonne et Grandes Rousses) à différentes altitudes (1800 et 2700 mètres) sur la période d'août 2020 à août 2021.

grain. Les sorties du modèle consistent en des propriétés physiques intégrées à l'échelle du manteau neigeux et les propriétés physiques internes pour toutes les couches numériques de neige. Parmi les sorties d'intérêt pour la suite, on peut citer la hauteur de neige, la température minimale et maximale sur la hauteur de neige, la TEL.[16] Par sa conception, CROCUS permet de traiter une variété de processus physiques tels que la conduction thermique, les changements de phase, le bilan d'énergie en surface, les métamorphoses de la neige, la percolation de l'eau liquide, et le tassement sous l'effet du poids des couches [50]. Un des nombreux avantages de ce modèle réside dans sa conception unidimensionnel, et un maillage large (à l'échelle des massifs alpins) qui permet une modélisation maîtrisée en terme de temps de calcul.

**Précision (16)**

Il est important de noter que sous CROCUS, la TEL est simulée pour chaque couche (en kg/m$^3$). Nous utilisons l'intégration sur la colonne, qui donne une information totale de TEL (en kg/m$^2$). Pour retrouver la TEL en pourcentage, il suffit d'utiliser l'intégration de la densité sur la colonne (en kg/m$^2$)

[50]: Carmagnola (2014), « Mesure, analyse et modélisation des processus physiques du manteau neigeux sec »

Un exemple de la variation temporelle de la TEL simulée par le modèle CROCUS est donné dans la figure 2.5. Le graphique illustre la variabilité au cours de l'hiver 2020-2021 de la TEL pour deux massifs alpins (Belledonne et Grandes Rousses) à différentes altitudes (1800 et 2700 mètres). La TEL est modélisée quotidiennement pour 18h. Nous pouvons remarquer pour la simulation à 2700m, un comportement proche de l'illustration 2.3. Après un épisode d'humidification à l'automne, plus présent dans le massif de Belledonne que dans celui des Grandes Rousses, nous pouvons observer l'humidification progressive des couches du manteau neigeux qui débute en même temps pour les deux massifs. Une fois la saturation en eau atteinte, fin mai, le manteau neigeux ruisselle et fond, marquant une diminution brusque de la TEL. Nous pouvons noter un comportement différent pour la simulation à 1800m, où l'augmentation de TEL est brusque, fin janvier soit près de deux mois avant la zone à 2700m d'altitude. La fonte quant à elle, se fait plus progressivement, avec plus de variations dénotant des épisodes de chute de neige fraîche.

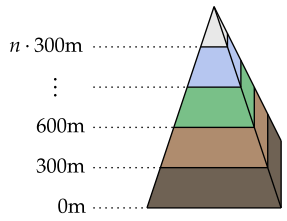



**Figure 2.6 –** Illustration du découpage spatiale de CROCUS sur la composante d'altitude. $n$ représente le multiple de 300 mètres.

L'hypothèse principale de ce modèle est l'homogénéité spatiale des massifs considérés, en particulier pour les précipitations. En conséquence, l'échelle spatiale considérée est celle d'une chaîne de montagnes, mais avec une division topographique de 300 mètres en altitude (de 900 à 3600 mètres, cf figure 2.6), pour huit orientations (avec un pas de 45°, cf figure 2.7) et trois pentes ( 0°, 20°, > 30°, cf figure 2.8). Ce découpage est réalisé pour tous les massifs montagneux

de France métropolitaine (comprenant 23 massifs alpins avec une superficie d'environ 400 km$^2$). Cette échelle exclut donc les effets locaux tels que ceux dus à l'accumulation de neige ou au transport par le vent.

La validation et la comparaison de CROCUS est nécessaire et de nombreux travaux ont ainsi pu valider le modèle par des campagnes de mesures [69] [61] [60], et comparer sa performance avec d'autres modèles [70]. La démarche d'estimation de l'incertitude a plutôt été étudiée sur la hauteur de neige [71]. Cependant l'incertitude sur les valeurs simulées de TEL et SWE, à notre connaissance, est inexistante, de par le manque d'observations de ces variables pour une quantification précise. L'incertitude est dominée par l'accumulation des incertitudes sur les analyses de précipitations. Par conséquent, en valeur absolue, cette incertitude augmente fortement avec la quantité de neige et donc avec l'altitude et avec l'avancée dans la saison. De plus, en raison de la variabilité intra-massif non simulée, l'incertitude est beaucoup plus grande si on raisonne à l'échelle locale que si l'on raisonne à l'échelle du massif pour une altitude et une pente donnée. À l'échelle du massif, les experts s'accordent qu'une incertitude relative de 20% est fréquente. Néanmoins à l'échelle locale, des incertitudes allant jusqu'à 100%, voire davantage dans les zones fortement affectées par le transport de neige par le vent, sont probables.

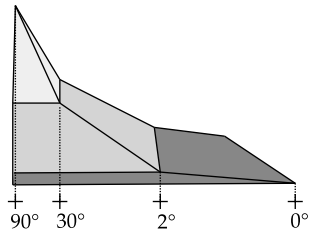



**Figure 2.7 –** Illustration du découpage spatiale de CROCUS sur la composante de pente

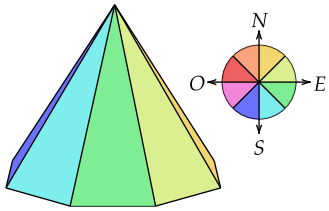



**Figure 2.8 –** Illustration du découpage spatiale de CROCUS sur la composante d'orientation. N, E, S, O représentent respectivement les orientations Nord, Est, Sud et Ouest.

De ces informations simulées, il est possible de dériver des cartes spatiales directement d'une variable ou d'indicateurs tels que la étendue de la couverture neigeuse (*Snow Cover Extent* SCE) [17] , par la simple règle si la hauteur de neige est supérieure à 0m

Le tableau 2.2 résume quatre méthodes utilisables pour obtenir des paramètres physiques du manteau neigeux notamment la TEL. Six variables y sont comparées. La résolution temporelle donne l'information de l'écart de temps entre deux mesures. Les modèles numériques et certaines stations automatiques ont les meilleures résolutions. L'historique des données informe la quantité d'archive sur les données. Ainsi les données obtenues par des campagnes de mesures sont limitées au moment de la mesure. La spécificité indique quelles mesures physiques sont accessibles par la méthode. Ainsi les campagnes de mesure ou les modèles sont dédiés à l'étude d'un phénomène ou d'une variable et possèdent une meilleure spécificité. Nous distinguons la résolution spatiale, qui s'attache à la finesse de la mesure, et densité spatiale, qui s'attache à la quantité de mesure. La résolution spatiale est médiocre pour les modèles numériques, de par le maillage requis, mais dépend intrinsèquement de l'emprise du modèle (soit de l'échelle : continent, massif). Cependant la densité est bonne contrairement à des stations de mesure, où l'information est

[69]: Larue et al. (2018), « Simulation and Assimilation of Passive Microwave Data Using a Snowpack Model Coupled to a Calibrated Radiative Transfer Model Over Northeastern Canada »

[61]: Vionnet et al. (2012), « The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2 »

[60]: Lehning et al. (1999), « SNOWPACK model calculations for avalanche warning based upon a new network of weather and snow stations »

[70]: Essery et al. (1999), « A comparison of four snow models using observations from an alpine site »

**(a)** catégories temporelles et spécificité de la mesure

| Données | résolution temporelle | historique des données | spécificité |
|---|---|---|---|
| campagne de mesure [73] | ++ | + | +++ |
| stations automatiques [74] | ++ / +++ | +++ | ++ |
| stations manuelles | ++ | ++ | ++/+++ |
| modèle numérique [62] | +++ | +++ | +++ |

**(b)** catégories spatiales

| Données | échelle | résolution spatiale | densité spatiale |
|---|---|---|---|
| campagne de mesure [73] | local | +++ | ++ / +++ |
| stations automatiques [74] | global | +++ | + |
| stations manuelles | global | +++ | + |
| modèle numérique [62] | global | + / ++ | +++ |

**Table 2.2 –** Comparaison des catégories spatio-temporelles de quatre accès à l'information TEL. Le signe "+" indique un niveau faible, "++" un niveau moyen et "+++" un niveau élevé. Adapté de [72]

très localisée (bonne résolution spatiale) mais trop peu répandue dans une zone (faible densité).

[71]: Haddjeri et al. (2023), « Exploring the sensitivity to precipitation, blowing snow, and horizontal resolution of the spatial distribution of simulated snow cover »

**Définition (17)**
*Snow Cover Extent:* SCE, étendue de la couverture neigeuse. Elle peut s'exprimer en terme de superficie ou se visualiser sous forme de carte géolocalisée.

[16]: Taylor et al. (2021), « Remote sensing of the mountain cryosphere »

[75]: Yang et al. (2013), « The role of satellite remote sensing in climate change studies »

[76]: Poursanidis et al. (2017), « Remote Sensing, natural hazards and the contribution of ESA Sentinels missions »

[77]: Chawla et al. (2020), « A review of remote sensing applications for water security : Quantity, quality, and extremes »

**Mesures satellitaires : Optique traditionnelle et multi-spectrales**

Afin de palier les limitations spatiales des mesures sur le terrain, accentuées par des accès parfois difficiles en milieu de montagne, ou à la complexité des modèles physiques, la télédétection (prise de vue aérienne et satellitaire) apporte une réponse intéressante [16] en permettant d'obtenir des mesures à l'échelle régionale ou globale.

Depuis le premier satellite dédié à l'observation de la Terre, Landsat-1, lancé en 1972, l'observation spatiale a prouvé son utilité pour l'étude globale de phénomène climatique [75], de gestion et de suivi des risques [76] et des ressources naturelles [77]. Il est nécessaire de distinguer deux types de satellites : les satellites actifs et passifs. Les satellites actifs, tels que les radars (ERS, Envisat ou Sentinel-1), émettent un signal qui est réfléchi par la surface terrestre et renvoyé au satellite. Les données obtenues permettent de caractériser la surface terrestre en fonction de sa réflectance. Les satellites passifs, tels que Landsat, SPOT, ou encore Sentinel-2, permettent d'obtenir des images de la surface terrestre en observant de manière passive le rayonnement solaire réfléchi par la surface. Ces satellites utilisent plusieurs bandes spectrales, allant du visible à l'infrarouge, pour caractériser la surface terrestre. Dans cette dernière catégorie, nous pouvons différencier l'imagerie multi-spectrale de l'imagerie hyper-spectrale. L'imagerie optique peut être considérée comme un sous ensemble de l'imagerie multispectrale, où le nombre de bandes est limité (entre 3 et 10), celles-ci étant larges (de l'ordre de 0.1 micromètres à 1 micromètres). L'imagerie hyper-spectrale, quant à elle, possède un nombre de bandes plus important (de l'ordre de 100 à 200) et des bandes plus étroites (de l'ordre de 10 nanomètres). Cette différence permet d'obtenir une information plus fine sur la composition de la surface observée, rendant accessible les informations sur la couverture neigeuse, l'albédo, la taille des grains, la présence d'eau liquide très près de la surface ou encore la température [78].

[78]: Dozier et al. (2004), « Multispectral and hyperspectral remote sensing of alpine snow properties »

Dans le domaine d'étude de la cryosphère, certains travaux se sont attachés à l'utilisation de l'imagerie hyper-spectrale, notamment pour quantifier les propriétés physiques de la neige [79] ou la surveillance des glaciers dans l'Himalaya [80], mais les contraintes de coût et de traitement de ces données ont limité leur utilisation.

[79]: Haq et al. (2019), « Artificial neural network-based modeling of snow properties using field data and hyperspectral imagery »

[80]: Gore et al. (2019), « Glacier surface characteristics derivation and monitoring using Hyperspectral datasets : a case study of Gepang Gath glacier, Western Himalaya »

Dans un contexte de données multi-spectrales, [81] a proposé un indice similaire au NDVI[18] (traditionnellement utilisé dans l'étude des végétaux), en se basant sur plusieurs bandes spectrales de l'optique et du proche infrarouge : le Normalized Difference Snow Index (NDSI). Ce dernier s'exprime par :

$$NDSI = (G - SWIR_1)/(G + SWIR_1), \tag{2.1}$$

avec $G$ le canal visible vert, 0.53 – 0.61 micromètres et $SWIR_1$ le canal infrarouge à courte longueur d'onde (*Short-Wave Infrared* SWIR), 1.55 – 1.75 micromètres. Cet indicateur permet de cartographier la neige. En effet, la neige est très réfléchissante dans les parties visibles du spectre électromagnétique, mais inversement très absorbante dans le proche infrarouge. La réflectance de la plupart des nuages reste élevée dans le proche infrarouge, ce qui permet une bonne séparation de la plupart des nuages et de la neige. Le seul inconvénient de cet indicateur est son impossibilité à fournir une cartographie en présence

[81]: Valovcin (1976), *Snow/cloud discrimination*

**Définition (18)**
*Normalized Difference Vegetation Index:* Indice de végétation normalisé, utilisé pour caractériser la végétation à partir d'une combinaison de bandes spectrales. Il est défini par la formule : $NDVI = (NIR - R)/(NIR + R)$, où $NIR$ est le canal proche infrarouge et $R$ le canal rouge.

de nuages qui font barrage au rayonnement visible. Cet indicateur fait partie des techniques de base pour l'exploitation des images multi-spectrales et continue encore à être optimisé par les récents développements en intelligence artificielle [82].

La diversité des satellites met en évidence l'utilité des satellites optiques à haute résolution de type SPOT dans l'analyse de la couverture neigeuse, notamment par l'utilisation du NDSI [83]. L'utilisation de ces données hautes résolution est nécessaire pour le suivi et l'identification des glaciers de montagne de faible surface [84]. L'intégration de ces données pour caractériser la cryosphère est solidement établie dans la recherche académique, soulignée par les liens entre différents satellites à disposition et leur contribution à une compréhension globale de la neige [85].

D'autres méthodes visant à affiner la segmentation peuvent également être employées pour obtenir des informations précises sur la couverture neigeuse [86], en proposant une segmentation des images optiques à l'aide d'un réseau convolutif. Ces approches d'apprentissage automatique, en raison de l'explosion du volume de données satellitaires au cours des dernières années, sur lesquelles ces méthodes s'appuient, sont de plus en plus utilisées dans le domaine de la télédétection et de la cryosphère [87].

L'utilisation d'images optiques, voire multi-spectrales permet d'interpréter facilement les résultats par leur proximité avec la vision humaine. Pour l'étude de la neige, une rapide analyse par temps clair permet de voir la présence et l'extension de la couverture neigeuse uniquement en observant les zones blanches. La combinaison de canaux incluant l'infrarouge permet de distinguer les nuages des zones enneigées. Ainsi, dans le cadre d'un apprentissage supervisé, la labellisation de zones enneigées est simplifiée. Cependant, dans le cadre d'une analyse spatio-temporelle, le signal optique est bloqué principalement par la présence de nuages mais aussi par le manque de lumière, limitant spatialement ou temporellement les zones à étudier. Pour pallier ces limitations, des méthodes de remplissage de données manquantes peuvent être utilisées [88]. Cependant, si l'information est manquante, il est complexe de la retrouver, et les méthodes de remplissage sont lourdes pour une estimation fiable, pouvant introduire des biais dans l'analyse. En milieu montagneux, tout comme dans les zones tropicales, la présence de nuages est fréquente, limitant l'utilisation de l'imagerie optique. De plus l'étude du suivi du manteau neigeux demande une concentration des mesures sur la saison hivernale, où la luminosité est faible et les conditions météorologiques difficiles.

Une manière de contourner ces limitations repose sur l'utilisation conjointe de données de modalités différentes. Ainsi cette complémentarité s'observe dans [89], où les auteurs cherchent à prédire l'incertitude sur le pourcentage de couverture neigeuse fournit par le satellite optique MODIS à partir de données de températures. Cette démarche utilise l'imagerie satellitaire pour inférer sur des caractéristiques liées à la cryosphère, et valide les résultats par des mesures au sol. Dans ce cadre, on remarque que seul le potentiel spatial des images est exploité, au détriment de la résolution temporelle des mesures au sol. D'une manière similaire, l'étude proposée dans [90] aborde la détection de la couverture neigeuse (totale, humide et sèche) en complétant des données météorologiques avec des caractéristiques radar (coefficient de rétrodiffusion, cohérence et décomposition polarimétrique) exploitées à partir d'images satellitaires et des informations topographiques (pente, altitude, aspect et orientation). Le résultat final est obtenu par une

[82]: Varshney et al. (2018), « Snow and cloud discrimination using convolutional neural networks »

[83]: Dedieu et al. (2016), « On the Importance of High-Resolution Time Series of Optical Imagery for Quantifying the Effects of Snow Cover Duration on Alpine Plant Habitat »

[84]: Leigh et al. (2019), « Identifying and mapping very small (<0.5 $km^2$) mountain glaciers on coarse to high-resolution imagery »

[85]: Dietz et al. (2012), « Remote sensing of snow – a review of available methods »

[86]: Guo et al. (2020), « Extraction of snow cover from high-resolution remote sensing imagery using deep learning on a small dataset »

[87]: Liu (2021), « A Review of Deep Learning for Cryospheric Studies »

[88]: Kandasamy et al. (2013), « A Comparison Of Methods For Smoothing And Gap Filling Time Series Of Remote Sensing Observations : Application To MODIS LAI Products »

[89]: Dong et al. (2014), « Using Air Temperature to Quantitatively Predict the MODIS Fractional Snow Cover Retrieval Errors over the Continental United States »

[90]: Tsai et al. (2019), « Wet and Dry Snow Detection Using Sentinel-1 SAR Data for Mountainous Areas with a Machine Learning Technique »

méthode d'apprentissage automatique de type Forêt Aléatoire (*Random Forest* RF). Cette approche permet d'introduire un autre type de données, les données radar, qui permettent de contourner les limitations de l'imagerie optique. Ce type de données est particulièrement intéressant pour le suivi hivernal de caractéristiques de la neige et de la glace, en raison de la faible influence des conditions météorologiques sur les mesures radar. Nous proposons de détailler les principes de la télédétection radar dans la suite de ce travail et son application à l'étude de la cryosphère.

## 2.2 Synthetic Aperture Radar (SAR) pour la cryosphère

Nous présentons dans cette partie les grands principes de l'imagerie RADAR à Synthèse d'Ouverture (RSO ou *Synthetic Aperture Radar* SAR) des points de vue mécanismes, statistiques et limitations du signal rétrodiffusé. Pour plus d'informations sur la formation des images (échantillonnage, filtrage, focalisation) nous renvoyons le lecteur à [91, 92].

[91]: Nicolas (2019), *Une introduction au Radar à Synthèse d'Ouverture satellitaire*
[92]: Meyer (2019), « Spaceborne Synthetic Aperture Radar »

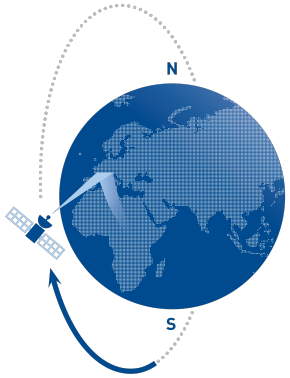



**Figure 2.9 –** Orbite ascendante d'un satellite SAR. Extrait des ressources pédagogiques de Tre Altamira

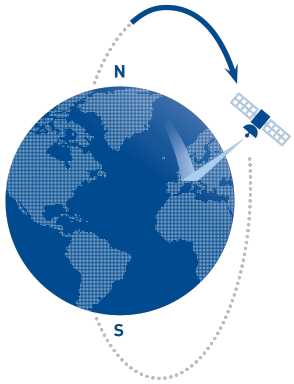
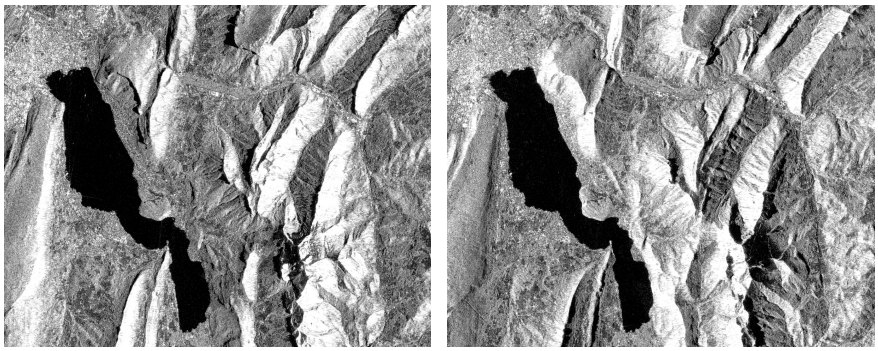



**Figure 2.10 –** Orbite descendante d'un satellite SAR. Extrait des ressources pédagogiques de Tre Altamira

### 2.2.1 Principes de la mesure SAR

Une autre manière de contourner le problème des données manquantes liées aux conditions météorologiques est d'utiliser l'imagerie SAR [93]. Les satellites SAR sont des systèmes actifs qui émettent des ondes électromagnétiques, une bande de fréquence autour d'une certaine fréquence centrale, qui sont réfléchies par la surface terrestre et renvoyées au capteur. En orbitant autour de la Terre, le satellite acquiert des images de la surface terrestre en balayant une zone donnée. On parle de période de revisite pour définir le temps nécessaire pour que le satellite repasse au-dessus de la même zone. De par la rotation de la Terre, nous pouvons différencier les passages du satellite en orbite descendante (passage du satellite du pôle Nord au pôle Sud, cf. figure 2.10) et en orbite ascendante (passage du satellite du pôle Sud au pôle Nord, cf. figure 2.9). Historiquement, les satellites ont leur fauchée d'acquisition orientée à droite (illustrée figure 2.11) afin de pouvoir observer l'océan arctique (au détriment du continent Antarctique). Les satellites de deuxième génération tel que TerraSAR-X se sont vus dotés d'une capacité de basculement pour pallier ce problème. L'angle de visée (aussi appelé angle off-nadir), soit l'angle formé entre la verticale et la direction du capteur, est typiquement compris entre 8° et 70°. Cet angle impacte directement l'empreinte au sol du capteur, et donc la résolution spatiale de l'image dans l'axe de visée. Respectivement les figure 2.11a et 2.11b illustrent une acquisition en passe descendante et ascendante obtenue avec Sentinel-1. Nous pouvons remarquer que les zones qui apparaissent sombres dans l'image en orbite descendante apparaissent claires dans l'image en orbite ascendante et inversement. Cette observation est particulièrement vraie en milieu de montagne, où les faces orientées vers le capteur dans une orbite sont comprimées alors qu'elles sont dilatées, voire invisibles dans l'autre orbite. Le lac (tache sombre centrale) apparaît sombre dans les deux orbites, dans ce cas le signal est réfléchi dans la direction opposée à celle du capteur.

[93]: Moreira et al. (2013), « A tutorial on synthetic aperture radar »



**(a)** Orbite descendante **(b)** Orbite ascendante

**Figure 2.11 –** Comparaison de deux images SAR acquises en orbite descendante (13 avril 2024) et ascendante (14 avril 2024) sur la région du lac d'Annecy par Sentinel-1. Les images sont acquises en polarisation VV et orthorectifiées. Extrait de Sentinel Hub.

La fréquence typique de ces systèmes actifs, dans le cadre satellitaire, varie de 0.3GHz (bande P) à quelques dizaines de GHZ (bande K), illustré dans la figure 2.12. C'est la longueur d'onde associée à la fréquence des satellites, très supérieure devant certains objets, qui explique principalement la pénétration des ondes électromagnétiques et la transparence des nuages à certaines fréquences dans les images obtenues. Une illustration du phénomène de pénétration est donnée pour les principales bandes de fréquences utilisées en SAR dans la figure 2.14. La bande P (100-30 cm) sera principalement employée pour cartographier et évaluer la biomasse végétale, grâce au lancement prochain de la mission Biomass (~ 2025). La bande L (30-15 cm) est souvent utilisée pour des applications nécessitant une résolution moyenne, telles que la surveillance géophysique et la cartographie de la biomasse et de la végétation, en raison de sa capacité de pénétration élevée. Quant à la bande S (15-7.5 cm), son utilisation croissante pour l'observation terrestre, notamment pour la surveillance agricole, étend les applications de la bande C à des densités de végétation plus élevées. La bande C (7.5-3.8 cm), considérée comme le pilier de l'imagerie SAR, est largement utilisée pour la cartographie globale, la détection des changements et la surveillance des zones avec une pénétration modérée, comme la glace et la navigation maritime océanique. La bande X (3.8-2.4 cm) est préférée pour les applications SAR nécessitant une résolution élevée, telles que la surveillance urbaine et la détection de glace et de neige, mais elle pénètre peu dans la couverture végétale. Enfin, les bandes K, Ku et Ka sont rarement utilisées en raison de diverses limitations telles que l'absorption atmosphérique élevée de l'eau ou la faible pénétration, les reléguant à des applications spécifiques ou à un usage rare dans le domaine SAR. Cependant pour la 7ème mission Earth Explorer de l'agence spatiale européenne (*European Space Agency* ESA), une des trois propositions en compétition était le satellite CoReH2O (COld REgions Hydrology High-resolution Observatory), dont le double capteur bande X et bande Ku, devait permettre d'améliorer la modélisation des cycles de l'eau et de l'énergie dans les hautes latitudes [94]. Malheureusement, cette mission n'a pas été retenue pour laisser place à la mission Biomass.

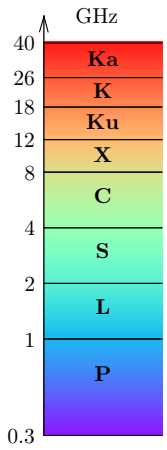



**Figure 2.12 –** Noms et bandes de fréquences des satellites radar (en GHz, échelle logarithmique)

[94]: Rott et al. (2008), « CoReH2O - A Ku- and X-Band SAR Mission for Snow and Ice Monitoring »

### Diffusion du signal radar

Il est possible de distinguer trois mécanismes de diffusion du signal radar :

1. La diffusion volumique
2. Le double rebond
3. La diffusion de surface

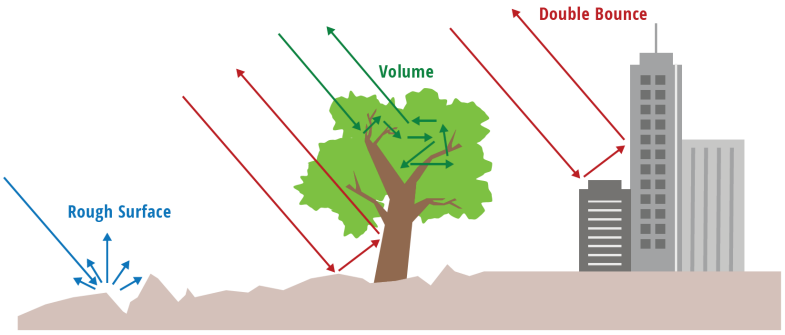



**Figure 2.13 –** Illustration des différents types de diffusion du signal radar. Extrait de [95]

Ces mécanismes sont illustrés dans la figure 2.13. La diffusion volumique est causée soit par la végétation, soit par des types de sols[19] qui permettent une pénétration du signal radar. Le signal est diffusé dans de multiples directions au fur et à mesure de sa propagation dans le milieu. Le double rebond qui renvoie directement le signal radar à l'antenne est causé par les angles droits. Il est souvent observé dans les zones urbaines (bâtiments, poteaux électriques, routes). Il peut aussi être causé par des objets naturels comme les troncs d'arbres. Cette réponse radar est forte puisque la majorité de l'énergie est renvoyée directement à l'antenne. Enfin la diffusion de surface est causée par la rugosité du sol. Cette diffusion est rencontrée dans la plupart des zones dont la surface est dénudée ou avec une végétation basse, dont le sol n'est pas poreux, et qui ne permet pas une pénétration du signal radar. En présence d'une rugosité faible, le signal radar rétrodiffusé est faible, ce cas est appelé réflexion spéculaire. En présence d'une rugosité élevée[20] , le signal radar rétrodiffusé est plus important, mais il est diffusé dans de multiples directions, ce cas est appelé diffusion diffuse.

**Précision (19)**
Les sols poreux et secs, tels que le sable ou la neige sèche par exemple, permettent une pénétration du signal radar.

**Précision (20)**
Nous donnons ici deux catégories de rugosité. La surface est caractérisée par l'écart-type de la hauteur des aspérités de la surface, noté $h$. Ici, la rugosité faible correspond à $h < \lambda/32$ et la rugosité élevée à $h > \lambda/2$ avec $\lambda$ la longueur d'onde du signal radar [95], basé sur le critère de Fraunhofer, pour un angle d'incidence de 0°. Il convient de noter que ce critère varie en fonction de l'angle d'incidence.

La pénétration du signal radar, illustrée dans la figure 2.14 pour une sélection de bandes de fréquences, indique la capacité de l'onde à pénétrer un milieu. Elle est déterminée par un ensemble de paramètres physiques, mais principalement par la fréquence du signal et la constante diélectrique du milieu. Dans le cas d'un matériau homogène, en considérant uniquement les pertes d'absorption, la profondeur de pénétration du signal $\delta_p$ est donnée, par la loi de Beer-Lambert, par : $\delta_p = 0.5\delta_e$, avec $\delta_e$ l'épaisseur de peau. Cette dernière fait intervenir la constante diélectrique complexe du milieu $\varepsilon_r = \varepsilon'_r - j\varepsilon''_r$, la longueur d'onde $\lambda$ du signal radar et l'angle de perte ou tangente de perte $\tan(\delta)$. Cette grandeur simplement définie par le rapport entre la partie imaginaire et la partie réelle de la constante diélectrique complexe, est souvent utilisée pour caractériser la perte d'énergie du signal radar lors de sa propagation dans un milieu ($\tan(\delta) = \varepsilon''_r/\varepsilon'_r$). En présence de tangente de perte très inférieure à 1 ($\varepsilon''_r \ll \varepsilon'_r$), $\delta_e$ peut être approximée par :

$$\delta_e \approx \frac{\lambda}{2\pi \tan(\delta)\sqrt{\varepsilon'_r}}.$$

Par exemple, dans le cas de glace polaire, la tangente de perte est inférieure à $10^{-3}$, pour une partie réelle de la constante diélectrique de l'ordre de $\varepsilon'_r \approx 3.2$ dans le domaine des micro-ondes [96], conduisant à une profondeur de pénétration variant de l'ordre de 1m pour un signal en bande X à près de 30m pour un signal en bande P. Dans le cas de matériaux hétérogènes, la profondeur de pénétration dépend aussi du coefficient d'extinction de diffusion, régit

[96]: Fujita et al. (2000), « A summary of the complex dielectric permittivity of ice in the megahertz range and its applications for radar sounding of polar ice sheets »

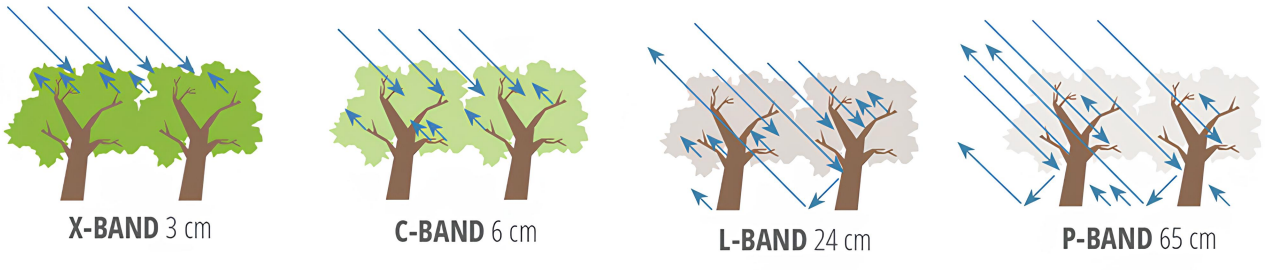



**Figure 2.14 –** Illustration de la pénétration des ondes électromagnétiques dans la canopée. Adaptée de [95]

par les théories de Mie ou de Rayleigh en fonction des tailles des particules présentes dans le milieu. Pour plus de détails sur les propriétés électriques de la glace et de la neige, nous renvoyons le lecteur à [97]. La permittivité complexe du milieu augmente avec la teneur en eau, menant à une diminution de la profondeur de pénétration du signal radar. En augmentant la fréquence du signal radar, la partie réelle de la constante diélectrique complexe diminue en opposition à la partie imaginaire, ce qui a pour conséquence une diminution de la profondeur de pénétration du signal radar. Ceci explique ainsi la figure 2.14 où l'onde électromagnétique pénètre plus profondément dans la canopée pour des fréquences plus basses. En parallèle, lorsque la largeur de bande, la résolution spatiale de l'image radar augmente. La largeur de bande est définie par la différence entre la fréquence maximale et minimale du signal radar. En première approximation nous pouvons dire que plus la fréquence centrale du signal radar augmente, plus la résolution spatiale de l'image radar augmente. Ceci est vrai par l'encombrement des fréquences dans le spectre (peut de place disponible à basse fréquence) et le rapport relatif entre la fréquence centrale et la largeur de bande.[21]

[97]: Glen et al. (1975), « The Electrical Properties of Snow and Ice »

**Précision (21)**
Pour une fréquence centrale de 100MHz, la largeur de bande correspondant à 10% équivaut à 10MHz, alors qu'à 10GHz, la même largeur de bande correspond à 1GHz.

À partir de la réflectivité du signal radar sur une cible, il est nécessaire de normaliser cette réflectivité pour remonter aux propriétés physiques de la cible. Il existe trois conventions pour exprimer la rétrodiffusion : beta naught, sigma zéro et gamma zéro ($\beta^0$, $\sigma^0$ et $\gamma^0$). Le coefficient $\beta^0$ est normalisé par unité de surface dans le plan perpendiculaire à la direction de propagation du signal radar. Ces valeurs ne prennent pas en compte l'incidence locale et donnent des variations d'intensité de rétrodiffusion en fonction de l'angle d'incidence. Le coefficient $\sigma^0$ est calibré sur l'antenne, par rapport à une unité de surface au sol avec le plan localement tangent à l'ellipsoïde. Le coefficient $\gamma^0$ est calibré par unité de surface dans le plan perpendiculaire à la direction de propagation du signal radar. Ces deux derniers coefficients[22] sont moins sensibles à la topographie que le coefficient $\beta_0$. Sans compensation les signatures plus subtiles induites par la couverture terrestre sont souvent masquées. C'est pourquoi, dans la suite de ce travail, nous nous concentrerons sur le coefficient $\sigma^0$, pour caractériser la rétrodiffusion du signal radar, qui est le plus souvent utilisé dans la littérature, et sa version logarithmique $\sigma^0_{\text{dB}} = 10\log_{10}(\sigma^0)$, pour des raisons de dynamique de valeurs.

**Précision (22)**
La relation entre $\beta^0$ et $\sigma^0$ est donnée par : $\sigma^0 = \beta^0 \sin(\theta_l)$,
la relation entre $\beta^0$ et $\gamma^0$ est donnée par : $\gamma^0 = \beta^0 \tan(\theta_l)$, avec $\theta_l$ l'angle local d'incidence.

### Données SAR

Les données SAR sont fournies par les agences spatiales ou les industrielles sous forme de produits. Usuellement, nous pouvons distinguer trois niveaux de produits :

1. Niveau 0 : données brutes (raw) non pré-traitées et non focalisées.

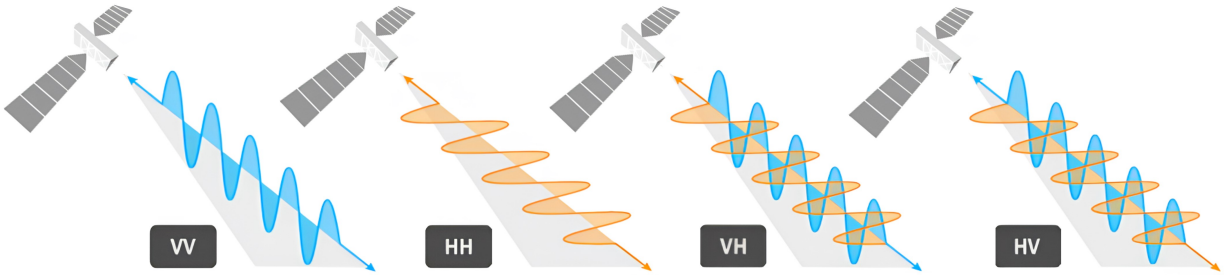



**Figure 2.15 –** Illustration de la polarisation du signal radar. Adaptée des ressources pédagogiques de ArcGIS

**Précision (23)**
Le détail complet du traitement des acquisitions Sentinel 1 est accessible en ligne : https://sentiwiki.copernicus.eu.

**Définition (24)**
*Focalisation SAR:* Technique consistant à combiner de manière cohérente des acquisitions effectuées à partir des positions différentes afin de permettre une amélioration de la résolution.

**Définition (25)**
*InSAR:* Interferometric Synthetic Aperture Radar est une technique qui utilise deux ou plusieurs images de radar à synthèse d'ouverture (SAR) pour générer des cartes de déformation de surface ou d'élévation, en exploitant la différence de phase des ondes. Cette technique peut potentiellement mesurer des changements de déformation à l'échelle du millimètre sur des périodes allant de quelques jours à quelques années[98].

**Définition (26)**
*Détection SAR:* Étape du traitement au cours de laquelle l'intensité du signal est déterminée pour chaque valeur de pixel. La détection supprime l'information de phase. Extrait de ESA.

**Définition (27)**
*Multi-vues SAR:* (Multi-looking en anglais) Technique de moyennage visant à la fois à réduire le bruit de speckle (cf 27). De plus, de par l'acquisition SAR, les pixels sont souvent rectangulaires (par exemple 1m x 3m). L'opération de multi-vues permet ainsi d'obtenir un pixel carré (par exemple 3m x 3m). La résolution finale est donc réduite pour une image moins bruitée.

2. Niveau 1 : données pré-traitées, contenant les informations de calibration et de géoréférencement. Dans ce niveau deux types de produits sont disponibles : les produits complexes et sans géoréférencement, et les produits calibrés, géo-référencés.
3. Niveau 2 : Obtenus à partir des données de niveau 1, ces produits fournissent des informations ou paramètres géophysiques.

Chaque agence utilise ses propres acronymes pour désigner les différents produits. Dans la suite de ce travail, nous utiliserons exclusivement des données de niveau 1. Les produits de niveau 1 de l'ESA sont fournis sous deux formes : Single Look Complex (SLC) et Ground Range Detected (GRD).

Les produits SLC de niveau 1[23] se composent de données SAR focalisées[24] géo-référencées à l'aide des données d'orbite et d'altitude du satellite, et fournies en géométrie radar. Ces données sont complexes (i.e de la forme $a+ib$), et possèdent à la fois l'information de phase et d'amplitude. L'utilisation de ces données est souvent liée à l'utilisation de la phase pour la construction de données d'interférométrie radar (*Interferometric Synthetic Aperture Radar* InSAR) [25] .

Les produits GRD de niveau 1, quant à eux, se composent de données SAR focalisées qui ont été détectées[26] , subit une opération de multi-vues[27] et projetées en géométrie sol en utilisant un modèle numérique de terrain (MNT). Les valeurs des pixels représentent l'amplitude détectée. Les informations de phase sont perdues. Le produit résultant a des pixels de résolution approximativement carrés par l'opération de multi-vues. Ceci permet d'obtenir une image avec un speckle réduit au détriment de la résolution spatiale réduite.

### Polarisation

Comme les satellites SAR possèdent leur propre source d'illumination de la scène considérée, ils ont la capacité de contrôler la polarisation de l'onde électromagnétique émise. La polarisation décrit l'orientation du plan d'oscillation de l'onde de signal en propagation. La majorité des capteurs SAR actuels sont polarisés linéairement et émettent des formes d'onde polarisées horizontalement et/ou verticalement. Dans ces systèmes à polarisation linéaire, l'orientation de ce plan d'oscillation est constante le long du trajet de propagation de l'onde électromagnétique. Il existe quatre types de polarisation linéaire : HH (horizontalement émis, horizontalement reçu), VV (verticalement émis, verticalement reçu), HV (horizontalement émis, verticalement reçu) et VH (verticalement émis, horizontalement reçu), illustrés dans la figure 2.15. Les données SAR peuvent être *single-pol*, *dual-pol* ou *quad-pol*, fournissant respectivement 1, 2 ou les 4 polarisations disponibles. Les satellites opérant en *dual-pol*, acquièrent des images en co-polarisation et polarisation croisée

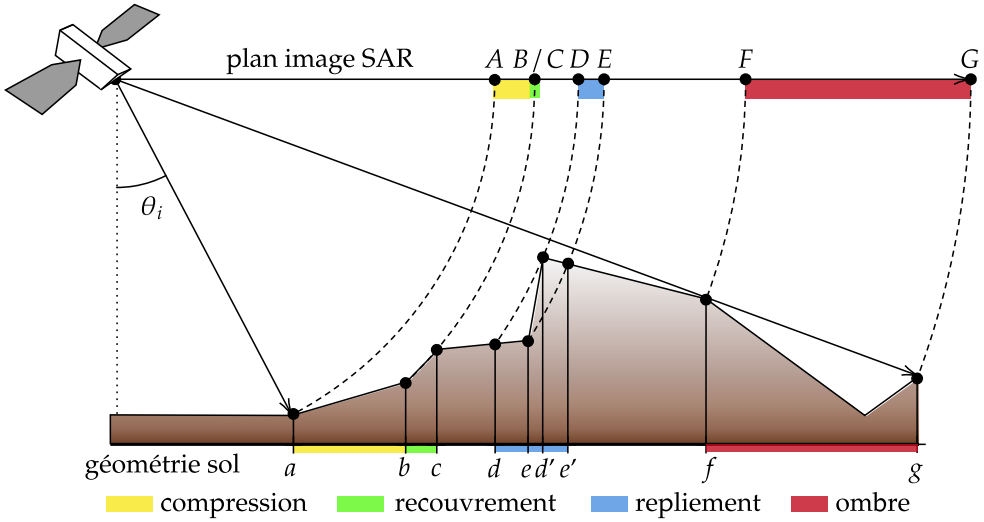



**Figure 2.16 –** Illustration des différentes distorsions géométriques du signal radar. Adapté de [103]

(soit les couples VV/VH et HH/HV). La polarimétrie donne une information supplémentaire sur la nature de la cible étudiée [99]. Soit $S_{XX}$ la rétrodiffusion en polarisation $XX$ avec $X \in \{H, V\}$, certains canaux polarimétriques sont plus sensibles à un type de diffusion [95] :

- Diffusion de surface rugueuse : $S_{VV} > S_{HH} > S_{HV}$ ou $S_{VH}$
- Double rebond : $S_{HH} > S_{VV} > S_{HV}$ ou $S_{VH}$
- Diffusion volumique : principalement $S_{HV}$ ou $S_{VH}$

Le lecteur pourra se référer à [100] pour une étude plus approfondie sur la polarimétrie SAR et ses applications.

[99]: Van Zyl et al. (2011), « Basic principles of SAR polarimetry »

[95]: Flores et al. (2019), « Synthetic Aperture Radar (SAR) Handbook »

[100]: Ferro-Famil et al. (2014), « Radar Polarimetry Basics and Selected Earth Remote Sensing Applications »

### Distorsion et bruit du signal radar

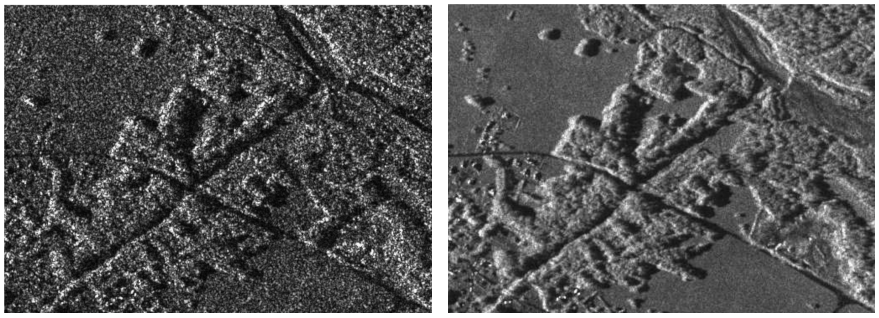
En raison de la géométrie d'acquisition et du caractère actif des satellites SAR, de nombreuses distorsions apparaissent dans les images SAR, d'autant plus marquées dans les zones montagneuses. Trois grandes catégories de distorsions peuvent être identifiées :

1. La distorsion atmosphérique, qui est causée par la propagation du signal radar dans l'atmosphère, et son interaction avec les particules atmosphériques (vapeur d'eau principalement[28] ).
2. La distorsion radiométrique, qui est causée par la variation la topographie de la zone étudiée. À cible équivalente (taille et nature) et à incidence équivalente, la rétrodiffusion radar varie en fonction de l'angle formé entre la surface et la direction du capteur. Cette distorsion peut être corrigée par des modèles de correction radiométrique [102].
3. La distorsion géométrique, qui est causée par la géométrie de l'acquisition (angle d'incidence) et la topographie de la zone étudiée. Elle se traduit par une déformation des pixels entre le plan radar et la projection géographique.

**Précision (28)**
Bien que dans le cadre de l'étude et analyse du coefficient de rétrodiffusion $\sigma_0$ cette distorsion est souvent négligée, elle se doit d'être prise en compte dans le cadre de l'analyse de la phase du signal radar et des techniques d'interférométrie [101]

[102]: Small (2011), « Flattening Gamma »

**Distorsion géométrique :** Nous présentons pour les besoins de la suite de ce travail, les distorsions géométriques du signal radar qui de par leur nature, ne peuvent être corrigées post-acquisition.



**(a)** Sans filtrage de speckle.

**(b)** Filtrage multi-temporel (20 images).

**Figure 2.17 –** Image SAR (TerraSAR-X) de zones agricoles homogènes polluées par du bruit de speckle (a) et filtrée par un filtre multi-temporel (20 images) (b). Extrait de [91].

On peut dénombrer trois grandes distorsions géométriques, illustrées dans la figure 2.16

1. Les zones de compression (en jaune) et d'étirement (en rouge) : le segment *ab* en géométrie sol se retrouve compressé en géométrie radar *AB*. Cet effet décroît en augmentant l'angle d'incidence $\theta_i$. Cette distorsion est aussi appelée *foreshortening* en anglais.
2. Les zones de repliement (en bleue) : le segment *ed'* en géométrie sol se retrouve inversé en géométrie radar *DE*, menant à une indétermination entre le haut et le bas de la montagne. Cette distorsion est aussi appelée *layover* en anglais. Un cas particulier de cette distorsion est le recouvrement (en vert), où la compression est maximale et le repliement n'est pas encore atteint : le segment *bc* en géométrie sol se retrouve compressé en un point *B/C* en géométrie radar. Cet effet décroît également en augmentant l'angle d'incidence $\theta_i$.
3. Les zones d'ombre (en rouge) : le segment *fg* en géométrie sol est invisible par le capteur radar. Ces zones sont ainsi reconnaissables dans les images SAR par un aspect sombre. Cet effet augmente en augmentant l'angle d'incidence $\theta_i$. Cette distorsion est aussi appelée *shadow* en anglais.

**Précision (29)**
Nous utiliserons dans la suite de ce manuscrit principalement le terme anglais *speckle* pour désigner le bruit de chatoiement.

**Définition (30)**
*Bruit additif:* Le bruit additif est un bruit qui s'ajoute au signal d'origine. Soit un signal $x$ et un bruit $n$, le signal observé bruité est donné par $y = x + n$. Le bruit additif $n$ est souvent modélisé par une distribution gaussienne de moyenne nulle $\mathcal{N}(0, \sigma^2)$.

**Définition (31)**
*Bruit multiplicatif:* Le bruit multiplicatif est un bruit qui est proportionnel au signal d'origine. Soit un signal $x$ et un bruit $n$, le signal observé bruité est donné par $y = x \cdot n$.

**Bruit du signal radar :** En plus des distorsions géométriques présentées, l'image SAR se voit polluée par du bruit. Ce bruit est principalement de deux types : le bruit thermique et le bruit de chatoiement ( *speckle*[29] ). Le bruit thermique, causé par l'agitation thermique des électrons dans les composants électroniques du capteur, est de nature additive[30] et visible principalement quand le signal est faible. La plupart des données fournissent soit une information sur la quantification de bruit, soit directement la donnée corrigée (correspondant au niveau 1 des données fournies par l'ESA). Le speckle quant à lui, est particulièrement complexe à traiter de par sa nature multiplicative[31] . Le lecteur pourra se référer à [104] pour un aperçu sur l'ensemble des modèles de bruit de speckle. Plusieurs stratégies peuvent être mises en place pour réduire le bruit de speckle [105]. Les méthodes les plus classiques sont basées sur du filtrage spatial (filtre de Lee) ou du filtrage temporel [106]. Les méthodes de filtrage multi-temporel, qui exploitent la corrélation temporelle des images, sont particulièrement efficaces pour réduire le bruit de speckle tout en préservant les détails de l'image. Plus récemment, des méthodes d'apprentissage profond ont été développées pour réduire le bruit de speckle, en particulier les réseaux de neurones convolutionnels (CNN) [107]. Les figures 2.17a et 2.17b illustrent l'impact du bruit de speckle sur une image SAR et l'effet d'un filtrage multi-temporel sur cette même image.

[104]: Yue et al. (2021), « Synthetic Aperture Radar Image Statistical Modeling »

[105]: Singh et al. (2021), « A Review on SAR Image and its Despeckling »

[106]: Quegan et al. (2000), « Multitemporal ERS SAR analysis applied to forest mapping »

[107]: Dalsasso et al. (2022), « As If by Magic : Self-Supervised Training of Deep Despeckling Networks With MERLIN »

### 2.2.2 Statistiques du signal

La statistique du signal radar peut être modélisée de plusieurs manières différentes. Nous nous attachons ici au détail des deux principales : les modèles de diffusion cohérente et les modèles empiriques.

**Modèles de diffusion cohérente :** La modélisation du bruit de speckle a été proposée par Goodman [6] dans le cadre de l'étude des lasers. Ce modèle, appelé modèle de diffuseur cohérent (*coherent scatterer model*), est basé sur l'hypothèse que le signal radar est la somme cohérente de $N$ diffuseurs élémentaires. Soit $N$ un nombre de diffuseurs idéaux dans une cellule de résolution. Si les hypothèses d'indépendance des diffuseurs ( i.i.d[32] ) et d'uniformité du gain de l'antenne sur la zone considérée sont vérifiées, la rétrodiffusion radar est donnée par la somme cohérente de ces diffuseurs :

[6]: Goodman (1976), « Some fundamental properties of speckle »

$$S = \sum_{i=1}^{N} a_i \exp^{j\phi_i} = A \exp^{j\Phi},$$

avec $a_i$ l'amplitude du diffuseur $i$ et $\phi_i$ sa phase. La simplification de ce modèle est le modèle de speckle de Rayleigh, si les hypothèses suivantes supplémentaires sont vérifiées :

- Les amplitudes ($a_i$) et les phases ($\phi_i$) des diffuseurs sont indépendants.
- Les diffuseurs sont aléatoirement distribués dans la cellule de résolution et la distance entre les diffuseurs est largement supérieure à la longueur d'onde du signal radar.
- Les amplitudes ($a_i$) sont i.i.d, et les premiers et seconds moments sont finis.
- Le nombre $N$ de diffuseurs est grand, permettant d'appliquer le théorème central limite[33] .

**Définition (32)**
*i.i.d:* Indépendant et identiquement distribué. Soit $X_1, X_2, \ldots, X_n$ une suite de variables aléatoires, elles sont dites i.i.d si elles sont indépendantes et suivent la même loi de probabilité.

Sous ces hypothèses, il découle que la densité de probabilité de la phase $\Phi$ est uniforme sur $[-\pi, \pi]$, l'intensité $I = |S|^2$ suit une loi exponentielle, et l'amplitude $A = \sqrt{I} = |S|$ suit une loi de Rayleigh. Les densités de probabilité associées sont données dans le tableau 2.3. La phase seule n'apporte pas d'information sur la nature de la cible, mais l'amplitude et l'intensité du signal radar sont des paramètres importants pour la classification des cibles. Une information statistique intéressante est le coefficient de variation (CV)

$$\mathrm{CV} = \frac{\sigma}{\mu}, \tag{2.2}$$

où $\sigma$ est l'écart-type et $\mu$ la moyenne du signal. Ce coefficient donne une indication sur l'homogénéité de la zone. Si CV = 1 en intensité ou CV $\approx 0.523$ en amplitude, nous nous trouvons dans le cas du speckle pleinement développé, et les fluctuations du speckle sont du même ordre de grandeur que la moyenne du signal. À mesure que le coefficient de variation augmente, l'hétérogénéité de la zone augmente. Si le nombre de diffuseurs diminue alors le coefficient de variation diminue, et le speckle passe d'un speckle pleinement développé à un speckle partiellement développé.

**Précision (33)**
Nous pouvons remarquer que l'hypothèse de proximité des diffuseurs au vu de la longueur d'onde est lié à l'hypothèse du nombre $N$ qui doit être *grand*. Par exemple, pour un pixel de résolution 10 $m \times 10$ $m$ (résolution de Sentinel-1), la longueur d'onde en bande C est de 6 $cm$. Si on considère que *largement supérieur* signifie minimum 10 fois la longueur d'onde, alors la distance minimale entre deux diffuseurs doit être de 60 $cm$, menant à un nombre de $N = 256$ diffuseurs. Typiquement, en présence d'un faible nombre de diffuseurs (zone urbaine homogène) ou lorsque la résolution du capteur est importante au vu de la longueur d'onde, le modèle de speckle de Rayleigh sort de son cadre d'application.

L'opération de multi-vues est souvent utilisée et de par ce processus de moyennage, modifie les densités de probabilité de l'intensité et de l'amplitude du signal radar. Le moyennage de distributions exponentielles donnant des distributions de type gamma et le moyennage de distributions de Rayleigh donnant des distributions de type Rayleigh-Nakagami [91] ou de type racine

[91]: Nicolas (2019), *Une introduction au Radar à Synthèse d'Ouverture satellitaire*

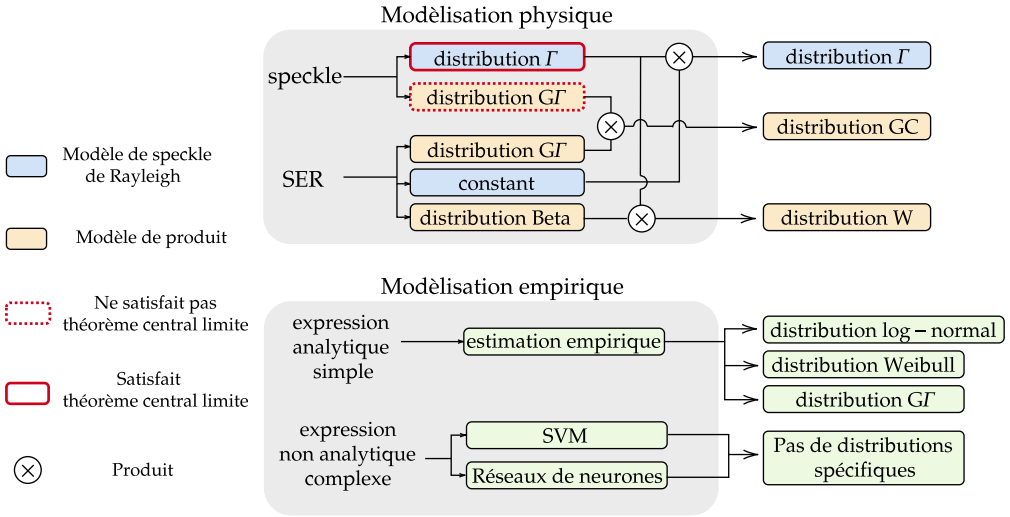



**Figure 2.18 –** Schéma illustrant les différentes hypothèses de modèle de densité de probabilité d'un pixel SAR en intensité. Adapté de [112]

carrée de gamma [108].

Une généralisation de ce modèle a été proposée par [109], sous la dénomination du modèle de produit. En effet le modèle de speckle de Rayleigh peut être vu comme le produit entre une section efficace radar (SER) constante et une distribution exponentielle modélisant le speckle. Le modèle de produit généralise cette idée en considérant une section efficace radar suivant une certaine distribution, et une distribution de speckle respectant les hypothèses du modèle de Rayleigh. Là où le modèle de produit permet de gagner en lien avec la physique de la cible, il fait intervenir des distributions plus complexes. La différence et les connexions entre ces modèles sont illustrées dans la figure 2.18 dans le cas de données multi-vues en intensité. L'utilisation d'une transformation logarithmique est courante sur les données SAR. L'intérêt est double, il permet de réduire la dynamique des images. Il est commun d'avoir des valeurs d'intensité ou d'amplitude comprise entre $10^{-4}$ et $10^{4}$, ce qui donne après transformation des valeurs comprises entre $-40dB$ et $40dB$, facilitant la visualisation des images. De plus le speckle étant un bruit de nature multiplicative, la transformation logarithmique permet de linéariser le modèle : $10\log_{10}(R * s) = 10\log_{10}(R) + 10\log_{10}(s)$, où $R$ est la rétrodiffusion radar et $s$ le speckle. Au premier ordre, nous pouvons dire que nous nous trouvons en présence d'un bruit blanchi additif [110]. Cependant, notamment avec un nombre de vue faible ou une seule vue, la distribution résultante est asymétrique vers les faibles valeurs et non centrées. La véritable distribution de la transformation logarithmique est donnée par une distribution de type Fisher-Tippett [111].

**Modèles empiriques :** Il est important de noter que les pixels SAR correspondent à une mesure physique moyenne. Ils sont donc strictement positifs et peuvent être modélisés par des distributions de probabilité positives sans a priori physique sur la cible, sans les contraintes des modèles de diffusion cohérente. Pour modéliser ces données, de nombreux modèles de probabilité ont été proposés [112, 113] pour tenir compte de la variabilité des scènes observées. Une sélection est donnée dans la figure 2.18. Les lois log normal et Weibull sont ainsi adaptées pour modéliser des images de moyenne résolution respectivement en amplitude pour des zones urbaines homogènes et ampli-

[108]: Oliver et al. (2004), *Understanding Synthetic Aperture Radar Images*

[109]: Ward (1981), « Compound representation of high resolution sea clutter »

[110]: Arsenault et al. (1976), « Properties of speckle integrated with a finite aperture and logarithmically transformed »

[111]: Xie et al. (2002), « Statistical properties of logarithmically transformed speckle »

[112]: Yue et al. (2021), « Synthetic Aperture Radar Image Statistical Modeling »
[113]: Deledalle et al. (2018), « Mathematical models for very high resolution SAR data and their applications »

tude ou intensité pour tout type de scène. Sans a priori de distribution sur les données, il est possible de traiter les images SAR en utilisant des méthodes non-paramétriques. Ainsi des méthodes de fenêtre de Parzen-Rosenblatt (estimation de la densité du noyau) ou méthodes Séparateurs à Vaste Marge (SVM ou *Support Vector Machine* SVM) ont pu être utilisées pour estimer la densité de probabilité des données SAR [114] dans le cadre de classification de cibles. œ Finalement, un récapitulatif des distributions utilisées dans ce travail est donné dans le tableau 2.3 et des exemples de ces distributions sont illustrés dans la figure 2.19. Le détail des fonctions spéciales utilisées dans ce travail est donné en annexe A.1.

[114]: Mantero et al. (2003), « Partially supervised classification of remote sensing images using SVM-based probability density estimation »

**Estimation et moments statistiques :** Parmi les statistiques mono-variées à notre disposition, nous rappelons les principaux moments statistiques utilisés dans la suite de ce travail :

- Moment d'ordre 1 : l'espérance $E[X] = \mu$.
- Moment centré d'ordre 2, la variance :

$$\sigma^2 = E[(X - \mu)^2],$$

  avec l'écart-type $\sigma = \sqrt{\sigma^2}$.
- Moment centré réduit d'ordre 3, le coefficient d'asymétrie ou skewness :

$$\gamma_1 = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right]. \tag{2.3}$$

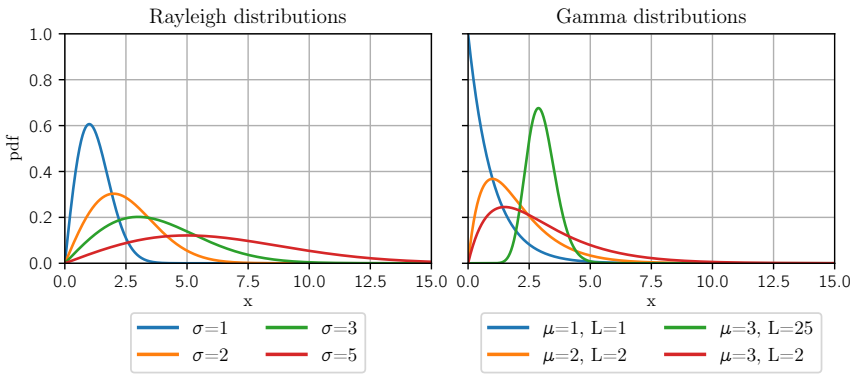
- Moment centré réduit d'ordre 4, l'aplatissement ou kurtosis :

$$\gamma_2 = E\left[\left(\frac{X - \mu}{\sigma}\right)^4\right]. \tag{2.4}$$

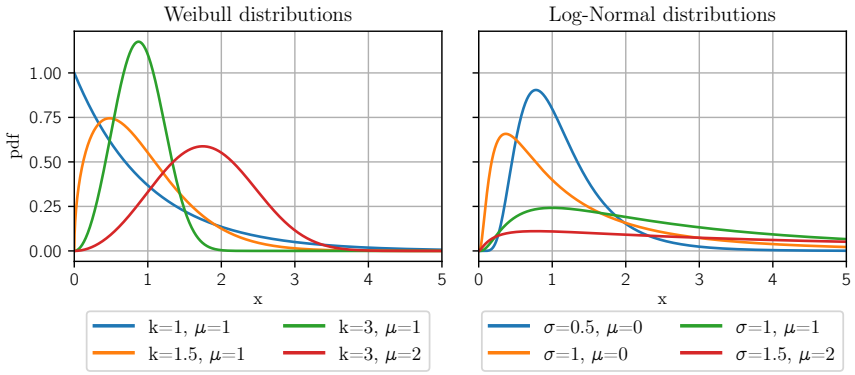
L'asymétrie et l'aplatissement sont des indicateurs de la forme de la distribution. L'asymétrie mesure la symétrie de la distribution autour de la moyenne. Une distribution symétrique a une asymétrie nulle. L'aplatissement mesure la concentration des valeurs autour de la moyenne. Une distribution normale a un aplatissement et une asymétrie nuls. L'estimation des paramètres des distributions est un point crucial dans l'analyse statistique des images SAR. Les méthodes classiques d'estimation des paramètres sont basées sur la méthode des moments ou la méthode du maximum de vraisemblance. Cependant leur

**Table 2.3 –** Tableau des distributions utilisées dans ce travail et leur fonction de densité de probabilité (pdf), pour un support $x > 0$.

| **Distribution** | **Densité de probabilité (PDF)** | **Paramètres** |
|---|---|---|
| Exponentielle | $\mathcal{E}(x;\lambda) = \lambda e^{-\lambda x}$ | Taux $\lambda > 0$ |
| Gamma | $\mathcal{G}(x;\mu,L) = \exp\left(-\frac{xL}{\mu}\right) \cdot \left(\frac{L}{\mu}\right)^L \cdot \frac{x^{L-1}}{\Gamma(L)}$ | Nombre de vues $L > 0$ et forme $\mu > 0$ |
| Log-normal | $\mathcal{O}(x;\mu,\sigma) = \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right] \cdot \frac{1}{x\sigma\sqrt{2\pi}}$ | Localisation $\mu > 0$ et variance $\sigma > 0$ |
| Rayleigh | $\mathcal{R}(x;\sigma) = \frac{x}{\sigma^2} \exp\left[-\frac{x^2}{2\sigma^2}\right]$ | Échelle $\sigma \quad x \geq 0, \quad \sigma > 0$ |
| Weibull | $\mathcal{W}(x;\mu,k) = \frac{k}{\mu} \cdot \left(\frac{x}{\mu}\right)^{k-1} \cdot \exp\left(-\frac{x^k}{\mu^k}\right)$ | Forme $k > 0$ et échelle $\mu > 0$ |




**(a)** Lois issues du modèle de speckle de Rayleigh (intensité et amplitude), Rayleigh à gauche et Gamma à droite




**(b)** Lois issues de modèle empirique, Weibull à gauche et log-normale à droite

**Figure 2.19 –** Graphiques des fonctions de densité de probabilité des distributions utilisées dans ce travail en fonction des deux types de modèles considérés, sur une sélection de paramètres.

utilisation peut s'avérer difficile. La Méthode des Log-Cumulant (*Method of Log-Cumulant* MoLC) [115], est une méthode alternative qui permet d'estimer les paramètres des distributions de manière robuste et avec une précision accrue [116]. L'estimation du log-cumulant d'ordre $r$ est donnée par la formule de récurrence :

$$\hat{\kappa}_r = \hat{m}_r - \sum_{i=1}^{r-1} \binom{r-1}{i-1} \hat{\kappa}_i \hat{m}_{r-i}, \tag{2.5}$$

avec $\hat{\kappa}_1 = \hat{m}_1$, où $\hat{m}_1$ est le moment d'ordre 1 (moyenne), $\hat{m}_r = \frac{1}{n}\sum_{i=1}^{n} \ln z_i^r$, $r \in \mathbb{N}, r > 1$ et $z_i$, étant un échantillon d'une variable aléatoire continue $Z$ avec une fonction de densité de probabilité définie sur $\mathbb{R}^+$. L'utilisation des log-cumulants permet l'introduction de l'opérateur de convolution de Mellin qui peut décrire une grande variété de lois statistiques à 3 paramètres (comme les distributions kappa ou Pearson I) uniquement à partir des lois élémentaires Gamma et Gamma inverse.

L'intérêt de la méthode MoLC est immédiat dans le cas par exemple de la distribution de Weibull (donnée dans le tableau 2.3). En effet la méthode de maximum de vraisemblance, sous hypothèse de $N$ réalisations $(x_i, i = 1, \ldots, N)$ issues de cette loi, donne :

[115]: Nicolas et al. (2002), « Introduction to second kind statistics : Application of log-moments and log-cumulants to the analysis of radar image distributions »

[116]: Krylov et al. (2013), « On the Method of Logarithmic Cumulants for Parametric Probability Density Function Estimation »

**Table 2.4 –** Tableau des principaux satellites radar de type SAR et leurs applications pour la cryosphère adapté de [16]

| Bande ($\lambda$ cm) | Fréquence (GHz) | Missions | Applications |
|---|---|---|---|
| P (30 - 100) | 0.3 - 1.0 | Biomass (~2025) | Tomographie |
| L (15 - 30) | 1.0 - 2.0 | NISAR (2024),<br>ROSE-L (~2027),<br>ALOS (2006-2011),<br>JERS-1 (1992-1998) | Tomographie, glace, SWE<br>carte de vitesses, carte de déplacements |
| S (7.5 - 15) | 2.0 - 4.0 | NISAR (2024) | Hydrologie de la neige |
| C (3.75 - 7.5) | 4.0 - 8.0 | Sentinel-1 (2014- ),<br>Envisat (2002-2012),<br>Radarsat-1/2 (1995-2013, 2007- ),<br>ERS-1/2 (1991-2000, 1995-2011) | Classification de la neige humide/sèche,<br>cartographie de la ligne du névé,<br>cartographie de la ligne de la neige,<br>vélocité |
| X (2.4 - 3.75) | 8.0 - 12.5 | TerraSAR-X (2007-),<br>TanDeM-X (2010-),<br>PAZ (2018-),<br>ICEYE (2018-) | Capacités 3D, changement d'altitude,<br>dynamique des masses,<br>vitesse des glaciers |

$$\begin{cases} (\hat{\mu})^k = \frac{1}{N}\sum_{i=1}^{N}(x_i)^k \\ \\ \frac{1}{\hat{\kappa}} = \frac{1}{N}\sum_{i=1}^{N}\log x_i\left[\left(\frac{x_i}{\mu}\right)^k - 1\right] \end{cases} \tag{2.6}$$

Et pour la méthode des moments :

$$\begin{cases} \hat{m}_1 = \Gamma\left(1+\frac{1}{k}\right)\mu \\ \\ \hat{m}_2 = \Gamma\left(1+\frac{2}{k}\right)\mu^2 \end{cases} \tag{2.7}$$

Nous pouvons remarquer que les deux méthodes ne proposent pas des systèmes d'équations simples à résoudre, et nécessitent une recherche numérique des deux paramètres $\mu$ et $k$. La méthode MoLC donne directement l'estimation des paramètres :

$$\begin{cases} \hat{k} = \sqrt{\frac{\Psi(1,1)}{\tilde{\kappa}_2}} \\ \\ \hat{\mu} = \exp\left(-\frac{\Psi(1)}{k}\right)\cdot\exp\tilde{\kappa}_1 \end{cases} \tag{2.8}$$

avec $\Psi(i,j)$ les fonctions polygamma, $\Psi(i)$ les fonctions digamma, et $\tilde{\kappa}_i$ les log-cumulants empiriques d'ordre $i$. Le détail des fonctions spéciales utilisées dans ce travail est donné en annexe A.1.

### 2.2.3 Analyse SAR pour la cryosphère

La synergie entre l'imagerie optique et le SAR, permet de surpasser les difficultés de l'optique en ajoutant une information provenant du SAR [117]. Ces chaînes de traitement reposent sur la fusion d'informations entre les deux approches et sont souvent utilisées dans la caractérisation spatiale de la cryosphère, de la couverture neigeuse [118], à la différenciation entre neige humide et neige sèche [119]. La performance de ces approches repose sur l'exploitation de données nombreuses et variées, souvent au détriment de la flexibilité, à partir de méthodes simples comme des seuillages ou index. [90] donne un

[117]: Pettinato et al. (2009), « Snow cover mapping by using optical and SAR data »

[118]: Wendleder et al. (2018), « Mapping Snow Cover Extent Using Optical and SAR Data »

[119]: He et al. (2017), « Dry and Wet Snow Cover Mapping in Mountain Areas Using SAR and Optical Remote Sensing Data »

[90]: Tsai et al. (2019), « Wet and Dry Snow Detection Using Sentinel-1 SAR Data for Mountainous Areas with a Machine Learning Technique »

aperçu intéressant dans le cadre de l'étude de la couverture neigeuse, sur la complexité de ces systèmes, qui accomplissent des performances supérieures à chaque approche séparément. Cette étude en complément avec [16] démontre l'intérêt de l'apport des mesures supplémentaires dans le cadre de l'étude de la neige notamment l'orientation, la pente et l'altitude des zones étudiées. Cette démarche est étayée dans [120] pour la détection des glaciers rocheux, où la comparaison de méthodes utilisant soit uniquement des données satellitaires soit des données satellitaires et des attributs topographiques, permet de diviser par 2 le nombre de faux positifs dans le second cas.

[16]: Taylor et al. (2021), « Remote sensing of the mountain cryosphere »

[120]: Brenning (2009), « Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection »

Cependant l'utilisation conjointe de données optiques et SAR reste limitée par la disponibilité des données optiques, souvent affectées par les nuages, et par la volonté de disposer des acquisitions au même instant. L'utilisation de données SAR seule a trouvé de nombreuses applications dans l'étude de la cryosphère, par la variété des capteurs disponibles, offrant diverses polarisations, fréquences et résolutions spatiales [36]. Un aperçu des principaux satellites SAR associés à leurs applications et bandes de fréquences est donné dans le tableau 2.4.

[36]: Tsai et al. (2019), « Remote Sensing of Snow Cover Using Spaceborne SAR »

En raison de la facilité de traitement avec un seul capteur, les données SAR ont été largement utilisées dans les études de neige, principalement avec un capteur en bande X [121] ou en bande C [122]. Néanmoins, l'utilisation de données SAR en bande C s'est avérée plus appropriée pour analyser les propriétés de la neige [16].

Le lancement des satellites Sentinel-1A et 1B en bande C, fonctionnant à une fréquence de 5.405GHz, a permis une surveillance continue de l'Europe occidentale et de vastes régions [123, 124], avec une résolution de l'ordre de 10m. Ces satellites[34] ont permis un accès gratuit aux données en offrant des acquisitions répétées tous les 6 jours. Les deux satellites produisant, des séries temporelles denses d'images qui sont particulièrement intéressantes pour un suivi régulier et sur le long terme de caractéristiques de la cryosphère [125, 126].

[123]: Lievens et al. (2019), « Snow depth variability in the Northern Hemisphere mountains observed from space »

[124]: Thakur et al. (2016), « Estimation of Snow Cover Area, Snow Physical Properties and Glacier Classification in Parts of Western Himalayas Using C-Band SAR Data »

**Précision (34)**

Depuis décembre 2021, le satellite Sentinel-1B est indisponible, rendant la période de revisite à 12 jours. Le satellite PAZ, quant à lui, propose une résolution de 3m, et une période de revisite de 11 jours.

En se concentrant sur le signal radar, plusieurs approches peuvent être choisies. Sur la compréhension de la physique du signal, des comparaisons ont été menées dans [127] et [128] entre la valeur du coefficient de rétrodiffusion $\sigma_0$ du satellite et celles obtenues par des modèles physiques. Ces approches ont un double intérêt : elles permettent de valider les modèles tels que celui proposé par [129], mais aussi de garder une explicabilité sur l'analyse du signal par des valeurs physiques. La diminution de $\sigma_0$ sur un manteau neigeux peut s'expliquer par l'augmentation de la teneur en eau et se caractérise par l'augmentation de sa permittivité diélectrique $\varepsilon_r$. Les compromis sont présentés dans [112] où les analyses basées sur des modèles électromagnétiques sont riches en informations physiques mais demeurent des modèles complexes où la capacité à remonter aux paramètres primaires du modèle pour différencier des objets n'est pas toujours garantie. A contrario, les analyses basées sur des modèles statistiques utilisent généralement des modèles plus simples mais où l'information physique a été perdue. Dans le cadre de l'analyse d'images de capteur SAR pour la cryosphère, nous proposons de nous concentrer sur deux approches : celles basées directement sur la rétrodiffusion radar et celles basées sur la polarimétrie. Les caractéristiques de ces deux approches pour l'étude du manteau neigeux sont résumées dans le tableau 2.5.

[125]: Longépé et al. (2007), « Snow wetness monitoring using multi-temporal polarimetric ASAR data. »

[126]: Beltramone et al. (2020), « Wet snow detection in Patagonian Andes with Sentinel-1 SAR temporal series analysis in GEE »

[127]: Veyssière (2019), « Apport des mesures du radar à synthèse d'ouverture de Sentinel-1 pour l'étude des propriétés du manteau neigeux »

[128]: Longépé (2008), « Apport de l'imagerie SAR satellitaire en bandes L et C pour la caractérisation du couvert neigeux »

[129]: Picard et al. (2018), « Smrt »

[112]: Yue et al. (2021), « Synthetic Aperture Radar Image Statistical Modeling »

**Table 2.5 –** Caractéristique des traitements SAR dans le cadre de deux approches, adapté de [36] pour l'étude de la neige

| Type de détection | Rétrodiffusion | PolSAR |
|---|---|---|
| Effet notoire | Baisse du coefficient de rétrodiffusion lors de l'humidification de la neige | Différence diffusion neige sèche/humide et surface |
| Complexité algorithmique | Faible | Élevée |
| Dépendance à l'angle d'incidence local | Élevée | Faible |
| Richesse information dérivée | Moyenne | Élevé |
| Bruit information dérivée | Élevé | Moyenne |
| Détection de type de neige | Humide | Humide/sèche/totale |
| Nombre d'images min. | 1 | ~2 |

### Polarimetric Synthetic Aperture Radar (PolSAR)

La plupart des satellites SAR permettent une acquisition au minimum en 2 polarisations. Dans le cas de Sentinel 1, l'onde électromagnétique est émise par le satellite avec une polarisation verticale et est réceptionnée avec une polarisation soit horizontale soit verticale, ce qui donne 2 images de polarisations VV et VH. Ces différents canaux sont souvent utilisés conjointement pour l'étude de la végétation [130] en décomposant l'information présente dans les différents canaux [131]. Les différentes décompositions de ces canaux permettent d'obtenir des caractéristiques supplémentaires en lien avec la physique du signal [132] pour améliorer la classification[35] . Ces décompositions possèdent une structure mathématique intéressante qui peut être employée à améliorer les performances des classifications. Dans [134], les auteurs exploitent les structures définie symétrique positive (*Symmetric Positive Definite* SPD) sur des variétés riemanniennes, tandis que les auteurs de [135] proposent un réseau de neurones complexe par l'utilisation de quaternions pour cartographier la neige humide avec l'information de l'angle d'incidence. Malgré l'intérêt de cette approche, les données en polarisation complète (4 canaux) sont moins répandues et l'utilisation de 2 canaux peut s'avérer être insuffisante pour certaines décompositions. Cependant récemment les auteurs de [136] ont démontré la validité d'une décomposition des 2 canaux de Sentinel 1 dans la cartographie des zones de neige humide et sèche [137], mais aussi dans la détection des bords de glaciers proposant ainsi l'étude de la dynamique des glaciers par la régularité des images Sentinel. Ces approches tirant parti des différentes réponses des canaux polarimétriques, sont ainsi prisées pour la classification de glaciers [138], en affinant le suivi des glaciers [139], mettant particulièrement en évidence les frontières entre les zones couvertes de neige et les zones rocheuses [140], ou dans le cadre de l'étude de la limite de neige sur glacier [141]. Une étude récente compare le potentiel de l'exploitation direct de coefficient de rétrodiffusion par rapport à deux décompositions polarimétriques, dans le cas de suivi de zones glaciaires [142].

### Rétrodiffusion SAR

La deuxième possibilité d'exploitation de l'imagerie SAR est l'étude du coefficient de rétrodiffusion. Comme ce dernier ne nécessite qu'une image SAR, les algorithmes développés sur cette approche sont souvent moins complexes et les techniques de traitement d'images classiques (filtrage, seuillage, décomposition en ondelettes, machine learning) peuvent facilement s'adapter.

[130]: Bhogapurapu et al. (2021), « Dual-polarimetric descriptors from Sentinel-1 GRD SAR data for crop growth assessment »

[131]: Dey et al. (2021), « A Model-Free Four Component Scattering Power Decomposition for Polarimetric SAR Data »

[132]: Yin et al. (2019), « Optimal combination of polarimetric features for vegetation classification in PolSAR image »

**Précision (35)**
Parmi toutes les décompositions existantes, on peut retenir les décompositions de Pauli, $H/A/\alpha$, Sinclair, Freeman-Durden. Ces décompositions cherchent à mettre en valeur, dans des canaux spécifiques, des informations propres à la dépolarisation du signal ou à des phénomènes de rétrodiffusion spécifiques. Ainsi la décomposition $H/A/\alpha$ [133] permet la classification de neuf zones distinctes dont les processus de rétrodiffusion sont clairement identifiés.

[134]: Formont et al. (2013), « On the Use of Matrix Information Geometry for Polarimetric SAR Image Classification »

[135]: Usami et al. (2016), « PolSAR Wet Snow Mapping With Incidence Angle Information »

[136]: Mascolo et al. (2022), « Model-Based Decomposition of Dual-Pol SAR Data »

[137]: Liu et al. (2023), « Identifying wet and dry snow with dual-polarized c-band SAR data based on Markov random field model »

[138]: Panwar et al. (2022), « Classification of glacier with supervised approaches using PolSAR data »

[139]: Callegari et al. (2016), « A Pol-SAR Analysis for Alpine Glacier Classification and Snowline Altitude Retrieval »

[140]: Huang et al. (2011), « Classification and snow line detection for glacial areas using the polarimetric SAR image »

[141]: Li et al. (2012), « Glacier Snow Line Detection on a Polarimetric SAR Image »

Néanmoins les performances à une échelle plus locale peuvent être dégradées à cause de la dépendance forte de ce coefficient à l'angle d'incidence, si elle n'est pas corrigée ou prise en compte, problème d'autant plus marqué en milieu de montagne comme le montre certaines études [143]. Un aperçu des caractéristiques du signal SAR pour la différenciation de la neige sèche et humide en fonction de la fréquence de fonctionnement du capteur est donné dans le tableau 2.6.

Les données GRD sont le plus souvent utilisées pour analyser le coefficient de rétrodiffusion ($\sigma_0$) [144] afin de détecter la neige humide. L'eau liquide dans la neige humide induit des pertes diélectriques significatives et réduit donc le coefficient de rétrodiffusion à mesure que la TEL augmente. Contrairement à la neige saisonnière sèche, qui permet aux signaux radar de pénétrer sur plusieurs mètres, la neige humide reflète et disperse principalement les signaux radar près de la surface et dans les couches supérieures du manteau neigeux [145] (cf figure 2.20 et 2.21). Cette distinction entraîne un contraste plus notable en intensité de rétrodiffusion entre les régions couvertes de neige humide et celles avec des conditions sèches ou sans neige [146]. Cependant, de nombreux paramètres physiques, tels que la densité de la neige ou l'angle d'incidence, peuvent affecter cette réduction du signal [41]. L'une des méthodes de détection de la neige les plus largement utilisées repose sur l'approche développée par Nagler et al. dans [145] dans le cadre de la bande C. Aucune information statistique n'est prise en compte, un simple seuillage est réaliser entre deux images radar, l'une de référence principalement prise en été $\sigma_{ref}$, sans présence de neige, et l'autre étant celle que l'on cherche à segmenter. La carte de neige humide est générée par la condition :

$$10 \cdot \log\left(\frac{\sigma_t}{\sigma_{ref}}\right) < \lambda,$$

où typiquement $\lambda = -3dB$. La détermination de ce seuil est empirique et dépend d'un grand nombre de paramètres physiques [147]. Cette première approche simple mais efficace a conduit au développement des produits Copernicus de cartographie de neige humide SAR Wet Snow (SWS).

