Synthèse des prévisions d'ensemble par scénarios physiquement cohérents : mise en pratique au moyen d'approches objets et d'apprentissage statistique

Arnaud Mounier

▶ **To cite this version:**

Arnaud Mounier. Synthèse des prévisions d'ensemble par scénarios physiquement cohérents : mise en pratique au moyen d'approches objets et d'apprentissage statistique. Sciences de la Terre. Institut National Polytechnique de Toulouse - INPT, 2022. Français. NNT : 2022INPT0070 . tel-04247716v2

HAL Id: tel-04247716

https://theses.hal.science/tel-04247716v2

Submitted on 18 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Océan, atmosphère, climat

Présentée et soutenue par :

M. ARNAUD MOUNIER

le lundi 12 décembre 2022

Titre :

Synthèse des prévisions d'ensemble par scénarios physiquement cohérents. Mise en pratique au moyen d'approches objets et d'apprentissage statistique.

Ecole doctorale :

Sciences de l'Univers de l'Environnement et de l'Espace (SDU2E)

Unité de recherche :

Centre national de recherches météorologiques (CNRM METEO)

Directeur(s) de Thèse :

M. MATTHIEU PLU

MME LUCIE ROTTNER

Rapporteurs :

MME CECILE MALLET, UNIVERSITE DE VERSAILLES

M. STEPHANE VANNITSEM, Institut Royal Météorologique Belgique

Membre(s) du jury :

MME BÉATRICE LAURENT-BONNEAU, INSA TOULOUSE, Président

M. MATTHIEU PLU, METEO FRANCE TOULOUSE, Membre

MME LAURE RAYNAUD, METEO FRANCE TOULOUSE, Invité(e)

MME LUCIE ROTTNER, METEO FRANCE TOULOUSE, Membre

M. RONAN FABLET, IMT ATLANTIQUE, Membre

Remerciements

Je souhaite commencer par remercier mes encadrants durant ces trois ans de thèse, Matthieu Plu, Lucie Rottner et Laure Raynaud. Cet encadrement collégial m'a permis de me familiariser progressivement au monde de la Recherche et ce malgré un environnement souvent compliqué à cause des années COVID. Ils m'ont en particulier aidé à trouver le bon équilibre entre recherche bibliographique, résultats, valorisations et interactions avec les prévisionnistes de Météo-France, indispensables durant cette thèse. Je souhaite remercier aussi Cécile Mallet et Stéphane Vannitsem qui ont lu ce (long) manuscrit en entier et ont rapporté cette thèse. Leurs regards extérieurs ainsi que ceux de Ronan Fablet et Béatrice Laurent-Bonneau, autres membres de ce jury de soutenance, ont permis des échanges fructueux et une amélioration du manuscrit existant.

Si ce sujet de thèse existe, c'est en grande partie grâce à Philippe Arbogast qui a encadré mon stage de fin d'étude dans les services de prévision, qui a participé à la genèse de ce sujet et qui a suivi les avancés des travaux pendant ces trois ans. Je le remercie donc grandement pour m'avoir orienté vers cette thématique passionnante qui attrait à l'exploitation la plus pertinente possible des prévisions d'ensemble par des méthodes innovantes. Je remercie aussi la directrice des études de l'ENM Isabelle Beau qui m'a orienté à bon escient dans cette voie mobilisant à la fois des connaissances en statistiques et en prévision météorologique, comme je le souhaitais.

Cette thèse est comme je l'indiquais plus haut le fruit d'une collaboration étroite avec les services de prévision. Ces travaux n'auraient pas été possibles sans la participation de services extérieurs au CNRM. Je peux citer en particulier Fabrice Guillemot qui a participé aux discussions orientant les travaux vers la détection des échos arqués. Je pense aussi à Michaël Kreitz qui a suivi mes travaux et m'a apporté une aide précieuse grâce à ses connaissances en convection. Je n'oublie pas non plus d'autres collègues à Toulouse comme Benoît Touzé, Karine Maynard ou encore Fabienne Dupont. L'implication des services en région a aussi été décisive commme Léo Mignan à Rennes ou Jean Imbert à Aix-en-Provence et bien d'autres avec qui j'ai eu la chance d'échanger à de nombreuses reprises sur mes résultats.

Les échanges avec les services de prévision étaient complémentaires avec ceux tout aussi importants en interne dans mon équipe de recherche, PREVISIBILITE. Je tiens ici à saluer Thomas Rieutord et Clément Brochet avec qui certains échanges ont abouti à de vrais virages méthodologiques concernant la synthèse par scénarios. Je n'oublie pas non plus la contribution de mes anciens co-bureaux, Meryl Wimmer qui m'a fait redécouvrir la météorologie dynamique et permis de mieux comprendre les prévisions d'ensemble, ou Bruno Joly avec qui je vais pouvoir

collaborer à partir de 2023 pour étendre les résultats de cette thèse à de nouveaux domaines météo-sensibles. Je tiens également à remercier Sylvain Galliau que j'ai eu la chance d'encadrer pour son stage de fin d'étude. Son volontarisme et ses travaux ont permis des avancées notables dans un contexte pourtant peu évident à cause à l'époque d'un télétravail encore massif et imposé.

Cette thèse a aussi été l'occasion pour moi de découvrir l'enseignement. Je remercie donc Pascal Laveau de m'avoir fait confiance pendant 2 ans pour assurer certains TPs en statistiques à l'ENM.

Enfin, je souhaite mentionner ici tous ces moments conviviaux passés avec l'équipe PREVISIBILITE (je pense ici à Carole, Laurent, Alexis, Gabriel, Jeremy, Léna, Grégory, Stéphane et bien d'autres). Je dois aussi mentionner les pauses cafés, de façon quasi-quotidienne, à la DIROP (je ne peux pas oublier Ronan, Isabelle, Michel, Marie ou encore Jessica). Ces moments ont été pour moi essentiels pour maintenir l'équilibre avec le travail intense que représente une thèse. Un autre moment de détente important pour le passionné de prévision que je suis réside dans ma participation au concours de prévision. Je souhaite donc finir en mentionnant l'équipe Team RADAR du concours de prévision, les autres membres se reconnaitront, avec qui j'espère continuer à jouer les premières places dans les années à venir.

Résumé

La prévision météorologique, basée sur la modélisation du système chaotique atmosphérique, comporte une part d'incertitude. Celle-ci peut être évaluée à l'aide d'ensembles de prévisions. La thèse s'intéresse au post-traitement de la prévision d'ensemble à échelle kilométrique de Météo-France (PE-AROME). L'objectif est de développer de nouveaux outils pour extraire des informations essentielles des modèles pour les prévisions. Ces outils s'appuient sur des méthodes d'apprentissage profond. Ce manuscrit commence par la présentation des prévisions d'ensemble, des méthodes de post-traitement existantes et des méthodes employées aujourd'hui pour expertiser les modèles météorologiques. Cette thèse propose d'enrichir la palette d'outils disponibles avec deux approches complémentaires pour extraire de nouvelles informations de la PE-AROME.

Dans un premier temps, une approche pour aider à la prévision de structures orageuses particulières et à fort impact, qu'on nomme les échos arqués, est développée. L'objectif est de les détecter dans les sorties de chaque membre PE-AROME. Cette détection est faite par un réseau de neurones convolutif (de type U-Net). Le réseau de neurones retenu montre une bonne capacité à discerner les échos arqués d'autres organisations orageuses. Les performances de ce U-Net ont permis de déboucher sur une production en mode recherche quotidienne synthétisant le risque d'échos arqués. L'évaluation objective des prévisions montre la plus-value de la PE-AROME comparée au modèle déterministe AROME au-delà de 15 heures d'échéance.

La deuxième approche développée durant cette thèse vise à mettre en œuvre une synthèse par scénarios des prévisions de pluies. Un scénario peut être défini comme un groupe de membres prévoyant des évènements météorologiques similaires. Cette approche ne s'intéresse pas à un évènement particulier et est donc moins spécialisée que la détection des échos arqués. Cette synthèse repose sur une réduction de dimension du champ de pluies à l'aide d'un autre réseau de neurones convolutif (de type autoencodeur). Chaque membre est ainsi exprimé et classé dans l'espace latent de l'autoencodeur qui possède quelques dizaines de dimensions. La pertinence de cette représentation sous forme de scénarios est discutée à l'aide de scores et d'études de cas. Les scores montrent une réelle complémentarité entre le modèle déterministe AROME et la PE-AROME. Ils montrent qu'il est plus intéressant de suivre les deux scénarios les plus peuplés de la PE-AROME que deux réseaux AROME. Mais ils montrent aussi qu'en cas de deux scénarios équiprobables dans la PE-AROME, le scénario qui est aussi celui d'une prévision AROME a plus de chance d'être correct que l'autre.

Abstract

Weather forecasting, based on the modelling of atmospheric chaotic system, has a degree of uncertainty. This uncertainty can be assessed by an ensemble prediction system. This PhD-thesis focuses on the post-processing of the Météo-France convection-permitting ensemble prediction system (AROME-EPS). The aim is to develop new tools to extract a relevant information from models for forecasting. These tools are based on deep learning methods. This manuscript starts with the presentation of ensemble forecasts, existing post-processing methods and work methods of forecasters. In order to enhance the forecasting tools currently available, this thesis proposes two complementary approaches to extract new and relevant information from the AROME-EPS.

Firstly, an approach to deal with the prediction of a specific type of damaging thunderstorms, called bow echoes, is developed. The objective is to detect them in output of each AROME-EPS member. This detection is made by a convolutional neural network (inspired from the U-Net architecture). The selected neural network demonstrates a good ability to distinguish bow echoes from other convective systems. The skill of this U-Net has led to a daily research mode production summarising the risk of bow echoes. The objective evaluation of forecasts shows the added value of the AROME-EPS compared to the deterministic AROME model beyond 15 lead times.

The second approach developed during this thesis aims at implementing a rainfall scenario synthesis. A scenario can be defined as a group of members predicting similar weather events. Since this approach doesn't focus on a particular event, it can be useful on a daily basis to quickly examine ensemble forecasts. This synthesis is based on a dimension reduction of rainfall fields using another convolutional neural network (autoencoder). Each member is expressed and classified in the latent space of the autoencoder which has tens of dimensions. The relevance of this scenario approach is discussed using scores and case studies. The scores show complementarity between the deterministic AROME model and the AROME-EPS. They show it is more interesting to watch the two biggest scenarios in the PE-AROME than two AROME runs. But they also show in case of two equally probable scenarios in the PE-AROME, the scenario that is also the one of an AROME forecast is more likely correct than the other.