[142]: Barzycka et al. (2023), « Comparison of Three Methods for Distinguishing Glacier Zones Using Satellite SAR Data »

[143]: Nagler et al. (2016), « Advancements for Snowmelt Monitoring by Means of Sentinel-1 SAR »

[144]: Karbou et al. (2021), « Monitoring Wet Snow Over an Alpine Region Using Sentinel-1 Observations »

[145]: Nagler et al. (2000), « Retrieval of wet snow by means of multitemporal SAR data »

[146]: Mann (2018), « Spatial and temporal variability of the snow environment in the Western Canadian Arctic »

[41]: Marin et al. (2020), « Use of Sentinel-1 radar observations to evaluate snowmelt dynamics in alpine regions »

[147]: Magagi et al. (2003), « Optimal conditions for wet snow detection using RADARSAT SAR data »

**Table 2.6 –** Caractéristique du signal SAR dans le cadre d'une différenciation neige sèche neige humide adapté de [36]

| **Type de neige** | **Neige sèche** | **Neige humide** |
|---|---|---|
| Source de rétrodiffusion | Diffusion volumique à partir du manteau neigeux, diffusion de surface à l'interface neige/sol. | Diffusion de surface à l'interface air/neige |
| Facteurs dominants de diffusion | Surface sous la neige (< 10 GHz), taille du grain (> 10 GHz) | Teneur en eau liquide (la plus importante), rugosité de surface |
| Coefficient de rétrodiffusion | Elevé | Faible |

De nombreuses variantes ont été proposées pour pallier les principales difficultés rencontrées par cette approche. Ainsi dans [143], Nagler propose une amélioration de ses travaux en intégrant un coefficient qui équilibre le précèdent ratio avec deux canaux polarimétriques (VV et VH) en fonction de l'angle de visée. Le traitement réalisé ne fait pas appel à une calibration, les variations sont dues à une augmentation relativement forte du signal de rétrodiffusion des surfaces de neige humides dans le cadre de la co-polarisation. Dans ce travail, le seuillage est réalisé à partir d'une moyenne pondérée des deux canaux et cette fois-ci le seuil empirique de détection est donnée par $\lambda = -2dB$.

D'après [36] 82% des études qui cartographient la neige humide utilisent des approches basées sur le seuillage proposé par Nagler dans [145], tel que proposé dans [9]. D'autres travaux ont su tirer parti de cette méthode pour remonter à la hauteur de neige [148] à une résolution de l'ordre de la centaine de mètres en réalisant une détection de changements à partir des coefficients de rétrodiffusion, la valider sur l'ensemble du massif alpin [149] ou étudier la dynamique de fonte [41]. L'atout principal de cette méthode est sa simplicité algorithmique, qui dépend de l'hypothèse de l'homogénéité du seuil quelle que soit la topographie étudiée (dans le cas d'un seuillage fixe). Ceci induit de bons résultats à l'échelle globale (massif montagneux), mais donne des résultats plus nuancés nécessitant de prendre en compte la topographie dans certaines zones avec un relief accidenté [8] ou l'atténuation observée empiriquement de l'onde électromagnétique en présence de forêts [150].

Une autre limitation provient du choix d'une image de référence. La recherche de cette image est longue et coûteuse, et certaines expériences de moyennages d'images se sont avérées infructueuses. Cette difficulté peut venir du caractère inadapté de la moyenne euclidienne comme outil de calcul de barycentre[36] .

Il est intéressant de noter que dans le cadre des produits Copernicus, la validation de la méthode a été faite sur 3 sites expérimentaux à l'échelle européenne. Sur cet aspect de validation des méthodologies proposées, [36] indique que près de 60% des études proposent une validation inexistante. La disponibilité et la densité de point de mesures spécifiques est problématique et pose la question de la pertinence de méthode non ou insuffisamment validée.

Dans ce cadre de la détection de neige humide, le travail de [8] propose une étude statistique en analysant les conditions sur le seuillage d'une image SAR par l'utilisation d'un modèle de transfert radiatif pour déterminer le seuil optimal. Cette approche, les travaux de [153, 154], ainsi que l'étude de [155], qui compare les sorties du modèle CROCUS à la rétrodiffusion radar de Sentinel 1, donnent un support pour l'inter-connectivité entre les modalités (rétrodiffusion et modélisation physique) et leur intégration dans un processus soit d'accès à une vérité terrain soit d'optimisation physique d'algorithme. Certains ont exploité les récentes méthodes d'apprentissage automatique [90] mais leur utilisation dans la communauté reste encore limitée. Les principales limitations de ces méthodes étant l'aspect opérationnel, et dans le cadre d'approches supervisées, la labellisation. En effet ces méthodes nécessitent une phase d'entraînement et d'inférence qui peuvent s'avérer très gourmandes en temps de calcul pour le traitement de larges zones. De plus, pour l'entraînement, une labellisation fiable, avec une densité spatio-temporelle suffisante (couvrant un massif, voir une région) est rarement disponible, rendant l'étiquetage partiel de données [52]. Ceci explique que la majorité des travaux s'attachent soit à produire des cartes de neige humide à large échelle basée sur des approches ne nécessitant pas de labellisation (méthode non-supervisée comme le seuil),

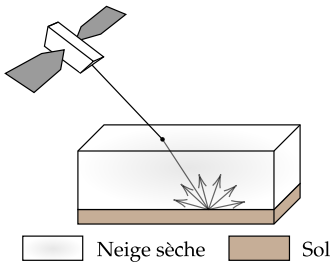



**Figure 2.20 –** Illustration de la rétrodiffusion de la neige sèche en bande C. Adaptée de [41]

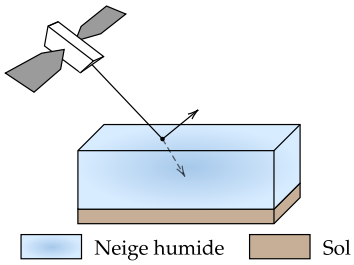



**Figure 2.21 –** Illustration de la rétrodiffusion de la neige humide en bande C. Adaptée de [41]

[36]: Tsai et al. (2019), « Remote Sensing of Snow Cover Using Spaceborne SAR »

[9]: Guiot et al. (2023), « Insights into Segmentation Methods Applied to Remote Sensing SAR Images for Wet Snow Detection »

[148]: Lievens et al. (2021), *Sentinel-1 snow depth retrieval at sub-kilometer resolution over theEuropean Alps*

[149]: Karbou et al. (2021), « Monitoring Wet Snow Over an Alpine Region Using Sentinel-1 Observations »

[41]: Marin et al. (2020), « Use of Sentinel-1 radar observations to evaluate snowmelt dynamics in alpine regions »

[8]: Besic et al. (2015), « Stochastic Approach in Wet Snow Detection Using Multitemporal SAR Data »

[150]: Zhu et al. (2019), « Snow Covered Area Estimation for Forest Area in Northeastern China Based on Multitemporal Sentinel-1 SAR Data »

**Précision (36)**

Dans le cas de la moyenne euclidienne des images de neige humide et de neige sèche peuvent être moyennées pour donner une image de neige moyenne, qui n'existe pas dans la réalité. Une piste intéressante serait de déterminer une image moyenne grâce à de multiples dates (ou multiple source [151]), comme proposée dans [152] pour une fusion d'images hyperspectrales et multispectrales.

soit l'étude plus locale de paramètres physiques du manteau neigeux comme la TEL ou le SWE.

L'utilisation de données SAR en bande C pour remonter à une caractéristique du manteau neigeux n'est pas nouvelle. En 1993, les travaux de [156] ont permis de mettre en pratique l'obtention de TEL à partir de données d'AIRSAR en polarisation complète. Cette mesure a été assez peu utilisée, notamment à la large échelle, les caractérisations spatiales les plus courantes étant la densité de la neige [157] et le SWE [158, 159]. L'estimation du SWE, et la cartographie associée nécessitent une validation par l'utilisation de mesures [160]. La méthodologie associée est basée sur une succession de régression linéaire, la première liant le coefficient de rétrodiffusion d'un ratio d'images pour s'affranchir de la topographie avec la résistance thermique de la neige. La deuxième régression permet de remonter depuis cette résistance thermique au SWE. Nous pouvons remarquer que la plupart des études utilisent la rétrodiffusion en bande C (pour sa capacité de pénétration), mais d'autres travaux [161] ont montré la capacité de l'utilisation conjointe de la bande C et X. De plus, il existe peu de travaux sur l'exploitation et la classification automatique de ces indicateurs. Néanmoins, les travaux et les données de [49] permettent une ouverture vers un aspect apprentissage automatique par les vérités terrain de SWE compilées couvrant de manière hétérogène l'hémisphère nord.

La capacité des satellites SAR pour l'observation et la classification des glaciers est significative. Grunblatt et al. ont utilisé 36 acquisitions de ERS-2 en avril 2009 pour cartographier la présence d'eau liquide sous la glace, en s'appuyant sur un réseau de près de 105 données bathymétriques [162]. La détection des frontières neige/glace est cruciale pour observer la dynamique de la fonte des neiges saisonnières [163]. Elle offre des informations essentielles sur le bilan massique du glacier en surveillant les changements dans les zones d'accumulation et d'ablation [31], lesquelles sont les principaux composants de ce bilan massique. Hartmann et al. ont utilisé un réseau de neurones convolutif pour segmenter les images SAR et distinguer ces zones [164]. L'estimation du bilan massique a été réalisée en utilisant des images SAR en bande X dans l'Himalaya central [165]. Taylor et al. ont montré que la bande X est particulièrement bien adaptée au suivi des informations cryosphériques, ce qui permet, entre autres, l'étude de l'élévation des glaciers et de la dynamique massique [16]. Cependant, la détection des frontières entre différents types de zones glaciaires en utilisant des images SAR demeure un défi important en raison du bruit de speckle [166].

**Analyse temporelle**

Tandis qu'un certain nombre de travaux s'attachent à caractériser le manteau neigeux ou classifier les glaciers dans les images SAR, la richesse de ces satellites permet d'exploiter des séries temporelles denses. L'utilisation de ces données repose sur la capacité de détecter des changements (apparition/disparition de neige, transformation de la neige en eau liquide) [167]. Le comportement et l'évolution temporelle de certains glaciers s'avère être discriminant au vu d'autres glaciers, et ce en s'intéressant à deux bandes de fréquences (X et C) [29]. Cette analyse permet de suivre le comportement de rétrodiffusion au cours du temps et de caractériser d'éventuelles changement de classe, telle que une zone d'accumulation se transformant graduellement en zone d'ablation, menant à une surveillance de l'emprise des glaciers au sol [168].

[8]: Besic et al. (2015), « Stochastic Approach in Wet Snow Detection Using Multitemporal SAR Data »

[153]: Longepe et al. (2009), « Snowpack Characterization in Mountainous Regions Using C-Band SAR Data and a Meteorological Model »

[154]: Lund et al. (2022), « Interpreting Sentinel-1 SAR Backscatter Signals of Snowpack Surface Melt/Freeze, Warming, and Ripening, through Field Measurements and Physically-Based SnowModel »

[155]: Veyssière et al. (2018), « Evaluation of Sub-Kilometric Numerical Simulations of C-Band Radar Backscatter over the French Alps against Sentinel-1 Observations »

[90]: Tsai et al. (2019), « Wet and Dry Snow Detection Using Sentinel-1 SAR Data for Mountainous Areas with a Machine Learning Technique »

[52]: Odry et al. (2020), « Using artificial neural networks to estimate snow water equivalent from snow depth »

[156]: Shi et al. (1993), « Deriving snow liquid water content using C-band polarimetric SAR »

[157]: Varade et al. (2020), « Modelling of early winter snow density using fully polarimetric C-band SAR data in the Indian Himalayas »

[158]: Tsang et al. (2022), « Review article : Global monitoring of snow water equivalent using high-frequency radar remote sensing »

[159]: Bernier et al. (1999), « Determination of snow water equivalent using RADARSAT SAR data in eastern Canada »

[160]: Shaobo et al. (2015), « Estimation and Analysis of Snow Water Equivalents Based on C-Band SAR Data and Field Measurements »

[161]: Shi et al. (2000), « Estimation of snow water equivalence using SIR-C/X-SAR. II. Inferring snow depth and particle size »

[49]: Fontrodona-Bach et al. (2023), « NH-SWE : Northern Hemisphere Snow Water Equivalent dataset based on in situ snow depth time series »

[162]: Grunblatt et al. (2014), « Mapping lakes for winter liquid water availability using SAR on the North Slope of Alaska »

[163]: Braun et al. (2000), « Using A Multiyear Data Archive Of Ers Sar Imagery For The Monitoring Of Firn Line Positions And Ablation Patterns On The King George Island Ice Cap (Antarctica) »

[31]: Engeset et al. (2002), « Change Detection and Monitoring of Glacier Mass Balance and Facies Using ERS SAR Winter Images over Svalbard »

[164]: Hartmann et al. (2021), « Bayesian U-Net for Segmenting Glaciers in Sar Imagery »

## 2.3 Caractéristiques temporelles de la rétrodiffusion : comparaison entre les bandes X et C pour la cryosphère

Suite à la présentation des caractéristiques radiométriques des images SAR et de leur interaction avec la cryosphère, nous présentons dans cette section une analyse du comportement de la rétrodiffusion en bande X et C sur une sélection de zones d'intérêt (*Regions of Interest* ROI) de nature différente sur le massif du Mont-Blanc. Cette analyse est basée sur le jeu de données CPAZMAL (dont la description et le détail de la construction sont donnés section 3.3). Nous avons complété ce dataset par des données SAR Sentinel-1 en bande C. Ces données ont été pré-traitées de la même manière que les données en bande X, avec en plus une correction des orbites.[37]

[165]: Bandyopadhyay et al. (2018), « Mass Balance Estimation Using Sar Data in Central Himalaya »

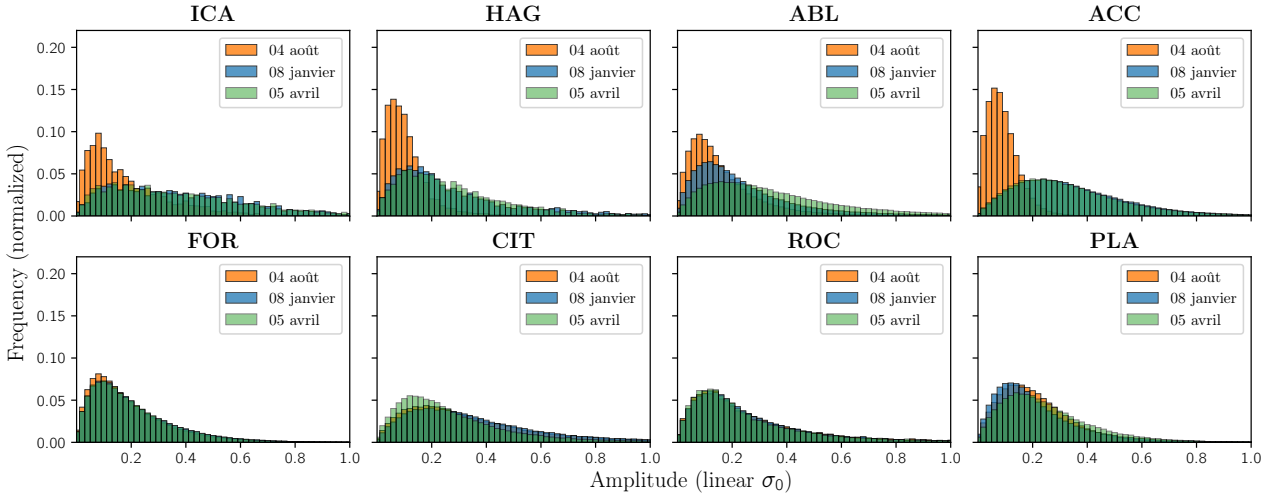
[16]: Taylor et al. (2021), « Remote sensing of the mountain cryosphere »

[166]: Kundu et al. (2015), « Delineation of glacial zones of Gangotri and other glaciers of Central Himalaya using RISAT-1 C-band dual-pol SAR »

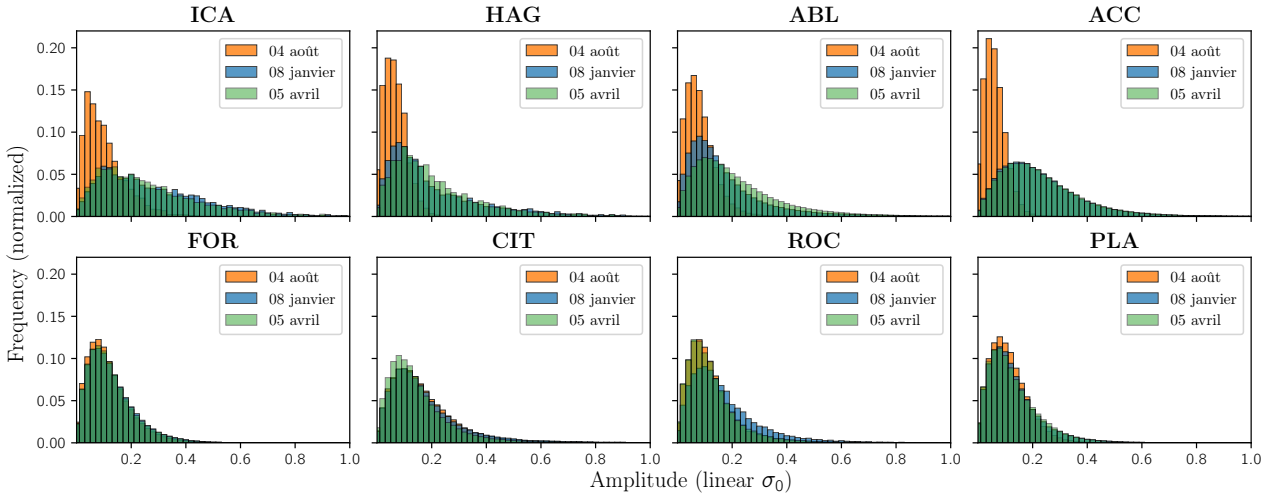
[167]: Beltramone et al. (2023), « Identification of Seasonal Snow Phase Changes From C-Band SAR Time Series With Dynamic Thresholds »

[29]: Kaushik et al. (2022), « Analysis of the Temporal Evolution of Ice Aprons in the Mont-Blanc Massif Using X and C-Band SAR Images »

[168]: Arigony-Neto et al. (2007), « A time series of SAR data for monitoring changes in boundaries of glacier zones



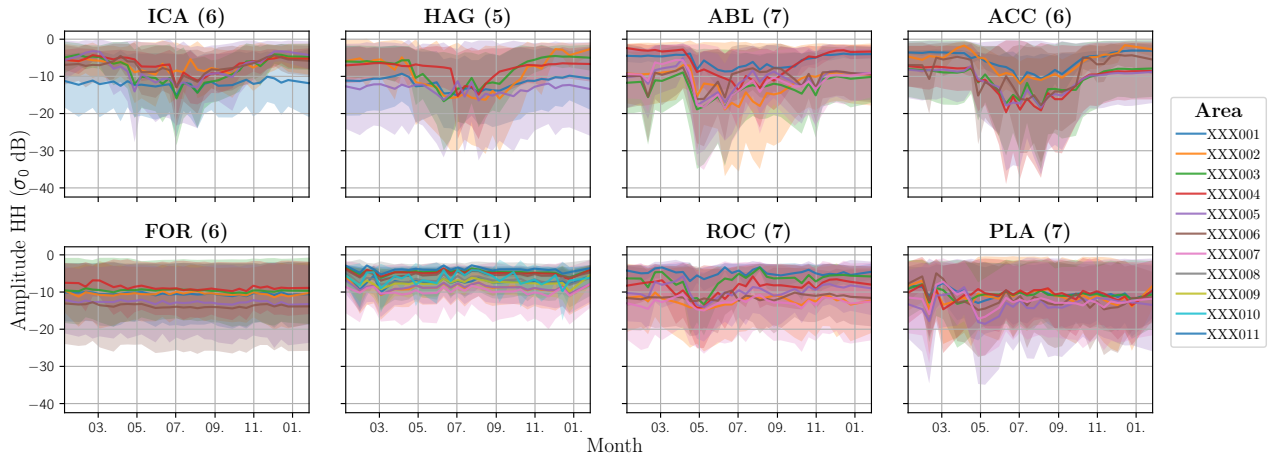

**(a)** Polarisation HH.



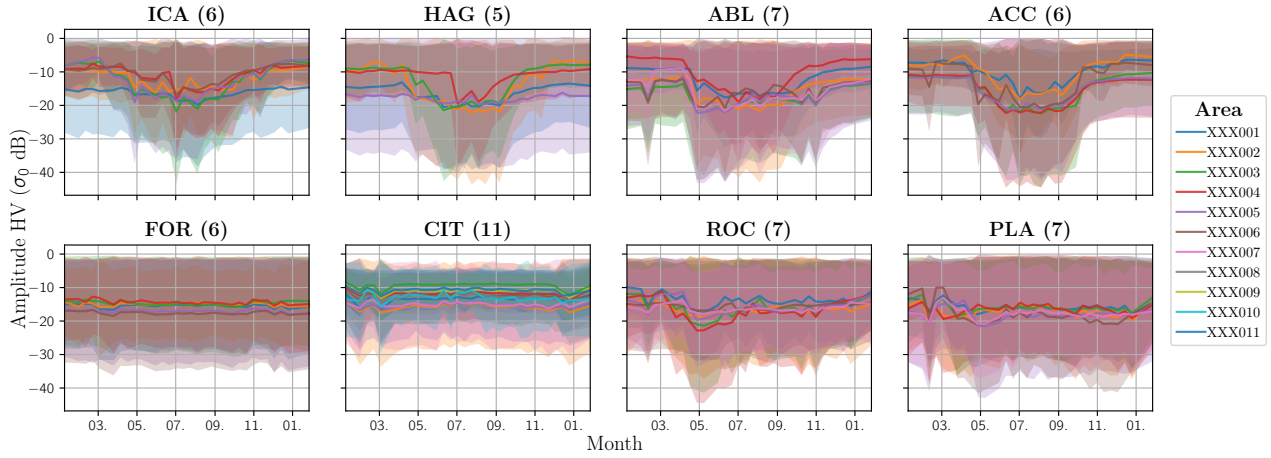

**(b)** Polarisation HV.

**Figure 2.22 –** Comparaison des distributions statistiques des 8 classes en bande X, pour 3 dates en 2020 (4 août, 8 janvier et 5 avril) pour les deux polarisations HH et HV.

### 2.3.1 Comportement temporel de la rétrodiffusion en bande X




**(a)** Polarisation HH avec moyenne et écart-type (log).




**(b)** Polarisation HV avec moyenne et écart-type (log).

**Figure 2.23 –** Comparaison temporelle et polarimétrique des 8 différentes classes en bande X sur l'année 2020 (valeurs moyennes et écarts-types) en fonction de la polarisation et du ROI.

Les figures 2.22a et 2.22b montrent la distribution en amplitude du coefficient de rétrodiffusion pour les 8 classe retenues en polarisation simple (HH) et croisée (HV), pour trois périodes de l'année : le 4 août (plein été), le 8 janvier (hiver) et le 5 avril (printemps, début de la fonte des neiges). Comme on peut le voir à partir des histogrammes pour les deux polarisations, les classes de tablier glaciaire, de glacier suspendu, de zone d'ablation du glacier et d'accumulation montrent une distribution relativement uniforme des valeurs d'histogramme en hiver et au début de l'été. Cependant, nous observons une augmentation de la fréquence des valeurs de faible coefficient de rétrodiffusion pour toutes les classes mentionnées ci-dessus pendant les mois d'été. Un autre aspect intéressant à noter ici concerne la classe d'ablation. Comme pour toutes les classes de glacier, une réduction des valeurs de rétrodiffusion est observée pendant la période estivale, mais la distribution de fréquence est plus basse. Les quatre classes restantes, c'est-à-dire la forêt, la ville (urbaine), les zones rocheuses et les plaines, montrent une distribution de fréquence constante des valeurs de rétrodiffusion tout au long de l'année.

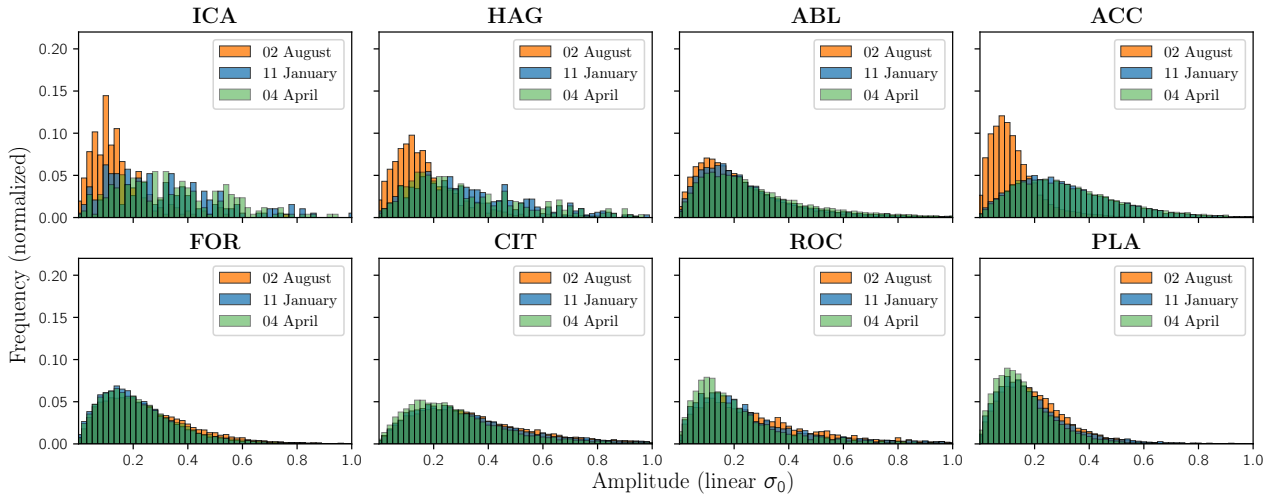
Une comparaison des résultats entre les deux polarisations montre que les ensembles de données en polarisation croisée (HV) sont plus sensibles aux changements physiques à la surface des corps de glacier par rapport à la polarisation simple (HH). Cela est particulièrement bien illustré pendant les mois d'été, où la distribution de fréquence montre un pic plus élevé de valeurs de rétrodiffusion plus faibles. À partir des observations, on peut suggérer que les ensembles de données en polarisation croisée sont plus adaptés à l'analyse des changements physiques de surface des corps de neige/glacier par rapport à la polarisation simple, un résultat également corroboré par des recherches antérieures [169].

[169]: Kaushik et al. (2023), « Temporal Evolution of X and C Band SAR Backscattering in theMont-Blanc Massif »

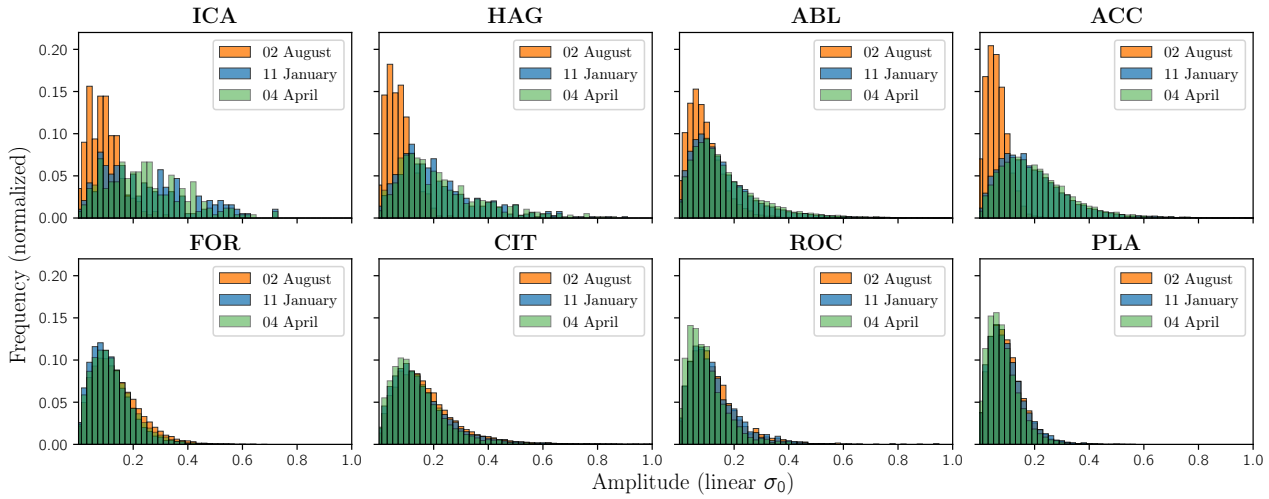
Les figures 2.23a et 2.23b montrent l'évolution annuelle des valeurs moyennes de rétrodiffusion de toutes les classes et de leurs écarts-types pour toutes les ROI. Les résultats sont basés sur la comparaison de plusieurs zones pour chaque classe (le nombre en haut de la classe dans la figure indique combien de zones pour cette classe particulière ont été prises en compte). La tendance annuelle confirme davantage les résultats de nos observations précédentes. Pour toutes les classes de glacier, une baisse du signal de rétrodiffusion est observée pendant les mois d'été. Les valeurs de rétrodiffusion restent presque constantes tout au long de l'année pour toutes les autres classes de terrains. Une tendance similaire est observée pour les deux polarisations. L'écart-type montre une large dispersion des valeurs autour de la moyenne pour toutes les classes de glacier et la ville (urbaine), tandis que pour les autres classes de terrain, l'écart-type est plus faible.

### 2.3.2 Comportement temporel de la rétrodiffusion en bande C

De manière similaire à la bande X, les figures 2.24a et 2.24b montrent la distribution en amplitude du coefficient de rétrodiffusion en bande C pour toutes les différentes classes en polarisation simple (VV) et croisée (VH), pour trois périodes de l'année : le 2 août (plein été), le 11 janvier (hiver) et le 4 avril (printemps, début de la fonte des neiges). Nous observons des histogrammes plus irréguliers pour les zones ICA et HAG, qui s'expliquent par une résolution spatiale plus faible en bande C et donc peu de pixels par zone. Sur les 8 classes observées, seules les classes de glacier (ICA, HAG, ABL et ACC) montrent une variation significative de la distribution entre l'été et l'hiver en polarisation croisée. Ceci est dû à la diminution de la rétrodiffusion causée par l'humidification de la surface. Là où trois surfaces glaciaires (ICA, HAG et ACC) montrent cette même diminution en polarisation simple, la classe d'ablation (ABL) ne montre pas de variation significative. Nous pouvons remarquer que les autres classes de terrain (forêt, ville, zone rocheuse et plaine) affichent une faible fluctuation de la distribution de rétrodiffusion en fonction de la période de l'année, similaire à ce que nous avons observé en bande X. Nous observons une tendance similaire pour le logarithme des valeurs moyennes de rétrodiffusion en bande C, comme illustré dans les figures 2.25a et 2.25b. Il est intéressant de noter l'avantage du satellite Sentinel-1, qui permet une acquisition plus fréquente des données, ce qui permet une meilleure caractérisation des changements de rétrodiffusion au cours de l'année. Le comportement temporel de la rétrodiffusion en bande C est similaire à celui en bande X, avec un signal relativement stable pour les classes de terrain autres que les glaciers, à l'exception de la classe des rochers, où l'on peut voir un comportement proche de celui observé pour les surfaces glaciaires mais avec un décalage temporel, la diminution du signal intervient entre mars




**(a)** Polarisation VV.



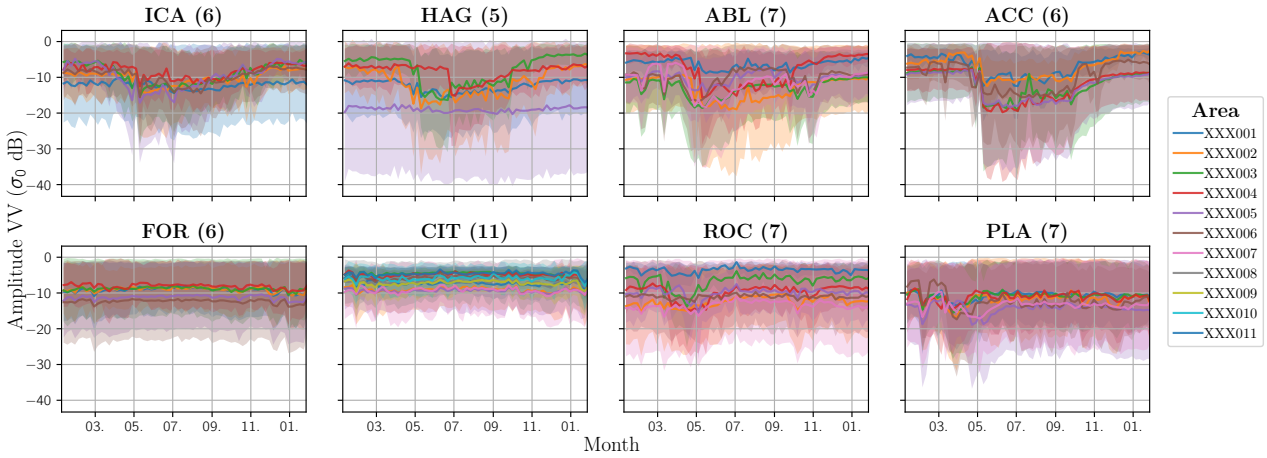

**(b)** Polarisation VH.

**Figure 2.24 –** Comparaison des distributions statistiques des 8 classes en bande C, pour 3 dates en 2020 (2 août, 11 janvier et 4 avril) pour les deux polarisations VV et VH.

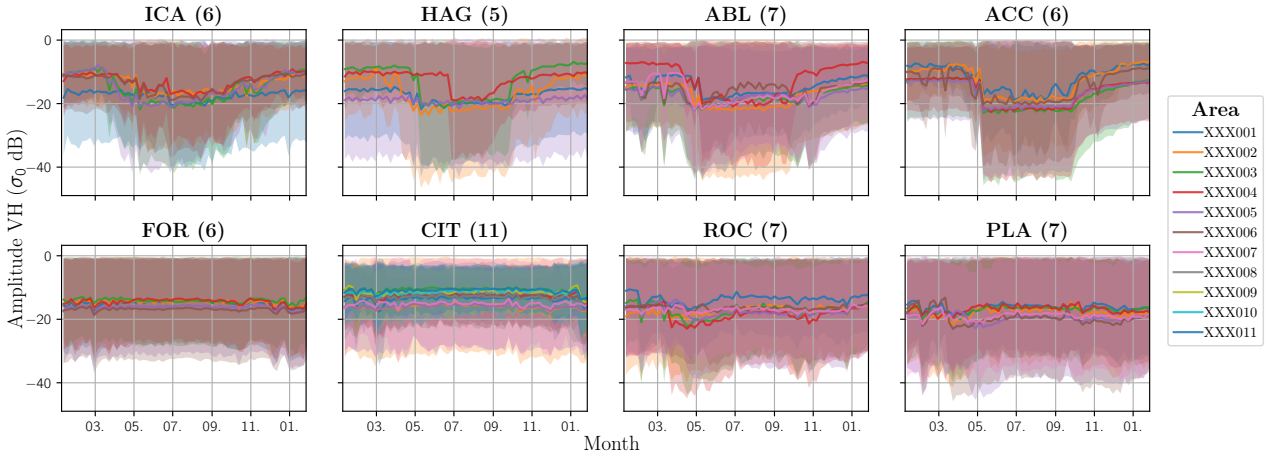
et mai avec une diminution plus faible que pour les glaciers. Ces derniers montrent un comportement similaire entre eux, mais avec des spécificités propres à chaque surface. Toutes les zones d'accumulation (ACC) voient leur rétrodiffusion diminuer en même temps (vers le mois de mai), pour les deux polarisations, là où nous pouvons remarquer plus de variabilité entre les zones d'ablation. Enfin la diminution du signal est plus faible pour les zones de glaciers suspendus (HAG) et de tabliers glaciaires (ICA) avec des différences de démarrage de la diminution de la rétrodiffusion entre les zones.

### 2.3.3 Analyse comparative pour les glaciers

À partir du jeu de données CPAZMAL (cf section 3.3), nous avons extrait les données SAR en bande X pour les 4 classes de glaciers (ICA, HAG, ABL et ACC) sur les régions d'étude 001 et 002 (cf tableau en annexe A.3), et extrait les mêmes données en bande C pour les mêmes régions.




**(a)** Polarisation VV avec moyenne et écart-type (log).




**(b)** Polarisation VH avec moyenne et écart-type (log).

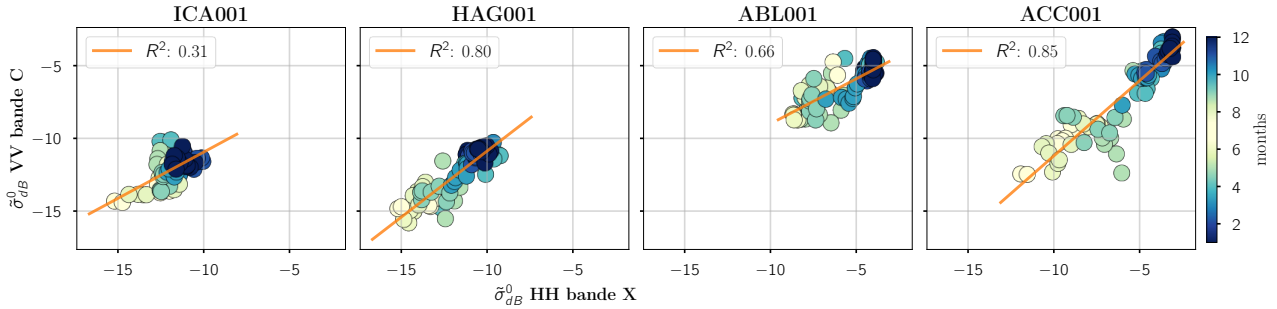
**Figure 2.25 –** Comparaison temporelle et polarimétrique des 8 différentes classes en bande C sur l'année 2020 (valeurs moyennes et écarts-types) en fonction de la polarisation et du ROI.

Nous proposons une sélection de résultats complémentaires en annexe A.2.1 pour les sites d'études 003 et 004. Nous avons sélectionné comme statistique de comparaison, le coefficient de rétrodiffusion moyen de l'image en intensité, en décibel.

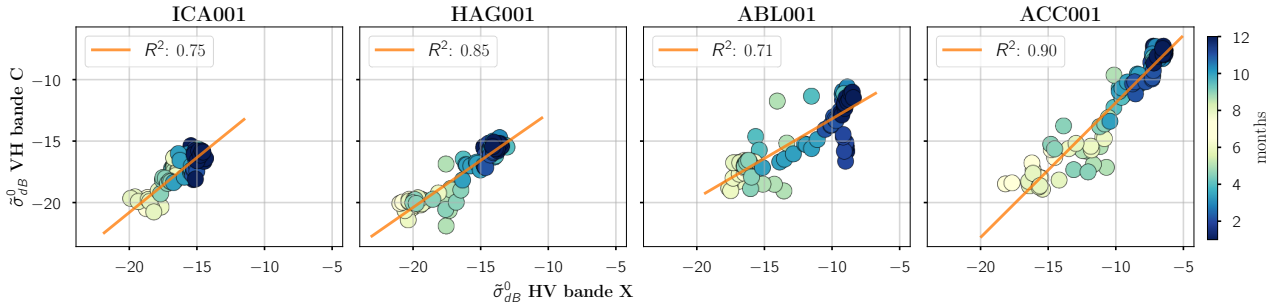
**Méthodologie :** Les images en bande X ont une meilleure résolution spatiale[38] que celle en bande C, cependant ces dernières ont une meilleure résolution temporelle.[39] Comme les emprises des ROI sont les mêmes pour les deux bandes, nous avons calculé le coefficient de rétrodiffusion moyen (moyenne des pixels) pour chaque classe et pour chaque acquisition et ainsi obtenue une série temporelle pour chaque classe et bande. Nous avons en bande C, 64 dates d'acquisition et en bande X, 30 dates d'acquisition. Les dates d'acquisition entre les deux bandes n'étant pas synchrones, nous avons réalisé une interpolation polynomial par parties, afin d'obtenir deux séries temporelles synchrones. Nous proposons deux études, la première sur la comparaison des données en co-polarisation (HH / VV) et la seconde sur la comparaison des données en polarisation croisée (HV / VH). Nous avons effectué pour chaque

**Précision (38)**
Dans notre cas, sans multi-vues, les pixels de PAZ en bande X sont de résolution ~2m×2.5m contre ~2m×10m pour Sentinel-1 en bande C.

**Précision (39)**
Pour l'année considérée, les acquisitions Sentinel-1 sont disponibles toutes les 6 jours contre une acquisition tous les 11 jours au maximum pour PAZ




**(a)** Co-polarisation (HH / VV)



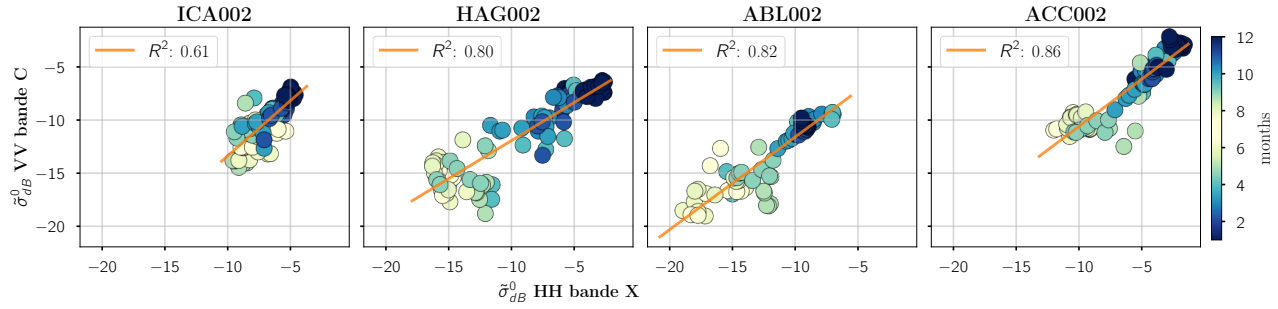

**(b)** Polarisation croisée (HV / VH)

**Figure 2.26 –** Comparaison des données SAR en bande C et en bande X pour les 4 classes de glaciers sur les sites d'étude "XXX001"

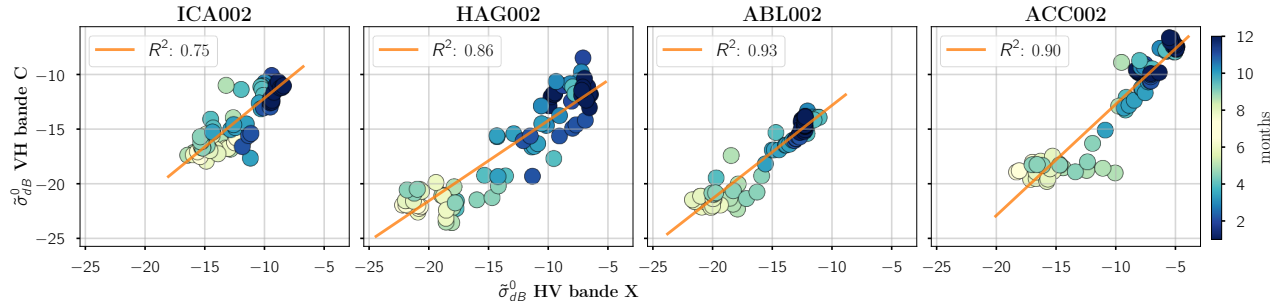
**Définition (40)**
*Coefficient de détermination:* Mesure de la qualité d'une prédiction par un modèle de régression linéaire. Il est définie pour $n$ observations par : $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$ avec $y_i$ les valeurs observées, $\hat{y}_i$ les valeurs prédites et $\bar{y}$ la moyenne des valeurs observées.

classe une régression linéaire entre les deux séries temporelles, et calculé le coefficient de détermination[40] $R^2$.

**Résultats :** Les résultats sont présentés dans les figures 2.26a et 2.27a respectivement pour les sites d'étude 001 et 002 en co-polarisation (VV *vs* HH) et dans les figures 2.26b et 2.27b en polarisation croisée (HV *vs* VH). La barre de couleur associée donne une information sur la date d'acquisition. Il est intéressant de noter que dans les 8 zones étudiées, la corrélation linéaire entre la bande X et C est plus forte en polarisation croisée qu'en co-polarisation, avec des différences comprises entre 5% et ce jusqu'à plus de 40% dans le cas du tablier glaciaire $ICA001$. De manière générale, les zones glaciaires de type accumulation et ablation montrent une bonne corrélation entre les deux bandes et pour les deux polarisations, avec des coefficients de détermination supérieurs à 0.8. Le comportement radiométrique au cours de la saison, correspond à une diminution de la rétrodiffusion en bande X et en bande C lors de la fonte des neiges. Nous pouvons observer que les zones de glaciers de type accumulation (ACC) présentent la rétrodiffusion la plus forte à un moment dans l'année différent des autres classes. Là où les classes HAG, ICA, et ABL, ont des rétrodiffusions maximales dans les deux bandes entre décembre et février, la classe ACC a une rétrodiffusion maximale plus tard (~avril). Ceci peut néanmoins s'expliquer par le fait que ces zones sont situées à des altitudes plus élevées et que la fonte des neiges y est plus tardive. Le comportement des zones de glaciers suspendus (HAG) et de type tablier glaciaire (ICA) est plus variable. Certaines zones comme $HAG001$ donnent des coefficients de détermination supérieurs à 0.8, tandis que d'autres ont des dynamiques de rétrodiffusion beaucoup plus faibles et ont un comportement plus circulaire entre bande X et C. Il est toutefois possible de noter une caractéristique particulière pour la classe des glaciers suspendus, où l'on observe deux régimes en polarisation croisée en




**(a)** Co-polarisation (HH / VV)




**(b)** Polarisation croisée (HV / VH)

**Figure 2.27 –** Comparaison des données SAR en bande C et en bande X pour les 4 classes de glaciers sur les sites d'étude "XXX002"

bande C (comme illustré dans *HAG*003). L'un correspondant aux dates plutôt hivernales avec des rétrodiffusions élevées et un autre pour les dates estivales avec des rétrodiffusions plus faibles, mais avec un écart important entre ces deux régimes. Cette caractéristique est moins marquée en bande X, où l'on observe une diminution de la rétrodiffusion au cours de la saison, mais sans distinction entre les dates hivernales et estivales.

Finalement, nous pouvons conclure que les données en bande X et en bande C ont des comportements radiométriques similaires sur les 4 types de glaciers étudiés. Un comportement qui est principalement linéaire, étayé par la valeur du coefficient de détermination $R^2$ qui est supérieur à 0.8 pour la majorité des zones d'étude. Cette comparaison est restreinte puisque uniquement menée sur les valeurs des rétrodiffusion moyenne des zones. L'avantage de la bande X est sa meilleure résolution spatiale, qui permet une meilleure caractérisation des zones de glaciers, notamment pour les zones de glaciers suspendus et les tabliers glaciaires qui sont des surfaces relativement faibles. Cependant, la bande C présente une meilleure résolution temporelle. De plus, les données en bande X proviennent du satellite PAZ[41] et sont le résultat d'un appel à projet sur la zone du Mont-Blanc. Dans le cadre d'une utilisation opérationnelle, les données Sentinel-1 en bande C sont plus adaptées, par leur accessibilité.

**Précision (41)**
Nous tenons à remercier l'Instituto Nacional de Técnica Aeroespacial (INTA) espagnol pour les images PAZ (Projet AO-001-051).

### 2.3.4 Analyse comparative pour la neige

Cette étude a été réalisée sur la zone d'étude *PLA*004 (cf tableau en annexe A.2), afin d'observer les corrélations entre des statistiques de rétrodiffusion en bande X et en bande C et des mesures ou des informations simulées du manteau neigeux. Le choix de cette zone particulière est motivé par une mesure manuelle et quotidienne de la hauteur de neige, obtenue grâce à

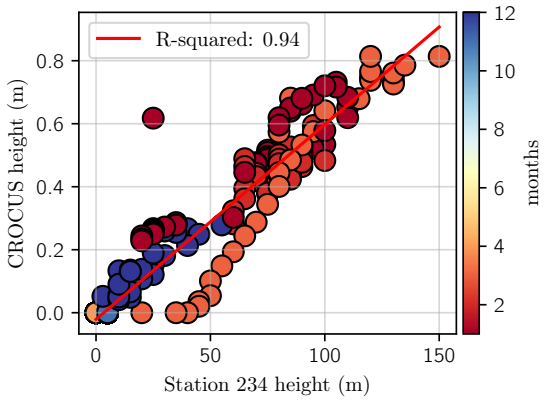



**Figure 2.28 –** Étude de la corrélation entre la hauteur de neige simulée par le modèle CROCUS et celle mesurée à l'emplacement 'PLA004'.

la collaboration avec le Centre de Recherche sur les Écosystèmes d'Altitude (CREA) basé à Chamonix.

**Méthodologie** Pour comparer les données, nous avons dans un premier temps procédé à un sur-échantillonnage spatial des données Sentinel-1, afin d'obtenir des images de mêmes dimensions entre les bandes X et C. Nous avons ensuite extrait des fenêtres de $32 \times 32$ pixels pour chaque date d'acquisition (compte tenu de la résolution spatiale et de l'emprise géographique des ROI, nous avons pu extraire trois fenêtres par date), sur lesquelles nous avons extrait les statistiques décrites dans le tableau 2.7. Nous avons moyenné ces statistiques sur l'ensemble des fenêtres pour chaque date d'acquisition. En plus de calculer ces statistiques sur les co-polarisations et les polarisations croisées (indiquées respectivement par les indices $\parallel$ et $\perp$), nous les avons également calculées sur les ratios des images par rapport à une image de référence (une par modalité bande X et bande C) en été sans neige (indiquées par les indices $^{\text{ref}}$). Les deux dates de référence, empiriquement sélectionnées, sont le 14 et le 15 août 2020 pour les bandes C et X respectivement. Ces ratios étant les primitives pour la caractéristique de Nagler en la bande C [145], nous les avons également calculés en bande X pour étudier les potentiels liens entre ce ratio et les caractéristiques du manteau neigeux.

[145]: Nagler et al. (2000), « Retrieval of wet snow by means of multitemporal SAR data »

Nous avons utilisé deux sources d'informations sur le manteau neigeux : une station de mesure située dans le polygone d'étude $PLA004$ fournissant la hauteur de neige (m) et quatre mesures de température (°C) à différentes hauteurs (-30cm, 0cm, 30cm et 200cm). La deuxième source de données est le modèle de neige CROCUS (cf 15), qui fournit la TEL et la température minimale du manteau neigeux. Nous avons également extrait la hauteur de neige simulée et présentons dans la figure 2.28, la régression linéaire entre la hauteur de neige simulée et la hauteur de neige mesurée ainsi que le coefficient de détermination $R^2$. Il faut noter que la simulation est faite pour le massif complet du Mont-Blanc, pour le pas d'altitude de $300m$ entre 1200m et 1500m, la pente moyenne de 20° et une orientation moyenne de 225°, soit les caractéristiques de la zone d'étude $PLA004$. Nous pouvons observer que la hauteur de neige simulée par CROCUS est en accord avec la hauteur de neige mesurée, avec un coefficient de détermination $R^2$ de 0.94. Cependant nous pouvons remarquer un décalage entre les deux mesures de près de 40cm, qui peut être expliqué par le caractère local de la mesure versus global à l'échelle du massif. Pour la suite nous utilisons la mesure de hauteur de neige. Nous avons ajouté l'indice $\Delta_{ACQ}$ qui correspond à la différence de hauteurs de neige

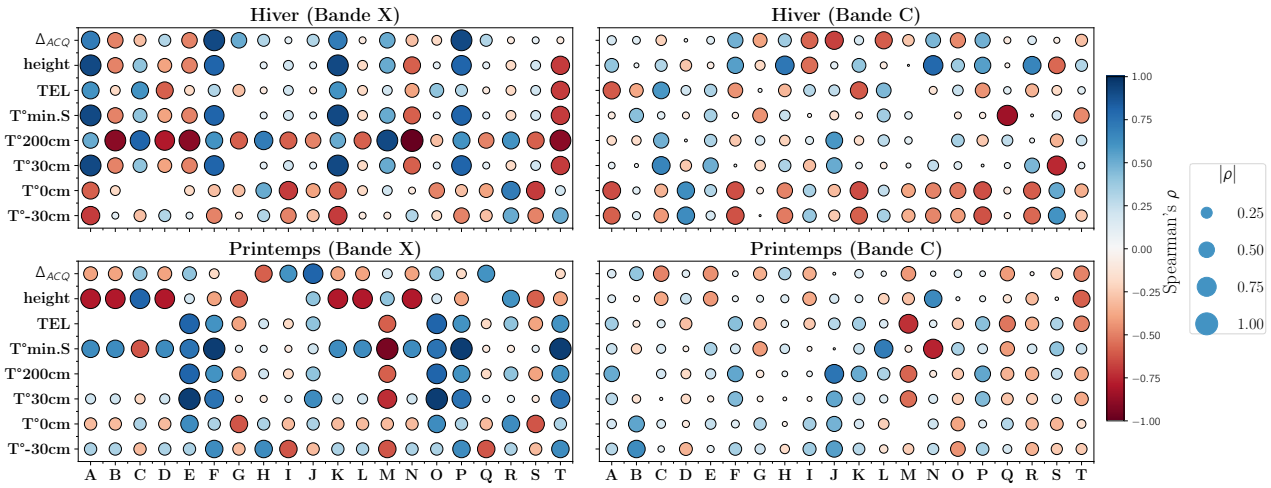



**Figure 2.29 –** Comparaison des corrélations de type Spearman entre les statistiques des images SAR (cf description dans le tableau 2.7) et les caractéristiques du manteau neigeux pour la période hivernale (première ligne) et pour la période printanière (seconde ligne), pour les deux bandes X (première colonne) et C (seconde colonne).

entre deux dates d'acquisition consécutives, pour observer l'évolution de la hauteur de neige entre deux dates.

| **Indice** | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **J** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Stats.** | $\mu_{\parallel}$ | $\sigma_{\parallel}$ | $\gamma 1_{\parallel}$ | $\gamma 2_{\parallel}$ | $CV_{\parallel}$ | $\mu_{\perp}$ | $\sigma_{\perp}$ | $\gamma 1_{\perp}$ | $\gamma 2_{\perp}$ | $CV_{\perp}$ |
| **Indice** | **K** | **L** | **M** | **N** | **O** | **P** | **Q** | **R** | **S** | **T** |
| **Stats.** | $\mu_{\parallel}^{\text{ref}}$ | $\sigma_{\parallel}^{\text{ref}}$ | $\gamma_{1\,\parallel}^{\text{ref}}$ | $\gamma_{2\,\parallel}^{\text{ref}}$ | $CV_{\parallel}^{\text{ref}}$ | $\mu_{\perp}^{\text{ref}}$ | $\sigma_{\perp}^{\text{ref}}$ | $\gamma_{1\,\perp}^{\text{ref}}$ | $\gamma_{2\,\perp}^{\text{ref}}$ | $CV_{\perp}^{\text{ref}}$ |

**Table 2.7 –** Tableau des statistiques extraites des images SAR en bande X et en bande C pour l'étude de la neige. Les statistiques $\gamma_1$ et $\gamma_2$ correspondent respectivement au coefficient d'asymétrie et au kurtosis définis dans la section 6.2. Les indices $^{\text{ref}}$ correspondent aux statistiques calculées sur les ratio avec une image de référence en été. Les indices $\parallel$ et $\perp$ correspondent respectivement à la co-polarisation et à la polarisation croisée.

**Résultats** Les données à notre disposition ont une fréquence de mesure quotidienne, nous les décimons en fonction des dates d'acquisition des images SAR. Nous distinguons deux périodes : la période hivernale couvrant les mois d'octobre à février, et la période printanière couvrant les mois de mars à juin. Cette séparation est faite pour distinguer les périodes de chutes de neige fraîche avec un manteau neigeux sec, des périodes de fonte de la neige et d'humidification du manteau neigeux. Nous avons calculé le coefficient de corrélation de Spearman[42] entre les différentes statistiques et les mesures correspondantes. Les résultats sont présentés dans la figure 2.29.

Nous observons que les comportements sont différents entre les bandes et entre les périodes. Les corrélations sont plus faibles en bande C qu'en bande X. En période hivernale, la bande C est principalement sensible aux températures au sol et enfouie. Les corrélations les plus fortes sont observées sur les statistiques de Nagler, avec une corrélation de $-0.75$ entre la température à 30cm (ce qui correspondrait en hiver à la température du manteau neigeux) et $\gamma_{2\,\perp}^{\text{ref}}$, et une corrélation de $-0.85$ entre la température minimale simulée du manteau neigeux et $\sigma_{\perp}^{\text{ref}}$. Une corrélation positive (0.79) est observable entre $\gamma_{2\,\parallel}^{\text{ref}}$ et la hauteur de neige. En période printanière, nous remarquons une baisse de toutes les corrélations. Les deux principales corrélations sont entre $CV_{\perp}$ et la température à 200cm (0.73) et entre $\gamma_{2\,\parallel}^{\text{ref}}$ et la température minimale simulée du manteau neigeux (0.77).

**Définition (42)**

*Corrélation de Spearman:* Mesure de dépendance non-paramétrique, comprise entre -1 et 1, qui mesure la relation monotone entre deux variables. Il consiste à estimer la corrélation entre les rangs des observations. Soit $X$ et $Y$ deux variables aléatoires et $\text{rg}_X$ et $\text{rg}_Y$ les rangs des observations de $X$ et $Y$. Le coefficient de corrélation de Spearman est donné par :

$$\rho = \frac{\text{cov}(\text{rg}_X, \text{rg}_Y)}{\sigma_{\text{rg}_X} \sigma_{\text{rg}_Y}},$$

avec cov la covariance et $\sigma$ l'écart-type.

Pour la bande X, en période hivernale les informations de moyennes ($\mu_{\|}$ et $\mu_{\perp}$) sont les plus corrélées avec l'ensemble des informations physiques. La température à 200cm, soit juste au dessus de la surface de la neige, est la plus corrélée avec les statistiques de rétrodiffusion, avec une corrélation négative de $-1$ pour $\gamma^{\text{ref}}_{2\,\|}$. L'information de TEL est moins bien corrélée (au maximum $-0.70$) que l'information de hauteur de neige (au maximum 0.90). En période printanière, les statistiques les plus corrélées sont les coefficients de variation ($CV_{\|}$ et $CV_{\perp}$) et les moyennes. Les informations physiques les plus corrélées sont la température minimale du manteau neigeux, la température à 30cm (soit la température du manteau neigeux mesurée). La hauteur de neige donne également des bonnes corrélations positives et négatives ($\pm 0.8$) en co-polarisation.

Il est intéressant de voir que la bande X semble bien adaptée pour remonter à des informations sur le manteau neigeux. Les corrélations sont moins fortes en bande C, mais nous pouvons observer dans les deux bandes une tendance a être fortement corrélé avec une information de température, soit la température du manteau neigeux ou la température à 30cm qui correspond en présence d'une haute neige supérieure à 30cm à la température du manteau neigeux. Les informations de hauteur de neige sont également bien corrélées avec les statistiques de rétrodiffusion, en particulier en bande X.

Il est important de noter que ces résultats sont obtenus sur une période d'un an, et sont basés sur des séries temporelles de statistiques mono-variées. De plus, l'estimation des statistiques est meilleure en bande X qu'en bande C, du fait de la meilleure résolution spatiale, mais la corrélation entre les séries temporelles est mieux estimée (corrélation plus significative au vu du nombre d'échantillons considérés) en bande C, du fait de la meilleure résolution temporelle. La description des images d'un point de vue texture ou multivariées n'est pas abordée.

Néanmoins ces résultats montrent que les données en bande X et en bande C ont des comportements radiométriques similaires pour les deux cas applicatifs étudiés, et sont adaptées pour l'étude de la neige et des glaciers. La pénétration en bande X rend son application plus adaptée pour les surfaces et hauteurs de neige.

## 2.4 Conclusion

Nous avons introduit dans ce chapitre les différents aspects de la cryosphère et de l'imagerie SAR, en particulier pour l'étude des glaciers et de la neige. Nous avons montré que l'imagerie SAR est un outil puissant pour l'étude de la cryosphère, en particulier pour les régions où les conditions météorologiques sont défavorables, ceci en présentant les caractéristiques et le comportement statistique des données SAR. Nous avons également montré les principales difficultés rencontrées, notamment la présence de bruit multiplicatif et de zones de distorsion rendant l'interprétation manuelle complexe.

Finalement nous avons proposé une première contribution sur l'analyse respective des données SAR en bande C et en bande X pour le suivi de la neige et des glaciers. Nous avons montré que dans le cadre de l'étude des glaciers, les deux bandes avaient un comportement fortement corrélé pour le coefficient de rétrodiffusion moyen, et ce pour les deux polarisations des données SAR. En revanche, pour l'étude de la neige, les deux bandes de fréquence ont

des comportements différents. La bande X et ses statistiques associées sont fortement corrélées à la hauteur de neige mais aussi à la température du manteau neigeux (étayée à la fois pas une information mesurée et simulée). La bande C, quant à elle, est plus sensible à l'humidité (TEL) et la température du manteaux neigeux avec des corrélations plus faibles en général. Cependant ces résultats sont à prendre avec précaution, car ils sont basés sur un nombre limité de données et de stations météorologiques mais aussi sur les configurations propres aux satellites (intégration du signal radar, incidence, puissance en émission). D'un point de vue opérationnel, la disponibilité des données est un élément à prendre en compte pour le choix de la bande de fréquence, là où la bande C offre une résolution temporelle plus élevée à un faible coût financier, la bande X offre une résolution spatiale plus fine avec un coût plus élevé.

Nous proposons dans le chapitre suivant de surmonter les difficultés de l'analyse manuelle en présentant l'apprentissage automatique dans sa globalité et en proposant une deuxième contribution sur la réalisation de deux jeux de données labellisés pour l'étude de la neige et des glaciers par imagerie SAR.

# Apprentissage automatique et jeux de données SAR

# 3


**3.1 Apprentissage statistique et automatique . . . . . 47**
**3.2 Jeu de données SAR en bande C pour l'étude de la neige humide . . . . . 67**
**3.3 Jeu de données SAR en bande X pour l'étude des glaciers . . . . . 77**
**3.4 Conclusion . . . . . 79**


Nous avons vu dans le chapitre précédent les caractéristiques de certains objets de la cryosphère au regard du signal de rétrodiffusion des satellites SAR. Nous présentons dans ce chapitre le contexte méthodologique pour identifier et caractériser ces objets d'études de manière automatique. Pour ce faire nous introduisons dans un premier temps les méthodes d'apprentissage statistique et automatique utilisées pour l'analyse de données SAR. Dans le cadre de développement de méthodes d'apprentissage automatique supervisé, il est nécessaire de disposer de données labellisées. Nous présentons notre deuxième contribution portant sur l'élaboration des jeux de données labellisés pour l'utilisation de méthodes d'apprentissage dans le cadre des deux applications que nous avons étudiées. Nous donnons une description des zones d'étude et des données utilisées, en présentant le jeu de données LSD4WSD, dédié à la caractérisation de la neige humide dans les Alpes françaises du Nord. Nous présentons dans la section 3.3 les données acquises par le satellite PAZ, utilisées pour la réalisation du dataset CPAZMAL, dont la construction permet l'étude de la classification de surfaces glaciaires au sein du massif du Mont-Blanc.

## 3.1 Méthodes d'apprentissage statistique et automatique pour l'analyse de données SAR

L'utilisation de l'imagerie SAR apporte des complications sur la méthodologie habituellement utilisée dans le traitement supervisé en vision par ordinateur. Ainsi, l'utilisation de méthodes d'apprentissage pour l'imagerie SAR est confrontée à plusieurs obstacles :

- Les distributions non gaussiennes (cf sous-section 6.2) qui compliquent la transposition des méthodes développées en optique au SAR, nécessitant une modification des hypothèses gaussiennes habituellement utilisées [170].
- La présence de bruit de speckle, qui de par sa nature multiplicative, nécessite des approches spécifiques pour son traitement [171]. De plus la présence de zones de distorsion vient masquer certaines parties de l'acquisition.
- La difficulté de labellisation[43] des données SAR. En effet, la labellisation manuelle des données SAR nécessite que les données soient en géométrie sol et interprétables, ce qui est rarement le cas. Une labellisation sur données optiques est possible, mais demande une projection précise des labels par l'intermédiaire d'un modèle numérique de terrain.

[170]: Voisin et al. (2013), « Classification of Very High Resolution SAR Images of Urban Areas Using Copulas and Texture in a Hierarchical Markov Random Field Model »

[171]: Baraha et al. (2022), « A systematic review on recent developments in non-local and variational methods for SAR image despeckling »

**Définition (43)**

*Labellisation:* Ou étiquetage, action de lier une donnée à des informations de haut niveau obtenues suite à une étape de classification par des experts ou à partir d'un autre procédé certifié correct par rapports aux spécifications d'usages

- Enfin, d'un point de vue opérationnel, les méthodes exploitant des jeux de données et de test, nécessitent toutes les précautions requises pour le traitement de zones spatiales, pour éviter les biais et les corrélations [14, 172].

[14]: Hansch et al. (2019), « The Truth About Ground Truth »
[172]: Schmitt et al. (2023), « There Are No Data Like More Data- Datasets for Deep Learning in Earth Observation »

À partir des moyens mis en œuvre pour l'observation de la cryosphère, il est nécessaire d'avoir des outils méthodologiques adéquats pour exploiter ces données [7]. On peut dissocier 3 grands axes méthodologiques dans notre cadre. Les deux premiers s'attachent à des études purement spatiales, soit par des méthodes paramétriques et soit par des méthodes non-paramétriques. Nous utilisons ici le terme de méthode paramétrique pour désigner une méthode qui repose sur une hypothèse de distribution des données, qui nécessite l'estimation de paramètres pour caractériser cette distribution. À l'inverse, le terme de méthode non-paramétrique, désigne dans notre contexte une méthode qui ne fait pas d'hypothèse sur la distribution des données, mais qui se base sur un ensemble d'échantillons pour apprendre des modèles sous-jacents ou des représentations des données. Le troisième axe s'attache à une étude spatio-temporelle afin de tirer parti de la temporalité des images SAR.

[7]: Passah et al. (2022), « SAR Image Classification »

Dans le cadre de la problématique de classification, pour la suite de ce travail, nous distinguons les méthodes supervisées des méthodes non-supervisées [173]. Les méthodes supervisées nécessitent un ensemble riche et varié de données labellisées pour entraîner un modèle, tandis que les méthodes non-supervisées n'ont pas besoin d'être guidées par ces labels. Les méthodes supervisées sont souvent plus performantes, mais nécessitent d'un jeu de données labellisées, ce qui peut être coûteux et difficile à obtenir. Les méthodes non-supervisées sont souvent utilisées pour explorer les données et identifier des structures sous-jacentes, mais peuvent être moins performantes pour la classification, n'étant pas guidées par l'a priori fort qu'est l'information de haut niveau contenue dans le label. Les autres méthodes, semi-supervisées, par renforcement, auto-supervisées, ne seront pas abordées dans ce travail.

[173]: Bruzzone et al. (2014), « A review of modern approaches to classification of remote sensing data »

### 3.1.1 Méthodes spatiales avec a priori de distribution

L'utilisation de distributions pour la classification ou la détection dans les images nécessite de modéliser correctement les données, d'introduire le bon test statistique (dans le cas de la détection) ou d'avoir la bonne mesure de distance entre les distributions pour définir l'appartenance à une classe.

#### Statistiques monovariées et multivariées

Dans le cadre de l'étude et de la classification des images SAR, de nombreux travaux se sont attachés à synthétiser l'information par un scalaire. Ce scalaire, basé sur les statistiques descriptives présentées dans la section6.2, garde un lien avec le modèle statistique des données et reste simple à analyser. Parmi ces indicateurs, les moments statistiques sont largement utilisés, la moyenne locale ($\mu$) [169] donne une information spatialisée sur le coefficient de rétrodiffusion ou encore le kurtosis ($\gamma_2$) [174], qui de par sa mesure de répartition des masses autour du centre de la distribution informe de la variabilité des rétrodiffuseurs au sein de la fenêtre considérée. Les distributions modélisant les pixels SAR, en amplitude ou en intensité, font partie des familles plus générales des lois de type Gamma, Gamma Inverse, Pearson. L'étude proposée dans [175]

[169]: Kaushik et al. (2023), « Temporal Evolution of X and C Band SAR Backscattering in theMont-Blanc Massif »
[174]: Lee et al. (2021), « Reduction of False Alarm Rate in SAR-MTI Based on Weighted Kurtosis »
[175]: Nicolas (2002), « 1 - Introduction aux Statistiques de deuxième espèce »

apporte des bases pour estimer les paramètres de ces lois statistiques par l'utilisation des log-cumulants. Ces outils robustes dans l'estimation des paramètres, permettent de plus, par la transformée de Mellin, de composer les lois de bases Gamma et Gamma Inverse afin de générer d'autres lois plus complexes. En utilisant le diagramme des log-cumulants $\kappa_2$, $\kappa_3$, les auteurs de [176] ont réalisé un test de détection de clutter maritime (échos parasites revenant de cibles définies comme étant sans intérêt).Dans [177], l'idée proposée dans [175] est étendue, pour analyser non plus une image mais plusieurs, basculant d'un paradigme mono-varié à un autre multi-varié. Cette proposition d'outil de statistiques complexes de type Mellin à matrice variable, permet de prendre en compte les corrélations (phase et amplitude) entre les différents canaux. L'application de ce travail est réalisée dans [178], par l'étude de textures dans des images polarimétriques. L'utilisation des modèles de distributions paramétriques est ainsi proposée dans l'étude de la neige humide [8], où le ratio normalisé entre deux images (suivant une loi de type Beta prime) est utilisé pour réaliser un seuillage sur la distribution cumulative. Ceci est réalisé après estimation des paramètres des lois Gammas des deux images et ce, avec respect du seuil modélisé en fonction de l'angle d'incidence. Ce travail s'ancre dans la continuité de l'approche par seuillage proposée dans [145] où la détection de pixels de neige humide est réalisée sans prise en compte de la distribution statistique associée, mais permet en utilisant une description statistique d'obtenir des résultats prenant en compte la variabilité de chaque image. La validation de cette approche reposant sur les résultats d'une modélisation multi-couches en amont qui peut introduire une incertitude sur le seuil. L'utilisation des modèles de speckle, comme celui multiplicatif, a permis de proposer des modèles de distributions plus adaptées ($\mathcal{G}^0_A$) pour la classification de zones hétérogènes [179].

[176]: Tao et al. (2012), « Sea clutter contamination test with log-cumulants »

[177]: Anfinsen et al. (2011), « Application of the Matrix-Variate Mellin Transform to Analysis of Polarimetric Radar Images »

[178]: Bombrun et al. (), « A Complete Coverage of Log-Cumulant Space in Terms of Distributions for Polarimetric SAR Data »

[8]: Besic et al. (2015), « Stochastic Approach in Wet Snow Detection Using Multitemporal SAR Data »

[145]: Nagler et al. (2000), « Retrieval of wet snow by means of multitemporal SAR data »

[179]: M. E. Mejail et al. (2003), « Classification of SAR images using a general and tractable multiplicative model »

L'exploitation de tous les canaux polarimétriques SAR ou d'un ensemble d'acquisitions temporelles nécessite de prendre en compte la dépendance entre les canaux. Les modèles multivariés permettent de prendre en compte cette dépendance, au détriment d'une complexité plus importante [112]. Dans le cadre de séries temporelles, la problématique est celle de la détection ou classification de changement. Ainsi les auteurs de [180], proposent l'utilisation de modèle Spherical Invariant Random Vector (SIRV), pour cette détection en caractérisant 3 tests statistiques. Le modèle SIRV est basé sur un vecteur gaussien composé (aussi appelé compound gaussian), défini comme le produit d'un vecteur gaussien $\mathbf{x} \in \mathbb{R}^m$ et de la racine carrée d'une variable aléatoire scalaire positive $\tau \in \mathbb{R}^+$ aussi appelé paramètre de texture. La résultante est donnée par :

$$\mathbf{z} = \sqrt{\tau} \cdot \mathbf{x}, \tag{3.1}$$

avec $\mathbf{x} \sim \mathcal{N}(0, \Sigma)$, $\Sigma$ étant la matrice de covariance. Cette approche a été déclinée en considérant $\tau$ déterministe ou suivant une distribution spécifique comme une loi log-normal [181]. L'intérêt de ces approches est de garder une structure gaussienne plus simple à manipuler, mais en introduisant une variabilité pour sortir du cadre strict gaussien. Une extension est proposée dans [182], en ajoutant la moyenne $\mu$ au vecteur gaussien $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$, et en dérivant la formule de divergence de Kullback-Leibler associée.

[112]: Yue et al. (2021), « Synthetic Aperture Radar Image Statistical Modeling »

[180]: Mian et al. (2019), « New Robust Statistics for Change Detection in Time Series of Multivariate SAR Images »

[181]: Chalabi et al. (2019), « Estimators of compound Gaussian clutter with log-normal texture »

[182]: Collas et al. (2023), « Riemannian Optimization for Non-Centered Mixture of Scaled Gaussian Distributions »

De manière similaire, mais plus proche des distributions usuelles en SAR, les travaux proposés dans[183] portent sur l'étude de détection de changement en utilisant des lois de type Gamma bivariées (appelées multisensor multivariate gamma distributions : MuMGDs). Les nombreux canaux polarimétriques se prêtent à la modélisation multivariée dans le cadre de la classification

[183]: Chatelain et al. (2008), « Change Detection in Multisensor SAR Images Using Bivariate Gamma Distributions »

non-supervisé [184] par l'utilisation de distributions de Wishart. Néanmoins, l'utilisation de modèle multivarié est complexe, dans la phase d'estimation et de test, si l'on souhaite des formes analytiques.

**Copules :**

Une autre manière de modéliser les caractéristiques ou les données brutes, repose sur l'utilisation des copules (voir la figure 3.1). Une copule $C$ est une fonction de distribution cumulative conjointe (CDF) définie sur le cube unité d-dimensionnel $[0\,;1]^d$, telle que sa projection sur tout axe (marginale) définisse une loi uniforme sur $[0;1]$. Cette approche permet de décrire des données multi-variées en séparant le comportement des marginales de la structure de dépendance (copule). Les copules sont régulièrement utilisées en finance [185] et en suivi climatique [186] pour leur capacité à modéliser les événements extrêmes (queue lourde des distributions).

Les copules ont ainsi pu être appliquées dans le cadre de la modélisation de la dépendance entre des labels et des caractéristiques pour une classification supervisée [187], permettre de modéliser les dépendances entre des paramètres polarimétriques [188] dans le cadre d'une classification de signaux radar météo ou encore en réalisant une fusion entre des images SAR aéroportées et satellitaires [189]. L'intégration de ces méthodes dans les modèles de réseaux profonds est récente et encore peu exploitée [190], mais offre des perspectives sur la notion de modélisation de dépendance au sein des réseaux.

Cette proposition peut ouvrir la porte à d'autres fusions d'images dans le cadre de la classification de la neige humide, comme la composition d'une image de référence par fusion de plusieurs, ou la fusion entre des images SAR en ascendant et en descendant.

Il est possible de placer la dépendance soit sur un axe spatial comme le propose [191] en examinant les performances d'une copule gaussienne dans le contexte d'une décomposition en ondelettes pour caractériser les textures. Cette approche est complémentée dans [192], où un problème similaire est abordée en appliquant des lois gammas aux marginales.Les auteurs de [170] utilise les copules pour effectuer des classifications dans des environnements urbains en se basant sur des données SAR haute résolution. La dépendance peut aussi être placée sur un axe temporel comme dans [193], pour une détection de changement à l'aide d'images SAR. Cette dimension temporelle est exploitée dans [194] et [195] pour la prédiction de la force du vent, ou dans un cadre de série temporelle d'images SAR dans [196].

Les copules bivariées sont particulièrement adaptées pour modéliser les données SAR en double polarisation, et permettent de garder un formalisme simple et des solutions analytiques. Dans le cas de copules de dimension supérieure à 2, il est courant de décomposer la structure en arbre et d'utiliser des copules bivariées pour les branches (Vine copula [197, 198]). Le lecteur pourra se référer à [199] pour une introduction aux copules et à [200] pour un détail des structures utilisées dans le cadre des sciences environnementales.

De la même manière qu'une distribution marginale (histogramme versus loi statistique), les copules peuvent être modélisées (estimation d'un paramètre spécifique à la copule) ou empiriquement estimées [201]. Nous nous intéressons dans la suite de ce travail aux copules basées sur des modèles de distributions

[184]: Liu et al. (2018), « Unsupervised Classification of Multilook Polarimetric SAR Data Using Spatially Variant Wishart Mixture Model with Double Constraints »

[185]: Bouyé et al. (2000), « Copulas for Finance - A Reading Guide and Some Applications »

[186]: Tootoonchi et al. (2022), « Copulas for Hydroclimatic Analysis »

[187]: Liu (), « Copula Multi-label Learning »

[188]: Yanovsky et al. (2016), « Multivariate copula approach for polarimetric classification in weather radar applications »

[189]: Wang et al. (2022), « Proposal-Copula-Based Fusion of Spaceborne and Airborne SAR Images for Ship Target Detection** »

[190]: Ling et al. (2020), « Deep archimedean copulas »

[191]: Lasmar et al. (2014), « Gaussian Copula Multivariate Modeling for Texture Image Retrieval Using Wavelet Transforms »

[192]: Stitou et al. (2009), « Copulas based multivariate gamma modeling for texture classification »

[170]: Voisin et al. (2013), « Classification of Very High Resolution SAR Images of Urban Areas Using Copulas and Texture in a Hierarchical Markov Random Field Model »

[193]: Mercier et al. (2008), « Conditional Copulas for Change Detection in Heterogeneous Remote Sensing Images »

[194]: Bessa (2016), « On the quality of the Gaussian copula for multi-temporal decision-making problems »

[195]: Valizadeh Haghi et al. (2015), « Spatiotemporal Modeling of Wind Generation for Optimal Energy Storage Sizing »

[196]: Cui et al. (2011), « Mining Satellite Image Time Series »

[197]: Bedford et al. (2002), « Vines–a new graphical model for dependent random variables »

[198]: Czado et al. (2022), « Vine Copula Based Modeling »

[199]: Nelsen (1999), *An Introduction to Copulas*

[200]: Bhatti et al. (2019), « Recent development in copula and its applications to the energy, forestry and environmental sciences »

[201]: Safaai et al. (2018), « Information Estimation Using Nonparametric Copulas »

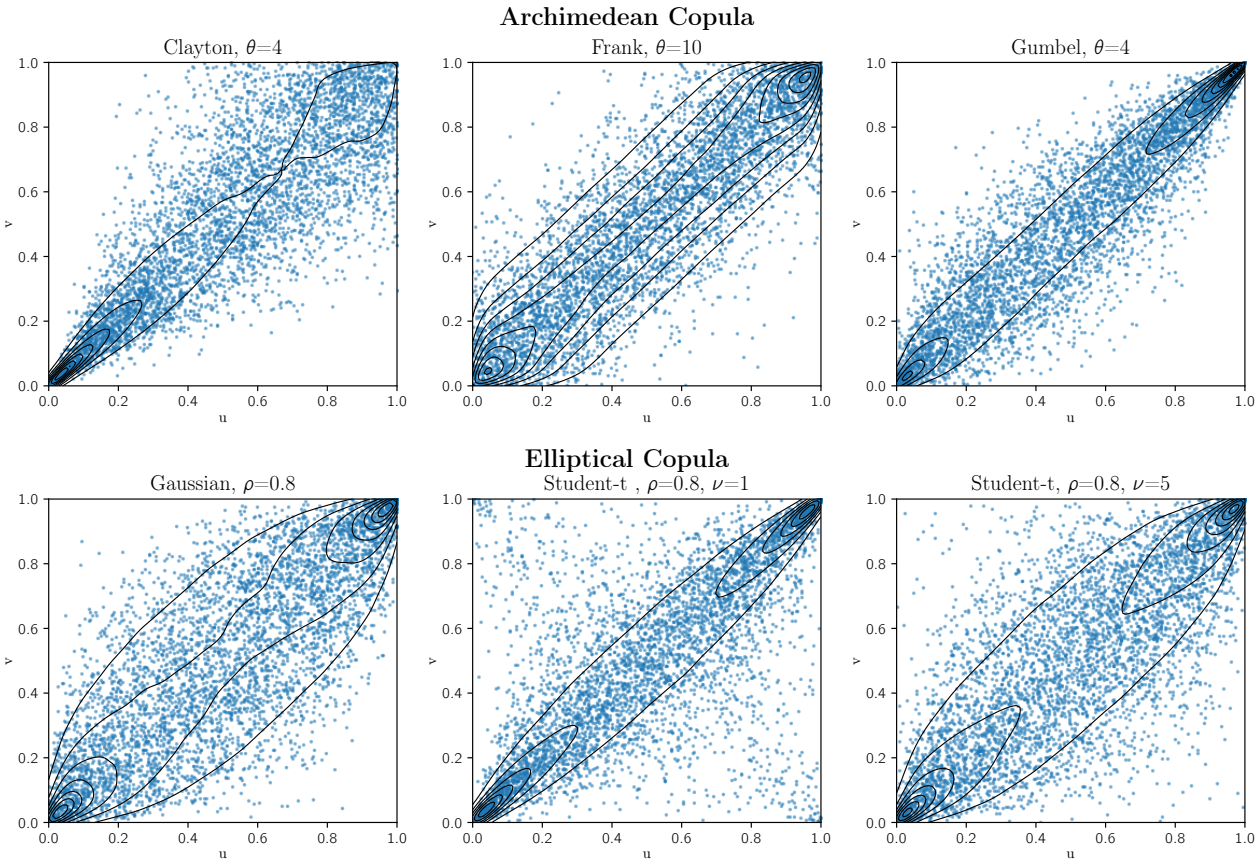



**Figure 3.1 –** Illustration des copules elliptique et archimédienne pour différentes paramétrisations. Les copules archimédiennes sont des copules asymétriques (sauf la copule de Frank) avec la capacité de modéliser des queues lourdes (hautes ou basses) tandis que les copules elliptiques sont symétriques et la modélisation des queues est réalisable par une copule de Student-t induisant une queue symétrique (haute et basse).

paramétriques. Deux grandes familles peuvent être distinguées : les copules archimédiennes et les copules elliptiques, illustrées dans la figure 3.1.

Les copules archimédiennes sont faciles à échantillonner et ont une forme mathématique simple dans le cas bivarié. Elles capturent les dépendances (ou concordances) au niveau de la queue. Plusieurs familles archimédiennes présentent une dépendance asymptotique dans au moins une des deux queues (dépendances à très long terme) axiales, les rendant particulièrement adaptées à la modélisation d'événements extrêmes [202]. La forme générale des copules archimédiennes, pour $u, v$ marginales monovariées uniformes sur $[0, 1]$ est donnée par :

$$C_\theta(u, v) = \psi\left(\psi^{-1}(u) + \psi^{-1}(v)\right), \tag{3.2}$$

où $\psi$ est une fonction, dite génératrice, continue strictement décroissante et convexe, $\psi^{-1}$ sa pseudo-inverse [203] et $\theta$ un paramètre de dépendance. La copule de Gumbel [204] est une copule asymétrique avec une probabilité plus élevée concentrée dans la queue supérieure, obtenue par $\phi(u) = (-\log(u))^\theta$, tandis que les copules de Clayton [205] ont une densité plus élevée dans la queue inférieure. Les copules de Frank [206], qui sont symétriques, ont tendance à avoir une densité plus faible dans les queues par rapport aux copules de Clayton et de Gumbel et sont capables de capturer des corrélations négatives également.

Les copules elliptiques sont capables de représenter les dépendances positives et négatives (dépendances globales), mais les formes classiques (gaussiennes

[202]: Salvadori et al. (2007), « Extremes in Nature : An Approach Using Copulas »

[203]: McNeil et al. (2009), « Multivariate Archimedean copulas, d-monotone functions and l1-norm symmetric distributions »

[204]: Gumbel (1960), « Bivariate Exponential Distributions »

[205]: CLAYTON (1978), « A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence »

[206]: Frank (1978), « On the simultaneous associativity of F(x, y) and x+y-F(x, y) »

et de Student-t) ne sont pas adaptées pour modéliser les queues lourdes et ne sont pas asymétriques. La forme générale des copules elliptiques est donnée par :

$$C_\theta(u, v) = \Phi(\Phi^{-1}(u), \Phi^{-1}(v)) \tag{3.3}$$

avec $\Phi$ une distribution multivariée appropriée, $\Phi^{-1}$ les inverses des distributions marginales univariées et $\theta$ les paramètres de la copule. Les deux types de copules elliptiques les plus largement utilisés sont les copules gaussiennes et de Student-t (ligne du bas de la figure 3.1). Les copules gaussiennes, tout comme les distributions gaussiennes bivariées, peuvent modéliser des dépendances positives et négatives et sont symétriques. Cependant, contrairement aux distributions gaussiennes bivariées, les marginales ne sont pas limitées à suivre uniquement une distribution gaussienne. La copule de Student-t se caractérise par une plus grande flexibilité de modélisation en termes de dépendance en queue, ce qui est rendu possible par un paramètre de degré de liberté ($\nu$) supplémentaire par rapport à la copule gaussienne. L'estimation des paramètres peut s'avérer complexe. Dans certains cas le paramètre $\theta$ peut être directement lié au tau de Kendall[44] . Dans les autres cas, la méthode de pseudo-maximum likelihood (PML) [207] se base sur la transformation des marginales en variables aléatoires uniformes sur $[0, 1]$ par une estimation non-paramétrique d'histogrammes [208] suivie par estimation Maximum Likelihood.

**Définition (44)**

*Tau de Kendall:* Cette statistique mesure la corrélation de rang entre deux variables et est définie par :

$$\tau = \frac{\text{conc.} - \text{disc.}}{0.5n \cdot (n-1)},$$

avec $textconc.$, disc. une paire d'observation, respectivement concordante et discordante. Soit les paires d'observations $(x_i, y_i)$ et $(x_j, y_j)$, elles sont dites concordantes si $x_i < x_j$ et $y_i < y_j$ ou $x_i > x_j$ et $y_i > y_j$ et discordantes si $x_i < x_j$ et $y_i > y_j$ou $x_i > x_j$ et $y_i < y_j$.

Nous avons présenté jusqu'à présent un ensemble de méthodes utilisant des modèles plus ou moins complexes mais en présence d'un milieu homogène formé d'un seul type de rétrodiffuseurs. Dans le cas où plusieurs cibles de natures différentes sont présentes, les modèles de mélange sont plus adaptés, en particulier les mélanges de gaussiennes [209] ou de gamma [210], en modélisant la distribution des pixels par une somme de plusieurs distributions de paramètres différents. De part le traitement séparé des marginales et de la dépendance, les copules sont particulièrement adaptées pour les images SAR multi-canaux, ou des données de modalités différentes. Leur utilisation dans le cadre de la classification de paires d'images SAR est abordée dans le chapitre Chapitre 6. L'utilisation de distributions nécessite ainsi de comparer les modèles entre eux ou de tester des hypothèses statistiques afin de réaliser les tâches de classification. Nous nous concentrons dans la suite de ce travail sur la comparaison de ces modèles, en utilisant des divergences statistiques.

[207]: GENEST et al. (1995), « A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions »

[208]: Jaworski et al. (2010), « Copula Theory and Its Applications »

[209]: Yang et al. (2019), « Unsupervised Change Detection of SAR Images Based on Variational Multivariate Gaussian Mixture Model and Shannon Entropy »

[210]: Li et al. (2015), « Unsupervised learning of generalized gamma mixture model with application in statistical modeling of high-resolution SAR images »

### Divergences et distance

Afin de quantifier la proximité de distributions de probabilités à partir d'observations, soit par une approche non-paramétrique (histogrammes par exemple) ou en utilisant des modèles de probabilité paramétriques, il est nécessaire de déterminer une mesure de distance ou de divergence entre ces distributions. Les divergences statistiques sont des outils permettant de mesurer la dissimilarité entre deux distributions de probabilité. Une divergence est une fonctionnelle qui évalue la dissimilarité entre deux lois de probabilité. Les divergences permettent d'étendre la notion de distance aux espaces statistiques mais ne possèdent pas certaines propriétés des distances, telles que la symétrie et l'inégalité triangulaire. Nous pouvons distinguer plusieurs familles de divergences, dont les deux d'intérêt pour ce travail sont données par :

- Les $f$-divergences, qui sont des divergences définies par une fonction convexe $f$. Soient $P$ et $Q$ deux variables aléatoires et $p$ et $q$ leurs fonctions

de densité de probabilité respectives, la $f$-divergence est définie par :

$$D_f(P||Q) = \int p(x) f\left(\frac{p(x)}{q(x)}\right) dx.$$

- Les $\alpha$-divergences, dérivées des $f$-divergences, introduisent une paramétrisation $\alpha$ permettant de contrôler le comportement de la divergence.

D'autres familles de divergences existent (divergence de Bregman, $\beta$-divergence), que nous ne traiterons pas dans ce travail. Cependant dans le cadre de classification sur des modèles statistiques, les méthodes basées sur la divergence statistique ont montré leur efficacité [211]. Les divergences ont été utilisées pour la classification de textures [212], dans le clustering [213] avec de belles propositions dans l'estimation de centroïdes avec des divergences de Jensen-Shannon [214]. Un des principaux avantages des divergences est la possibilité d'obtenir une forme analytique pour caractériser la dissimilarité entre deux distributions, par exemple entre des distributions monovariées de type Weibull [215], ou sur des modèles multivariés plus complexes comme les copules [191]. Cependant, ces formes analytiques ne sont pas toujours disponibles et il est nécessaire de recourir à des méthodes numériques pour les estimer, qui peuvent s'avérer coûteuses en temps de calcul. Pour pallier à ces problèmes, des méthodes d'apprentissage profond ont été proposées, afin d'apprendre soit la fonctionnelle d'une divergence de Bregman [216] soit la divergence en elle même [217], et ce à partir de données d'apprentissage.

Leur utilisation dans le contexte d'images SAR, est principalement centrée sur la détection de changement entre deux dates [218] ou dans le cadre de la construction de matrices de divergences sur l'ensemble d'une série temporelle [219]. Nous retrouvons également des travaux sur la classification de données polarimétriques modélisées par des lois de Wishart [220] ou par des lois gammas généralisées [221].

Nous nous intéressons dans la suite de ce travail à la divergence de Kullback-Leibler (KLD) et à ces symétrisations, ainsi qu'à la $\alpha$-divergence de Rényi [222], qui est une généralisation de la KLD. Ce choix est motivé par les nombreuses formes analytiques disponibles pour ces divergences, l'utilisation extensive de la KLD dans la littérature et la capacité de la divergence de Rényi à généraliser un grand nombre de divergences bien connues [223].

La divergence de Kullback-Leibler (KLD), notée $\mathcal{K}$ dans la suite, est définie par :

$$\mathcal{K}(P||Q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx. \tag{3.4}$$

Nous introduisons la version symétrique de la KLD, aussi appelée divergence de Jeffreys,[45] définie par :

$$\mathcal{J}(P||Q) = \mathcal{K}(P||Q) + \mathcal{K}(Q||P). \tag{3.5}$$

La divergence de Rényi d'ordre $\alpha$ entre deux variables Q $P$ et $Q$ est définie par :

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \int p(x)^\alpha q(x)^{1-\alpha} dx, \tag{3.6}$$

avec $\alpha \in \mathbb{R}_+^*$ et $\alpha \neq 1$. Un aperçu des divergences de Rényi pour différentes valeurs de $\alpha$ est donné dans le tableau A.1. Il est important de noter que lorsque $\alpha \to 0$ le résultat n'est pas une divergence [223]. La distance de Bhattacharyya,

[211]: Basseville (2013), « Divergence measures for statistical data processing–An annotated bibliography »

[212]: Kwitt et al. (2008), « Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval »

[213]: Kampffmeyer et al. (2019), « Deep divergence-based approach to clustering »

[214]: Nielsen (2020), « On a Generalization of the Jensen–Shannon Divergence and the Jensen–Shannon Centroid »

[215]: Bauckhage (2013), *Computing the Kullback-Leibler Divergence between two Weibull Distributions*

[191]: Lasmar et al. (2014), « Gaussian Copula Multivariate Modeling for Texture Image Retrieval Using Wavelet Transforms »

[216]: Rezaei et al. (2023), « Deep Bregman divergence for self-supervised representations learning »

[217]: Cilingir et al. (2020), « Deep divergence learning »

[218]: Xu et al. (2013), « Change detection on SAR images by a parametric estimation of the KL-divergence between Gaussian Mixture Models »

[219]: Atto et al. (2013), « Multidate Divergence Matrices for the Analysis of SAR Image Time Series »

[220]: Silva et al. (2013), « Classification of segments in PolSAR imagery by minimum stochastic distances between Wishart distributions »

[221]: Qin et al. (2015), « Region-based classification of SAR images using Kullback–Leibler distance between generalized gamma distributions »

[222]: Rényi (1961), « On measures of entropy and information »

[223]: Van Erven et al. (2014), « Rényi divergence and Kullback-Leibler divergence »

**Précision (45)**

Une autre façon de considérer la symétrisation de la divergence de Kullback-Leibler est la divergence de Jensen-Shannon, définie par $\mathcal{JS}(X_p, X_q) = \frac{1}{2}\mathcal{K}(X_p||M) + \frac{1}{2}\mathcal{K}(X_q||M)$, où $M = \frac{1}{2}(X_p + X_q)$.

**Table 3.1 –** Divergences de Rényi pour différentes valeurs de $\alpha$, adapté de [224]

| $\alpha$ | **Divergence de Rényi** | **Commentaires** |
|---|---|---|
| $\alpha \to 1$ | $\int p(x) \log \frac{p(x)}{q(x)}\, dx$ | Divergence de Kullback-Leibler |
| $\alpha = 0.5$ | $-2 \log \int \sqrt{p(x)q(x)}\, dx$ | Distance de Bhattacharyya |
| $\alpha = 2$ | $\log(1 + \chi^2(P\|\|Q))$ | Homogène à divergence $\chi^2$ |
| $\alpha \to +\infty$ | $\log \max \frac{P}{Q}$ | Longueur minimale de description |

en dépit de son nom, n'est pas une distance car elle ne respecte pas l'inégalité triangulaire, cependant elle possède la particularité d'être symétrique dans ses arguments. Assez peu de travaux utilisent la divergence $\chi^2$ et ses variantes de Pearson ou de Neyman, mais plus de détails sur cette divergence peuvent être trouvés dans [225]. Les différentes formes particulières de la divergence de Rényi ont pu être utilisées pour la segmentation d'images SAR [226], la reconnaissance de motifs [227] ou encore la détection de cible dans les images SAR [228]. Le lecteur pourra trouver un descriptif exhaustif avec leur application en problème inverse dans [229]

Dans cette sous section, nous avons abordé les méthodes paramétriques pour la classification d'images SAR, en se basant sur la modélisation mathématique (issue de la compréhension physique du signal) ou empirique des acquisitions considérées. Cependant, il est parfois restrictif de considérer uniquement l'image ou l'acquisition brute, et la recherche de caractéristiques ou de décompositions peut s'avérer plus efficace.

### 3.1.2 Caractéristiques et décompositions

Afin d'enrichir, de simplifier ou d'orienter l'apprentissage, il peut être intéressant de travailler sur des caractéristiques extraites directement des données [230]. Pour les données SAR, les statistiques simples comme la moyenne, présentées dans la section 6.2, peuvent être vues comme caractéristiques, dont l'intérêt supplémentaire est de réduire la dimension des données. Ainsi le calcul et l'utilisation d'histogramme empirique généralise cette approche en réduisant la dimension des données tout en conservant une information importante. Le découpage de l'histogramme est crucial et des méthodes ont pu être développées pour le déterminer automatiquement à partir des données. Une des méthodes les plus courantes est la méthode de Freedman-Diaconis [231] où le nombre optimal d'intervalles $H_{bins}$, est calculé à partir d'un échantillon de données $x$ de taille $n$ et de quartiles $q_{25\%}$ et $q_{75\%}$ de $x$ :

$$H_{bins} = \left\lceil \frac{\sqrt[3]{n}(\max \mathbf{x} - \min \mathbf{x})}{2 \cdot (q_{75\%} - q_{25\%})} \right\rceil, \tag{3.7}$$

La classification de zone en SAR peut se rapprocher de l'étude des textures, et les techniques développées dans ce cadre y sont applicables [232]. Ainsi nous retrouvons dans cette tâche les extracteurs de caractéristiques tel que SIFT[46] [234] particulièrement adapté pour une mise en correspondance entre images ou la détection de cible [235]. D'autres méthodes telles que l' HOG (Histogram of Oriented Gradients), SURF (Speeded-Up Robust Features) ou encore GLCM

[225]: Nielsen et al. (2014), « On the chi square and higher-order chi distances for approximating f-divergences »

[226]: Galland et al. (2003), « Minimum description length synthetic aperture radar image segmentation »

[227]: Gao et al. (2022), « An improved belief chi2 divergence for Dempster–Shafer theory and its applications in pattern recognition »

[228]: Tang et al. (2022), « Large-difference-scale target detection using a revised Bhattacharyya distance in SAR images »

[229]: Lanteri (2019), « Divergences. divergences invariantes. applications aux problèmes inverses linéaires, nmf et déconvolution aveugle »

[230]: Prats-Montalbán et al. (2011), « Multivariate Image Analysis »

[231]: Freedman et al. (1981), « On the histogram as a density estimator :L2 theory »

[232]: Fekri-Ershad (2019), « Texture Analysis Using Machine Learning Methods - A Review »

**Définition (46)**

*SIFT:* Scale-Invariant Feature Transform [233], méthode permettant de détecter et décrire les caractéristiques distinctives d'une image, comme les coins et les bords, de manière invariante à l'échelle et à la rotation. Elle fonctionne en détectant d'abord les points d'intérêt à partir des variations de luminosité, puis en affinant leur localisation pour une correspondance précise entre images, et enfin en décrivant ces points par des vecteurs de caractéristiques robustes aux transformations géométriques, facilitant ainsi la comparaison et la correspondance entre les points d'intérêt.

[234]: Sedaghat et al. (2015), « Remote sensing image matching based on adaptive binning SIFT descriptor »

[235]: Dong et al. (2021), « Keypoint-Based Local Descriptors for Target Recognition in SAR Images »

(gray-level co-occurrence matrix) sont également utilisées pour extraire des informations texturales. Ainsi dans le cadre d'une seule polarisation, [236] et [237] proposent une classification respectivement des côtes maritimes et des zones d'eau en utilisant la caractéristique GLCM. Dans le cadre de classification avec plusieurs canaux polarimétriques, [132] étudie la combinaison optimale d'une centaine de caractéristiques polarimétriques comme les matrices de covariance ou l'index de végétation radar. Nous pouvons y voir une approche de type BOV (Bag of Visual Words) ou par dictionnaire, où les caractéristiques sont extraites puis regroupées pour former un dictionnaire auquel chaque image est comparée [238].

L'utilisation de techniques de filtrage pour décomposer une information multi-échelle, est également une approche courante, notamment avec les décompositions en ondelettes [239]. Ces décompositions permettent également de minimiser l'impact du bruit de speckle [240]. Il est possible de voir les décompositions en ondelettes comme une forme simplifiée des réseaux de neurones, où le noyau est fixe. Par conséquent, certains travaux se sont attachés à utiliser les premières couches des réseaux profonds pré-entraînés pour extraire des caractéristiques pertinentes pour la classification [241].

Enfin, notamment dans le domaine de la télédétection, des informations complémentaires peuvent être ajoutés afin d'affiner les résultats, comme l'ajout d'informations topographiques [90].

La classification peut se faire en cherchant à les modéliser ces caractéristiques. C'est le cas dans [242], où le modèle de texture associé est le SIRV. De manière similaire, [241] réalise une classification sur une variété riemannienne en se basant sur la nature des matrices de covariances de ces caractéristiques, également adaptée dans le cadre de données GPR (Ground Penetrating Radar) [243]. Cette utilisation de la géométrie de l'information pour décrire au mieux les données est retrouvée dans [134] pour la classification d'images PolSAR avec le modèle SIRV. [244] réalise une classification de texture SAR en mêlant graphe et entropie de permutation, pour obtenir une approche efficace avec seulement 2 caractéristiques. On remarque que ces études peuvent être plus complexes que des réseaux de neurones, mais chaque étape est explicable et cherche à exploiter au maximum la nature des données utilisées. Dans l'approche de [244] une attention particulière est portée sur la capacité de ces méthodes à obtenir des performances respectables tout en gardant un coût calculatoire maîtrisé. Ceci fait sens dans un contexte de big-data, où la quantité d'information spatiale et temporelle pour l'imagerie satellitaire est très importante et où la facilité de l'utilisation des réseaux de neurones peut augmenter considérablement les temps et les coûts de calcul.

[236]: Wang et al. (2011), « Coastline detection in SAR images using multi-feature and SVM »

[237]: Wentao-Lv et al. (2010), « Water extraction in SAR images using GLCM and Support Vector Machine »

[132]: Yin et al. (2019), « Optimal combination of polarimetric features for vegetation classification in PolSAR image »

[238]: Feng et al. (2011), « Bag-of-Visual-Words Based on Clonal Selection Algorithm for SAR Image Classification »

[239]: Mian et al. (2019), « Design of New Wavelet Packets Adapted to High-Resolution SAR Images With an Application to Target Detection »

[240]: Fonseca et al. (2023), « Wavelet Spatio-Temporal Change Detection on Multitemporal SAR Images »

[241]: Akodad et al. (2020), « Ensemble Learning Approaches Based on Covariance Pooling of CNN Features for High Resolution Remote Sensing Scene Classification »

[90]: Tsai et al. (2019), « Wet and Dry Snow Detection Using Sentinel-1 SAR Data for Mountainous Areas with a Machine Learning Technique »

[242]: Bombrun et al. (2011), « Multivariate texture retrieval using the geodesic distance between elliptically distributed random variables »

[243]: Gallet et al. (2022), « Classification of GPR Signals via Covariance Pooling on CNN Features within a Riemannian Framework »

[134]: Formont et al. (2013), « On the Use of Matrix Information Geometry for Polarimetric SAR Image Classification »

[244]: Chagas et al. (2021), « Analysis and Classification of SAR Textures Using Information Theory »

### 3.1.3 Méthodes spatiales sans a priori de distribution

Par l'explosion des données et le nombre croissant de missions satellitaires variées, les méthodes sans a priori de distribution sont particulièrement adaptées. Dans ce cadre, la principale motivation est l'apprentissage par la donnée, de modèles sous-jacents complexes, sans ou avec peu d'a priori sur la distribution des données. Nous pouvons distinguer les méthodes d'apprentissage dites classiques et celles d'apprentissage profond. Les premières ont posé les bases théoriques et pratiques de l'apprentissage automatique, tandis que les secondes ont permis de résoudre des problèmes plus complexes et ont été démocratisées par l'augmentation de la puissance de calcul et la disponibilité

de données. Elles peuvent être utilisées pour des tâches de classification, de segmentation, de détection de changements, de débruitage, de reconstruction, de super-résolution, de détection d'anomalies, de prédiction, de génération de données, de réduction de dimension, de clustering. Nous nous intéressons ici à la classification et présentons plus en détail les méthodes dédiées à cette tâche.

### Apprentissage automatique classique

Les descriptions théoriques des algorithmes peuvent être trouvées dans [245] et une brève présentation de leur utilisation sur les problèmes de télédétection dans [246]. Un des principaux avantages de ces méthodes est leur faible nombre de variables à apprendre (en comparaison aux réseaux de neurones profonds) et d' hyperparamètres[47] à régler. Dans la variété des algorithmes d'apprentissage automatique, nous pouvons distinguer trois catégories principales pour la tâche de classification :

- Les algorithmes basés sur l'apprentissage d'un ensemble de classifieurs faibles, tels que les arbres de décision, les forêts aléatoires [247] ou Adaboost [248]. Les forêts aléatoires agrègent un grand nombre d'arbres de décision pour réduire le sur-ajustement et améliorer la généralisation. En revanche, AdaBoost itère sur des modèles d'apprentissage faibles, en donnant plus de poids aux observations mal classées à chaque itération, ce qui concentre l'effort d'apprentissage sur les exemples difficiles. La décision finale est prise en agrégeant l'information soit par la majorité des votes (forêt aléatoire) soit par une pondération adaptative (AdaBoost), permettant une interprétation des résultats obtenus. L'utilisation d'un grand nombre de classifieurs permet un apprentissage rapide [249] (voire dans certains cas parallélisable).
- Les algorithmes basés sur le concept de séparation ou de distance. Certains algorithmes réalisent la classification en projetant les données pour mieux distinguer les classes, en maximisant la séparation ou la distance entre elles à travers l'apprentissage d'un hyperplan de séparation, comme le font les Séparateurs à Vaste Marge (SVM ou *Support Vector Machine* SVM) [250]. D'autres méthodes déterminent directement la classe d'un nouvel échantillon en fonction d'une distance (euclidienne, riemannienne, divergence) des échantillons voisins et de leurs étiquettes comme la méthode K plus proches voisins (*K-Nearest Neighbors* KNN) [251], illustrée en figure 3.2.
- La dernière catégorie couvre les algorithmes d'apprentissage automatique neuronal. Ils sont basés sur l'apprentissage de poids pour une combinaison linéaire des entrées ou des caractéristiques et sont couplés avec des fonctions non linéaires. Alors que la régression logistique (*Logistic Regression* LR) [252] n'apprend qu'un seul niveau de poids, le perceptron multicouche (*Multi-Layer Perceptron* MLP) [253] est composé d'un certain nombre de couches d'apprentissage successives.

Cette dernière catégorie est la base des réseaux de neurones profonds.

### Réseaux de neurones et apprentissage profond

Naturellement, certains travaux ont exploité la montée des réseaux profonds (Deep Neural Networks) pour réaliser des tâches de classification ou de segmentation dans un cadre d'observation de la Terre [254]. L'article de [255]

[245]: Hastie et al. (2009), *The Elements of Statistical Learning*

[246]: Lary et al. (2016), « Machine learning in geosciences and remote sensing »

**Définition (47)**

*Hyperparamètres:* Les hyperparamètres des réseaux de neurones profonds sont des paramètres fixés avant l'entraînement, influençant le processus d'apprentissage et la structure du réseau. Contrairement aux paramètres du modèle (poids et biais) appris pendant l'entraînement, les hyperparamètres ne sont pas modifiés durant ce processus. Ces hyperparamètres sont souvent optimisés par des méthodes de validation croisée ou de recherche hyperparamétrique pour minimiser l'erreur de généralisation du modèle. Parmi les hyperparamètres les plus courants, on peut citer : le taux d'apprentissage (*learning rate*) qui est le pas de mise à jour des poids, la taille des mini-lots (*batch size*) qui est le nombre d'échantillons utilisés pour une mise à jour des poids, le nombre d'époques (*epochs*) qui est le nombre de fois que l'ensemble des données est utilisé pour l'entraînement, l'architecture du réseau (nombre de couches, nombre de neurones par couche, fonctions d'activation).

[247]: Breiman (2001), « Random forests »

[248]: Freund et al. (1995), « A desicion-theoretic generalization of on-line learning and an application to boosting »

[249]: Sani et al. (2018), « Computational Complexity Analysis of Decision Tree Algorithms »

[250]: Cortes et al. (1995), « Support-Vector Networks »

[251]: Cover et al. (1967), « Nearest neighbor pattern classification »

[252]: Anderson (1972), « Separate sample logistic discrimination »

[253]: Rosenblatt (1958), « The perceptron : a probabilistic model for information storage and organization in the brain. »

[254]: Zhu et al. (2017), « Deep Learning in Remote Sensing »

[255]: Reichstein et al. (2019), « Deep learning and process understanding for data-driven Earth system science »

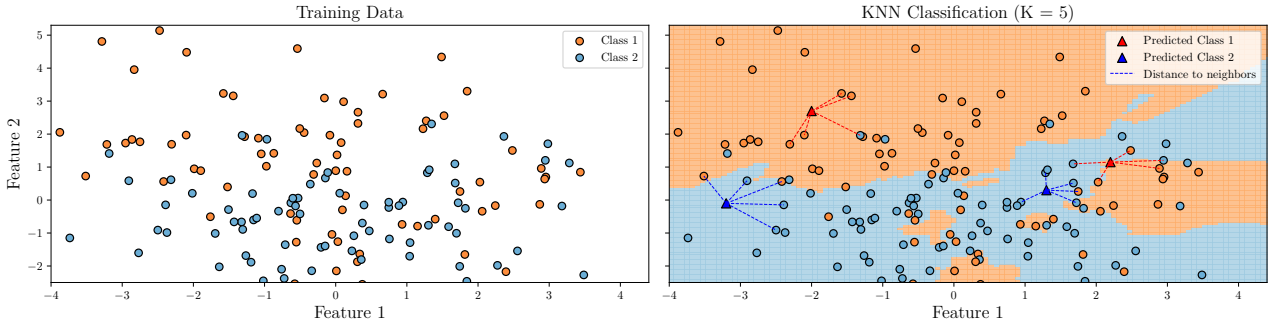



**Figure 3.2 –** Illustration de la méthode K-Nearest Neighbors en 2D (avec $K = 5$)

donne une vision de l'apprentissage profond (*deep-learning*) dans ce domaine, en expliquant un ajustement de ces réseaux en ajoutant une connaissance physique dans ces modèles. Le concept d'apprentissage profond est basé sur l'empilement de couches de neurones et d'opérateurs (linéaires et non-linéaires). Les architectures sont nombreuses et variées, mais les plus courantes sont les réseaux de neurones convolutifs (CNN) et les réseaux de neurones récurrents (RNN). Les CNN sont particulièrement adaptés pour l'analyse d'images, car ils capturent la structure spatiale des données en utilisant des opérations de convolution. Les principales tâches dans pour l'observation de la Terre sont la segmentation [11] et la classification [256]. Les RNN, quant à eux, sont utilisés pour traiter des données séquentielles, comme des séries temporelles. Les travaux de [257] donnent un aperçu de l'utilisation de l'apprentissage profond pour le SAR et [12] répertorie les nombreuses architectures pour la classification d'images SAR.

[11]: Zhang et al. (2016), « Deep Learning for Remote Sensing Data »

[256]: Bahri et al. (2020), « Remote Sensing Image Classification via Improved Cross-Entropy Loss and Transfer Learning Strategy Based on Deep Convolutional Neural Networks »

[257]: Zhu et al. (2021), « Deep Learning Meets SAR »

[12]: Parikh et al. (2020), « Classification of SAR and PolSAR images using deep learning »

**Fonctions d'activations et regroupements :** Les fonctions d'activation sont cruciales dans les réseaux de neurones, introduisant la non-linéarité qui permet aux modèles d'apprendre des relations complexes dans les données. Parmi les fonctions d'activation les plus couramment utilisées, on trouve :

- **ReLU (Rectified Linear Unit)** : Cette fonction est largement utilisée pour sa simplicité et son efficacité. Elle est définie comme :

$$f(x) = \max(0, x)$$

  Remplaçant les valeurs négatives par zéro, elle favorise la sparsité dans les activations des neurones.
- **Sigmoid** : Utilisée souvent en sortie de réseau pour des prédictions binaires, la fonction sigmoïde transforme les valeurs en un intervalle [0, 1], permettant une interprétation probabiliste des prédictions. Elle est définie par :

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

- **Softmax** : La fonction Softmax est généralement utilisée en sortie de réseau pour des problèmes de classification multiclasse. Elle normalise les valeurs de sortie en une distribution de probabilités sur les différentes classes, facilitant ainsi la prise de décision. Pour un vecteur d'entrée $x = (x_1, x_2, ..., x_n)$, la fonction Softmax est définie comme :

$$f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$$

Pour apporter plus de flexibilité, à l'instar d'amplificateur en électronique, certains ont proposé des versions paramétrisées de ces fonctions [258, 259]. Les opérations de regroupement (*pooling*) sont utilisées pour réduire la dimension des données. On retrouve ainsi les opérations de max-pooling, min-pooling et average-pooling. Dans les réseaux de neurones convolutifs, l'opération de regroupement local et spatial moyenné (*average-pooling*) correspond à une convolution avec un noyau unitaire de taille fixe (mis à l'échelle par un facteur de normalisation).

[258]: Lee et al. (2020), « Estimation of COVID-19 spread curves integrating global data and borrowing information »
[259]: Atto et al. (2021), « Parametric Rectified Power Sigmoid Units : Learning Nonlinear Neural Transfer Analytical Forms »

**Adaptation de domaines :** L'étape la plus complexe étant l'apprentissage, un axe de recherche est de transférer les connaissances d'un domaine à un autre [260]. Ceci permettant d'éviter de réapprendre des connaissances déjà acquises sur des jeux de données plus larges. L'adaptation de domaine permet également de s'affranchir dans une certaine mesure du problème de labellisation des données par des approches non-supervisées [261, 262]. Dans les réseaux profonds, le transfert d'apprentissage (*transfert-learning*) repose sur des réseaux profonds (Xception, MobileNet-V2) composés de plusieurs centaines de couches et pré-entraînés sur des bases d'images, comme ImageNet, de plus d'un million d'images et plusieurs centaines de classes. À partir de ces réseaux, le transfert se fait en modifiant les dernières couches des réseaux pour les entraîner à résoudre un autre problème spécifique. De cette manière, l'apprentissage ce fait sur quelques milliers de paramètres, bien moins que les millions de paramètres de la totalité de ces réseaux. Appliqué à tous les domaines, le transfert d'apprentissage a été utilisé plus récemment dans l'imagerie SAR haute résolution dans les travaux de [263] et [264]. Bien que cette approche soit simple, le travail de [265] dans le cadre de la détection de cible, pointe les difficultés pour transférer un réseau entraîné sur des images optiques vers des images SAR.

[260]: Tuia et al. (2016), « Domain adaptation for the classification of remote sensing data : An overview of recent advances »
[261]: Kouw et al. (2021), « A Review of Domain Adaptation without Target Labels »
[262]: Shi et al. (2021), « Unsupervised domain adaptation for SAR target detection »

[263]: Huang et al. (2021), « Classification of Large-Scale High-Resolution SAR Images With Deep Transfer Learning »
[264]: Geng et al. (2020), « Transfer Learning for SAR Image Classification Via Deep Joint Distribution Adaptation Networks »
[265]: Huang et al. (2020), « What, Where and How to Transfer in SAR Target Recognition Based on Deep CNNs »
[266]: Geng et al. (2018), « SAR Image Classification via Deep Recurrent Encoding Neural Networks »

**Labellisation :** Le cadre idéal et classique de la classification d'images par réseaux profonds est l'utilisation d'images optiques avec des classes équilibrées. Il est donc nécessaire d'adapter les techniques en classification pour qu'elles soient efficaces en SAR [266] mais aussi prendre en compte les éventuels déséquilibres de certaines classes étudiées, principalement causés par la difficulté de labelliser les images SAR. Pour pallier ce problème, des travaux ont été proposés utilisant certains biais (*contrastive*, *balanced*) pour permettre d'apprendre à partir de datasets déséquilibrés et de ne pas biaiser les résultats [267]. Une méthode est proposée dans [268] à base de *contrastive learning*[48] et de pseudo-labels[49] pour surmonter les limitations de labellisation des images SAR. Quelle que soit la méthode d'apprentissage, le déséquilibre des classes est un problème récurrent, et peut être simplement résolu par des méthodes de sous-échantillonnage, mais qui peut être contre-productif dans le cas d'apprentissage profond en raison d'une diminution drastique de la variabilité des données.

[267]: Hong et al. (2021), « Unbiased Classification through Bias-Contrastive and Bias-Balanced Learning »
[268]: Wang et al. (2022), « SAR Image Classification Using Contrastive Learning and Pseudo-Labels With Limited Data »

**Définition (48)**
*Contrastive learning:* Le contrastive learning vise à apprendre des représentations en basse dimension des données en contrastant des échantillons similaires et dissimilaires. Concrètement, il rapproche les échantillons similaires dans l'espace de représentation et éloigne les dissimilaires en utilisant une mesure de distance.

### Fonctions de coût et optimisation

Pour toutes méthodes d'apprentissage automatique, les fonctions de coût sont nécessaires pour évaluer la performance des modèles et guider l'optimisation des paramètres. Elles décrivent l'objectif à atteindre par le modèle et permettent

de quantifier l'écart entre les prédictions du modèle et les valeurs réelles des données. À partir de leur formulation et de l'architecture de l'algorithme, le processus d'apprentissage se base sur l'optimisation (minimisation) de la fonction de coût pour ajuster les paramètres du modèle et améliorer sa capacité à généraliser sur de nouvelles données. Un aperçu des fonctions de coût et de leur utilisation est donné dans [269].

Nous présentons ici deux des principales fonctions de coût utilisées dans la suite de ces travaux. Soit $\mathbf{y}$ le vecteur des valeurs réelles et $\hat{\mathbf{y}}$ le vecteur des valeurs prédites, tous les deux de dimension $n$, et $\hat{y}_i$ la valeur prédite pour l'échantillon $i$ et $y_i$ sa valeur réelle pour $i = 1, 2, ..., n$.

- **Erreur Quadratique Moyenne (MSE - Mean Squared Error)** : Cette fonction peut être utilisée pour les problèmes de régression ou de classification, calculant la moyenne des carrés des différences entre les valeurs prédites et les valeurs réelles. Elle est définie par :

$$\mathrm{MSE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{3.8}$$

  L'avantage de cette fonction est sa forme simple (et convexe), mais elle est sensible aux valeurs aberrantes. La dérivée de la fonction de coût par rapport à la prédiction est donnée par le gradient suivant :

$$\nabla_{\mathrm{MSE}} = \frac{2}{n} (\hat{\mathbf{y}} - \mathbf{y}). \tag{3.9}$$

- **Entropie Croisée (Cross Entropy)** : Adaptée aux problèmes de classification, l'entropie croisée mesure la divergence entre la distribution réelle et la distribution prédite. Elle est définie par :

$$\mathscr{H}(\mathbf{y}, \hat{\mathbf{y}}) = -\frac{1}{n} \sum_{i=1}^{n} y_i \log(\hat{y}_i). \tag{3.10}$$

  Dans le cas où le nombre de classes est limité à 2 (ou classification binaire), la fonction d'entropie croisée est équivalente à la fonction de perte logistique :

$$\mathscr{L}(\mathbf{y}, \hat{\mathbf{y}}) = -\frac{1}{n} \sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i). \tag{3.11}$$

  Cette fonction requiert que les prédictions soient des probabilités. Elle est souvent utilisée en combinaison avec la fonction d'activation softmax[50] pour les problèmes de classification multi-classes dans les réseaux de neurones profonds. La dérivée de l'entropie croisée par rapport à la prédiction est donnée par le gradient suivant :

$$\nabla_{\mathscr{H}} = -\frac{1}{n} \mathbf{y} \oslash \hat{\mathbf{y}}, \tag{3.12}$$

  avec $\oslash$ l'opérateur de division élément par élément. La dérivée de la fonction de perte logistique par rapport à la prédiction est donnée par le gradient suivant :

$$\nabla_{\mathscr{L}} = -\frac{1}{n} \left[ \mathbf{y} \oslash \hat{\mathbf{y}} - (1 - \mathbf{y}) \oslash (1 - \hat{\mathbf{y}}) \right]. \tag{3.13}$$

Il est intéressant de noter le lien de l'entropie croisée avec la divergence de

**Définition (49)**
*Pseudo labellisation:* La pseudo labellisation (*pseudo-labeling*), aussi connu sous le nom d'auto-apprentissage, est une technique d'apprentissage semi-supervisée qui combine des éléments des méthodes supervisées et non-supervisées. Elle débute avec un petit ensemble de données étiquetées utilisées pour entraîner un modèle initial. Ce modèle est ensuite utilisé pour prédire des étiquettes pour des instances non-étiquetées, les transformant ainsi en échantillons pseudo-étiquetés. Ces échantillons pseudo-étiquetés sont ajoutés à l'ensemble d'entraînement, et le modèle est ré-entraîné sur cet ensemble étendu. Ce processus itératif se poursuit, permettant au modèle d'apprendre de ses propres prédictions et de s'améliorer au fil du temps.

[269]: Wang et al. (2020), « A Comprehensive Survey of Loss Functions in Machine Learning »

**Précision (50)**
La couche softmax permet la conversion de l'activation neuronale en probabilité catégorielle (qui décrit la probabilité qu'une variable aléatoire discrète prenne l'une parmi un nombre fini de valeurs (ou catégories).)

Kullback-Leibler (KL). Soient $p$ et $q$ deux distributions de probabilité sur le même espace probabiliste, le lien entre ces deux mesures est donné par :

$$\mathcal{K}(p,q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right) = \mathcal{H}(p,q) - \mathcal{H}(p). \tag{3.14}$$

L'optimisation des fonctions de coût peut être faite par différentes méthodes. Dans le cas où la fonction décrivant le modèle est strictement convexe[51] , tout minimum local est un minimum global. Ainsi si la dérivée de cette fonction est différentiable, des méthodes analytiques peuvent être utilisées pour trouver le minimum global, comme illustré dans la figure 3.3.

**Définition (51)**
*Convexité:* Une fonction est dite convexe si la droite joignant deux points de la courbe est toujours au-dessus de la courbe.

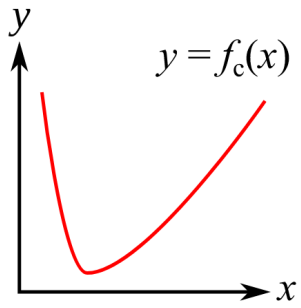



**Figure 3.3 –** Illustration d'une fonction strictement convexe différentiable, extrait de Wikimedia Commons

Cependant, dans le cas général, les fonctions de coût sont non-convexes (illustré dans la figure 3.4). Des méthodes d'optimisation itératives sont ainsi utilisées pour trouver un minimum local ou global.

Les méthodes les plus couramment utilisées sont les méthodes de descente de gradient et leurs variantes (stochastique, par mini-lots). D'autres méthodes sont également utilisées, telles que l'optimisation par essaims particulaires, bayésienne, par recuit simulé, par méthode proximale. Néanmoins dans le cadre de l'apprentissage automatique, plus particulièrement celui décrit jusqu'à présent, la descente de gradient est la méthode la plus couramment utilisée [270]. Ces méthodes sont basées sur le calcul des gradients de la fonction de coût par rapport aux paramètres du modèle, et la mise à jour itérative de ces paramètres pour minimiser la fonction de coût. Soit $\theta$ les paramètres du modèle et $\alpha$ le taux d'apprentissage et $\mathcal{L}$ la fonction de coût à minimiser. La descente de gradient est définie par :

$$\theta_{t+1} = \theta_t - \alpha \nabla_\theta \mathcal{L}(\theta) \tag{3.15}$$

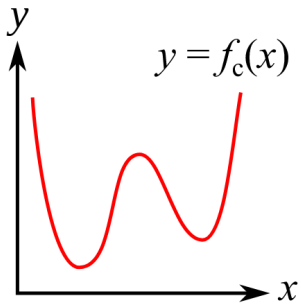



**Figure 3.4 –** Illustration d'une fonction non convexe différentiable, extrait de Wikimedia Commons

Avec $\nabla_\theta \mathcal{L}(\theta)$ le gradient de la fonction de coût par rapport aux paramètres du modèle. Le choix du taux d'apprentissage est crucial pour la convergence de l'algorithme, et plusieurs méthodes adaptatives ont été proposées pour ajuster ce taux automatiquement, telles que Adam [271]. Une autre approche consiste à optimiser ce pas en dérivant la fonction de coût par rapport au taux d'apprentissage. La rétro-propagation de l'erreur est utilisée pour calculer les gradients de la fonction de coût par rapport aux variables du modèle, en utilisant la règle de la chaîne. Cette méthode est particulièrement adaptée pour les modèles complexes, tels que les réseaux de neurones profonds, où la fonction globale du modèle est la résultante d'une composition de fonctions. Soit un modèle scalaire simple paramétré par $\theta$ et défini par $\hat{y} = f \circ g \circ h(x, \theta)$, où $h$ est la fonction d'entrée pour une donnée $x$, $g$ une fonction cachée et $f$ la fonction de sortie. En posant $\hat{y} = f(u)$, $u = g(v)$ et $v = h(x, \theta)$, la règle de la chaîne nous donne le gradient de la fonction de coût par rapport au paramètre $\theta$ du modèle :

$$\frac{\partial \hat{y}}{\partial \theta} = \left.\frac{\partial \hat{y}}{\partial u}\right|_{u=g(h(x,\theta))} \times \left.\frac{\partial u}{\partial v}\right|_{v=h(x,\theta)} \times \left.\frac{\partial v}{\partial \theta}\right|_{\theta} \tag{3.16}$$

[270]: Sun et al. (2020), « A Survey of Optimization Methods From a Machine Learning Perspective »

[271]: Kingma et al. (2014), « Adam : A Method for Stochastic Optimization »

### Agrégation de modèles

L'agrégation d'informations (données ou modèles) devient de plus en plus importante à mesure que le nombre de sources de données et de modèles continue d'augmenter.

Ces dernières années, les techniques d'agrégation de données ont suscité un intérêt croissant dans divers domaines, trouvant leur importance dans des applications telles que l'apprentissage fédéré [272], la santé [273], et la télédétection [274, 275]. Elles permettent la préservation de la confidentialité des données et la gestion efficace de différentes sources de données ou d'expertises [276]. Les recherches ont également abordé l'agrégation de données temporelles, comme illustré dans [277], ainsi que dans le domaine de la communication sans fil [278]. Côté agrégation de modèles, les méthodes d'apprentissage ensembliste, telles que les forêts aléatoires, sont particulièrement utilisées en télédétection [279], par leur capacité à améliorer la précision des modèles individuels, tout en permettant la parallélisation de l'apprentissage. De plus leur structure permet de remonter à une interprétabilité des résultats par l'utilisation de méthode de permutation afin de mesurer l'importance des variables apprises. Dans [280] une comparaison des méthodes d'apprentissage ensembliste est proposée pour différents scénarios.

Dans la classification d'images, l'agrégation peut s'appliquer soit à partir d'un type de classifieurs dont l'apprentissage est fait sur différentes caractéristiques, soit dans l'agrégation d'une variété de classifieurs dont l'apprentissage est réalisé sur les mêmes données [281]. Cette dernière approche peut être comparée au processus d'agrégation d'informations dans les réseaux neuronaux, en particulier dans les couches denses[52] ou *fully connected* des réseaux convolutifs. Récemment, les auteurs de [283] ont proposé de généraliser ce processus en utilisant des opérateurs de la moyenne généralisée. Nous pouvons noter que cette approche de généralisation est proche de celle proposée dans [284] et [285] qui cherchent à apprendre l'agrégation optimale des couches, appelée pooling[53] , dans les réseaux neuronaux convolutifs.

**Agrégation floue :** L'agrégation est particulièrement intéressante dans le domaine de la logique floue en raison de sa capacité à tenir compte des relations entre les classifieurs et à représenter divers opérateurs d'agrégation dans des décisions incertaines. Cette approche de l'agrégation d'informations n'est pas nouvelle [286] et utilise des opérateurs non linéaires tels que l'intégrale de Choquet [287]. Cette dernière reste un outil significatif dans les concepts de fusion non linéaire à partir de mesures floues.

Les travaux de [288] ont montré que l'inclusion de mesures floues dans l'apprentissage automatique s'est révélée bénéfique pour traiter des données très incertaines. L'agrégation floue en apprentissage automatique peut être abordée de deux manières. La première approche implique des entrées floues basées sur la logique floue et l'application d'opérateurs associés, ce qui conduit à la construction de systèmes d'inférence neuro-flous adaptatifs (*Adaptive Neuro-Fuzzy Inference System* ANFIS) [289]. La deuxième option se concentre sur la construction et l'utilisation d'une mesure floue dans un opérateur d'agrégation, tel que l'intégrale de Choquet. Certaines études se focalisent sur l'apprentissage de la mesure floue [290], tandis que d'autres optimisent l'intégrale de Choquet [291]. Cet axe est à mettre en parallèle avec l'apprentissage de métriques (metric learning). Dans ce cas, la mesure floue optimale est recherchée pour des tâches de classification supervisée ou non-supervisée [292], résolues à l'aide d'algorithmes méta-heuristiques [293] ou de la représentation de Möbius [290]. Cependant, ces approches sont confrontées à des défis imposés par des contraintes nécessaires pour respecter la logique floue, le nombre de contraintes augmentant de manière exponentielle avec le nombre d'entrées. Pour pallier à ces contraintes, des approches alternatives

[272]: Sah et al. (2022), « Aggregation Techniques in Federated Learning : Comprehensive Survey, Challenges and Opportunities »

[273]: Opaliński et al. (2019), « Medical data exploration based on the heterogeneous data sources aggregation system »

[274]: Dong et al. (2022), « Multiscale Context Aggregation Network for Building Change Detection Using High Resolution Remote Sensing Images »

[275]: M. D. Nelson et al. (2009), « Effects of satellite image spatial aggregation and resolution on estimates of forest land area »

[276]: Mertzani et al. (2023), « Expertise, Social Influence, and Knowledge Aggregation in Distributed Information Processing »

[277]: Böhlen et al. (2006), « Multidimensional Aggregation for Temporal Data »

[278]: Morell et al. (2016), « Data Aggregation and Principal Component Analysis in WSNs »

[279]: Lu et al. (2020), « Glacier mapping based on Random Forest algorithm : A case study over the Eastern Pamir »

[280]: Campagner et al. (2023), « Aggregation Models in Ensemble Learning »

[281]: Skublewska-Paszkowska et al. (2022), « Aggregation of Tennis Groundstrokes on the Basis of the Choquet Integral and Its Generalizations »

**Définition (52)**

*Couches denses:* Dans un réseau de neurones artificiels, couches où chaque neurone dans une couche est connecté à chaque neurone de la couche précédente et de la couche suivante. Elles sont définies par :

$$\mathbf{y} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b}),$$

où $\mathbf{W}$ est la matrice de poids, $\mathbf{x}$ le vecteur d'entrée, $\mathbf{b}$ le vecteur de biais et $\sigma$ une fonction d'activation [282].

[283]: Terziyan et al. (2022), « Hyperflexible Convolutional Neural Networks based on Generalized Lehmer and Power Means »

[284]: McFee et al. (2018), « Adaptive Pooling Operators for Weakly Labeled Sound Event Detection »

[285]: Lee et al. (2018), « Generalizing Pooling Functions in CNNs : Mixed, Gated, and Tree »

**Définition (53)**
*Pooling:* Opération d'agrégation de l'information dans un réseau de neurones, permettant de réduire la dimensionnalité des données, par moyennage, maximum ou autre opération sur une fenêtre d'analyse glissante.

ont été proposées, consistant à considérer une mesure fixe et à analyser ou à apprendre ses paramètres [294]. Cette approche hybride permet une paramétrisation spécifique aux données de la mesure floue tout en conservant les principes de la logique floue mais avec moins de contraintes. Des mesures floues telles que la mesure de puissance ou la moyenne pondérée [295] ont trouvé des applications dans la détection de contours [296] sans apprentissage de paramètres. Au lieu de cela, elles explorent différentes mesures floues et proposent une nouvelle forme d'agrégation appelée pré-agrégation, basée sur des t-normes paramétriques [297], conduisant à une généralisation de l'intégrale de Choquet.

L'utilisation de la logique floue dans la classification d'images a été explorée en télédétection, notamment avec des données multispectrales [298, 299]. Cette technique implique l'introduction d'une étape de *fuzzification* pour permettre l'apprentissage à travers des opérateurs de logique floue. L'étude de [300] présente une approche alternative avec une forêt de SVM et des sorties floues des classifieurs pour produire une décision finale.

En imagerie SAR, des techniques de filtrage des données polarimétriques ont été explorées en utilisant des opérateurs flous [301]. Des applications telles que la détection de contours dans des données multi-temporelles ont également été explorées [302]. L'utilisation de l'agrégation floue est bien adaptée à l'imagerie SAR dans un cadre de classification, exploitant des données étiquetées avec incertitude et une variété d'algorithmes offrant une expertise variée [303]. L'étude de [304] présente une approche pour la détection de navires en SAR qui agrège des informations multi-niveaux grâce à un regroupement flou de C-moyennes. De plus, dans [305] une approche de segmentation est proposée, qui agrège l'entropie et l'intensité des informations dans les images SAR à travers un système d'inférence floue basée sur des fonctions d'appartenance floues. La stratégie est mise en œuvre de manière similaire à un ANFIS pour détecter les déversements de pétrole en mer en fonction des caractéristiques extraites des images SAR [306].

À partir de l'ensemble de ces modèles ou techniques d'apprentissage il est nécessaire de construire une stratégie de gestion des données et d'évaluation en évitant les biais et les corrélations spatiales ou temporelles.

### Données et métriques d'évaluation

De nombreux paramètres entrent en jeux dans la construction de méthodes d'apprentissage automatique notamment dans le cadre de l'observation de la Terre [172]. Nous présentons ici brièvement les principaux points que nous avons pris en compte.

**Gestion des données :** En apprentissage automatique, un jeu de données peut être divisé en deux ensembles distincts : un ensemble d'entraînement et un ensemble de test. L'ensemble d'entraînement est utilisé pour entraîner le modèle, tandis que l'ensemble de test est utilisé pour évaluer les performances du modèle sur des données non vues. Il est donc essentiel que cette division soit faite judicieusement pour éviter une corrélation spatiale ou temporelle [14]. Les valeurs typiques pour cette division varient entre 70% et 90% pour l'ensemble d'entraînement et entre 10% et 30% pour l'ensemble de test. Différentes techniques [307] existent afin d'obtenir une vision exhaustive des performances des modèles sur l'ensemble des données. Les techniques de

[286]: Pedrycz (1992), « Fuzzy neural networks with reference neurons as pattern classifiers »

[287]: Campos et al. (1992), « Characterization and comparison of sugeno and choquet integrals »

[288]: Hüllermeier (2005), « Hüllermeier, E. : Fuzzy-Methods in Machine Learning and Data Mining : Status and Prospects. Fuzzy Sets and Systems 156(3), 387-407 »

[289]: Walia et al. (2015), « Anfis »

[290]: Türkarslan et al. (2022), « Measure Identification for the Choquet Integral : A Python Module »

[291]: Beliakov et al. (2021), « Choquet Integral Optimisation with Constraints and the Buoyancy Property for Fuzzy Measures »

[292]: Beliakov et al. (2011), « Learning Choquet-Integral-Based Metrics for Semisupervised Clustering »

[293]: Wang et al. (2011), « Particle swarm optimization for determining fuzzy measures from data »

[290]: Türkarslan et al. (2022), « Measure Identification for the Choquet Integral : A Python Module »

[294]: Islam et al. (2018), « Data-Driven Compression and Efficient Learning of the Choquet Integral »

[295]: Lucca et al. (2016), « Preaggregation Functions »

[296]: Marco-Detchart et al. (2021), « Neuro-inspired edge feature fusion using Choquet integrals »

[297]: Alsina et al. (2006), *Associative functions*

[298]: Du et al. (2019), « Multiple Instance Choquet Integral Classifier Fusion and Regression for Remote Sensing Applications »

[299]: Melgani et al. (2000), « An Explicit Fuzzy Supervised Classification Method for Multispectral Remote Sensing Images »

[300]: Moustakidis et al. (2012), « SVM-Based Fuzzy Decision Trees for Classification of High Spatial Resolution Remote Sensing Images »

[301]: Torres et al. (2016), « Chapter 12 - Learning OWA filters parameters for SAR imagery with multiple polarizations »

[302]: Chanussot et al. (1999), « Fuzzy fusion techniques for linear features detection in multitemporal SAR images »

K-Fold et ses dérivées (Stratified K-Fold, Group K-Fold), divisent les données en K sous-ensembles, et utilisent K-1 sous-ensembles pour l'entraînement et le dernier pour la validation. Ces techniques permettent ainsi de balayer l'ensemble des données et d'obtenir une notion de variabilité des performances des modèles. Pour les dérivées de K-Fold, la stratification [308] permet de conserver la distribution des classes dans chaque sous-ensemble, tandis que le Group K-Fold permet de conserver la distribution des groupes dans chaque sous-ensemble. Ce dernier est particulièrement intéressant pour les données temporelles ou spatiales, où l'on peut définir un groupe comme une région dans l'acquisition et éviter de voir les échantillons d'une même région dans l'ensemble d'entraînement et de test.

Un aperçu des besoins et des caractéristiques des jeux de données pour l'observation de la Terre est donnée dans [172]. Un troisième ensemble, l'ensemble de validation, est souvent utilisé pour ajuster les hyper-paramètres du modèle. Dans notre cas, l'ensemble de validation[54] est utilisé pour spatialiser les résultats ou en tirer des indicateurs temporels (approche qualitative ou comparative). Ceci est à mettre en opposition à l'ensemble de test qui, dans nos travaux, est utilisé pour extraire des métriques quantitatives sur les performances des modèles.

Dans une majorité des jeux de données, les classes sont déséquilibrées, c'est-à-dire que certaines classes sont sous-représentées par rapport à d'autres. Cela peut entraîner un biais dans l'apprentissage du modèle, qui peut être atténué en utilisant des techniques de sur-échantillonnage (SMOTE [309]) ou de sous-échantillonnage (aléatoire). Cependant ces techniques peuvent introduire un biais dans les données par la sélection réalisée. Dans ce travail, nous nous plaçons principalement dans un équilibrage par sous-échantillonnage, en raison de la taille des jeux de données. Pour évaluer de manière fiable le modèle en présence de classes déséquilibrées, nous utilisons des métriques adaptées, telles que le F1-score, la courbe ROC et l'AUC, qui prennent en compte à la fois les faux positifs et les faux négatifs.

**Métrique d'évaluation :** Pour évaluer les performances de chaque méthode, associée à des transformations ou des données auxiliaires, nous nous concentrons d'abord sur une analyse quantitative. Un grand nombre d'indicateurs existent pour caractériser chaque modèle avec $T_P, T_N$ les vrais positifs et négatifs, $F_P, F_N$ les faux positifs et négatifs et $N$ le nombre d'échantillons considérés, la précision,[55] rappel,[56] la spécificité.[57] La relation entre les taux vrai, faux positifs et négatifs est illustrée dans la matrice de confusion (table 3.2). Nous proposons d'utiliser dans la suite de ce travail quatres les métriques spécifiques :

- Exactitude (*Accuracy*) : C'est le rapport entre le nombre de prédictions correctes (vrais positifs et vrais négatifs) et le nombre total d'échantillons. Elle mesure la capacité globale du modèle à prédire correctement les classes. Sa formule est donnée par :

$$\text{Accuracy} = \frac{T_P + T_N}{N}. \tag{3.17}$$

- F1-score : Le F1-score est la moyenne harmonique de la précision et du rappel. Il est calculé en fonction des vrais positifs, des faux positifs et

[303]: Blonda et al. (2001), « Combination of multiple classifiers by fuzzy integrals »

[304]: Sun et al. (2023), « Ship Detection in SAR Images Based on Multilevel Superpixel Segmentation and Fuzzy Fusion »

[305]: Lalchhanhima et al. (2020), « Single Look SAR Image Segmentation Using Local Entropy, Median Low-Pass Filter and Fuzzy Inference System »

[306]: Lounis et al. (2020), « Adapting Neuro Fuzzy System for Oil Spills Detection From Sea Sar Images »

[172]: Schmitt et al. (2023), « There Are No Data Like More Data- Datasets for Deep Learning in Earth Observation »

[14]: Hansch et al. (2019), « The Truth About Ground Truth »

[307]: Kohavi (1995), « A study of cross-validation and bootstrap for accuracy estimation and model selection »

[308]: Sechidis et al. (2011), « On the stratification of multi-label data »

[172]: Schmitt et al. (2023), « There Are No Data Like More Data- Datasets for Deep Learning in Earth Observation »

**Précision (54)**
Il est intéressant de noter que dans le cas de l'étude de la neige, [36] indique que près de 60% des études proposent une validation inexistante.

[309]: Fernández et al. (2018), « SMOTE for learning from imbalanced data : progress and challenges, marking the 15-year anniversary »

**Table 3.2 –** Matrice de confusion

| | | Vérité | |
|---|---|---|---|
| | | positif | négatif |
| Prédiction | positif | $T_P$ | $F_P$ |
| | négatif | $F_N$ | $T_N$ |

**Définition (55)**
*Précision:* Le taux de précision est défini par : $\frac{T_P}{T_P + F_P}$

**Définition (56)**
*Rappel:* Le taux de rappel est défini par : $\frac{T_P}{T_P + F_N}$

**Définition (57)**
*Spécificité:* Le taux de spécificité est défini par : $\frac{T_N}{T_N + F_P}$

des faux négatifs. Sa formule est donnée par :

$$\text{F1-score} = \frac{2 \times T_P}{2 \times T_P + F_N + F_P}. \tag{3.18}$$

L'avantage du F1-score par rapport à l'accuracy est qu'il prend en compte à la fois les faux positifs et les faux négatifs, ce qui en fait une mesure plus représentative des performances lorsque les classes sont déséquilibrées.

- Kappa : Le coefficient Kappa [310] est une mesure de la concordance entre les classifications observées et les classifications attendues par le hasard. Il tient compte à la fois de l'accord entre les classes observées et attendues, ainsi que de la probabilité de concordance due au hasard. Une valeur élevée de kappa indique un accord élevé au-delà de ce qui pourrait être dû au hasard.
- Courbes Receiver Operating Characteristic (ROC) : Les courbes ROC sont des représentations graphiques de la performance d'un modèle de classification à différents seuils de classification. Une illustration de la courbe ROC est donnée dans la figure 3.5. Elles représentent le taux de vrais positifs (sensibilité ou probabilité de détection) en fonction du taux de faux positifs (1 - spécificité ou probabilité de fausse alarme). Les courbes ROC sont souvent utilisées pour évaluer et comparer les performances des modèles de classification, en particulier lorsqu'il y a un déséquilibre entre les classes. L'aire sous la courbe ROC (*Area Under the ROC Curve* AUC) est une mesure de la performance globale du modèle, où une valeur de 1 indique une performance parfaite et une valeur de 0.5 indique une performance aléatoire sous hypothèse de classes équilibrées (illustrée par la courbe en pointillés dans la figure 3.5).

[310]: Landis et al. (1977), « The measurement of observer agreement for categorical data. »

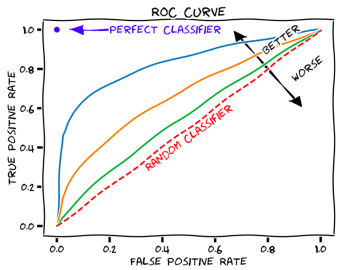



**Figure 3.5 –** Illustration du fonctionnement des courbes ROC. Tiré de WikiMedia

En outre, pour une évaluation plus détaillée, les matrices de confusion (figure 3.6) peuvent être utilisées. Ces matrices fournissent un résumé des résultats de classification à différents seuils de classification. Elles montrent le nombre de vrais positifs ($T_P$), de vrais négatifs ($T_N$), de faux positifs ($F_P$) et de faux négatifs ($F_N$). Les matrices de confusion offrent une vue plus granulaire des performances du modèle, ce qui peut être utile pour diagnostiquer les erreurs de classification spécifiques. Pour accéder à une information résumé d'une classe, un calcul un contre tous (*One vs All*) peut être réalisé.

Pour des données multi-classes, les métriques précédentes peuvent être étendues à des cas plus complexes. Par exemple, le F1-score peut être généralisé à une moyenne pondérée par le nombre d'échantillons de chaque classe, appelée F1-weighted.

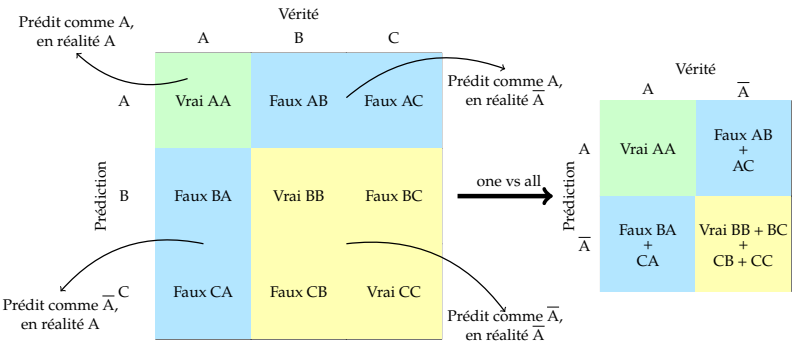



**Figure 3.6 –** Illustration de la matrice de confusion pour un problème de classification à trois classes.

**Application à la cryosphère**

Les méthodes d'apprentissage automatique ont été largement utilisées pour l'analyse des données de télédétection de la cryosphère, par imagerie SAR ou optique. Les algorithmes de forêt aléatoire sont particulièrement prisés dans l'étude et la classification de zones glaciaires [311] ou diverses caractéristiques des glaciers [10, 120, 312], par leur capacité à gérer des données hétérogènes et à fournir des résultats interprétables. Ainsi [90] utilise à la fois l'imagerie SAR et optique pour créer des cartes de neige humide, et visualise les caractéristiques topographiques d'intérêt par analyse de variables apprises. Les méthodes SVM sont également utilisées, que ce soit pour la classification de débris glaciaires [313] ou la classification de neige humide dans [119] ou [73], mais pour ce dernier, en affinant les résultats par un champ de Markov. La combinaison optique et SAR a été utilisée pour la détection de débris glaciaires avec une comparaison exhaustive de près de sept algorithmes [314]. L'estimation d'indicateurs glaciaires (superficie et volume) a été étudiée par KNN [315], même si nous retrouvons assez peu de travaux utilisant cette méthode, par la contrainte de construction d'une matrice de distance (coûts en temps et en mémoire importants pour traiter de larges zones). Les réseaux de neurones artificiels sont plus récemment utilisés pour l'estimation de l'épaisseur de glaciers [316], ou l'extraction d'informations physiques sur le manteau neigeux (hauteur, densité) [317]. Nous pouvons remarquer l'intérêt de ces approches par leur capacité à traiter des données hétérogènes, en ajoutant des caractéristiques supplémentaires pour la couverture de neige [318]. De plus, la grande flexibilité des réseaux de neurones artificiels permet de remonter à des informations complexes. Néanmoins, ils exigent à la fois une grande quantité de données pour l'apprentissage et une bonne labellisation de ces données.

Dans le cadre de l'imagerie SAR, de nombreux challenges résistent encore et toujours à la communauté. Les approches pour réaliser les tâches de classification, segmentation des images SAR sont variées, mais comme le souligne [170], *"les méthodes de classification standard qui ont été validées pour les données optiques ne donnent pas de résultats satisfaisants lorsqu'elles sont appliquées directement aux images SAR."*.

### 3.1.4 Méthodes d'apprentissage temporel

Les séries temporelles d'images SAR, largement accessibles grâce à des missions telles que Sentinel-1, offrent une opportunité précieuse pour exploiter la temporalité. La richesse apportée par cette dimension permet le suivi des transformations spatio-temporelles des territoires. Les séries temporelles sont souvent exploitées pour la détection de changement, notamment dans le contexte de la surveillance urbaine [319], de la déforestation [320] et du suivi des cultures [321] ou encore de la glace de mer [322]. Ces séries temporelles sont particulièrement adaptées à la surveillance des zones glaciaires [323] et au suivi de leur dynamique [168].

Un des principaux défis lié à l'imagerie SAR réside dans la gestion de la présence du bruit multiplicatif : le speckle. De nombreux travaux font usage de la décomposition en ondelettes pour filtrer le speckle [324, 325], puis utilisent cette méthode pour mener des études sur les changements [180, 326]. Une autre difficulté réside dans la capacité des modèles à évaluer les divers types de changements, que ce soit par regroupement ou classification [327]. En effet, ces changements, de natures diverses, peuvent être spatiaux, temporels ou les

[311]: Senf et al. (2012), « Using MODIS time series and random forests classification for mapping land use in South-East Asia »

[10]: Bouchayer et al. (2022), « A Machine Learning Framework to Automate the Classification of Surge-Type Glaciers in Svalbard »

[120]: Brenning (2009), « Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection »

[312]: Zhang et al. (2019), « Glacier Facies Mapping Using a Machine-Learning Algorithm : The Parlung Zangbo Basin Case Study »

[90]: Tsai et al. (2019), « Wet and Dry Snow Detection Using Sentinel-1 SAR Data for Mountainous Areas with a Machine Learning Technique »

[313]: Khan et al. (2020), « Machine-Learning Algorithms for Mapping Debris-Covered Glaciers : The Hunza Basin Case Study »

[119]: He et al. (2017), « Dry and Wet Snow Cover Mapping in Mountain Areas Using SAR and Optical Remote Sensing Data »

[73]: Liu et al. (2023), « Identifying Wet and Dry Snow With Dual-Polarized C-Band SAR Data Based on Markov Random Field Model »

[314]: Alifu et al. (2020), « Machine-learning classification of debris-covered glaciers using a combination of Sentinel-1/-2 (SAR/optical), Landsat 8 (thermal) and digital elevation data »

[315]: Ambinakudige et al. (2022), « Estimation of area and volume change in the glaciers of the Columbia Icefield, Canada using machine learning algorithms and Landsat images »

[316]: Haq et al. (2021), « Efficiency of artificial neural networks for glacier ice-thickness estimation : a case study in western Himalaya, India »

[317]: Palermo et al. (2023), « Using Artificial Neural Networks to Couple Satellite C-Band Synthetic Aperture Radar Interferometry and Alpine3D Numerical Model for the Estimation of Snow Cover Extent, Height, and Density »

[318]: Nijhawan et al. (2019), « A hybrid of deep learning and hand-crafted features based approach for snow cover mapping »

[170]: Voisin et al. (2013), « Classification of Very High Resolution SAR Images of Urban Areas Using Copulas and Texture in a Hierarchical Markov Random Field Model »

[319]: Ban et al. (2012), « Multitemporal Spaceborne SAR Data for Urban Change Detection in China »

deux, et nécessitent une vision multi-échelle pour leurs réponses respectives lors du traitement d'une série d'images satellite couvrant de vastes zones [328]. Le travail de [240] propose une méthode de détection non-supervisée basée sur la décomposition en ondelettes des images, fournissant des niveaux d'information de changement multi-temporels. Cette méthode, appelée Wavelet Energy Correlation Screening (WECS), repose sur la corrélation entre les coefficients d'ondelettes de la série temporelle décomposée. Les auteurs démontrent que WECS surpasse les méthodes de l'état de l'art, telles que le seuillage de la différence absolue agrégée ou le *K*-Means Clustering sur la décomposition en Analyse en Composantes Principales [326], lorsque la corrélation saisonnière affecte la corrélation locale d'évolution spatiale.

Certains auteurs ont intégré l'utilisation de séries temporelles SAR en conjonction avec la méthode d'apprentissage automatique KNN pour des tâches de classification [329]. Cependant, la principale difficulté de ces approches réside à la fois dans la complexité de ces méthodes, qui peuvent être difficiles à interpréter, et dans l'irrégularité temporelle des données. Cette irrégularité peut être due à l'utilisation de capteurs avec des dates d'acquisition différentes, à la présence d'acquisitions commerciales irrégulières, et à celles soumises à des projets ou des demandes. Pour réaliser la classification de signaux temporels, la mesure de similarité DTW (Dynamic Time Warping) apporte une solution intéressante. Initialement proposé par [330], le DTW est une mesure de similarité entre deux séries temporelles qui vise à les aligner. Cette mesure est construite en minimisant un coût de déformation basé sur une distance euclidienne entre deux séries temporelles. Cette mesure est invariante par rapport au décalage et à l'échelle du temps. Cependant, il faut noter que cette mesure n'est pas une distance car elle ne satisfait pas l'inégalité triangulaire. Elle est largement utilisée dans des applications aussi variées que la finance [331] ou la médecine [332]. Soit $X$ et $Y$ deux séries temporelles de même longueur $n$. Nous définissons la matrice de coût $C$ de taille $n \times n$ comme suit :

$$C_{i,j} = (X_i - Y_j)^2 \tag{3.19}$$

La distance DTW entre $X$ et $Y$ est alors définie comme suit :

$$DTW(X,Y) = \min_{\pi} \sqrt{\sum_{(i,j)\in\pi} C_{i,j}} \tag{3.20}$$

Avec $\pi$ un chemin d'alignement entre $X$ et $Y$ qui satisfait les contraintes données par [330] [333]. Nous pouvons affiner la formulation en ajoutant une contrainte de rayon ou de largeur proposée par Sakoe-Chiba [334]. Cette contrainte limite la recherche de la distance minimale à un voisinage de taille $w$ autour du point courant, par exemple entre $Y_{i-w}$ et $Y_{i+w}$. Une illustration est donnée dans la figure 3.7 avec en comparaison la distance euclidienne. Cette dernière est définie par :

$$Euc(X,Y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} \tag{3.21}$$

[320]: Ygorra et al. (2021), « Deforestation Monitoring Using Sentinel-1 SAR Images in Humid Tropical Areas »

[321]: Silva-Perez et al. (2021), « Multitemporal Polarimetric SAR Change Detection for Crop Monitoring and Crop Type Classification »

[322]: Gao et al. (2019), « Sea Ice Change Detection in SAR Images Based on Convolutional-Wavelet Neural Networks »

[323]: Winsvold et al. (2018), « Using SAR satellite data time series for regional glacier mapping »

[168]: Arigony-Neto et al. (2007), « A time series of SAR data for monitoring changes in boundaries of glacier zones on the Antarctic Peninsula »

[324]: Gupta et al. (2007), « Despeckle and geographical feature extraction in SAR images by wavelet transform »

[325]: Atto et al. (2016), « Wavelet Operators and Multiplicative Observation Models–Application to SAR Image Time-Series Analysis »

[180]: Mian et al. (2019), « New Robust Statistics for Change Detection in Time Series of Multivariate SAR Images »

[326]: Celik et al. (2010), « Unsupervised Change Detection for Satellite Images Using Dual-Tree Complex Wavelet Transform »

[327]: Jia et al. (2016), « SAR Image Change Detection Based on Multiple Kernel K-Means Clustering With Local-Neighborhood Information »

[328]: Moser et al. (2011), « Multiscale Unsupervised Change Detection on Optical Images by Markov Random Fields and Wavelets »

[240]: Fonseca et al. (2023), « Wavelet Spatio-Temporal Change Detection on Multitemporal SAR Images »

[326]: Celik et al. (2010), « Unsupervised Change Detection for Satellite Images Using Dual-Tree Complex Wavelet Transform »

[329]: Xiao et al. (2021), « Temporal Series Crop Classification Study in Rural China Based on Sentinel-1 SAR Data »

[330]: Sakoe et al. (1971), « A Dynamic Programming Approach to Continuous Speech Recognition »

[331]: Li et al. (2020), « Dynamic Time Warping Application for Financial Pattern Recognition »

[332]: Hebbrecht et al. (2020), « Understanding personalized dynamics to inform precision medicine : a dynamic time warp analysis of 255 depressed inpatients »

[330]: Sakoe et al. (1971), « A Dynamic Programming Approach to Continuous Speech Recognition »

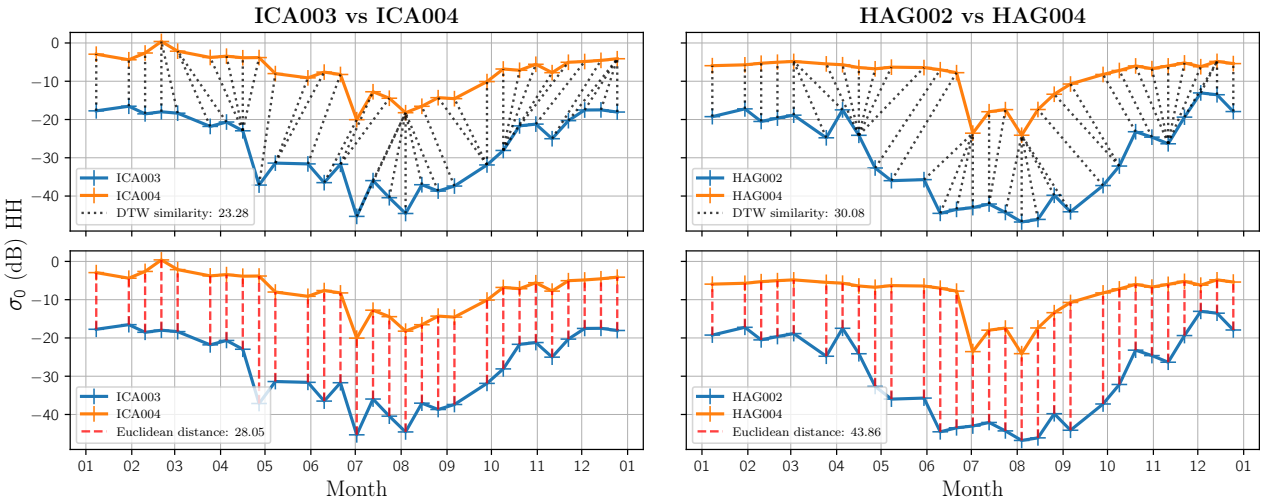



**Figure 3.7 –** Comparaison de la distance entre deux séries temporelles avec la distance euclidienne et la similarité DTW. Les comparaisons sont réalisées sur les séries temporelles de $\sigma_0$ en dB entre deux zones ice aperon (ICA) et deux zones de hanging glacier (HAG) en 2020 (cf table A.3).

À partir de la mesure de DTW, les auteurs de [335] propose un moyennage temporel respectant les retards et les caractéristiques propres à chaque signal DBA (Dynamic Time Warping Barycenter Averaging). De nombreuses évolutions ont été proposées notamment dans [336] où une approche du DTW par réseau de neurones est proposée, améliorant l'extraction des caractéristiques temporelles.

[333]: Tavenard (2021), *An introduction to Dynamic Time Warping*

[334]: Sakoe et al. (1978), « Dynamic programming algorithm optimization for spoken word recognition »

[335]: Petitjean et al. (2011), « A global averaging method for dynamic time warping, with applications to clustering »

[336]: Cai et al. (2019), « DTWNet : a Dynamic Time Warping Network »

## 3.2 Jeu de données SAR en bande C pour l'étude de la neige humide

Le jeu de données LSD4WSD (Learning SAR Dataset for Wet Snow Detection) a été créé pour permettre l'utilisation et le développement de méthodes d'apprentissage automatique pour la caractérisation de la neige à large échelle spatiale. Il se décline en deux versions : la version complète (LSD4WSD VX [337]), avec plus de données et d'informations associées, et une version plus légère (LSD4WSD V1 [338]) pour une utilisation dédiée à la détection de neige humide. Dans un premier temps, nous présentons les données communes utilisées pour la création de ces jeux de données, puis nous examinons en détail les caractéristiques et les choix respectifs de création des deux datasets.

[337]: Gallet et al. (2023), *LSD4WSD : An Open Dataset for Wet Snow Detection with SAR Data and Physical Labelling*

### 3.2.1 Zone d'étude

L'étude est conduite au sein d'un segment de la tuile 31TGL, un découpage résultant du système de tuile de Sentinel-2 basé sur le Système de Référence de Grille Militaire (MGRS) afin de faciliter la gestion des données sur l'ensemble du globe. Cette tuile, couvrant une superficie de $110 \times 110$ km$^2$, est située dans les Alpes françaises du Nord. Sept massifs alpins français sont complètement contenus dans cette tuile : Beaufortain, Belledonne, Chartreuse, Maurienne, Vanoise, Bauges et Grandes Rousses. Ces massifs font partie de la sélection réduite pour le jeu de données LSD4WSD V1. Afin d'ajouter de la diversité topographique, nous avons aussi considéré 6 autres massifs dans le cadre de la version étendue du dataset. Ces massifs sont partiellement contenus dans

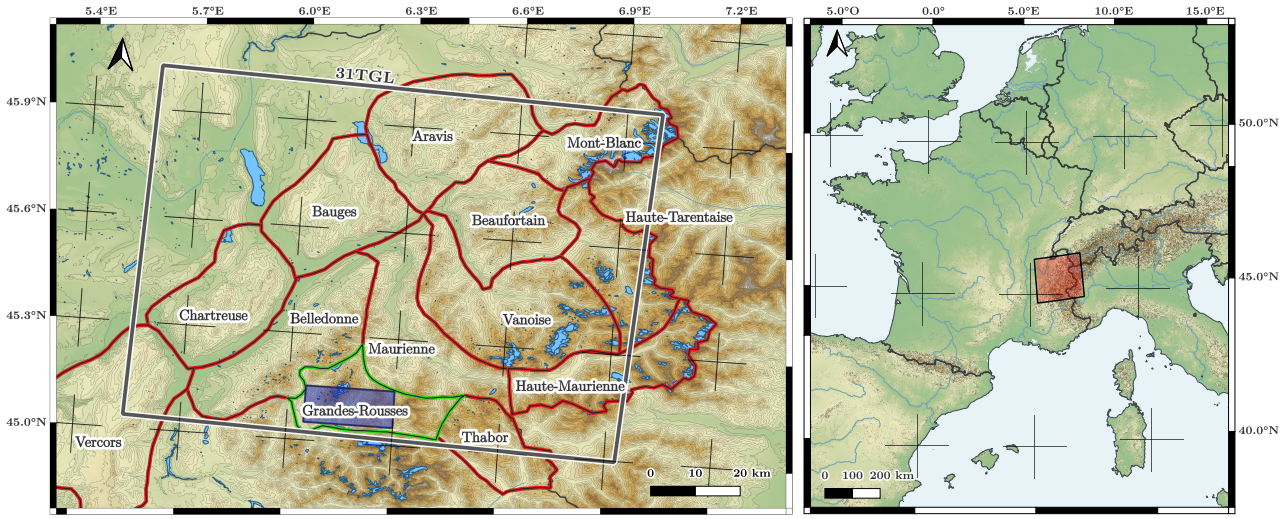



**Figure 3.8 –** Visualisation de l'empreinte au sol de l'acquisition du satellite Sentinel-1 (tuile 31TGL) en orbite descendante et les massifs alpins français utilisés. Le massif de validation est en vert, le rectangle bleu foncé représente l'emprise des cartes de résultats. Les zones d'eau sont en bleu (lacs et glaciers principaux référencés par l'IGN).

la tuile : Aravis, Haute-Maurienne, Haute-Tarentaise, Mont-Blanc, Thabor et Vercors. La position et la répartition géographique de ces massifs sont illustrées dans la figure 3.8.

### 3.2.2 Données satellitaire

**Données SAR :** Nous utilisons des données GRD double polarisation (VV et VH) obtenues, sur l'orbite ascendante numéro 161 de Sentinel-1, désignée par A161, et l'orbite descendante numéro 139 (D139). Les orbites ascendantes et descendantes ont respectivement une heure de passage local de 17h30 UTC et 5h30 UTC. L'angle d'incidence est compris entre 26° et 44°. Les images résultantes ont subi un processus multi-vues (par une fenêtre de taille $5 \times 1$) afin de produire des pixels carrés de 10m de résolution. Les acquisitions utilisent Sentinel-1A et 1B pour une résolution temporelle de 6 jours.

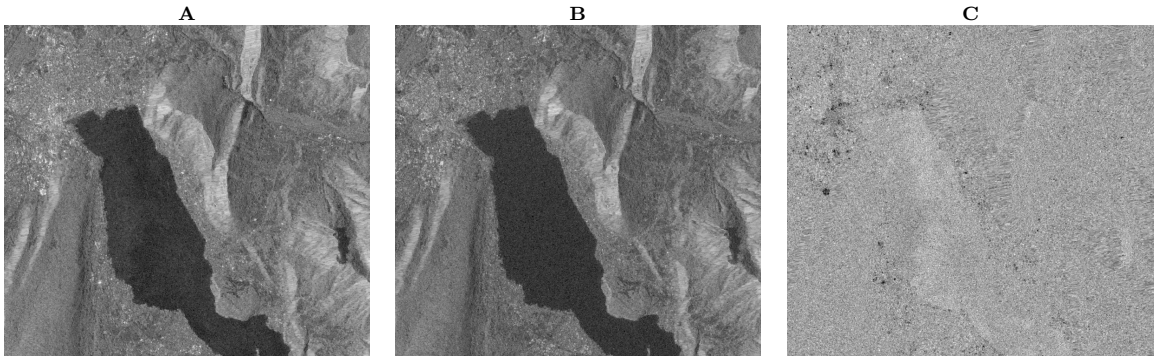



**Figure 3.9 –** Exemple d'une image SAR, centrée sur le lac d'Annecy (cf figure 3.10 ), avec les canaux VV (**A**), VH (**B**) et le rapport VV/VH (**C**) en dB.

Nous avons ainsi une série de 70 images en passe ascendante, entre le 4 juillet 2020 et le 22 août 2021 et une série de 70 images en passe descendante, entre le 3 juillet 2020 et le 27 août 2021. Ces images n'ont pas subi de filtrage supplémentaire et sont de $10980 \times 10980$ pixels. Chaque image dispose de deux canaux polarimétriques, VV et VH, et d'un troisième canal correspondant au ratio des deux autres : VV/VH. Un exemple de ces images est donné dans la figure 3.9, où nous pouvons observer un signal plus fort en polarimétrie VV dans les zones urbaines, qui se détache notamment dans le résultat du ratio VV/VH.

Les distorsions géométriques (recouvrement, ombres et zones de compression) sont inhérentes à l'imagerie SAR. Pour éviter de sélectionner les zones de distorsions irréversibles dans le jeu de données, nous utilisons un masque binaire, préalablement calculé à partir de l'angle d'incidence local et de la topographie, mappant les zones de distorsion.

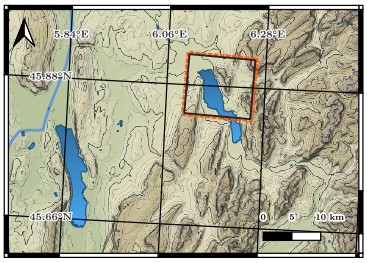



**Figure 3.10 –** Situation géographique de la zone d'illustration de l'image SAR, centrée sur le lac d'Annecy.

**Données topographiques :** Afin de compléter les images SAR, et de permettre l'étude de l'impact d'autres données par les algorithmes proposés dans la suite, nous ajoutons des informations dérivées de données topographiques. Nous utilisons un Modèle Numérique de Terrain (MNT ou *Digital Elevation Model* DEM) [58] de résolution 10m fourni par l'Institut National de l'Information Géographique et Forestière français (IGN).