Table des matières

Remerciements i

Résumé iii

Abstract v

Introduction 1

I Prévision météorologique : méthodes et limites actuelles 5

1 Modélisation de l'atmosphère 7
1.1 Description de l'atmosphère 8
1.1.1 Quelques variables importantes 8
1.1.2 Les différentes échelles de l'atmosphère 11
1.2 Les grandes étapes de l'observation à la prévision 13
1.2.1 Observation de l'atmosphère 14
1.2.2 Assimilation de données 15
1.2.3 Modèle météorologique 16
1.3 Prévisibilité et modèles météorologiques : des avancées conjointes . 17
1.3.1 Les premiers modèles météorologiques (années 1950-1960) . . 17
1.3.2 Questionnements autour de la notion de prévisibilité (années 1950-1960) 18
1.3.3 L'amélioration des modèles grandes échelles (années 1970-2020) 19
1.3.4 La construction des prévisions ensemblistes (années 1970-2020) 20
1.3.5 L'arrivée en condition opérationnelle des modèles méso-échelle (années 2000-2020) 23
1.3.6 Prévisibilité de l'échelle convective 24
1.4 La Prévision d'Ensemble AROME (PE-AROME) 26
1.4.1 Méthodes de perturbation 26
1.4.2 Évolutions de la PE-AROME de 2016 à 2022 28

2 Méthodes de visualisation des systèmes ensemblistes 35
2.1 Synthétisation en un point . . . 37
2.2 Synthétisation spatialisée . . . 40
2.2.1 Représentation de l'incertitude . . . 41
2.2.2 Risque d'évènements extrêmes . . . 47
2.2.3 Regroupement de membres . . . 49
2.3 Synthétisation pour des événements particuliers . . . 55

3 Synthétiser les ensembles pour la prise de décision 61
3.1 Expertise du prévisionniste . . . 62
3.1.1 Les structures cohérentes . . . 62
3.1.2 La prise de décision en contexte incertain . . . 64
3.2 Approche objet . . . 66
3.2.1 Représentation des structures cohérentes . . . 66
3.2.2 Solution à la problématique de la double peine . . . 67
3.3 Échos arqués . . . 68
3.3.1 Caractéristiques générales . . . 69
3.3.2 Formation . . . 70
3.3.3 Outil de synthèse adapté . . . 72
3.4 Synthèse par scénarios . . . 74
3.4.1 Définition . . . 74
3.4.2 Méthode développée . . . 76

II Deep learning : des outils pertinents pour traiter les données atmosphériques 81

4 Deep learning : concept et applications 83
4.1 Définitions et principes . . . 83
4.1.1 Définitions des concepts d'IA, machine learning et deep learning . . . 83
4.1.2 Deep learning : principes de fonctionnement . . . 85
4.1.3 Phase d'apprentissage : notion de batch, d'epoch et de learning rate . . . 86
4.1.4 Base d'apprentissage, de validation et de test . . . 87
4.2 Couches d'un réseau de neurones et fonction coût . . . 88
4.2.1 Couches d'un réseau de neurones . . . 89
4.2.2 Fonction coût . . . 95
4.2.3 Récapitulatif des hyperparamètres . . . 96
4.3 Applications en météorologie . . . 97
4.3.1 Recherches en cours sur la chaîne observation, assimilation, prévision . . . 97
4.3.2 Limites et challenges liés aux méthodes de deep learning . . . 99

5 Détection des échos arqués avec un U-Net 105
5.1 Construction des jeux de données . . . 105
5.1.1 Labellisation des échos arqués . . . 105
5.1.2 Bases d'apprentissage et de validation . . . 107
5.1.3 Choix du prédicteur . . . 108

5.2 Architecture du réseau de neurones . . . 108
5.3 Entraînement du réseau de neurones . . . 109
5.3.1 Pré-traitement de la base d'apprentissage . . . 109
5.3.2 Mise en place de l'entraînement . . . 111
5.4 Prédiction sur la grille complète . . . 112
5.5 Scores orientés objets . . . 113
5.5.1 Attributs des objets échos arqués . . . 113
5.5.2 Scores . . . 114

6 Réduction de dimension avec un autoencodeur 117
6.1 Construction des jeux de données . . . 118
6.1.1 Base d'apprentissage . . . 118
6.1.2 Sélection par échantillonnage préférentiel (importance sampling) . . . 119
6.1.3 Base de validation pour les hyperparamètres . . . 121
6.2 Construction de l'autoencodeur . . . 122
6.2.1 Architecture . . . 122
6.2.2 Entraînement . . . 123
6.2.3 Récapitulatif des hyperparamètres . . . 123
6.3 Scores orientés objets . . . 123
6.3.1 Méthode de détection des objets . . . 124
6.3.2 Scores à partir des attributs . . . 125
6.4 Sélection des hyperparamètres . . . 127
6.4.1 Taille des batchs . . . 127
6.4.2 Importance sampling . . . 128
6.4.3 Espace latent à 10 dimensions . . . 129
6.4.4 Espace latent à 20 dimensions . . . 130
6.4.5 Espace latent à 40 dimensions . . . 131
6.4.6 Scores région par région . . . 132
6.5 Comparaison avec des méthodes plus classiques . . . 133
6.5.1 Principes généraux de l'ACP et des ondelettes . . . 133
6.5.2 Comparaison des méthodes . . . 136

III Prévision du risque d'échos arqués 139

7 Synthèse des détections dans PE-AROME et AROME 141
7.1 Résumé de l'article . . . 141
7.2 Article . . . 143
7.3 Supplément de l'article . . . 166

8 Performance des modèles 183
8.1 Détection des échos arqués dans les observations . . . 183
8.2 Scores objectifs PE-AROME et AROME . . . 185
8.2.1 Méthodologie . . . 185
8.2.2 Échéances courtes ($\leqslant$ 15h) . . . 187
8.2.3 Échéances longues ($>$ 15h) . . . 189
8.2.4 Étude des fausses alarmes . . . 190
8.3 Bilan de la saison 2021 . . . 190

8.4 Ajout d'un graphique rafale de vent . . . 192
8.4.1 Construction du graphique . . . 192
8.4.2 Exemple sur une situation . . . 193
8.5 Comparaison anciennes et nouvelles versions . . . 195

IV Construction des scénarios de pluie 199

9 Étude approfondie de l'espace latent 201
9.1 Étude comparative des attributs d'objets . . . 202
9.1.1 Caractéristiques générales . . . 202
9.1.2 Caractéristiques des non détections . . . 204
9.1.3 Caractéristiques des fausses alarmes . . . 206
9.2 Structures pluvieuses idéalisées . . . 206
9.2.1 Génération des structures . . . 207
9.2.2 Modification de la taille . . . 208
9.2.3 Modification de l'intensité . . . 208
9.2.4 Modification de la position . . . 208
9.3 Étude dimension par dimension . . . 213
9.3.1 Répartition selon les échéances . . . 215
9.3.2 Répartition selon les zones régionales . . . 215
9.3.3 Répartition selon les mois . . . 217
9.3.4 Champ moyen des pluies associées au Q1 ou au Q99 . . . 218
9.3.5 Champ de pluie associé aux extrêmes de la distribution . . . 219

10 Structures météorologiques récurrentes 221
10.1 Base de données . . . 221
10.2 Méthodes de classification . . . 222
10.2.1 Méthodes par partitionnement . . . 223
10.2.2 Méthodes hiérarchiques . . . 224
10.2.3 Méthodes par densité . . . 225
10.2.4 Méthodes par mélange gaussien . . . 226
10.2.5 Méthode de classification retenue . . . 226
10.3 Représentation synthétique des classes . . . 228
10.3.1 Répartition selon les mois et années . . . 228
10.3.2 Répartition selon les heures de validité . . . 230
10.3.3 Moyenne/Écart-type . . . 230
10.3.4 Champ de pluie le plus proche et le plus éloigné du centroïde 230
10.4 Sélection de la taille de l'espace latent . . . 231
10.4.1 Sensibilité de la classification à la taille de l'espace latent . . 231
10.4.2 Taille de l'espace latent retenue . . . 233
10.5 Sélection du nombre de classes . . . 236
10.5.1 Scores pour fixer le nombre de classes . . . 236
10.5.2 Nombre de scénarios différents . . . 237

11 Synthèse par scénarios : études de cas et scores 241
11.1 Représentation de la synthèse par scénarios . . . 242
11.1.1 Synthèse globale . . . 242
11.1.2 Représentation des scénarios . . . 245

11.2 Études de cas 246
11.2.1 14 septembre 2021 246
11.2.2 2 octobre 2020 251
11.2.3 3-4 octobre 2021 255
11.3 Évaluation de la méthode et des modèles 259
11.3.1 Répartition ANTILOPE parmi les 20 classes 260
11.3.2 Taille des scénarios 262
11.3.3 Intérêt du multi-réseaux 263
11.3.4 Fiabilité 263
11.3.5 Comment choisir un scénario de prévision ? 265
11.3.6 Statistiques sur les scénarios corrects 266
11.3.7 Comparaison anciennes et nouvelles versions 270