Nous avons ainsi accès en tout point de la tuile 31TGL à une information d'altitude. De l'information d'élévation, nous dérivons les caractéristiques de pente et d'orientation. Une illustration sur la même zone que pour les acquisitions SAR est donnée dans la figure 3.12.




**Figure 3.11 –** Illustration de la différence entre le MNT et le MNS. En bleu, le MNT et en rouge le MNS. Extrait de [339]

**Définition (58)**
*Modèle Numérique de Terrain:* Il s'agit d'une représentation numérique de la surface terrestre qui indique l'élévation ou la hauteur de chaque point par rapport à un niveau de référence, généralement le niveau moyen de la mer. Le Modèle Numérique de Surface (MNS), est similaire au MNT, mais il inclut non seulement le relief naturel de la surface terrestre, mais également les éléments artificiels. Une illustration est donnée figure 3.11

**Caractéristique de Nagler :** Dans l'objectif de l'étude et de la caractérisation de la neige humide, nous avons montré dans la section 2.2.3, l'intérêt de la méthode de Nagler pour la caractérisation de la neige humide. Nous avons donc ajouté cette caractéristique, en nous basant sur deux références temporelles, le 8 août 2020 et le 9 août 2020, respectivement pour les passes descendantes et ascendantes. Le ratio est calculé pour toutes les acquisitions et pour les 3 canaux polarimétriques VV, VH et VV/VH.

Une illustration de cette caractéristique pour le 31 mars 2021 est donnée dans la figure 3.13. La dynamique des images est comprise entre -7 et 7 dB. Les zones de distorsions géométriques irréversibles sont mises en évidence en

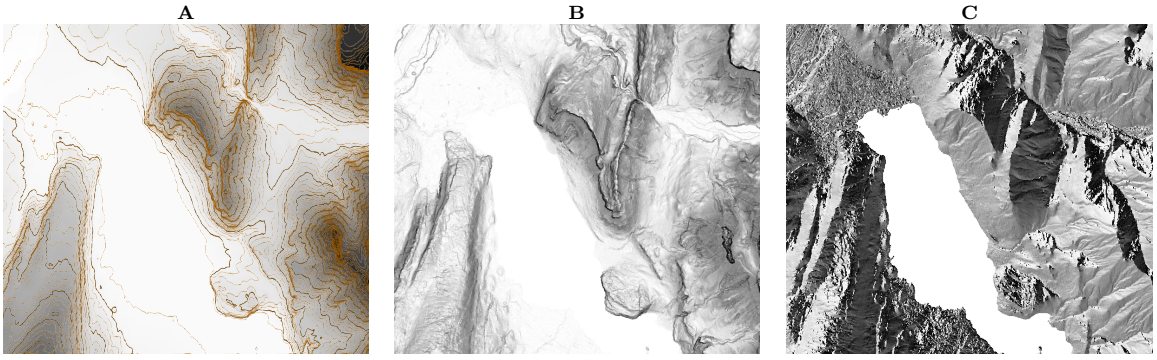



**Figure 3.12 –** Illustration des données topographiques sur la zone du lac d'Annecy. **A.** Modèle Numérique de Terrain (MNT) de l'IGN. **B.** Pente dérivée du MNT. **C.** Orientation dérivée du MNT.

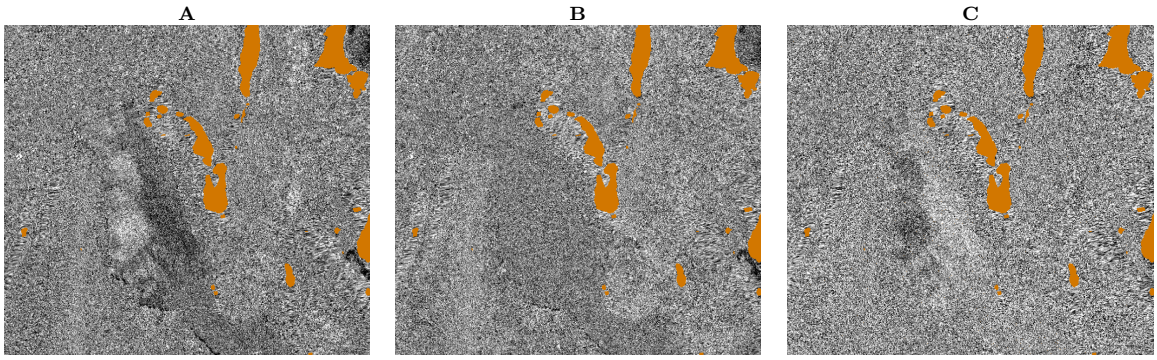



**Figure 3.13 –** Illustration de la caractéristique de Nagler (en dB) en passe ascendante pour le 31 mars 2021, avec une acquisition de référence du 9 août 2020. **A.** $VV/VV_{ref}$, **B.** $VH/VH_{ref}$, **C.** $VVVH_{ref}/VHVV_{ref}$. En orange, les zones de distorsions géométriques irréversibles.

orange. Nous pouvons observer trois zones d'intérêt. La première sur le lac, où le signal est faible dans le canal VV, indiquant selon le critère de Nagler une présence éventuelle de neige humide. Cette zone n'apparaît pas dans le canal VH, démontrant l'utilité de l'information conjointe VV, VH. Les deux autres zones se situent en bordure droite de l'image, où l'on peut clairement voir une zone plus sombre sur les deux canaux polarimétriques. Les zones enneigées pour cette date sont données dans la figure 3.14 par le produit optique de fraction de couverture neigeuse (*Fractionnal Snow Cover* FSC) [59] de Copernicus. Il est intéressant de noter que la zone de neige humide détectée par la caractéristique de Nagler est en accord avec une présence de neige dans le produit FSC.

**Définition (59)**
*FSC:* Le FSC représente la fraction (entre 0% et 100%) de neige sur un pixel de résolution 20m au niveau de la canopé



**Figure 3.14 –** Illustration du produit FSC pour le 30 mars 2021, sur la zone du lac d'Annecy (cf figure 3.10 et figure 3.13).

### 3.2.3 LSD4WSD V1

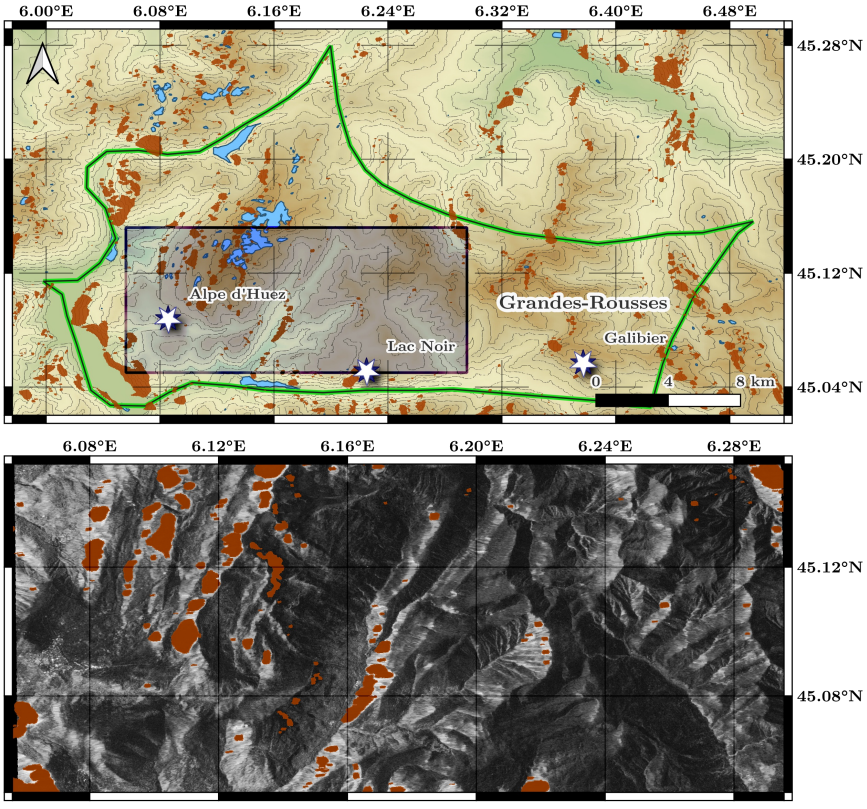



**Figure 3.15 –** Visualisation de la zone de test (en vert), de validation (rectangle bleu foncé) et des stations de mesures comparatives (étoiles) sur le massif des Grandes Rousses. En bas, span ($\sqrt{VV^2 + VH^2}$) de la zone de validation avec en orange les zones de distorsions géométriques.

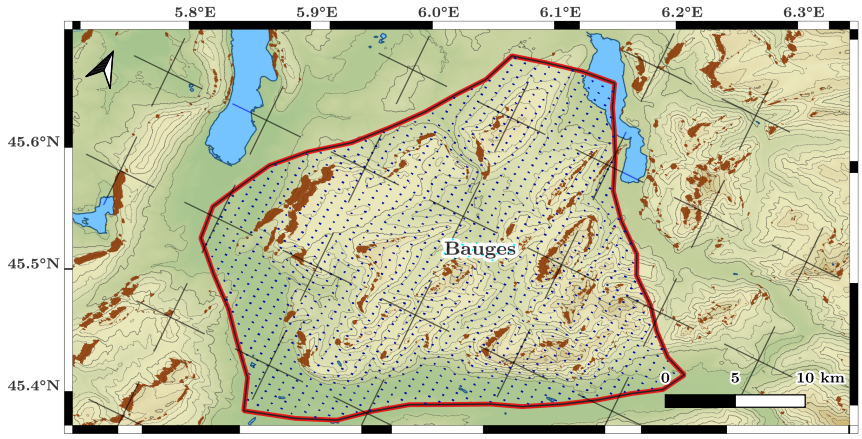



**Figure 3.16 –** Détail du processus de sélection des échantillons (ou pavage) dans le massif des Bauges, en marron, les masques de distorsion géométrique irréversibles.

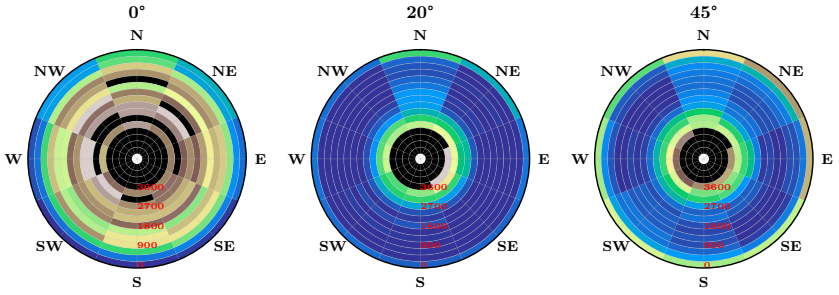
Pour une date d'acquisition donnée, nous utilisons les 2 canaux polarimétriques de Sentinel-1 et leur rapport (VV/VH) provenant de l'orbite ascendante, 3 canaux topographiques (élévation, orientation et pente) et les trois canaux de caractéristiques de Nagler. Ces 3 canaux supplémentaires permettent d'éviter l'étape de calcul non-linéaire que certains algorithmes, tels que ceux basés sur les arbres de décision, ne peuvent effectuer, mais aussi d'étudier les décisions prises par l'algorithme en présence de ces 3 canaux uniquement par rapport au seuil donné par [143]. Tous les canaux sont mis à l'échelle pour avoir une résolution de pixel de 10m. Les pixels des canaux polarimétriques et de rapport sont transformés en dB. Toutes les dates sont utilisées dans la création de cette base de données. La séparation du jeu de données d'entraînement et de test est purement spatiale et permet d'éviter de biaiser le classifieur par des données de test fortement corrélées spatialement aux données d'entraînement [14]. La chaîne de création du jeu de données d'entraînement et de test est basée sur 9 canaux.

[14]: Hansch et al. (2019), « The Truth About Ground Truth »

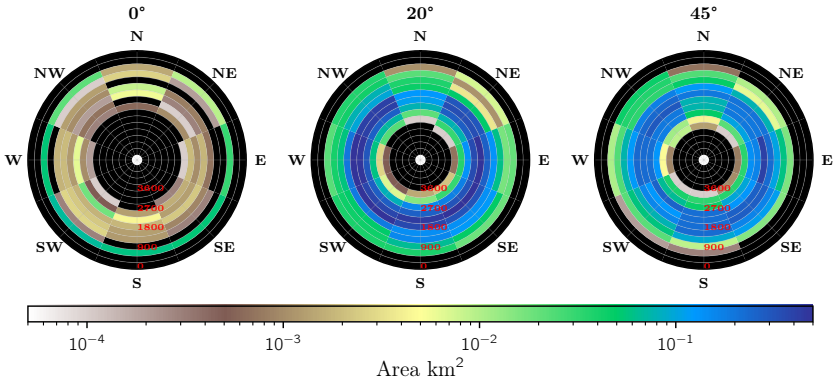
### Sélection des zones

Nous divisons les données en trois catégories. Les données d'entraînement utilisées pour l'apprentissage, les données de test nécessaires pour l'évaluation quantitative du modèle (figure 3.8) et enfin la zone de validation utilisée pour l'évaluation qualitative des méthodes d'un point de vue spatial et temporel et comparaison avec les produits Copernicus existants, illustrée figure 3.15. La délimitation en vert dans la figure donne l'emprise dans laquelle les échantillons de test sont sélectionnés, tandis que la délimitation en bleu foncé donne l'emprise sur laquelle les résultats sont cartographiés à une date donnée. Les trois étoiles correspondent aux stations de mesure de Météo France, exploitables aux dates considérées, permettant la comparaison temporelle avec les résultats.

Nous avons sélectionné les sept massifs qui étaient complètement contenus dans la tuile 31TGL pour la création du jeu de données LSD4WSD V1. 6 massifs pour construire le jeu de données d'entraînement (Beaufortain, Belledonne, Chartreuse, Maurienne et Vanoise) et 1 massif pour le test et la validation (Grandes Rousses). La sélection de 6 massifs pour entraîner le modèle permet au modèle d'être robuste aux caractéristiques intrinsèques de chaque massif, sans recourir à des techniques d'adaptation de domaine.




**(a)** Distribution des zones d'entraînement pour 3 pentes




**(b)** Distribution des zones de test pour 3 pentes

**Figure 3.17** – Distributions topographiques des échantillons de test et d'entraînement selon la discrétisation donnée par CROCUS, pour la pente, l'orientation et l'élévation

Nous découpons chaque massif en utilisant une fenêtre de $16 \times 16$, et sélectionnons une fenêtre sur 5 comme illustré dans la figure 3.16. La taille de la fenêtre de $16 \times 16$ pixels a été choisie empiriquement par une étude de validation croisée expérimentale. Elle permet d'avoir assez d'échantillons dans le patch pour être statistiquement significatif, sa taille, étant un multiple de 2, offre des perspectives sur l'utilisation d'extractions ou de décompositions de caractéristiques possibles (Transformée en ondelettes discrète (*Discrete Wavelet Transform* DWT), CNN). Ainsi, nous conservons l'étendue spatiale et la diversité de la base de données, mais nous réduisons le nombre d'échantillons pour réduire le temps de calcul et réduire la corrélation spatiale entre chaque échantillon. Ce pavage représente respectivement 1.4% et 1.1% de la surface d'entraînement et de test des massifs. Les données de validation utilisent le jeu de données de test avec une découpe similaire à celle utilisée pour le jeu de données de test mais en utilisant une fenêtre glissante sur l'ensemble de la zone. La décision étant rendue sur le pixel central, soit dans ce cas avec un décalage d'un demi pixel.

Le choix du massif de test résulte de la comparaison faite dans la figure 3.17. Cette dernière illustre la distribution des zones en km$^2$ pour les 3 paramètres topographiques du modèle CROCUS (élévation, orientation et pente). Nous notons l'intérêt d'utiliser 6 massifs pour l'entraînement par la variété des caractéristiques topographiques qu'ils représentent ensemble (figure 3.17a). Nous notons que le massif de test des Grandes Rousses est homogène en termes de distribution de ces zones pour toutes les orientations et principalement pour les pentes moyennes à fortes (Figure 3.17b). Cependant, pour les faibles

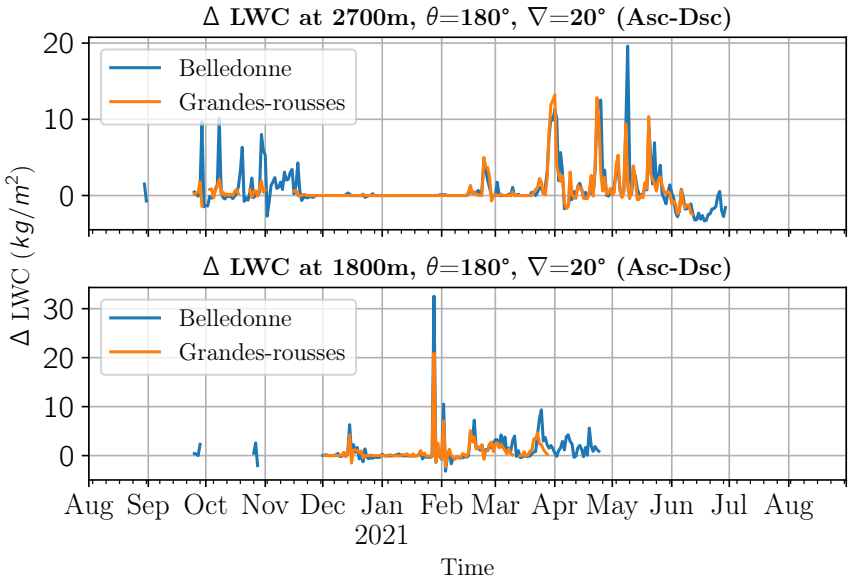



**Figure 3.18 –** Différence temporelle de la teneur en eau liquide (TEL) pour le modèle CROCUS entre les données simulées à 6h et 18h pour deux massifs alpins (Belledonne et Grandes Rousses) à différentes altitudes (1800 et 2700 mètres) sur la période d'août 2020 à août 2021.

pentes (de l'ordre de 0°), certaines zones n'existent pas dans le jeu de données d'entraînement mais apparaissent dans le jeu de données de test. L'impact de cet écart est mesuré par la taille moyenne de ces zones (de l'ordre de $10^{-4}$ km$^2$), ce qui est bien plus petit que pour des pentes plus grandes (de l'ordre de $10^{-2}$ km$^2$). Étant donné la Figure 3.17a, le jeu de données d'entraînement est également homogène avec environ 10 fois plus de surfaces pour chaque caractéristique topographique étudiée, que la zone de test sélectionnée.

### Labellisation des données

Cependant, le principal obstacle au développement et à la généralisation des méthodes d'apprentissages automatiques sur de grandes zones est l'étiquetage des données requis pour entraîner les modèles. En outre, la nature spatio-temporelle des données nécessite une plus grande attention dans le processus de création et de sélection des zones d'entraînement et de test pour éviter des corrélations spatiales ou temporelles [340].

[340]: Hänsch (2021), « The Trap Of Random Sampling and How to Avoid It - Alternative Sampling Strategies for a Realistic Estimate of the Generalization Error in Remote Sensing »

Pour ce jeu de données dont l'objectif est la tâche de détection de neige humide, nous nous sommes concentrés sur l'orbite ascendante de Sentinel-1. Il est important de noter que la neige se réchauffe après une journée complète, ce qui augmente la teneur en eau liquide. En analysant la teneur en eau liquide à 18h et à 6h à partir du modèle CROCUS, nous pouvons voir la différence entre les temps d'orbite descendante et ascendante, illustrée dans la figure 3.18 pour les massifs de Belledonne et des Grandes Rousses à deux altitudes différentes. La valeur médiane de cette différence est de 0.15 kg/m$^2$ d'octobre à fin mars, où nous pouvons considérer un comportement global de type hivernal. De même, cette différence est de 1.28 kg/m$^2$ entre avril et fin septembre. Ces valeurs confirment une présence plus significative d'eau liquide en fin de journée. Choisir cette option aide à cartographier des zones plus étendues de neige, car le manteau neigeux est plus fortement associé à l'eau liquide. De plus cette option permet d'obtenir plus de données.

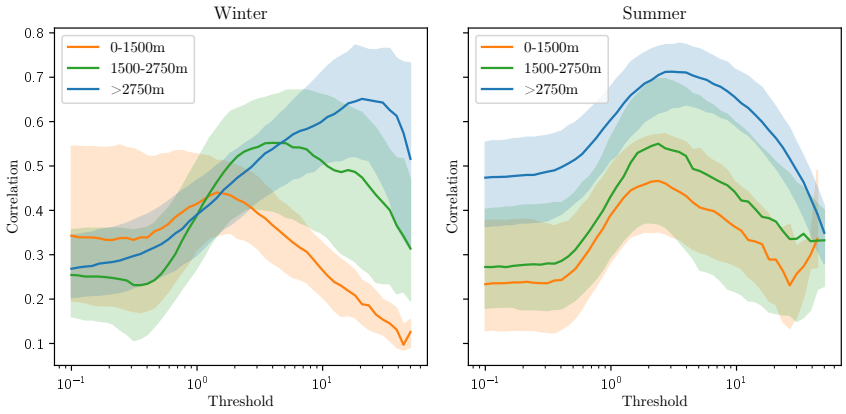



**Figure 3.19 –** Corrélation entre la règle de détection de neige humide basée sur la température minimale et la règle basée sur le seuillage de la TEL pour deux périodes dans l'année 2020-2021.

À partir des données de CROCUS, nous avons accès, entre autres, à la TEL et la température minimale et la hauteur du manteau neigeux, et ce en tout point de la tuile 31TGL, en fonction de l'altitude, de l'orientation, de la pente et du massif considéré. Nous utilisons ainsi ce modèle physique, qui permet d'exploiter tous les échantillons au détriment des erreurs de modélisation et des incertitudes locales importantes. La neige peut être considérée comme humide en fonction de la température du manteau neigeux [341]. Selon cette définition, la neige est considérée comme humide si la température du manteau neigeux est supérieure à 0°. Comparée aux définitions de la neige humide basées sur la teneur en eau liquide, cette formulation présente l'avantage de rester proche de la réalité physique tout en étant compréhensible pour les non-experts avec moins d'incertitude sur la variable TEL du modèle. Nous proposons la règle de relation suivante pour les étiquettes de neige humide, basée sur la température minimale ($T_{min}$) en degrés Celsius et la hauteur de neige ($H_s$) en mètres :

[341]: Salcedo et al. (2014), « Snow Cover Area Estimation Using Radar and Optical Satellite Information »

$$\text{Wet Snow} = T_{min} \geq 0 \text{ et } H_s \geq 0.4 \tag{3.22}$$

En incluant une exigence de hauteur de neige, nous nous assurons que le signal radar est principalement dû à l'existence de neige humide, et non à la réponse du sol sous une fine couche de neige ou des zones rocheuses avec une couverture neigeuse limitée. Cette limitation garantit également d'avoir des zones homogènes couvertes de neige. Nous présentons une analyse dans la figure 3.19 sur les deux règles de détection de neige humide données dans [36]. À partir des données de CROCUS, nous avons réalisé la corrélation entre la règle de détection basée sur la température minimale (cf (3.22)) et la règle basée sur le seuillage de la TEL, en faisant varier ce seuil de 0.1 à 50 kg/m$^2$. Nous avons séparé les échantillons collectés pendant la période de neige sèche (de juin à décembre) de ceux pendant la période de la neige humide (de janvier à juin). Nos résultats montrent que la variabilité du seuil reste élevée, intrinsèquement liée à la topographie de la zone. De plus, cette corrélation est valable pour les échantillons situés à des latitudes plus élevées mais avec un seuil plus élevé que ceux documentés dans la littérature. En analysant les altitudes plus basses, cette corrélation diminue et la corrélation optimale peut être obtenue avec la valeur du seuil. Cependant, pour des altitudes moyennes entre 1500m et 2750m, la corrélation maximale est observée pour un seuil de TEL d'environ 2%[60] de TEL, ce qui est plus cohérent avec les résultats présentés dans [34].

[36]: Tsai et al. (2019), « Remote Sensing of Snow Cover Using Spaceborne SAR »

**Précision (60)**

La sortie de CROCUS est en kg/m$^2$, la conversion en (%) est faite en considérant une densité comprise entre 100 et 300 kg/m$^3$

[34]: Techel et al. (2010), « Point observations of liquid water content in wet snow - investigating methodical, spatial and temporal aspects »

Le label (donné dans (3.22)) est associé à tous les échantillons sélectionnés à

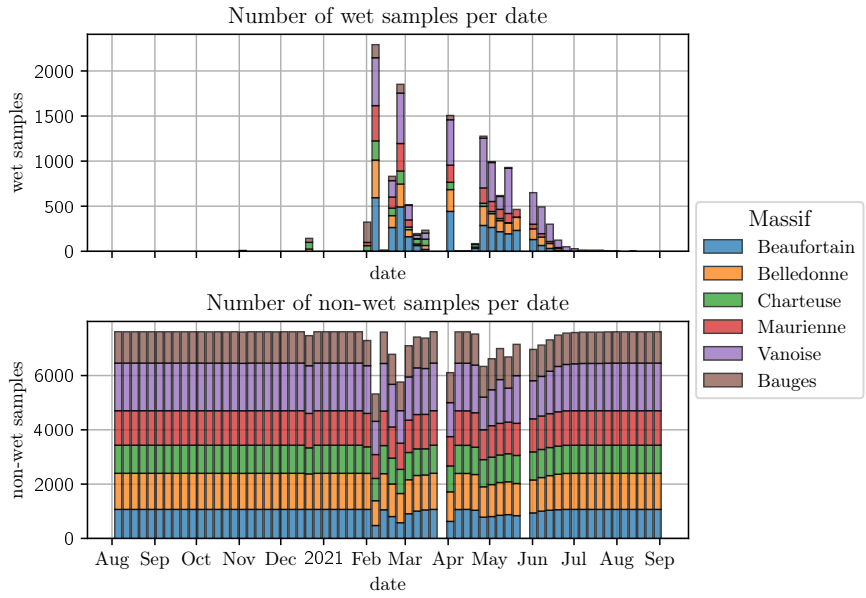



**Figure 3.20 –** Distribution du nombre d'échantillons entre août 2020 et 2021, selon les 6 massifs d'entraînement considérés pour les deux classes *humide* (wet) et *non-humide* (non wet).

partir des données topographiques de l'échantillon. La distribution mensuelle des échantillons du jeu de données d'entraînement selon les classes humide et non-humide est donnée dans la figure 3.20.

Il est intéressant de noter que la classe non-humide est répartie uniformément sur tous les massifs, tandis que la classe humide est principalement représentée par les massifs de la Vanoise et du Beaufortain. Le mois de février a également été associé à un grand nombre d'échantillons d'étiquettes humides. Cela peut s'expliquer par le temps doux en février 2021 [342], et la présence de poussière saharienne, qui accélère le réchauffement du manteau neigeux [343]. Le jeu de données final est disponible en ligne [344].

[342]: France (2021), « Hiver 2020-2021 : très arrosé avec une alternance de temps hivernal et de douceur printanière »

[343]: Dick et al. (2023), « Can Saharan dust deposition impact snowpack stability in the French Alps ? »

[344]: Gallet et al. (2023), *LSD4WSD : An Open Dataset for Wet Snow Detection with SAR Data and Physical Labelling*

### Données de validation

Pour comparer les résultats obtenus, nous avons sélectionné deux autres produits de haut niveau fournis par le programme Copernicus. Le premier est la Couverture Neigeuse Fractionnée (FSC) au-dessus du couvert végétal qui donne le pourcentage de couverture neigeuse pour chaque pixel avec une résolution de 20m par pixel à partir des données multispectrales Sentinel-2 en utilisant l'Indice de Neige Normalisé (NDSI cf (2.1)). Le second produit fourni par Copernicus est le SAR Wet Snow (SWS) obtenu par la méthode de Nagler [145] avec les données de Sentinel-1. Le produit final fait appel à un filtrage supplémentaire pour éliminer les fausses détections, de plus de nombreuses zones sont masquées, comme illustré dans la figure 3.21. Les zones urbaines, de forêts et les zones de distorsions géométriques irréversibles sont masquées. Le produit final est donné en détection binaire (neige humide ou non), avec une résolution de 60m par pixel. Un exemple est donné dans la figure 3.21 couvrant les massifs de Belledonne, Chartreuse et Bauges. Comme nous utilisons uniquement des données ascendantes, nous avons comparé nos résultats avec les produits SWS obtenus uniquement lors de passages ascendants. Deux orbites couvraient la zone de validation (trajectoire 88 et trajectoire 161), et ont été utilisées.

[145]: Nagler et al. (2000), « Retrieval of wet snow by means of multitemporal SAR data »

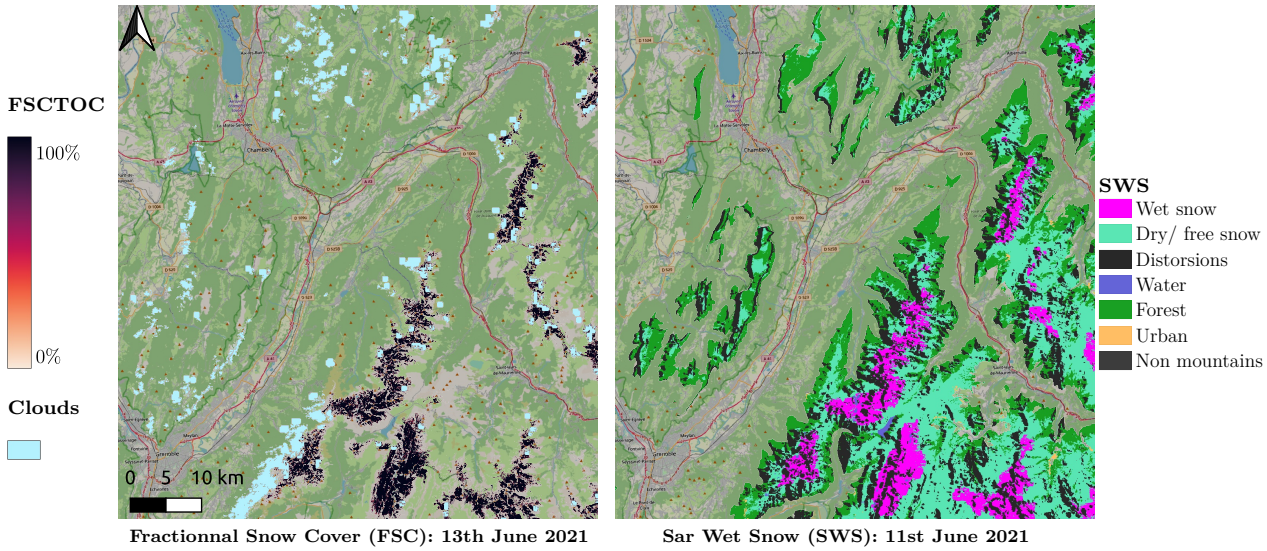



**Figure 3.21 –** Comparaison entre le produit FSC et SWS pour respectivement le 13 et 11 juin 2021, regroupant les massifs de Belledonne, Chartreuse et Bauges

### 3.2.4 LSD4WSD VX

La version étendue du jeu de données LSD4WSD est construite sur la même base que la version V1, mais avec des données supplémentaires. Nous avons ajouté 6 massifs alpins supplémentaires pour la création du jeu de données d'entraînement (Aravis, Haute-Maurienne, Haute-Tarentaise, Mont-Blanc, Thabor et Vercors). Ces massifs sont néanmoins partiellement contenus dans la tuile 31TGL. Nous considérons la même série temporelle et les mêmes données auxiliaires que pour la version V1 : les trois canaux polarimétriques (VV, VH et VV/VH), les trois canaux topographiques (élévation, orientation et pente) et les trois canaux de caractéristiques de Nagler. Cependant, nous avons ajouté les acquisitions en passe descendante de Sentinel-1, avec une série de 70 images entre le 3 juillet 2020 et le 27 août 2021. Les images en passe descendante sont acquises à 5h30, et leur traitement est identique à celui des images en passe ascendante. La seule différence est la date de référence utilisée pour la caractéristique de Nagler, qui est le 8 août 2020. Le jeu de données final est disponible en ligne [337]. Nous avons fait le choix de ne pas utiliser de transformation logarithmique (dB) pour les canaux polarimétriques, afin de laisser le choix à l'utilisateur de la méthode de traitement des données. Nous avons également modifié la taille (impaire sur chaque dimension spatiale) des échantillons à $15 \times 15$ pixels pour permettre l'évaluation d'une décision algorithmique sur le pixel central. Les fenêtres sont espacées de 500m pour éviter la corrélation spatiale entre les échantillons.

[337]: Gallet et al. (2023), *LSD4WSD : An Open Dataset for Wet Snow Detection with SAR Data and Physical Labelling*

#### Labellisation des données

Nous avons fait le choix d'élargir les informations disponibles pour chaque échantillon, en ne faisant pas de restriction sur les zones d'entraînement et de test ou de validation. Ainsi chaque échantillon est associé à 9 métadonnées permettant une sélection spécifique des échantillons pour l'entraînement et la validation et un plus grand nombre de tâches. Ces métadonnées sont regroupées en 3 catégories : topographiques, contextuelles et physiques. Les métadonnées topographiques sont l'altitude, la pente et l'orientation moyenne

de l'échantillon. Les métadonnées contextuelles sont la date de l'acquisition, le massif considéré et le type de passage (ascendant ou descendant). Enfin, les métadonnées physiques sont la TEL, la hauteur de neige et la température minimale, simulées par CROCUS[61] .

**Précision (61)**
Nous fournissons le code permettant de sélectionner tout ou partie des échantillons en fonction de ces métadonnées sur le stockage Zenedo

## 3.3 Jeu de données SAR en bande X pour l'étude des glaciers

Nous avons réalisé le dataset CPAZMAL (Cryosphere PAZ satellite MAchine Learning) afin d'exploiter des acquisitions SAR haute résolution du satellite PAZ en bande X sur la région du Mont-Blanc pour la classification des glaciers.

### 3.3.1 Zone d'étude

La zone d'étude pour ce travail est le massif du Mont-Blanc (MBM), qui couvre une superficie de 550 km$^2$ sur trois pays : la France, l'Italie et la Suisse. Les glaciers couvrent environ 145 km$^2$, ce qui en fait le massif le plus haut et le plus glacé des Alpes françaises [345]. 28 des 82 sommets supérieurs à 4000 m d'altitude dans les Alpes se trouvent dans le MBM, formant la ligne de partage des eaux entre les bassins du Rhône et du Pô. Le massif se compose de 12 grands glaciers de plus de 5 km$^2$, bordés de parois abruptes. La zone d'accumulation de certains de ces glaciers dépasse 4000 m d'altitude, et certaines langues terminales peuvent atteindre 1500 m d'altitude. Les vallées profondes sur les flancs du MBM sont principalement constituées de roches très fracturées. La géologie principale du MBM est dominée par des paragneiss[62] , des orthogneiss[63] et des migmatites[64] , avec de grandes intrusions de roches granitiques [346]. Le massif est fortement fracturé à grande échelle dans plusieurs plans avec une densité et une direction variables. La géologie et la topographie complexe du terrain, combinées à de longues périodes d'érosion glaciaire, fournissent des conditions pertinentes pour le développement de nombreuses caractéristiques glaciaires. Une analyse modélisée par [347] montre que sur les 85 km$^2$ de pentes raides (> 40°) du côté français du MBM, presque certainement, environ 45% à 79% de la zone seraient recouverts de pergélisol à des altitudes supérieures à 1900 m d'altitude sur les faces nord et au-dessus de 2600 m sur les faces sud où les conditions structurelles sont favorables. Le pergélisol serait plus continu au-dessus de 2600 m. Au-dessus de 3600 m, le pergélisol occuperait très certainement l'intégralité des parois rocheuses escarpées, quelle que soit l'orientation. Cela fait du MBM un site intéressant pour étudier différentes caractéristiques glaciaires. Cette étude [347] a analysé différentes caractéristiques des glaciers telles que les zones d'accumulation et d'ablation, les glaciers suspendus et les tabliers glaciaires à l'intérieur du MBM. De plus, les zones autour du MBM, en particulier la ville de Chamonix, ont également été étudiées pour d'autres classes d'utilisation des terres comme les zones urbaines, les forêts, les plaines et les formations rocheuses.

[345]: Gardent et al. (2014), « Multitemporal glacier inventory of the French Alps from the late 1960s to the late 2000s »

**Définition (62)**
*Paragneiss:* Roche métamorphique (dont la formation est due à la solidification d'une roche dans des conditions de température et de pression élevées sur une longue période de temps) dérivée de roche sédimentaire (provenant de l'accumulation de sédiments).

**Définition (63)**
*Orthogneiss:* Roche métamorphique dérivée de roche magmatique (résultat de la solidification d'un magma avec ou sans cristallisation complète).

**Définition (64)**
*Migmatites:* Roche métamorphique résultant majoritairement de la fusion partielle de roches préexistantes dans la croûte terrestre.

[346]: Raumer et al. (1993), « The Pre-Alpine Evolution of the External Massifs »

[347]: Magnin et al. (2017), « Modelling rock wall permafrost degradation in the Mont Blanc massif from the LIA to the end of the 21st century »

### 3.3.2 Données SAR

Le jeu de données étudié est basé sur 57 acquisitions du satellite espagnol PAZ[65] acquises entre le 08 janvier 2020 et le 20 novembre 2021. Le satellite

**Précision (65)**
satellite SAR espagnol de type commercial en bande X (9.65GHz) lancé en 2018, avec un période de révisite de 11 jours en polarisation HH et HV.

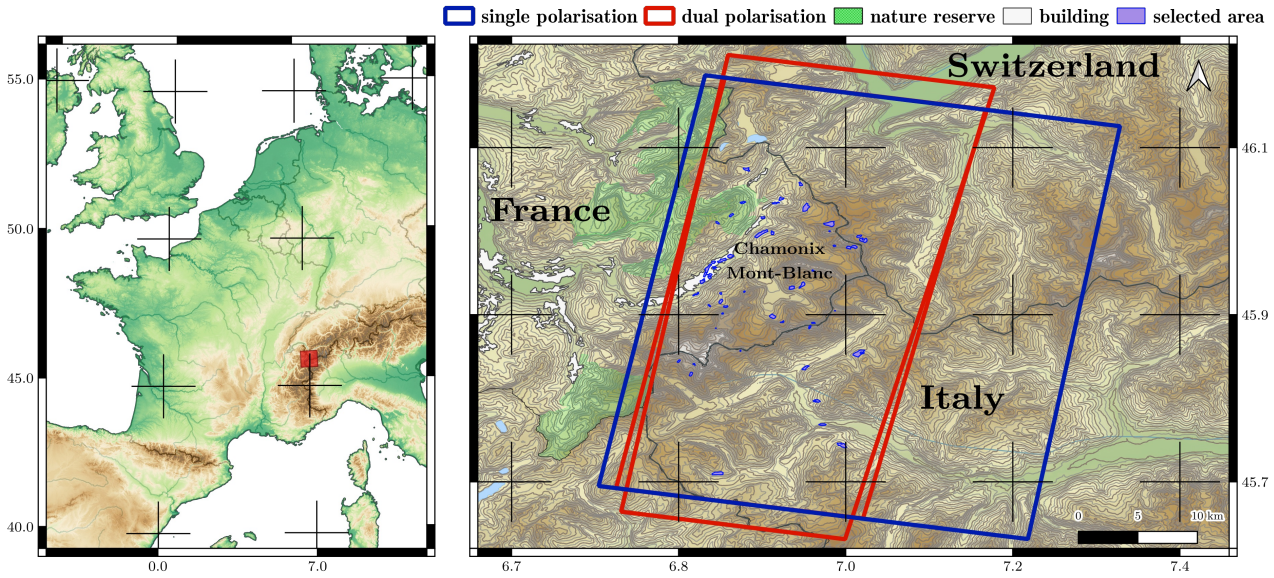



**Figure 3.22 –** Visualisation de l'acquisition du satellite PAZ en orbite descendante sur le massif du Mont-Blanc et des polygones de vérité terrain.

fonctionne en bande X à une longueur d'onde de 9,65 GHz et fournit des acquisitions tous les 11 jours. Les acquisitions sont centrées sur le MBM, couvrant une superficie comprise entre 980 et 1600 km$^2$ à la frontière entre la Suisse, l'Italie et la France, comme le montre la figure 3.22. Notre étude s'est concentrée exclusivement sur les données descendantes en raison de l'angle d'incidence constant (37.86°) maintenu tout au long de la période d'acquisition. Il convient de noter qu'après le 27 janvier 2020, il y a eu une transition dans le type d'acquisition de données, passant de la polarimétrie double (HH et HV) à la polarisation simple (HH). Cette transition a principalement affecté la résolution spatiale et l'étendue de la zone d'acquisition, divisant par deux la résolution azimutale de 6.6 à 3.3m. L'emprise au sol des acquisitions est montrée dans la figure 3.22. Un traitement de base de la série temporelle d'images a été effectué à l'aide du logiciel SARScape à partir des données de niveau L1 complexe fournies par PAZ. Les étapes de traitement impliquées sont le recalage au niveau sub-pixel, la calibration radiométrique, la détection et la normalisation. À la fin des étapes de prétraitement, le produit final donne la valeur du coefficient de rétrodiffusion $\sigma^0$ en amplitude en géométrie radar.

### 3.3.3 Labellisation des données

Pour l'analyse, 8 classes distinctes ont été définies, dont 4 de nature générale (CIT, FOR, PLA, ROC, correspondant aux zones urbaines, aux forêts, aux

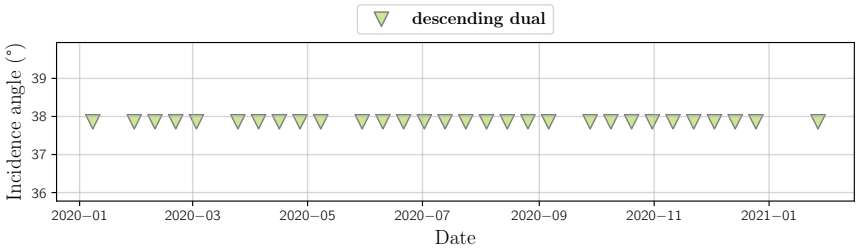



**Figure 3.23 –** Répartition temporelle des acquisitions de PAZ

plaines et aux terrains rocheux, respectivement) et 4 spécifiquement adaptées à l'analyse des glaciers (ABL, ACC, ICA et HAG, représentant les zones d'ablation, les zones d'accumulation, les tabliers glaciaires et les glaciers suspendus, respectivement). Nous avons délimité manuellement entre 5 et 7 régions d'intérêt (ROIs) pour chaque classe en utilisant des images optiques de Pléiades et des données sur l'occupation du sol de Copernicus. L'utilisation d'images radar à ouverture synthétique (SAR) a introduit des distorsions géométriques telles que le recouvrement et l'ombre. Ces zones ont été masquées pour éviter de sélectionner des ROIs dans des zones affectées par ces distorsions. Nous avons créé un nouveau jeu de données basé sur les acquisitions de PAZ en projetant ces ROIs dans la géométrie radar en utilisant un Modèle Numérique de Terrain (MNT) Copernicus de résolution 25 mètres. Nous avons choisi de projeter les ROI dans la géométrie radar plutôt que de géocoder les images dans la géométrie au sol pour minimiser les transformations géométriques. Bien que des MNT avec des résolutions inférieures à 10 m existent, ils sont souvent spécifiques à chaque pays. Par conséquent, un MNT de résolution plus basse couvrant les trois pays présents dans les images de PAZ a été sélectionné. D'un point de vue distorsion de l'image, il est plus intéressant de projeter une région d'intérêt (ROI) de faible taille dans la géométrie radar plutôt que de géocoder l'image dans sa globalité en géométrie sol. En conséquence, pour chaque date de la série d'images de PAZ, nous avons extrait 55 ROIs, dont les détails topographiques et de positions sont fournis dans les tableaux A.3 et A.2. Il est important de noter que chaque ROI varie en taille et en forme, allant d'environ 1 hectare (ha) à plus de 30 hectares.

Le jeu de données utilisé dans l'étude est fourni en libre accès, avec le code Python associé permettant la création d'une base de données composée de fenêtres de même taille associées à des informations de classe et de groupe. Cette information de groupe permet l'évaluation des méthodes d'apprentissage automatique en séparant spatialement les données d'entraînement des données de test [340]. Le jeu de données brut est librement accessible sur Zenodo [348].

## 3.4 Conclusion

Dans ce chapitre, nous avons présenté les bases théoriques et méthodologiques pour l'analyse des données SAR. Nous avons vu comment les méthodes d'apprentissage automatique peuvent être utilisées, avec a priori de distribution de probabilité, supervisé et non-supervisé. Nous avons également présenté les différentes étapes de la classification des données SAR, notamment la préparation des données, la sélection des caractéristiques, la sélection du modèle et l'évaluation des performances, notamment pour l'étude de la cryosphère. Ce contexte nous a amenés à développer des jeux de données labellisés pour l'étude de la neige et des glaciers. Nous avons présenté le jeu de données LSD4WSD pour l'étude de la neige humide dans les Alpes françaises du Nord et le jeu de données CPAZMAL pour la classification des glaciers dans le massif du Mont-Blanc. Ces jeux de données ont été acquis par les satellites SAR en bande C Sentinel-1 et en bande X PAZ respectivement. Ces derniers ont été soigneusement préparés, manuellement labellisés dans le cadre du jeu de données CPAZMAL et automatiquement labellisés par le modèle physique du manteau neigeux de Météo-France (CROCUS) pour le jeu de données LSD4WSD. Ces jeux de données sont utilisés pour l'évaluation des méthodes d'apprentissage automatique dans les chapitres suivants.

# Détection de la neige humide par apprentissage automatique

# 4


**4.1 Modèles d'apprentissage automatique . . . . . 81**
**4.2 Résultats quantitatifs . . . . . 85**
**4.3 Résultats qualitatifs et analyse temporelle . . . . . 90**
**4.4 Conclusion . . . . . 98**


Nous proposons dans ce chapitre d'étudier les capacités de différents modèles d'apprentissage automatique à détecter la neige humide à partir de données radar et d'informations topographiques. Pour cela nous utilisons le jeu de données LSD4WSD décrit dans le chapitre précédent, section 3.2.3. Pour rappel, ce jeu de données est composé de 9 caractéristiques radar et topographiques, et d'une étiquette binaire indiquant la présence ou l'absence de neige humide obtenue à partir de la température minimale du manteau neigeux simulée par le modèle CROCUS. L'objectif de cette étude est d'observer la validité du jeu de données LSD4WSD pour la détection de la neige humide et étudier les caractéristiques pertinentes pour cette tâche. Nous évaluons, dans un premier temps, les performances de ces modèles au vu de métriques propres à la détection ou classification (détaillées dans la section 53). Dans un second temps, nous analysons les résultats cartographiques obtenus par une sélection de ces modèles, et la tendance temporelle de l'indicateur de probabilité de neige humide.

## 4.1 Modèles d'apprentissage automatique

Nous présentons dans cette section les 8 modèles d'apprentissage automatique utilisés pour la détection de la neige humide, les différentes combinaisons de canaux utilisées et les métriques de performance utilisées pour comparer les modèles. Nous détaillons également les différentes étapes de prétraitement des données et les résultats obtenus pour chaque modèle.

### 4.1.1 Modèles

Les algorithmes proposés sont les suivants : Adaboost, Random-Forest (RF), Support Vector Machine (SVM) avec noyau gaussien, K-Nearest Neighbors (KNN), Régression Logistique (RL), Perceptron Multi-Couche (MLP) et une architecture de Réseau de Neurones Convolutionnels (CNN). Ce choix met en évidence trois catégories d'algorithmes. Les algorithmes basés sur l'apprentissage d'un ensemble de classifieurs faibles (Adaboost et RF). La catégorie d'algorithmes basée sur le concept de séparation ou de distance (SVM, KNN). La dernière catégorie couvre les algorithmes d'auto-apprentissage neuronaux (CNN, MLP). L'apprentissage a été réalisé en utilisant les bibliothèques Scikit-Learn et Keras de Python[66] . Les principaux paramètres des algorithmes sont donnés dans le tableau 4.1.

**Précision (66)**
Les codes sont disponibles en ligne sur Github : ML-WetSnowSAR

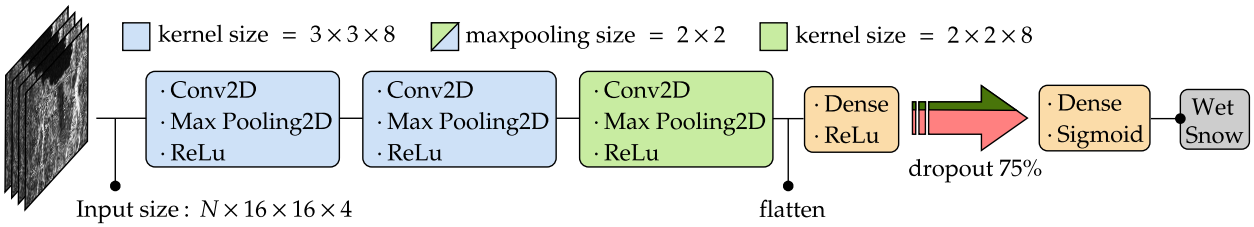



**Figure 4.1 –** Architecture CNN proposée pour la classification de la neige humide.

**Réseau de Neurones Convolutionnels (CNN)**

[12]: Parikh et al. (2020), « Classification of SAR and PolSAR images using deep learning »

L'architecture du CNN est basée sur l'étude des architectures CNN pour les images SAR, proposée par [12] et illustrée dans la figure 4.1. Le choix est porté par la volonté d'obtenir une architecture simple avec un nombre de paramètres limité (5497 paramètres à apprendre). Ceci est réalisé dans notre cas en ne gardant que 3 blocs (en bleu et vert dans la figure) consistant en une cascade d'une couche de convolution 2D, un max-pooling (permettant de réduire le nombre de paramètres) et une activation ReLu. La deuxième partie du réseau (en orange dans la figure) se concentre sur l'apprentissage des combinaisons de ces décompositions par des couches entièrement connectées. Deux couches denses composent cette partie afin de réduire progressivement le nombre de paramètres au nombre de classes. Pour éviter le sur-apprentissage, nous avons ajouté une couche de *dropout* qui ne laisse passer que 25% des valeurs de la couche précédente aléatoirement pendant l'apprentissage.

### 4.1.2 Prétraitement et représentation des données

Nous normalisons les données entre 0 et 1. Le minimum et le maximum de tous les échantillons sont utilisés pour normaliser les 6 premiers canaux. Pour les 3 canaux construits sur le ratio avec une image de référence sans neige, les valeurs extrêmes utilisées pour la normalisation correspondent à la moyenne du minimum et du maximum de tous les échantillons. Nous rappelons que les échantillons sont de type GRD en échelle logarithmique. À partir d'un ensemble de données dans $\mathbb{R}^{N,W_x,W_y,C}$, avec $N$ le nombre d'échantillons, $W_x, W_y$ la taille des fenêtres et $C$ le nombre de canaux ou caractéristiques, nous étudions une étape de prétraitement en amont de ces algorithmes basés sur deux solutions :

**Table 4.1 –** Paramétrisation des algorithmes utilisés

| Algorithmes | Paramètres |
|---|---|
| KNN | K voisins = 50 |
| SVM | noyau = Fonction de base radiale (RBF) |
| RF | estimateurs = 200, critère = entropie |
| AdaBoost | estimateurs = 200 |
| MLP | couches cachées = 100, terme de régularisation L2 = 0.01 |
| Régression Logistique | pénalité L2 |
| CNN | 3 blocs de convolution 2D + Maxpooling + ReLu |

La première approche consiste à calculer l'histogramme de chaque caractéristique pour passer de $\mathbb{R}^{N,W_x,W_y,C}$ à une représentation $\mathbb{R}^{N,H_{bins}}$. Ce traitement permet de synthétiser l'information du patch et a été largement utilisé dans la classification d'images [349] [350]. Dans le contexte du traitement de grandes images ou de grands ensembles de données, cette solution nous permet de conserver une certaine frugalité computationnelle tout en offrant des performances intéressantes. Cependant, cette stratégie ne prend pas en compte la structure spatiale potentielle. Pour que l'histogramme soit statistiquement fiable et homogène, un nombre optimal de bins $H_{bins} = 16$ est utilisé. Ce choix provient de la règle de Freedman-Diaconis. Nous utilisons la valeur entière maximale de $H_{bins}$ sur l'ensemble de données et les canaux. Le calcul de l'histogramme est effectué sur une plage dynamique entre 0 et 1, ce qui garantit que tous les histogrammes des échantillons ont le même support. Cette approche est utilisée pour la comparaison des modèles, sauf pour le CNN où les images sont directement données en utilisant les $C$ canaux considérés, ainsi que pour l'étude des différents canaux d'entrée. Cette approche est utilisée pour comparer les différents modèles d'apprentissage mais aussi pour évaluer l'impact des différentes caractéristiques radar et topographiques sur la classification de la neige humide. Pour ce faire, nous présentons dans le tableau 4.2, les différentes combinaisons de canaux testées. Les canaux de rétrodiffusion sont notés VV et VH, les indices de Nagler sont notés $VV_{ratio}$ et $VH_{ratio}$ et les canaux de topographie sont notés élévation, orientation et pente. Les combinaisons de canaux sont notées de A à J.

[349]: Dai et al. (2011), « Multilevel Local Pattern Histogram for SAR Image Classification »

[350]: Kayabol (2015), « Histogram-Based Contextual Classification of SAR Images »

La deuxième étude se concentre sur l'utilisation de différents paramètres statistiques pour chaque canal. Il est donc intéressant d'étudier quelle information statistique prédomine sur la décision de classification, et si c'est le cas, quelles statistiques sont significatives pour distinguer les classes. Dans ce contexte, nous nous intéressons à la moyenne $\mu$, à l'écart-type $\sigma$, à l'asymétrie $\gamma_1$ et à l'aplatissement $\gamma_2$. L'intérêt de cette analyse est sa capacité à donner des résultats explicables, à partir d'une distance euclidienne entre un petit nombre de variables (maximum 4) directement corrélées à la nature du signal, et à analyser la perspective de l'utilisation de méthodes paramétriques. Les quatre combinaisons statistiques testées sont résumées dans le tableau 4.3.

**Table 4.2 –** Description des combinaisons de canaux testées

| Identifiant | Combinaisons de canaux |
|---|---|
| **A** | VV, VH |
| **B** | **A**, VV/VH |
| **C** | **A**, élévation |
| **D** | **A**, orientation |
| **E** | **A**, pente |
| **F** | **A**, élévation, orientation, pente |
| **G** | $VV_{ratio}$, $VH_{ratio}$ |
| **H** | **A**, $VV_{ratio}$, $VH_{ratio}$ |
| $\mathbf{H_r}$ | **A**, VV/VH ,$VV_{ratio}$, $VH_{ratio}$, $VV_{ratio}/VH_{ratio}$ |
| **I** | **A**, $VV_{ratio}$, $VH_{ratio}$, orientation |
| $\mathbf{I_D}$ | **A**, $VV_{ratio}$, $VH_{ratio}$, élévation |
| $\mathbf{I_P}$ | **A**, $VV_{ratio}$, $VH_{ratio}$, pente |
| **J** | **F**, **G**, VV/VH, $VV_{ratio}/VH_{ratio}$ |

**Table 4.3 –** Description des combinaisons statistiques testées, où $\mu, \sigma$ sont la moyenne et l'écart-type et $\gamma_1, \gamma_2$ sont donnés dans Eq.(2.3) et (2.4)

| Identifiant | Combinaisons de statistiques |
|---|---|
| **L** | $\mu$ |
| **M** | $\mu, \sigma$ |
| **N** | $\mu, \sigma, \gamma_1$ |
| **O** | $\mu, \sigma, \gamma_1, \gamma_2$ |

### 4.1.3 Gestion des données et scénarios d'apprentissage

Le jeux de données LSD4WSD est divisé en deux ensembles : un ensemble d'entraînement et un ensemble de test, respectivement composés de 487157 échantillons de taille 16 par 16 par 9 pour l'entraînement et 3668 pour les tests. L'équilibre de l'ensemble de données est important et afin d'éviter un biais d'apprentissage, nous effectuons un équilibrage de l'ensemble de données d'entraînement entre les classes. Cela peut être fait par sous-échantillonnage aléatoire pour équilibrer la classe majoritaire *non humide*. Cette technique nous permet d'avoir un nombre égal d'échantillons dans les deux classes, mais si le déséquilibre est significatif, elle réduit considérablement la classe majoritaire, passant dans notre cas de 474K échantillons à 33K.

Le jeu de données de test est équilibré pour avoir un nombre égal d'échantillons des classes humides et non humides. Cependant, équilibrer le jeu de données d'entraînement par sous-échantillonnage entraînerait la suppression d'une proportion significative d'échantillons et donc la réduction de la variabilité du jeu de données (la classe humide étant minoritaire avec 13975 échantillons, nous réduirions la taille du jeu de données de près de 95%). Pour résoudre ce problème, nous proposons une variante de la technique K-Fold [351] : le K-Fold[67] stratifié équilibré. Nous divisons la classe majoritaire du jeu de données en K sous-ensembles de taille égale au nombre d'échantillons de la classe minoritaire. Chaque sous-ensemble de la classe majoritaire est combiné avec la classe minoritaire pour former un pli. De cette façon, nous pouvons balayer l'ensemble du jeu de données et étudier la variabilité des modèles sur tous les plis équilibrés, comme illustré dans la figure 4.2. Dans notre cas, nous obtenons $K = 34$ plis avec 13975 échantillons de chaque classe.

[351]: Arlot et al. (2010), « A survey of cross-validation procedures for model selection »

**Précision (67)**
Nous parlons ici de fold ou pli, qui est une division de l'ensemble de données en K sous-ensembles.

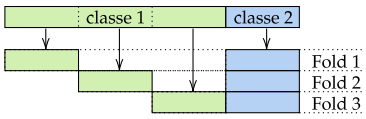



**Figure 4.2 –** K-Fold stratifié équilibré

Nous introduisons deux indicateurs supplémentaires : la meilleure précision sur la ROC (*Best Accuracy on the ROC* BAROC) et le taux de faux positifs constant sur la ROC (*False positive rate Constant on the ROC* FCROC). En regardant la sortie du modèle non seuillé, c'est-à-dire directement la probabilité d'appartenir à la classe *humide*, nous étudions quelle valeur de seuil sur ces probabilités donne respectivement la meilleure précision et un taux de faux positifs constant. Nous fixons le taux de faux positifs à 5%. Ainsi, nous appliquons ce nouveau seuillage pour affiner les résultats fournis par le modèle selon notre utilisation. Garder le taux de faux positifs bas et constant assure que la quantité de neige *humide* ou *non humide* n'est pas surestimée. Ce post-traitement donne ainsi un encadrement entre surestimation et sous-estimation de la neige humide.

Nous proposons deux niveaux d'évaluation. Le premier est quantitatif et est réalisé de la manière suivante :

- Le F1-score (cf (3.18)) est utilisé pour évaluer la performance des modèles, le choix des caractéristiques et le pré-traitement des données. La technique de K-Fold stratifié équilibré est utilisée sur le jeu de données d'entraînement pour évaluer la performance moyenne des modèles sur l'ensemble du jeu de données et fournir la variabilité des performances.

- Les courbes ROC, et les deux indicateurs FRCROC et BAROC, sont estimées sur l'ensemble de test, soit le massif des Grandes Rousses, pour évaluer la performance du meilleur modèle, et ce à partir d'un apprentissage sur les massifs d'entraînement.

Le deuxième niveau d'évaluation est qualitatif et est réalisé de la manière suivante :

- Les cartes de probabilité de neige humide sont générées, en utilisant les deux indicateurs FCROC à 5% et BAROC, pour une sélection de modèles et de combinaisons de canaux. Ces cartes sont comparées visuellement pour évaluer la qualité des prédictions et mises en perspective avec les produits Copernicus, FSC et SWS, sur la zone de validation (cf figure 3.15).
- L'évolution temporelle de la probabilité de neige humide est étudiée sur le meilleur modèle et une sélection de canaux. Cette étude est réalisée sur trois emplacements de la carte, situés à la station météorologique de l'Alpe d'Huez, celle du Lac Noir, et du Galibier, illustrée dans la figure 3.15. Nous comparons les résultats obtenus avec 3 mesures terrains sur ces sites : température de l'air, hauteur de neige et précipitation. Nous ajoutons également la tendance temporelle du produit SWS sur ces sites.

Il est important de noter que nous nous plaçons dans un cadre d'apprentissage global, en évaluant les capacités de généralisation des différents algorithmes. Dans ce cadre l'enjeu est de pouvoir appliquer un modèle appris sur des dates différentes et des zones géographiques différentes, où le modèle CROCUS n'est pas utilisé.

## 4.2 Résultats quantitatifs

### 4.2.1 Évaluation des modèles

Les résultats de l'évaluation des sept modèles pour les 34 plis avec les deux bandes VV et VH sont donnés dans le tableau 4.4. Nous pouvons voir que les meilleurs résultats pour les trois métriques sont obtenus avec le modèle KNN. Dans notre cas, nous utilisons ce modèle avec $K = 50$, le nombre de voisins[68] . Il est intéressant de noter que les résultats du modèle KNN sont meilleurs que les résultats du modèle CNN, bien que le modèle CNN soit un modèle d'apprentissage profond, qui est plus complexe que le modèle KNN.

**Précision (68)**
La détermination du nombre optimal de voisin a été réalisée de manière empirique.

De plus, la variabilité des deux modèles de la famille des modèles d'apprentissage profond (CNN et MLP) est plus importante que celle des autres modèles. Ce facteur est assez important, car il montre le sur-apprentissage de certains

| | Adaboost | RF | KNN | CNN |
|---|---|---|---|---|
| Précision | 70.9 ± 0.3 | 71.8 ± 0.3 | **73.2 ± 0.3** | 71.5 ± 1.9 |
| Score F1 | 74.7 ± 0.2 | 75.4 ± 0.2 | **75.9 ± 0.3** | 70.7 ± 2.4 |
| Kappa | 41.8 ± 0.7 | 43.6 ± 0.6 | **46.5 ± 0.6** | 43.1 ± 3.7 |

| | Régression Logistique | MLP | SVM |
|---|---|---|---|
| Précision | 70.4 ± 0.2 | 72.0 ± 2.6 | 72.6 ± 0.2 |
| Score F1 | 74.5 ± 0.1 | 75.0 ± 1.0 | 75.6 ± 0.1 |
| Kappa | 40.8 ± 0.4 | 43.9 ± 5.1 | 45.1 ± 0.3 |

**Table 4.4 –** Résultats pour 3 métriques pour les 7 classifieurs considérés pour le problème de détection de neige en utilisant K-fold, avec $K = 34$

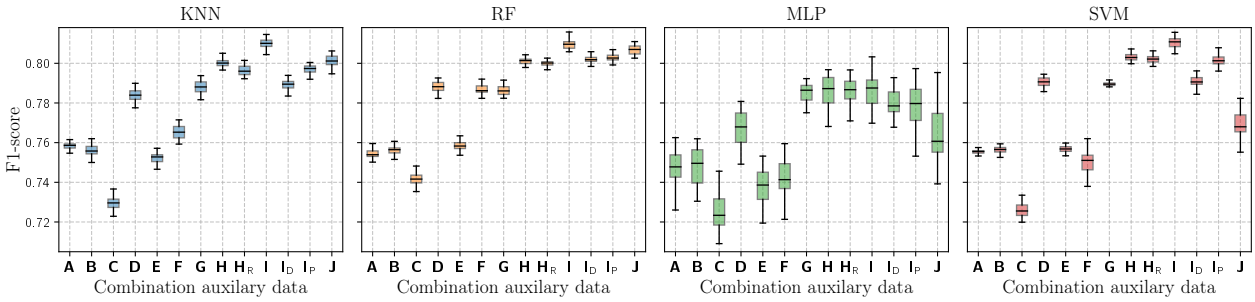



**Figure 4.3 –** Évaluation du score F1 pour les 4 meilleurs algorithmes obtenus dans le tableau 4.4 en fonction des combinaisons de canaux considérées

des modèles. Cela peut s'expliquer par le fait que pour chaque pli, une instance d'un classifieur est formée puis testée sur le même ensemble de données. Si la variabilité entre les plis est grande, cela signifie que l'apprentissage de l'instance du classifieur se concentre sur la maximisation des métriques spécifiques à l'ensemble de données du pli plutôt que par la recherche d'un modèle suffisamment général. Cette réduction de la variabilité va de pair avec la généralisation du modèle sur tous les plis, mais entraîne une légère diminution de certains critères tels que la précision par rapport à d'autres modèles tels que le KNN. Sur cet aspect purement métrique, le modèle KNN et le modèle SVM sont les meilleurs modèles pour la détection de la neige humide, avec une performance moyenne de plus de 75% pour le score F1 et ce avec une variabilité comprise entre 0.3 et 0.1%.

Compte tenu de ces résultats, nous avons choisi de sélectionner les 4 meilleurs algorithmes en termes de score F1 (SVM, RF, MLP et KNN) pour évaluer la performance de ces modèles en fonction des combinaisons des bandes d'entrées (voir Tableau 4.2). Cette sélection détaille le comportement des 3 principales familles proposées. Des résultats complémentaires peuvent être trouvés dans [352], sur les performances des modèles de forêt aléatoire et de CNN.

[352]: Gallet et al. (2023), « CNN Classification of Wet Snow By Physical Snowpack Model Labeling »

### 4.2.2 Évaluation des données auxiliaires

Les résultats de l'évaluation des combinaisons de canaux pour les 4 meilleurs algorithmes sont donnés dans la figure 4.3. Pour l'algorithme KNN, nous pouvons noter que la solution utilisant les canaux VV et VH (**A**) seuls donne une base pour les autres combinaisons avec une valeur autour de 76% et une faible variabilité. L'ajout du ratio VV/VH (**B**) dégrade légèrement les performances. La combinaison la plus frappante est l'ajout d'informations topographiques. L'élévation en particulier ne permet pas d'affiner les résultats, contrairement à l'orientation qui améliore le score F1 de plus de 2%. On peut voir que la combinaison (**G**) composée uniquement des ratios VV et VH des dates actuelles par rapport à la date de référence donne de meilleurs résultats que les combinaisons utilisant à la fois les canaux polarimétriques simples et les informations topographiques. La tentative d'utiliser tous les canaux pour avoir la meilleure performance possible n'est pas la meilleure solution, mais plutôt les solutions **H** et **I**, qui offrent un compromis entre la meilleure performance et une moindre variabilité sur tous les K-Folds. On observe que la performance des algorithmes considérés est assez constante sur les canaux examinés. Cependant, l'approche MLP présente un degré significatif de variabilité lorsqu'elle est appliquée à des combinaisons. L'algorithme de forêt

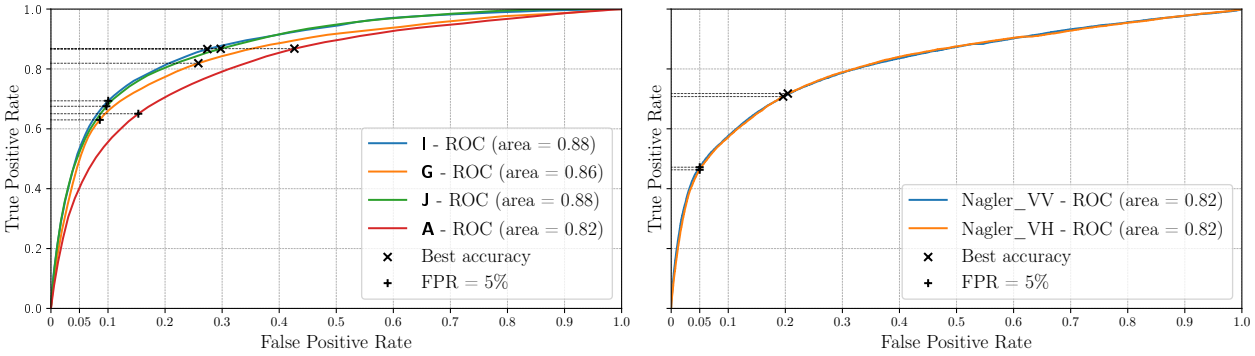



**Figure 4.4 –** Courbe ROC pour 4 combinaisons de canaux d'entrée pour le classifieur KNN et leur métrique BAROC et FCROC associée à 5% (à gauche). Courbe ROC pour le classifieur de type seuillage simple sur la moyenne de chaque échantillon pour les canaux VV et VH (seuillage Nagler), avec les métriques BAROC et FCROC (à droite).

aléatoire est remarquable pour ses bonnes performances à grande échelle et sa variabilité minimale sur tous les plis. Cette méthodologie a été précédemment utilisée dans [90].

[90]: Tsai et al. (2019), « Wet and Dry Snow Detection Using Sentinel-1 SAR Data for Mountainous Areas with a Machine Learning Technique »

La figure 4.4 donne les courbes ROC pour une sélection de combinaisons pour la classe humide en utilisant le modèle KNN. Sur la figure de gauche, les seuils BAROC (×) et FCROC (+) ont été ajoutés. Ces seuils ont été calculés à partir des données d'entraînement. Cela explique pourquoi les seuils FCROC, calculés pour un taux de faux positifs de 5%, sont décalés. Les meilleurs classifieurs sur ce critère sont ceux qui ont le plus faible taux de faux positifs de test de 5%. La combinaison **G** est la meilleure sur ce critère avec un taux de faux positifs de test de 8% contre ∼ 10% et 16% pour les autres. Sur le critère de l'aire sous la courbe (AUC), les deux combinaisons **I** et **J** donnent les meilleures performances en termes de détection, avec un léger avantage pour la combinaison **I** qui donne sa précision maximale avec un taux de faux positifs inférieur à celui de la combinaison **J**.

Ces résultats peuvent être liés à la partie droite de la figure 4.4 où nous avons étudié la capacité de la rétrodiffusion moyenne sur une fenêtre des canaux VV et VH à détecter la classe de neige humide. Pour ce faire, nous avons utilisé l'ensemble de données d'entraînement et calculé la valeur moyenne de chaque échantillon pour VV et VH. Nous remarquons que les deux seuils sont obtenus pour un taux de vrais positifs inférieur aux méthodes proposées dans la partie droite de la figure, démontrant l'avantage des canaux supplémentaires et d'une classification basée sur un algorithme plus complexe qu'un seuillage.

| | VV | VH |
|---|---|---|
| min/max (dB) | -35.96 / 36.42 | -33.30 / 29.53 |
| BAROC (dB) | -0.99 | -1.33 |
| FCROC 5% (dB) | **-2.61** | **-3.40** |

**Table 4.5 –** Valeur de seuil pour les métriques BAROC et FCROC pour l'apprentissage sur la moyenne de l'échantillon pour VV et VH

### Sur le seuillage de Nagler

L'algorithme de seuillage peut alors être considéré et la courbe ROC ainsi que les seuils BAROC et FCROC associés à cette moyenne peuvent être calculés. La valeur FCROC donne une valeur de seuil entre 0 et 1 pour laquelle seulement 5% de faux positifs sont obtenus (et de même pour la valeur BAROC). Ainsi, à partir de ces valeurs de seuil, et en connaissant la dynamique utilisée pour

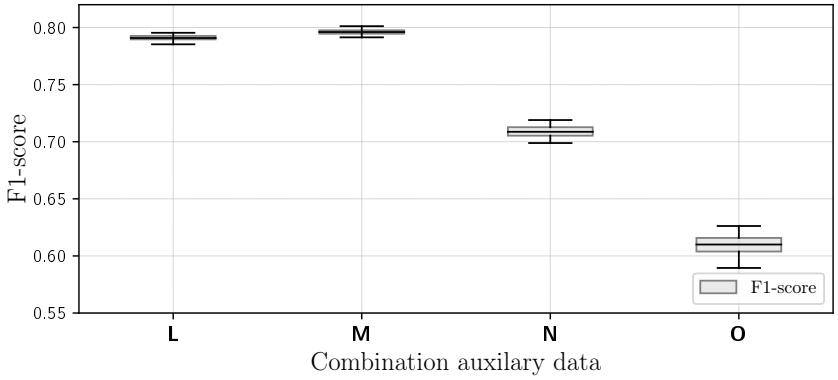



**Figure 4.5 –** Score F1 pour le classifieur KNN en fonction de la combinaison de statistiques décrites dans le tableau 4.3 sur les combinaisons de canaux **H**

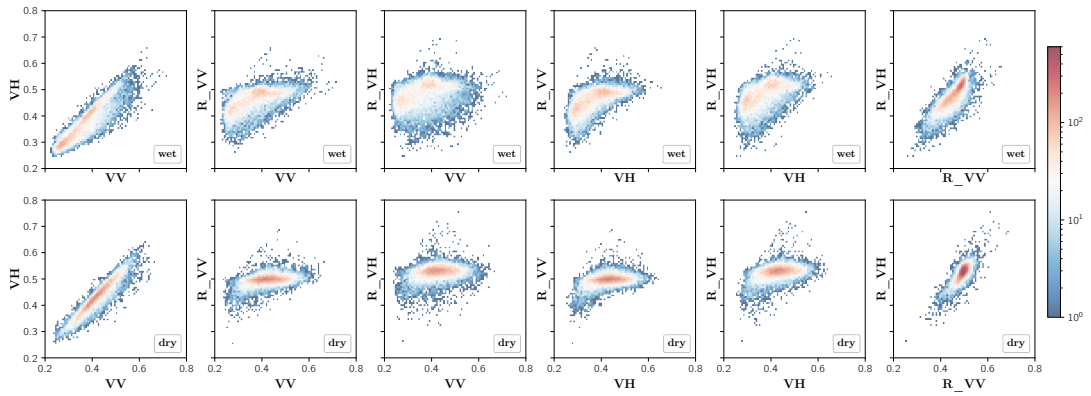
normaliser les données, nous pouvons arriver à une valeur de seuil en dB qui peut être comparée à celle traditionnellement utilisée par la méthode Nagler. Les résultats obtenus sont donnés dans le tableau 4.5. On peut voir que la valeur de seuillage traditionnelle de -3dB est proche du seuil obtenu avec le seuillage FCROC. Le seuillage basé sur la moyenne d'une fenêtre de $16 \times 16$ pixels, pour obtenir un taux de faux positifs de la classe de neige humide par rapport à la règle d'étiquetage donnée dans 3.22 est similaire au seuil proposé dans [145]. Ce résultat met l'accent sur la pertinence de l'étiquetage proposé basé sur CROCUS, où en moyenne la neige est considérée comme humide par une approche de seuillage sur les données avec une valeur proche de celle utilisée par Nagler. Nous remarquons également que lorsque nous sélectionnons le seuillage BAROC qui donne le seuillage pour la précision maximale, le seuil de détection tend à être supérieur à -3dB. Par conséquent, nous pouvons déduire que dans le cadre de cet étiquetage, le seuil optimal pour détecter des échantillons respectant l'étiquetage proposé est d'environ 2dB supérieur au seuil de Nagler. Cela se traduit par la détection d'une zone plus grande. Cette conclusion est basée sur un apprentissage du seuil respectant l'étiquetage CROCUS à partir de données ascendantes.

[145]: Nagler et al. (2000), « Retrieval of wet snow by means of multitemporal SAR data »

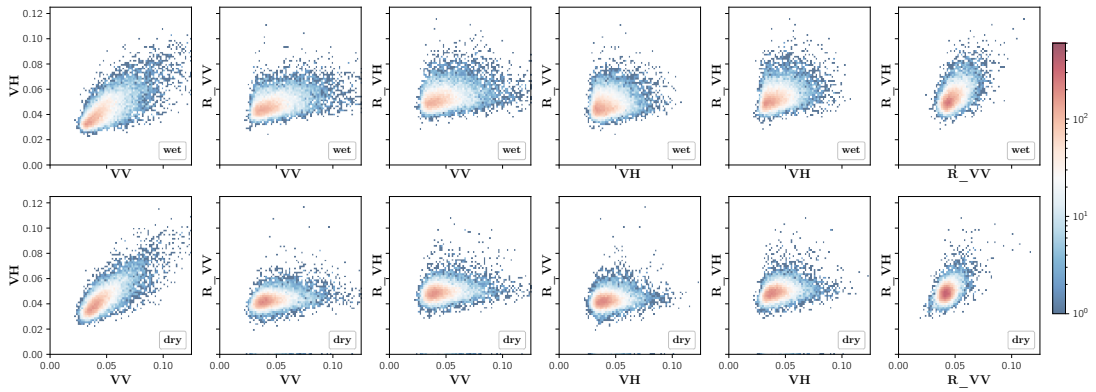
### 4.2.3 Ensemble de statistiques

À partir des résultats obtenus, modèles et combinaisons de canaux, nous proposons d'explorer la capacité du modèle KNN à détecter la neige humide en utilisant des statistiques sur la combinaison de bandes **H**, afin d'observer des résultats qui dépendent uniquement des données polarimétriques. Les résultats obtenus sont donnés dans la figure 4.5.

Nous remarquons d'abord que l'utilisation simple de la moyenne de chacun des 4 canaux considérés (**L** : 79.1 ± 0.2) donne une diminution de ~ 2% sur le score F1 par rapport à l'utilisation de l'histogramme. Cependant, nous avons réduit le nombre de caractéristiques pour chaque échantillon par 16 pour effectuer cette tâche avec seulement 4 caractéristiques. L'ajout de moments du second ordre (**M** : 79.6 ± 0.2) améliore les performances. L'utilisation de moments d'ordres supérieurs, respectivement **N** et **O**, réduit le score F1 (respectivement 70.9 ± 0.5 et 60.9 ± 1), et ce en augmentant la variabilité sur l'ensemble des plis. Ainsi l'utilisation des 4 premiers moments (**O**) sur les 4 canaux donne un score F1 de 61%, soit une réduction de ~ 20% par rapport à l'utilisation de l'histogramme complet. Le meilleur score est obtenu en utilisant la combinaison de la moyenne et de l'écart type (**N**) de chaque canal. Il est intéressant de noter qu'avec seulement les deux premiers moments calculés




**(a)** Comparaison de la moyenne pour chaque échantillon des quatre canaux considérés. La première ligne de graphiques correspond à la classe *humide* et la seconde à la classe *non-humide*.




**(b)** Comparaison de l'écart type pour chaque échantillon des quatre canaux considérés. La première ligne de graphiques correspond à la classe *humide* et la seconde à la classe *non-humide*.

**Figure 4.6 –** Comparaison de l'histogramme 2D pour la moyenne et l'écart type de chaque échantillon, pour les deux classes *humide* et *non humide*, et entre les différents canaux pour ($VV$,$VH$ $R_VV$ et $R_VH$). $R_VV$ et $R_VH$ représentent le ratio entre l'acquisition actuelle et l'image de référence décrite dans 3.2.2 respectivement en polarisation $VV$ et $VH$.

pour chaque canal, nous obtenons ~ 99% du score F1 obtenu en utilisant l'histogramme complet sur chaque canal.

Nous observons que les moments d'ordre supérieur à 2 réduisent drastiquement le score F1. Cela peut être dû à deux facteurs. Le premier est l'imprécision qui augmente lorsqu'on essaie d'estimer des moments d'ordres supérieurs. Le second est dû à la nature des distributions des deux classes, où le caractère gaussien prend le dessus, donnant lieu à des performances supérieures uniquement grâce aux deux premiers moments.

Pour étudier cette hypothèse, nous proposons dans la figure 4.6 une illustration des distributions jointes de chacune des 6 combinaisons possibles pour les 4 canaux en fonction de la moyenne et de l'écart type, c'est-à-dire l'histogramme 2D pour la moyenne de l'échantillon (4.6a) et l'écart type de l'échantillon (4.6b). Il est important de noter que la couleur est en échelle logarithmique.

Nous pouvons observer que les distributions associées à la moyenne des échantillons sont plus distinctes pour les deux classes que celles associées à l'écart type. Cependant, ces dernières montrent des différences significatives de forme (**R_VV** versus **VV**) et même de mode (**R_VV** versus **R_VH**). Les distributions liées à la moyenne sont plus différentes entre les deux classes : alors que la classe sèche semble fortement gaussienne, ce qui peut s'expliquer

par un moyennage de 256 pixels de résolution moyenne suivant des distributions quasi gaussiennes, celle de la classe humide fait apparaître un aspect bimodal.

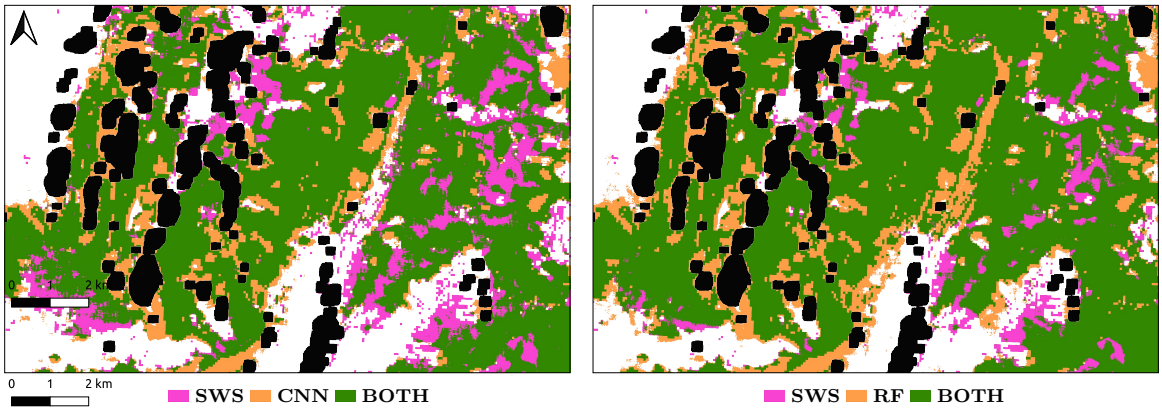
Il existe des différences marquées dans le mode de certaines des distributions (**R_VV** versus **R_VH**), mais surtout nous pouvons observer un aspect bimodal, surtout dans les distributions entre les canaux polarimétriques primaires (VV,VH) et les ratios. Nous avons analysé la composition des modes présents dans la distribution formée par les canaux **VH** et **R_VV**. Nous avons sélectionné 5000 échantillons de la classe humide autour des 2 modes principaux $M_{b1}$ et $M_{b2}$ dont les barycentres sont situés respectivement autour de $(x_{M_{b1}}, y_{M_{b1}}) = (\sim 0.3, \sim 0.42)$ et $(x_{M_{b2}}, y_{M_{b2}}) = (\sim 0.4, \sim 0.48)$. La première différence entre les échantillons est la date d'acquisition. Alors que les échantillons composant le premier mode $M_{b1}$ sont répartis de manière homogène entre février et juin, les échantillons du second mode $M_{b2}$ sont principalement concentrés en février. La deuxième distinction entre les deux modes est la position géographique des échantillons. Nous notons que les échantillons appartenant au mode $M_{b1}$ sont principalement situés dans des zones rocheuses, sur des plaines ou des terrains nus. En revanche, les échantillons du mode $M_{b2}$ sont principalement situés dans des zones forestières. Il est intéressant de noter que la position de $M_{b2}$ est très proche du mode unique pour les échantillons de la classe non humide. Ceci peut s'expliquer par la réduction significative de la pénétration de l'onde en présence de forêts. Cela rend difficile la distinction fiable entre les surfaces couvertes de neige humide et celles qui ne le sont pas, lorsqu'on considère la moyenne du signal sur la cellule de résolution. Cependant, la présence de tels échantillons peut permettre l'utilisation de l'histogramme pour différencier certaines zones avec une faible couverture forestière.