Conclusion **275**

Liste des figures **xvii**

Liste des tableaux **xx**

Liste des acronymes **xxiii**

Annexes **xxv**

A Synthétiser les ensembles pour la prise de décision **i**

B Réduction de dimension avec un autoencodeur **iii**

C Performance des modèles **vii**

D Étude approfondie de l'espace latent **xi**

E Structures météorologiques récurrentes **lxi**

Introduction

Prévoir le temps est un enjeu majeur pour la sécurité des personnes et des biens et pour de nombreux acteurs (institutionnels et économiques) comme par exemple l'aviation, l'énergie ou l'agriculture. Cette forte météo-sensibilité explique une exigence toujours croissante de la société envers ces prévisions pour qu'elles soient le plus juste possible. Des avancées majeures ont été faites en prévision depuis plusieurs décennies grâce à nos connaissances qui ont progressé, le développement de modèles plus performants, des moyens d'observations accrus et grâce à une puissance de calcul disponible toujours plus importante. Ces progrès ont permis de mettre en évidence le comportement chaotique de l'atmosphère et la notion de prévisibilité qui montre qu'une unique prévision ne permettra jamais de faire une prévision parfaite au-delà d'un certain nombre d'heures, de jours ou de semaines selon les échelles considérées. Ces découvertes ont motivé le développement de prévisions d'ensemble. Ces prévisions d'ensemble ne fournissent pas une unique prévision mais plusieurs qui permettent de mettre en évidence la prévisibilité d'une situation en explorant les possibles futurs états de l'atmosphère. Développer des prévisions d'ensemble nécessite de lourds efforts en modélisation, en puissance de calcul disponible et en archivage des données. Cette masse de données est alors mise à disposition des prévisionnistes pour qu'ils établissent leurs prévisions. Mais cette masse de données pose la question de son utilisation. Comment prendre en compte toutes ces informations dans le temps très contraint de la prise de décision ? Cette masse de données devient aujourd'hui tellement importante que les prévisionnistes n'ont plus le temps de l'explorer dans son intégralité. Il est alors nécessaire de développer des méthodes de post-traitement des prévisions d'ensemble afin de synthétiser l'information. Les méthodes les plus classiques reposent sur des cartes de probabilités ou des quantiles en mesurant par exemple la probabilité que le vent dépasse un certain seuil en un point pour le lendemain. Cette approche purement statistique est cependant problématique pour une pleine utilisation en contexte opérationnel, en particulier parce qu'elle ne correspond pas à la façon de travailler des prévisionnistes qui s'appuient sur des structures météorologiques cohérentes pour expertiser les modèles.

L'objectif de la présente thèse est donc de lever ces freins pour la prévision d'ensemble à fine échelle PE-AROME, opérationnelle à Météo-France, en proposant des méthodes de post-traitement plus cohérentes physiquement. Pour cela, deux grandes approches sont explorées. Tout d'abord une approche pour détecter automatiquement dans les prévisions PE-AROME des objets nommés échos arqués, une organisation orageuse particulière et potentiellement responsable de dégâts importants. Ce type de synthèse vise à faciliter l'utilisation de la PE-AROME en cas d'évènements à fort impact bien identifiés. Une deuxième approche vise

à développer une synthèse par scénarios des prévisions de pluies, qui est moins spécialisée en comparaison de l'approche pour les échos arqués. Ces scénarios se définissent comme un regroupement de membres proposant des prévisions similaires. La construction de ces scénarios repose sur une réduction de dimension puis sur la création de structures météorologiques récurrentes qui permettront dans une utilisation opérationnelle d'assigner chaque membre à une de ces structures. Tous les membres associés à la même structure météorologique seront alors considérés comme un scénario de l'ensemble.

La partie I de ce manuscrit décrira les différentes étapes pour établir une prévision allant de l'observation du temps, la modélisation de celui-ci jusqu'à la prise de décision pour établir la carte de vigilance. Le chapitre 1 présentera les différentes échelles de l'atmosphère et la construction des prévisions d'ensemble en retraçant l'évolution de la modélisation de l'atmosphère depuis le milieu du XX^{e} siècle. Ce chapitre se terminera sur un descriptif plus détaillé de la PE-AROME. Le chapitre 2 présentera les nombreuses méthodes de post-traitement existantes pour visualiser des prévisions d'ensemble et leurs limites. Le chapitre 3 présentera comment un prévisionniste apporte son expertise aux sorties des modèles, quelles méthodes statistiques semblent les plus en adéquation avec cette expertise et détaillera les deux approches novatrices de cette thèse.

Ces deux approches sont rendues possibles par l'essor des méthodes d'apprentissage profond (ou deep learning) dans les sciences atmosphériques depuis quelques années. Ces méthodes de deep learning seront détaillées spécifiquement dans la partie II de ce manuscrit. Cette partie montrera en quoi les réseaux de neurones convolutifs sont des outils pertinents pour nos problématiques (chapitre 4) et comment ils peuvent apprendre à détecter des échos arqués en utilisant un ensemble de cas où un expert a labellisé préalablement à la main ces objets convectifs sur des images (chapitre 5). Les nombreuses étapes de pré-traitement, nécessaires en raison de la rareté de l'évènement dans le temps et l'espace, seront présentées. De même les autoencodeurs convolutifs, qui permettent une réduction de dimension des champs de pluies pour construire les scénarios, seront décrits (chapitre 6).

Dans la partie III, les performances des réseaux de neurones entraînés à détecter les échos arqués seront présentées au travers de scores objets tout comme les cartes de synthèse des détections (chapitre 7). Des collaborations avec les services opérationnels ont également permis une évaluation subjective du réseau de neurones par des prévisionnistes et le lancement d'une production expérimentale en mode recherche à partir du printemps 2021. Les enseignements tirés de cette production expérimentale, les performances des modèles sur la capacité à correctement anticiper les échos arqués et des premiers travaux sur la détection des échos arqués dans les données observées seront également présentés (chapitre 8).

La construction de la synthèse par scénarios sera détaillée dans la partie IV de ce manuscrit. Le chapitre 9 s'intéressera aux espaces latents des autoencodeurs définis dans le chapitre 6 afin de mieux comprendre comment les autoencodeurs compressent l'information dans un espace à quelques dizaines de dimensions. Après avoir étudié en profondeur ces espaces latents, des classifications seront faites dans le chapitre 10 pour définir les structures météorologiques récurrentes permettant d'utiliser la synthèse par scénarios sur les sorties quotidiennes de la PE-AROME. Une représentation synthétique des scénarios, des études de cas et des scores sur

un an de données pour comprendre comment utiliser au mieux cette synthèse concluront ce manuscrit avec le chapitre 11.

Première partie

Prévision météorologique : méthodes et limites actuelles

Chapitre 1

Modélisation de l'atmosphère

L'Humanité s'intéresse à l'état du ciel depuis maintenant plusieurs millénaires. Leur prêtant au départ des volontés divines, les colères du ciel ont petit à petit bénéficié d'une approche plus scientifique pour les expliquer. Le terme *météorologie* voit le jour dès l'Antiquité grâce à Aristote pour désigner l'étude des météores. La définition moderne de la météorologie recouvrant l'étude des phénomènes atmosphériques est plus récente. Depuis l'Antiquité et les premiers travaux d'Aristote jusqu'au XX^{e} siècle, plusieurs avancées majeures en mécanique des fluides ont permis de décrire de plus en plus précisément l'état de l'atmosphère. C'est le cas par exemple des équations de Navier-Stokes pour décrire l'écoulement d'un fluide. L'essor de la météorologie moderne a permis de définir des équations pour tenter de prévoir le temps pour les heures ou jours à venir. Cependant, ces équations n'ont pas de solution analytique à ce jour. A titre d'exemple l'existence d'une solution régulière (qu'on pourrait qualifier schématiquement de physiquement acceptable) aux équations de Navier-Stokes est l'un des 7 problèmes du millénaire définis par l'institut de mathématiques Clay en 2000 (Fefferman, 2006). En l'absence prouvée de solution analytique, il est cependant possible de prévoir ces états futurs de proche en proche à l'aide de méthodes numériques, ce que font aujourd'hui les modèles météorologiques. Dans la première partie de ce chapitre, nous décrirons rapidement l'atmosphère, son organisation et les différentes échelles qui la composent. Dans un second temps, nous évoquerons comment une prévision météorologique est produite à partir d' observations de l'atmosphère. Un bref historique concernant les modèles météorologiques des années 1950 à nos jours et leurs améliorations sera alors proposé dans une troisième partie. Nous verrons aussi comment parallèlement à ces améliorations, des études sur la prévisibilité à différentes échelles ont révélé la nécessité d'utiliser la prévision probabiliste et non une seule prévision en raison de l'aspect chaotique de l'atmosphère. Ces travaux ont conduit à la construction de prévisions d'ensemble. Nous verrons alors plus particulièrement dans une dernière partie comment la prévision d'ensemble (PE) pour la fine échelle ($\sim$1km) est construite à Météo-France.

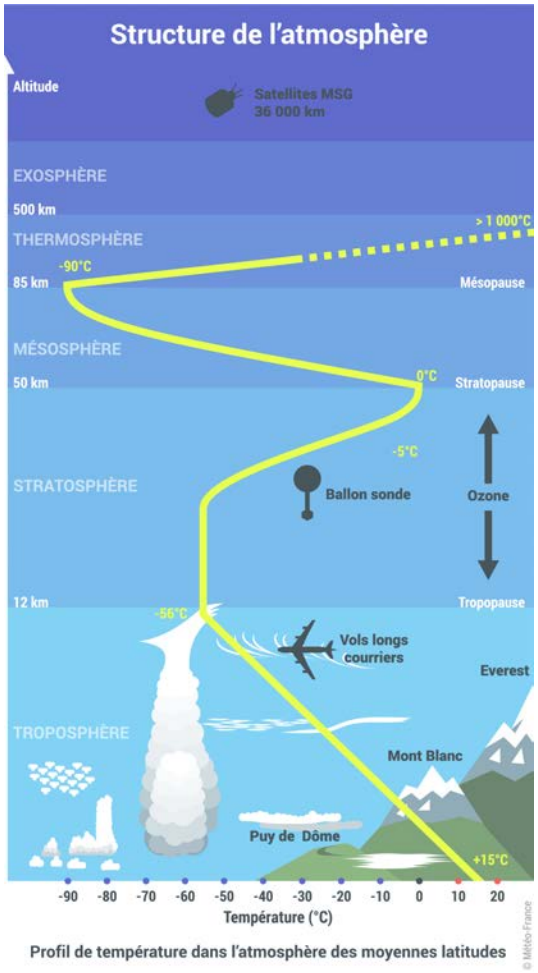


FIGURE 1.1 – *Présentation des différentes couches de l'atmosphère de la troposphère à l'exosphère et profil vertical de température (Source : Météo-France).*

1.1 Description de l'atmosphère

L'atmosphère désigne l'enveloppe gazeuse d'une planète de façon générale. Il est ici question de l'atmosphère terrestre qui est constituée de différentes strates (figure 1.1). Les phénomènes météorologiques se produisant majoritairement près de la surface et nous impactant directement, c'est l'étude de la troposphère qui concentre l'attention des météorologues. Celle-ci va de la surface du globe jusqu'à une altitude d'environ 12km mais cette altitude est variable selon la latitude. La troposphère est plus épaisse près de l'équateur et plus fine près des pôles, avec respectivement des épaisseurs moyennes de 17 et 9 km (Malardel, 2009).

1.1.1 Quelques variables importantes

Avant de décrire plus précisément la circulation de l'atmosphère, il est nécessaire de rappeler quelques grandeurs physiques qui permettent cette description.

- La pression (P) : En météorologie, cette variable s'exprime généralement en hectopascal (hPa). On utilise bien souvent la pression réduite au niveau de la mer (P_{mer}), construite par correction de mesures au sol avec l'altitude. Cette pression est en moyenne de 1013 hPa sur le globe, des valeurs

inférieures à cette moyenne indiquant des conditions dépressionnaires ou cycloniques (la pression peut descendre sous 900 hPa au cœur des plus forts cyclones, Dunnavan and Diercks 1980). A l'inverse des valeurs de pression au-dessus de cette moyenne désignent des conditions anticycloniques (culminant à 1080 hPa sous les plus puissants anticyclones). La pression est aussi utilisée pour étudier l'atmosphère à différents niveaux par les météorologues : si on souhaite étudier des phénomènes en milieu de troposphère, on examine des paramètres à 500 hPa par exemple.

- Le géopotentiel (Φ ou Z) : Cette variable est très utilisée en météorologie pour étudier les circulations atmosphériques en altitude. La pression au niveau de la mer donne un aperçu de la colonne d'air totale, le géopotentiel permet d'étudier des niveaux intermédiaires. Le géopotentiel (Φ) à une altitude h se définit sous la forme suivante :

$$\Phi = \int_0^h g(\phi, z)\, dz \approx GZ \tag{1.1}$$

  Avec $g(\phi, z)$ l'accélération de pesanteur locale à une latitude ϕ et une altitude z. Elle est souvent approchée sous la forme du produit GZ avec G un coefficient constant adimensionnel et c'est alors Z qu'on désigne comme le géopotentiel et qui s'exprime en mètre géopotentiel (mgp) ou $J \cdot kg^{-1}$. C'est cette notation Z qui sera utilisée dans la suite de ce manuscrit. L'étude des structures atmosphériques d'échelle synoptique se fait alors fréquemment grâce au géopotentiel à 500 hPa (Z500). On peut alors étudier des hauts et des bas géopotentiels, comme des zones de basses et de hautes pressions sont étudiées près du sol grâce à la pression.

- La température (T) : La température n'est pas homogène sur tout le globe à cause principalement d'un rayonnement solaire inégal entre l'équateur et les pôles. Elle s'exprime généralement en degrés Celsius (°C) avec une température moyenne sur le globe de 15°C. Des variations importantes sont aussi observées selon l'altitude (figure 1.1). Le gradient vertical de température est un bon marqueur pour caractériser la limite entre troposphère et stratosphère, appelée tropopause (Dameris, 2003). En effet, à cause de la couche d'ozone et de sa réaction avec le rayonnement solaire, la température ne décroît plus dans la stratosphère avec l'altitude.

- Le vent (u,v,w) : La plupart des fluides s'étudient à travers leurs écoulements. C'est également le cas pour l'atmosphère avec trois composantes : u (direction Ouest-Est en général), v (Sud-Nord) et w pour les mouvements verticaux. On parle d'ascendance si le mouvement est montant (vers la tropopause), de subsidence dans le cas contraire. On les exprime généralement en $m \cdot s^{-1}$ ou en $km \cdot h^{-1}$. La force du vent qui est la norme 2 du vecteur horizontal est aussi très étudiée. On différencie la force du vent moyen sur 10 minutes et les rafales de vent, ces dernières étant par exemple plus utiles pour matérialiser le risque de dégât sur une situation donnée que le vent moyen.

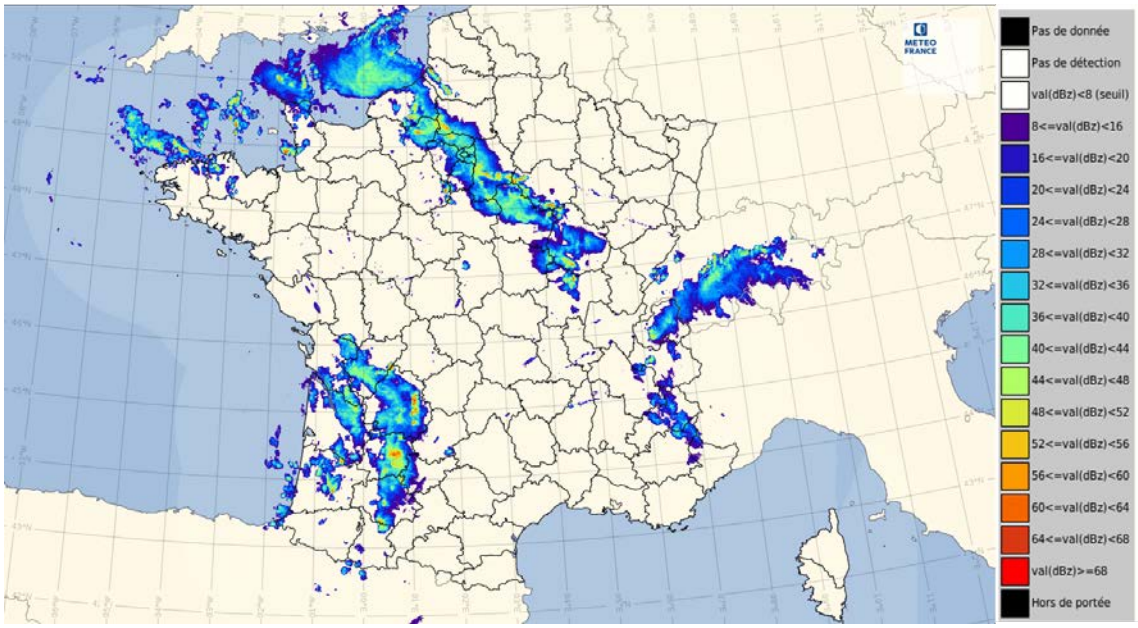


FIGURE 1.2 – *Mosaïque de réflectivités SERVAL pour la France métropolitaine. Image du 15 mai 2022 à 22h UTC. Plus les couleurs sont chaudes (vert/jaune), plus les dBZ sont importants et les intensités fortes. L'absence de couleur indique l'absence de précipitation (Source : Météo-France, outil Synopsis).*

D'autres variables plus spécifiques aux précipitations seront utilisées dans ce manuscrit :

- La réflectivité : Elle est mesurée par les radars météorologiques dans la gamme des microndes avec une longueur d'onde de 1 à 10 cm. Ils mesurent le signal renvoyé par des zones précipitantes à 360 degrés autour d'eux afin d'en déduire la nature des précipitations ou encore leurs intensités à un instant donné. Ce signal se mesure en dBZ. Il peut être converti avec quelques approximations en millimètre par heure ($mm \cdot h^{-1}$) grâce à une loi « Z-R » comme celle de Marshall and Palmer (1948). Après conversion, on parle plutôt de pseudo-réflectivités. En combinant les différents radars recouvrant le territoire français, une mosaïque radar SERVAL est construite (Système d'Elaboration des produits Radar et de Visualisation centrALisé, figure 1.2 extraite de l'outil Synopsis [1]).

- Le cumul de pluie (RR) : On étudie également la quantité de précipitation tombée sur une période de temps donnée. Les cumuls s'expriment en millimètre (mm). 1 mm correspondent à cette hauteur d'eau sur 1 m^2 (donc 1L d'eau). Cette quantité est mesurée à l'aide de pluviomètres au sol, elle peut aussi être estimée à l'aide des mesures radars, la combinaison de ces deux sources d'information permet de définir des produits de fusion comme l'ANalyse par spaTIaLisation hOraire des PrEcipitations (ANTILOPE) à Météo-France (Lorentin, 2013) afin d'estimer au mieux ce paramètre (figure 1.3).

1. Outil de travail des prévisionnistes à Météo-France

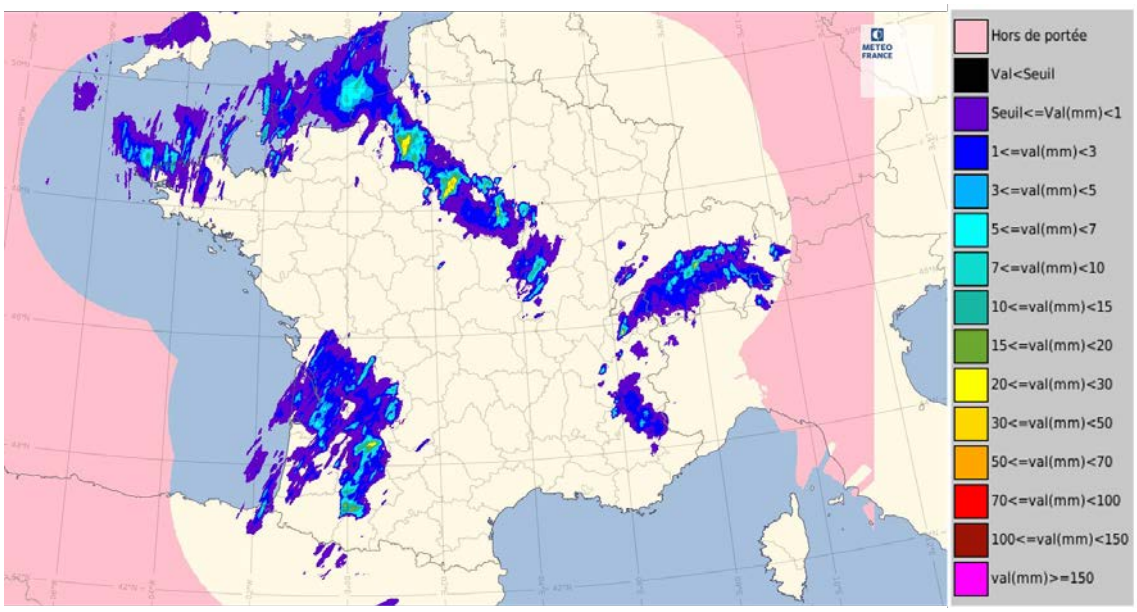


FIGURE 1.3 – *Cumul de précipitations (produit ANTILOPE) combinant les informations issues des mesures des pluviomètres et des données radars. Cumul de précipitation du 15 mai 2022 de 21h à 22h UTC (Source : Météo-France, outil Synopsis).*

Cette liste de paramètre n'est absolument pas exhaustive. Il existe bien d'autres variables importantes comme le contenu en eau liquide, l'humidité relative qui ne sont pas présentés.