## 4.3 Résultats qualitatifs et analyse temporelle

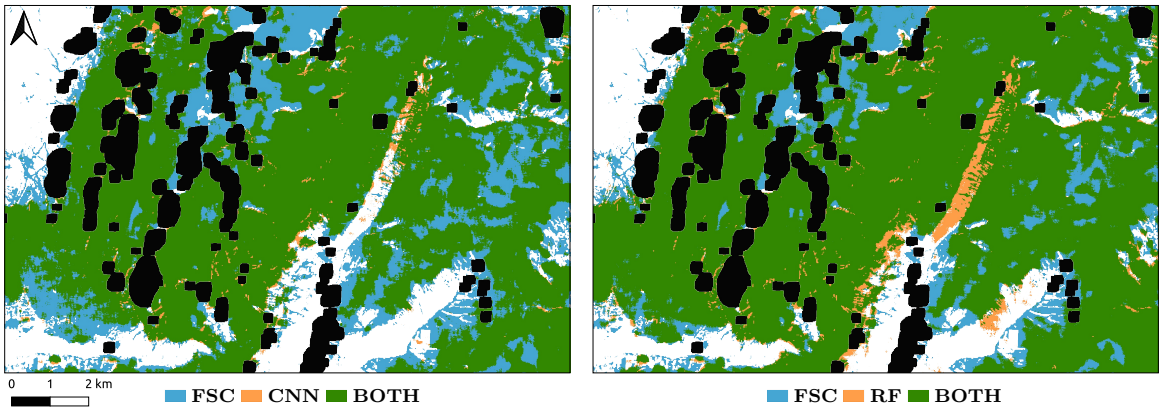
Nous proposons dans cette section d'évaluer les performances d'une sélection de modèles et de bandes, par l'utilisation des histogrammes, sur la qualité des cartes de neige humide estimées en comparaison avec les produits Copernicus FSC et SWS et leur évolution temporelle. La zone d'évaluation et la position des stations de mesure sont présentées à la figure 3.15.

### 4.3.1 Cartographie en période de fonte

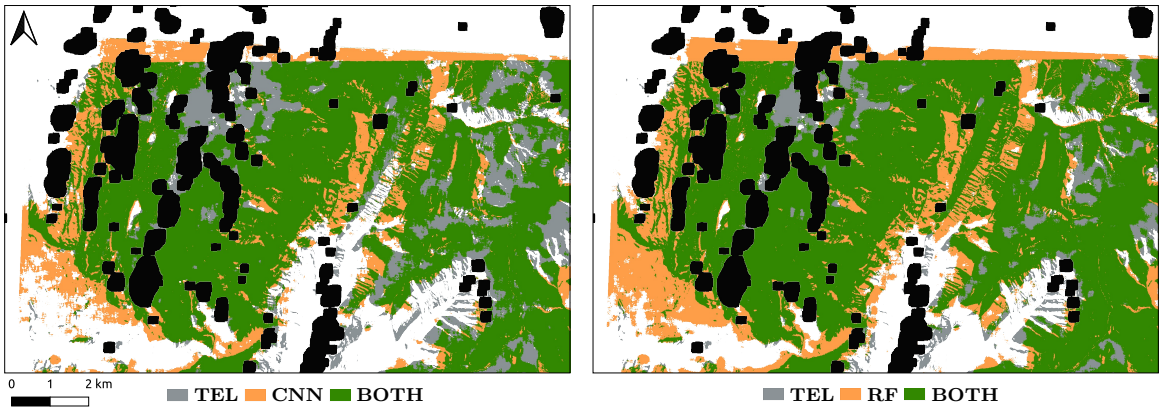
Nous évaluons dans un premier temps, les cartes estimées pour une acquisition en période de fonte, le 31 mars 2021. À cette période, la fonte a déjà commencé et le manteau considéré comme humide est proche de l'extension maximale de la couverture totale neigeuse. Le produit SWS de Copernicus est obtenu pour cette même date, tandis que le produit FSC est obtenu pour la date la plus proche soit le 30 mars 2021. Pour comparer les résultats avec le produit FSC, nous avons seuillé ce dernier et considéré tous les pixels ayant une fraction supérieure à 0%. La figure 4.7 présente les résultats obtenus par le RF et le CNN sur la combinaison **I** pour le 31 mars 2021. Les résultats sont ceux obtenus par le seuillage BAROC. Les résultats sont comparés au produit SWS (figure 4.7a), au produit FSC (figure 4.7b) et à la règle de labellisation basée sur CROCUS ( 4.7c). Dans les 6 cartes présentées, la zone en vert correspond à la correspondance entre la détection de neige humide par le modèle considéré et le produit auquel il est comparé. En orange les zones seulement détectées par




**(a)** Comparaison entre les résultats de classification obtenus par RF et CNN et les produits Copernicus SWS




**(b)** Comparaison entre les résultats de classification obtenus par RF et CNN et les produits Copernicus FSC (seuillage >0%)




**(c)** Comparaison entre les résultats de classification obtenus par RF et CNN et la spatialisation de la règle de CROCUS (cf (3.22))

**Figure 4.7 –** Cartes de neige humide pour le 31 mars 2021, obtenues avec le RF (à droite) et CNN (à gauche) sur la combinaison **I** et en comparant avec les produits Copernicus FSC et SWS et la règle de labellisation à partir du modèle CROCUS, pour une portion du massif des Grandes Rousses

un des deux modèles, CNN ou RF (respectivement les cartes de gauche et de droite). En noir, les zones de distorsions géométriques exclues de l'évaluation et en blanc les zones sans neige.

Globalement, les cartes obtenues par les deux méthodes sont proches. Le CNN va être plus détaillé sur les contours des zones, quitte à être plus hétérogène. Le RF va être plus lisse, mais va avoir une tendance à surestimer les zones de neige humide. En comparant les modèles au produit SWS, nous pouvons remarquer que les 2 modèles couvrent les mêmes zones, avec une sous-estimation sur la partie droite de la carte. En revanche le RF sur estime une zone linéaire en diagonale de l'image. Cette zone représente une barre de roche sous une crête. Une hypothèse plausible est que la réponse de cette zone peut être due à une rugosité plus grande due à la présence de roche et à l'affleurement de roche ou à une zone qui peut être presque considérée comme une distorsion de type ombre. La sur estimation des modèles par rapport à SWS est justifiée par une corrélation plus forte entre les modèles et le FSC, notamment en bordure de zones détectées, soit au niveau des lignes de fontes. Ainsi les modèles ont pu capturer plus de caractéristiques fines sur la nature de la neige pour approcher la couverture totale de neige, même si elle n'est pas encore complètement humide. Enfin la comparaison avec la règle de labellisation montre une zone (en bas à gauche) détectée par les modèles et non par la règle. Cette zone correspond à une zone urbaine avec des pentes faibles, qui ont pu être capturées par les modèles en dépit d'une mauvaise labellisation puisqu'étant vue par les deux produits Copernicus.

### 4.3.2 Comparaison entre la période de fonte et de début de saison

Nous comparons ici les résultats obtenus par le meilleur classifieur (KNN) et pour la meilleure combinaison (**I**), et ce pour deux dates, le 18 janvier 2021 et le 31 mars 2021. Sur la base de cette combinaison, l'ensemble des données d'entraînement nous permet de calculer les seuils BAROC et FCROC. Traditionnellement, les sorties du classifieur sont seuillées à 0.5, les valeurs supérieures étant considérées comme appartenant à la classe en question. Il est proposé d'ajouter les deux seuils supplémentaires BAROC et FCROC. La précision maximale est atteinte à 79.7% avec un seuil BAROC de 0.48. Nous avons obtenu une précision de 75.2% avec un taux de faux positifs constant de 5%, pour un seuil FCROC de 0.72. Les résultats de classification sont effectués à l'aide d'une fenêtre glissante. La valeur de classification du pixel central est respectivement 0,1,2 ou 3 s'il n'y a pas de neige humide, de la neige humide obtenue par seuillage BAROC, de la neige humide par seuillage classique à 0,5 et de la neige humide par seuillage FCROC. Les résultats sont donnés dans la figure 4.8. Nous remarquons que l'ajout du seuil BAROC n'apporte aucune information aux résultats obtenus avec le KNN par rapport au seuillage habituel (0.5). Cependant, nous pouvons voir l'avantage d'utiliser les seuils ensemble, avec des valeurs de pixels résultantes plus élevées dans les zones où le seuil au-dessus duquel le classifieur d'entraînement a fait peu de mauvaises classifications. Nous devons être prudents lors de la comparaison directe des résultats du 18 janvier 2021. En effet, à cette date, le modèle CROCUS classe environ 5% des pixels de cette zone comme de la neige humide selon le critère proposé. Le produit FSC ne détaille pas la différence entre la neige humide et la neige sèche et donne le résultat global de la couverture de neige, tandis que le produit SWS, en raison de son seuil,

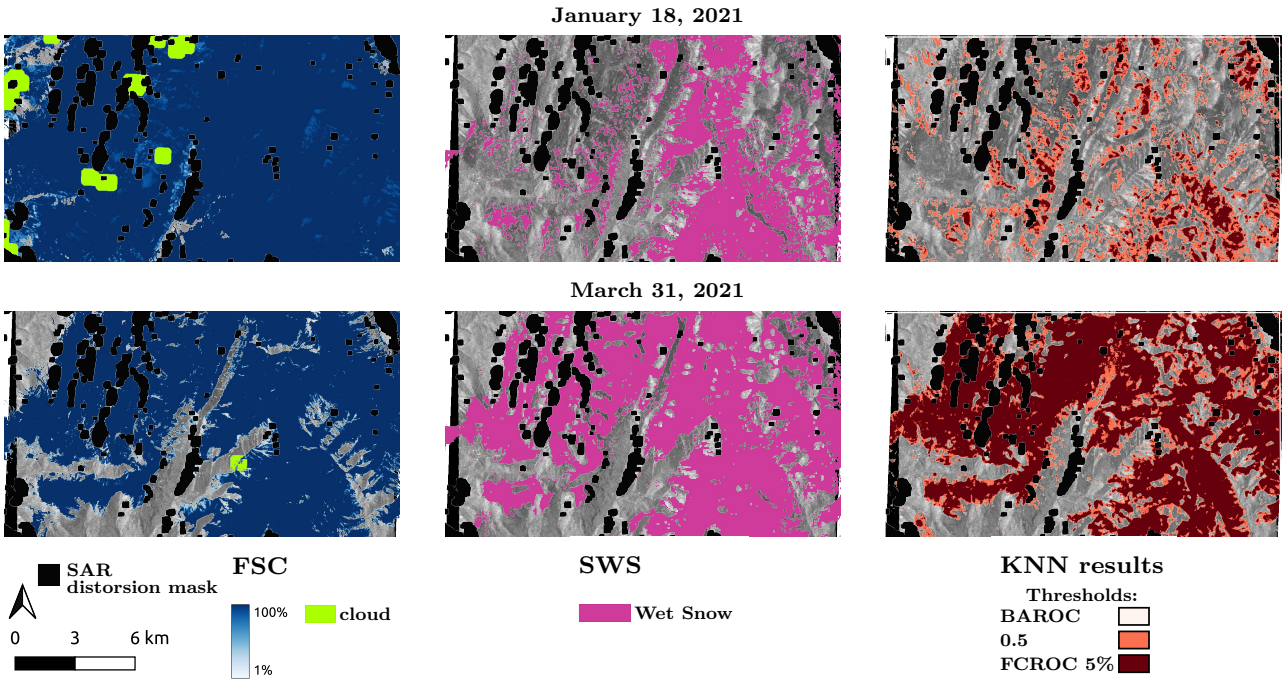



**Figure 4.8 –** Cartes des résultats de classification de la neige pour deux dates en comparaison avec les produits Copernicus FSC et SWS

a tendance à surestimer notre critère. À cette date, le résultat obtenu montre que de nombreuses zones sont encore peu validées par le critère du FCROC. À la date du 31 mars 2021, cette fois nous sommes dans la période de fonte et nous pouvons voir une corrélation plus marquée entre les 3 résultats. Là où le produit SWS donne de grandes zones homogènes, nous obtenons des zones plus hétérogènes, mais donnant lieu à un découpage plus fin des lignes de fontes. Nous retrouvons le comportement étudié dans la sous section 4.3.1. La figure 4.9 représente un zoom de la figure 4.8 sur les Alpes d'Huez pour le 31 mars 2021. Cette zone est hétérogène par la présence de forêts, de zones urbaines et de plaines. Pour les produits SWS, nous avons ajouté les catégories de masques de produit. La construction du produit SWS se fait en masquant à la fois les zones de ville et de forêt. De manière intéressante, le résultat du KNN offre une segmentation plus fine de la neige humide. Nous présentons ici les résultats pour la combinaison **I** et **J**. Ce détail montre une bonne homogénéité des résultats avec une surestimation avec la combinaison **J**, qui peut s'expliquer par l'utilisation de données topographiques non pertinentes. La bordure des zones détectées est dans les deux cas progressive, par la présence de zone seuillées à 0.5.

Il est important de noter que l'information utilisée pour l'étiquetage est imparfaite. Cependant, la diversité des échantillons permet une bonne généralisation de l'apprentissage. Nous notons que les cartes créées à l'aide de l'algorithme KNN ont un degré d'hétérogénéité plus élevé que les produits FSC et SWS. Cependant, la résolution du SWS en particulier est de 60m par pixel, alors que nous proposons des cartes avec 10m par pixel, induisant un aspect visuel plus détaillé sur les frontières des cartes, tel que décrit en amont. Il peut être intéressant de réaliser une analyse dans des travaux ultérieurs sur les scores et la qualité de la classification en fonction du type de sol. Nous pouvons noter que la solution utilisant tous les canaux (**J**) a tendance à surestimer les zones considérées comme de la neige humide. Cela est clairement visible sur la figure 4.9, où des zones qui ne sont même pas détectées par le produit FSC

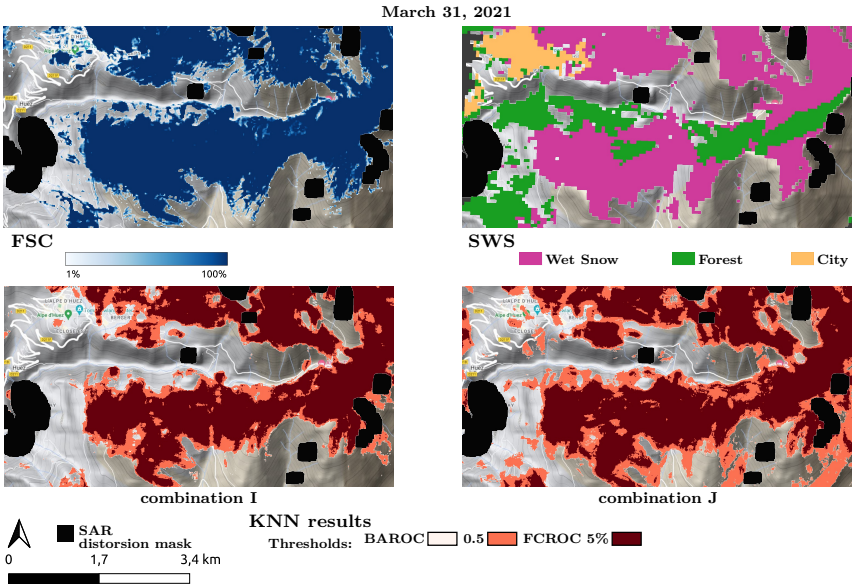



**Figure 4.9 –** Détails de la carte et vue agrandie des résultats de classification de la neige pour les deux meilleures combinaisons **I** et **J** comparant les produits Copernicus FSC et SWS dans l'Alpe d'Huez, France pendant la période de fonte le 31 mars 2021

comme ayant de la neige sont détectées comme de la neige humide. Cependant les zones de ville sont correctement prises en compte par rapport au produit SWS, tout comme les zones de forêt.

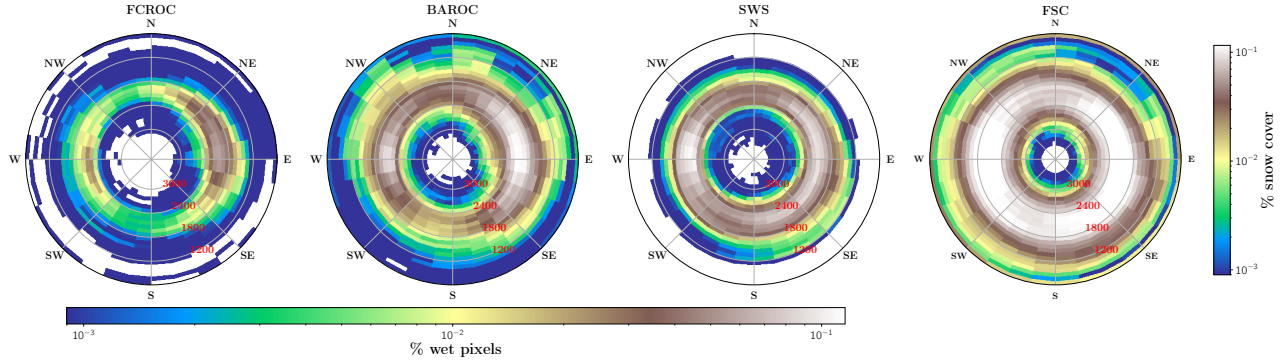
### 4.3.3 Répartition topographique de la neige humide

Afin d'observer la capacité de généralisation du modèle KNN à correctement se transférer, nous avons évalué la distribution des pixels considérés comme neige humide pour les paramètres topographiques d'orientation et d'altitude, pour les deux dates (31 mars 2021 et 18 janvier 2021). Nous avons considéré un pas de 100m pour l'altitude et 12.5° pour l'orientation. Pour simplifier l'analyse, nous n'avons pas pris en compte la variable de pente, et avons calculé cette distribution sur toutes les pentes inférieures à 45°. Cette évaluation a été réalisée sur les mêmes résultats que dans la section 4.3.4, c'est-à-dire KNN avec la combinaison de bandes **I**. Nous avons étudié séparément les résultats obtenus par le seuillage FCROC, celui de BAROC, et comparé avec la répartition topographique calculée pour les produits SWS. Nous avons également comparé la distribution de la couverture de neige du produit FSC[69] . La représentation à la figure 4.10 donne en pourcentage pour chaque cellule sur l'ensemble de la zone considérée sans nuages pour le FSC et sans distorsion géométrique.

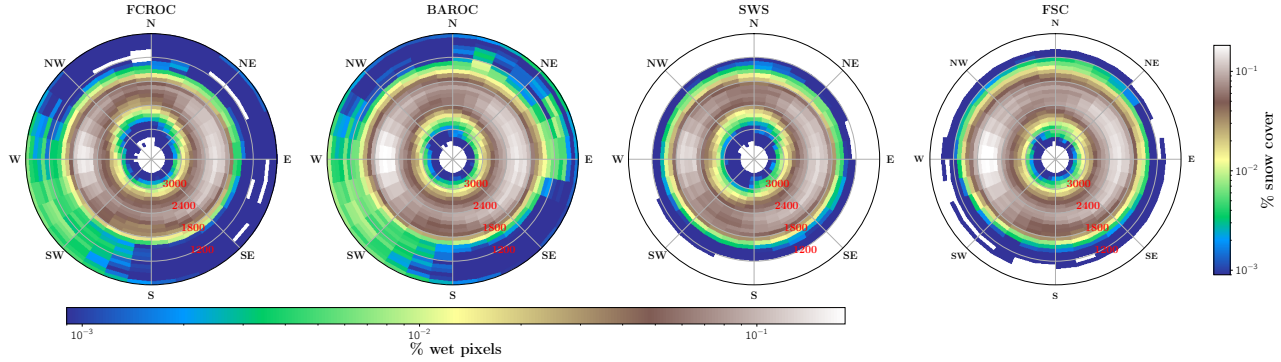
**Précision (69)**
Comme précédemment nous considérons la version seuillée, avec tous les pixels ayant une fraction strictement au dessus 0%. Cette proposition est discutable puisqu'elle surestime les pixels enneigés, là où notre règle exclue les zones n'ayant pas au moins 40cm de neige donné par le modèle CROCUS

Pour la date du 18 janvier 2021, nous constatons que les trois résultats basés sur le SAR et donc sensibles à la neige humide, montrent une répartition plus localisée et moins homogène que l'estimation de l'extension de la neige basée sur l'optique. Nous pouvons observer que le FCROC, BAROC et SWS donnent un maximum de neige humide pour des altitudes entre 1800m et 2400m face à l'Est.

Pour la date du 31 mars 2021, nous remarquons que le diagramme du seuil BAROC est plus proche de celui du FSC que le SWS, notamment en termes de pourcentage, nous trouvons les 2 modes principaux pour des altitudes entre 1800 et 2400m avec une orientation Est et Ouest, et cela de manière marquée




**(a)** 18 janvier 2021




**(b)** 31 mars 2021

**Figure 4.10 –** Diagrammes de la distribution des pixels humides (FRCOC, BAROC, SWS) et de la couverture neigeuse (FSC) en fonction de l'altitude et de l'orientation pour une pente inférieure à 45°, pour deux dates

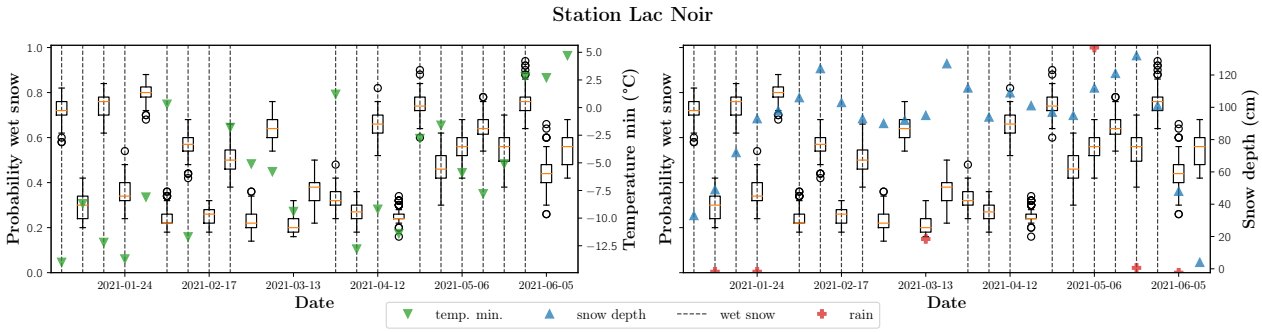
dans les 2 diagrammes, alors que le SWS est légèrement plus homogène sur toutes les orientations.

Un certain nombre de pixels sont détectés comme étant de la neige humide pour des altitudes inférieures à 1800m pour des orientations entre l'Ouest et le Sud-Ouest, mais seulement dans les résultats FCROC et BAROC. Ces zones correspondent à la détection de 2 lacs dans la partie sud du massif des Grandes Rousses, ces zones n'ayant pas été masquées dans l'apprentissage sont considérées comme de la neige humide, à l'état extrême de la fonte.

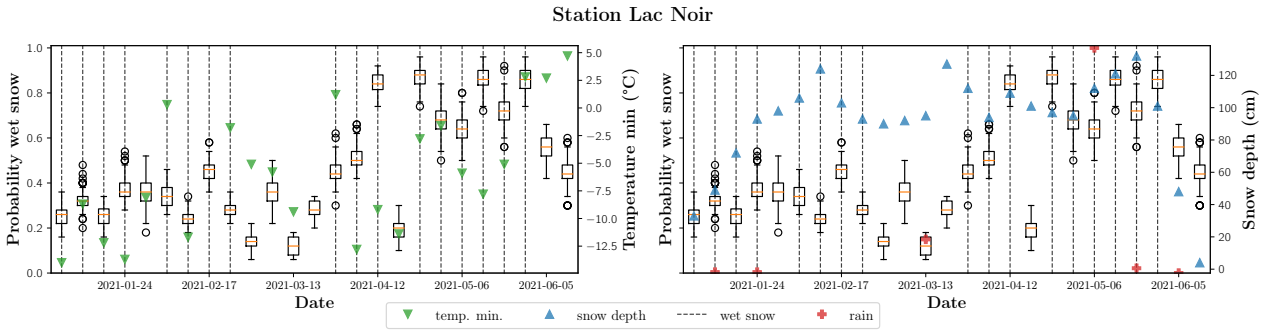
### 4.3.4 Comparaison in-situ

Nous proposons une comparaison avec trois stations de mesure qui donnent la hauteur de neige et la température minimale. Mais cela reste une comparaison et non une validation des informations in situ auxquelles nous avons accès car nous n'avons pas accès aux mêmes informations. Nous nous référons au rapport d'évaluation de la qualité du produit des zones humides de Copernicus, qui couvre la majeure partie de l'Europe, où l'évaluation est réalisée par comparaison avec les résultats obtenus par Landsat8 pendant la période de fonte, de manière similaire aux résultats obtenus par le produit FSC. De plus, le rapport inclut une comparaison avec une seule station de mesure in situ (Conny Alm) fournissant des informations sur la hauteur de neige et la température de l'air. Nous considérons dans un premier temps la station météorologique Lac Noir, fournissant les mesures in-situ quotidiennes suivantes : profondeur de la neige (cm), température minimale quotidienne (degrés Celsius) et précipitations (mm). Pour faciliter la lecture des figures,

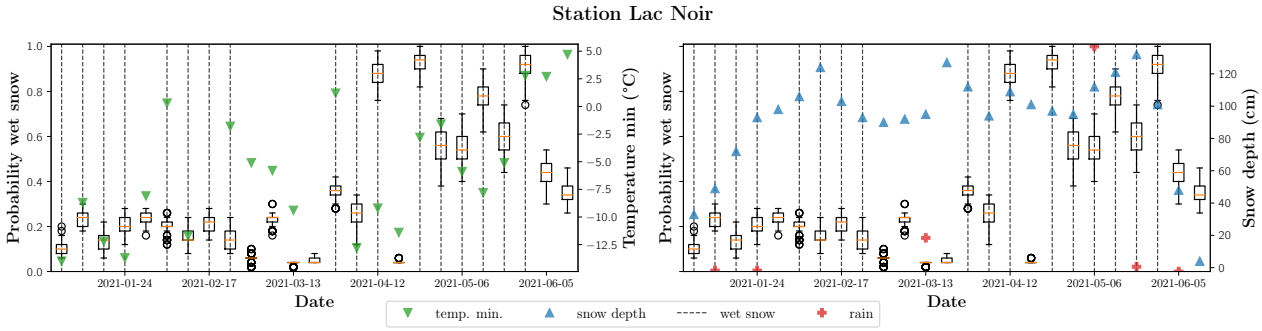
nous avons normalisé les précipitations entre 0 et 1, où 1 est la précipitation maximale pendant la période considérée (correspondant à une précipitation de 23,5mm). Les figures 4.11a, 4.11b montrent la variation temporelle de




**(a)** Prédiction du KNN en utilisant les combinaisons de canaux **A**.




**(b)** Prédiction du KNN en utilisant les combinaisons de canaux **I**.




**(c)** Prédiction du KNN en utilisant les combinaisons de canaux **J**.

**Figure 4.11 –** Comparaison des mesures de la station Lac Noir avec la prédiction du KNN.

la prédiction de neige humide par la méthode KNN, en fonction des trois informations in situ pour trois combinaisons de canaux (**A, I, J**). Dans cette comparaison, nous avons directement utilisé les valeurs de probabilité de neige humide plutôt que les sorties seuillées. La fenêtre temporelle est entre janvier 2020 et début juin 2020. La prédiction est faite sur une fenêtre de 2 par 2 pixels centrée sur la station. Nous avons ajouté aux figures les dates où le produit SWS de Copernicus indiquait la présence de neige humide. Il est intéressant de noter que la combinaison **A** donne des résultats assez homogènes sans tendances particulières. Si nous fixons le seuil de probabilité à 0.5, nous retrouvons l'information donnée par le produit SWS. Il est difficile de voir une corrélation claire entre ces données et les mesures in-situ. À partir d'avril, nous remarquons qu'une tendance se dessine et le nombre de dates

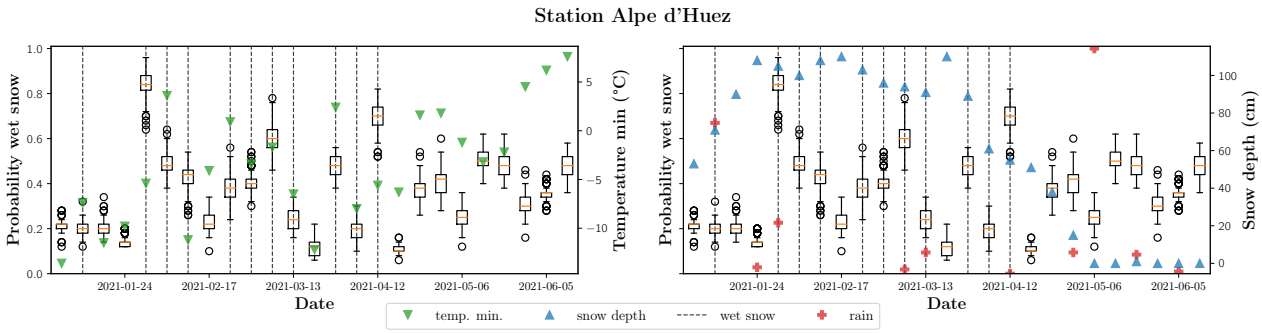
et la probabilité associée de neige humide augmentent. Ceci s'explique par l'augmentation progressive de la température du manteau neigeux et le début de la fonte des neiges. La figure 4.11b donne la variation pour la combinaison donnant les meilleurs résultats dans les tests proposés dans ce manuscrit. Il y a clairement deux régimes, un avant mi-mars lorsque la probabilité est faible même lorsqu'il y a une augmentation soudaine de la température. L'autre régime, après mi-mars, montre une augmentation de la probabilité de détection, corrélée à l'augmentation permanente de la température. Cette probabilité est supérieure à celle donnée par la combinaison **A**. Dans tous les scénarios, la baisse de la probabilité de neige humide au début de juin s'explique par une réduction significative du manteau neigeux, à la limite ou même en dessous du seuil d'étiquetage (40cm).

Les résultats présentés à la figure 4.11b sont intéressants car la méthode Copernicus montre clairement que la neige est détectée tôt dans la saison où, pour une profondeur de neige de plus de 40cm, la température minimale est basse, principalement en dessous de 0 voire $-2.5°C$, alors que nos propositions donnent des taux de probabilité faibles. Cela peut s'expliquer par l'impact du gel sur le signal radar. Des études générales ont montré la sensibilité du signal radar en bande C à la détection du sol gelé [353], provoquant une baisse du signal rétrodiffusé de manière similaire à la conséquence de la présence de neige humide [354]. Nous pouvons supposer que la méthode de seuillage pour les produits Copernicus est sensible à ce gel et le détecte comme de la neige humide, alors que notre méthode, en apprenant à partir d'un grand nombre de caractéristiques associées à des données d'entraînement riches, ne considère pas ces zones comme de la neige humide. Les résultats présentés à la figure 4.11c montrent quant à eux une baisse sur l'incertitude des moments de présence de neige humide en période de neige fraîche jusqu'au mois de mars, avec des probabilités inférieures à 25%. Cependant, nous observons une plus grande variabilité en période de fonte avec des probabilités légèrement plus basses que pour la combinaison **I** (environ 10% de moins). Cela peut s'expliquer par l'ajout de canaux non pertinents pour la détection de la neige humide, qui peuvent ajouter du bruit à la classification et ne se basent plus uniquement que sur une information de rétrodiffusion.

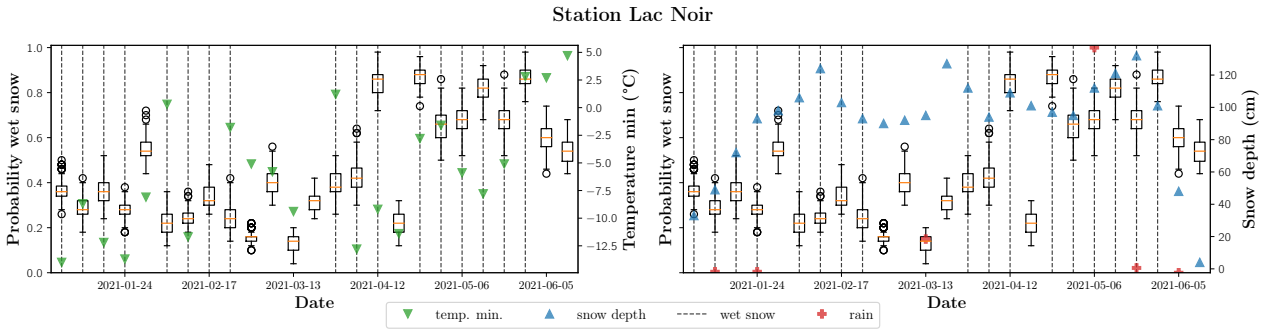
[353]: Jagdhuber et al. (2014), « Identification of Soil Freezing and Thawing States Using SAR Polarimetry at C-Band »

[354]: Baghdadi et al. (2018), « Detection of Frozen Soil Using Sentinel-1 SAR Data »

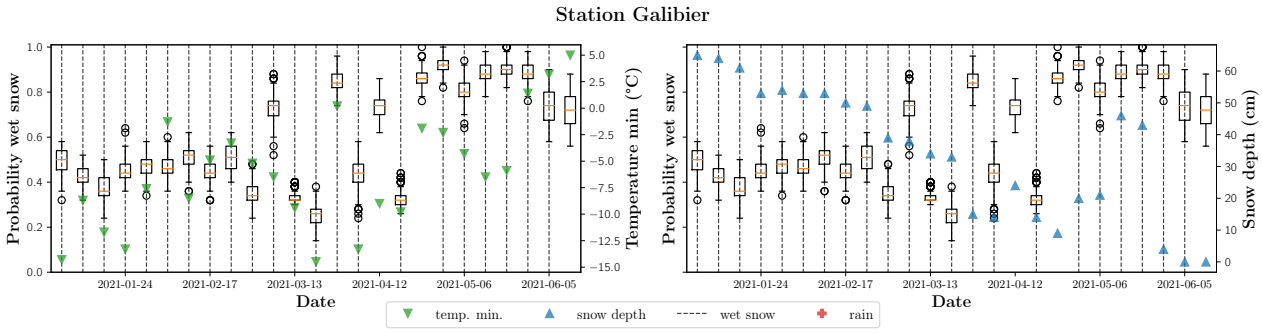
La figure 4.12 montre pour la combinaison **H**, soit les 2 canaux de polarisations et les 2 canaux de ratio de type Nagler, l'évolution temporelle pour les trois stations météos de la zone des Grandes Rousses. La zone de la station de l'Alpe d'Huez (figure 4.12a) présente la forte variabilité et une difficulté à déterminer les moments de neige humide. Avant le 12 avril, nous pouvons considérer les échantillons comme ayant un seuil de probabilité supérieur à 40% comme étant de la neige humide pour être en adéquation avec le produit SWS (ligne pointillée). Cependant à partir du 06 mai il n'y a plus de neige mais la probabilité est toujours de 40%. Ceci peut s'expliquer par la position de cette station, à proximité de zones urbaines. La figure 4.12b montre la même tendance que celle observée dans la figure 4.11b. Il est intéressant de noter une différence entre les deux combinaisons sur les deux dernières acquisitions. L'ajout de la pente (combinaison **I**) permet de diminuer la probabilité de détection de neige humide lorsque la hauteur de neige est nulle. Sur la zone de la station du Galibier (figure 4.12c), nous pouvons observer que la plupart des acquisitions sont considérées comme humide par SWS, et ce dès le début de la saison. Au vu de la hauteur de neige (plus de 50cm) et de la température de l'air ($\sim -15°C$), le phénomène observé est principalement dû à un gel du manteau neigeux plutôt qu'à une humidification. Ainsi ce cas montre la supériorité de




**(a)** Prédiction du KNN sur la station de l'Alpe d'Huez.




**(b)** Prédiction du KNN sur la station du Lac Noir.




**(c)** Prédiction du KNN sur la station du Galibier.

**Figure 4.12 –** Comparaison des mesures pour les combinaisons de canaux **H**,pour trois stations de mesures, Alpe d'Huez, Lac Noir et Galibier.

notre apprentissage permettant de rejeter des scénarios dans lesquels d'autres phénomènes que l'humidification font baisser la rétrodiffusion. Il est toutefois intéressant de noter que nous retrouvons cette probabilité de détection de neige humide en fin de saison, et ce sans neige. Ceci peut être du à la présence d'humidité au sol qui fait diminuer la rétrodiffusion.

## 4.4 Conclusion

Nous avons présenté dans ce chapitre notre troisième contribution, portant sur l'analyse de la détection de la neige humide par apprentissage automatique. Nous avons pu étudier la pertinence du jeu de données LSD4WSD pour cette tâche et nous avons évalué 7 algorithmes différents. Parmi ces algorithmes, tous ont démontré de bonnes performances, avec une détection supérieure à la règle de Nagler (dans le cadre de notre labellisation). Nous avons testé l'impact de

l'utilisation de différentes caractéristiques autres que celles polarimétriques, et avons observé que les caractéristiques topographiques n'apportent pas de gain significatif, à l'exception de la pente. L'utilisation du ratio de Nagler permet d'améliorer significativement la détection (de l'ordre de 4% en moyenne sur les algorithmes) par rapport à l'utilisation des seuls canaux VV et VH. Nous avons pu comparer spatialement les résultats de détection obtenus par une sélection des modèles et de caractéristiques optimales sur une zone de validation du massif des Grandes Rousses, et avons observé une forte corrélation avec les produits de Copernicus (optique et SAR), avec un détail plus fin sur les zones de transition pour les différentes approches testées. Il est toutefois important de noter que la variabilité entre les massifs montagneux est très importante, et que la généralisation peut être difficile. Il pourrait être intéressant de comparer les performances des modèles sur des sous ensembles de données présentant des caractéristiques topographiques plus proches. Enfin, nous avons comparé la probabilité de détection de la neige humide au cours du temps, et ce sur trois zones d'étude différentes dotées de stations de mesures, avec le résultat du produit de Copernicus (SWS), la hauteur de neige et la température de l'air mesurée par les stations. Ces résultats montrent une corrélation entre la probabilité de détection et les mesures de température, et une tendance à la sous-estimation de la neige humide par les modèles, notamment en début de saison.

Nous nous sommes placés dans ce chapitre sur l'évaluation de méthodes de détection de la neige humide sur notre dataset, néanmoins une perspective naturelle est la classification multi-classes. En effet il pourrait être intéressant de distinguer des sous classes d'humidité de la neige, autrement dit de remonter à une information de TEL. Cette problématique peut s'aborder de deux manières : soit en cherchant à modéliser plus finement le comportement des différentes classes par des distributions de probabilité, soit en cherchant à affiner les caractéristiques utilisées pour la classification. Dans le premier cas, nous avons vu que la moyenne et l'écart type des canaux radar sont des caractéristiques pertinentes pour la détection de la neige humide, l'utilisation de modèles statistiques multivariables pourrait être une piste intéressante, bien que l'on observe un comportement multi-modal sur les histogrammes de ces caractéristiques. Dans le second cas, l'apprentissage de décompositions pourrait permettre d'extraire des caractéristiques plus discriminantes.

Cependant, de tous ces algorithmes et de toutes ces combinaisons, nous pouvons nous interroger sur le choix le plus judicieux pour la détection de la neige humide. En effet, nous avons pu observer que spatialement des différences de comportement existent entre les modèles. Il est donc naturel de se demander si une combinaison de ces modèles pourrait améliorer la détection. Nous proposons dans le chapitre suivant d'étudier cette question en utilisant des méthodes d'ensemble et de logique floue pour agréger les sorties de modèles.

# Méthodes d'ensemble floue et application pour la cryosphère

# 5


**5.1 Contexte: logique floue et opérateurs d'agrégation . . . . . 101**
**5.2 Apprentissage flou: intégrale de Choquet et pré-agrégation . . . . . 104**
**5.3 Expérimentations . . . . . 107**
**5.4 Résultats quantitatifs sur l'apprentissage flou . . . . . 111**
**5.5 Spatialisation des résultats . . . . . 114**
**5.6 Conclusion . . . . . 116**


Nous avons étudié dans le chapitre Chapitre 4 l'utilisation de nombreux algorithmes de classification pour la détection de la neige humide, et avons remarqué l'homogénéité des performances de ces modèles. Malgré ces performances similaires, nous avons pu observer des différences notables dans les cartes de prédiction. De plus, nous avons évoqué le principal obstacle à cette détection, à savoir la confiance et la validité du label associé. Ainsi, il est naturel de se demander si l'agrégation de ces modèles peut améliorer les performances de détection, et ce, dans un contexte où les labels sont bruités. Nous proposons dans ce chapitre de nous intéresser à l'agrégation de forêt de classifieurs. Cette nouvelle contribution consiste à combiner des classifieurs sans a priori statistiques en prenant en compte la variabilité des labels et le bruit des images par l'utilisation de la logique floue. Nous évaluons ensuite cette proposition sur des données SAR, dans le cadre de la distinction entre les zones glaciaires d'accumulation et d'ablation.

Comme décrit dans la sous-section 3.1.3, dans la partie liée à l'agrégation, l'utilisation de la logique floue est particulièrement pertinente pour l'agrégation d'informations incertaines, dont l'intégrale de Choquet est un exemple pour la fusion. Nous proposons ici une utilisation conjointe de la logique floue et de l'apprentissage automatique pour l'agrégation de forêts de classifieurs.

## 5.1 Contexte : logique floue et opérateurs d'agrégation

Dans cette section, nous rappelons la définition des mesures floues, de l'intégrale de Choquet et des opérateurs d'agrégation T-norme. Nous présentons également les mesures floues et les T-normes sélectionnées dans le cadre proposé.

### 5.1.1 Mesures floues

**Definition 5.1.1** *Soit $X = \{x_1, x_2, \dots x_m\}$ un ensemble fini. Une fonction d'ensemble $\mu : \mathcal{P}(X) \rightarrow [0, 1]$ est appelée mesure floue si elle respecte (voir [287] par exemple) :*

[287]: Campos et al. (1992), « Characterization and comparison of sugeno and choquet integrals »

- $\forall\, a, b \subseteq X$, *si* $a \subseteq b$ *alors* $\mu(a) \leq \mu(b)$ *(monotonie),*
- $\mu(\varnothing) = 0$ *et* $\mu(X) = 1$ *(bornes).*

Pour un ensemble $\Omega \subseteq X$, de nombreuses mesures floues existent telles que la moyenne pondérée ordonnée [355], ou la mesure uniforme. Dans ce qui suit, nous considérons les deux mesures suivantes :

[355]: Yager (1993), « On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decisionmaking »

- mesure de puissance :

$$\left(\frac{|\Omega|}{m}\right)^q \text{ avec } q > 0, \tag{5.1}$$

- moyenne pondérée :

$$\sum_{i\in\Omega} w_i \text{ avec } \sum_{i=1}^{m} w_i = 1. \tag{5.2}$$

Le symbole $|\cdot|$ désigne la cardinalité de l'ensemble $\Omega$ et la variable $m$ représente le nombre total d'éléments de l'ensemble $X$. Nous sélectionnons la mesure de puissance en raison de son évaluation dans [295], du faible nombre de paramètres (un seul) et de certains travaux sur l'apprentissage du paramètre [356]. La moyenne pondérée est une mesure intéressante par son interprétabilité concernant les contributions au résultat final de l'agrégation. De plus, son comportement est proche des méthodes d'agrégation traditionnelles telles que la couche dense[70] dans les réseaux neuronaux.

[295]: Lucca et al. (2016), « Preaggregation Functions »

[356]: Tartas et al. (2013), « Using the Choquet Integral in the Fuzzy Reasoning Method of Fuzzy Rule-Based Classification Systems »

**Précision (70)**
La couche dense calcule une somme pondérée des entrées plus un biais, suivie d'une fonction d'activation. Les poids ne sont pas obligés de sommer à 1, et les poids et le biais peuvent être positifs ou négatifs.

### 5.1.2 Intégrale de Choquet

Nous présentons l'intégrale de Choquet, considérée comme une intégrale non additive, généralisant la notion de valeur attendue. L'intégrale de Choquet englobe plusieurs techniques d'agrégation conventionnelles, y compris la moyenne arithmétique, la moyenne pondérée et la moyenne ordonnée pondérée, lorsqu'on utilise une mesure floue du type moyenne pondérée comme indiqué dans l'équation (5.2). Cette méthode d'intégration non linéaire permet de considérer les interrelations entre divers classifieurs.

L'intégrale de Choquet est définie comme suit :

**Definition 5.1.2** *Soit* $\mu$ *une mesure floue telle que* $\mu : \mathcal{P}(X) \to [0,1]$. *L'intégrale de Choquet discrète* $C_\mu$ *de* $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_m) \in [0,1]^m$ *par rapport à* $\mu$ *est définie comme suit :*

$$C_\mu(\mathbf{x}) = \sum_{k=1}^{m} \mathbf{p}_k \cdot \mu(\Omega_k), \tag{5.3}$$

*avec* $\mathbf{p}$ *le vecteur ordonné et différencié de* $\mathbf{x}$, *c'est-à-dire* $\mathbf{p} = [\mathbf{x}_1 - \mathbf{x}_0, \ldots, \mathbf{x}_m - \mathbf{x}_{m-1}]$ *; où* $\mathbf{x}_0 \leq \mathbf{x}_1 \leq \cdots \leq \mathbf{x}_m$. *Par convention, nous posons* $\mathbf{x}_0 = 0$. $\Omega_k$ *représente l'ensemble des indices k jusqu'à m et correspond au sous-ensemble des indices pour les* $m - k + 1$ *plus grands éléments de* $\mathbf{x}$.

**Example 5.1.1** Soit $\mathbf{x} = [0.23, 0.12, 0.45]$ et $\mu$ une mesure de puissance avec

**Table 5.1 –** Détails des T-normes paramétrées sélectionnées et de leurs domaines de définition

| Forme | $T_\alpha(x,y)$ | domaine de définition de $\alpha$ | Forme spécifique |
|---|---|---|---|
| $T_\alpha^3$ | $\dfrac{xy}{1-\alpha(1-x)(1-y)}$ | $\alpha \in ]-\infty, 1]$ | $\begin{cases} F_z & \text{si } \alpha \to -\infty \\ F_p & \text{si } \alpha = 0 \\ F_h & \text{si } \alpha = 1 \end{cases}$ |
| $T_\alpha^6$ | $1 - \left[ \overbrace{\underbrace{(1-x)^\alpha}_{A} + \underbrace{(1-y)^\alpha}_{B} - (1-x)^\alpha(1-y)^\alpha}^{Z} \right]^{\frac{1}{\alpha}}$ | $\alpha \in ]0, \infty[$ | $\begin{cases} F_z & \text{si } \alpha \to 0 \\ F_p & \text{si } \alpha = 1 \\ F_m & \text{si } \alpha \to \infty \end{cases}$ |

$q = 2$. Nous avons :

$$\begin{aligned} C_\mu(\mathbf{x}) &= \mathbf{p}_1 \cdot \mu(\Omega_1) + \mathbf{p}_2 \cdot \mu(\Omega_2) + \mathbf{p}_3 \cdot \mu(\Omega_3) \\ &= 0.12 \cdot \left(\frac{3}{3}\right)^2 + 0.11 \cdot \left(\frac{2}{3}\right)^2 + 0.22 \cdot \left(\frac{1}{3}\right)^2 \\ &= 0.0144 + 0.0121 + 0.0484 \\ &= 0.0709 \end{aligned}$$

### 5.1.3 T-normes et opérateur de pré-agrégation

Nous décrivons ici une classe d'opérateurs d'agrégation appelés T-normes (Normes Triangulaires) qui sont utilisés pour définir l'opérateur de pré-agrégation.

**Definition 5.1.3** *Pour une fonction d'agrégation multivariée donnée $T : [0,1]^m \to [0,1]$, cette fonction est considérée comme une t-norme si pour tout $x$, $y$, et $z \in [0,1]$, $T$ doit satisfaire les propriétés suivantes [297] :*

- $T(x,y) = T(y,x)$ *(commutativité),*
- $T(x, T(y,z)) = T(T(x,y), z)$ *(associativité),*
- $T(x,1) = x$ *(condition de borne).*

[297]: Alsina et al. (2006), *Associative functions*

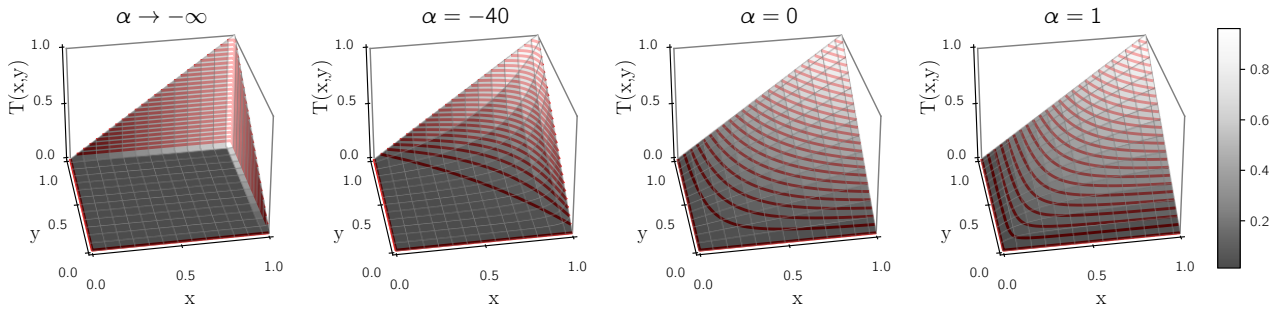
Une étude exhaustive des T-normes et de leurs propriétés est présentée dans [297]. Les T-normes peuvent être utilisées dans le contexte de la logique floue pour définir la pré-agrégation basée sur l'intégrale de Choquet. Cette généralisation a été proposée dans [295]. Cette modification permet la construction d'une nouvelle façon d'agréger l'information et a la capacité de généraliser l'intégrale de Choquet classique. La pré-agrégation est définie comme suit :

[295]: Lucca et al. (2016), « Preaggregation Functions »

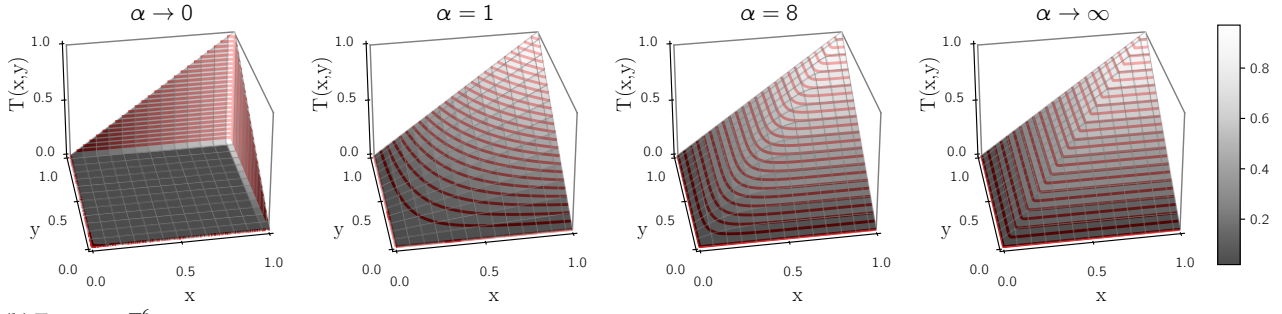
$$C_\mu^F(\mathbf{x}) = \sum_{k=1}^{m} T(\mathbf{p}_k, \mu(\Omega_k)) \tag{5.4}$$

On peut observer que lorsque $T(x,y) = xy$, qui est le produit algébrique et fait partie de la famille des T-normes, la fonction de pré-agrégation correspond à la définition classique de l'intégrale de Choquet. Ce nouvel opérateur de pré-agrégation a été étudié dans [295] et montre que le choix de la T-norme est crucial pour le processus d'apprentissage. Les auteurs démontrent que le choix de la T-norme peut entraîner une variation moyenne de 3% sur les 27 ensembles de données utilisés dans leur étude pour une mesure floue pondérée.

[295]: Lucca et al. (2016), « Preaggregation Functions »




**(a)** T-norme : $T_\alpha^3$




**(b)** T-norme : $T_\alpha^6$

**Figure 5.1 –** Vue bivariée des deux T-normes sélectionnées en fonction de la valeur de $\alpha$. La carte de couleur représente la valeur de la T-norme. La ligne rouge représente l'isocontour de la fonction.

[297]: Alsina et al. (2006), *Associative functions*

Dans [297], les auteurs fournissent une sélection de T-normes paramétrées. Ces formulations ont l'avantage d'englober plusieurs formes spécifiques de T-normes en fonction de la valeur du paramètre $\alpha$. Nous choisissons d'utiliser deux T-normes paramétrables, correspondant aux expressions (3) et (6) proposées dans [297]. Ces deux formes seront désignées par $T_\alpha^3$ et $T_\alpha^6$. Leur forme, leurs caractéristiques et le domaine de définition du paramètre $\alpha$ sont résumés dans le tableau 5.1. Une visualisation de ces deux T-normes pour quatre valeurs de $\alpha$ est présentée à la figure 5.1. Ces deux formes ont été choisies pour deux raisons. La première est liée au domaine de définition du paramètre $\alpha$, qui n'est pas défini par parties, ce qui rend l'apprentissage de ce paramètre plus simple. La seconde est que ces deux formes généralisent le produit algébrique (forme standard de l'intégrale de Choquet) et ont des formes particulières intéressantes :

- $F_z(x,y) = \begin{cases} x & \text{si } y = 1 \\ y & \text{si } x = 1, \\ 0 & \text{sinon} \end{cases}$
- $F_p(x,y) = xy$ (produit algébrique),
- $F_h(x,y) = \dfrac{xy}{x+y-xy}$ (produit de Hamacher),
- $F_m(x,y) = \min(x,y)$.

## 5.2 Apprentissage flou : intégrale de Choquet et pré-agrégation

Soit $\mathbf{p} \in [0,1]^m$ un vecteur représentant les sorties de $m$ classifieurs (voir la Définition 5.1.2). Afin d'optimiser les paramètres des deux mesures floues considérées, nous proposons de reformuler l'intégrale de Choquet de la

**Table 5.2 –** Détails de la forme reformulée pour les mesures floues de puissance et de moyenne pondérée

| **Mesure** | $g(\mathbf{h}, \theta)$ | $\mathbf{h}$ | $\theta$ | $\mathbf{r} = \partial g / \partial \theta_j$ |
|---|---|---|---|---|
| Puissance | $\mathbf{h}^\theta$ | $[m/m, m-1/m, \ldots, 1/m] \in \mathbb{R}^{m,1}$ | $\exp(b) \in \mathbb{R}^+$ | $\exp(b)\log(\mathbf{h}) \odot \mathbf{h}^{\exp(b)}$ |
| Moyenne pondérée | $\mathbf{h} \cdot \theta$ | $T_m(\mathbb{Z}/2\mathbb{Z}) \in \mathbb{Z}^{m \times m}$ | $\frac{\mathbf{w}}{\sum_i \mathbf{w}_i} \in [0,1]^m$ | $\frac{\mathbf{h}}{\sum_i \mathbf{w}_i{}^2} \cdot (\mathbf{v} - \mathbf{w})$ avec $\mathbf{v}_k = \begin{cases} 0 & \text{si } j \neq k \\ \sum_i \mathbf{w}_i \end{cases}$ |

manière suivante :

$$\begin{aligned} C_g(\mathbf{p}, \theta; \mathbf{h}) &= \sum_{k=1}^{m} \mathbf{p}_k \cdot g(\theta; \mathbf{h})_k, \\ &= \mathbf{p}^T \cdot g(\theta; \mathbf{h}). \end{aligned} \tag{5.5}$$

Dans cette reformulation, nous représentons la mesure floue avec la fonction $g(\cdot)$. Cette fonction est paramétrée par une matrice constante $\mathbf{h}$, dont la forme est spécifique à la mesure floue considérée, et par l'ensemble des variables à optimiser $\theta$. Les contraintes associées aux paramètres $\theta$, c'est-à-dire la stricte positivité pour la mesure de puissance et la somme des poids égale à 1 pour la mesure pondérée, sont directement intégrées en appliquant une fonction de reparamétrisation judicieusement sélectionnée pour garantir le domaine de définition. Une description détaillée des fonctions, des paramètres[71] et de la reparamétrisation pour prendre en compte les contraintes est donnée dans le tableau 5.2.

**Précision (71)**
Dans le tableau 5.2, pour la formulation de $\mathbf{h}$ dans la mesure de moyenne pondérée, $T_m(\mathbb{Z}/2\mathbb{Z})$ dénote l'ensemble de toutes les matrices triangulaires supérieures $m \times m$ avec des entrées binaires.

### 5.2.1 Apprentissage des paramètres de l'intégrale de Choquet

Nous considérons que les mesures floues présentées ci-dessus sont paramétrées par un ensemble de paramètres $\theta$. Dans le cas de la mesure de puissance $\theta = q$, et dans le cas des paramètres de moyenne pondérée $\theta = \mathbf{w}$. Nous sélectionnons une fonction de perte quadratique, basée sur la norme $\ell_2$ comme présenté dans (5.6) et symbolisée par $\mathbf{\Sigma}_\theta$. Dans le cadre de cet apprentissage supervisé, en présence de $N$ échantillons, nous avons les données de sortie des classifieurs $\mathbf{X} \in \mathbb{R}^{N,m}$, associées aux étiquettes de vérité terrain $\mathbf{y}$. Nous considérons $\mathbf{P} \in [0,1]^{N,m}$ comme la matrice des probabilités de sortie des $m$ classifieurs, triées et différenciées comme décrit dans la Définition 5.1.2.

L'objectif est de trouver les valeurs optimales de $\theta$ qui minimisent une fonction de perte entre la vérité terrain $\mathbf{y}$ utilisée pour entraîner les $m$ classifieurs et l'agrégation $\hat{\mathbf{y}} \in \mathbb{R}^{N,1}$, où $\mathbf{y} = [C_g(\mathbf{P}_0, \theta; \mathbf{h}), \ldots, C_g(\mathbf{P}_N, \theta; \mathbf{h})]$, avec la mesure floue associée $g(\mathbf{h}, \theta)$. Nous avons donc la minimisation suivante :

$$\arg\min_\theta \mathbf{\Sigma} = \arg\min_\theta \frac{1}{2N} ||\mathbf{y} - \hat{\mathbf{y}}||_2^2 . \tag{5.6}$$

Nous cherchons donc le gradient qui minimise la perte par rapport au paramètre $\theta$ de la mesure $g$. En utilisant la règle de la chaîne, nous avons :

$$\frac{\partial \mathbf{\Sigma}}{\partial \theta_j} = \frac{1}{N} \sum_{i=0}^{N} \frac{\partial \mathbf{\Sigma}}{\partial \hat{\mathbf{y}}_i} \cdot \frac{\partial \hat{\mathbf{y}}_i}{\partial g} \cdot \frac{\partial g}{\partial \theta_j} \tag{5.7}$$

**Table 5.3 –** Dérivées des T-normes par rapport aux contraintes $\alpha$ et $g$. $A$, $B$ et $Z$ sont des variables intermédiaires données dans le tableau 5.1.

| **Forme** | $\partial T_\alpha / \partial g$ | $\partial T_\alpha / \partial \beta$ |
|---|---|---|
| $T_\alpha^3(x, g)$ | $\dfrac{x\left[x(-e^\beta(x-1)+x)\right]}{\left[e^\beta(x-1)(g-1)-xg+x+g\right]^2}$ | $-\dfrac{e^\beta(x-1)(g-1)xg}{\left[e^\beta(x-1)(g-1)-xg+x+g\right]^2}$ |
| $T_\alpha^6(x, g)$ | $B^{\alpha-1}(1-A^\alpha)\left[B^\alpha(1-A^\alpha)+A^\alpha\right]^{\frac{1-\alpha}{\alpha}}$ | $-\frac{Z^{\frac{1}{\alpha}}}{\alpha}\left[\dfrac{A^\alpha e^\beta \log(A)(1-B^\alpha)+B^\alpha e^\beta \log(B)(1-A^\alpha)}{Z}-\log(Z)\right]$ |

La dérivée $\partial g / \partial \theta_j$ est donnée dans le tableau 5.2. En considérant que les agrégats sont indépendants entre chaque échantillon, nous avons :

$$\frac{\partial \mathbf{\Sigma}}{\partial \hat{\mathbf{y}}_i} = \frac{1}{N}(\hat{\mathbf{y}}_i - \mathbf{y}_i) \tag{5.8}$$

Pour alléger la formulation, nous définissons $\mathbf{r} = \frac{\partial g}{\partial \theta_j} \in \mathbb{R}^{m,1}$. La forme analytique de $\mathbf{r}$ en fonction de la mesure floue utilisée est donnée dans le tableau 5.2. Enfin, le gradient de la perte par rapport au paramètre $\theta_j$ pour une mesure floue $g$ est donné par :

$$\nabla \mathbf{\Sigma}_{\theta_j} = \frac{1}{N}(\hat{\mathbf{y}} - \mathbf{y})^T \cdot (\mathbf{P} \cdot \mathbf{r}) \tag{5.9}$$

La minimisation est effectuée par une descente de gradient sur l'ensemble des paramètres $\theta$. À l'itération $t+1$, la mise à jour de $\theta$ par rapport à un pas de descente de gradient $\gamma$ est réalisée grâce à :

$$\theta_j^{[t+1]} = \theta_j^{[t]} - \gamma \nabla \mathbf{\Sigma}_{\theta_j} \tag{5.10}$$

### 5.2.2 Apprentissage des T-normes paramétriques

En utilisant l'équation (5.5) et l'opérateur T-norme $T_\alpha$, nous pouvons reformuler la pré-agrégation avec la même notation que dans la section précédente :

$$C_g^T(\mathbf{p}, \theta; \mathbf{h}) = \sum_{k=1}^{m} T_\alpha \left(\mathbf{p}k, g(\theta; \mathbf{h})k\right), \tag{5.11}$$

Les deux choix de T-normes que nous proposons d'étudier sont basés sur un paramètre $\alpha$ dont le domaine de définition est donné dans le tableau 5.1. Nous proposons de contourner la contrainte associée au domaine de définition en réalisant une reparamétrisation. Nous définissons $\beta \in \mathbb{R}$ comme le paramètre à optimiser. Nous utilisons les formes suivantes :

- Pour $T\alpha^3$, nous définissons $\alpha = 1 - e^\beta$.
- Pour $T\alpha^6$, nous définissons $\alpha = e^\beta$.

De cette manière l'optimisation est réalisée sur $\beta$ et non sur $\alpha$, qui appartient à $\mathbb{R}$, mais dont la reparamétrisation permet de respecter le domaine de définition de $\alpha$. La dérivée de la T-norme par rapport à $\beta$ est donnée dans le tableau 5.3, Avec

$\mathbf{y} = [C_g^T(\mathbf{P}0, \theta; \mathbf{h}), \ldots, C_g^T(\mathbf{P}N, \theta; \mathbf{h})]$, nous avons la minimisation suivante :

$$\arg\min \theta, \beta\mathbf{\Sigma} = \arg\min \theta, \beta \frac{1}{2N} ||\mathbf{y} - \hat{\mathbf{y}}|| 2^2. \tag{5.12}$$

Cela conduit au gradient suivant pour $\theta$ et $\beta$ :

$$\begin{aligned} \frac{\partial \mathbf{\Sigma}}{\partial \theta_j} &= \frac{1}{N} \sum_{i=0}^{N} \frac{\partial \mathbf{\Sigma}}{\partial \mathbf{y}i} \cdot \frac{\partial \mathbf{y}i}{\partial T\alpha} \cdot \frac{\partial T\alpha}{\partial g} \cdot \frac{\partial g}{\partial \theta_j} \\ \frac{\partial \mathbf{\Sigma}}{\partial \beta} &= \frac{1}{N} \sum_{i=0}^{N} \frac{\partial \mathbf{\Sigma}}{\partial \mathbf{y}i} \cdot \frac{\partial \mathbf{y}i}{\partial T\alpha} \cdot \frac{\partial T\alpha}{\partial \beta} \end{aligned} \tag{5.13}$$

Les deux nouvelles dérivées $\partial T_\alpha / \partial g$ et $\partial T_\alpha / \partial \beta$ sont données dans le tableau 5.1. Le gradient de la perte par rapport au paramètre $\theta_j$ et $\beta$ pour une T-norme $T_\alpha$ est donné par :

$$\begin{aligned} \nabla \mathbf{\Sigma} \theta_j &= \frac{1}{N} (\hat{\mathbf{y}} - \mathbf{y})^T \cdot \left( \frac{\partial T\alpha}{\partial g} \cdot \mathbf{r} \right) \\ \nabla \mathbf{\Sigma} \beta &= \frac{1}{N} \frac{\partial T_\alpha}{\partial \beta} \cdot (\hat{\mathbf{y}} - \mathbf{y})^T \end{aligned} \tag{5.14}$$

La dérivée $\partial \mathbf{y}i / \partial T\alpha$ est simplement une dérivée d'une somme de monômes, donc elle est égale à 1. De manière similaire à la section précédente, la minimisation est effectuée par une descente de gradient sur l'ensemble des paramètres $\theta$ et $\beta$ en utilisant le pas de descente de gradient $\gamma$.

## 5.3 Expérimentations

### 5.3.1 Description du jeu de données

Pour évaluer la méthode proposée, nous nous sommes appuyés sur une image SAR en bande X (9,65 GHz) du satellite PAZ acquise le 4 août 2020. Cette acquisition est centrée sur le massif du Mont-Blanc. L'acquisition est en double polarisation (HH et HV) et a été obtenue à partir d'une orbite descendante avec un angle d'incidence de 37,86°. Respectivement 7 et 6 régions d'intérêt (ROI) ont été sélectionnées pour chacune des 2 classes considérées : zones d'ablation et d'accumulation des glaciers, à partir des images optiques Pléiades, respectivement abrégées *ABL* et *ACC* dans ce qui suit. Les ROI sont par la suite désignées comme groupes. La figure 5.2 illustre la couverture au sol et la distribution des groupes. Nous avons effectué une calibration radiométrique sur l'image affichée dans cette figure. Les ROI de chaque classe ont été transformées d'une géométrie au sol à une géométrie radar pour produire la figure directement en géométrie radar, excluant les zones de distorsion géométrique. Les données ont été transformées en échelle logarithmique. Dans chaque groupe, nous avons extrait des patchs non superposés. Les patchs extraits sont réalisés sur les deux polarisations HH et HV, conduisant à des jeux de données de taille $N \times 2 \times K^2$ pixels avec $N$ étant le nombre

| **taille** | 3×3 | 5×5 | 7×7 | 9×9 |
|---|---|---|---|---|
| *ABL* | 25959 | 9100 | 4527 | 2673 |
| *ACC* | 18600 | 6503 | 3208 | 1891 |

**Table 5.4 –** Nombre de patchs dans le jeu de données en fonction de la taille de la fenêtre.

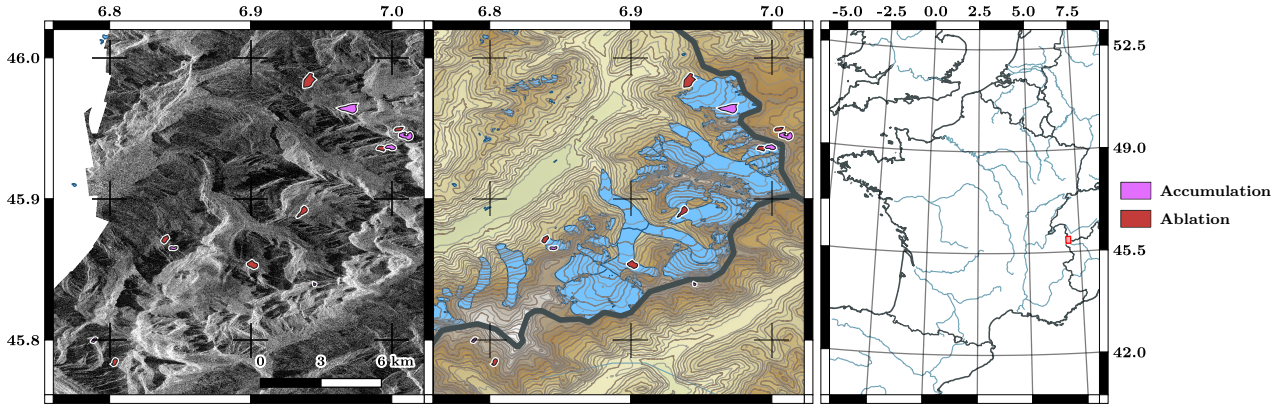



**Figure 5.2 –** Position des ROI pour les 2 classes *ABL* et *ACC* sur le massif du Mont Blanc. À gauche l'image SAR PAZ, au centre illustration de la topographie et répartition des zones. La présence d'eau (y compris les glaciers) est donnée en bleu.

de patchs et $K$ la taille de la fenêtre. Nous avons créé 4 jeux de données en utilisant des fenêtres carrées de tailles 3×, 5×5, 7×7 et 9×9 pixels, représentant approximativement une taille de fenêtre entre 9 mètres et 27 mètres. De cette manière, nous étudions l'impact de la taille de la fenêtre, et donc de la quantité d'informations associées, sur la classification. Le nombre de patchs pour chaque jeu de données est présenté dans le tableau 5.4.

### 5.3.2 Algorithmes et stratégie d'apprentissage

Nous avons employé une stratégie de K-fold stratifiée, avec K=4, pour évaluer différentes approches sur notre jeu de données. Cette stratégie divise le jeu de données en quatre parties, dont trois sont utilisées pour l'apprentissage tandis que la partie restante est utilisée pour les tests. L'évaluation se fait de manière itérative sur les 4 plis pour évaluer l'ensemble du jeu de données. Avant chaque processus d'apprentissage, nous avons équilibré les classes en utilisant une méthode de sous-échantillonnage aléatoire. Aucun équilibrage n'a été effectué sur l'ensemble de test. Nous proposons de déployer notre approche sur une forêt de classifieurs évalués sur les mêmes données d'entrée. Les données sont ainsi vectorisées pour chaque polarisation et concaténées. Nous visons à étudier l'agrégation des décisions fournies par les classifieurs décrits dans le tableau 5.5. La sélection des classifieurs est basée sur le désir d'obtenir une variété de comportements en raison de leur nature fondamentalement différente. De plus, une sélection de ces algorithmes a été étudiée sur l'évaluation de leur capacité à classifier la neige humide et la neige sèche dans [72], donnant des résultats similaires mais avec des comportements différents. Un aperçu concis de leur application dans les problèmes de télédétection est donné dans [246]. Trois catégories d'algorithmes sont mises en avant :

- la première implique des algorithmes tels qu'Adaboost et Random Forest (RF), centrés sur l'apprentissage de classifieurs faibles, particulièrement adaptés à la classification binaire avec l'avantage supplémentaire de l'interprétabilité des résultats.
- La deuxième catégorie tourne autour de la notion de séparation ou de distance, utilisant des méthodes comme SVM pour projeter les données pour une discrimination optimale des classes. Alternativement, les algorithmes de cette catégorie, comme KNN, déterminent directement la classe d'un nouvel échantillon basé sur les distances avec les échantillons voisins et leurs étiquettes.

[72]: Gallet et al. (2024), « Wet Snow Detection From Satellite SAR Images by Machine Learning With Physical Snowpack Model Labeling »

[246]: Lary et al. (2016), « Machine learning in geosciences and remote sensing »

| Classifieur | Type | Paramètres | Référence |
|---|---|---|---|
| RF | Ensemble | 100 arbres | [247] |
| Adaboost | Ensemble | 100 arbres | [248] |
| Arbre de Décision | Arbre | | [357] |
| Gradient Boosting | Ensemble | 100 étapes de boost | [358] |
| SVM | Séparation | noyau linéaire | [250] |
| SVM | Séparation | noyau RBF, C=1 | [250] |
| KNN | Séparation | K=50 | [251] |
| MLP | Neuronal | 100 couches cachées | [253] |
| Régression Logistique | Neuronal | pénalité $\ell_2$ | [252] |

**Table 5.5 –** Détails des classifieurs et de leurs paramètres utilisés dans le processus d'apprentissage.

- La dernière catégorie englobe les algorithmes d'auto-apprentissage neuronaux, qui emploient l'apprentissage des poids pour des combinaisons linéaires et des fonctions non linéaires. Tandis que la régression logistique apprend un seul niveau de poids, le MLP consiste en plusieurs couches d'apprentissage successives.

L'apprentissage de l'appartenance à une classe se fait en deux parties, en utilisant l'ensemble de données d'entraînement $S_{train}$ et l'ensemble de données de test $S_{test}$. Un diagramme du processus d'apprentissage est donné dans la figure 5.3.

- La première partie consiste à estimer les paramètres des 9 classifieurs à partir de l'ensemble de données $S_{train}$. Les probabilités d'appartenance à la classe $ACC$ de $S_{train}$ sont récupérées pour former l'ensemble de données de probabilités d'entraînement $p_{train}$ pour la partie d'agrégation. Cet ensemble de données est de taille $N \times 9$. Nous construisons l'ensemble de test $p_{test}$ en estimant les probabilités d'appartenance à la classe $ACC$ de $S_{test}$ pour chaque classifieur.
- La deuxième partie consiste à estimer les paramètres des diverses méthodes d'agrégation proposées à partir de $p_{train}$ et à évaluer leurs performances sur $p_{test}$. Cette étape est décrite dans l'algorithme 1.

**Algorithm 1:** Descente de gradient pour l'apprentissage des paramètres d'agrégation

**entrée :** Données d'entraînement $p_{train}$, Données de test $p_{test}$, Paramètres initiaux $\theta$ et $\beta$, Taux d'apprentissage $\gamma$, Tolérance $\epsilon$
**sortie :** Paramètres optimisés $\theta$ et $\beta$
Initialiser $\theta$ et $\beta$
Calculer $\mathbf{P}$, le vecteur trié et différencié de $p_{train}$
Calculer $\mathbf{h}$, le vecteur ou la matrice pour construire la mesure floue
**while** $\|\nabla_\theta f(\theta, \beta)\| > \epsilon$ *ou* $\|\nabla_\beta f(\theta, \beta)\| > \epsilon$ *et* $t < n$ **do**
    Calculer la mesure floue $g(\mathbf{h}, \theta)$
    Calculer $\hat{\mathbf{y}} = [C_g^T(\mathbf{P}_0, \theta; \mathbf{h}), \ldots, C_g^T(\mathbf{P}_N, \theta; \mathbf{h})]$
    Calculer le gradient $\nabla\mathbf{\Sigma}_\theta$ et $\nabla\mathbf{\Sigma}_\beta$
    $\theta^{t+1} \leftarrow \theta^t - \gamma\nabla\mathbf{\Sigma}_\theta$
    $\beta^{t+1} \leftarrow \beta^t - \gamma\nabla\mathbf{\Sigma}_\beta$
**return** $\theta, \beta$
Évaluer la performance de l'agrégation sur $p_{test}$

Pour l'initialisation des paramètres appris, nous utilisons une distribution aléatoire suivant une distribution uniforme entre 0 et 1. Cela est fait pour les paramètres $\theta$ de la mesure floue utilisée, et pour le paramètre re-paramétrisé $\beta$ de

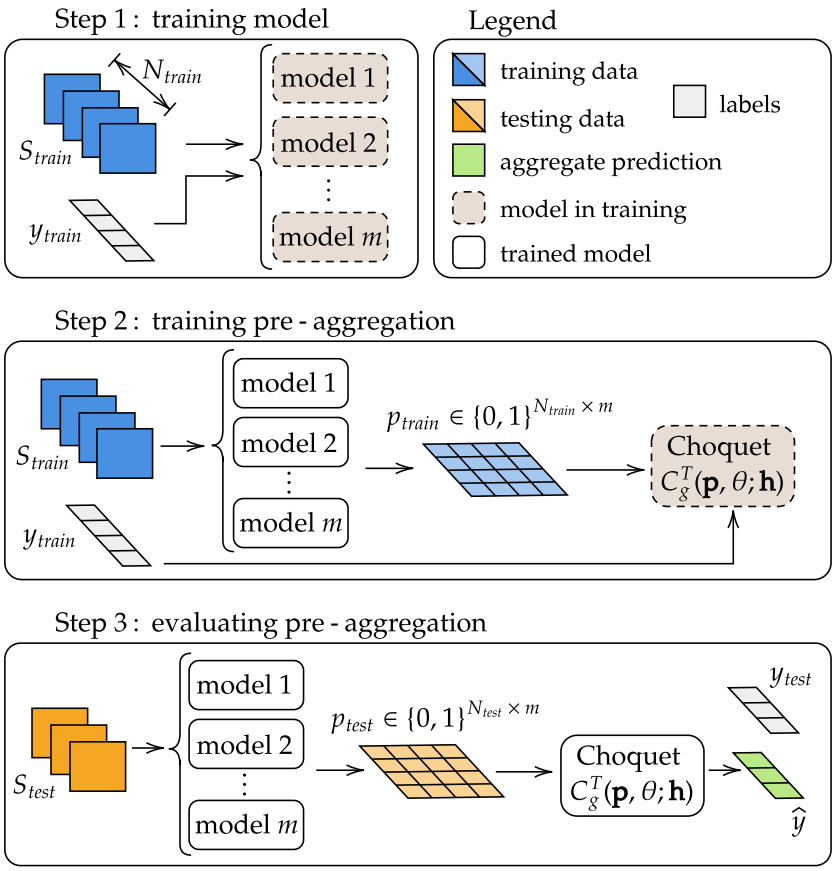



**Figure 5.3 –** Schéma du processus d'apprentissage en deux étapes. Une troisième étape est illustrée pour l'évaluation des performances des différentes méthodes d'agrégation

la T-norme dans le cas de la pré-agrégation. Dans les résultats expérimentaux, le nombre maximal d'itérations est fixé à $n = 1000$, le taux d'apprentissage est fixé à $\gamma = 0.005$ pour les classifieurs flous basés sur les poids et $\gamma = 0.001$ pour les classifieurs flous basés sur les puissances. Nous étudions 6 de agrégations :

- Intégrale de Choquet avec mesure floue pondérée (ChqW),
- Intégrale de Choquet avec mesure floue de puissance (ChqP),
- Pré-agrégation avec mesure floue pondérée et T-norme $T^3$ (ChqW-Tn3),
- Pré-agrégation avec mesure floue pondérée et T-norme $T^6$ (ChqW-Tn6),
- Pré-agrégation avec mesure floue de puissance et T-norme $T^3$ (ChqP-Tn3),
- Pré-agrégation avec mesure floue de puissance et T-norme $T^6$ (ChqP-Tn6).

Nous comparons les résultats obtenus avec ceux d'une régression logistique (LR). Il s'agit d'un opérateur d'agrégation courant, qui simule la couche dense (cf Définition 51) d'un réseau neuronal et est donné par :

$$\hat{y} = \sigma(\mathbf{w}^T \mathbf{x} + b) \tag{5.15}$$

où $\sigma$ est la fonction sigmoïde, $\mathbf{w}$ est le vecteur de poids, et $b$ est le biais. Nous effectuons les évaluations en utilisant les 4 plis de l'ensemble de données et 20 graines différentes. L'évaluation des résultats se fait en utilisant le score F1 pondéré pour tenir compte du déséquilibre de classe dans l'ensemble de test. Les développements sont réalisés en Python.

| **size** | 3×3 | 5×5 | 7×7 | 9×9 |
|---|---|---|---|---|
| LR | 61.9 ± 0.3 | 63.7 ± 0.6 | 65.7 ± 0.6 | 67.4 ± 1.1 |
| ChqW | 62.6 ± 0.3 | 63.8 ± 0.9 | 66.2 ± 0.5 | 68.6 ± 1.2 |
| ChqP | **69.3 ± 0.3** | **73.2 ± 0.3** | **75.3 ± 0.9** | **76.5 ± 1.5** |

**Table 5.6 –** F1-score pondéré pour les méthodes d'agrégation floue avec des mesures floues moyenne pondérée et de puissance, et la régression logistique.

## 5.4 Résultats quantitatifs sur l'apprentissage flou

### 5.4.1 Apprentissage de la mesure floue

Les résultats obtenus sont présentés dans le tableau 5.6. Nous observons que les classifieurs flous basés sur les puissances (ChqP) obtiennent de meilleurs résultats que les classifieurs flous basés sur des combinaisons pondérées (ChqW).

La méthode la moins efficace sur tous les ensembles de données considérés est l'apprentissage d'une combinaison de classifieurs par la méthode de régression logistique, avec près de ∼ 9% de différence par rapport au classifieur flou ChqP. Le comportement entre la régression logistique et le classifieur ChqW semble similaire, ce qui peut s'expliquer par le fait que la mesure floue utilisée généralise la moyenne pondérée. Il est également à noter que les performances augmentent avec la taille de la fenêtre, ce qui est attendu car cela permet d'augmenter la quantité d'informations disponibles pour la classification, et cela se produit de manière similaire pour tous les classifieurs.

### 5.4.2 Apprentissage des T-normes

Le tableau 5.7 présente les résultats obtenus pour les méthodes de pré-agrégation floue avec mesures floues pondérées et de puissance pour les deux T-normes (ChqP $T^3$, ChqP $T^6$, ChqW $T^3$, ChqW $T^6$). Les performances sont meilleures pour la T-norme $T^3$ que pour la T-norme $T^6$ dans le contexte de la pré-agrégation pondérée (ChqW), et vice versa pour la pré-agrégation de puissance (ChqP). Nous observons que les performances sont similaires en termes de valeur pour les cas considérés par rapport à l'apprentissage de la mesure floue sans T-norme. Cependant, il convient de noter que dans le contexte de ChqP, les performances présentent plus de variance pour les deux T-normes, mais avec une valeur plus cohérente entre les ensembles de données qu'avec l'apprentissage de la T-norme.

| **size** | | 3×3 | 5×5 | 7×7 | 9×9 |
|---|---|---|---|---|---|
| LR | | 61.9 ± 0.3 | 63.7 ± 0.6 | 65.7 ± 0.6 | 67.4 ± 1.1 |
| ChqW | $T^3$ | **70.6 ± 0.5** | **73.5 ± 0.2** | **75.2 ± 1.3** | **76.4 ± 1.4** |
| | $T^6$ | 68.3 ± 1.5 | 71.0 ± 1.4 | 72.8 ± 1.6 | 74.1 ± 1.8 |
| ChqP | $T^3$ | 68.6 ± 1.6 | 71.4 ± 1.8 | 74.2 ± 2.1 | 74.4 ± 2.1 |
| | $T^6$ | 69.9 ± 1.3 | 73.1 ± 1.1 | 74.8 ± 1.4 | 76.0 ± 1.7 |

**Table 5.7 –** F1-score pondéré pour les méthodes d'agrégation floue robustes avec des mesures floues moyenne pondérée et de puissance pour les deux T-normes et la régression logistique.

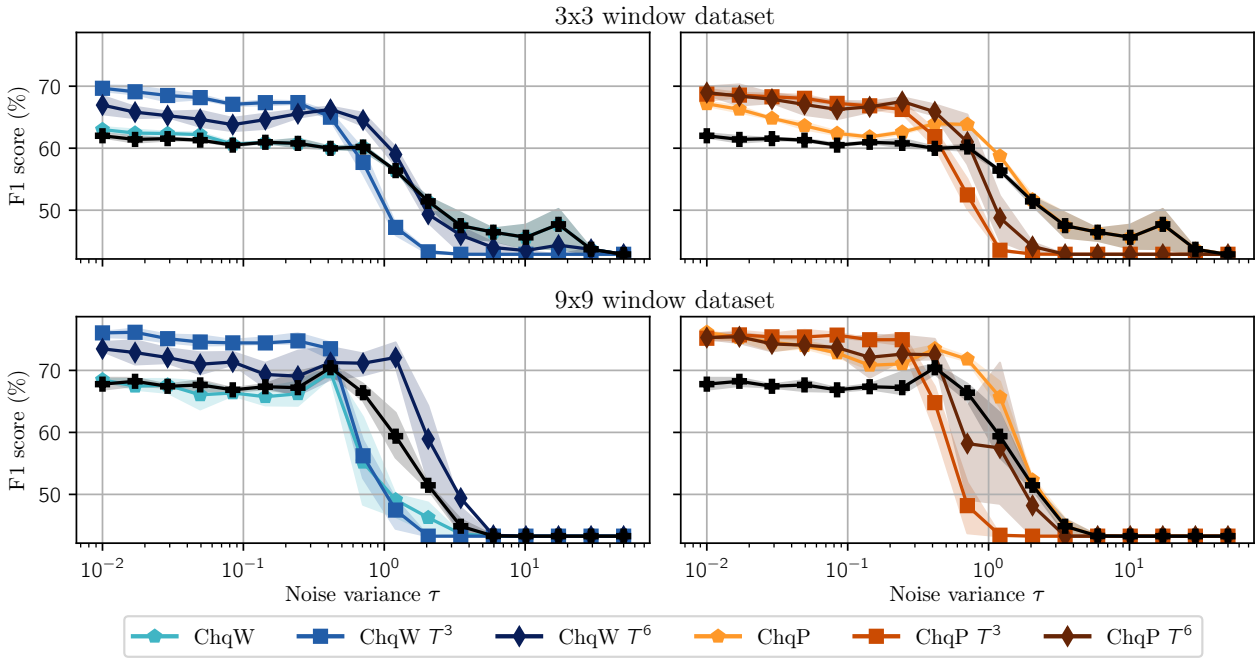



**Figure 5.4 –** Évaluation du score F1 (pondéré) pour les 2 classes *ABL* et *ACC*, pour les 3 classifieurs flous pondérés (à gauche et bleu), les 3 classifieurs flous de puissance (à droite et orange) et la régression logistique (noir) en fonction de la variance d'un bruit multiplicatif.

### 5.4.3 Robustesse aux perturbations

Nous visons à étudier la robustesse de différentes agrégations pour les deux ensembles de données de tailles différentes (3×3 et 9×9) par rapport à deux types de perturbations. Les résultats sont présentés sous forme de boîtes à moustaches simplifiées, avec la médiane symbolisée par un carré et les barres d'erreur fournissant des informations sur les premier et troisième quartiles.

#### Bruit multiplicatif

La première perturbation étudiée est le bruit multiplicatif, similaire à la nature du bruit de speckle présent dans les images SAR. La perturbation est introduite dans les données d'entraînement comme suit : $X_{bruit} = X + \sqrt{\tau} \times \mathcal{N}(0,1) \times X$, avec $\tau$ étant la variance du bruit, et $\mathcal{N}(0,1)$ une distribution normale standard. Les résultats sont fournis dans la figure 5.4.

Dans le cas des classifieurs flous basés sur les poids, nous notons la claire valeur ajoutée de l'utilisation des T-normes dans l'apprentissage, permettant un gain de performance entre 5 et 10% jusqu'à une variance de 0,4, pour les deux ensembles de données. Dans le cas de l'ensemble de données de taille3×3, le classifieur ChqW est légèrement plus efficace que la régression, tandis que pour l'ensemble de données de taille 9×9, les performances sont équivalentes, voire légèrement inférieures pour le classifieur ChqW. Cela peut s'expliquer par le fait que l'ensemble de données de taille 9×9 est plus riche en informations, poussant le classifieur hors de sa zone d'utilisation. Le classifieur le plus adapté doit être choisi selon la tâche à accomplir. Le classifieur ChqW $T^3$ offre la meilleure performance pour un faible niveau de bruit ($\tau < 0,4$) avec une bonne stabilité, tandis que le classifieur ChqW $T^6$ est plus volatil mais reste plus efficace que la régression logistique jusqu'à un niveau de bruit $\tau = 2$.

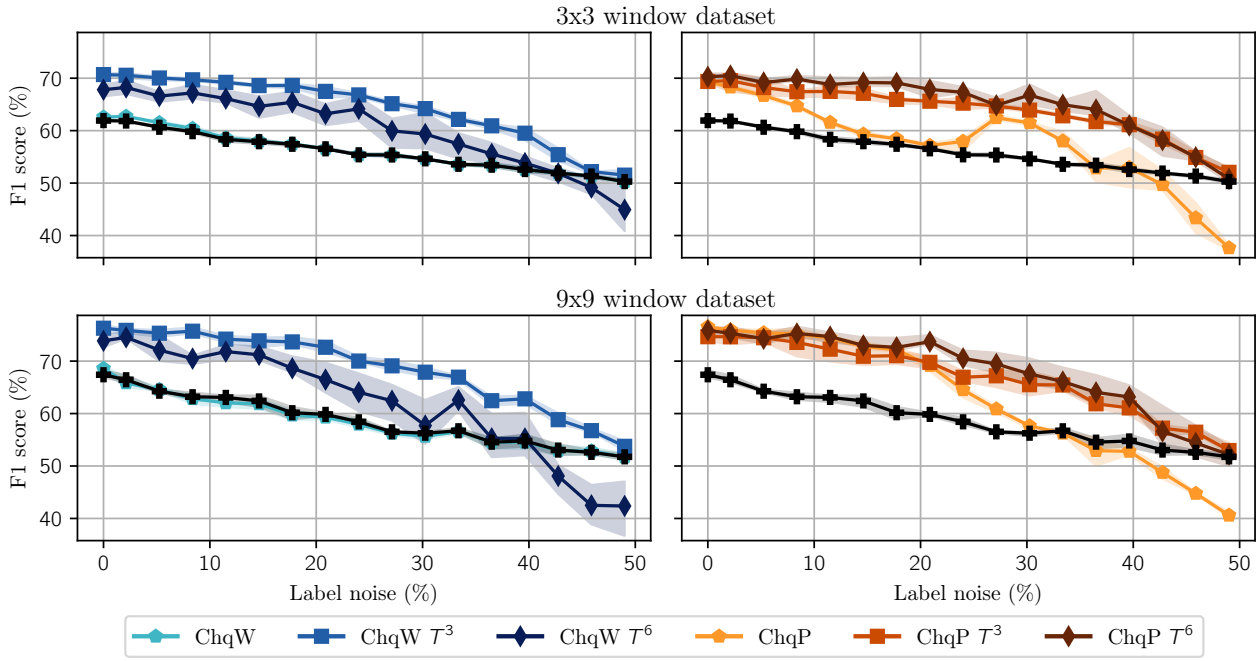



**Figure 5.5 –** Évaluation du score F1 (pondéré) pour les 2 classes *ABL* et *ACC*, pour les 3 classifieurs flous pondérés (à gauche et en bleu), les 3 classifieurs flous de puissance (à droite et en orange) et la régression logistique (en noir) en fonction du pourcentage de bruit d'étiquetage.

Nous observons que les classifieurs basés sur les mesures de puissance (ChqP), avec ou sans T-norme, donnent des résultats supérieurs à ceux basés sur les mesures de poids (ChqW) mais avec une plus grande variance. De manière similaire à l'analyse des résultats du classifieur ChqW, le choix de la meilleure forme dépend de ce qui est désiré. Cependant, dans ce cas, même la forme sans T-norme est intéressante, car elle offre des performances supérieures à la régression logistique pour un niveau de bruit $\tau > 1$. Les formes ChqP $T^3$ et ChqP $T^6$ sont les plus efficaces avec des différences entre $\sim 5\%$ et $\sim 10\%$ par rapport à la régression logistique mais jusqu'à un niveau de bruit inférieur au classifieur ChqP (jusqu'à $\tau = 0,3$).

### Bruit de label

La deuxième perturbation étudiée est la bruit d'étiquetage. Nous visons à évaluer la performance de la méthode d'agrégation dans le contexte de l'étiquetage incorrect des données d'origine. Pour ce faire, nous inverserons de manière aléatoire un pourcentage d'étiquettes dans l'ensemble de données d'origine. Nous entraînons ensuite la forêt de classifieurs et effectuons l'agrégation sur ces données bruitées. Enfin, nous évaluons les performances sur l'ensemble de données de test, où nous avons conservé les étiquettes originales, non bruitées. Les résultats sont présentés dans la figure 5.5. Nous avons observé que lorsqu'il s'agit d'étiquettes bruitées, l'utilisation d'une T-norme ou d'un opérateur de pré-agrégation au lieu d'une simple intégrale de Choquet améliore significativement les performances globales des mesures floues. Cette amélioration peut être de plus de 10%. Nous avons également trouvé des résultats similaires lors du traitement de données avec du bruit de speckle. Pour les classifieurs basés sur la mesure floue pondérée, le comportement sans T-norme est le même que celui de la régression logistique. Cependant, l'utilisation de la forme ChqW $T^3$ donne un comportement plus stable et des performances plus élevées que

la forme ChqW $T^3$. Cette dernière est plus adaptée pour agréger des données avec plus d'informations (9×9). L'utilisation de T-normes réduit la variance des performances pour les classifieurs basés sur la mesure floue de puissance, en particulier pour l'ensemble de données3×3. La forme ChqP $T^6$ donne les meilleurs résultats avec des performances entre 5% et 12% supérieures à la régression logistique, jusqu'à une pollution des étiquettes de 45%. La forme ChqP $T^3$ est moins efficace mais offre une variabilité plus faible en présence de niveaux élevés de pollution des étiquettes ($\geq$ 30%).

## 5.5 Spatialisation des résultats

Dans cette section, nous visons à spatialiser les résultats sur un glacier spécifique, le glacier d'Argentière, situé dans le massif du Mont-Blanc, afin de comparer visuellement les approches d'agrégation proposées. La figure 5.6 montre le glacier d'Argentière, à la fois en images optique et radar, avec les lignes d'altitude correspondantes. Pour cette partie, nous utilisons les ensembles de données de pixels 9×9 et3×3 détaillés dans la section 5.3.1 pour entraîner les classifieurs et les modèles d'agrégation. Nous découpons manuellement un ROI centré sur le glacier d'Argentière à partir de l'image radar du 04/08/2020. Le choix de ce glacier est motivé par le fait qu'aucune des données labellisées ne provient de ce glacier, et il se présente comme un glacier de vallée, avec des zones d'accumulation et d'ablation distinctes. Nous avons ensuite estimé, via une fenêtre glissante de taille 9×9 pixels, la probabilité d'appartenance aux classes d'ablation et d'accumulation pour chacun des classifieurs, puis agrégé ces probabilités via les différentes méthodes d'agrégation. La décision pour une fenêtre donnée est prise pour le pixel central de la fenêtre. Dans cette partie, nous n'avons pas de vérité terrain. Par conséquent, la comparaison des résultats est visuelle avec une connaissance experte des glaciers. Nous proposons de visualiser les résultats pour les deux meilleures méthodes de pré-agrégation, à savoir ChqW$T^3$ et ChqP$T^6$, avec la régression logistique dans les figures 5.7 et 5.8 Pour chaque carte, nous comparons un classifieur de type pré-agrégation (noté Chq) avec la régression logistique (LR). Les pixels bleu foncé et orange correspondent respectivement aux pixels classés comme appartenant aux classes d'accumulation et d'ablation, par les deux classifieurs (Chq et LR). Les pixels orange clair correspondent aux pixels classés comme appartenant à la classe d'ablation par le classifieur Chq et à la classe d'accumulation par le classifieur LR. Les pixels bleu clair correspondent aux

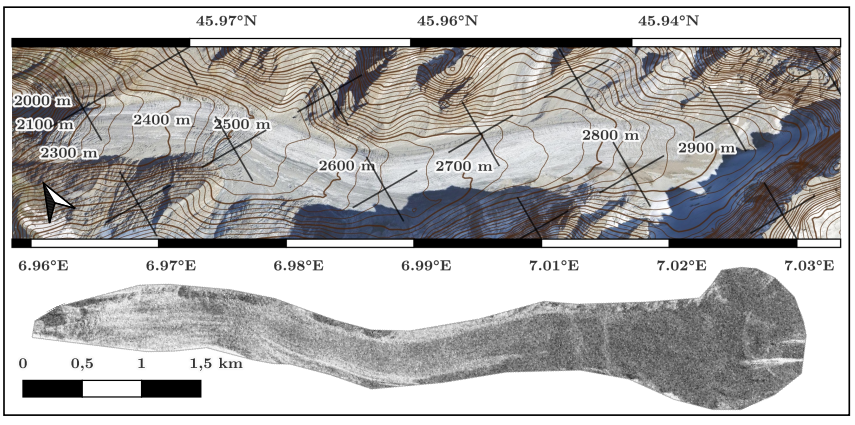



**Figure 5.6 –** Image optique (en haut) et image radar (en bas) du glacier d'Argentière. L'image SAR considérée correspond à la moyenne des canaux HH et HV en dB pour la date du 04/08/2020.

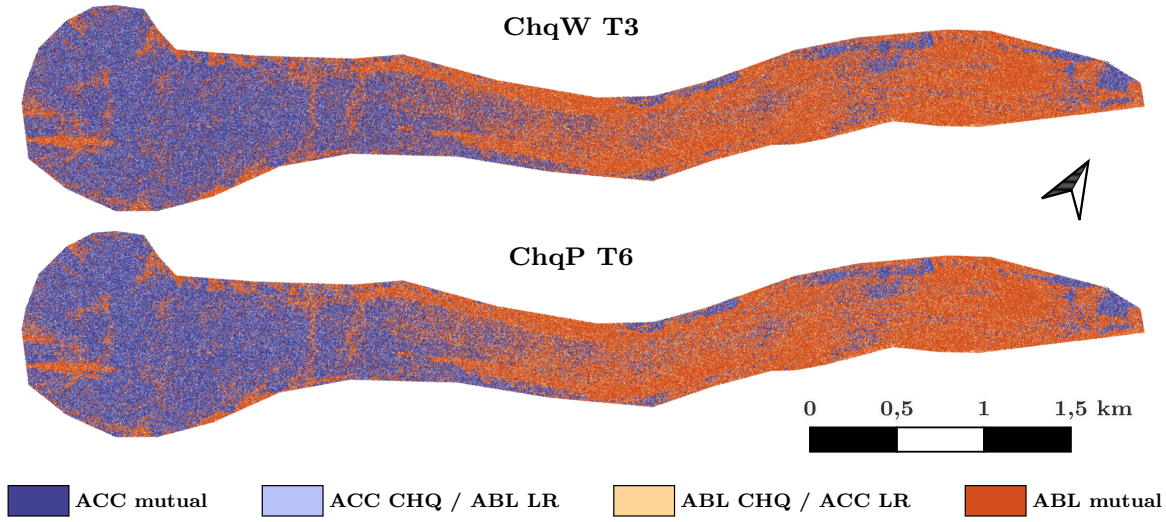



**Figure 5.7 –** Ensemble de données en fenêtre3×3. Spatialisation des résultats (en géométrie radar) pour le glacier d'Argentière pour les deux meilleures méthodes de pré-agrégation, Chq$W^3$ et Chq$P^6$, en comparaison avec l'agrégation de régression logistique.

pixels classés comme appartenant à la classe d'accumulation par le classifieur Chq et à la classe d'ablation par le classifieur LR.

Nous remarquons que les résultats basés sur l'ensemble de données de taille3×3 dans la figure 5.7 sont beaucoup plus bruités que ceux basés sur l'ensemble de données de taille 9×9 dans la figure 5.8. Cela s'explique par le plus petit nombre de pixels considérés pour l'apprentissage des classifieurs. Cependant, nous pouvons remarquer que les résultats obtenus par les classifieurs Chq sont plus homogènes que ceux obtenus par la régression logistique. En effet, nous observons un nombre significatif de pixels bleu clair en amont du glacier, à haute altitude et correspondant à une zone d'accumulation. De nombreux pixels orange clair sont également observés en aval du glacier, à des altitudes plus basses et correspondant à une zone d'ablation.

En observant les cartes obtenues avec l'ensemble de données de taille 9×9, nous remarquons que les résultats sont plus homogènes entre Chq et LR. Nous remarquons qu'il y a très peu de pixels bleu clair, ce qui signifie que la zone d'accumulation est vue de manière similaire par les deux classifieurs. Cependant, nous pouvons voir un nombre significatif de pixels orange clair, sur le bord de la zone considérée comme une ablation par les deux classifieurs. Cela indique que les classifieurs Chq ont tendance à donner des résultats plus larges et plus homogènes que la régression logistique. Cette observation est cohérente avec l'aval du glacier où nous sommes presque certains que la zone est d'ablation.

Il est important de noter qu'en raison de la sélection manuelle dans la géométrie radar, les bords des cartes ne correspondent pas exactement à la géométrie du glacier. De plus, le contour de l'amont du glacier est particulièrement difficile à réaliser et est plus large que la réalité. Cela explique la présence de zones horizontales (à l'extrémité gauche des cartes) considérées comme une ablation par les deux classifieurs et qui sont vraiment des zones de distorsion de l'image radar. Nous pouvons également observer des zones verticales en haut à gauche des cartes, qui sont probablement des crevasses ou des zones de rupture de

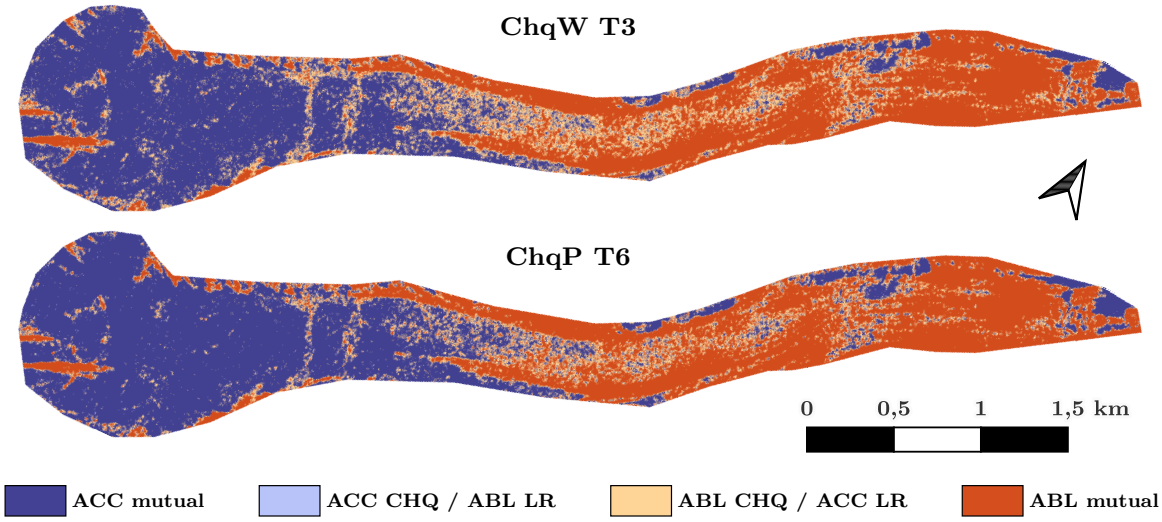



**Figure 5.8 –** Ensemble de données en fenêtre 9×9. Spatialisation des résultats (en géométrie radar) pour le glacier d'Argentière pour les deux meilleures méthodes de pré-agrégation, Chq$WT^3$ et Chq$PT^6$, en comparaison avec l'agrégation de régression logistique.

pente et d'accélération du glacier, perçu comme une zone d'ablation par les classifieurs.

## 5.6 Conclusion

Nous avons présenté dans cette section une application de la logique floue et de ses agrégateurs associés pour élaborer un cadre de classification des zones glaciaires dans les images SAR à l'aide d'une forêt de classifieurs. Cette nouvelle approche implique l'adaptation de l'intégrale de Choquet pour permettre un apprentissage autonome, nous permettant d'apprendre les paramètres des mesures floues constitutives. De plus, nous avons étendu cette méthode en utilisant des T-normes pour former des pré-agrégations, afin de renforcer la robustesse aux différents types de bruit (de données et d'étiquettes) rencontrés dans les images SAR.

Grâce à un ensemble de tests sur une zone du massif du Mont-Blanc, nous avons démontré la validité expérimentale de notre approche et sa valeur pratique. En analysant les résultats obtenus en utilisant différentes tailles de fenêtre, nous avons obtenu une amélioration significative de près de 10 % par rapport à une agrégation basée sur un opérateur neuronal standard : la régression logistique. Cette découverte souligne le potentiel tangible de notre méthode pour améliorer la précision de la classification des zones glaciaires. Il est clair que l'utilisation de millions de paramètres pour décomposer une image de 3×3 ou 9×9 n'est pas raisonnable, et puisque nous visons à rester à l'échelle du voisinage de pixels, les réseaux neuronaux convolutifs profonds (CNN) ne sont pas une option appropriée, surtout lorsque de bons résultats peuvent être obtenus avec beaucoup moins de paramètres explicables, rendant ainsi cette approche préférable à un réseau profond surdimensionné pour le problème en question.

Nous avons également évalué la capacité de nos méthodes à gérer les deux principaux défis de l'imagerie SAR : le bruit multiplicatif (speckle) et la fiabilité

des étiquettes. Les résultats obtenus montrent une robustesse face à ces deux types de bruit. De plus, nous avons démontré les avantages de l'utilisation des opérateurs de pré-agrégation, qui ont amélioré les résultats de classification par rapport à l'utilisation seule de l'intégrale de Choquet. De nombreuses perspectives existent pour ce travail, la principale étant de déterminer les conditions optimales d'utilisation de ces méthodes. Une approche pourrait être envisagée à travers les probabilités approximativement correctes (*Probabilities Approximately Correct* PAC) [359], de manière similaire à [360], permettant une quantification de la fiabilité des résultats obtenus. Dans ce travail, nous avons proposé une approche en deux étapes. Cependant, il pourrait être intéressant d'explorer la voie de l'optimisation en une seule étape. Cette proposition reste complexe en raison des opérateurs de type tri, dépendant des entrées. Au niveau des applications, une perspective serait de se concentrer sur l'utilisation de données provenant de différents capteurs de la même modalité ou même de modalités différentes.

[359]: Guedj (2019), *A Primer on PAC-Bayesian Learning*

[360]: Zantedeschi et al. (2021), « Learning Stochastic Majority Votes by Minimizing a PAC-Bayes Generalization Bound »

# Apprentissage d'ensemble de combinaisons de distances et divergences

# 6

**6.1 Apprentissage d'ensemble de combinaisons de distances et divergences 119**
**6.2 Apprentissage d'ensemble de divergences paramétriques . . . . . 127**
**6.3 Conclusion . . . . . 137**

Nous avons vu dans le chapitre Chapitre 5 l'utilisation de méthodes d'ensemble pour agréger des classifieurs de nature différente, dans le cadre de la classification d'images SAR. Cette approche est prometteuse, néanmoins elle se base sur un cadre sans a priori statistiques sur les données, nécessitant une quantité de données suffisante pour l'apprentissage du modèle sous-jacent. Afin de réduire le nombre de paramètres appris et d'apporter une robustesse supplémentaire face aux perturbations, nous pouvons introduire des distributions a priori sur les données, toujours en gardant le cadre de l'apprentissage automatique et la problématique du choix du modèle. Nous proposons dans ce chapitre, notre dernière contribution : une approche plus frugale en termes de paramètres et de données, mais également explicable sur la décision prise par le modèle. Cette méthode s'appuie sur diverses distances calculées à partir de plusieurs distributions caractérisant les données. Nous proposons une combinaison optimale de ces distances pour classifier des paires d'images SAR. Enfin, nous élargissons cette approche en introduisant des divergences paramétriques. Dans ce chapitre, nous évaluons les différentes approches sur des données SAR en bande X, extraites du dataset CPAZMAL décrit dans la section 3.3.

## 6.1 Apprentissage d'ensemble de combinaisons de distances et divergences

Nous proposons d'étudier dans cette section, une approche de classification binaire de paires d'images, basée sur l'apprentissage de la combinaison optimale d'une variété de distances entre des descripteurs statistiques de nature différente. Cette approche est motivée par la volonté d'intégrer des connaissances a priori sur les caractéristiques statistiques des données, mais sans pour autant imposer un modèle spécifique. L'utilisation de distances nécessite une approche de classification par paire d'images, où la décision finale est donnée pour la paire, et non pour chaque image individuellement. Dans le cadre de la classification supervisé, à partir d'un ensemble d'images labellisées, nous construisons toutes les paires possibles de ces images. Si un couple d'images est composé de deux images de la même classe, la labellisation du couple est la classe des images. Si le couple est composé d'images de classes différentes, la labellisation du couple est la classe dissemblable.

Ceci mène ainsi dans le cadre de classification binaire, à deux classes thématiques de nature similaire et à la création d'une troisième classe : les paires

dissimilaires. Nous nous plaçons dans le cadre d'une classification d'images SAR en amplitude, dont la distribution est strictement positive par définition.

### 6.1.1 Descripteurs statistiques a priori

Nous présentons dans cette partie les différents modèles statistiques a priori utilisés et les mesures de similarité associées.

#### Mesures de statistiques descriptives

En considérant un ensemble de $n$ observations $x_1, \ldots, x_n$, appartenant à une variable aléatoire $X$ de loi $f_X$ et de support $[0\,;+\infty[$, on s'intéresse aux mesures de statistiques descriptives de $X$ suivantes : la moyenne arithmétique $\mu$, l'écart-type $\sigma$ et le coefficient de variation CV, défini dans l'équation (2.2). L'utilisation de ces mesures de statistiques descriptives est motivée par le fait qu'elles sont significatives dans l'étude du signal SAR. La moyenne est ainsi un coefficient de rétrodiffusion moyen à l'échelle de la fenêtre considérée, qui peut servir à discriminer des zones de nature différente [361], de même que le coefficient de variation qui permet de caractériser l'homogénéité de la rétrodiffusion dans une zone. Dans ce cadre, on caractérise la distance $\mathcal{D}$ entre deux distributions $f_X$ et $f_Y$ par la norme $\ell_2$ entre chacune de ces mesures de statistiques descriptives.

[361]: Dobson et al. (1997), « Use of SAR image texture in terrain classification »

#### Distributions marginales positives

Soit $\mathbb{1}_{\mathcal{S}}$ la fonction indicatrice du set $\mathcal{S}$, $\Gamma(\cdot)$ et $\Gamma(\cdot,\cdot)$ les fonctions spéciales respectivement appelées gamma et gamma incomplète supérieure (cf Annexe A.1). Nous considérons des fonctions de densité de probabilité (PDF) paramétriques ayant un support sur $\mathbb{R}^+$. Ces modèles de PDF seront utilisés pour décrire des données positives associées à l'intensité ou l'amplitude de données SAR, ou encore à des caractéristiques de modèle CNN. Nous pouvons considérer l'opération de rectification propre aux fonctions d'activation des réseaux de neurones, qui vise à mettre à 0 les parties négatives, ou celle de troncation qui réduit l'intervalle d'étude entre 2 bornes $a$ et $b$. Dans la suite de ce travail, nous considérons uniquement la troncation d'une fonction de densité de probabilité $g$ sur le support $]0\,;+\infty[$. Cette transformation se fait en deux étapes :

- considérer la restriction $g_+$ de $g$ définie par $g_+ = g\mathbb{1}_{\mathbb{R}^+}$ ;
- calculer la PDF tronquée comme $f = \lambda g_+$ où $\lambda$ correspond à un paramètre de normalisation spécifié pour satisfaire la condition de somme unitaire sur $f$.

La facilité de l'estimation des paramètres de la gaussienne et sa capacité à modéliser aux premiers ordres une distribution positive, nous amène à considérer la troncature de la distribution Gaussienne généralisée (GGau) comme modèle de référence pour la modélisation de données positives. À partir de la formulation de troncation, et en considérant la distribution Gaussienne généralisée (GGau) comme $g$, nous obtenons :

**Definition 6.1.1** (GGau tronquée : RGGau) *La RGGau associée avec un paramètre de position ou décalage $\mu \in \mathbb{R}$, un paramètre d'échelle $\sigma > 0$ et un paramètre*

*de forme $\beta > 0$ est définie par :*

$$f^{RGGau}_{\mu,\sigma,\beta}(x) = \frac{1}{\Lambda(\mu,\sigma,\beta)} \exp -\left(\frac{|x-\mu|}{\sigma}\right)^{\beta} \mathbb{1}_{\mathbb{R}^+}(x), \tag{6.1}$$

*où $\Lambda(\mu,\sigma,\beta)$ est la fonction de normalisation de la RGGau.*

**Proposition 6.1.1** (Constante de normalisation de la RGGau) *Pour le modèle RGGau défini par* (6.1)*, la constante de normalisation $\Lambda(\mu,\sigma,\beta)$ est donnée par :*

$$\Lambda(\mu,\sigma,\beta) = \begin{cases} \sigma\Gamma\left[1+\frac{1}{\beta}\right] + \frac{\sigma}{\beta}\left(\Gamma\left[\frac{1}{\beta}\right] - \Gamma\left[\frac{1}{\beta}, \left(\frac{\mu}{\sigma}\right)^{\beta}\right]\right) & \text{si } \mu \geqslant 0 \\ \frac{\sigma}{\beta}\left(\Gamma\left[\frac{1}{\beta}, \left(\frac{-\mu}{\sigma}\right)^{\beta}\right]\right) & \text{sinon} \end{cases} \tag{6.2}$$

En plus de (6.1), nous considérons la classe des distributions à deux paramètres Weibull (Wb) dont le support par construction est $\mathbb{R}^+$. Ce choix de distribution est motivé par le fait que la distribution de Weibull donne la distribution exponentielle décroissante lorsque $\beta = 1$, et la distribution de Rayleigh lorsque $\beta = 2$ et $\lambda = \sqrt{2}\alpha$[72] . Ces deux distributions sont respectivement utilisées pour décrire l'intensité et l'amplitude des données SAR et leur PDF est donnée tableau 2.3.

**Précision (72)**
$\alpha$ étant le paramètre de forme de la distribution de Rayleigh

Il est intéressant de noter que lorsque $\mu < 0$ dans (6.1), alors RGGau se réduit à une courbe exponentielle décroissante qui coïncide avec certaines formes de Weibull. Dans ce cadre joint où nous considérons également cette distribution, nous exclurons ces fonctions redondantes en spécifiant $\mu \geqslant 0$.

**Proposition 6.1.2** *La restriction de la RGGau définie par* (6.1) *à $\mu \geqslant 0$ est donnée par :*

$$\Lambda(\mu,\sigma,\beta) = \sigma\Gamma\left[1+\frac{1}{\beta}\right] + \frac{\sigma}{\beta}\left(\Gamma\left[\frac{1}{\beta}\right] - \Gamma\left[\frac{1}{\beta}, \left(\frac{\mu}{\sigma}\right)^{\beta}\right]\right) \tag{6.3}$$

Par la suite, on considère le paramètre de localisation entre deux distributions RGGau $p$ et $q$, $\overline{\mu} = 0.5(\mu_p + \mu_q)$ comme le centroïde des paramètres de localisation $\mu_{(\cdot)}$. En effet, sans cette contrainte, les formes suivantes de divergences n'admettent pas à notre connaissance de formulation analytique. Afin de garder l'information de localisation, nous ajoutons la distance sur les paramètres décrits dans la section 6.1.1 : $\mathcal{D}_{loc} = ||\mu_p - \mu_q||_2$. Pour la divergence entre deux RGGau, nous supposons un alignement de variables latentes, comme si les variables RGGau étaient distribuées par rapport à un paramètre de localisation de centroïde $\overline{\mu}$.

Pour quantifier la distance ou dissimilarité entre deux distributions positives, nous considérons la version symétrique de la divergence de Kullback-Leibler : divergence de Jeffreys dont l'équation est donnée par (3.5). Dans la suite, nous considérons les variables aléatoires $X_k^{\text{RGGau}}$ associées à la PDF $f^{\text{RGGau}}_{\mu_k,\sigma_k,\beta_k}$.

**Proposition 6.1.3** *À partir du calcul de l'entropie et de l'entropie croisée, la divergence de Kullback-Leibler entre deux variables aléatoires $X_p$ et $X_q$ suivant une*

*loi RGGau de paramètres* $(\overline{\mu}, \sigma_p, \beta_p)$ *et* $(\overline{\mu}, \sigma_q, \beta_q)$ *est donnée par :*

$$\begin{aligned}\mathcal{K}(X_p^{RGGau}||X_q^{RGGau}) &= \frac{\sigma_p}{\beta_p \cdot \Lambda(\overline{\mu}, \sigma_p, \beta_p)} \\ &\cdot \left[M(p,q) - N(p,q)\right] + \log\left[\frac{\Lambda(\overline{\mu}, \sigma_q, \beta_q)}{\Lambda(\overline{\mu}, \sigma_p, \beta_p)}\right], \\ \text{avec } M(p,q) &= \left(\frac{\sigma_p}{\sigma_q}\right)^{\beta_q}\left[2\Gamma\left(\frac{1+\beta_q}{\beta_p}\right) - \Gamma\left(\frac{1+\beta_q}{\beta_p}, \left(\frac{\overline{\mu}}{\sigma_p}\right)^{\beta_p}\right)\right] \\ \text{et } N(p,q) &= \frac{2}{\beta_p}\Gamma\left(\frac{1}{\beta_p}\right) - \Gamma\left(1+\frac{1}{\beta_p}, \left(\frac{\overline{\mu}}{\sigma_p}\right)^{\beta_p}\right)\end{aligned} \tag{6.4}$$

De cette proposition, nous en déduisons :

**Proposition 6.1.4** *La divergence de Jeffreys entre deux variables aléatoires* $X_p$ *et* $X_q$ *suivant une loi RGGau de paramètres* $(\overline{\mu}, \sigma_p, \beta_p)$ *et* $(\overline{\mu}, \sigma_q, \beta_q)$ *est donnée par :*

$$\begin{aligned}\mathcal{J}(X_p^{RGGau}||X_q^{RGGau}) = &\frac{\sigma_p}{\beta_p \cdot \Lambda(\overline{\mu}, \sigma_p, \beta_p)} \cdot \left[M(p,q) - N(p,q)\right] + \\ &\frac{\sigma_q}{\beta_p \cdot \Lambda(\overline{\mu}, \sigma_q, \beta_q)} \cdot \left[M(q,p) - N(q,p)\right]\end{aligned} \tag{6.5}$$

*avec* $M(p,q)$ *et* $N(p,q)$ *définis dans la proposition 6.1.3.*

La deuxième divergence symétrique considérée est celle portant sur les marginales suivant une distribution de Weibull.

**Definition 6.1.2** *La divergence de Jeffreys entre deux variables aléatoires* $X_p$ *et* $X_q$ *suivant une loi de Weibull est définie par [212] :*

$$\begin{aligned}\mathcal{J}(X_p^{Wbl}, X_q^{Wbl}) = &\Gamma\left(1+\frac{\beta_q}{\beta_p}\right)\left(\frac{\alpha_p}{\alpha_q}\right)^{\beta_q} \\ &+ \Gamma\left(1+\frac{\beta_p}{\beta_q}\right)\left(\frac{\alpha_q}{\alpha_p}\right)^{\beta_p} \\ &+ (\beta_p - \beta_q)\cdot\log\left(\frac{\alpha_p}{\alpha_q}\right) + \varepsilon\cdot\left(\frac{\beta_p}{\beta_q} + \frac{\beta_q}{\beta_p} - 2\right) - 2\end{aligned},$$

*où* $\varepsilon$ *est la constante d'Euler-Mascheroni*[73] .

[212]: Kwitt et al. (2008), « Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval »

**Définition (73)**

*Constante d'Euler-Mascheroni:* Définie par

$$\varepsilon = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \ln(n)\right),$$

elle est approchée par $\varepsilon \approx 0.577215$. Cette constante est liée à la fonction $\Gamma$ par : $\varepsilon = \Gamma'(1)$.

### Copules sur données bivariées

Nous proposons d'utiliser des modèles de copules pour modéliser les dépendances des données bivariées. Comme détaillé dans la partie sur les copules de la sous-section 3.1.1, les copules permettent de séparer les contributions de chaque marginale et de la dépendance entre elles. Une copule $C$ est une fonction de distribution cumulative conjointe (CDF) définie sur le cube unité d-dimensionnel $[0\,;1]^d$, telle que chaque marginale soit uniforme sur $[0;1]$. La restriction à des copules bivariées vient de l'utilisation dans ce travail, de données SAR à deux composantes polarimétriques, comme Sentinel-1 ou PAZ.

Pour cette modélisation de dépendance, nous avons choisi de travailler avec des copules de type EFGM, pour sa forme simple et sa capacité à approximer au premier ordre des copules plus complexes comme les copules de type Frank et Plackett [199].

[199]: Nelsen (1999), *An Introduction to Copulas*

Pour l'estimation et par extension le calcul de divergence entre les copules, nous avons choisi d'utiliser la méthode non-paramétrique basée sur les histogrammes [208] pour transformer les marginales en variables aléatoires uniformes sur $[0,1]$.

[208]: Jaworski et al. (2010), « Copula Theory and Its Applications »

**Definition 6.1.3** (Copule EFGM) *Soient deux variables aléatoires uniformes $U$ et $V$ sur $[0,1]$, sous l'hypothèse d'une copule bivariée de type EFGM paramétrée par $\rho$, la PDF du couple $(U,V)$ est donnée par [362] :*

[362]: (2010), *Copula Theory and Its Applications*

$$c(u,v) = 1 + \rho(1-2u)(1-2v)\mathbb{1}_{[0,1]\times[0,1]}(u,v) \tag{6.6}$$

Nous proposons d'utiliser la divergence entre la modélisation de la dépendance des marginales par la copule EFGM. Il est important de noter que cette divergence caractérise uniquement la dépendance entre les marginales des distributions de EFGM et non la structure complète de la copule. Nous avons exploré la piste du calcul de la divergence de Kullback-Leibler sur la copule dans sa totalité, mais cette approche n'a pas abouti à une formulation analytique.

**Proposition 6.1.5** (Jeffreys divergence EFGM) *La divergence de Jeffreys entre deux variables aléatoires $X_p$ et $X_q$ suivant une copule EFGM est donnée par :*

$$\mathcal{J}(\mathbf{X}_p^{EFGM}, \mathbf{X}_q^{EFGM}) = \frac{3}{2}\left(\frac{\rho_p}{\rho_q} + \frac{\rho_q}{\rho_p} - 2\right) + \frac{1}{2}\sum_{k=1}^{3}\mathbb{Q}_k(\rho_p,\rho_q), \tag{6.7}$$

*Avec les fonctions $\mathbb{Q}_k$ définies par :*

$$\mathbb{Q}_1(\rho_p,\rho_q) = \left|\rho_p - \rho_q\right|\left|\tanh^{-1}(\rho_p) - \tanh^{-1}(\rho_q)\right|, \tag{6.8}$$

$$\begin{aligned}\mathbb{Q}_2(\rho_p,\rho_q) = &\left(1-\frac{\rho_q}{\rho_p}\right)\frac{\tanh^{-1}(\rho_p)}{\rho_p} \\ &+\left(1-\frac{\rho_p}{\rho_q}\right)\frac{\tanh^{-1}(\rho_q)}{\rho_q},\end{aligned} \tag{6.9}$$

$$\begin{aligned}\mathbb{Q}_3(\rho_p,\rho_q) = &\left(1-\frac{\rho_q}{\rho_p}\right)\frac{Li_2(\rho_p) - Li_2(-\rho_p)}{\rho_p} \\ &+\left(1-\frac{\rho_p}{\rho_q}\right)\frac{Li_2(\rho_q) - Li_2(-\rho_q)}{\rho_q},\end{aligned} \tag{6.10}$$

*avec $Li_2$ la fonction poly-logarithme de degré 2 (Di-logarithme cf Annexe A.1).*

Nous complétons cette divergence avec la distance $\ell_2$ entre les distributions de FGM définie par :

**Proposition 6.1.6** (EFGM $\ell_2$ distance)

$$\ell_2(\mathbf{X}_p^{EFGM}, \mathbf{X}_q^{EFGM}) = \frac{1}{9}(\rho_p - \rho_q)^2 \tag{6.11}$$

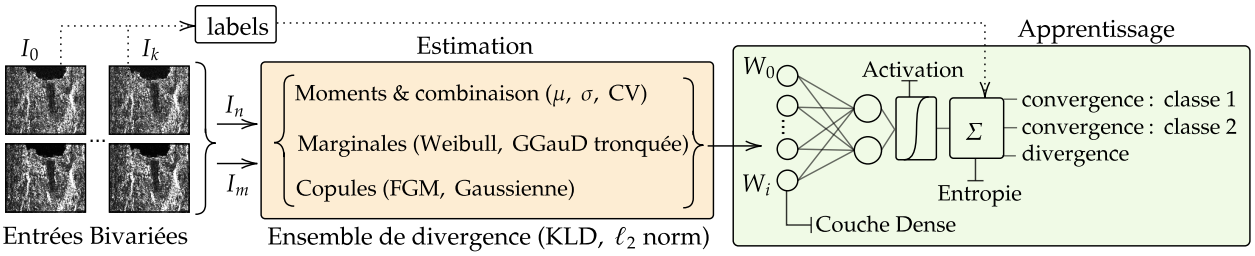



**Figure 6.1 –** Schéma de l'architecture de l'apprentissage d'ensembles de divergence.

Afin de parfaire notre ensemble de divergence, nous proposons d'utiliser la divergence de Kullback-Leibler entre 2 copules Gaussiennes bivariée [191]. Ce choix est motivé par le fait que la copule Gaussienne est une copule de référence simple à mettre en oeuvre et à estimer, mais qui permet de couvrir un grand nombre de cas [200]. De plus cette divergence permet de séparer la contribution des marginales de celle de la copule, comme indiqué dans l'équation 14 de [191]. Cependant, de manière similaire à la divergence de Jeffreys sur les copules EFGM, nous restreignons notre analyse à la contribution de la copule.

[191]: Lasmar et al. (2014), « Gaussian Copula Multivariate Modeling for Texture Image Retrieval Using Wavelet Transforms »

[200]: Bhatti et al. (2019), « Recent development in copula and its applications to the energy, forestry and environmental sciences »

**Proposition 6.1.7** *La divergence de Jeffreys entre deux variables aléatoires $X_p$ et $X_q$ suivant une copule Gaussienne bivariée est donnée par :*

$$\mathcal{J}(\mathbf{X}_p^{GauC}, \mathbf{X}_q^{GauC}) = (1 - \rho_p \rho_q)\left(\frac{1}{1-\rho_p^2} + \frac{1}{1-\rho_q^2}\right), \tag{6.12}$$

*avec $\rho_p, \rho_q$ la corrélation entre les marginales de la copule respectivement pour les variables, $\mathbf{X}_p^{GauC}$ et $\mathbf{X}_q^{GauC}$.*

Cette restriction sur l'utilisation de la seule modélisation de la dépendance à un double intérêt. Elle permet d'obtenir des formes analytiques simples. De plus, dans le cas de la copule gaussienne, l'utilisation des divergences sur les marginales présentées dans la section 6.1.1, permet d'obtenir une divergence sur la copule complète par la somme entre la divergence sur la copule gaussienne et n'importe quelles marginales.

### 6.1.2 Apprentissage de l'ensemble de divergence

Nous considérons un cadre d'apprentissage supervisé pour l'agrégation des distances et divergences estimées dans la partie 6.1.1, tel qu'illustré dans la figure 6.1. À partir de l'ensemble de divergences, estimées sur les paires d'images d'entrée, nous cherchons ainsi à minimiser la fonction suivante

$$\mathcal{L} = \mathcal{H}\left[\mathbf{y}, \mathcal{A} \circ f_\theta(\mathbf{d})\right]$$

avec $\mathcal{H}$ l' entropie croisée[74] , $\mathcal{A}$ une fonction d'activation softmax[75] , $\mathbf{d}$ le vecteur d'entré constitué de toutes les distances et divergences et $f_\theta$ une architecture neuronale caractérisée par un jeu de paramètres $\theta$.

**Précision (74)**
Voir l'équation (3.10) de la partie 49

**Précision (75)**
Voir l'équation 47 de la partie 47

Nous explorons des architectures neuronales de type perceptron multi-couches (MLP) uni-dimensionnelles pour apprendre la combinaison optimale des

| SAR | stat. desc. | marginales | copules |
|---|---|---|---|
| HH | $\mu, \sigma$, CV | RGGau, Wbl | EFGM, Gaussienne |
| HV | $\mu, \sigma$, CV | RGGau, Wbl | EFGM |
| HH/HV | $\mu, \sigma$, CV | RGGau, Wbl | |

**Table 6.1 –** Récapitulatif des modèles utilisés en fonction des données polarimétriques SAR, en bleu les distances basées sur les divergences de Jeffreys et en noir celle sur les normes $\ell_2$

distances et divergences. Nous considérons 3 architectures neuronales $f_\theta$ différentes :

**a.** une couche dense (*fully connected*) ($\mathbf{y} = \mathbf{X}\mathbf{w} + b$, $\mathbf{w}$ et $b$ étant les paramètres d'apprentissages),
**b.** une couche dense avec activation suivie d'un minimum pooling et d'une couche dense,
**c.** deux couches denses successives.

L'objectif est d'apprendre la combinaison optimale de l'ensemble de divergence dans le cadre d'une classification binaire. Le résultat de l'apprentissage permet de déterminer soit la classe à laquelle appartient le couple d'images en entrée, soit s'il existe une différence de classe entre les deux images. Nous contraignons la positivité des paramètres $\theta$ afin de garder une explicabilité dans la décision finale de classification. L'opérateur de minimum pooling permet de réduire la dimensionnalité d'un vecteur en donnant la valeur minimale sur une fenêtre glissante de taille fixe. Le choix du minimum se justifie par l'utilisation de distances ou de leurs combinaisons, plus la distance est faible, plus les images sont similaires. Dans le cas de l'utilisation successive de couches denses, il est évident que le chemin d'explicabilité est plus complexe à suivre, mais peut être encore utilisable pour comprendre la décision finale. Nous utilisons la fonction d'activation softmax pour obtenir une probabilité de classe pour chaque couple d'images. Dans ce cadre, la classe de dissimilarité est considérée comme une classe supplémentaire.

### 6.1.3 Résultats expérimentaux

**Précision (76)**
Les codes de ce travail sont donnés sur Github

Les développements ont été effectués conjointement en Matlab et Python[76] . Nous réalisons l'évaluation de cette approche sur un ensemble de données SAR en bande X du satellite PAZ, acquises entre juillet 2020 et août 2020 sur le massif du Mont-Blanc. Au total, 774 images de dimension 32 par 32 pixels sur 2 canaux polarimétriques (HH et HV) ont été sélectionnées. Les canaux HH, HV sont en amplitude linéaire. Nous complétons les canaux par l'ajout de la valeur absolue du log-ratio HH/HV. Un descriptif des mesures de similarité utilisées est donné dans le tableau 6.1, identifiant pour chaque canal la mesure utilisée pour le descripteur statistique.

Ces échantillons sont répartis en 2 classes : forêt et prairie, respectivement (390 et 384 échantillons). La labellisation des images a été réalisée en utilisant le modèle de couverture de sol CORINE fournit par Copernicus [363]. Le choix de ces classes repose sur l'homogénéité des zones considérées et la différence d'interactions polarimétriques entre une rétrodiffusion volumique et de surface. De plus, cette sélection en été permet d'assurer un signal relativement stable, sans perturbation due à la neige ou à la végétation en phase de croissance. La construction de l'ensemble de divergences se fait en estimant 18 divergences ou distances, détaillées dans le tableau 6.1, pour chaque paire d'images du dataset. Le label associé à un couple d'images étant : *forêt*, *prairie* ou *divergent*.

[363]: Kosztra et al. (2019), *Updated CLC Illustrated Nomenclature Guidelines*

**Table 6.2 –** Score d'apprentissage des différentes combinaisons

| **Architecture** | **a** | **b** | **c** |
|---|---|---|---|
| **F1-score (%)** | 89.4 ± 0.1 | 94.7 ± 0.4 | 91.2 ± 0.2 |
| **Nombre paramètres** | 57 | 195 | 201 |

Les résultats des différentes architectures sont donnés dans le tableau 6.2 en réalisant une validation croisée par la méthode du K-Fold (avec K=5) permettant de parcourir l'ensemble du dataset avec la métrique F1-score.

Nous remarquons que les 3 architectures donnent des résultats moyens de l'ordre de ∼ 91.7%. L'architecture **b** reste la plus performante avec une précision de près de 5% supérieure à celle de l'architecture **a**, de par l'utilisation d'un nombre plus important de paramètres (195). Cependant, l'architecture **a** est intéressante par une variabilité plus faible que les autres architectures sur l'ensemble de folds et un faible nombre de paramètres (57). Nous pouvons néanmoins noter que l'architecture **b** donne une variabilité plus importante que les deux autres architectures. L'avantage principal de l'architecture **a** est son faible nombre de paramètres. Ces paramètres sont directement utilisables pour comprendre la décision finale puisqu'ils correspondent à la combinaison des différentes divergences. Cette explicabilité est représentée dans la figure 6.2 où nous donnons les poids et le biais associés à chaque distance et ce pour chaque classe. Les correspondances entre les indices des distances et divergences utilisées dans la figure 6.2 et les mesures statistiques utilisées sont données dans le tableau 6.3. Cette analyse post-apprentissage permet de remarquer que la classe *divergence* utilise l'ensemble des distances pour rendre sa décision, avec un poids plus important pour les distances basées sur la moyenne et l'écart-type en polarisation HV et la copule EFGM. Les autres classes rejettent certaines distances, ainsi la classe *prairie* se concentre moins sur une distance entre Weibull sur le canal HV et se concentre sur la divergence entre copules EFGM. Il est intéressant de noter l'importance de l'utilisation d'un modèle a priori sur les distances pour HV. Les descripteurs de moyenne et d'écart-type ont les poids à 0, là où le poids sur la distribution de Weibull est le plus important pour cette classe. La classe *forêt* utilise majoritairement les statistiques descriptives en prenant peu en compte l'interaction entre les canaux HV et HH.

Il est toutefois important de noter que les distances et divergences sont faibles en présence d'une paire similaire. Ainsi l'analyse des poids appris pour combiner ces informations n'est pas si simple à interpréter puisqu'un poids important sur une similarité en moyenne faible peut être équivalent à un poids faible sur une similarité en moyenne importante. Plus qu'une analyse des poids, il conviendrait d'étudier l'importance des caractéristiques de manière similaire aux méthodes utilisées pour les forêts aléatoires, en utilisant par exemple la méthode de permutation de Shapley.

**Table 6.3 –** Tableau des correspondances entre les indices des distances et divergences utilisés dans la figure 6.2 et les mesures statistiques utilisées. La première lettre représentant la nature de la distance ($\mathcal{H}$ ou $\ell_2$), le deuxième bloc le modèle paramétrique (RGGA, Weibull, copule gaussienne ou EFGM) et le troisième bloc le canal considéré (HH, HV ou R= HH/HV).

| **Indice** | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Mesures.** | $\mathcal{H}$ GAUSS COP | $\mathcal{H}$ EFGM COP | $\ell_2$ EFGM COP | $\ell_2$ CV HH | $\ell_2$ CV HV | $\ell_2$ CV R | $\mathcal{H}$ WBUL HH | $\mathcal{H}$ WBUL HV | $\mathcal{H}$ WBUL R | |
| **Indice** | **J** | **K** | **L** | **M** | **N** | **O** | **P** | **Q** | **R** | **S** |
| **Mesures.** | $\ell_2$ MU HH | $\ell_2$ MU HV | $\ell_2$ MU R | $\mathcal{H}$ RGGA HH | $\mathcal{H}$ RGGA HV | $\mathcal{H}$ RGGA R | $\ell_2$ VAR HH | $\ell_2$ VAR HV | $\ell_2$ VAR R | BIAIS |

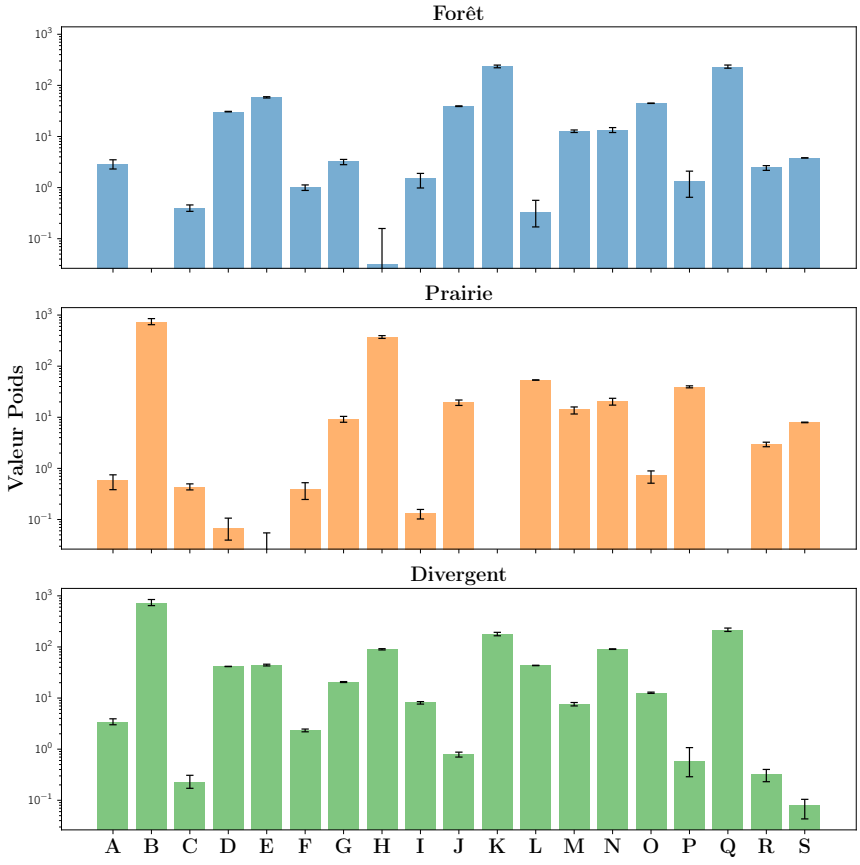



**Figure 6.2 –** Valeur des poids et du biais pour les 3 classes considérées (2 thématiques et la classe divergence) en fonction des distances considérées. Les identifiants sont donnés dans le tableau 6.3

## 6.2 Apprentissage d'ensemble de divergences paramétriques

Nous avons pu observer dans la section précédente l'intérêt de l'utilisation de modèle a priori sur les données pour diminuer le nombre de paramètres à apprendre, tout en gardant une explicabilité par la contrainte de positivité des poids, qui informe de la contribution de chaque distribution et distances utilisées dans la classification.

Cependant le choix de la distance à utiliser pour la classification reste un problème ouvert et un compromis doit être trouvé entre un dictionnaire très large de distances et de distributions, et donc de paramètres à apprendre, et un dictionnaire trop restreint qui ne permettrait pas de capturer la variabilité des données. Lorsqu'il s'agit de classification sur des modèles statistiques, les méthodes basées sur la divergence statistique sont populaires [211]. Cependant, les limitations de ces approches résident dans le fait qu'elles reposent soit sur des distributions a priori et calculent des distances entre des divergences empiriques [217], soit sur un seul modèle de probabilité a priori [220] alors qu'il est souvent difficile de modéliser la variabilité des données à partir d'une seule famille de distributions de probabilité. Dans le second cas, les réseaux de neurones sont capables d'apprendre les caractéristiques des champs imagés grâce à une grande quantité de données. Cependant, dans le contexte SAR, ces réseaux manquent souvent de généralisation [15] et de transparence décisionnelle en raison du nombre important de paramètres impliqués.

Nous proposons dans cette section, tout en gardant la structure présentée dans

[211]: Basseville (2013), « Divergence measures for statistical data processing–An annotated bibliography »

[217]: Cilingir et al. (2020), « Deep divergence learning »

[220]: Silva et al. (2013), « Classification of segments in PolSAR imagery by minimum stochastic distances between Wishart distributions »

[15]: Hänsch et al. (2017), « Correct and Still Wrong »

la section précédente, de s'intéresser au choix optimal de la distance à utiliser pour la classification par l'utilisation de la divergence paramétrique de Rényi. Étant donné une sélection de distributions propre à la modélisation de données SAR, nous proposons d'apprendre le paramètre $\alpha$ de la divergence à partir des données et de dériver les gradients nécessaires pour effectuer une optimisation de bout en bout, ce qui, à notre connaissance, n'a pas encore été étudié dans ce scénario. Nous fournissons notamment des expressions analytiques pour la dérivée pour les distributions log-normale, gamma et Rayleigh. Nous réalisons l'évaluation de cette proposition sur un jeu de données SAR réel et démontrons à la fois la nature explicable des paramètres appris et la meilleure robustesse de l'approche, en comparaison avec les algorithmes CNN et Random Forest en étendant le cadre de la classification binaire à la classification multi-classe.

**Modèlisations statistiques des données SAR**

Nous simplifions le dictionnaire de caractérisation présenté dans la section 6.1, en nous concentrant uniquement sur des distributions statistiques monovariées, avec un ou deux paramètres. Ce choix est motivé par deux principaux facteurs :

1. la volonté de réduire le nombre de caractéristiques de nature différente, en se concentrant uniquement sur celle permettant l'utilisation de la divergence paramétrique de Rényi,
2. la volonté de garder une structure simple pour la classification, en utilisant des distributions monovariées, dont l'estimation des paramètres est simple et dont la forme analytique de la divergence de Rényi est connue.

[113]: Deledalle et al. (2018), « Mathematical models for very high resolution SAR data and their applications »

Nous considérons des images SAR en amplitude, ce qui signifie que les valeurs des pixels appartiennent à $\mathbb{R}^+$. Pour modéliser ces données, de nombreux modèles probabilistes ont été proposés [113] afin de tenir compte de la variabilité des scènes observées. Nous considérons les distributions gamma, log-normale et Rayleigh, dont les fonctions de densité de probabilité sont données dans le tableau 2.3. Ces dernières ont montré leur capacité à modéliser avec précision les données SAR en amplitude avec peu de paramètres (1 ou 2) facilement estimables [112]. Pour estimer les paramètres de ces modèles probabilistes, nous nous appuyons sur une estimation des log-cumulants des distributions qui sont directement liés aux paramètres statistiques. Cette considération permet de trouver un compromis entre une bonne estimation et un coût de calcul très faible [364].

[112]: Yue et al. (2021), « Synthetic Aperture Radar Image Statistical Modeling »

[364]: Nicolas (2018), « Application de la transformée de Mellin : étude des lois statistiques de l'imagerie cohérente »

L'estimation du log-cumulant d'ordre $r$ est donnée par la formule de récurrence donnée dans la sous section par l'équation 2.5. Dans notre cas, le lien entre les paramètres des trois distributions et les log-cumulants est donné par :

**Definition 6.2.1** *Le lien entre les log-cumulants $\hat{\kappa}_1$ et $\hat{\kappa}_2$ empiriquement estimés et les paramètres de la distribution gamma $\mu$ et $L$ est donné par :*

$$\begin{cases} \hat{\kappa}_1 &= \log(\mu) + \Psi(L) - \log(L) \\ \hat{\kappa}_2 &= \Psi(1, L) \end{cases} \tag{6.13}$$

*avec $\Psi(.)$ la fonction digamma et $\Psi(.,.)$ la fonction polygamma, dont le détail est donné dans le tableau A.1.*

La résolution se fait d'abord en résolvant de manière numérique $\hat{\kappa}_2 = \Psi(1, L)$. Cette étape est peu complexe, par la valeur de $L$ qui est supérieure ou égale à 1 (nombre de vues) induisant la nature monotone des fonctions polygamma dans ce cas.

**Definition 6.2.2** *Le lien entre les log-cumulants $\hat{\kappa}_1$ et $\hat{\kappa}_2$ empiriquement estimés et les paramètres de la distribution log-normale $\mu$ et $\sigma$ est donné par :*

$$\begin{cases} \hat{\kappa}_1 &= \mu \\ \hat{\kappa}_2 &= \sigma^2 \end{cases} \tag{6.14}$$

L'estimation dans le cas de la distribution log-normale est direct.

**Definition 6.2.3** *Le lien entre le log-cumulant $\hat{\kappa}_1$ empiriquement estimé et le paramètre de la distribution Rayleigh $\mu$ est donné par :*

$$\hat{\kappa}_1 = \log(\mu) + \frac{\psi(1)}{2} \tag{6.15}$$

*Avec $\psi(1)$ la fonction digamma évaluée en 1 soit $\psi(1) = -\varepsilon$ avec $\varepsilon$ la constante d'Euler-Mascheroni$\approx 0.5772$.*

L'estimation dans le cas de la distribution Rayleigh est tout aussi directe.

### 6.2.1 Divergences de Rényi paramétriques

[223]: Van Erven et al. (2014), « Rényi divergence and Kullback-Leibler divergence »

Concernant les divergences, nous nous appuyons sur la paramétrisation de Rényi pour ses grandes capacités de généralisation par rapport à une large classe de divergences bien connues [223]. La divergence de Rényi discrète d'ordre $\alpha$ entre deux distributions de probabilité $P, Q \in \mathbb{R}^n$ peut être exprimée comme suit :

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \ln \sum_i p_i^\alpha q_i^{1-\alpha}, \tag{6.16}$$

avec $\alpha > 0$ et $\alpha \neq 1$. Cette divergence paramétrique est intéressante en raison de sa capacité à représenter la divergence de Kullback-Leibler lorsque $\alpha$ tend vers 1. Elle englobe des cas particuliers tels que la distance de Bhattacharyya lorsque $\alpha = 0,5$ et devient proportionnelle à une divergence $\chi^2$ lorsque $\alpha = 2$ (cf tableau A.1).

[365]: Gil et al. (2013), « Rényi divergence measures for commonly used univariate continuous distributions »

Contrairement aux approches qui déterminent et fixent $\alpha$ pour une utilisation ultérieure, nous proposons un cadre où $\alpha$ est un paramètre apprenable par rapport aux données et aux distributions de classes, permettant ainsi l'utilisation des métriques appropriées pour notre problème d'optimisation. Grâce à [365], nous disposons des formes analytiques de la divergence de Rényi pour les trois distributions considérées, en fonction de leurs paramètres et du paramètre de la divergence de Rényi $\alpha$.

Les divergences de Rényi pour les distributions gamma, log-normale et Rayleigh proposées dans [365] sont données et adaptées dans le tableau 6.4 au vu des pdf données dans le tableau 2.3. Avec les paramètres $\left(\sigma^2\right)^*_\alpha$, $\theta^*_\alpha$ et $k_\alpha$

définis comme suit :

$$\begin{aligned}\left(\sigma^2\right)_\alpha^* &= \alpha\sigma_j^2 + (1-\alpha)\sigma_i^2 \\ \theta_\alpha^* &= \alpha\frac{\mu_j}{L_j} + (1-a)\frac{\mu_i}{L_i} \\ k_\alpha &= \alpha L_i + (1-\alpha)L_j\end{aligned} \tag{6.17}$$

La divergence de Rényi n'est pas définie pour $\alpha = 1$, pour contourner cette limitation, nous restreignons l'ensemble des valeurs possibles à $\alpha \in [0, 1[$, soit la partie basse du domaine de définition. Cette restriction permet néanmoins de conserver la propriété de convergence de la divergence de Rényi vers la divergence de Kullback-Leibler lorsque $\alpha$ tend vers 1 et la distance de Bhattacharyya. Nous faisons le choix de ne pas utiliser la partie haute du domaine de définition de $\alpha$ (i.e. $\alpha \in ]1, +\infty[$) simplement pour éviter un trop grand nombre de paramètres à apprendre dans un premier temps et éviter aussi les doublons potentiels lorsque $\alpha \longrightarrow 1^-$ et $\alpha \longrightarrow 1^+$. La restriction est réalisée par une reparamétrisation :

$$\hat{\alpha} = \frac{1}{1+e^{-\alpha}} \tag{6.18}$$

### 6.2.2 Cadre d'apprentissage par divergence

Nous considérons un problème de classification supervisée multi-classes. Étant donné un ensemble de $n$ données étiquetées $E = \{(\mathbf{X}_i, \mathbf{y}_i) \in \mathbb{R}^{v\times 2} \times \{0,1\}^m : 1 \leq i \leq n\}$, où $\mathbf{X}_i$ correspond à une paire de vecteurs de $v$ caractéristiques qui sont les paramètres estimés des distributions a priori. Ils sont associés à un vecteur d'étiquetage one-hot $\mathbf{y}_i(c) = [\delta_{1c}, \ldots, \delta_{mc}]^{\mathrm{T}}$, $m$ étant le nombre de classes, $\delta$ le symbole de Kronecker avec $c \in \{1, \ldots, m\}$, correspondant à la classe de la paire. L'une de ces classes correspond à une classe de dissimilarité entre les deux images de la paire.

**Précision (77)**
Informant sur le nombre de canaux considérés, dans le cas de données dual-pol, $j = 2$

Soit $\mathbf{I}_{i,j}$ une paire de patchs d'images SAR $j$-variées,[77] iid. Nous construisons la paire de vecteurs de caractéristiques $\hat{\mathbf{x}}_{i,j}$ comme suit :

$$\hat{\mathbf{x}}_{i,j} = \left[\theta_{\mathcal{G}}(\mathbf{I}_{i,j}), \theta_{\mathcal{O}}(\mathbf{I}_{i,j}), \theta_{\mathcal{R}}(\mathbf{I}_{i,j})\right]^{\mathrm{T}}, \forall i, j, \tag{6.19}$$

où $\theta_{\mathcal{G}}(\mathbf{I}_{i,j})$, $\theta_{\mathcal{O}}(\mathbf{I}_{i,j})$ et $\theta_{\mathcal{R}}(\mathbf{I}_{i,j})$ sont les paramètres des trois distributions sélectionnées et estimés sur l'amplitude du patch SAR $\mathbf{I}_{i,j}$. De cette formulation,

**Table 6.4 –** Divergences de Rényi pour les distributions gamma, log-normale et Rayleigh.

| **Distribution** | $D_\alpha(P_i \Vert Q_j)$ | **Conditions** |
|---|---|---|
| Gamma $\mathcal{G}$ | $\ln\left[\frac{\Gamma(L_j)}{\Gamma(L_i)}\left(\frac{\mu_j}{L_j}\right)^{L_j}\left(\frac{\mu_i}{L_i}\right)^{-L_i}\right] + \frac{1}{\alpha-1}\ln\left[\left(\frac{\mu_i}{L_i}\right)^{-L_i}\frac{\Gamma(k_\alpha)}{\Gamma(L_i)}\left(\frac{\mu_i\mu_j}{L_iL_j\theta_\alpha^*}\right)^{k_\alpha}\right]$ | $\theta_\alpha^* > 0$ and $k_\alpha > 0$ |
| Log-normal$\mathcal{O}$ | $\ln\frac{\sigma_j}{\sigma_i} + \frac{1}{2(\alpha-1)}\ln\left(\frac{\sigma_j^2}{(\sigma^2)_\alpha^*}\right) + \frac{1}{2}\frac{\alpha\left(\mu_i-\mu_j\right)^2}{(\sigma^2)_\alpha^*}$ | $\left(\sigma^2\right)_\alpha^* > 0$ |
| Rayleigh $\mathcal{R}$ | $2\ln\frac{\sigma_j}{\sigma_i} + \frac{1}{\alpha-1}\ln\left(\frac{\sigma_j^2}{(\sigma^2)_\alpha^*}\right)$ | $\left(\sigma^2\right)_\alpha^* > 0$ |

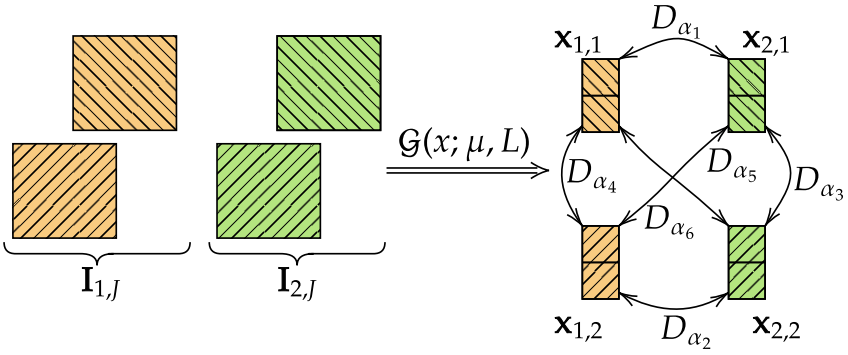



**Figure 6.3 –** Illustration de l'ensemble des divergences calculées sur des distributions Gamma estimées à partir d'une paire d'images SAR bivariées ($J = 2$).

nous obtenons un ensemble de caractéristiques $\mathbf{X}$ défini par

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2) = \begin{pmatrix} \hat{\mathbf{x}}_{1,1} & \hat{\mathbf{x}}_{2,1} \\ \vdots & \vdots \\ \hat{\mathbf{x}}_{1,J} & \hat{\mathbf{x}}_{2,J} \end{pmatrix} \tag{6.20}$$

Pour simplifier les formulations et pour l'adaptation de la proposition à des données SAR, nous considérons dans ce qui suit le cas bivarié ($J = 2$), illustré dans la figure 6.3 dans le cas d'une seule distribution.

Soit $\mathbf{f} = \left[D_{\alpha_1}(\mathbf{x}_1^1, \mathbf{x}_2^1), \ldots, D_{\alpha_p}(\mathbf{x}_1^p, \mathbf{x}_2^p)\right]^{\mathrm{T}}$ un vecteur composé d'un ensemble de $p$ divergences[78] de Rényi avec les paramètres $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p]^{\mathrm{T}} \in (0,1)^p$. Nous construisons un réseau, comme illustré dans la figure 6.4, qui calcule pour chaque classe $c$, un vecteur de divergences, étant donné un ensemble de paramètres $\boldsymbol{\alpha}_c$ entre la paire et les combine avec :

$$\begin{aligned} d_c(\mathbf{X}_i, \boldsymbol{\alpha}_c) &= \mathbf{w}_c^{\mathrm{T}} \mathbf{f}(\mathbf{X}_i, \boldsymbol{\alpha}_c) + b_c, \\ &\text{avec } \mathbf{w}_c \in \mathbb{R}_+^p, b_c \in \mathbb{R}_+. \end{aligned} \tag{6.21}$$

**Précision (78)**
Le nombre total de divergences est donnée par $p = \mathcal{S} \times \binom{2\times J}{2}$, $\mathcal{S}$ étant le nombre de distributions considérées (soit 3 dans notre cas), $J$ le nombre de canaux considérés pour une paire d'images

La positivité des poids et des biais est imposée pour préserver l'explicabilité de la combinaison finale. Ensuite, toutes les combinaisons obtenues sont introduites dans un opérateur softmax pour obtenir une distribution de probabilité sur les $m$ classes :

$$\sigma(\mathbf{d}_c) = \frac{\mathrm{e}^{d_c}}{\sum_i^m \mathrm{e}^{d_i}} \tag{6.22}$$

Soit $\hat{y}, y \in \mathbb{R}^m$ étant respectivement la prédiction d'un classifieur et sa vérité

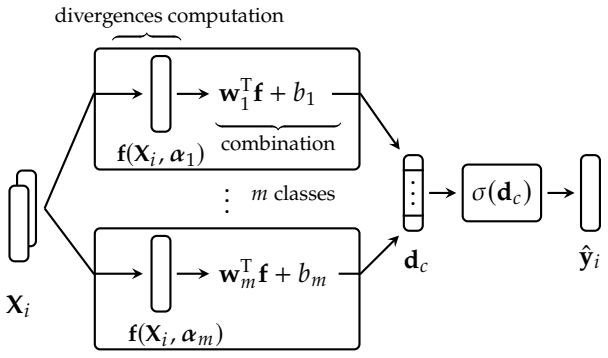



**Figure 6.4 –** Résumé de l'approche de classification proposée.

**Table 6.5 –** Dérivée de Rényi, avec $\lambda_{ij} = (e^{\alpha_l} L_i + L_j)/(1 + e^{\alpha_l})$, $\beta_{ij} = e^{\alpha_l} L_i \mu_j + L_j \mu_i$, $\Sigma_{ij} = \sigma_j^2 + \sigma_i^2$ et $\gamma_{ij} = e^{\alpha_l} \mu_j^2$.

| **Divergence** | **dérivée** $\partial D_{\hat{\alpha}_{(\cdot)}} / \partial \alpha_{(\cdot)}$ |
|---|---|
| $D_{\hat{\alpha}_l}(\mathcal{G}_i \| \mathcal{G}_j)$ | $e^{\alpha_l} \frac{L_i \mu_j - L_j \mu_i}{\lambda_{ij} \beta_{ij}} - e^{\alpha_l} L_i \log\left[\frac{(e^{\alpha_l}+1)\mu_i \mu_j}{\beta_{ij}}\right] - e^{\alpha_l} \log\left[\frac{\left(\frac{\mu_i}{L_i}\right)^{-L_i} \Gamma(\lambda_{ij})}{\Gamma(L_i)}\right] + e^{\alpha_l} \frac{(L_j - L_i)\psi^{(0)}(\lambda_{ij})}{e^{\alpha_l}+1}$ |
| $D_{\hat{\alpha}_l}(\mathcal{O}_i \| \mathcal{O}_j)$ | $\frac{e^{\alpha_l}}{2\left(e^{\alpha_l}\Sigma_{ij}\right)^2}\left[e^{\alpha_l}\sigma_j^2(\sigma_j^2 - \sigma_i^2) + \sigma_i^2\left[(\mu_i - \mu_j)^2 - \Sigma_{ij}\right] - \left(e^{\alpha_l}\Sigma_{ij}\right)^2 \log\left(\frac{(e^{\alpha_l}+1)\sigma_j^2}{e^{\alpha_l}\Sigma_{ij}}\right)\right]$ |
| $D_{\hat{\alpha}_l}(\mathcal{R}_i \| \mathcal{R}_j)$ | $\frac{\gamma_{ij} - e^{\alpha_l}\mu_i^2}{\gamma_{ij} + \mu_i^2} - e^{\alpha_l} \log\left[\frac{\gamma_{ij} + \mu_j^2}{\gamma_{ij} + \mu_i^2}\right]$ |

terrain associée, nous considérons la fonction de coût d'entropie croisée définie par :

$$H(y, \hat{y}) = -\sum_{c}^{m} y_c \log(\hat{y}_c), \tag{6.23}$$

Cela nous donne le problème de minimisation suivant :

$$\underset{\substack{\forall c \in \{1,\dots,m\},\\ \boldsymbol{\alpha}_c \in (0,1)^p,\\ \mathbf{w}_c \in \mathbb{R}_+^p, b_c \in \mathbb{R}_+}}{\arg\min} \frac{1}{n} \sum_{i=0}^{n} \underbrace{-\sum_{c}^{m} \mathbf{y}_i(c) \log\left[\sigma \circ d_c(\mathbf{X}_i, \boldsymbol{\alpha}_c)\right]}_{\mathscr{L}_i}. \tag{6.24}$$

Pour résoudre ce problème, nous utilisons une approche de descente de gradient sur les poids $\mathbf{w}_c$ et les biais $b_c$. Afin de respecter les contraintes de domaine sur les variables, nous utilisons une reparamétrisation[79] de la manière suivante :

- $\hat{\mathbf{w}} = \mathrm{e}^{\mathbf{w}}$
- $\hat{b} = \mathrm{e}^{b}$

**Précision (79)**
Nous avons fait le choix de reparamétriser les paramètres contraints dans un domaine donné par l'utilisation d'une tierce fonction respectant ce domaine. Cette solution n'est pas la plus idéale car elle introduit des fonctions de type exponentielle qui peuvent déstabiliser le gradient. D'autres solutions existent comme l'utilisation de gradient projeté ou de descente miroir [366], mais sont plus complexes à mettre en place.

Pour des raisons de concision, nous gardons la notation $\mathscr{L}_i$ tout en remplaçant les paramètres par leurs reparamétrisations. Pour calculer le gradient, nous calculons les dérivées de la perte, grâce à la règle de la chaîne :

$$\frac{\partial \mathscr{L}_i}{\partial d_c} = \hat{y}_c - y_c, \frac{\partial \mathscr{L}_i}{\partial \mathbf{w}_c} = \mathbf{f} \cdot \mathrm{e}^{\mathbf{w}_c} \cdot (\hat{y}_c - y_c),$$
$$\frac{\partial \mathscr{L}_i}{\partial b_c} = \mathrm{e}^{b_c} \cdot (\hat{y}_c - y_c),$$

Pour les paramètres de Rényi, nous avons :

$$\frac{\partial \mathscr{L}_i}{\partial \boldsymbol{\alpha}_c} = \frac{\partial \mathscr{L}_i}{\partial d_c} \frac{\partial d_c}{\partial \mathbf{f}} \frac{\partial \mathbf{f}}{\partial \boldsymbol{\alpha}_c}, \text{ où } \frac{\partial d_c}{\partial \mathbf{f}} = \mathbf{w}_c,$$

et $\frac{\partial \mathbf{f}}{\partial \boldsymbol{\alpha}_c}$ est une matrice de dérivées calculées grâce aux résultats donnés dans le tableau 6.5, avec les variables intermédiaires suivantes :

$$\begin{aligned} \lambda_{ij} &= (e^{\alpha_l} L_i + L_j)/(1 + e^{\alpha_l}) \\ \beta_{ij} &= e^{\alpha_l} L_i \mu_j + L_j \mu_i \\ \Sigma_{ij} &= \sigma_j^2 + \sigma_i^2 \\ \gamma_{ij} &= e^{\alpha_l} \mu_j^2 \end{aligned} \tag{6.25}$$

### 6.2.3 Expérimentations

#### Jeu de données SAR réel

Pour évaluer notre cadre, nous créons un jeu de données SAR en bande X à partir du satellite PAZ. Opérant à une fréquence de 9,65 GHz, ce satellite fournit des images à des intervalles multiples de 11 jours, offrant une résolution d'environ 3 mètres par pixel en deux polarisations, HH et HV. Nous examinons spécifiquement six acquisitions d'amplitude en linéaire allant du 7 juillet 2020 au 26 août 2020, couvrant le massif du Mont-Blanc. Ce choix s'explique par la volonté dans un premier temps de minimiser les impacts des variabilités saisonnières et climatiques. Au total, nous avons extrait 645 fenêtres de $32 \times 32$ pixels au cours de cette période. Ces fenêtres sont catégorisées en cinq classes distinctes : zone d'accumulation de glaciers, ville, forêt, plaine ou terre et zone rocheuse, respectivement nommées ACC, CIT, FOR, PLA et ROC pour les résultats. Chaque classe est également représentée par un jeu de 129 échantillons. La labellisation des images a été effectuée en utilisant le modèle de couverture du sol CORINE fourni par Copernicus[80] . Le cadre de ce travail repose sur l'évaluation de distance entre deux images. Par conséquent nous proposons de réaliser la tâche de classification entre tous les couples d'images de notre jeu de données. Nous considérons toutes les paires dans ce jeu de données en ajoutant la classe dissimilaire (DIS). Cette dernière est constitué de toutes les paires dissimilaires entre les 5 classes. Les données sont mélangées et équilibrées pour chacune des 6 classes.

**Précision (80)**
© Union européenne, Service de surveillance des terres de Copernicus 2018, Agence européenne pour l'environnement (AEE)

#### Architectures de modèles et paramétrisation

Nous comparons notre approche avec deux autres méthodes : RandomForest (RF) et CNN. RandomForest est efficace en termes de calcul et offre une transparence décisionnelle grâce à l'analyse des seuils, mais peut nécessiter de nombreux nœuds en fonction des caractéristiques, rendant l'interprétation complexe et sensible au bruit et aux valeurs aberrantes. Dans notre cas, nous vectorisons la paire d'images bivariées et appliquons une Analyse en Composantes Principales (ACP ou *Principal Component Analysis* PCA) pour expliquer 95% de la variance. Cette transformation réduit la dimensionnalité de 4096 pixels à 200. Nous utilisons ensuite une forêt aléatoire avec 100 arbres pour la classification.

D'autre part, les réseaux de neurones convolutifs (CNN) capturent des informations spatiales détaillées mais nécessitent un grand nombre d'échantillons et perdent en interprétabilité avec l'augmentation de la profondeur du réseau. En raison du nombre limité de données d'échantillons disponibles, nous avons créé un CNN peu profond comme proposé dans [12]. Le CNN prend en entrée une paire d'images bivariées, concaténées sur l'axe des caractéristiques, soit une taille de $32 \times 32 \times 4$ suivi de 3 blocs successifs. Chaque bloc comprend une convolution spatiale de $3 \times 3$ suivie d'une opération de max-pooling de $2 \times 2$. Le réseau implique ensuite une couche d'agrégation dense de taille 256, suivie d'une couche dense finale pour la prédiction de classe. Au total, ce CNN comprend 226 406 paramètres.

[12]: Parikh et al. (2020), « Classification of SAR and PolSAR images using deep learning »

Notre approche *Rényi Divergence Learning* (RDL) fonctionne sur les divergences entre les distributions de canaux d'image (gamma, log-normal et Rayleigh). Nous définissons ces combinaisons comme $\{\|; =; \perp\}, ij$, où $ij$ spécifie les images pour lesquelles la divergence est calculée :

- $\|$ représente les divergences de copolarisation, i.e. $D(HH_i, HH_j)$
- $=$ dénote les divergences de cross-polarisation, i.e. $D(HV_i, HV_j)$
- $\perp$ représente les divergences croisées entre copolarisation et cross-polarisation, entre les images $i$ et $j$ ou au sein d'une même image, i.e. $D(HH_i, HV_j)$, $D(HV_i, HH_j)$, $D(HH_i, HV_i)$ et $D(HH_j, HV_j)$

Cela fournit 18 divergences pour 6 classes, pondérées par 6 vecteurs avec 18 poids et 6 biais. Le nombre de paramètres et la taille des entrées sont décrits dans le tableau 6.6.