1.1.2 Les différentes échelles de l'atmosphère

Les phénomènes atmosphériques peuvent être divisés en différentes catégories présentées figure 1.4 selon leurs tailles caractéristiques (Orlanski, 1975). La taille d'un phénomène est aussi très corrélée à la durée de vie de celui-ci (plus le phénomène est petit, plus sa durée de vie est courte). Ces différentes échelles n'impliquent cependant pas l'absence d'interaction, elles s'influencent en effet les unes les autres formant ainsi un continuum (Malardel, 2009). Ces différentes échelles sont rapidement présentées avec quelques exemples de phénomènes atmosphériques en gras.

échelle planétaire

Cette catégorie englobe les phénomènes hémisphériques (taille caractéristique supérieure 2000 km) et avec une durée de vie très longue excédant les 10 jours. Ces phénomènes sont souvent liés à la rotation de la Terre et au gradient thermique entre l'équateur et les pôles. En résultent de fortes ascendances au niveau de la Zone de Convergence Intertropicale (**ZCIT**, Fletcher 1945) près de l'équateur, formant une branche des **cellules de Hadley** (1735) qui permettent le transport du surplus d'énergie vers les pôles. Cependant ces cellules n'atteignent pas les pôles à cause de la force de Coriolis issue de la rotation de la Terre sur elle-même. Le gradient thermique est aussi responsable de puissants **courants jet** (zone de vents

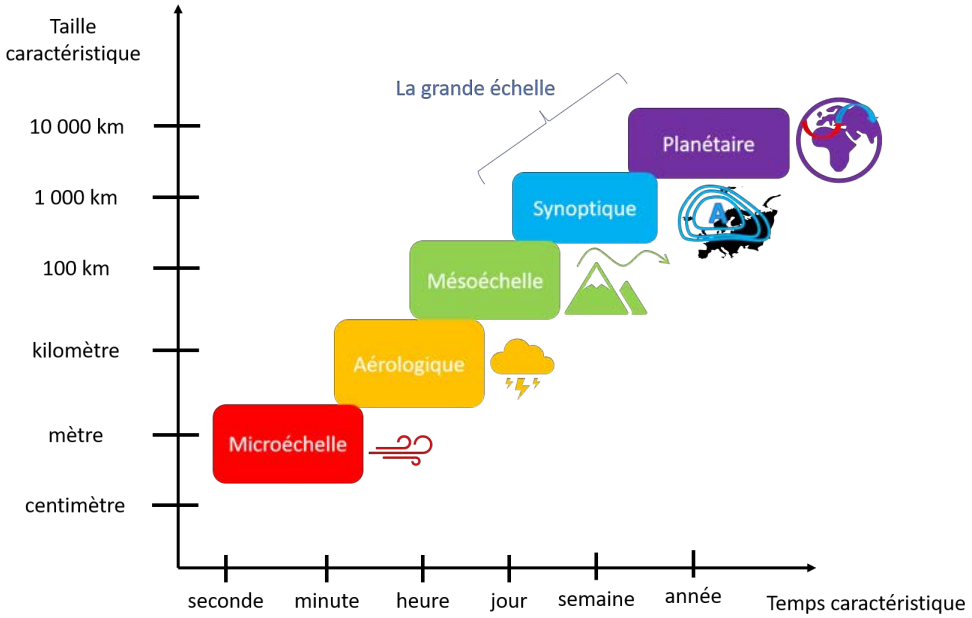


FIGURE 1.4 – *Représentation schématique des différentes échelles météorologiques avec un exemple pour chacune des échelles : l'échelle planétaire (circulation générale), l'échelle synoptique (anticyclone, représenté à l'aide d'isobares en P_{mer}), la mésoéchelle (ondes de relief), l'échelle aérologique (cellule orageuse) et la microéchelle (rafale de vent). Inspiré de Malardel (2009).*

forts) qui sont proches de la tropopause. Ils délimitent dans l'hémisphère Nord les masses d'air douce ou chaude au sud, des masses d'air froide au Nord. La France se situe généralement à la limite de ces masses d'air, proche du courant jet polaire. On parle des moyennes latitudes. Ce courant jet ne circule pas cependant de façon zonale (Ouest-Est). Ses oscillations Nord-Sud forment alors les **ondes de Rossby** (1939) qui appartiennent également à l'échelle planétaire. Ces phénomènes sont très étudiés pour comprendre par exemple comment le réchauffement climatique peut bouleverser ces équilibres (Lu et al., 2007). Ces changements de circulation sont capables de changer le climat d'une région du globe.

échelle synoptique

L'échelle synoptique est celle des phénomènes de l'ordre du millier de kilomètres avec des durées de vie de l'ordre de quelques jours à la semaine. Aux moyennes latitudes, c'est en particulier dans cette catégorie que se classent les **dépressions** et les **anticyclones** qui peuvent donner lieu respectivement à des tempêtes ou à des canicules par exemple (Malardel, 2009). Les positions de ces objets météorologiques dépendent fortement du courant jet en altitude et des ondes de Rossby et constituent un exemple d'interaction entre l'échelle planétaire et l'échelle synoptique. Les **fronts** délimitant les masses d'airs chaudes et froides qui s'enroulent autour d'un centre dépressionnaire font aussi partie de cette catégorie (Calas, 2013).

Le terme de grande échelle est souvent utilisé pour désigner à la fois l'échelle planétaire et l'échelle synoptique.

méso-échelle

La méso-échelle désigne les phénomènes météorologiques de l'ordre de 10 à 100 km avec des durées de vie de quelques heures. C'est en particulier à cette échelle qu'appartient la catégorie des **ondes de relief** (Bretherton, 1969) ou ondes orographiques qui sont des ondes de gravité initiées par le soulèvement contraint de la masse d'air au passage d'un relief. Plusieurs organisations orageuses (regroupement de plusieurs cellules) appartiennent aussi à cette catégorie comme les **lignes de grains** (Trapp et al., 2005).

échelle aérologique

L'échelle aérologique est celle des tailles caractéristiques allant de quelques centaines de mètres à une dizaine de kilomètres avec des temps caractéristiques courts de quelques dizaines de minutes à quelques heures. On considère que cette échelle est principalement celle de la convection (Malardel, 2009). Une **cellule orageuse isolée** appartient à cette catégorie comme les **tornades** qui sont parmi les plus petits éléments de cette échelle.

microéchelle

La microéchelle est la catégorie des phénomènes dont la taille est de seulement quelques mètres avec des temps caractéristiques de l'ordre de la seconde à la minute. On retrouve à cette échelle les **tourbillons de poussière**, les **rafales de vent**. Cette échelle est plus généralement celle de la turbulence.

La difficulté de prévision d'un phénomène est reliée à ces échelles caractéristiques car les processus physiques prépondérants ne sont pas les mêmes (Malardel, 2009). C'est en particulier pour cette raison que la prévision des orages est difficile si on la compare par exemple à la prévision d'une canicule que nous pouvons anticiper quelques jours avant. La partie suivante va présenter rapidement les grandes étapes nécessaires pour établir ces prévisions météorologiques.

1.2 Les grandes étapes de l'observation à la prévision

Pour établir des prévisions météorologiques, trois grandes étapes sont nécessaires (figure 1.5). De façon schématique, il est tout d'abord nécessaire de construire un modèle qui permet d'estimer l'état de l'atmosphère de proche en proche à l'aide des équations d'évolution de l'atmosphère. Cette construction de proche en proche nécessite de créer un point de départ qu'on appelle l'état initial et qui doit correspondre au mieux à l'état de l'atmosphère à cet instant. Pour estimer cet état initial, les systèmes de prévision doivent s'appuyer sur des observations. Ces observations sont cependant très insuffisantes pour former un état initial complet. C'est pourquoi cette estimation s'appuie également sur les anciennes prévisions faites par le modèle (figure 1.5, partie observations et ancienne prévision). L'étape suivante a pour but d'utiliser ces données afin de créer l'état

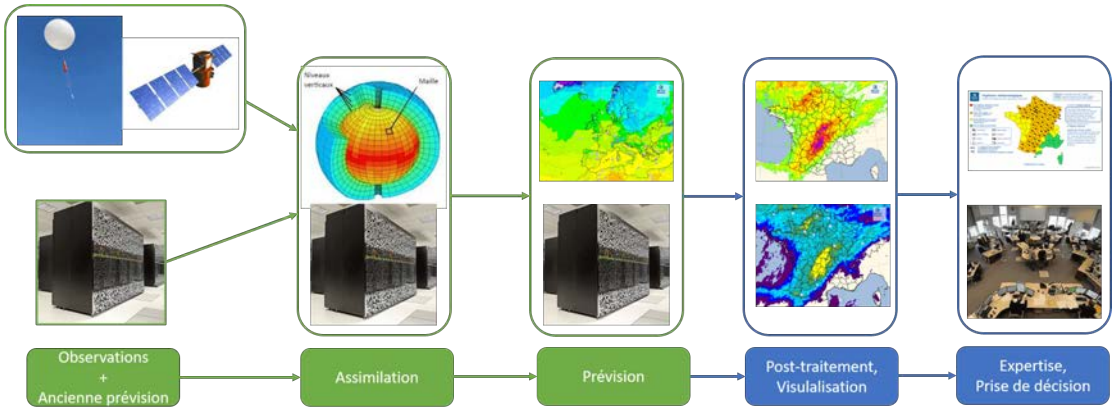


FIGURE 1.5 – *Représentation schématique de la chaîne de prévision. Le chapitre 1 s'intéresse à la partie en vert à travers : l'observation du temps présent à l'aide de différents instruments et l'utilisation d'anciennes prévisions (à gauche), la prise en compte de ces données par le modèle à l'aide de l'assimilation de données pour créer un état initial (au centre) et enfin la prévision des états futurs à l'aide du modèle (à droite).*

initial dans le modèle météorologique. Celui-ci représente l'atmosphère sous une forme discrétisée à l'aide d'une grille en trois dimensions (figure 1.5, partie centrale). Dans chaque maille de la grille, l'atmosphère est supposée homogène (une seule valeur pour chaque variable). Il est nécessaire de fournir à chaque maille une valeur pour chaque paramètre à l'état initial. L'objectif du module d'assimilation de données est alors de construire un état initial le plus fidèle possible avec les données disponibles et qui reste physiquement cohérent. Lorsque cet état initial est construit, les états futurs de l'atmosphère sont estimés à l'aide du modèle décrit précédemment (figure 1.5, partie prévision).