**Table 6.6 –** Nombre de paramètres et taille des entrées utilisées

| | RF | CNN | RDL |
|---|---|---|---|
| taille d'entrée | $200 \times 1$ | $32 \times 32 \times 4$ | $10 \times 2$ |
| paramètres | $\sim 152,000$ | 226,406 | 222 |

Il est intéressant de noter que malgré la volonté de minimiser le nombre de paramètres pour le CNN, notre approche nécessite 1000 fois moins de paramètres. Cependant il est nécessaire de garder à l'esprit que notre approche estime dans un premier temps des paramètres de distributions et à chaque pas de gradient calcule des divergences plus complexes que des convolutions.

### 6.2.4 Résultats et discussion

#### Evaluation de la performance

**Précision (81)**
Élements diagonaux de la matrice de confusion (cf sous-section 3.1.3 sur la partie métriques) normalisé par le nombre d'échantillons de la classe.

Nous évaluons d'abord nos résultats en utilisant une validation croisée stratifiée (K-fold) avec K=5 pour maintenir l'équilibre des classes tout en analysant l'ensemble du jeu de données. Les résultats sont présentés dans le tableau 6.7, qui affiche les pourcentages de classification moyens[81] pour chaque classe selon les trois méthodes. Nous observons que le CNN et notre approche donnent des résultats similaires, surpassant le RF, mais avec des forces distinctes. Le CNN excelle dans la détection des glaciers, des zones rocheuses et des paires dissemblables, tandis que notre méthode se montre supérieure pour la classification des villes, des forêts et des plaines. Ceci s'explique bien, par le fait que les CNN sont plus adaptés pour capturer des informations spatiales détaillées (textures, motifs géométriques) en apprenant les filtres permettant de distinguer les classes. A l'inverse, notre approche se concentre sur une modélisation statistique a priori sans prendre en compte les dépendances spatiales. Ainsi entre les classes *forêt* et *plaine*, la variation des distributions de canaux est plus discriminante qu'une modélisation de textures. Nous pouvons également noter que notre approche présente en moyenne une variabilité plus faible ($\hat{\sigma} \approx 6.5\%$) que le CNN ($\hat{\sigma} \approx 8.3\%$), ce qui peut être attribué à la régularisation induite par la minimisation des divergences de Rényi. Nous pouvons observer pour la solution RF, la même tendance qu'observée pour le CNN sur l'ensemble des classes, mais avec des performances inférieures, avec une performance sur la classe *forêt* particulièrement faible (environ 7 fois inférieure à notre approche).

La disparité la plus importante en termes de performance est la classe dissemblable, où le CNN a obtenu des résultats environ deux fois meilleurs que notre approche. Cette différence peut s'expliquer par notre utilisation des divergences. En effet, nous utilisons uniquement les divergences entre les mêmes distributions. Par conséquent, en présence d'un échantillon dissemblable, la modélisation et la distance associée ne sont pas adaptées, là où le CNN et le

**Table 6.7 –** Pourcentage de bonne classification avec un K-Fold stratifié avec K=5 pour chaque classe.

| | ACC | CIT | FOR | PLA | ROC | DIS |
|---|---|---|---|---|---|---|
| RF | 65.3 ± 13.0 | 70.8 ± 9.2 | 11.9 ± 1.2 | 33.6 ± 5.8 | 55.9 ± 9.0 | 82.7 ± 4.1 |
| CNN | **83.5 ± 7.0** | 61.1 ± 16.5 | 45.1 ± 8.1 | 49.5 ± 13.0 | **72.3 ± 1.2** | **82.9 ± 4.3** |
| Renyi | 59.1 ± 11.1 | **83.2 ± 4.2** | **80.5 ± 6.9** | **67.3 ± 3.7** | 62.7 ± 12.0 | 45.3 ± 1.2 |

RF apprennent des caractéristiques plus discriminantes sur la fenêtre sans a priori.

### Paramètres appris et explicabilité

Une des particularités de notre approche est la possibilité d'interpréter les paramètres appris. Nous examinons respectivement les valeurs des pondérations des divergences $\mathbf{w}_c$ dans la figure 6.5 et les paramètres des divergences $\alpha$ dans la figure 6.6. La figure 6.5 nous montre que la plupart des classes valorisent l'utilisation de Rayleigh entre les canaux de co-polarisation et de polarisation croisée. Ceci est cohérent avec les caractéristiques des images SAR, où la distribution d'une image mono-vue (*single look*) en amplitude est modélisée par cette loi. Nous pouvons observer l'importance des canaux polarimétriques et de leurs corrélations (même si cette dernière n'est pas correctement modélisée), par l'utilisation de divergences croisées, soit au sein d'une même image ($\perp, 11$ et $\perp, 22$) ou entre les images de la paire ($\perp, 12$ et $\perp, 21$). Nous pouvons retrouver des contributions dans toutes les divergences pour la classe dissemblable, ce qui rejoint l'explication des performances médiocres pour cette classe, le modèle cherche à utiliser toutes les distributions pour correctement la représenter. Nous retrouvons une forte proportion pour les distributions log-normale pour la classe d'accumulation, ce qui peut être expliqué par la période de l'année où les images ont été acquises, où la fonte du glacier recouvre d'une pellicule d'eau les zones de glace, diminuant la rétrodiffusion, mais également la queue de la distribution.

La figure 6.6 nous renseigne sur la distribution des paramètres de Rényi $\alpha$ appris pour chaque classe. Nous avons fait le choix de nous intéresser uniquement au sous domaine $[0, 1[$, ainsi il est intéressant de voir une forte concentration autour de 0,5, homogène à la distance de Bhattacharyya, utilisée de manière cohérente par toutes les classes. Ceci va dans le sens du choix de

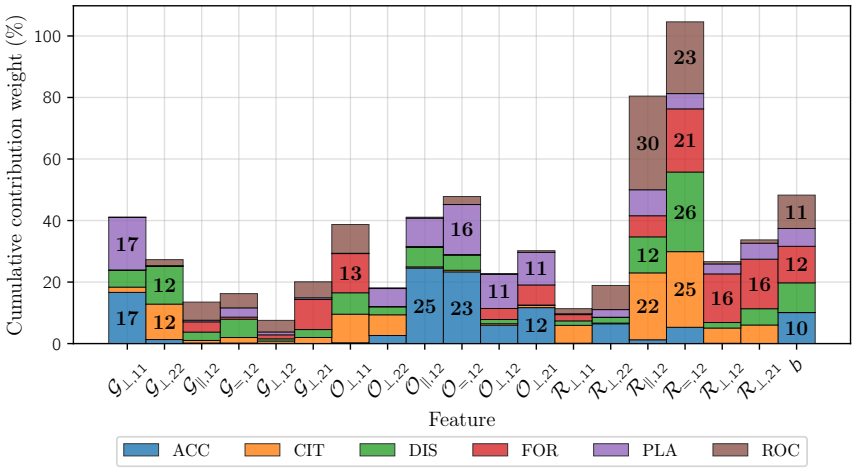



**Figure 6.5 –** Poids moyens associés à chaque divergence pour chaque classe (normalisés par la somme des poids pour chaque classe).

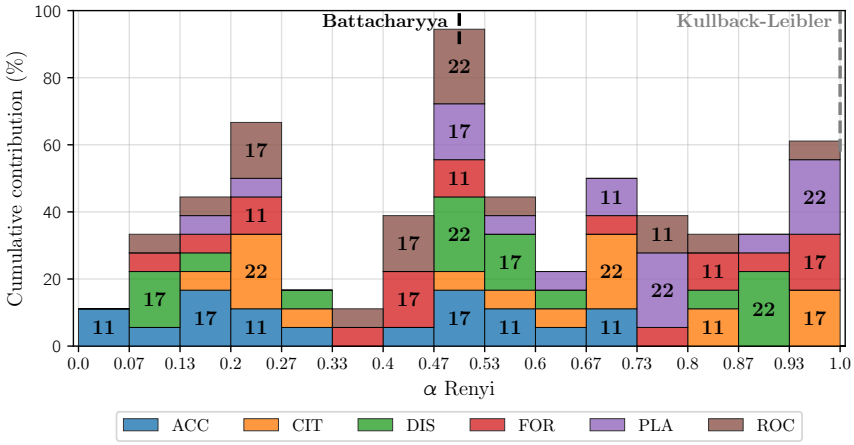



**Figure 6.6 –** Répartition des paramètres de Rényi appris pour chaque classe.

cette distance dans le cadre de classification d'images SAR notamment dans [228]. Nous constatons que la classe *ville* utilise des divergences peu standard (22% de ces paramètres sont autour de 0.25 et 0.7). Finalement nous observons l'utilisation, principalement pour les classes forêt et plaine, d'une distance approchant la divergence de Kullback-Leibler. Il est intéressant de noter que malgré la liberté de choix de la divergence par l'apprentissage, ~ 25% des paramètres appris sont proches soit de la distance de Bhattacharyya, soit de la divergence de Kullback-Leibler.

[228]: Tang et al. (2022), « Large-difference-scale target detection using a revised Bhattacharyya distance in SAR images »

### Sensibilité aux perturbations

Le but de cette approche est double, d'une part proposer une méthode de classification explicative avec peu de paramètres, mais également de proposer une méthode qui soit résiliente aux perturbations. Nous évaluons notre approche sous deux scénarios de perturbation, illustré dans la figure 6.7. Nous mesurons la performance en utilisant la moyenne du score F1 pondéré sur toutes les classes et employons quatre initialisations aléatoires différentes pour la variation des scores.

Dans la figure 6.7.a, nous examinons la sensibilité des méthodes aux variations des données d'entraînement. La perturbation est introduite en modifiant les étiquettes des classes de manière aléatoire, ainsi 30% de bruit signifie que 30% des étiquettes ont été modifiées et ne sont plus correctes. Bien que les scores soient proches entre le CNN et notre approche (plus de 10% de plus que le RF) en l'absence de perturbation, notre approche montre une meilleure résistance face à la perturbation, pour un écart de plus de 10% par rapport aux autres méthodes avec une perturbation de 40%. L'écart diminue et au delà de 75% de perturbation, les performances des trois méthodes convergent.

La figure 6.7.b illustre l'évolution des performances face à une diminution dans le nombre d'échantillons d'entraînement. Nous observons que notre approche et le CNN ont des performances similaires, mais que notre approche est plus résiliente face à la diminution des données d'entraînement. En effet, notre approche maintient une performance stable jusqu'à 20% de données d'entraînement maintenu, tandis que le CNN commence à décliner à environ 35% de données d'entraînement maintenu. Cette résilience est cohérente avec les hypothèses statistiques que nous introduisons, qui atténuent l'impact de l'apprentissage par les données.

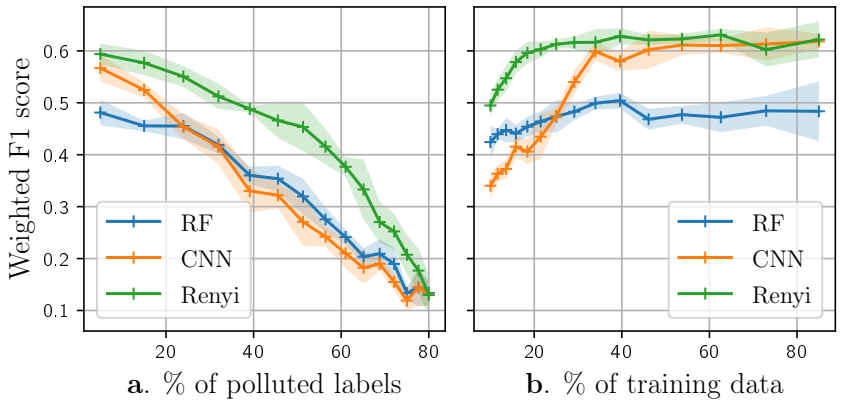



**Figure 6.7 –** Comparaison de la performance (moyenne du score F1 pondéré sur toutes les classes en fonction) de deux perturbations, **a**. Pourcentage de perturbation des labels et **b**. Pourcentage des données d'entraînement maintenu

Cette robustesse observée est la résultante de l'utilisation conjointe d'un a priori statistique et d'un apprentissage automatique. L'a priori statistique permet de limiter l'impact des données d'entraînement sur les paramètres appris, tandis que l'apprentissage automatique permet de capturer les caractéristiques discriminantes des classes.

## 6.3 Conclusion

Dans ce chapitre nous avons proposé l'utilisation de méthodes d'ensemble, par l'agrégation de distances et divergences sur des distributions statistiques a priori, pour la classification d'images SAR. Cette contribution a montré des performances supérieures à des classifieurs classiques (CNN et RF), notamment dans des scénarios de perturbations (bruit sur les labels, diminution du nombre de données, bruit multiplicatif sur les données). Nous avons à travers deux propositions, mis en évidence l'intérêt du couplage entre des a priori statistiques variés et nouveau (Kullback-Leibler entre copules EFGM) avec des combinaisons neuronales basées sur l'apprentissage automatique. La deuxième proposition, a démontré l'intérêt de l'utilisation de divergences paramétriques, notamment la divergence de Renyi, avec un nombre de paramètres réduit et une robustesse accrue face aux perturbations et une meilleure explication de la décision finale. Les perspectives sont nombreuses, et peuvent être réparties en deux axes. D'une part, il est possible d'améliorer la modélisation des distributions, en gardant un compromis entre des distributions plus complexes et des estimations robustes et rapides. Cette amélioration vise à mieux capturer les variabilités entre les classes, prendre en compte la corrélation entre les canaux et explorer les divergences entre distributions différentes. D'autre part, l'apprentissage par la donnée nous montre bien les avantages de cette approche, notamment sa capacité à capturer des caractéristiques spatiales. Il peut être intéressant d'intégrer une partie de cette information dans notre approche, en utilisant des caractéristiques spatiales pour pondérer les divergences ou en utilisant des réseaux de neurones pour apprendre des caractéristiques spatiales.

# Conclusion & perspectives 7

## Conclusion

Les travaux présentés dans ce manuscrit ont porté sur l'analyse de données d'images SAR pour la classification et la caractérisation de surfaces cryosphériques. À travers ce manuscrit, nous avons pu proposer plusieurs solutions et analyses pour répondre aux différentes problématiques abordées dans le chapitre Chapitre 1.

Nous avons restreint le cadre applicatif à deux cas : la classification de glaciers et la caractérisation de la neige humide. Nous avons ainsi proposé dans un premier temps une analyse conjointe des données SAR en bande X et en bande C pour comprendre les informations discriminantes entre les fréquences pour les deux applications. Sur l'ensemble des surfaces glaciaires considérées, la corrélation entre le coefficient de rétrodiffusion en bande X et en bande C est très forte. Cependant l'analyse plus spécifique d'un ensemble de statistiques estimées sur les deux fréquences montre une sensibilité différente dans le cadre de l'étude manteau neigeux : la bande C est plus sensible à la présence d'eau dans le manteau neigeux, là où la bande X est plus sensible à la hauteur de neige. Une différence majeure entre les deux bandes réside dans les acquisitions. La bande X est acquise par un satellite commercial (PAZ), avec une résolution temporelle de 11 jours pour une résolution spatiale de l'ordre de 3m, tandis que la bande C est acquise par un satellite de la mission Sentinel-1, avec une résolution temporelle de 6 jours pour une résolution spatiale de 10m. Cette différence de résolution spatio-temporelle et de disponibilité des données nous a amenés à réfléchir à la construction de jeu de données pour l'apprentissage automatique spécifique aux deux cas d'étude.

Nous avons construit deux jeux de données pour la classification des glaciers et la caractérisation de la neige humide. Pour la classification des glaciers, nous avons construit un jeu de données à partir de données SAR en bande X acquises par le satellite PAZ : CPAZMAL. Ce jeu de données est composé de 4 classes de surfaces glaciaires : zone d'accumulation, zone d'ablation, glacier suspendu et couverture glacio-nivale, et de 4 classes de surfaces non-glaciaires : forêt, ville, plaine et rocher. La labellisation de ces classes a été réalisée manuellement à partir de données optiques sur une zone d'étude centrée sur le massif du Mont-Blanc. Pour la caractérisation de la neige humide, nous avons construit un jeu de données à partir de données SAR en bande C acquises par le satellite Sentinel-1 et couvrant les Alpes françaises du Nord : LSD4WSD. Ce jeu de données comprend deux variantes, la première (V1) exclusivement pour la détection de neige humide, dont la labellisation a été réalisée à partir du modèle de manteau neigeux CROCUS, et la seconde (VX) pour la caractérisation de la neige humide, avec un ensemble de variables contextuelles (topographiques, sorties du modèle CROCUS), permettant l'utilisation de plus nombreuses méthodes d'apprentissage automatique différentes.

Du jeu de données LSD4WSD, nous avons proposé une analyse exhaustive des performances de différentes méthodes d'apprentissage automatique supervisé pour la détection de neige humide. Nous avons montré que de manière globale les performances étaient similaires, avec des spécificités propres à chaque

méthode. Nous avons également cherché à mettre en lumière l'impact de caractéristiques autres que les canaux polarimétriques simples (VV et VH) sur les résultats de classification au regard des méthodes existantes (SWS basée sur l'approche de Nagler). La spatialisation des résultats a montré une bonne correspondance avec les zones identifiées comme étant de la neige par l'imagerie optique (FSC) et SAR (SWS). Il est complexe de valider ces résultats avec une vérité terrain exhaustive, mais nous avons montré que les résultats étaient cohérents avec des données de stations météorologiques disponibles.

Ces résultats ont permis d'introduire les principales contributions méthodologiques de cette thèse, en proposant d'utiliser la variabilité et la spécificité des modèles d'apprentissage automatique pour construire un schéma d'agrégation de classifieurs. Nous avons proposé cette approche à travers le prisme de la logique floue afin de prendre en compte la variabilité des classifieurs et de la qualité de la labellisation. Nous avons montré que cette approche permettait d'améliorer les performances de classification, notamment dans des scénarios de perturbations (bruit sur les labels, diminution du nombre de données, bruit multiplicatif sur les données), et notamment dans le cas de données de faible dimension spatiale. L'application de cette méthode sur la distinction entre les zones d'ablation et d'accumulation a montré une amélioration significative des performances par rapport à l'agrégation classique utilisée dans la littérature.

De l'agrégation de forêts de classifieurs, entièrement basée sur l'apprentissage par la donnée, nous avons progressivement cherché à introduire des informations a priori sur la nature statistique des données. Nous avons ainsi proposé une approche hybride d'agrégation de distances et divergences. Hybride par l'utilisation conjointe d'une part de distances ou divergences, dont les formes sont analytiques et dérivées des paramètres des distributions statistiques sélectionnées, et d'autre part de combinaisons de ces distances ou de leur paramétrisation, estimées par l'apprentissage automatique sur les données. Cette approche a également mis en évidence, de par sa structure, une frugalité en termes de paramètres et de données, donnant lieu à des performances supérieures à des classifieurs de type CNN et RF, notamment en présence d'un faible nombre de données d'apprentissage. Nous avons montré que de par le faible nombre de paramètres et les contraintes de positivité sur les poids, la décision finale était explicable et interprétable.

## Perspectives

Nous pouvons distinguer deux axes majeurs de perspectives. Le premier est applicatif. En effet, comme nous avons pu le soulever dans le chapitre Chapitre 2 et Chapitre 4, la problématique majeure est la validation avec des données terrain. Celles-ci sont la plupart du temps ponctuelles et ne fournissent pas toujours la mesure physique que l'on cherche à identifier. Au delà de l'information de présence de neige humide, qui est une mesure binaire s'attachant à quantifier une évolution progressive, il serait intéressant de se focaliser sur des quantités physiques comme la TEL ou le SWE. Récemment les travaux de [49], ont permis l'élaboration d'un jeu de données de SWE pour l'hémisphère Nord, basé sur des mesures de profondeur de neige. Une reprise de notre dataset en intégrant l'information de SWE modélisée par CROCUS pourrait permettre de mieux valider les approches de caractérisation de la neige humide, et surtout d'introduire une quantification de l'erreur sur l'apprentissage. Néanmoins, sur la zone d'étude des Alpes françaises du

[49]: Fontrodona-Bach et al. (2023), « NH-SWE : Northern Hemisphere Snow Water Equivalent dataset based on in situ snow depth time series »

Nord, nous avons observé que les données SWE n'étaient pas disponibles pour la période de notre jeu de données, et s'arrêtaient en 2019. Il serait donc intéressant de prolonger temporellement le jeu de données LSD4WSD pour intégrer ces informations.

Nous travaillons, au moment où ces lignes sont écrites, à l'amélioration du jeu de données CPAZMAL, pour l'exploitation de l'information temporelle pour la classification des surfaces glaciaires. Les travaux préliminaires sur l'utilisation de barycentres temporels pour la classification des glaciers avec CPAZMAL que nous avons réalisés, ont montré des performances encourageantes tout en permettant de comprendre les résultats de classification et de permettre une interaction entre l'expert et le modèle [367]. De plus, nous avons observé une synergie entre les acquisitions en bande X et en bande C, qui peut permettre une utilisation conjointe des deux fréquences pour améliorer la classification des surfaces glaciaires voire éventuellement de remonter à une information de couverture de neige totale. Cette dernière proposition nécessite de pallier à la différence de résolution spatiale et temporelle entre les acquisitions, mais pourrait être une base intéressante pour une agrégation multi-sources de données SAR voire aussi multi-spectrales.

[367]: Lin-Kwong-Chon et al. (2024), « Supervised Classification For Analysis Of Cryospheric Zones Using SAR Statistical Timeseries »

De ces travaux, le second axe de perspective est méthodologique. Ceci s'illustre sur la problématique de caractérisation de la neige humide de différentes manières. Dans un premier temps de nombreuses pistes pointent vers la stratégie d'apprentissage. Comment réaliser un apprentissage optimal en gérant des jeux de données déséquilibrés et très hétérogènes. Le découpage entre jeu d'entraînement et de test à partir des massifs [340], permet d'éviter un biais dans la validation et le calcul des métriques, mais le déséquilibre très important entre le nombre d'échantillons des classes, nécessite des mesures, qui dans notre cas ont été principalement de réaliser une stratégie de sous-échantillonnage de la classe majoritaire. Nous pouvons étudier l'impact des stratégies d'équilibrage de classe sur les performances de classification, et notamment comparer entre les stratégies qui impactent les données d'apprentissage et celles qui impactent les classifieurs par une modification de leur fonction de coût [368].

[340]: Hänsch (2021), « The Trap Of Random Sampling and How to Avoid It - Alternative Sampling Strategies for a Realistic Estimate of the Generalization Error in Remote Sensing »

[368]: Haixiang et al. (2017), « Learning from class-imbalanced data : Review of methods and applications »

L'hétérogénéité est visible sur la différence de comportement temporel d'une année à l'autre pour une même zone d'étude. Cette variabilité est liée aux modifications météorologiques ou dans certains cas aux transformations du territoire. Nous travaillons sur la prise en compte de cet aspect dans le cadre de la classification des surfaces glaciaires avec la méthode DTW, qui permet de comparer des séries temporelles de manière élastique. La dérivation de méthodologies de classification basées sur la DTW permet de garder une explicabilité des résultats, puisque la distance entre les séries temporelles est interprétable.

Nous pouvons observer également que l'hétérogénéité est spatiale, entre les massifs, mais aussi entre l'échelle locale et globale. Le comportement à une date donnée entre deux massifs peut différer, ainsi, quelle stratégie adopter pour établir un modèle d'apprentissage suffisamment généraliste pour être applicable sur des zones géographiques différentes ? Une belle perspective est l'intégration des méthodes d'adaptation de domaine permettant de transférer les connaissances d'un domaine source à un domaine cible, en prenant en compte les différences de distribution des données entre les deux domaines [369]. Sur la base des résultats proposés dans le chapitre Chapitre 4, plus précisément sur la figure 4.6, nous proposons de tester l'utilisation de l'algorithme KNN en utilisant une métrique autre que la norme euclidienne

[369]: Courty et al. (2016), « Optimal transport for domain adaptation »

entre les caractéristiques, mais plutôt en utilisant la divergence de Kullback-Leibler.

La divergence est calculée sur trois scénarios :

- *Histogramme*, directement sur les *Histogramme*s de la combinaison **H** (cf tableau 4.2)
- *Gaussienne multi* À partir d'une hypothèse gaussienne, en utilisant la forme analytique de la divergence de Kullback-Leibler entre deux gaussiennes multivariées,
- *Gaussienne composée NC*, À partir d'une hypothèse de gaussiennes composées, en utilisant la forme analytique de la divergence de Kullback-Leibler entre deux gaussiennes composées non centrées avec une texture déterministe (proposée par [182]).

[182]: Collas et al. (2023), « Riemannian Optimization for Non-Centered Mixture of Scaled Gaussian Distributions »

La problématique est la classification de la TEL à partir du jeu de données LSD4WSD VX, en divisant le spectre de la TEL en 3 classes : faible (entre 3 et 20 $kg/m^2$), moyenne (entre 20 et 30 $kg/m^2$) et forte (entre 30 et 100 $kg/m^2$). Des premiers tests ont été effectués avec des tailles de fenêtres de 32 pixels en échelle logarithmique (dB). Nous avons étudié les performances pour deux massifs (Vanoise et Haute Tarentaise) pour deux orientations (Nord et Sud), pour des élévations supérieures à 2500m (considéré comme étant la haute montagne). Les résultats préliminaires sont donnés dans le tableau 7.1 avec l'utilisation d'un K-fold (K=5) et 25 voisins.

**Table 7.1 –** Performance moyenne du F1-score pour la classification de la TEL à partir de la divergence de Kullback-Leibler,pour les 3 scénarios proposés en fonction de l'orientation et du massif considéré.

| Orientation | Massif | Méthode | F1 score |
|---|---|---|---|
| Sud | VANOISE | *Histogramme* | 39.46 ± 0.20 |
| Sud | HTE-TARENT | *Histogramme* | 29.03 ± 0.23 |
| Sud | VANOISE | *Gaussienne multi* | 43.07 ± 0.30 |
| Sud | HTE-TARENT | *Gaussienne multi* | 29.03 ± 0.51 |
| Sud | VANOISE | *Gaussienne composée NC* | 38.90 ± 0.12 |
| Sud | HTE-TARENT | *Gaussienne composée NC* | 33.33 ± 0.29 |
| Nord | VANOISE | *Histogramme* | 24.53 ± 0.16 |
| Nord | HTE-TARENT | *Histogramme* | 22.97 ± 0.68 |
| Nord | VANOISE | *Gaussienne multi* | 26.29 ± 0.39 |
| Nord | HTE-TARENT | *Gaussienne multi* | 22.25 ± 1.50 |
| Nord | VANOISE | *Gaussienne composée NC* | 25.89 ± 0.48 |
| Nord | HTE-TARENT | *Gaussienne composée NC* | 20.94 ± 0.77 |

Les résultats montrent que la divergence de Kullback-Leibler est une métrique intéressante pour la classification de la TEL. Cependant nous pouvons remarquer que les performances sont médiocres (un tirage aléatoire donnerait 33%). Nous voyons directement dans ces résultats l'illustration de la variabilité spatiale, d'un point de vue du massif ou de l'orientation. Enfin nous pouvons observer que la méthode *Gaussienne multi* est plus performante que les deux autres méthodes, l'utilisation du paramètre de texture et son estimation est liée aussi au nombre de canaux considérés (4) peuvent expliquer ces résultats. Ainsi nous proposons de continuer à étudier cette approche, en cherchant à

affiner la modélisation des distributions pour chaque classe, voire à intégrer une décomposition de l'information SAR.

Sur les méthodes proposées dans ce manuscrit, nous avons évoqué en conclusion des chapitres les perspectives et améliorations. En perspective du chapitre Chapitre 5, l'agrégation de classifieurs flous peut être intégrée dans des réseaux de neurones, et ainsi être résolue en une étape d'apprentissage. Pour ce faire le principal obstacle reste la construction d'une matrice binaire de permutation, permettant d'ordonner les probabilités et de construire le gradient pour la rétro-propagation. Enfin sur cette approche, la quantification de l'erreur de classification est un point important, et l'introduction des PAC dans l'agrégation floue est une piste intéressante. Des travaux ont déjà porté sur l'intégration de la théorie Probabilities Approximately Correct (PAC) pour maîtriser l'erreur d'apprentissage notamment dans le cadre simple de vote majoritaire [359]. Enfin les travaux portant sur l'apprentissage de divergences proposées dans le chapitre Chapitre 6 peuvent être étendus aux réseaux de neurones, avec une modélisation des distributions et de corrélations entre les filtres, permettant de diminuer le nombre de caractéristiques et d'introduire une modélisation des caractéristiques spatiales.

[359]: Guedj (2019), *A Primer on PAC-Bayesian Learning*

# A

# Annexes

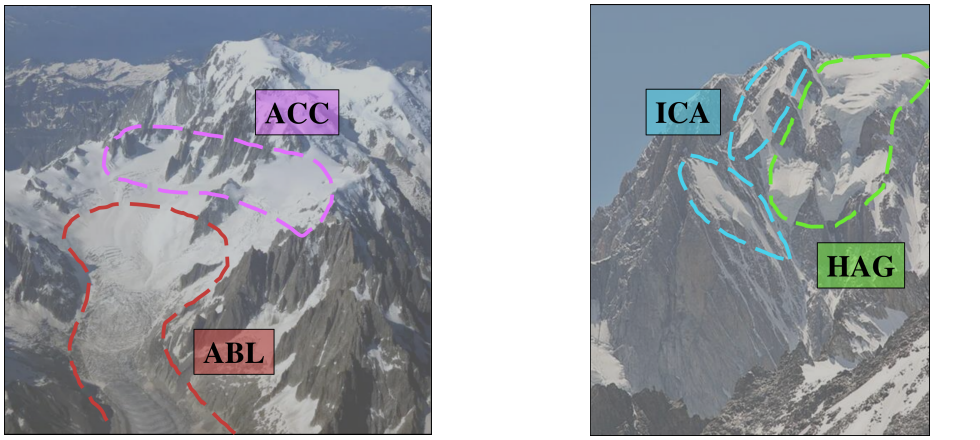
## A.1 Compléments sur les fonctions spéciales

**Table A.1 –** Fonctions spéciales utilisées dans le manuscrit.

| Nom | Fonction | Commentaires |
|---|---|---|
| Gamma | $\Gamma(L) = \int_0^\infty x^{L-1}e^{-x}dx$ | $\mathbb{R}(L) > 0$ |
| Gamma incomplète supérieure | $\Gamma(L,t) = \int_t^\infty x^{L-1}e^{-x}dx$ | $\mathbb{R}(L) > 0, t > 0$ |
| Gamma incomplète inférieure | $\gamma(L,t) = \int_0^t x^{L-1}e^{-x}dx$ | $\mathbb{R}(L) > 0, t > 0$ |
| Di-gamma | $\psi(L) = \Gamma'(L)/\Gamma(L)$ | $L \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}, \psi(1) = -\varepsilon$ |
| Poly-gamma | $\psi^{(n)}(L) = \left(\frac{d}{dL}\right)^n \psi(L)$ | $L \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}, n \in \mathbb{N}, \psi^{(0)}(L) = \psi(L)$ |
| Poly-logarithme | $\mathrm{Li}_s(z) = \sum_{k=1}^\infty \frac{z^k}{k^s}$ | $\forall s, z \in \mathbb{C}, \lvert z\rvert < 1$, $\mathrm{Li}_2$ est appelée dilogarithme |

La constante $\varepsilon$ est la constante d'Euler-Mascheroni.

## A.2 Détails des régions d'étude pour la classification des glaciers




**(a)** Illustration des zones d'accumulation (ACC) et d'ablation (ABL) (*Tofig, Rashidov*. Face Est du Mont Blanc).

**(b)** Illustration des zones de couverture glacio nivale (ICA) et de glacier suspendu (HAG) (*Simo, Räsänen*. Face Sud-Est du Mont Blanc).

**Figure A.1 –** Photographies des zones d'étude pour la classification des glaciers sur le massif du Mont-Blanc.

**Table A.2 –** Détails des polygones de vérité terrain du jeu de données CPAZMAL pour les classes de couverture du sol dans le massif du Mont-Blanc utilisés avec l'orbite descendante du satellite PAZ.

| **Class** | **Group** | **Elevation** (m) | **Orientation** (°) | **Slope** (°) | **Area** (ha) | **Latitude** | **Longitude** |
|---|---|---|---|---|---|---|---|
| CIT | 001 | 1265 | 211 | 4 | 5.43 | 45.98 | 6.93 |
| CIT | 002 | 1080 | 283 | 3 | 13.72 | 45.94 | 6.90 |
| CIT | 003 | 1041 | 283 | 2 | 6.35 | 45.92 | 6.87 |
| CIT | 004 | 1042 | 267 | 3 | 5.37 | 45.92 | 6.87 |
| CIT | 005 | 1063 | 261 | 2 | 16.50 | 45.94 | 6.89 |
| CIT | 006 | 1057 | 129 | 9 | 19.95 | 45.93 | 6.87 |
| CIT | 007 | 1047 | 195 | 2 | 5.14 | 45.93 | 6.88 |
| CIT | 008 | 1064 | 146 | 9 | 6.97 | 45.94 | 6.88 |
| CIT | 009 | 1035 | 295 | 5 | 12.11 | 45.92 | 6.87 |
| CIT | 010 | 1044 | 112 | 4 | 1.50 | 45.92 | 6.87 |
| CIT | 011 | 1044 | 119 | 5 | 7.85 | 45.92 | 6.87 |
| FOR | 001 | 1709 | 219 | 28 | 11.78 | 45.79 | 6.99 |
| FOR | 002 | 1239 | 331 | 16 | 5.60 | 45.90 | 6.86 |
| FOR | 003 | 1103 | 253 | 39 | 9.22 | 45.76 | 7.00 |
| FOR | 004 | 1897 | 82 | 27 | 24.08 | 45.74 | 7.02 |
| FOR | 005 | 1709 | 323 | 35 | 31.47 | 45.97 | 6.93 |
| FOR | 006 | 1628 | 315 | 40 | 12.62 | 45.82 | 6.98 |
| PLA | 001 | 1930 | 321 | 15 | 7.61 | 45.97 | 6.94 |
| PLA | 002 | 1525 | 239 | 11 | 8.68 | 46.01 | 6.95 |
| PLA | 003 | 1945 | 179 | 21 | 12.15 | 46.04 | 6.91 |
| PLA | 004 | 1440 | 224 | 7 | 6.90 | 46.00 | 6.95 |
| PLA | 005 | 2313 | 301 | 18 | 41.53 | 45.84 | 7.03 |
| PLA | 006 | 2293 | 178 | 26 | 26.51 | 45.71 | 6.88 |
| PLA | 007 | 2202 | 275 | 15 | 7.19 | 46.02 | 6.97 |
| ROC | 001 | 2100 | 48 | 25 | 6.65 | 46.00 | 6.91 |
| ROC | 002 | 2496 | 212 | 41 | 4.74 | 45.98 | 6.88 |
| ROC | 003 | 2362 | 127 | 16 | 6.64 | 45.99 | 6.90 |
| ROC | 004 | 2695 | 157 | 28 | 13.14 | 45.91 | 6.96 |
| ROC | 005 | 2261 | 36 | 12 | 3.05 | 45.90 | 6.88 |
| ROC | 006 | 3701 | 176 | 62 | 2.84 | 45.87 | 6.99 |
| ROC | 007 | 2287 | 318 | 29 | 2.35 | 45.89 | 6.87 |

La figure A.1a illustre les zones d'accumulation et d'ablation (ACC/ABL) sur le massif du Mont-Blanc. La figure A.1b illustre les zones de couverture glacio-nivale et de glacier suspendu (ICA/HAG) sur le massif du Mont-Blanc.

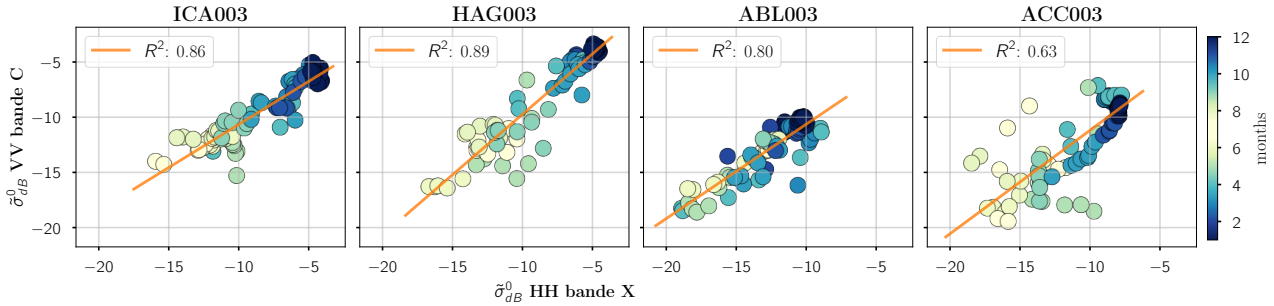
Nous présentons dans les tableaux A.2 et A.3 les détails des polygones de vérité terrain utilisés pour la classification des glaciers sur le massif du Mont-Blanc avec les données du satellite PAZ en orbite descendante. Les classes sont les suivantes : ABL (Ablation), ACC (Accumulation), CIT (Ville), FOR (Forêt), HAG (Hanging Glacier), ICA (Ice Aperon), PLA (Plaine), ROC (Roche). Les groupes sont numérotés de 001 à maximum 007 pour chaque classe. Les caractéristiques topographiques sont l'élévation, l'orientation, la pente, la superficie, la latitude et la longitude. Les superficies sont exprimées en hectares. Les orientations sont exprimées en degrés par rapport au nord. Les pentes sont exprimées en degrés. Les latitudes et longitudes sont exprimées en degrés décimaux.

**Table A.3 –** Détails des polygones de vérité terrain du jeu de données CPAZMAL pour la classification des glaciers dans le massif du Mont-Blanc utilisés avec l'orbite descendante du satellite PAZ.

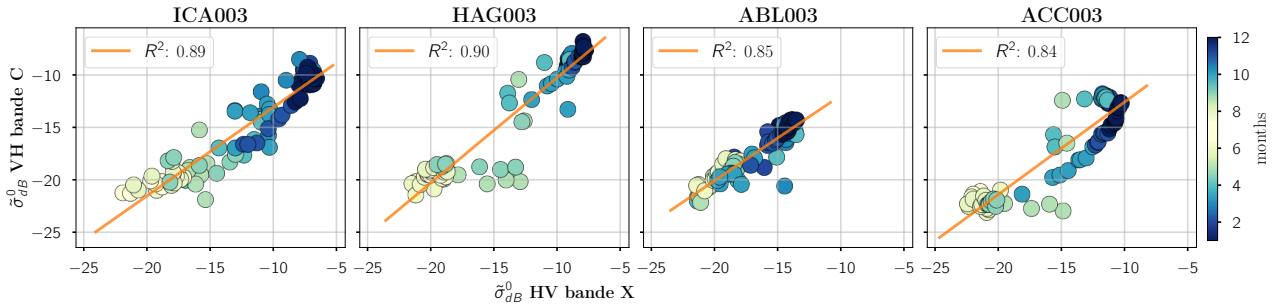
| **Class** | **Group** | **Elevation** (m) | **Orientation** (°) | **Slope** (°) | **Area** (ha) | **Latitude** | **Longitude** |
|---|---|---|---|---|---|---|---|
| ABL | 001 | 3038 | 89 | 17 | 7.74 | 45.96 | 7.03 |
| ABL | 002 | 3267 | 227 | 16 | 8.75 | 45.95 | 7.02 |
| ABL | 003 | 2483 | 292 | 26 | 32.47 | 45.99 | 6.98 |
| ABL | 004 | 3302 | 213 | 21 | 8.00 | 45.82 | 6.85 |
| ABL | 005 | 2635 | 194 | 8 | 14.73 | 45.91 | 6.97 |
| ABL | 006 | 2346 | 313 | 11 | 10.05 | 45.89 | 6.88 |
| ABL | 007 | 2494 | 315 | 22 | 16.93 | 45.88 | 6.94 |
| ACC | 001 | 3731 | 223 | 44 | 1.55 | 45.87 | 6.98 |
| ACC | 002 | 3475 | 217 | 17 | 4.28 | 45.83 | 6.84 |
| ACC | 003 | 3110 | 220 | 12 | 19.93 | 45.96 | 7.04 |
| ACC | 004 | 3128 | 295 | 13 | 33.21 | 45.98 | 7.00 |
| ACC | 005 | 3360 | 248 | 14 | 11.44 | 45.95 | 7.03 |
| ACC | 006 | 2549 | 268 | 33 | 5.63 | 45.89 | 6.89 |
| HAG | 001 | 3239 | 220 | 35 | 2.76 | 45.95 | 7.01 |
| HAG | 002 | 3474 | 183 | 44 | 0.39 | 45.88 | 7.01 |
| HAG | 003 | 2927 | 260 | 28 | 3.05 | 45.87 | 6.95 |
| HAG | 004 | 3534 | 220 | 47 | 1.60 | 45.88 | 6.89 |
| HAG | 005 | 3423 | 320 | 43 | 0.37 | 45.87 | 6.88 |
| ICA | 001 | 3611 | 193 | 24 | 0.99 | 45.97 | 7.00 |
| ICA | 002 | 2839 | 346 | 51 | 1.35 | 45.91 | 6.91 |
| ICA | 003 | 3468 | 192 | 43 | 1.98 | 45.93 | 6.99 |
| ICA | 004 | 3927 | 224 | 36 | 1.04 | 45.87 | 6.98 |
| ICA | 005 | 3484 | 190 | 43 | 0.60 | 45.84 | 6.91 |
| ICA | 006 | 3943 | 214 | 45 | 1.57 | 45.83 | 6.86 |

### A.2.1 Complément sur la comparaison des données SAR en bande C et en bande X pour les glaciers

Nous proposons dans les figures A.2 et A.3 des comparaisons des données SAR en bande C et en bande X pour 8 zones supplémentaires de surface glaciaire. Les sites d'étude "XXX003" et "XXX004" font référence au nom de la classe et au groupe considéré tel que présenté dans le tableau A.3. Les données SAR en bande C (VV et VH) sont acquises par les satellites Sentinel-1 tandis que les données SAR en bande X (HH et HV) sont acquises par le satellite PAZ. Nous retrouvons une forte corrélation entre les deux bandes avec un coefficient de détermination $R^2$ supérieur à 0.8. Les deux exceptions notables sont les sites "ICA004", et "ACC003" avec un étalement temporel pour la zone d'accumulation.

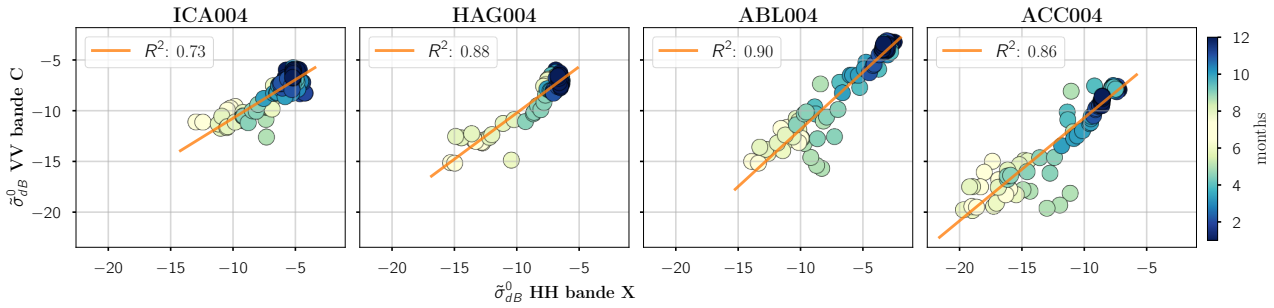



**(a)** Co-polarisation (HH / VV)

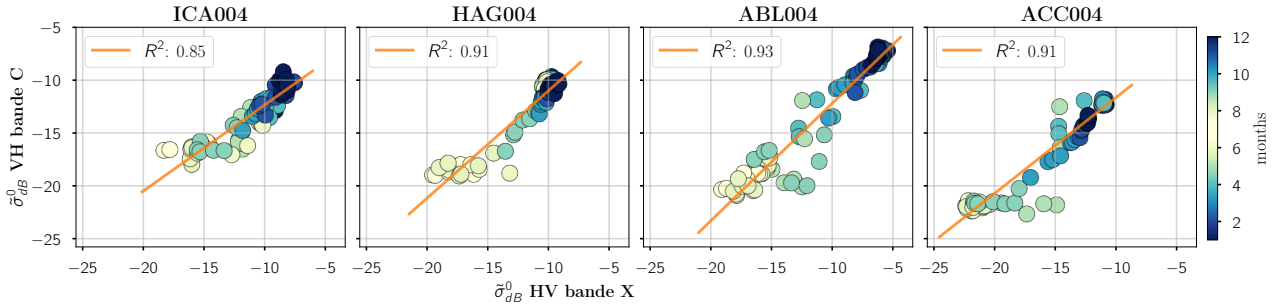



**(b)** Polarisation croisée (HV / VH)

**Figure A.2 –** Comparaison des données SAR en bande C et en bande X pour les 4 classes de glaciers sur les sites d'étude "XXX003"




**(a)** Co-polarisation (HH / VV)




**(b)** Polarisation croisée (HV / VH)

**Figure A.3 –** Comparaison des données SAR en bande C et en bande X pour les 4 classes de glaciers sur les sites d'étude "XXX004"

# Table des figures


2.1 Schéma des différentes études de la cryosphère adapté de [16] . . . 7
2.2 Illustration des 4 types de surfaces glaciaires étudiées dans ce travail. . . . 8
2.3 Schéma de l'évolution générale de deux indicateurs du manteau neigeux (SWE et TEL) au cours d'une saison hivernale. Adapté de [41]. . . . 10
2.5 Variation temporelle de teneur en eau liquide (TEL) pour le modèle CROCUS simulé quotidiennement à 18h pour deux massifs alpins (Belledonne et Grandes Rousses) à différentes altitudes (1800 et 2700 mètres) sur la période d'août 2020 à août 2021. . . . 14
2.12 Bandes de fréquences des satellites radar . . . 19
2.13 Illustration des différents types de diffusion du signal radar. Extrait de [95] . . . 20
2.14 Illustration de la pénétration des ondes électromagnétiques dans la canopée. Adaptée de [95] . 21
2.16 Illustration des différentes distorsions géométriques du signal radar. Adapté de [103] . . . 23
2.17 Image SAR (TerraSAR-X) de zones agricoles homogènes polluées par du bruit de speckle (a) et filtrée par un filtre multi-temporel (20 images) (b). Extrait de [91]. . . . 24
2.18 Schéma illustrant les différentes hypothèses de modèle de densité de probabilité d'un pixel SAR en intensité. Adapté de [112] . . . 26
2.22 Comparaison des distributions statistiques des 8 classes en bande X, pour 3 dates en 2020 (4 août, 8 janvier et 5 avril) pour les deux polarisations HH et HV. . . . 35
2.23 Comparaison temporelle et polarimétrique des 8 différentes classes en bande X sur l'année 2020 (valeurs moyennes et écarts-types) en fonction de la polarisation et du ROI. . . . 36
2.24 Comparaison des distributions statistiques des 8 classes en bande C, pour 3 dates en 2020 (2 août, 11 janvier et 4 avril) pour les deux polarisations VV et VH. . . . 38
2.25 Comparaison temporelle et polarimétrique des 8 différentes classes en bande C sur l'année 2020 (valeurs moyennes et écarts-types) en fonction de la polarisation et du ROI. . . . 39
2.26 Comparaison des données SAR en bande C et en bande X pour les 4 classes de glaciers sur les sites d'étude "XXX001" . . . 40
2.27 Comparaison des données SAR en bande C et en bande X pour les 4 classes de glaciers sur les sites d'étude "XXX002" . . . 41
2.28 Étude de la corrélation entre la hauteur de neige simulée par le modèle CROCUS et celle mesurée à l'emplacement 'PLA004'. . . . 42
2.29 Comparaison des corrélations de type Spearman entre les statistiques des images SAR (cf description dans le tableau 2.7) et les caractéristiques du manteau neigeux pour la période hivernale (première ligne) et pour la période printanière (seconde ligne), pour les deux bandes X (première colonne) et C (seconde colonne). . . . 43

3.1 Illustration des copules elliptique et archimédienne pour différentes paramétrisations. Les copules archimédiennes sont des copules asymétriques (sauf la copule de Frank) avec la capacité de modéliser des queues lourdes (hautes ou basses) tandis que les copules elliptiques sont symétriques et la modélisation des queues est réalisable par une copule de Student-t induisant une queue symétrique (haute et basse). . . . 51
3.8 Visualisation de l'empreinte au sol de l'acquisition du satellite Sentinel-1 (tuile 31TGL) en orbite descendante et les massifs alpins français utilisés. Le massif de validation est en vert, le rectangle bleu foncé représente l'emprise des cartes de résultats. Les zones d'eau sont en bleu (lacs et glaciers principaux référencés par l'IGN). . . . 68
3.9 Exemple d'une image SAR, centrée sur le lac d'Annecy (cf figure 3.10 ), avec les canaux VV (**A**), VH (**B**) et le rapport VV/VH (**C**) en dB. . . . 68
3.10 Situation géographique de la zone d'illustration de l'image SAR, centrée sur le lac d'Annecy. . . 69
3.12 Illustration des données topographiques sur la zone du lac d'Annecy. **A.** Modèle Numérique de Terrain (MNT) de l'IGN. **B.** Pente dérivée du MNT. **C.** Orientation dérivée du MNT. . . . 69

3.13 Illustration de la caractéristique de Nagler (en dB) en passe ascendante pour le 31 mars 2021, avec une acquisition de référence du 9 août 2020. **A.** VV/VV$_{ref}$, **B.** VH/VH$_{ref}$, **C.** VVVH$_{ref}$/VHVV$_{ref}$. En orange, les zones de distorsions géométriques irréversibles. . . . . 70
3.14 Illustration du produit FSC pour le 30 mars 2021, sur la zone du lac d'Annecy (cf figure 3.10 et figure 3.13). . . . . 70
3.15 Visualisation de la zone de test (en vert), de validation (rectangle bleu foncé) et des stations de mesures comparatives (étoiles) sur le massif des Grandes Rousses. En bas, span ($\sqrt{VV^2 + VH^2}$) de la zone de validation avec en orange les zones de distorsions géométriques. . . . . 70
3.16 Détail du processus de sélection des échantillons (ou pavage) dans le massif des Bauges, en marron, les masques de distorsion géométrique irréversibles. . . . . 71
3.17 Distributions topographiques des échantillons de test et d'entraînement selon la discrétisation donnée par CROCUS, pour la pente, l'orientation et l'élévation . . . . . 72
3.18 Différence temporelle de la teneur en eau liquide (TEL) pour le modèle CROCUS entre les données simulées à 6h et 18h pour deux massifs alpins (Belledonne et Grandes Rousses) à différentes altitudes (1800 et 2700 mètres) sur la période d'août 2020 à août 2021. . . . . 73
3.19 Corrélation entre la règle de détection de neige humide basée sur la température minimale et la règle basée sur le seuillage de la TEL pour deux périodes dans l'année 2020-2021. . . . . 74
3.20 Distribution du nombre d'échantillons entre août 2020 et 2021, selon les 6 massifs d'entraînement considérés pour les deux classes *humide* (wet) et *non-humide* (non wet). . . . . 75
3.21 Comparaison entre le produit FSC et SWS pour respectivement le 13 et 11 juin 2021, regroupant les massifs de Belledonne, Chartreuse et Bauges . . . . . 76
3.22 Visualisation de l'acquisition du satellite PAZ en orbite descendante sur le massif du Mont-Blanc et des polygones de vérité terrain. . . . . 78
3.23 Répartition temporelle des acquisitions de PAZ . . . . . 78

4.1 Architecture CNN proposée pour la classification de la neige humide. . . . . 82
4.2 K-Fold stratifié équilibré . . . . . 84
4.3 Évaluation du score F1 pour les 4 meilleurs algorithmes obtenus dans le tableau 4.4 en fonction des combinaisons de canaux considérées . . . . . 86
4.4 Courbe ROC pour 4 combinaisons de canaux d'entrée pour le classifieur KNN et leur métrique BAROC et FCROC associée à 5% (à gauche). Courbe ROC pour le classifieur de type seuillage simple sur la moyenne de chaque échantillon pour les canaux VV et VH (seuillage Nagler), avec les métriques BAROC et FCROC (à droite). . . . . 87
4.5 Score F1 pour le classifieur KNN en fonction de la combinaison de statistiques décrites dans le tableau 4.3 sur les combinaisons de canaux **H** . . . . . 88
4.6 Comparaison de l'histogramme 2D pour la moyenne et l'écart type de chaque échantillon, pour les deux classes *humide* et *non humide*, et entre les différents canaux pour ($VV$,$VH$ $R_VV$ et $R_VH$). $R_VV$ et $R_VH$ représentent le ratio entre l'acquisition actuelle et l'image de référence décrite dans 3.2.2 respectivement en polarisation $VV$ et $VH$. . . . . 89
4.7 Cartes de neige humide pour le 31 mars 2021, obtenues avec le RF (à droite) et CNN (à gauche) sur la combinaison **I** et en comparant avec les produits Copernicus FSC et SWS et la règle de labellisation à partir du modèle CROCUS, pour une portion du massif des Grandes Rousses . . 91
4.8 Cartes des résultats de classification de la neige pour deux dates en comparaison avec les produits Copernicus FSC et SWS . . . . . 93
4.9 Détails de la carte et vue agrandie des résultats de classification de la neige pour les deux meilleures combinaisons **I** et **J** comparant les produits Copernicus FSC et SWS dans l'Alpe d'Huez, France pendant la période de fonte le 31 mars 2021 . . . . . 94
4.10 Diagrammes de la distribution des pixels humides (FRCOC, BAROC, SWS) et de la couverture neigeuse (FSC) en fonction de l'altitude et de l'orientation pour une pente inférieure à 45°, pour deux dates . . . . . 95
4.11 Comparaison des mesures de la station Lac Noir avec la prédiction du KNN. . . . . 96

4.12 Comparaison des mesures pour les combinaisons de canaux **H**,pour trois stations de mesures, Alpe d'Huez, Lac Noir et Galibier. . . . . 98

5.1 Vue bivariée des deux T-normes sélectionnées en fonction de la valeur de $\alpha$. La carte de couleur représente la valeur de la T-norme. La ligne rouge représente l'isocontour de la fonction. . . . 104
5.2 Position des ROI pour les 2 classes *ABL* et *ACC* sur le massif du Mont Blanc. À gauche l'image SAR PAZ, au centre illustration de la topographie et répartition des zones. La présence d'eau (y compris les glaciers) est donnée en bleu. . . . . 108
5.3 Schéma du processus d'apprentissage en deux étapes. Une troisième étape est illustrée pour l'évaluation des performances des différentes méthodes d'agrégation . . . . . 110
5.4 Évaluation du score F1 (pondéré) pour les 2 classes *ABL* et *ACC*, pour les 3 classifieurs flous pondérés (à gauche et bleu), les 3 classifieurs flous de puissance (à droite et orange) et la régression logistique (noir) en fonction de la variance d'un bruit multiplicatif. . . . . 112
5.5 Évaluation du score F1 (pondéré) pour les 2 classes *ABL* et *ACC*, pour les 3 classifieurs flous pondérés (à gauche et en bleu), les 3 classifieurs flous de puissance (à droite et en orange) et la régression logistique (en noir) en fonction du pourcentage de bruit d'étiquetage. . . . . 113
5.6 Image optique (en haut) et image radar (en bas) du glacier d'Argentière. L'image SAR considérée correspond à la moyenne des canaux HH et HV en dB pour la date du 04/08/2020. . . . . 114
5.7 Ensemble de données en fenêtre3×3. Spatialisation des résultats (en géométrie radar) pour le glacier d'Argentière pour les deux meilleures méthodes de pré-agrégation, ChqW$T^3$ et ChqP$T^6$, en comparaison avec l'agrégation de régression logistique. . . . . 115
5.8 Ensemble de données en fenêtre 9×9. Spatialisation des résultats (en géométrie radar) pour le glacier d'Argentière pour les deux meilleures méthodes de pré-agrégation, ChqW$T^3$ et ChqP$T^6$, en comparaison avec l'agrégation de régression logistique. . . . . 116

6.1 Schéma de l'architecture de l'apprentissage d'ensembles de divergence. . . . . 124
6.2 Valeur des poids et du biais pour les 3 classes considérées (2 thématiques et la classe divergence) en fonction des distances considérées. Les identifiants sont donnés dans le tableau 6.3 . . . . . 127
6.3 Illustration de l'ensemble des divergences calculées sur des distributions Gamma estimées à partir d'une paire d'images SAR bivariées ($J = 2$). . . . . 131
6.4 Résumé de l'approche de classification proposée. . . . . 131
6.5 Poids moyens associés à chaque divergence pour chaque classe (normalisés par la somme des poids pour chaque classe). . . . . 135
6.6 Répartition des paramètres de Rényi appris pour chaque classe. . . . . 136
6.7 Comparaison de la performance (moyenne du score F1 pondéré sur toutes les classes en fonction) de deux perturbations, **a**. Pourcentage de perturbation des labels et **b**. Pourcentage des données d'entraînement maintenu . . . . . 137

A.1 Photographies des zones d'étude pour la classification des glaciers sur le massif du Mont-Blanc. 145
A.2 Comparaison des données SAR en bande C et en bande X pour les 4 classes de glaciers sur les sites d'étude "XXX003" . . . . . 148
A.3 Comparaison des données SAR en bande C et en bande X pour les 4 classes de glaciers sur les sites d'étude "XXX004" . . . . . 148

# Liste des tableaux


2.1 Classification de la neige en fonction de sa teneur en eau liquide (TEL). Adapté de [35] . . . . . 10
2.2 Comparaison des catégories spatio-temporelles de quatre accès à l'information TEL. Le signe "+" indique un niveau faible, "++" un niveau moyen et "+++" un niveau élevé. Adapté de [72] . . . 15
2.3 Tableau des distributions utilisées dans ce travail et leur fonction de densité de probabilité (pdf), pour un support $x > 0$. . . . . 27
2.4 Tableau des principaux satellites radar de type SAR et leurs applications pour la cryosphère adapté de [16] . . . . . 29
2.5 Caractéristique des traitements SAR dans le cadre de deux approches, adapté de [36] pour l'étude de la neige . . . . . 31
2.6 Caractéristique du signal SAR dans le cadre d'une différenciation neige sèche neige humide adapté de [36] . . . . . 32
2.7 Tableau des statistiques extraites des images SAR en bande X et en bande C pour l'étude de la neige. Les statistiques $\gamma_1$ et $\gamma_2$ correspondent respectivement au coefficient d'asymétrie et au kurtosis définis dans la section 6.2. Les indices $^{\text{ref}}$ correspondent aux statistiques calculées sur les ratio avec une image de référence en été. Les indices $\parallel$ et $\perp$ correspondent respectivement à la co-polarisation et à la polarisation croisée. . . . . 43

3.1 Divergences de Rényi pour différentes valeurs de $\alpha$, adapté de [224] . . . . . 54
3.2 Matrice de confusion . . . . . 63

4.1 Paramétrisation des algorithmes utilisés . . . . . 82
4.2 Description des combinaisons de canaux testées . . . . . 83
4.3 Description des combinaisons statistiques testées, où $\mu, \sigma$ sont la moyenne et l'écart-type et $\gamma_1, \gamma_2$ sont donnés dans Eq.(2.3) et (2.4) . . . . . 84
4.4 Résultats pour 3 métriques pour les 7 classifieurs considérés pour le problème de détection de neige en utilisant K-fold, avec $K = 34$ . . . . . 85
4.5 Valeur de seuil pour les métriques BAROC et FCROC pour l'apprentissage sur la moyenne de l'échantillon pour VV et VH . . . . . 87

5.1 Détails des T-normes paramétrées sélectionnées et de leurs domaines de définition . . . . . 103
5.2 Détails de la forme reformulée pour les mesures floues de puissance et de moyenne pondérée . 105
5.3 Dérivées des T-normes par rapport aux contraintes $\alpha$ et $g$. $A, B$ et $Z$ sont des variables intermédiaires données dans le tableau 5.1. . . . . 106
5.4 Nombre de patchs dans le jeu de données en fonction de la taille de la fenêtre. . . . . 107
5.5 Détails des classifieurs et de leurs paramètres utilisés dans le processus d'apprentissage. . . . . 109
5.6 F1-score pondéré pour les méthodes d'agrégation floue avec des mesures floues moyenne pondérée et de puissance, et la régression logistique. . . . . 111
5.7 F1-score pondéré pour les méthodes d'agrégation floue robustes avec des mesures floues moyenne pondérée et de puissance pour les deux T-normes et la régression logistique. . . . . 111

6.1 Récapitulatif des modèles utilisés en fonction des données polarimétriques SAR, en bleu les distances basées sur les divergences de Jeffreys et en noir celle sur les normes $\ell_2$ . . . . . 125
6.2 Score d'apprentissage des différentes combinaisons . . . . . 126
6.3 Tableau des correspondances entre les indices des distances et divergences utilisés dans la figure 6.2 et les mesures statistiques utilisées. La première lettre représentant la nature de la distance ($\mathcal{K}$ ou $\ell_2$), le deuxième bloc le modèle paramétrique (RGGA, Weibull, copule gaussienne ou EFGM) et le troisième bloc le canal considéré (HH, HV ou R= HH/HV). . . . . 126

6.4 Divergences de Rényi pour les distributions gamma, log-normale et Rayleigh. . . . . 130
6.5 Dérivée de Rényi, avec $\lambda_{ij} = (e^{\alpha_l} L_i + L_j)/(1 + e^{\alpha_l})$, $\beta_{ij} = e^{\alpha_l} L_i \mu_j + L_j \mu_i$, $\Sigma_{ij} = \sigma_j^2 + \sigma_i^2$ et $\gamma_{ij} = e^{\alpha_l} \mu_j^2$. 132
6.6 Nombre de paramètres et taille des entrées utilisées . . . . . 134
6.7 Pourcentage de bonne classification avec un K-Fold stratifié avec K=5 pour chaque classe. . . . . 135

7.1 Performance moyenne du F1-score pour la classification de la TEL à partir de la divergence de Kullback-Leibler,pour les 3 scénarios proposés en fonction de l'orientation et du massif considéré. 142

A.1 Fonctions spéciales utilisées dans le manuscrit. . . . . 145
A.2 Détails des polygones de vérité terrain du jeu de données CPAZMAL pour les classes de couverture du sol dans le massif du Mont-Blanc utilisés avec l'orbite descendante du satellite PAZ. . . . . 146
A.3 Détails des polygones de vérité terrain du jeu de données CPAZMAL pour la classification des glaciers dans le massif du Mont-Blanc utilisés avec l'orbite descendante du satellite PAZ. . . . . 147

# Bibliographie


[1] IPCC. *The Ocean and Cryosphere in a Changing Climate : Special Report of the Intergovernmental Panel on Climate Change*. en. 1re éd. Cambridge University Press, avr. 2022. (Visité le 22/04/2022) (cf. p. 1).

[2] IPCC. *Climate Change 2022 : Impacts, Adaptation and Vulnerability*. en. Fév. 2022. URL : https://www.ipcc.ch/report/ar6/wg2/ (visité le 22/04/2022) (cf. p. 1).

[3] Josef ASCHBACHER. « ESA's earth observation strategy and Copernicus ». In : *Satellite earth observations and their impact on society and policy* (2017), p. 81-86 (cf. p. 1).

[4] Herbert J KRAMER et al. *Observation of the Earth and its Environment : Survey of Missions and Sensors*. T. 1982. Springer, 2002 (cf. p. 1).

[5] Susan L. USTIN et Elizabeth M. MIDDLETON. « Current and near-term advances in Earth observation for ecological applications ». In : *Ecological Processes* 10 (2021) (cf. p. 1).

[6] Joseph W. GOODMAN. « Some fundamental properties of speckle ». In : *Journal of the Optical Society of America* 66 (1976), p. 1145-1150 (cf. p. 2, 25).

[7] Alicia PASSAH et al. « SAR Image Classification : A Comprehensive Study and Analysis ». en. In : *IEEE Access* 10 (2022), p. 20385-20399. DOI : 10.1109/access.2022.3151089. (Visité le 12/04/2022) (cf. p. 2, 48).

[8] Nikola BESIC et al. « Stochastic Approach in Wet Snow Detection Using Multitemporal SAR Data ». en. In : *IEEE Geosci. Remote Sensing Lett.* 12.2 (fév. 2015). Number : 2, p. 244-248. DOI : 10.1109/lgrs.2014.2334355. (Visité le 13/09/2021) (cf. p. 2, 33, 34, 49).

[9] Ambroise GUIOT et al. « Insights into Segmentation Methods Applied to Remote Sensing SAR Images for Wet Snow Detection ». In : *Geosciences* 13.7 (2023). DOI : 10.3390/geosciences13070193 (cf. p. 2, 33).

[10] C. BOUCHAYER et al. « A Machine Learning Framework to Automate the Classification of Surge-Type Glaciers in Svalbard ». In : *Journal of Geophysical Research : Earth Surface* 127.7 (2022). e2022JF006597 2022JF006597, e2022JF006597. DOI : https://doi.org/10.1029/2022JF006597 (cf. p. 2, 65).

[11] Liangpei ZHANG, Lefei ZHANG et Bo DU. « Deep Learning for Remote Sensing Data : A Technical Tutorial on the State of the Art ». en. In : *IEEE Geoscience and Remote Sensing Magazine* 4.2 (juin 2016), p. 22-40. DOI : 10.1109/mgrs.2016.2540798. (Visité le 11/04/2022) (cf. p. 2, 57).

[12] Hemani PARIKH, Samir PATEL et Vibha PATEL. « Classification of SAR and PolSAR images using deep learning : a review ». en. In : *International Journal of Image and Data Fusion* 11.1 (2 jan. 2020). Number : 1, p. 1-32. DOI : 10.1080/19479832.2019.1655489. (Visité le 14/10/2021) (cf. p. 2, 57, 82, 133).

[13] Bing ZHANG et al. « Progress and Challenges in Intelligent Remote Sensing Satellite Systems ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022), p. 1814-1822. DOI : 10.1109/jstars.2022.3148139 (cf. p. 2).

[14] Ronny HANSCH et Olaf HELLWICH. « The Truth About Ground Truth : Label Noise in Human-Generated Reference Data ». In : *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. Ieee, juill. 2019, p. 5594-5597. DOI : 10.1109/igarss.2019.8898003. (Visité le 29/01/2024) (cf. p. 2, 48, 62, 63, 71).

[15] R. HÄNSCH, A. LEY et O. HELLWICH. « Correct and still wrong : The relationship between sampling strategies and the estimation of the generalization error ». In : *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Ieee, juill. 2017, p. 3672-3675. DOI : 10.1109/igarss.2017.8127795. (Visité le 20/11/2022) (cf. p. 3, 127).

[16] Liam S TAYLOR et al. « Remote sensing of the mountain cryosphere : Current capabilities and future opportunities for research ». en. In : *Progress in Physical Geography : Earth and Environment* 45.6 (30 juin 2021), p. 931-964. DOI : 10.1177/03091333211023690. (Visité le 09/11/2021) (cf. p. 7, 16, 29, 30, 34, 35).

[17] Peter H. GLEICK. « Water in crisis : a guide to the world's fresh water resources ». In : 1993 (cf. p. 7).

[18] Martin Beniston. « Climatic Change in Mountain Regions : A Review of Possible Impacts ». In : *Climatic Change* 59 (2003), p. 5-31 (cf. p. 7).

[19] H. François et al. « Crossing numerical simulations of snow conditions with a spatially-resolved socio-economic database of ski resorts : A proof of concept in the French Alps ». In : *Cold Regions Science and Technology* 108 (2014), p. 98-112. doi : https://doi.org/10.1016/j.coldregions.2014.08.005 (cf. p. 8).

[20] Clare Morrison et Catherine M. Pickering. « Perceptions of climate change impacts, adaptation and limits to adaption in the Australian Alps : the ski-tourism industry and key stakeholders ». In : *Journal of Sustainable Tourism* 21.2 (2013), p. 173-191. doi : 10.1080/09669582.2012.681789 (cf. p. 8).

[21] C. Viel et al. « How seasonal forecast could help a decision maker : an example of climate service for water resource management ». In : *Advances in Science and Research* 13 (2016), p. 51-55. doi : 10.5194/asr-13-51-2016 (cf. p. 8).

[22] Gabriel Rondeau-Genesse, Mélanie Trudel et Robert Leconte. « Monitoring snow wetness in an Alpine Basin using combined C-band SAR and MODIS data ». In : *Remote Sensing of Environment* 183 (2016), p. 304-317. doi : https://doi.org/10.1016/j.rse.2016.06.003 (cf. p. 8).

[23] Ala Bahrami, Kalifa Goï¨ta et Ramata Magagi. « Analysing the contribution of snow water equivalent to the terrestrial water storage over Canada ». In : *Hydrological Processes* 34.2 (2020), p. 175-188. doi : https://doi.org/10.1002/hyp.13625 (cf. p. 8).

[24] Stefano Baggi et Juerg Schweizer. « Characteristics of wet-snow avalanche activity : 20 years of observations from a high alpine valley (Dischma, Switzerland) ». In : *Natural Hazards* 50 (juill. 2008), p. 97-108. doi : 10.1007/s11069-008-9322-7 (cf. p. 8).

[25] John J. Clague et Jim E. O'Connor. « Chapter 14 - Glacier-related outburst floods ». In : *Snow and Ice-Related Hazards, Risks, and Disasters (Second Edition)*. Sous la dir. de Wilfried Haeberli et Colin Whiteman. Second Edition. Hazards and Disasters Series. Elsevier, 2021, p. 467-499. doi : https://doi.org/10.1016/B978-0-12-817129-5.00019-6 (cf. p. 8).

[26] Intergovernmental Panel on Climate Change (IPCC). « Ocean, Cryosphere and Sea Level Change ». In : *Climate Change 2021 – The Physical Science Basis : Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2023, p. 1211-1362 (cf. p. 8).

[27] Grégoire Guillet et Ludovic Ravanel. « Variations in surface area of six ice aprons in the Mont-Blanc massif since the Little Ice Age ». In : *Journal of Glaciology* 66.259 (2020), p. 777-789. doi : 10.1017/jog.2020.46 (cf. p. 8).

[28] Douglas Benn et David J. A. Evans. *Glaciers and Glaciation, 2nd Edition*. 2e éd. London : Routledge, juin 2010 (cf. p. 8).

[29] S. Kaushik et al. « Analysis of the Temporal Evolution of Ice Aprons in the Mont-Blanc Massif Using X and C-Band SAR Images ». In : *Frontiers in Remote Sensing* 3 (2022). doi : 10.3389/frsen.2022.930021 (cf. p. 8, 9, 34, 35).

[30] J. Faillettaz, M. Funk et M. Vagliasindi. « Time forecast of a break-off event from a hanging glacier ». In : *The Cryosphere* 10.3 (2016), p. 1191-1200. doi : 10.5194/tc-10-1191-2016 (cf. p. 9).

[31] R. V. Engeset et al. « Change Detection and Monitoring of Glacier Mass Balance and Facies Using ERS SAR Winter Images over Svalbard ». In : *International Journal of Remote Sensing* 23.10 (jan. 2002), p. 2023-2050. doi : 10.1080/01431160110075550. (Visité le 03/02/2024) (cf. p. 9, 34).

[32] C. Ancey. *Guide neige et avalanches : connaissances, pratiques, sécurité*. [Departement_IRSTEA]EEE [TR1_IRSTEA]E31-Ingénierie et prévention des risques naturels en montagne. Edisud, 1996, p. 315 (cf. p. 9).

[33] S.L. Dingman. *Physical Hydrology*. Physical Hydrology ptie. 2. Prentice Hall, 2002 (cf. p. 9).

[34] Frank Techel et Christine Pielmeier. « Point observations of liquid water content in wet snow - investigating methodical, spatial and temporal aspects ». In : *The Cryosphere* 5 (2010), p. 405-418 (cf. p. 9, 11, 74).

[35] Association internationale des sciences cryosphériques et al. « The International classification for seasonal snow on the ground ». English. In : *Technical documents in hydrology*. Technical documents in hydrology, 83 ; IACS contribution, no.1. Incl. bibl. 2009, p. 90 (cf. p. 9, 10).

[36] Ya-Lun S. Tsai et al. « Remote Sensing of Snow Cover Using Spaceborne SAR : A Review ». en. In : *Remote Sensing* 11.12 (19 juin 2019). Number : 12, p. 1456. doi : 10.3390/rs11121456. (Visité le 09/11/2021) (cf. p. 9, 10, 30-33, 63, 74).

[37] H. A. P. Hapuarachchi, Q. J. Wang et T. C. Pagano. « A review of advances in flash flood forecasting ». In : *Hydrological Processes* 25.18 (2011), p. 2771-2784. doi : https://doi.org/10.1002/hyp.8040 (cf. p. 10).

[38] C. Vera Valero et al. « Modeling the influence of snow cover temperature and water content on wet-snow avalanche runout ». In : *Natural Hazards and Earth System Sciences* 18.3 (2018), p. 869-887. doi : 10.5194/nhess-18-869-2018 (cf. p. 10).

[39] Sascha Bellaire et al. « On forecasting wet-snow avalanche activity using simulated snow cover data ». In : *Cold Regions Science and Technology* 144 (2017). International Snow Science Workshop 2016 Breckenridge, p. 28-38. doi : https://doi.org/10.1016/j.coldregions.2017.09.013 (cf. p. 10).

[40] N. Wever, Cesar Valero et Charles Fierz. « Assessing wet snow avalanche activity using detailed physics based snowpack simulations : SIMULATING WET SNOW AVALANCHE ACTIVITY ». In : *Geophysical Research Letters* 43 (juin 2016). doi : 10.1002/2016gl068428 (cf. p. 10).

[41] Carlo Marin et al. « Use of Sentinel-1 radar observations to evaluate snowmelt dynamics in alpine regions ». en. In : *The Cryosphere* 14.3 (12 mars 2020). Number : 3, p. 935-956. doi : 10.5194/tc-14-935-2020. (Visité le 21/10/2021) (cf. p. 10, 32, 33).

[42] David R. DeWalle et Albert Rango. « Snowmelt-runoff processes ». In : *Principles of Snow Hydrology*. Cambridge University Press, 2008, p. 235-265 (cf. p. 11).

[43] Ismail Gultepe. « Chapter Three - Mountain Weather : Observation and Modeling ». In : t. 56. Advances in Geophysics. Elsevier, 2015, p. 229-312. doi : https://doi.org/10.1016/bs.agph.2015.01.001 (cf. p. 11).

[44] S. Kaushik et al. « Monitoring Hanging Glacier Dynamics From Sar Images Using Corner Reflectors And Field Measurements In The Mont-blanc Massif ». en. In : *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* V-3-2022 (mai 2022), p. 325-332. doi : 10.5194/isprs-annals-V-3-2022-325-2022 (cf. p. 11).

[45] N. J. Kinar et J. W. Pomeroy. « Measurement of the physical properties of the snowpack ». In : *Reviews of Geophysics* 53.2 (2015), p. 481-544. doi : https://doi.org/10.1002/2015RG000481 (cf. p. 11).

[46] Franziska Koch et al. « Measuring Snow Liquid Water Content with Low-Cost GPS Receivers ». In : *Sensors* 14.11 (2014), p. 20975-20999. doi : 10.3390/s141120975 (cf. p. 11).

[47] Ladina Steiner et al. « (Near) Real-Time Snow Water Equivalent Observation Using GNSS Refractometry and RTKLIB ». In : *Sensors* 22.18 (2022). doi : 10.3390/s22186918 (cf. p. 11).

[48] Angela Lundberg, Nils Granlund et David Gustafsson. « Towards automated 'Ground truth' snow measurements–a review of operational and new measurement methods for Sweden, Norway, and Finland ». In : *Hydrological Processes* 24.14 (2010), p. 1955-1970. doi : https://doi.org/10.1002/hyp.7658 (cf. p. 12).

[49] A. Fontrodona-Bach et al. « NH-SWE : Northern Hemisphere Snow Water Equivalent dataset based on in situ snow depth time series ». In : *Earth System Science Data* 15.6 (2023), p. 2577-2599. doi : 10.5194/essd-15-2577-2023 (cf. p. 12, 34, 140).

[50] Carlo Maria Carmagnola. « Mesure, analyse et modélisation des processus physiques du manteau neigeux sec ». Theses. Université de Grenoble, nov. 2014 (cf. p. 12, 14).

[51] Dominique Tapsoba et al. « Apport de la technique du krigeage avec dérive externe pour une cartographie raisonnée de l'équivalent en eau de la neige : Application aux bassins de la rivière Gatineau ». In : *Canadian Journal of Civil Engineering* 32 (2005), p. 289-297 (cf. p. 12).

[52] J. Odry et al. « Using artificial neural networks to estimate snow water equivalent from snow depth ». en. In : *Canadian Water Resources Journal / Revue canadienne des ressources hydriques* 45.3 (juill. 2020), p. 252-268. doi : 10.1080/07011784.2020.1796817. (Visité le 12/04/2022) (cf. p. 12, 33, 34).

[53] Guilhem Marsy et al. « Monitoring Mountain Cryosphere Dynamics by Time Lapse Stereo Photogrammetry ». In : *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (2020), p. 459-466 (cf. p. 12).

[54] Matthias Kopp, Ye Tuo et Markus Disse. « Fully automated snow depth measurements from time-lapse images applying a convolutional neural network ». en. In : *Science of The Total Environment* 697 (déc. 2019), p. 134213. doi : 10.1016/j.scitotenv.2019.134213. (Visité le 12/04/2022) (cf. p. 12).

[55] M. Stefko et al. « Coherent backscatter enhancement in bistatic Ku- and X-band radar observations of dry snow ». In : *The Cryosphere* 16.7 (2022), p. 2859-2879. doi : 10.5194/tc-16-2859-2022 (cf. p. 12, 13).

[56] Richard Armstrong et Eric Brun. « Snow and Climate : Physical Processes, Surface Energy Exchange and Modeling ». In : *Snow and Climate : Physical Processes, Surface Energy Exchange and Modeling, Edited by R.L. Armstrong and E. Brun. Cambridge : Cambridge University Press, 2008.* (jan. 2008) (cf. p. 12, 13).

[57] Richard Essery et al. « A comparison of 1701 snow models using observations from an alpine site ». In : *Advances in Water Resources* 55 (2013). Snow–Atmosphere Interactions and Hydrological Consequences, p. 131-148. doi : https://doi.org/10.1016/j.advwatres.2012.07.013 (cf. p. 12, 13).

[58] Satoru Yamaguchi, Atsushi Sato et Michael Lehning. « Application of the numerical snowpack model (SNOWPACK) to the wet-snow region in Japan ». In : *Annals of Glaciology* 38 (2004), p. 266-272 (cf. p. 13).

[59] Ulrich Strasser, Michael Warscher et Glen E. Liston. « Modeling Snow–Canopy Processes on an Idealized Mountain ». In : *Journal of Hydrometeorology* 12.4 (2011), p. 663-677. doi : 10.1175/2011JHM1344.1 (cf. p. 13).

[60] Michael Lehning et al. « SNOWPACK model calculations for avalanche warning based upon a new network of weather and snow stations ». In : *Cold Regions Science and Technology* 30.1-3 (déc. 1999), p. 145-157. doi : 10.1016/s0165-232x(99)00022-1. (Visité le 29/01/2024) (cf. p. 13, 15).

[61] V Vionnet et al. « The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2 ». en. In : *Geosci. Model Dev.* (2012), p. 19 (cf. p. 13, 15).

[62] E. Brun et al. « A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting ». In : *Journal of Glaciology* 38.128 (1992), p. 13-22. doi : 10.3189/s0022143000009552 (cf. p. 13, 15).

[63] Daniel Günther et al. « Uncertainties in Snowpack Simulations—Assessing the Impact of Model Structure, Parameter Choice, and Forcing Data Error on Point-Scale Energy Balance Snow Model Performance ». In : *Water Resources Research* 55.4 (2019), p. 2779-2800. doi : https://doi.org/10.1029/2018WR023403 (cf. p. 13).

[64] M. S. Raleigh, J. D. Lundquist et M. P. Clark. « Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework ». In : *Hydrology and Earth System Sciences* 19.7 (2015), p. 3153-3179. doi : 10.5194/hess-19-3153-2015 (cf. p. 13).

[65] Francesco Avanzi et al. « Model complexity and data requirements in snow hydrology : seeking a balance in practical applications ». In : *Hydrological Processes* 30.13 (2016), p. 2106-2118. doi : https://doi.org/10.1002/hyp.10782 (cf. p. 13).

[66] Michael Engel et al. « Snow model sensitivity analysis to understand spatial and temporal snow dynamics in a high-elevation catchment ». In : *Hydrological Processes* 31.23 (2017), p. 4151-4168. doi : https://doi.org/10.1002/hyp.11314 (cf. p. 13).

[67] Jordi Bolibar et al. « Deep learning applied to glacier evolution modelling ». en. In : *The Cryosphere* 14.2 (fév. 2020), p. 565-584. doi : 10.5194/tc-14-565-2020. (Visité le 12/04/2022) (cf. p. 13).

[68] Y. Durand et al. « A meteorological estimation of relevant parameters for snow models ». In : *Annals of Glaciology* 18 (1993), p. 65-71. doi : 10.3189/s0260305500011277 (cf. p. 13).

[69] F. Larue et al. « Simulation and Assimilation of Passive Microwave Data Using a Snowpack Model Coupled to a Calibrated Radiative Transfer Model Over Northeastern Canada ». In : *Water Resources Research* 54.7 (juill. 2018), p. 4823-4848. doi : https://doi.org/10.1029/2017WR022132. (Visité le 29/01/2024) (cf. p. 15).

[70] Richard L.H. Essery et al. « A comparison of four snow models using observations from an alpine site ». In : *Climate Dynamics* 15 (1999), p. 583-593 (cf. p. 15).

[71] A. Haddjeri et al. « Exploring the sensitivity to precipitation, blowing snow, and horizontal resolution of the spatial distribution of simulated snow cover ». In : *EGUsphere* 2023 (2023), p. 1-44. doi : 10.5194/egusphere-2023-2604 (cf. p. 15, 16).

[72] Matthieu Gallet et al. « Wet Snow Detection From Satellite SAR Images by Machine Learning With Physical Snowpack Model Labeling ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17 (2024), p. 2901-2917. doi : 10.1109/JSTARS.2023.3342990 (cf. p. 15, 108).