1.2.1 Observation de l'atmosphère

Pour connaître au mieux l'état présent de l'atmosphère, le maximum d'information est recueilli que ce soit sur les conditions près de la surface mais également en altitude jusqu'à plusieurs dizaines de kilomètres. Cette quantité de données recueillie augmente régulièrement depuis plusieurs décennies (figure 1.6). Le premier type de mesure qui est le plus ancien est la mesure par des instruments au sol à l'aide de thermomètres, baromètres, pyranomètres (mesure du rayonnement solaire),... Ces stations peuvent être sur terre, sur des bouées en mer ou même sur certains navires. Les radars météorologiques sont aussi d'une aide précieuse pour estimer les précipitations mais aussi le vent en se basant sur l'effet Doppler (Saltikoff et al., 2019). Pour acquérir des données en altitude, des ballons sondes (figure 1.5, en haut à gauche) sont régulièrement envoyés. Des données issues des avions de ligne sont également utilisées. La chute du trafic aérien en mars 2020 est d'ailleurs perceptible sur la zone verte en bas de la figure 1.6. Enfin les satellites météorologiques constituent aujourd'hui la masse de données la plus importante (figure 1.6), ces satellites pouvant être défilants ou géostationnaires (Eyre et al., 2020, 2022). L'arrivée de nouveaux satellites mais également l'augmentation de résolution des modèles et leur capacité à assimiler plus d'observations (comme en

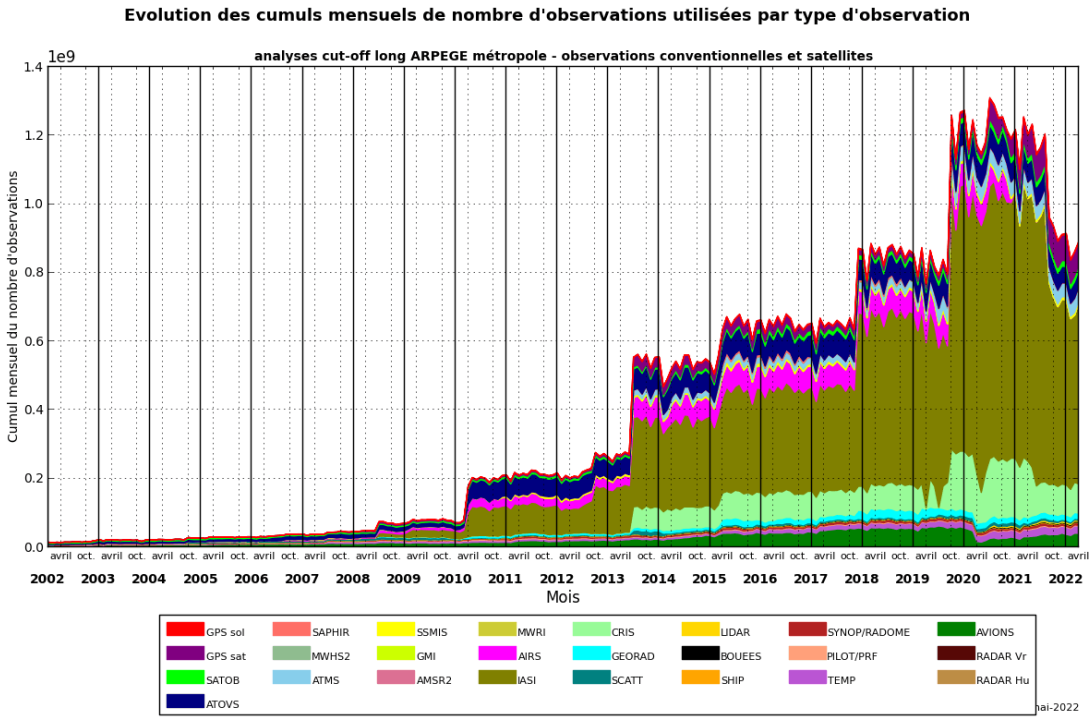


FIGURE 1.6 – *Évolution du nombre d'observations utilisées par le modèle français ARPEGE entre 2002 et 2022 par type de données (Source : équipe COMPAS, Météo-France).*

juillet 2019) expliquent les principales augmentations du nombre d'observations utilisées ces dernières années. La disparition du satellite Metop-A en 2021 a en revanche fait chuter ce nombre.

1.2.2 Assimilation de données

L'assimilation de données a pour objectif de prendre en compte les observations disponibles afin de créer un état initial. Mais l'assimilation n'utilise pas seulement les observations car leur nombre est trop faible comparé à la grille des modèles. De plus certaines observations ne sont pas directement liées à une variable du modèle (exemple : radiance satellitaire, réflectivité radar) et ne sont donc pas utilisables directement pour estimer l'état initial. C'est pourquoi, l'assimilation de données utilise aussi une prévision antérieure faite par le modèle qu'on espère la plus juste possible. Cette prévision présente l'avantage de fournir une valeur pour chaque variable météorologique et pour chaque point de grille, on l'appelle l'ébauche (notée x_b). Les observations sont alors utilisées pour corriger cette ébauche, on parle de schéma prédicteur-correcteur. La comparaison entre observations (y) et données modèles (x) est faite à l'aide d'un opérateur d'observation qu'on note usuellement $\mathcal{H}$ avec la relation suivante :

$$y = \mathcal{H}(x) \tag{1.2}$$

Cet opérateur est linéarisé et noté $\mathbf{H}$. L'objectif est donc d'obtenir un état initial qui minimise l'erreur commise par rapport à l'ébauche x_b et par rapport aux observations y_{obs}. Ce champ qui minimise l'erreur aux observations et à l'ébauche

est appelé l'analyse et noté x_a. L'approche BLUE pour Best Linear Unbiased Estimator est l'une des approches les plus communes pour estimer x_a (Bouttier and Courtier, 2002). x_a s'écrit sous la forme :

$$x_a = x_b + \mathbf{K}(y_{obs} - \mathbf{H}(x_b)) \tag{1.3}$$

L'analyse s'interprète alors comme une correction de l'ébauche par sa différence aux observations, pondérée par la matrice de gain $\mathbf{K}$ qui se déduit de la matrice $\mathbf{H}$, de la matrice des covariances d'erreurs d'ébauche notée $\mathbf{B}$ et de la matrice des covariances d'erreurs d'observations notée $\mathbf{R}$:

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \tag{1.4}$$

Cette analyse x_a obtenue par le BLUE est optimale si on suppose les observations et l'ébauche non biaisées et si les distributions d'erreurs sont gaussiennes. L'estimation de la matrice $\mathbf{R}$ est faite à partir des connaissances du système d'observation (erreur de mesure concernant les instruments par exemple). Elle est souvent supposée diagonale avec donc des termes de covariance nuls. La matrice $\mathbf{B}$ est elle souvent estimée de façon climatologique ou à l'aide d'un ensemble d'assimilation (Pereira and Berre, 2006). Pour résoudre le BLUE, x_a peut être déterminé à l'aide d'un formalisme dit variationnel qui a alors pour objectif de minimiser une fonction coût. Ce formalisme est nommé 3D-Var ou 4D-Var selon l'ajout ou non de la variable temporelle pour optimiser une analyse dans le temps (figure 1.7). L'intervalle dans lequel les observations sont prises en compte est appelé la fenêtre d'assimilation. La fonction coût J s'écrit sous la forme :

$$J(x) = \frac{1}{2} \cdot (x - x_b)^T \cdot \mathbf{B}^{-1} \cdot (x - x_b) + \frac{1}{2} \cdot (y_{obs} - \mathbf{H}(x))^T \cdot \mathbf{R}^{-1} \cdot (y_{obs} - \mathbf{H}(x)) \tag{1.5}$$

Le premier terme concerne la minimisation de l'erreur par rapport à l'ébauche, le deuxième par rapport aux observations. D'autres méthodes existent mais ne seront pas détaillées comme le filtre de Kalman (Carrassi et al., 2018). A l'issue de cette étape, une analyse x_a, état optimal obtenu par combinaison des observations et de l'ébauche, est créée pour former notre état initial dans la grille du modèle.

1.2.3 Modèle météorologique

A partir de cette analyse x_a le rôle du modèle va être alors de produire une prévision $x(t)$. Plus précisément, si on représente le comportement de l'atmosphère par un opérateur non-linéaire $\mathcal{M}$, un modèle de Prévision Numérique du Temps (PNT) tente d'intégrer l'équation suivante de l'instant initial 0 à l'instant t :

$$\frac{\partial x}{\partial t} = \mathcal{M}(x) \tag{1.6}$$

En pratique, l'opérateur $\mathcal{M}$ est scindé en une partie dynamique $\mathcal{D}$ que le modèle peut résoudre et une partie paramétrisation $\mathcal{P}$. Le deuxième terme est rendu nécessaire par la présence de phénomènes sous-mailles (Stensrud, 2009). En effet, la maille du modèle détermine ce que le modèle est en capacité de résoudre explicite-

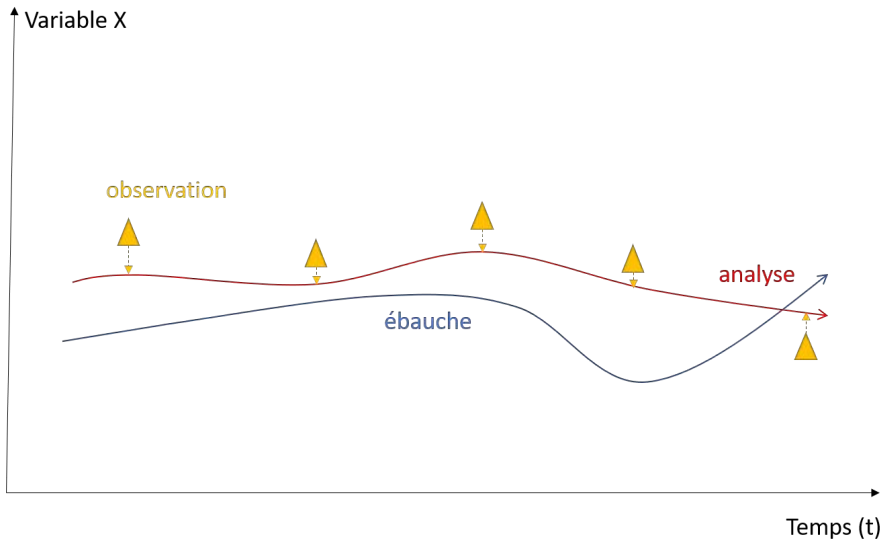


FIGURE 1.7 – *Représentation schématique de la construction de l'état analysé (rouge) à partir de l'ébauche (bleu) et des observations (orange) pour une assimilation 4D-Var.*

ment. Par exemple un modèle avec une maille de plusieurs dizaines de kilomètres sera incapable de modéliser une cellule orageuse ayant une taille caractéristique de l'ordre de quelques kilomètres. Pour un modèle avec cette maille, il s'agit donc d'un phénomène sous-maille. Mais ce phénomène influençant aussi les échelles plus grandes, il est nécessaire de le prendre en compte, c'est le rôle des paramétrisations. La turbulence, la convection peu profonde, la microphysique sont plusieurs exemples de phénomènes à paramétrer dans un modèle météorologique. L'état $x(t+\Delta t)$ de l'atmosphère est alors déduit à partir de $x(t)$ et du terme de tendance obtenu après calcul de la partie dynamique et des paramétrisations.

1.3 Prévisibilité et modèles météorologiques : des avancées conjointes

Cette partie retrace l'histoire de la PNT depuis la fin de la deuxième guerre mondiale jusqu'à nos jours. Les évolutions liées à la prise en compte des incertitudes des prévisions sont également proposées en parallèle, ces avancées ayant directement influencé la construction de nos modèles actuels.

1.3.1 Les premiers modèles météorologiques (années 1950-1960)

Les premiers modèles météorologiques se développent conjointement aux premiers ordinateurs après la seconde guerre mondiale (Lewis, 2005), un des premiers étant celui de Jule Charney et son équipe aux Etats-Unis. Ces modèles sont bien sûr beaucoup plus simples que ceux d'aujourd'hui avec une maille de plus de 700 km couvrant l'Amérique du Nord, une seule couche à 500 hPa (Charney, 1948).

Ces modèles supposent également l'atmosphère barotrope [2] et l'assimilation de données est faite manuellement. Produisant des prévisions à 24h, les performances sont très inégales d'un jour à l'autre. L'intérêt pour les modèles météorologiques se développe rapidement dans les années 1950 avec le premier modèle en condition opérationnelle à l'Institut Suédois de Météorologie et d'Hydrologie (SMHI) en 1954 et aux Etats-Unis en 1955 (Lynch and Moor, 2008). Plusieurs variables météorologiques importantes ne sont cependant pas encore présentes dans ces modèles (comme la pluie ou les nuages). Au début des années 1960, des modèles basés sur les équations de Navier-Stokes deviennent possibles grâce aux progrès de l'informatique, proposant des variables météorologiques en sortie correspondant mieux au temps sensible comme la pluie. Plusieurs de ces modèles deviennent rapidement opérationnels au milieu des années 1960 (Lynch and Moor, 2008). Les modèles dits de circulation générale ou GCM, destinés principalement à l'étude du climat [3], suivent rapidement. On en dénombre huit dans cinq pays différents au milieu des années 1960 que sont les Etats-Unis, le Royaume-Uni, l'URSS, l'Australie et la France (Lewis, 2005).

1.3.2 Questionnements autour de la notion de prévisibilité (années 1950-1960)

En parallèle de ces développements, des chercheurs s'intéressent aux erreurs commises par ces modèles, surtout aux échéances longues. En effet pour d'autres systèmes physiques comme le mouvement d'un pendule simple dans le cadre de la physique newtonienne, nous savons qu'une légère erreur de position commise à l'instant initial n'entraînera pas une croissance de cette erreur dans la prévision future. Ainsi l'erreur commise à l'instant initial n'aura que des conséquences limitées sur la trajectoire prévue (comparée à la trajectoire observée). Cependant, cette caractéristique est loin d'être systématique pour l'ensemble des systèmes physiques et semble particulièrement fausse pour l'atmosphère. La théorie de l'instabilité barocline [4] en est un exemple et est développée conjointement par Jule Charney et Eric Eady à la fin des années 1940 (Lewis, 2005). Cette idée que les erreurs commises sur l'état de l'atmosphère peuvent croître rapidement jusqu'à rendre la prévision complètement inutilisable est exprimée clairement dans l'article de Eady (1951) :

« [...] we never know what small perturbations may exist below a certain margin of error. Since the perturbations may grow at an exponential rate, the margin of error in the forecast (final) state will grow exponentially as the period of forecast is increased, and this possible error is unavoidable whatever our method of forecasting. »

Il prône déjà alors dans ce même article une approche probabiliste pour traiter ces questions. Thompson (1957) montre quelques années plus tard que le manque de données d'observation dans le Pacifique entraîne un accroissement significatif des

2. Cette hypothèse impose en particulier qu'à un niveau de pression donné, la température est uniforme.

3. Ces modèles se sont diversifiés depuis. Ils sont également utilisés pour étudier l'atmosphère des autres planètes du système solaire par exemple.

4. L'instabilité barocline est un phénomène clé aux moyennes latitudes qui est à l'origine en particulier des tempêtes.

erreurs sur la prévision 48h plus tard au Canada et aux Etats-Unis, confirmant les travaux d'Eady et militant pour une extension des systèmes d'observation sur les zones maritimes. Ces travaux posent les bases du concept d'horizon de prévisibilité. L'horizon de prévisibilité correspond à l'échéance à partir de laquelle, à cause de la croissance des erreurs dans la prévision, celle-ci perd toute valeur informative. Pour la grande échelle, qui est la seule à être représentée dans les modèles à l'époque, Thompson estime un horizon de prévisibilité à une semaine (Thompson, 1957). Cette limite sera réévaluée à deux semaines dans des travaux ultérieurs (Novikov, 1959; Lewis, 2005).

Parallèlement aux progrès réalisés par les modèles météorologiques, des travaux importants sur la prévisibilité sont menés par Edward Lorenz sur un modèle idéalisé, se rapprochant cependant des équations qui régissent l'atmosphère (Lorenz, 1963). Lorenz (1963) montre dans son système l'existence d'un attracteur étrange [5](figure 1.8). L'orbite du système évolue alors sur cette attracteur en s'enroulant autour d'un point fixe (instable) puis change de point de fixe et s'enroule autour de ce deuxième point et ainsi de suite. Les enroulements autour de ces deux points fixes forment les deux ailes du célèbre papillon de Lorenz qui sera largement repris pour illustrer la théorie du chaos à partir des années 1970. On peut noter que ces transitions d'une aile à l'autre ne sont absolument pas régulières et semblent chaotiques. Lorenz (1963) montre également la grande sensibilité aux conditions initales des orbites. Même si on tombera toujours dans cet attracteur étrange, il montre que deux états initiaux proches deviendront aussi différents que des états choisis au hasard en un temps fini. Ces résultats montrent les limites de la prévision déterministe qui ne sera plus informative après un temps fini à cause de la sensibilité du système aux conditions initiales. Ils fermant la porte au rêve d'une prévision parfaite sur de longues échéances, confirment la notion importante d'horizon de prévisibilité et le besoin d'établir des prévisions probabilistes. Quelques années plus tard, Lorenz (1969) s'intéresse aux erreurs dans un système à plusieurs échelles. Il montre que l'horizon de prévisibilité dépend de l'échelle des phénomènes considérés : plus l'échelle est grande, plus l'horizon de prévisibilité peut être repoussé. Il faut cependant noter que ces conclusions sont obtenues sur un modèle idéalisé, comme le rappelle Lorenz (1969) lui-même. Mais Ruelle and Takens (1971) montreront qu'un système plus complexe gouverné par les équations de Navier-Stokes obéit à une dynamique chaotique, ce qui rend la conclusion de Lorenz plausible à l'atmosphère.