[73] Chang Liu et al. « Identifying Wet and Dry Snow With Dual-Polarized C-Band SAR Data Based on Markov Random Field Model ». In : *IEEE Geoscience and Remote Sensing Letters* 20 (2023), p. 1-5. doi : 10.1109/lgrs.2023.3266399 (cf. p. 15, 65).

[74] Arnab Muhuri et al. « Snow Cover Mapping Using Polarization Fraction Variation With Temporal RADARSAT-2 C-Band Full-Polarimetric SAR Data Over the Indian Himalayas ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11.7 (juill. 2018). Conference Name : IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, p. 2192-2209. doi : 10.1109/jstars.2018.2817687 (cf. p. 15).

[75] Jun Yang et al. « The role of satellite remote sensing in climate change studies ». In : *Nature Climate Change* 3 (2013), p. 875-883 (cf. p. 16).

[76] Dimitris Poursanidis et Nektarios Chrysoulakis. « Remote Sensing, natural hazards and the contribution of ESA Sentinels missions ». In : *Remote Sensing Applications : Society and Environment* 6 (2017), p. 25-38. doi : https://doi.org/10.1016/j.rsase.2017.02.001 (cf. p. 16).

[77] Ila Chawla, L. Karthikeyan et Ashok K. Mishra. « A review of remote sensing applications for water security : Quantity, quality, and extremes ». In : *Journal of Hydrology* 585 (2020), p. 124826. doi : https://doi.org/10.1016/j.jhydrol.2020.124826 (cf. p. 16).

[78] Jeff Dozier et Thomas H. Painter. « Multispectral and hyperspectral remote sensing of alpine snow properties ». In : *Annual Review of Earth and Planetary Sciences* 32.Volume 32, 2004 (2004), p. 465-494. doi : https://doi.org/10.1146/annurev.earth.32.101802.120404 (cf. p. 16).

[79] Mohd Anul Haq et al. « Artificial neural network-based modeling of snow properties using field data and hyperspectral imagery ». In : *Natural Resource Modeling* 32.4 (2019), e12229. doi : https://doi.org/10.1111/nrm.12229 (cf. p. 16).

[80] Akshay Gore et al. « Glacier surface characteristics derivation and monitoring using Hyperspectral datasets : a case study of Gepang Gath glacier, Western Himalaya ». In : *Geocarto International* 34.1 (2019), p. 23-42. doi : 10.1080/10106049.2017.1357766 (cf. p. 16).

[81] Francis R Valovcin. *Snow/cloud discrimination*. 349. Air Force Geophysics Laboratories, Air Force Systems Command, United States . . ., 1976 (cf. p. 16).

[82] D. Varshney et al. « Snow and cloud discrimination using convolutional neural networks ». In : *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* Iv-5 (2018), p. 59-63. doi : 10.5194/isprs-annals-IV-5-59-2018 (cf. p. 17).

[83] Jean-Pierre Dedieu et al. « On the Importance of High-Resolution Time Series of Optical Imagery for Quantifying the Effects of Snow Cover Duration on Alpine Plant Habitat ». en. In : *Remote Sensing* 8.6 (juin 2016), p. 481. doi : 10.3390/rs8060481. (Visité le 02/02/2022) (cf. p. 17).

[84] J R Leigh et al. « Identifying and mapping very small (<0.5 $km^2$) mountain glaciers on coarse to high-resolution imagery ». en. In : *J. Glaciol.* 65.254 (déc. 2019), p. 873-888 (cf. p. 17).

[85] Andreas Juergen Dietz et al. « Remote sensing of snow – a review of available methods ». en. In : *International Journal of Remote Sensing* 33.13 (juill. 2012), p. 4094-4134. doi : 10.1080/01431161.2011.640964. (Visité le 18/03/2022) (cf. p. 17).

[86] Xuejun Guo et al. « Extraction of snow cover from high-resolution remote sensing imagery using deep learning on a small dataset ». en. In : *Remote Sensing Letters* 11.1 (jan. 2020), p. 66-75. DOI : 10.1080/2150704x.2019.1686548. (Visité le 12/04/2022) (cf. p. 17).

[87] Lin Liu. « A Review of Deep Learning for Cryospheric Studies ». In : *Deep Learning for the Earth Sciences*. John Wiley et Sons, Ltd, 2021. Chap. 17, p. 258-268. DOI : https://doi.org/10.1002/9781119646181.ch17 (cf. p. 17).

[88] S. Kandasamy et al. « A Comparison Of Methods For Smoothing And Gap Filling Time Series Of Remote Sensing Observations : Application To MODIS LAI Products ». In : *Biogeosciences* 10.6 (2013), p. 4055-4071. DOI : 10.5194/bg-10-4055-2013 (cf. p. 17).

[89] Jiarui Dong et al. « Using Air Temperature to Quantitatively Predict the MODIS Fractional Snow Cover Retrieval Errors over the Continental United States ». In : *Journal of Hydrometeorology* 15.2 (2014), p. 551-562. DOI : 10.1175/jhm-d-13-060.1 (cf. p. 17).

[90] Tsai et al. « Wet and Dry Snow Detection Using Sentinel-1 SAR Data for Mountainous Areas with a Machine Learning Technique ». en. In : *Remote Sensing* 11.8 (12 avr. 2019). Number : 8, p. 895. DOI : 10.3390/rs11080895. (Visité le 13/09/2021) (cf. p. 17, 29, 33, 34, 55, 65, 87).

[91] J.M. Nicolas. *Une introduction au Radar à Synthèse d'Ouverture satellitaire*. Rapp. tech. 2019, p. 306. (Visité le 16/04/2024) (cf. p. 18, 24, 25).

[92] Franz Meyer. « Spaceborne Synthetic Aperture Radar : Principles, Data Access, and Basic Processing Techniques ». In : (2019). DOI : 10.25966/ez4f-mg98. (Visité le 29/01/2024) (cf. p. 18).

[93] Alberto Moreira et al. « A tutorial on synthetic aperture radar ». en. In : *IEEE Geoscience and Remote Sensing Magazine* 1.1 (mars 2013), p. 6-43. DOI : 10.1109/mgrs.2013.2248301. (Visité le 03/12/2021) (cf. p. 18).

[94] Helmut Rott et al. « CoReH2O - A Ku- and X-Band SAR Mission for Snow and Ice Monitoring ». In : *7th European Conference on Synthetic Aperture Radar*. 2008, p. 1-4 (cf. p. 19).

[95] Africa Flores et al. « Synthetic Aperture Radar (SAR) Handbook : Comprehensive Methodologies for Forest Monitoring and Biomass Estimation ». In : (avr. 2019). Publisher : NASA. DOI : 10.25966/nr2c-s697. (Visité le 30/07/2023) (cf. p. 20, 21, 23).

[96] Shuji Fujita et al. « A summary of the complex dielectric permittivity of ice in the megahertz range and its applications for radar sounding of polar ice sheets ». In : 2000 (cf. p. 20).

[97] J. W. Glen et J. G. Paren. « The Electrical Properties of Snow and Ice ». In : *Journal of Glaciology* 15.73 (1975), p. 15-38. DOI : 10.3189/S0022143000034249 (cf. p. 21).

[98] Shaowei Li, Wenbin Xu et Zhiwei Li. « Review of the SBAS InSAR Time-series algorithms, applications, and challenges ». In : *Geodesy and Geodynamics* 13.2 (2022). Advances in InSAR algorithms, deformation monitoring and source parameter inversion, p. 114-126. DOI : https://doi.org/10.1016/j.geog.2021.09.007 (cf. p. 22).

[99] JJ Van Zyl et Y Kim. « Basic principles of SAR polarimetry ». In : *Synthetic Aperture Radar Polarimetry*. Wiley, 2011, p. 23-72 (cf. p. 23).

[100] Laurent Ferro-Famil et Eric Pottier. « Radar Polarimetry Basics and Selected Earth Remote Sensing Applications ». In : *Academic Press Library in Signal Processing*. T. 2. Elsevier, 2014, p. 1119-1244. DOI : 10.1016/b978-0-12-396500-4.00021-1. (Visité le 13/02/2023) (cf. p. 23).

[101] Bochen Zhang et al. « A review of methods for mitigating ionospheric artifacts in differential SAR interferometry ». In : *Geodesy and Geodynamics* 13.2 (2022). Advances in InSAR algorithms, deformation monitoring and source parameter inversion, p. 160-169. DOI : https://doi.org/10.1016/j.geog.2021.12.001 (cf. p. 23).

[102] David Small. « Flattening Gamma : Radiometric Terrain Correction for SAR Imagery ». In : *IEEE Transactions on Geoscience and Remote Sensing* 49.8 (août 2011), p. 3081-3093. DOI : 10.1109/tgrs.2011.2120616. (Visité le 29/01/2024) (cf. p. 23).

[103] Renaud Fallourd. « Suivi des glaciers alpins par combinaison d'informations hétérogènes : images SAR Haute Résolution et mesures terrain ». Theses. Université de Grenoble, avr. 2012 (cf. p. 23).

[104] Dong-Xiao Yue et al. « Synthetic Aperture Radar Image Statistical Modeling : Part One-Single-Pixel Statistical Models ». en. In : *IEEE Geosci. Remote Sens. Mag.* 9.1 (mars 2021). Number : 1, p. 82-114. DOI : 10.1109/mgrs.2020.3004508. (Visité le 09/11/2021) (cf. p. 24).

[105] Prabhishek Singh et al. « A Review on SAR Image and its Despeckling ». In : *Archives of Computational Methods in Engineering* 28 (2021), p. 4633-4653 (cf. p. 24).

[106] S. Quegan et al. « Multitemporal ERS SAR analysis applied to forest mapping ». In : *IEEE Transactions on Geoscience and Remote Sensing* 38.2 (2000), p. 741-753. DOI : 10.1109/36.842003 (cf. p. 24).

[107] Emanuele Dalsasso, Loïc Denis et Florence Tupin. « As If by Magic : Self-Supervised Training of Deep Despeckling Networks With MERLIN ». In : *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), p. 1-13. DOI : 10.1109/TGRS.2021.3128621 (cf. p. 24).

[108] C. Oliver et S. Quegan. *Understanding Synthetic Aperture Radar Images*. SciTech radar and defense series. SciTech Publ., 2004 (cf. p. 26).

[109] K.D. Ward. « Compound representation of high resolution sea clutter ». English. In : *Electronics Letters* 17 (16 août 1981), 561-563(2) (cf. p. 26).

[110] H. H. Arsenault et G. April. « Properties of speckle integrated with a finite aperture and logarithmically transformed ». In : *J. Opt. Soc. Am.* 66.11 (nov. 1976), p. 1160-1163. DOI : 10.1364/JOSA.66.001160 (cf. p. 26).

[111] Hua Xie, L.E. Pierce et F.T. Ulaby. « Statistical properties of logarithmically transformed speckle ». In : *IEEE Transactions on Geoscience and Remote Sensing* 40.3 (2002), p. 721-727. DOI : 10.1109/TGRS.2002.1000333 (cf. p. 26).

[112] Dong-Xiao Yue et al. « Synthetic Aperture Radar Image Statistical Modeling : Part Two-Spatial Correlation Models and Simulation ». en. In : *IEEE Geoscience and Remote Sensing Magazine* 9.1 (mars 2021), p. 115-138. DOI : 10.1109/mgrs.2020.3027609. (Visité le 18/03/2022) (cf. p. 26, 30, 49, 128).

[113] Charles-Alban Deledalle et al. « Mathematical models for very high resolution SAR data and their applications ». In : *Mathematical Models for Remote Sensing Image Processing*. Springer, 2018 (cf. p. 26, 128).

[114] P. Mantero, G. Moser et S.B. Serpico. « Partially supervised classification of remote sensing images using SVM-based probability density estimation ». In : *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003.* 2003, p. 327-336. DOI : 10.1109/WARSD.2003.1295212 (cf. p. 27).

[115] JM. Nicolas et S. N. Anfinsen. « Introduction to second kind statistics : Application of log-moments and log-cumulants to the analysis of radar image distributions ». In : *Trait. Signal* 19.3 (2002), p. 139-167 (cf. p. 28).

[116] Vladimir A. Krylov et al. « On the Method of Logarithmic Cumulants for Parametric Probability Density Function Estimation ». In : *IEEE Transactions on Image Processing* 22.10 (oct. 2013), p. 3791-3806. DOI : 10.1109/TIP.2013.2262285 (cf. p. 28).

[117] S. Pettinato et al. « Snow cover mapping by using optical and SAR data ». In : *Image and Signal Processing for Remote Sensing XV*. T. 7477. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Sept. 2009, 74771n, 74771n. DOI : 10.1117/12.830593 (cf. p. 29).

[118] Anna Wendleder, Andreas J. Dietz et Katharina Schork. « Mapping Snow Cover Extent Using Optical and SAR Data ». en. In : *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. Valencia : Ieee, juill. 2018, p. 5104-5107. DOI : 10.1109/igarss.2018.8518374. (Visité le 18/03/2022) (cf. p. 29).

[119] Guangjun He et al. « Dry and Wet Snow Cover Mapping in Mountain Areas Using SAR and Optical Remote Sensing Data ». en. In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.6 (15 mars 2017). Number : 6, p. 2575-2588. DOI : 10.1109/jstars.2017.2673409. (Visité le 13/09/2021) (cf. p. 29, 65).

[120] Alexander Brenning. « Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection ». en. In : *Remote Sensing of Environment* 113.1 (jan. 2009), p. 239-247. DOI : 10.1016/j.rse.2008.09.005. (Visité le 12/04/2022) (cf. p. 30, 65).

[121] Silvan Leinss, Giuseppe Parrella et Irena Hajnsek. « Snow Height Determination by Polarimetric Phase Differences in X-Band SAR Data ». In : *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* 7.9 (9 sept. 2014). Number : 9, p. 3794-3810. doi : 10.1109/jstars.2014.2323199. (Visité le 31/08/2021) (cf. p. 30).

[122] Jarkko T. Koskinen et al. « Monitoring of Snow-Cover Properties During the Spring Melting Period in Forested Areas ». In : *IEEE Transactions on Geoscience and Remote Sensing* 48.1 (2010), p. 50-58. doi : 10.1109/tgrs.2009.2024755 (cf. p. 30).

[123] Hans Lievens et al. « Snow depth variability in the Northern Hemisphere mountains observed from space ». en. In : *Nature Communications* 10.1 (déc. 2019), p. 4629. doi : 10.1038/s41467-019-12566-y. (Visité le 09/03/2022) (cf. p. 30).

[124] Praveen Thakur et al. « Estimation of Snow Cover Area, Snow Physical Properties and Glacier Classification in Parts of Western Himalayas Using C-Band SAR Data ». In : *Journal of the Indian Society of Remote Sensing* 45.3 (juill. 2016). doi : 10.1007/s12524-016-0609-y. (Visité le 29/01/2024) (cf. p. 30).

[125] Nicolas Longépé, S. Allain et Eric Pottier. « Snow wetness monitoring using multi-temporal polarimetric ASAR data. » In : *2007 IEEE International Geoscience and Remote Sensing Symposium*. Ieee, août 2007, p. 2644-2647. doi : 10.1109/igarss.2007.4423388. (Visité le 29/01/2024) (cf. p. 30).

[126] Giuliana Beltramone et al. « Wet snow detection in Patagonian Andes with Sentinel-1 SAR temporal series analysis in GEE ». In : *2020 IEEE Congreso Bienal de Argentina (ARGENCON)*. 2020, p. 1-8. doi : 10.1109/argencon49523.2020.9505487 (cf. p. 30).

[127] Gaëlle Veyssière. « Apport des mesures du radar à synthèse d'ouverture de Sentinel-1 pour l'étude des propriétés du manteau neigeux ». 2019INPT0025. Thèse de doct. 2019 (cf. p. 30).

[128] Nicolas Longépé. « Apport de l'imagerie SAR satellitaire en bandes L et C pour la caractérisation du couvert neigeux ». These de doctorat. Rennes 1, jan. 2008. (Visité le 21/04/2022) (cf. p. 30).

[129] Ghislain Picard, Melody Sandells et Henning Löwe. « SMRT : an active–passive microwave radiative transfer model for snow with multiple microstructure and scattering formulations (v1.0) ». en. In : *Geoscientific Model Development* 11.7 (juill. 2018). Number : 7, p. 2763-2788. doi : 10.5194/gmd-11-2763-2018. (Visité le 27/11/2021) (cf. p. 30).

[130] Narayanarao Bhogapurapu et al. « Dual-polarimetric descriptors from Sentinel-1 GRD SAR data for crop growth assessment ». en. In : *ISPRS Journal of Photogrammetry and Remote Sensing* 178 (août 2021), p. 20-35. doi : 10.1016/j.isprsjprs.2021.05.013. (Visité le 27/04/2022) (cf. p. 31).

[131] Subhadip Dey et al. « A Model-Free Four Component Scattering Power Decomposition for Polarimetric SAR Data ». en. In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), p. 3887-3902. doi : 10.1109/jstars.2021.3069299. (Visité le 09/11/2021) (cf. p. 31).

[132] Qiang Yin et al. « Optimal combination of polarimetric features for vegetation classification in PolSAR image ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12.10 (2019), p. 3919-3931 (cf. p. 31, 55).

[133] S.R. Cloude et E. Pottier. « An entropy based classification scheme for land applications of polarimetric SAR ». In : *IEEE Transactions on Geoscience and Remote Sensing* 35.1 (1997), p. 68-78. doi : 10.1109/36.551935 (cf. p. 31).

[134] Pierre Formont, Jean-Philippe Ovarlez et Frédéric Pascal. « On the Use of Matrix Information Geometry for Polarimetric SAR Image Classification ». en. In : *Matrix Information Geometry*. Springer Berlin Heidelberg, 2013, p. 257-276. doi : 10.1007/978-3-642-30232-9_10. (Visité le 15/10/2021) (cf. p. 31, 55).

[135] Naoto Usami et al. « PolSAR Wet Snow Mapping With Incidence Angle Information ». en. In : *IEEE Geosci. Remote Sensing Lett.* 13.12 (déc. 2016). Number : 12, p. 2029-2033. doi : 10.1109/lgrs.2016.2621891. (Visité le 14/10/2021) (cf. p. 31).

[136] Lucio Mascolo, Shane R. Cloude et Juan M. Lopez-Sanchez. « Model-Based Decomposition of Dual-Pol SAR Data : Application to Sentinel-1 ». en. In : *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), p. 1-19. doi : 10.1109/tgrs.2021.3137588. (Visité le 07/04/2022) (cf. p. 31).

[137] Chang Liu et al. « Identifying wet and dry snow with dual-polarized c-band SAR data based on Markov random field model ». In : *IEEE Geoscience and Remote Sensing Letters* (2023) (cf. p. 31).

[138] R. Panwar et G. Singh. « Classification of glacier with supervised approaches using PolSAR data ». en. In : *Environ. Monit. Assess.* 195.1 (nov. 2022), p. 58 (cf. p. 31).

[139] Mattia Callegari et al. « A Pol-SAR Analysis for Alpine Glacier Classification and Snowline Altitude Retrieval ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.7 (juill. 2016), p. 3106-3121. doi : 10.1109/jstars.2016.2587819. (Visité le 03/02/2024) (cf. p. 31).

[140] Lei Huang et al. « Classification and snow line detection for glacial areas using the polarimetric SAR image ». en. In : *Remote Sensing of Environment* 115.7 (juill. 2011). Number : 7, p. 1721-1732. doi : 10.1016/j.rse.2011.03.004. (Visité le 13/09/2021) (cf. p. 31).

[141] Zhen Li et al. « Glacier Snow Line Detection on a Polarimetric SAR Image ». In : *IEEE Geoscience and Remote Sensing Letters* 9.4 (juill. 2012), p. 584-588. doi : 10.1109/lgrs.2011.2175697. (Visité le 03/02/2024) (cf. p. 31).

[142] B. Barzycka et al. « Comparison of Three Methods for Distinguishing Glacier Zones Using Satellite SAR Data ». In : *Remote Sensing* 15.3 (2023). doi : 10.3390/rs15030690 (cf. p. 31, 32).

[143] Thomas Nagler et al. « Advancements for Snowmelt Monitoring by Means of Sentinel-1 SAR ». en. In : *Remote Sensing* 8.4 (20 avr. 2016). Number : 4, p. 348. doi : 10.3390/rs8040348. (Visité le 13/09/2021) (cf. p. 32, 33, 71).

[144] F. Karbou et al. « Monitoring Wet Snow Over an Alpine Region Using Sentinel-1 Observations ». In : *Remote Sensing* 13.381 (2021), p. 2575-2588. doi : 110.3390/rs13030381 (cf. p. 32).

[145] T. Nagler et H. Rott. « Retrieval of wet snow by means of multitemporal SAR data ». In : *IEEE Transactions on Geoscience and Remote Sensing* 38.2 (avr. 2000), p. 754-765. doi : 10.1109/36.842004. (Visité le 29/01/2024) (cf. p. 32, 33, 42, 49, 75, 88).

[146] Philip Mann. « Spatial and temporal variability of the snow environment in the Western Canadian Arctic ». In : (2018) (cf. p. 32).

[147] R Magagi et M Bernier. « Optimal conditions for wet snow detection using RADARSAT SAR data ». en. In : *Remote Sensing of Environment* 84.2 (fév. 2003). Number : 2, p. 221-233. doi : 10.1016/s0034-4257(02)00104-9. (Visité le 09/11/2021) (cf. p. 32).

[148] Hans Lievens et al. *Sentinel-1 snow depth retrieval at sub-kilometer resolution over theEuropean Alps*. en. preprint. Snow/Remote Sensing, 12 mars 2021. doi : 10.5194/tc-2021-74. (Visité le 09/11/2021) (cf. p. 33).

[149] Fatima Karbou et al. « Monitoring Wet Snow Over an Alpine Region Using Sentinel-1 Observations ». en. In : *Remote Sensing* 13.3 (2021), p. 22. doi : 10.3390/rs13030381 (cf. p. 33).

[150] Xiaoxin Zhu, Lingjia Gu et Ruizhi Ren. « Snow Covered Area Estimation for Forest Area in Northeastern China Based on Multi-temporal Sentinel-1 SAR Data ». en. In : (déc. 2019), p. 7. doi : 10.1109/PIERS-Fall48861.2019.9021756. (Visité le 29/01/2024) (cf. p. 33).

[151] Eduardo Fernandes Montesuma et Fred Maurice Ngole Mboula. « Wasserstein Barycenter for Multi-Source Domain Adaptation ». en. In : *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Ieee, juin 2021, p. 16780-16788. doi : 10.1109/cvpr46437.2021.01651. (Visité le 28/04/2022) (cf. p. 33).

[152] Jamila Mifdal et al. « Hyperspectral and multispectral wasserstein barycenter for image fusion ». en. In : *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Ieee, juill. 2017, p. 3373-3376. doi : 10.1109/igarss.2017.8127721. (Visité le 28/04/2022) (cf. p. 33).

[153] Nicolas Longepe et al. « Snowpack Characterization in Mountainous Regions Using C-Band SAR Data and a Meteorological Model ». In : *IEEE Transactions on Geoscience and Remote Sensing* 47.2 (2009), p. 406-418. doi : 10.1109/TGRS.2008.2006048 (cf. p. 33, 34).

[154] Jewell Lund et al. « Interpreting Sentinel-1 SAR Backscatter Signals of Snowpack Surface Melt/Freeze, Warming, and Ripening, through Field Measurements and Physically-Based SnowModel ». In : *Remote Sensing* 14.16 (2022). doi : 10.3390/rs14164002 (cf. p. 33, 34).

[155] Gaëlle Veyssière et al. « Evaluation of Sub-Kilometric Numerical Simulations of C-Band Radar Backscatter over the French Alps against Sentinel-1 Observations ». en. In : *Remote Sensing* 11.1 (20 déc. 2018). Number : 1, p. 8. doi : 10.3390/rs11010008. (Visité le 27/11/2021) (cf. p. 33, 34).

[156] Jiancheng Shi, J. Dozier et H. Rott. « Deriving snow liquid water content using C-band polarimetric SAR ». In : *Proceedings of IGARSS '93 - IEEE International Geoscience and Remote Sensing Symposium*. 1993, 1038-1041 vol.3. doi : 10.1109/IGARSS.1993.322643 (cf. p. 34).

[157] Divyesh Varade et al. « Modelling of early winter snow density using fully polarimetric C-band SAR data in the Indian Himalayas ». In : *Remote Sensing of Environment* 240 (2020), p. 111699. doi : https://doi.org/10.1016/j.rse.2020.111699 (cf. p. 34).

[158] L. Tsang et al. « Review article : Global monitoring of snow water equivalent using high-frequency radar remote sensing ». In : *The Cryosphere* 16.9 (2022), p. 3531-3573. doi : 10.5194/tc-16-3531-2022 (cf. p. 34).

[159] Monique Bernier et al. « Determination of snow water equivalent using RADARSAT SAR data in eastern Canada ». In : *Hydrological Processes* 13.18 (1999), p. 3041-3051. doi : https://doi.org/10.1002/(SICI)1099-1085(19991230)13:18<3041::AID-HYP14>3.0.CO;2-E (cf. p. 34).

[160] Sun Shaobo et al. « Estimation and Analysis of Snow Water Equivalents Based on C-Band SAR Data and Field Measurements ». In : *Arctic, Antarctic, and Alpine Research* 47.2 (2015), p. 313-326. doi : 10.1657/AAAR00C-13-135 (cf. p. 34).

[161] J. Shi et J. Dozier. « Estimation of snow water equivalence using SIR-C/X-SAR. II. Inferring snow depth and particle size ». In : *IEEE Transactions on Geoscience and Remote Sensing* 38.6 (2000), p. 2475-2488. doi : 10.1109/36.885196 (cf. p. 34).

[162] Jess Grunblatt et Don Atwood. « Mapping lakes for winter liquid water availability using SAR on the North Slope of Alaska ». In : *International Journal of Applied Earth Observation and Geoinformation* 27 (2014). Special Issue on Polar Remote Sensing 2013, p. 63-69. doi : https://doi.org/10.1016/j.jag.2013.05.006 (cf. p. 34).

[163] Matthias Braun et Frank Rau. « Using A Multi-year Data Archive Of Ers Sar Imagery For The Monitoring Of Firn Line Positions And Ablation Patterns On The King George Island Ice Cap (Antarctica) ». In : 1 (2000) (cf. p. 34).

[164] Andreas Hartmann et al. « Bayesian U-Net for Segmenting Glaciers in Sar Imagery ». In : *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. Juill. 2021, p. 3479-3482. doi : 10.1109/IGARSS47720.2021.9554292. (Visité le 03/02/2024) (cf. p. 34).

[165] Debmita Bandyopadhyay et Gulab Singh. « Mass Balance Estimation Using Sar Data in Central Himalaya ». In : *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. Juill. 2018, p. 8522-8524. doi : 10.1109/igarss.2018.8517297. (Visité le 03/02/2024) (cf. p. 34, 35).

[166] Sanchayita Kundu et Manab Chakraborty. « Delineation of glacial zones of Gangotri and other glaciers of Central Himalaya using RISAT-1 C-band dual-pol SAR ». In : *International Journal of Remote Sensing* 36.6 (2015), p. 1529-1550. doi : 10.1080/01431161.2015.1014972 (cf. p. 34, 35).

[167] Giuliana Beltramone et al. « Identification of Seasonal Snow Phase Changes From C-Band SAR Time Series With Dynamic Thresholds ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 (2023), p. 6995-7008. doi : 10.1109/JSTARS.2023.3281149 (cf. p. 34, 35).

[168] J. Arigony-Neto et al. « A time series of SAR data for monitoring changes in boundaries of glacier zones on the Antarctic Peninsula ». In : *Annals of Glaciology* 46 (2007), p. 55-60. doi : 10.3189/172756407782871387 (cf. p. 34, 35, 65, 66).

[169] Suvrat Kaushik et al. « Temporal Evolution of X and C Band SAR Backscattering in theMont-Blanc Massif ». In : *International Geoscience and Remote Sensing Symposium (IGARSS 2023)*. Juill. 2023 (cf. p. 37, 48).

[170] Aurelie Voisin et al. « Classification of Very High Resolution SAR Images of Urban Areas Using Copulas and Texture in a Hierarchical Markov Random Field Model ». en. In : *IEEE Geoscience and Remote Sensing Letters* 10.1 (jan. 2013), p. 96-100. doi : 10.1109/lgrs.2012.2193869. (Visité le 18/03/2022) (cf. p. 47, 50, 65).

[171] Satyakam Baraha, Ajit Kumar Sahoo et Sowjanya Modalavalasa. « A systematic review on recent developments in nonlocal and variational methods for SAR image despeckling ». In : *Signal Processing* 196 (2022), p. 108521. doi : https://doi.org/10.1016/j.sigpro.2022.108521 (cf. p. 47).

[172] Michael Schmitt et al. « There Are No Data Like More Data- Datasets for Deep Learning in Earth Observation ». In : *IEEE Geoscience and Remote Sensing Magazine* 11.3 (sept. 2023), p. 63-97. doi : 10.1109/mgrs.2023.3293459. (Visité le 29/01/2024) (cf. p. 48, 62, 63).

[173] Lorenzo Bruzzone et Begüm Demir. « A review of modern approaches to classification of remote sensing data ». In : *Land Use and Land Cover Mapping in Europe : Practices & Trends* (2014), p. 127-143 (cf. p. 48).

[174] Myung-Jun Lee et al. « Reduction of False Alarm Rate in SAR-MTI Based on Weighted Kurtosis ». In : *IEEE Transactions on Geoscience and Remote Sensing* 59.4 (2021), p. 3122-3135. doi : 10.1109/tgrs.2020.3010582 (cf. p. 48).

[175] Jean-Marie Nicolas. « 1 - Introduction aux Statistiques de deuxième espèce : applications des Logs-moments et des Logs-cumulants à l'analyse des lois d'images radar ». en. In : *undefined* (2002). (Visité le 18/03/2022) (cf. p. 48, 49).

[176] Ding Tao, Anthony P. Doulgeris et Camilla Brekke. « Sea clutter contamination test with log-cumulants ». en. In : Edinburgh, United Kingdom, nov. 2012, p. 85360i. doi : 10.1117/12.976879. (Visité le 30/03/2022) (cf. p. 49).

[177] Stian Normann Anfinsen et Torbjørn Eltoft. « Application of the Matrix-Variate Mellin Transform to Analysis of Polarimetric Radar Images ». en. In : *IEEE Transactions on Geoscience and Remote Sensing* 49.6 (juin 2011), p. 2281-2295. doi : 10.1109/tgrs.2010.2103945. (Visité le 30/03/2022) (cf. p. 49).

[178] Lionel Bombrun, Stian Normann Anfinsen et Olivier Harant. « A Complete Coverage of Log-Cumulant Space in Terms of Distributions for Polarimetric SAR Data ». en. In : (), p. 8 (cf. p. 49).

[179] A. C. Frery M. E. Mejail J. C. Jacobo-Berlles et O. H. Bustos. « Classification of SAR images using a general and tractable multiplicative model ». In : *International Journal of Remote Sensing* 24.18 (2003), p. 3565-3582. doi : 10.1080/0143116021000053274 (cf. p. 49).

[180] Ammar Mian et al. « New Robust Statistics for Change Detection in Time Series of Multivariate SAR Images ». en. In : *Ieee Tsp* 67.2 (jan. 2019). Number : 2, p. 520-534. doi : 10.1109/tsp.2018.2883011. (Visité le 18/10/2021) (cf. p. 49, 65, 66).

[181] Izzeddine Chalabi et Amar Mezache. « Estimators of compound Gaussian clutter with log-normal texture ». In : *Remote Sensing Letters* 10.7 (2019), p. 709-716. doi : 10.1080/2150704X.2019.1601275 (cf. p. 49).

[182] Antoine Collas et al. « Riemannian Optimization for Non-Centered Mixture of Scaled Gaussian Distributions ». In : *IEEE Transactions on Signal Processing* 71 (2023), p. 2475-2490. doi : 10.1109/tsp.2023.3290354. (Visité le 29/01/2024) (cf. p. 49, 142).

[183] Florent Chatelain, Jean-Yves Tourneret et Jordi Inglada. « Change Detection in Multisensor SAR Images Using Bivariate Gamma Distributions ». In : *IEEE Transactions on Image Processing* 17.3 (2008), p. 249-258. doi : 10.1109/TIP.2008.916047 (cf. p. 49).

[184] Chi Liu et al. « Unsupervised Classification of Multilook Polarimetric SAR Data Using Spatially Variant Wishart Mixture Model with Double Constraints ». In : *IEEE Transactions on Geoscience and Remote Sensing* 56.10 (2018), p. 5600-5613. doi : 10.1109/TGRS.2018.2819995 (cf. p. 50).

[185] Eric Bouyé et al. « Copulas for Finance - A Reading Guide and Some Applications ». In : *SSRN Electronic Journal* (2000). doi : 10.2139/ssrn.1032533. (Visité le 29/01/2024) (cf. p. 50).

[186] Faranak Tootoonchi et al. « Copulas for Hydroclimatic Analysis : A Practice-oriented Overview ». In : *WIREs Water* 9.2 (mars 2022). doi : 10.1002/wat2.1579. (Visité le 30/08/2022) (cf. p. 50).

[187] Weiwei Liu. « Copula Multi-label Learning ». en. In : (), p. 10 (cf. p. 50).

[188] F. J. Yanovsky, A. N. Rudiakova et R. B. Sinitsyn. « Multivariate copula approach for polarimetric classification in weather radar applications ». en. In : *2016 17th International Radar Symposium (IRS)*. Ieee, mai 2016, p. 1-5. doi : 10.1109/irs.2016.7497371. (Visité le 18/03/2022) (cf. p. 50).

[189] Xueqian Wang et al. « Proposal-Copula-Based Fusion of Spaceborne and Airborne SAR Images for Ship Target Detection** ». en. In : *Information Fusion* 77 (jan. 2022), p. 247-260. DOI : 10.1016/j.inffus.2021.07.019. (Visité le 09/05/2022) (cf. p. 50).

[190] Chun Kai Ling, Fei Fang et J. Zico Kolter. « Deep archimedean copulas ». In : *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Nips'20. Curran Associates Inc., déc. 2020, p. 1535-1545. (Visité le 09/05/2022) (cf. p. 50).

[191] Nour-Eddine Lasmar et Yannick Berthoumieu. « Gaussian Copula Multivariate Modeling for Texture Image Retrieval Using Wavelet Transforms ». en. In : *IEEE Trans. on Image Process.* 23.5 (mai 2014). Number : 5, p. 2246-2261. DOI : 10.1109/tip.2014.2313232. (Visité le 18/10/2021) (cf. p. 50, 53, 124).

[192] Youssef Stitou, N. Lasmar et Yannick Berthoumieu. « Copulas based multivariate gamma modeling for texture classification ». en. In : *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. Ieee, avr. 2009, p. 1045-1048. DOI : 10.1109/icassp.2009.4959766. (Visité le 18/10/2021) (cf. p. 50).

[193] G. Mercier, G. Moser et S.B. Serpico. « Conditional Copulas for Change Detection in Heterogeneous Remote Sensing Images ». en. In : *IEEE Trans. Geosci. Remote Sensing* 46.5 (mai 2008), p. 1428-1441. DOI : 10.1109/tgrs.2008.916476. (Visité le 12/10/2022) (cf. p. 50).

[194] Ricardo J. Bessa. « On the quality of the Gaussian copula for multi-temporal decision-making problems ». en. In : *2016 Power Systems Computation Conference (PSCC)*. Ieee, juin 2016, p. 1-7. DOI : 10.1109/pscc.2016.7541001. (Visité le 09/05/2022) (cf. p. 50).

[195] Hamed Valizadeh Haghi et Saeed Lotfifard. « Spatiotemporal Modeling of Wind Generation for Optimal Energy Storage Sizing ». In : *IEEE Transactions on Sustainable Energy* 6.1 (jan. 2015), p. 113-121. DOI : 10.1109/tste.2014.2360702. (Visité le 09/05/2022) (cf. p. 50).

[196] Shiyong Cui et Mihai Datcu. « Mining Satellite Image Time Series : Statistical Modeling and Evolution Analysis ». en. In : *2011 International Symposium on Image and Data Fusion*. Ieee, août 2011, p. 1-4. DOI : 10.1109/isidf.2011.6024237. (Visité le 18/03/2022) (cf. p. 50).

[197] Tim Bedford et Roger M Cooke. « Vines–a new graphical model for dependent random variables ». In : *The Annals of Statistics* 30.4 (2002), p. 1031-1068 (cf. p. 50).

[198] Claudia Czado et Thomas Nagler. « Vine Copula Based Modeling ». In : *Annual Review of Statistics and Its Application* 9.1 (mars 2022), p. 453-477. DOI : 10.1146/annurev-statistics-040220-101153. (Visité le 12/10/2022) (cf. p. 50).

[199] Roger B. Nelsen. *An Introduction to Copulas*. en. Réd. par P. Bickel et al. 2nd ed. T. 139. Lecture Notes in Statistics. New York, NY : Springer, 1999. (Visité le 30/03/2023) (cf. p. 50, 123).

[200] M. Ishaq Bhatti et Hung Quang Do. « Recent development in copula and its applications to the energy, forestry and environmental sciences ». In : *International Journal of Hydrogen Energy* 44.36 (juill. 2019), p. 19453-19473. DOI : 10.1016/j.ijhydene.2019.06.015. (Visité le 30/08/2022) (cf. p. 50, 124).

[201] Houman Safaai et al. « Information Estimation Using Nonparametric Copulas ». In : *Physical Review E* 98.5 (nov. 2018), p. 053302. DOI : 10.1103/PhysRevE.98.053302. (Visité le 29/01/2024) (cf. p. 50).

[202] Gianfausto Salvadori et al. « Extremes in Nature : An Approach Using Copulas ». In : *Extremes in Nature : An Approach Using Copulas, By G. Salvadori, C. De Michele, N.T. Kottegoda, and R. Rosso. Berlin : Springer, 2007.* 49 (jan. 2007) (cf. p. 51).

[203] Alexander J. McNeil et Johanna Nešlehová. « Multivariate Archimedean copulas, d-monotone functions and l1-norm symmetric distributions ». In : *The Annals of Statistics* 37.5B (2009), p. 3059-3097. DOI : 10.1214/07-AOS556 (cf. p. 51).

[204] E. J. Gumbel. « Bivariate Exponential Distributions ». In : *Journal of the American Statistical Association* 55.292 (1960), p. 698-707. DOI : 10.1080/01621459.1960.10483368 (cf. p. 51).

[205] D. G. CLAYTON. « A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence ». In : *Biometrika* 65.1 (avr. 1978), p. 141-151. DOI : 10.1093/biomet/65.1.141 (cf. p. 51).

[206] M. J. Frank. « On the simultaneous associativity of F(x, y) and x+y-F(x, y) ». In : *aequationes mathematicae* 19 (1978), p. 194-226 (cf. p. 51).

[207] C. GENEST, Kilani Ghoudi et Louis-Paul Rivest. « A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions ». In : *Biometrika* 82 (sept. 1995). doi : 10.1093/biomet/82.3.543 (cf. p. 52).

[208] Piotr Jaworski et al. « Copula Theory and Its Applications ». In : 2010 (cf. p. 52, 123).

[209] Gang Yang et al. « Unsupervised Change Detection of SAR Images Based on Variational Multivariate Gaussian Mixture Model and Shannon Entropy ». In : *IEEE Geoscience and Remote Sensing Letters* 16.5 (2019), p. 826-830. doi : 10.1109/LGRS.2018.2879969 (cf. p. 52).

[210] Heng-Chao Li et al. « Unsupervised learning of generalized gamma mixture model with application in statistical modeling of high-resolution SAR images ». In : *IEEE Transactions on Geoscience and Remote Sensing* 54.4 (2015), p. 2153-2170 (cf. p. 52).

[211] Michéle Basseville. « Divergence measures for statistical data processing–An annotated bibliography ». In : *Signal Processing* 93.4 (2013), p. 621-633 (cf. p. 53, 127).

[212] Roland Kwitt et Andreas Uhl. « Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval ». In : *2008 15th IEEE International Conference on Image Processing*. 2008, p. 933-936. doi : 10.1109/icip.2008.4711909 (cf. p. 53, 122).

[213] Michael Kampffmeyer et al. « Deep divergence-based approach to clustering ». In : *Neural Networks* 113 (2019), p. 91-101 (cf. p. 53).

[214] Frank Nielsen. « On a Generalization of the Jensen–Shannon Divergence and the Jensen–Shannon Centroid ». In : *Entropy* 22.2 (fév. 2020), p. 221. doi : 10.3390/e22020221. (Visité le 29/01/2024) (cf. p. 53).

[215] Christian Bauckhage. *Computing the Kullback-Leibler Divergence between two Weibull Distributions*. 2013 (cf. p. 53).

[216] Mina Rezaei et al. « Deep Bregman divergence for self-supervised representations learning ». In : *Computer Vision and Image Understanding* 235 (2023), p. 103801. doi : https://doi.org/10.1016/j.cviu.2023.103801 (cf. p. 53).

[217] Hatice Kubra Cilingir, Rachel Manzelli et Brian Kulis. « Deep divergence learning ». In : *International Conference on Machine Learning*. Pmlr. 2020, p. 2027-2037 (cf. p. 53, 127).

[218] Qian Xu et Lina J. Karam. « Change detection on SAR images by a parametric estimation of the KL-divergence between Gaussian Mixture Models ». In : *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2013, p. 2109-2113. doi : 10.1109/ICASSP.2013.6638026 (cf. p. 53).

[219] Abdourrahmane Mahamane Atto et al. « Multidate Divergence Matrices for the Analysis of SAR Image Time Series ». en. In : *IEEE Transactions on Geoscience and Remote Sensing* 51.4 (avr. 2013). Number : 4, p. 1922-1938. doi : 10.1109/tgrs.2012.2210228. (Visité le 18/10/2021) (cf. p. 53).

[220] Wagner B Silva et al. « Classification of segments in PolSAR imagery by minimum stochastic distances between Wishart distributions ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6.3 (2013), p. 1263-1273 (cf. p. 53, 127).

[221] Xianxiang Qin et al. « Region-based classification of SAR images using Kullback–Leibler distance between generalized gamma distributions ». In : *IEEE Geoscience and Remote Sensing Letters* 12.8 (2015), p. 1655-1659 (cf. p. 53).

[222] Alfréd Rényi. « On measures of entropy and information ». In : *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1 : contributions to the theory of statistics*. T. 4. University of California Press. 1961, p. 547-562 (cf. p. 53).

[223] Tim Van Erven et Peter Harremos. « Rényi divergence and Kullback-Leibler divergence ». In : *IEEE Transactions on Information Theory* 60.7 (2014), p. 3797-3820 (cf. p. 53, 129).

[224] Noor Sajid et al. « Bayesian Brains and the Rényi Divergence ». In : *Neural Computation* 34.4 (mars 2022), p. 829-855. doi : 10.1162/neco_a_01484 (cf. p. 54).

[225] Frank Nielsen et Richard Nock. « On the chi square and higher-order chi distances for approximating f-divergences ». In : *IEEE Signal Processing Letters* 21.1 (jan. 2014), p. 10-13. doi : 10.1109/lsp.2013.2288355 (cf. p. 54).

[226] F. Galland, N. Bertaux et P. Refregier. « Minimum description length synthetic aperture radar image segmentation ». In : *IEEE Transactions on Image Processing* 12.9 (2003), p. 995-1006. doi : 10.1109/TIP.2003.816005 (cf. p. 54).

[227] Xueyuan Gao et Fuyuan Xiao. « An improved belief chi2 divergence for Dempster–Shafer theory and its applications in pattern recognition ». In : *Computational and Applied Mathematics* 41 (sept. 2022). doi : 10.1007/s40314-022-01975-3 (cf. p. 54).

[228] Jiaxin Tang et al. « Large-difference-scale target detection using a revised Bhattacharyya distance in SAR images ». In : *IEEE Geoscience and Remote Sensing Letters* 19 (2022), p. 1-5 (cf. p. 54, 136).

[229] Henri Lanteri. « Divergences. divergences invariantes. applications aux problèmes inverses linéaires, nmf et déconvolution aveugle ». In : (2019) (cf. p. 54).

[230] J.M. Prats-Montalbán, A. de Juan et A. Ferrer. « Multivariate Image Analysis : A Review with Applications ». In : *Chemometrics and Intelligent Laboratory Systems* 107.1 (mai 2011), p. 1-23. doi : 10.1016/j.chemolab.2011.03.002. (Visité le 10/08/2022) (cf. p. 54).

[231] David A. Freedman et Persi Diaconis. « On the histogram as a density estimator :L2 theory ». In : *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57 (1981), p. 453-476 (cf. p. 54).

[232] Shervan Fekri-Ershad. « Texture Analysis Using Machine Learning Methods - A Review ». en. In : 2 (2019), p. 29 (cf. p. 54).

[233] D.G. Lowe. « Object recognition from local scale-invariant features ». In : *Proceedings of the Seventh IEEE International Conference on Computer Vision*. T. 2. 1999, 1150-1157 vol.2. doi : 10.1109/ICCV.1999.790410 (cf. p. 54).

[234] Amin Sedaghat et Hamid Ebadi. « Remote sensing image matching based on adaptive binning SIFT descriptor ». In : *IEEE transactions on geoscience and remote sensing* 53.10 (2015), p. 5283-5293 (cf. p. 54).

[235] Ganggang Dong, Hongwei Liu et Jocelyn Chanussot. « Keypoint-Based Local Descriptors for Target Recognition in SAR Images : A Comparative Analysis ». en. In : *IEEE Geoscience and Remote Sensing Magazine* 9.1 (mars 2021), p. 139-166. doi : 10.1109/mgrs.2020.3005597. (Visité le 18/03/2022) (cf. p. 54).

[236] Yufan Wang et al. « Coastline detection in SAR images using multi-feature and SVM ». en. In : *2011 4th International Congress on Image and Signal Processing*. Shanghai, China : Ieee, oct. 2011, p. 1227-1230. doi : 10.1109/cisp.2011.6100487. (Visité le 18/03/2022) (cf. p. 55).

[237] Wentao-Lv, Qiuze-Yu et Wenxian-Yu. « Water extraction in SAR images using GLCM and Support Vector Machine ». en. In : *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS*. Beijing : Ieee, oct. 2010, p. 740-743. doi : 10.1109/icosp.2010.5655766. (Visité le 18/03/2022) (cf. p. 55).

[238] Jie Feng et al. « Bag-of-Visual-Words Based on Clonal Selection Algorithm for SAR Image Classification ». In : *IEEE Geoscience and Remote Sensing Letters* 8.4 (2011), p. 691-695. doi : 10.1109/LGRS.2010.2100363 (cf. p. 55).

[239] Ammar Mian et al. « Design of New Wavelet Packets Adapted to High-Resolution SAR Images With an Application to Target Detection ». en. In : *IEEE Transactions on Geoscience and Remote Sensing* 57.6 (juin 2019). Number : 6, p. 3919-3932. doi : 10.1109/tgrs.2018.2888993. (Visité le 18/10/2021) (cf. p. 55).

[240] Rodney V. Fonseca et al. « Wavelet Spatio-Temporal Change Detection on Multitemporal SAR Images ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 (2023), p. 4013-4023. doi : 10.1109/jstars.2023.3268601. (Visité le 02/07/2023) (cf. p. 55, 66).

[241] Sara Akodad et al. « Ensemble Learning Approaches Based on Covariance Pooling of CNN Features for High Resolution Remote Sensing Scene Classification ». In : *Remote Sensing* 12.20 (2020). doi : 10.3390/rs12203292 (cf. p. 55).

[242] Lionel Bombrun et al. « Multivariate texture retrieval using the geodesic distance between elliptically distributed random variables ». en. In : *2011 18th IEEE International Conference on Image Processing*. Ieee, sept. 2011, p. 3637-3640. doi : 10.1109/icip.2011.6116506. (Visité le 18/10/2021) (cf. p. 55).

[243] Matthieu Gallet et al. « Classification of GPR Signals via Covariance Pooling on CNN Features within a Riemannian Framework ». In : *International Geoscience and Remote Sensing Symposium*. Juill. 2022 (cf. p. 55).

[244] Eduarda T. C. Chagas et al. « Analysis and Classification of SAR Textures Using Information Theory ». en. In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), p. 663-675. doi : 10.1109/jstars.2020.3031918. (Visité le 09/11/2021) (cf. p. 55).

[245] Trevor Hastie, Robert Tibshirani et Jerome Friedman. *The Elements of Statistical Learning*. Springer New York, 2009 (cf. p. 56).

[246] David J. Lary et al. « Machine learning in geosciences and remote sensing ». en. In : *Geoscience Frontiers* 7.1 (jan. 2016), p. 3-10. doi : 10.1016/j.gsf.2015.07.003. (Visité le 12/04/2022) (cf. p. 56, 108).

[247] Leo Breiman. « Random forests ». In : *Machine learning* 45 (2001), p. 5-32 (cf. p. 56, 109).

[248] Yoav Freund et Robert E Schapire. « A desicion-theoretic generalization of on-line learning and an application to boosting ». In : *European conference on computational learning theory*. Springer. 1995, p. 23-37 (cf. p. 56, 109).

[249] Habiba Muhammad Sani, Ci Lei et Daniel Neagu. « Computational Complexity Analysis of Decision Tree Algorithms ». In : *Artificial Intelligence XXXV*. Cham : Springer International Publishing, 2018, p. 191-197 (cf. p. 56).

[250] Corinna Cortes et Vladimir Naumovich Vapnik. « Support-Vector Networks ». In : *Machine Learning* 20 (1995), p. 273-297 (cf. p. 56, 109).

[251] T. Cover et P. Hart. « Nearest neighbor pattern classification ». In : *IEEE Transactions on Information Theory* 13.1 (1967), p. 21-27. doi : 10.1109/tit.1967.1053964 (cf. p. 56, 109).

[252] J. A. D. Anderson. « Separate sample logistic discrimination ». In : *Biometrika* 59 (1972), p. 19-35 (cf. p. 56, 109).

[253] Frank Rosenblatt. « The perceptron : a probabilistic model for information storage and organization in the brain. » In : *Psychological review* 65.6 (1958), p. 386 (cf. p. 56, 109).

[254] Xiao Xiang Zhu et al. « Deep Learning in Remote Sensing : A Comprehensive Review and List of Resources ». en. In : *IEEE Geoscience and Remote Sensing Magazine* 5.4 (déc. 2017), p. 8-36. doi : 10.1109/mgrs.2017.2762307. (Visité le 12/04/2022) (cf. p. 56).

[255] Markus Reichstein et al. « Deep learning and process understanding for data-driven Earth system science ». en. In : *Nature* 566.7743 (fév. 2019), p. 195-204. doi : 10.1038/s41586-019-0912-1. (Visité le 18/03/2022) (cf. p. 56).

[256] Ali Bahri et al. « Remote Sensing Image Classification via Improved Cross-Entropy Loss and Transfer Learning Strategy Based on Deep Convolutional Neural Networks ». en. In : *IEEE Geoscience and Remote Sensing Letters* 17.6 (juin 2020), p. 1087-1091. doi : 10.1109/lgrs.2019.2937872. (Visité le 11/04/2022) (cf. p. 57).

[257] Xiao Xiang Zhu et al. « Deep Learning Meets SAR ». en. In : *arXiv :2006.10027 [cs, eess, stat]* (jan. 2021). arXiv : 2006.10027. (Visité le 22/10/2021) (cf. p. 57).

[258] Se Yoon Lee, Bowen Lei et Bani Mallick. « Estimation of COVID-19 spread curves integrating global data and borrowing information ». In : *Plos One* 15.7 (juill. 2020), p. 1-17. doi : 10.1371/journal.pone.0236860 (cf. p. 58).

[259] Abdourrahmane Mahamane Atto et al. « Parametric Rectified Power Sigmoid Units : Learning Nonlinear Neural Transfer Analytical Forms ». In : *ArXiv* abs/2101.09948 (2021) (cf. p. 58).

[260] Devis Tuia, Claudio Persello et Lorenzo Bruzzone. « Domain adaptation for the classification of remote sensing data : An overview of recent advances ». In : *IEEE geoscience and remote sensing magazine* 4.2 (2016), p. 41-57 (cf. p. 58).

[261] Wouter M. Kouw et Marco Loog. « A Review of Domain Adaptation without Target Labels ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.3 (2021), p. 766-785. DOI : 10.1109/TPAMI.2019.2945942 (cf. p. 58).

[262] Yu Shi, Lan Du et Yuchen Guo. « Unsupervised domain adaptation for SAR target detection ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), p. 6372-6385 (cf. p. 58).

[263] Zhongling Huang et al. « Classification of Large-Scale High-Resolution SAR Images With Deep Transfer Learning ». en. In : *IEEE Geoscience and Remote Sensing Letters* 18.1 (jan. 2021), p. 107-111. DOI : 10.1109/lgrs.2020.2965558. (Visité le 21/04/2022) (cf. p. 58).

[264] Jie Geng et al. « Transfer Learning for SAR Image Classification Via Deep Joint Distribution Adaptation Networks ». en. In : *IEEE Transactions on Geoscience and Remote Sensing* 58.8 (août 2020), p. 5377-5392. DOI : 10.1109/tgrs.2020.2964679. (Visité le 21/04/2022) (cf. p. 58).

[265] Zhongling Huang, Zongxu Pan et Bin Lei. « What, Where and How to Transfer in SAR Target Recognition Based on Deep CNNs ». en. In : *IEEE Transactions on Geoscience and Remote Sensing* 58.4 (avr. 2020). arXiv : 1906.01379, p. 2324-2336. DOI : 10.1109/tgrs.2019.2947634. (Visité le 21/04/2022) (cf. p. 58).

[266] Jie Geng et al. « SAR Image Classification via Deep Recurrent Encoding Neural Networks ». en. In : *IEEE Transactions on Geoscience and Remote Sensing* 56.4 (avr. 2018), p. 2255-2269. DOI : 10.1109/tgrs.2017.2777868. (Visité le 12/04/2022) (cf. p. 58).

[267] Youngkyu Hong et Eunho Yang. « Unbiased Classification through Bias-Contrastive and Bias-Balanced Learning ». In : *Advances in Neural Information Processing Systems*. Sous la dir. de M. Ranzato et al. T. 34. Curran Associates, Inc., 2021, p. 26449-26461 (cf. p. 58).

[268] Chenchen Wang, Hong Gu et Weimin Su. « SAR Image Classification Using Contrastive Learning and Pseudo-Labels With Limited Data ». en. In : *IEEE Geoscience and Remote Sensing Letters* 19 (2022), p. 1-5. DOI : 10.1109/lgrs.2021.3069224. (Visité le 12/04/2022) (cf. p. 58).

[269] Qi Wang et al. « A Comprehensive Survey of Loss Functions in Machine Learning ». In : *Annals of Data Science* 9 (2020), p. 187-212 (cf. p. 59).

[270] Shiliang Sun et al. « A Survey of Optimization Methods From a Machine Learning Perspective ». In : *IEEE Transactions on Cybernetics* 50.8 (2020), p. 3668-3681. DOI : 10.1109/TCYB.2019.2950779 (cf. p. 60).

[271] Diederik P. Kingma et Jimmy Ba. « Adam : A Method for Stochastic Optimization ». In : *CoRR* abs/1412.6980 (2014) (cf. p. 60).

[272] Mukund Prasad Sah et Amritpal Singh. « Aggregation Techniques in Federated Learning : Comprehensive Survey, Challenges and Opportunities ». In : *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. 2022, p. 1962-1967. DOI : 10.1109/icacite53722.2022.9823702 (cf. p. 61).

[273] A. Opaliński et al. « Medical data exploration based on the heterogeneous data sources aggregation system ». In : *2019 Federated Conference on Computer Science and Information Systems (FedCSIS)*. 2019, p. 591-597. DOI : 10.15439/2019F258 (cf. p. 61).

[274] Jie Dong, Wufan Zhao et Shuai Wang. « Multiscale Context Aggregation Network for Building Change Detection Using High Resolution Remote Sensing Images ». In : *IEEE Geoscience and Remote Sensing Letters* 19 (2022), p. 1-5. DOI : 10.1109/LGRS.2021.3121094 (cf. p. 61).

[275] G. R. Holden M. D. Nelson R. E. McRoberts et M. E. Bauer. « Effects of satellite image spatial aggregation and resolution on estimates of forest land area ». In : *International Journal of Remote Sensing* 30.8 (2009), p. 1913-1940. DOI : 10.1080/01431160802545631 (cf. p. 61).

[276] Asimina Mertzani et al. « Expertise, Social Influence, and Knowledge Aggregation in Distributed Information Processing ». In : *Artificial Life* 29.1 (2023), p. 37-65. DOI : 10.1162/artl_a_00387 (cf. p. 61).

[277] Michael Böhlen, Johann Gamper et Christian S. Jensen. « Multi-dimensional Aggregation for Temporal Data ». In : *Advances in Database Technology - EDBT 2006*. Springer Berlin Heidelberg, 2006, p. 257-275 (cf. p. 61).

[278] Antoni Morell et al. « Data Aggregation and Principal Component Analysis in WSNs ». In : *IEEE Transactions on Wireless Communications* 15.6 (2016), p. 3908-3919. DOI : 10.1109/twc.2016.2531041 (cf. p. 61).

[279] Y. Lu, Z. Zhang et D. Huang. « Glacier mapping based on Random Forest algorithm : A case study over the Eastern Pamir ». en. In : *Water (Basel)* 12.11 (nov. 2020), p. 3231 (cf. p. 61).

[280] Andrea Campagner, Davide Ciucci et Federico Cabitza. « Aggregation models in ensemble learning : A large-scale comparison ». In : *Information Fusion* 90 (fév. 2023), p. 241-252. DOI : https://doi.org/10.1016/j.inffus.2022.09.015. (Visité le 29/01/2024) (cf. p. 61).

[281] Maria Skublewska-Paszkowska et al. « Aggregation of Tennis Groundstrokes on the Basis of the Choquet Integral and Its Generalizations ». In : *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. 2022, p. 1-8. DOI : 10.1109/fuzz-ieee55066.2022.9882592 (cf. p. 61).

[282] S.H. Shabbeer Basha et al. « Impact of fully connected layers on performance of convolutional neural networks for image classification ». In : *Neurocomputing* 378 (fév. 2020), p. 112-119. DOI : https://doi.org/10.1016/j.neucom.2019.10.008. (Visité le 06/02/2024) (cf. p. 61).

[283] Vagan Terziyan et al. « Hyper-flexible Convolutional Neural Networks based on Generalized Lehmer and Power Means ». In : *Neural Networks* 155 (2022), p. 177-203. DOI : https://doi.org/10.1016/j.neunet.2022.08.017 (cf. p. 61).

[284] Brian McFee, Justin Salamon et Juan Pablo Bello. « Adaptive Pooling Operators for Weakly Labeled Sound Event Detection ». In : *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26.11 (2018), p. 2180-2193. DOI : 10.1109/taslp.2018.2858559 (cf. p. 61).

[285] Chen-Yu Lee, Patrick Gallagher et Zhuowen Tu. « Generalizing Pooling Functions in CNNs : Mixed, Gated, and Tree ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40.4 (2018), p. 863-875. DOI : 10.1109/tpami.2017.2703082 (cf. p. 61).

[286] W. Pedrycz. « Fuzzy neural networks with reference neurons as pattern classifiers ». In : *IEEE Trans. Neural Netw.* 3.5 (sept. 1992), p. 770-775. DOI : 10.1109/72.159065. (Visité le 25/02/2023) (cf. p. 61, 62).

[287] Luis M. de Campos et M. Jorge. « Characterization and comparison of sugeno and choquet integrals ». In : *Fuzzy Sets and Systems* 52.1 (nov. 1992), p. 61-67. DOI : 10.1016/0165-0114(92)90037-5. (Visité le 25/02/2023) (cf. p. 61, 62, 101).

[288] Eyke Hüllermeier. « Hüllermeier, E. : Fuzzy-Methods in Machine Learning and Data Mining : Status and Prospects. Fuzzy Sets and Systems 156(3), 387-407 ». In : *Fuzzy Sets and Systems* 156 (déc. 2005), p. 387-406. DOI : 10.1016/j.fss.2005.05.036 (cf. p. 61, 62).

[289] Navneet Walia, Harsukhpreet Singh et Anurag Sharma. « ANFIS : Adaptive Neuro-Fuzzy Inference System- A Survey ». In : *International Journal of Computer Applications* 123.13 (août 2015), p. 32-38. DOI : 10.5120/ijca2015905635. (Visité le 06/02/2024) (cf. p. 61, 62).

[290] Ezgi Türkarslan et Vicenç Torra. « Measure Identification for the Choquet Integral : A Python Module ». In : *International Journal of Computational Intelligence Systems* 15 (2022) (cf. p. 61, 62).

[291] Gleb Beliakov et Simon James. « Choquet Integral Optimisation with Constraints and the Buoyancy Property for Fuzzy Measures ». In : *Information Sciences* 578 (nov. 2021), p. 22-36. DOI : 10.1016/j.ins.2021.07.032. (Visité le 29/01/2024) (cf. p. 61, 62).

[292] Gleb Beliakov, Simon James et Gang Li. « Learning Choquet-Integral-Based Metrics for Semisupervised Clustering ». In : *IEEE Transactions on Fuzzy Systems* 19.3 (juin 2011), p. 562-574. DOI : 10.1109/tfuzz.2011.2123899. (Visité le 06/02/2024) (cf. p. 61, 62).

[293] Xi-Zhao Wang et al. « Particle swarm optimization for determining fuzzy measures from data ». In : *Inf. Sci.* 181 (oct. 2011), p. 4230-4252. DOI : 10.1016/j.ins.2011.06.002 (cf. p. 61, 62).

[294] Muhammad Aminul Islam et al. « Data-Driven Compression and Efficient Learning of the Choquet Integral ». In : *IEEE Transactions on Fuzzy Systems* 26.4 (août 2018), p. 1908-1922. DOI : 10.1109/tfuzz.2017.2755002. (Visité le 06/02/2024) (cf. p. 62).

[295] Giancarlo Lucca et al. « Preaggregation Functions : Construction and an Application ». In : *IEEE Trans. Fuzzy Syst.* 24.2 (avr. 2016), p. 260-272. DOI : 10.1109/tfuzz.2015.2453020. (Visité le 22/02/2023) (cf. p. 62, 102, 103).

[296] Cedric Marco-Detchart et al. « Neuro-inspired edge feature fusion using Choquet integrals ». In : *Information Sciences* 581 (déc. 2021), p. 740-754. DOI : 10.1016/j.ins.2021.10.016. (Visité le 22/02/2023) (cf. p. 62).

[297] Claudi Alsina, B. Schweizer et Maurice J. Frank. *Associative functions : triangular norms and copulas*. OCLC : ocm68045908. Singapore ; Hackensack, NJ : World Scientific, 2006. 237 p. (cf. p. 62, 103, 104).

[298] Xiaoxiao Du et Alina Zare. « Multiple Instance Choquet Integral Classifier Fusion and Regression for Remote Sensing Applications ». In : *IEEE Trans. Geosci. Remote Sensing* 57.5 (mai 2019), p. 2741-2753. DOI : 10.1109/tgrs.2018.2876687. (Visité le 22/02/2023) (cf. p. 62).

[299] F. Melgani, B.A.R. Al Hashemy et S.M.R. Taha. « An Explicit Fuzzy Supervised Classification Method for Multispectral Remote Sensing Images ». In : *IEEE Transactions on Geoscience and Remote Sensing* 38.1 (jan. 2000), p. 287-295. DOI : 10.1109/36.823921. (Visité le 06/02/2024) (cf. p. 62).

[300] Serafeim Moustakidis et al. « SVM-Based Fuzzy Decision Trees for Classification of High Spatial Resolution Remote Sensing Images ». In : *IEEE Transactions on Geoscience and Remote Sensing* 50.1 (jan. 2012), p. 149-169. DOI : 10.1109/tgrs.2011.2159726. (Visité le 06/02/2024) (cf. p. 62).

[301] L. Torres et al. « Chapter 12 - Learning OWA filters parameters for SAR imagery with multiple polarizations ». In : *Bio-Inspired Computation and Applications in Image Processing*. Academic Press, 2016, p. 269-284. DOI : https://doi.org/10.1016/B978-0-12-804536-7.00012-0 (cf. p. 62).

[302] J. Chanussot, G. Mauris et P. Lambert. « Fuzzy fusion techniques for linear features detection in multitemporal SAR images ». In : *IEEE Transactions on Geoscience and Remote Sensing* 37.3 (1999), p. 1292-1305. DOI : 10.1109/36.763290 (cf. p. 62).

[303] P. Blonda et al. « Combination of multiple classifiers by fuzzy integrals : an application to synthetic aperture radar (SAR) data ». In : *10th IEEE International Conference on Fuzzy Systems. (Cat. No.01CH37297)*. 10th Annual IEEE Conference on Fuzzy Systems. T. 3. Melbourne, Vic., Australia : Ieee, 2001, p. 944-947. DOI : 10.1109/fuzz.2001.1009112. (Visité le 25/02/2023) (cf. p. 62, 63).

[304] Qian Sun et al. « Ship Detection in SAR Images Based on Multilevel Superpixel Segmentation and Fuzzy Fusion ». In : *IEEE Transactions on Geoscience and Remote Sensing* 61 (2023), p. 1-15. DOI : 10.1109/tgrs.2023.3266373. (Visité le 06/02/2024) (cf. p. 62, 63).

[305] R. Lalchhanhima, Debdatta Kandar et R. Chawngsangpuii. « Single Look SAR Image Segmentation Using Local Entropy, Median Low-Pass Filter and Fuzzy Inference System ». In : *Machine Learning, Image Processing, Network Security and Data Sciences*. T. 1240. Communications in Computer and Information Science. Springer, 2020, p. 149-159. DOI : 10.1007/978-981-15-6315-7_12. (Visité le 12/02/2024) (cf. p. 62, 63).

[306] B. Lounis et al. « Adapting Neuro Fuzzy System for Oil Spills Detection From Sea Sar Images ». In : *2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*. Ieee, mars 2020, p. 184-187. DOI : 10.1109/m2garss47143.2020.9105243. (Visité le 06/02/2024) (cf. p. 62, 63).

[307] Ron Kohavi. « A study of cross-validation and bootstrap for accuracy estimation and model selection ». In : *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*. IJCAI'95. Montreal, Quebec, Canada : Morgan Kaufmann Publishers Inc., 1995, p. 1137-1143 (cf. p. 62, 63).

[308] Konstantinos Sechidis, Grigorios Tsoumakas et Ioannis Vlahavas. « On the stratification of multi-label data ». In : *Machine Learning and Knowledge Discovery in Databases : European Conference, ECML PKDD 2011, Athens, Greece, September 5-9, 2011, Proceedings, Part III 22*. Springer. 2011, p. 145-158 (cf. p. 63).

[309] Alberto Fernández et al. « SMOTE for learning from imbalanced data : progress and challenges, marking the 15-year anniversary ». In : *J. Artif. Int. Res.* 61.1 (2018), p. 863-905 (cf. p. 63).

[310] J Richard Landis et Gary G. Koch. « The measurement of observer agreement for categorical data. » In : *Biometrics* 33 1 (1977), p. 159-74 (cf. p. 64).

[311] Cornelius Senf, Patrick Hostert et Sebastian van der Linden. « Using MODIS time series and random forests classification for mapping land use in South-East Asia ». In : *2012 IEEE International Geoscience and Remote Sensing Symposium*. Ieee. 2012, p. 6733-6736 (cf. p. 65).

[312] Jingxiao Zhang et al. « Glacier Facies Mapping Using a Machine-Learning Algorithm : The Parlung Zangbo Basin Case Study ». In : *Remote Sensing* 11.4 (2019). doi : 10.3390/rs11040452 (cf. p. 65).

[313] Aftab Ahmed Khan et al. « Machine-Learning Algorithms for Mapping Debris-Covered Glaciers : The Hunza Basin Case Study ». In : *IEEE Access* 8 (2020), p. 12725-12734. doi : 10.1109/access.2020.2965768 (cf. p. 65).

[314] Haireti Alifu et al. « Machine-learning classification of debris-covered glaciers using a combination of Sentinel-1/-2 (SAR/optical), Landsat 8 (thermal) and digital elevation data ». In : *Geomorphology* 369 (2020), p. 107365. doi : https://doi.org/10.1016/j.geomorph.2020.107365 (cf. p. 65).

[315] Shrinidhi Ambinakudige et Adjoa Intsiful. « Estimation of area and volume change in the glaciers of the Columbia Icefield, Canada using machine learning algorithms and Landsat images ». In : *Remote Sensing Applications : Society and Environment* 26 (2022), p. 100732. doi : https://doi.org/10.1016/j.rsase.2022.100732 (cf. p. 65).

[316] Mohd Anul Haq, Mohd Farooq Azam et Christian Vincent. « Efficiency of artificial neural networks for glacier ice-thickness estimation : a case study in western Himalaya, India ». In : *Journal of Glaciology* 67.264 (2021), p. 671-684. doi : 10.1017/jog.2021.19 (cf. p. 65).

[317] Gianluca Palermo et al. « Using Artificial Neural Networks to Couple Satellite C-Band Synthetic Aperture Radar Interferometry and Alpine3D Numerical Model for the Estimation of Snow Cover Extent, Height, and Density ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 (2023), p. 2868-2888. doi : 10.1109/jstars.2023.3253804. (Visité le 29/01/2024) (cf. p. 65).

[318] Rahul Nijhawan, Josodhir Das et Balasubramanian Raman. « A hybrid of deep learning and hand-crafted features based approach for snow cover mapping ». en. In : *International Journal of Remote Sensing* 40.2 (17 jan. 2019). Number : 2, p. 759-773. doi : 10.1080/01431161.2018.1519277. (Visité le 09/11/2021) (cf. p. 65).

[319] Yifang Ban et Osama A. Yousif. « Multitemporal Spaceborne SAR Data for Urban Change Detection in China ». In : *Ieee Jstars* 5.4 (août 2012), p. 1087-1094. doi : 10.1109/jstars.2012.2201135. (Visité le 13/02/2023) (cf. p. 65).

[320] B. Ygorra et al. « Deforestation Monitoring Using Sentinel-1 SAR Images in Humid Tropical Areas ». In : *Ieee Igarss*. 2021, p. 5957-5960. doi : 10.1109/igarss47720.2021.9554698 (cf. p. 65, 66).

[321] Cristian Silva-Perez et al. « Multitemporal Polarimetric SAR Change Detection for Crop Monitoring and Crop Type Classification ». In : *Ieee Jstars* 14 (2021), p. 12361-12374. doi : 10.1109/jstars.2021.3130186. (Visité le 13/02/2023) (cf. p. 65, 66).

[322] Feng Gao et al. « Sea Ice Change Detection in SAR Images Based on Convolutional-Wavelet Neural Networks ». In : *Ieee Grsl* 16.8 (août 2019), p. 1240-1244. doi : 10.1109/lgrs.2019.2895656. (Visité le 13/02/2023) (cf. p. 65, 66).

[323] S. H. Winsvold et al. « Using SAR satellite data time series for regional glacier mapping ». In : *The Cryosphere* 12.3 (2018), p. 867-890. doi : 10.5194/tc-12-867-2018 (cf. p. 65, 66).

[324] Karunesh K. Gupta et Rajiv Gupta. « Despeckle and geographical feature extraction in SAR images by wavelet transform ». In : *ISPRS Journal of Photogrammetry and Remote Sensing* 62.6 (2007), p. 473-484. doi : https://doi.org/10.1016/j.isprsjprs.2007.06.001 (cf. p. 65, 66).

[325] Abdourrahmane Mahamane Atto et al. « Wavelet Operators and Multiplicative Observation Models–Application to SAR Image Time-Series Analysis ». In : *Ieee Tgrs* 54.11 (nov. 2016), p. 6606-6624. doi : 10.1109/tgrs.2016.2587626. (Visité le 13/02/2023) (cf. p. 65, 66).

[326] Turgay Celik et Kai-Kuang Ma. « Unsupervised Change Detection for Satellite Images Using Dual-Tree Complex Wavelet Transform ». In : *Ieee Tgrs* 48.3 (mars 2010), p. 1199-1210. doi : 10.1109/tgrs.2009.2029095. (Visité le 13/02/2023) (cf. p. 65, 66).

[327] Lu Jia et al. « SAR Image Change Detection Based on Multiple Kernel K-Means Clustering With Local-Neighborhood Information ». In : *Ieee Grsl* 13.6 (juin 2016), p. 856-860. DOI : 10.1109/lgrs.2016.2550666. (Visité le 13/02/2023) (cf. p. 65, 66).

[328] Gabriele Moser, Elena Angiati et Sebastiano B. Serpico. « Multiscale Unsupervised Change Detection on Optical Images by Markov Random Fields and Wavelets ». In : *Ieee Grsl* 8.4 (juill. 2011), p. 725-729. DOI : 10.1109/lgrs.2010.2102333. (Visité le 13/02/2023) (cf. p. 66).

[329] Xiao Xiao et al. « Temporal Series Crop Classification Study in Rural China Based on Sentinel-1 SAR Data ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), p. 2769-2780. DOI : 10.1109/jstars.2021.3057903. (Visité le 29/01/2024) (cf. p. 66).

[330] Hiroaki Sakoe et Seibi Chiba. « A Dynamic Programming Approach to Continuous Speech Recognition ». In : 1971 (cf. p. 66).

[331] Yangqi Li et Taihua Hu. « Dynamic Time Warping Application for Financial Pattern Recognition ». In : *SSRN Electronic Journal* (2020). DOI : 10.2139/ssrn.3658339 (cf. p. 66).

[332] K. Hebbrecht et al. « Understanding personalized dynamics to inform precision medicine : a dynamic time warp analysis of 255 depressed inpatients ». In : *BMC Medicine* 18.1 (déc. 2020). DOI : 10.1186/s12916-020-01867-5 (cf. p. 66).

[333] Romain Tavenard. *An introduction to Dynamic Time Warping*. https://rtavenar.github.io/blog/dtw.html. 2021 (cf. p. 66, 67).

[334] H. Sakoe et S. Chiba. « Dynamic programming algorithm optimization for spoken word recognition ». In : *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26.1 (fév. 1978), p. 43-49. DOI : 10.1109/tassp.1978.1163055 (cf. p. 66, 67).

[335] François Petitjean, Alain Ketterlin et Pierre Gançarski. « A global averaging method for dynamic time warping, with applications to clustering ». en. In : *Pattern Recognition* 44.3 (mars 2011), p. 678-693. DOI : 10.1016/j.patcog.2010.09.013. (Visité le 05/01/2022) (cf. p. 67).

[336] Xingyu Cai et al. « DTWNet : a Dynamic Time Warping Network ». In : 32 (2019). Sous la dir. de H. Wallach et al. (cf. p. 67).

[337] Matthieu Gallet et al. *LSD4WSD : An Open Dataset for Wet Snow Detection with SAR Data and Physical Labelling*. Version 2.0.0. Zenodo, oct. 2023. DOI : 10.5281/zenodo.10046730 (cf. p. 67, 76).

[338] Matthieu Gallet et al. *LSD4WSD : An Open Dataset for Wet Snow Detection with SAR Data and Physical Labelling*. Version 1.0.0. Zenodo, juill. 2023. DOI : 10.5281/zenodo.8111485 (cf. p. 67).

[339] Stéphanie Bonnet. « Un modèle numérique de canopée pour l'estimation de la hauteur dominante des peuplements résineux en Région wallonne ». In : *Forêt Wallonne* 98 (jan. 2009), p. 53-59 (cf. p. 69).

[340] Ronny Hänsch. « The Trap Of Random Sampling and How to Avoid It - Alternative Sampling Strategies for a Realistic Estimate of the Generalization Error in Remote Sensing ». In : *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. Ieee, juill. 2021, p. 2020-2023. DOI : 10.1109/igarss47720.2021.9554961. (Visité le 29/01/2024) (cf. p. 73, 79, 141).

[341] Ana Paula Salcedo et Marisa Gloria Cogliati. « Snow Cover Area Estimation Using Radar and Optical Satellite Information ». In : *Applied Categorical Structures* 4 (2014), p. 514-523 (cf. p. 74).

[342] Météo France. « Hiver 2020-2021 : très arrosé avec une alternance de temps hivernal et de douceur printanière ». In : *Les bilans climatiques* (9 mars 2021) (cf. p. 75).

[343] O. Dick et al. « Can Saharan dust deposition impact snowpack stability in the French Alps ? » In : *The Cryosphere* 17.4 (2023), p. 1755-1773. DOI : 10.5194/tc-17-1755-2023 (cf. p. 75).

[344] Matthieu Gallet et al. *LSD4WSD : An Open Dataset for Wet Snow Detection with SAR Data and Physical Labelling*. Version 1.0.0. Zenodo, juill. 2023. DOI : 10.5281/zenodo.8111485 (cf. p. 75).

[345] Marie Gardent et al. « Multitemporal glacier inventory of the French Alps from the late 1960s to the late 2000s ». In : *Global and Planetary Change* 120 (2014), p. 24-37. DOI : https://doi.org/10.1016/j.gloplacha.2014.05.004 (cf. p. 77).

[346] J. F. von Raumer et al. « The Pre-Alpine Evolution of the External Massifs ». In : *Pre-Mesozoic Geology in the Alps*. Springer Berlin Heidelberg, 1993, p. 221-240. doi : 10.1007/978-3-642-84640-3_13 (cf. p. 77).

[347] F. Magnin et al. « Modelling rock wall permafrost degradation in the Mont Blanc massif from the LIA to the end of the 21st century ». In : *The Cryosphere* 11.4 (2017), p. 1813-1834. doi : 10.5194/tc-11-1813-2017 (cf. p. 77).

[348] Matthieu Gallet et al. « CPAZMAL : Cryosphere PAZ satellite MAchine Learning ». Version 1.0.0. In : (2023). doi : 10.5281/zenodo.10401360 (cf. p. 79).

[349] Dengxin Dai, Wen Yang et Hong Sun. « Multilevel Local Pattern Histogram for SAR Image Classification ». In : *IEEE Geoscience and Remote Sensing Letters* 8.2 (2011), p. 225-229. doi : 10.1109/lgrs.2010.2058997 (cf. p. 83).

[350] Koray Kayabol. « Histogram-Based Contextual Classification of SAR Images ». In : *IEEE Geoscience and Remote Sensing Letters* 12.1 (2015), p. 33-37. doi : 10.1109/lgrs.2014.2325220 (cf. p. 83).

[351] Sylvain Arlot et Alain Celisse. « A survey of cross-validation procedures for model selection ». In : *Statistics Surveys* 4.none (jan. 2010), p. 40-79. doi : 10.1214/09-ss054. (Visité le 29/01/2024) (cf. p. 84).

[352] Matthieu Gallet et al. « CNN Classification of Wet Snow By Physical Snowpack Model Labeling ». In : *International Geoscience and Remote Sensing Symposium (IGARSS 2023)*. Juill. 2023 (cf. p. 86).

[353] Thomas Jagdhuber et al. « Identification of Soil Freezing and Thawing States Using SAR Polarimetry at C-Band ». In : *Remote Sensing* 6.3 (2014), p. 2008-2023. doi : 10.3390/rs6032008 (cf. p. 97).

[354] Nicolas Baghdadi et al. « Detection of Frozen Soil Using Sentinel-1 SAR Data ». In : *Remote Sensing* 10.8 (2018). doi : 10.3390/rs10081182 (cf. p. 97).

[355] Ronald R. Yager. « On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decisionmaking ». In : *Readings in Fuzzy Sets for Intelligent Systems*. Morgan Kaufmann, 1993, p. 80-87. doi : https://doi.org/10.1016/B978-1-4832-1450-4.50011-0 (cf. p. 102).

[356] Edurne Barrenechea Tartas et al. « Using the Choquet Integral in the Fuzzy Reasoning Method of Fuzzy Rule-Based Classification Systems ». In : *Axioms* 2 (2013), p. 208-223 (cf. p. 102).

[357] L. Breiman et al. « Classification and Regression Trees ». In : *Biometrics* 40 (1984), p. 874 (cf. p. 109).

[358] Jerome H. Friedman. « Stochastic gradient boosting ». In : *Computational Statistics & Data Analysis* 38 (2002), p. 367-378 (cf. p. 109).

[359] Benjamin Guedj. *A Primer on PAC-Bayesian Learning*. 2019 (cf. p. 117, 143).

[360] Valentina Zantedeschi et al. « Learning Stochastic Majority Votes by Minimizing a PAC-Bayes Generalization Bound ». In : *Advances in Neural Information Processing Systems*. Sous la dir. de M. Ranzato et al. T. 34. Curran Associates, Inc., 2021, p. 455-467 (cf. p. 117).

[361] M.C. Dobson et al. « Use of SAR image texture in terrain classification ». In : *IGARSS'97. 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing - A Scientific Vision for Sustainable Development*. T. 3. 1997, 1180-1183 vol.3. doi : 10.1109/igarss.1997.606390 (cf. p. 120).

[362] *Copula Theory and Its Applications : Proceedings of the Workshop Held in Warsaw, 25-26 September 2009*. T. 198. Lecture Notes in Statistics. Springer Berlin Heidelberg, 2010. (Visité le 29/01/2024) (cf. p. 123).

[363] B. Kosztra et al. *Updated CLC Illustrated Nomenclature Guidelines*. Final Report. European Environment Agency, 2019 (cf. p. 125).

[364] Jean-Marie Nicolas. « Application de la transformée de Mellin : étude des lois statistiques de l'imagerie cohérente ». In : (oct. 2018) (cf. p. 128).

[365] Manuel Gil, Fady Alajaji et Tamas Linder. « Rényi divergence measures for commonly used univariate continuous distributions ». In : *Information Sciences* 249 (nov. 2013), p. 124-131. doi : 10.1016/j.ins.2013.06.018. (Visité le 05/09/2023) (cf. p. 129).

[366] Yingxiang Yang et al. « Learning Positive Functions with Pseudo Mirror Descent ». In : *Advances in Neural Information Processing Systems*. Sous la dir. de H. Wallach et al. T. 32. Curran Associates, Inc., 2019 (cf. p. 132).

[367] Christophe Lin-Kwong-Chon et al. « Supervised Classification For Analysis Of Cryospheric Zones Using SAR Statistical Timeseries ». In : *International Geoscience and Remote Sensing Symposium (IGARSS 2023)*. Athens, Greece, juill. 2024 (cf. p. 141).

[368] Guo Haixiang et al. « Learning from class-imbalanced data : Review of methods and applications ». In : *Expert Systems with Applications* 73 (2017), p. 220-239. doi : https://doi.org/10.1016/j.eswa.2016.12.035 (cf. p. 141).

[369] Nicolas Courty et al. « Optimal transport for domain adaptation ». In : *IEEE transactions on pattern analysis and machine intelligence* 39.9 (2016), p. 1853-1865 (cf. p. 141).