1.3.3 L'amélioration des modèles grandes échelles (années 1970-2020)

Le développement et l'amélioration des modèles déterministes [6] se poursuivent dans les années 70 et au-delà, portés par des machines de plus en plus performantes, un réseau d'observation de plus en plus dense (arrivée des premiers satellites destinés aux services météorologiques, partage des observations au niveau mondial) et une meilleure connaissance de l'atmosphère et des façons de la modéliser. Les

5. espace dans lequel le système évolue après un temps suffisamment long, quelque soit les conditions initiales

6. déterministes car ils ne produisent qu'une seule prévision à un instant donné

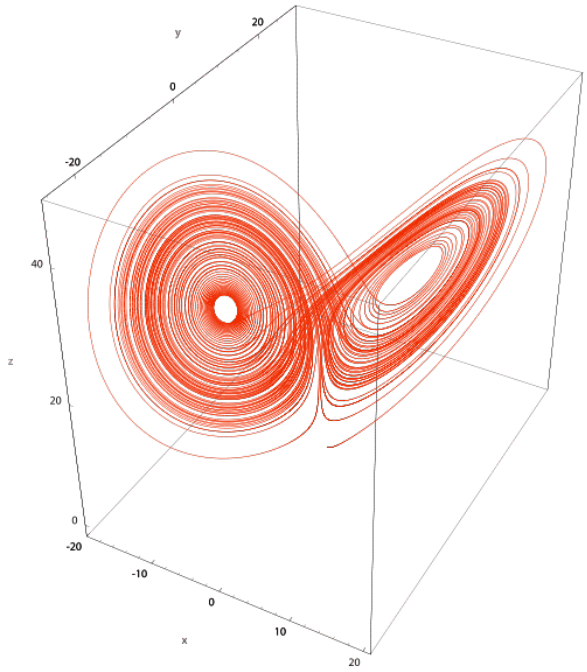


FIGURE 1.8 – *Attracteur de Lorenz (Source : https://commons.wikimedia.org/w/index.php?curid=3700799)*

mailles horizontales, de plusieurs centaines de kilomètres il y a 50 ans, mesurent autour de 10 km aujourd'hui et couvrent l'ensemble du globe avec une centaine de niveaux sur la verticale. Ces progrès peuvent se matérialiser sous forme de gain de prévision en nombre de jours. Si les erreurs de prévisions à 72h sont les mêmes que celles qui étaient commises il y a X années à 48h, nous pouvons dire que un jour de prévision a été gagné entre temps. La figure 1.9 montre que nous avons gagné un jour de prévision tous les dix ans sur les quarante dernières années (Bauer et al., 2015). La figure 1.9 illustre également comment l'amélioration des prévisions a permis d'envisager de réaliser des prévisions à plus longues échéances, la qualité des prévisions à 10 jours était en effet trop mauvaise dans les années 1980-1990 pour être exploitable. Enfin, cette figure montre des performances très inégales entre l'hémisphère Sud et Nord dans les années 1980-1990. Cette différence s'explique largement par un réseau d'observation bien moins dense que dans l'hémisphère Nord. Mais l'utilisation de plus en plus importante de données satellites sur l'ensemble du globe permettra des performances équivalentes à partir des années 2000.

1.3.4 La construction des prévisions ensemblistes (années 1970-2020)

A la fin des années 1960, de premières prévisions probabilistes voient le jour grâce à des méthodes de Monte Carlo appliquées à l'état initial (Epstein, 1969). Dans ces travaux, la variable météorologique n'est plus déterministe, mais une variable aléatoire à laquelle est associée une loi de probabilité. Il est seulement

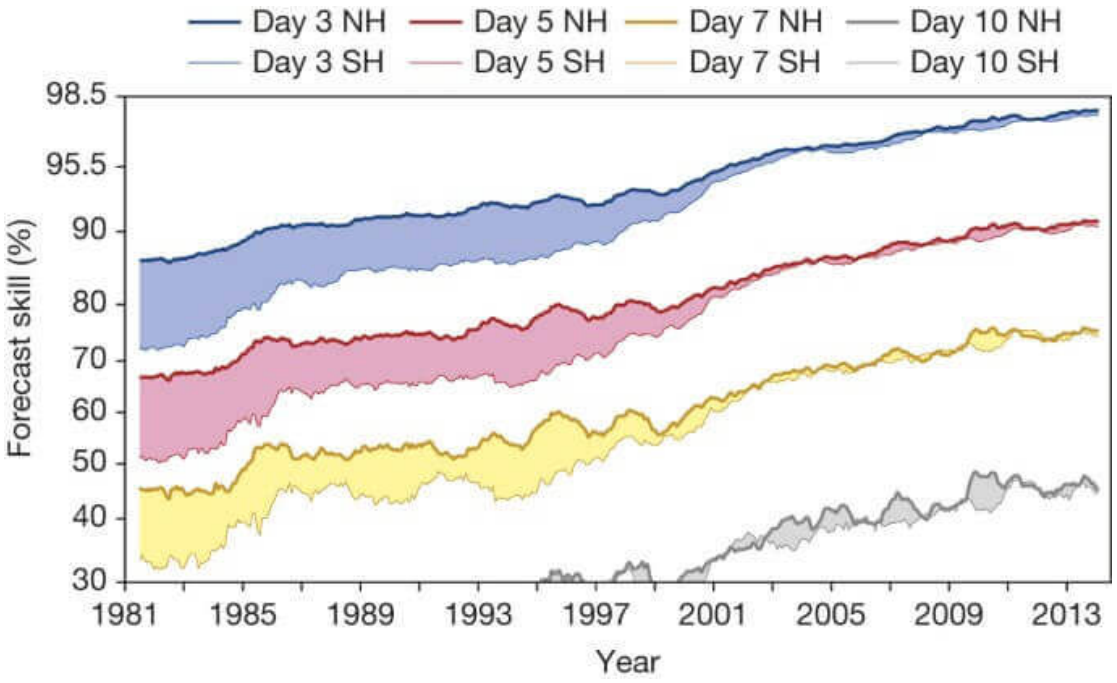


FIGURE 1.9 – *Qualité de prévision du géopotentiel à 500 hPa entre les années 1980 et 2010 à 3 jours (bleu), 5 jours (rouge), 7 jours (jaune) et 10 jours (gris). La courbe en trait gras représente les prévisions sur l'héminsphère Nord, en trait plus fin sur l'hémisphère Sud. D'après Bauer et al. (2015).*

question ici de modèle-jouet pas encore de modèle opérationnel. C'est également dans les années 1970-1980 que Lorenz poursuivra ses travaux sur la prévisibilité et confirmera, sur des modèles plus réalistes, que pour une vision déterministe il existe un horizon de prévisibilité limité qui s'établit à environ 2 semaines pour la grande échelle (Lorenz, 1982). Ces travaux nourrissent un intérêt de plus en plus croissant pour une prévision probabiliste en contexte opérationnel à l'aide d'une prévision d'ensemble (figure 1.10). Les travaux de Lorenz incitent en particulier les services météorologiques à travailler sur les erreurs commises lors de la création de l'état initial (Molteni et al., 1996). En échantillonnant différents états initiaux, plusieurs prévisions sont lancées pour échantillonner les potentiels états futurs à l'aide d'une distribution. On appelle ces prévisions des membres de la prévision d'ensemble.

La prévision probabiliste (et en particulier la prévision d'ensemble) en contexte opérationnel tardera cependant à venir. En effet, réaliser plusieurs prévisions pour explorer les incertitudes demande des ressources informatique bien plus importantes. Les améliorations portant sur la parallélisation des codes seront un atout majeur pour y parvenir. D'autres questions se posent également concernant ces prévisions ensemblistes. En particulier, combien de membres sont nécessaires ? En effet si nous cherchons à échantillonner une distribution, se pose la question du nombre minimum de trajectoires à calculer pour y parvenir convenablement. Des travaux (Leith, 1974) tendent à démontrer qu'un nombre de membres assez restreint ($\sim$ 10 membres) permet déjà de proposer des résultats intéressants. Une autre question importante porte sur la ou les méthode(s) de perturbation. En quoi va différer l'état initial d'un membre de l'ensemble par rapport à son voisin ? Une méthode portée principalement par le Centre Européen pour les Prévisions

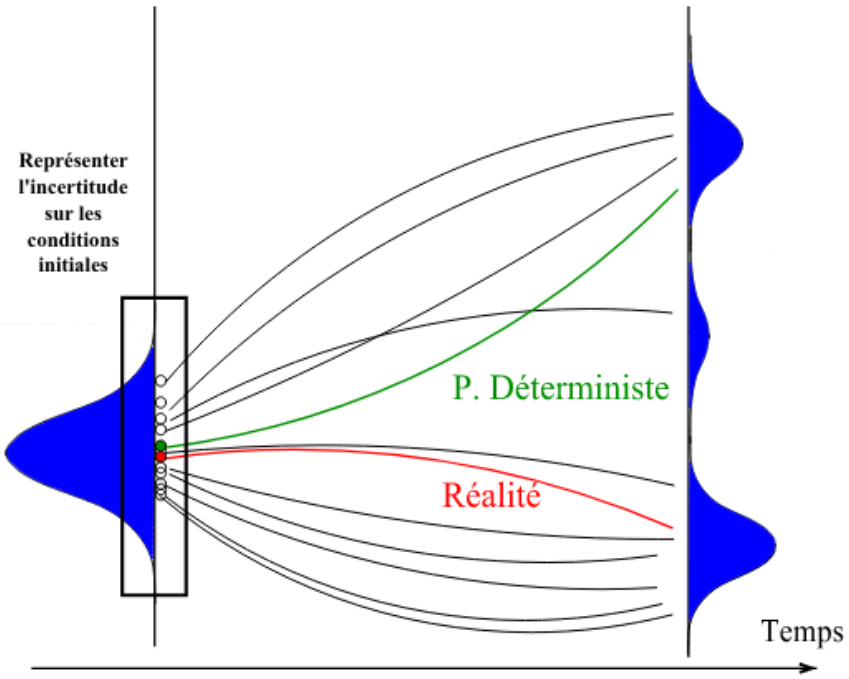


FIGURE 1.10 – *Représentation schématique d'une prévision d'ensemble (schéma de Carole LABADIE, Météo-France). Chaque trajectoire noire est un membre de la prévision d'ensemble, la courbe verte la prévision déterministe et en rouge l'état observé.*

météorologiques à moyen terme (CEPMMT) est la méthode des vecteurs singuliers (Lacarra and Talagrand, 1988). Les vecteurs singuliers sont les modes instables qui ont le plus tendance à s'amplifier au cours du temps. Cette méthode est particulièrement efficace pour représenter les modes d'instabilité barocline aux moyennes latitudes. Le National Centers for Environmental Prediction (NCEP) aux Etats-Unis fait un choix différent et utilisera une méthode dite de « breeding » (Tracton and Kalnay, 1993). Une perturbation aléatoire est ajoutée à l'analyse et des prévisions sur un jour sont lancés avec et sans cette perturbation. On utilise la différence des deux prévisions pour créer une nouvelle perturbation et, après normalisation, on réitère le processus pour créer de nouvelles perturbations (figure 1.11).

Les premiers systèmes de prévision ensembliste opérationnels voient le jour au CEPMMT, au NCEP et au Centre Météorologique Canadien (CMC) dans les années 1990 (Buizza et al., 2005). Ils vont là aussi se multiplier rapidement dans différents centres météorologiques dans le monde, dont Météo-France avec la Prevision d'Ensemble ARPEGE (PEARP). Ces modèles vont connaître comme leurs homologues déterministes de significatives améliorations dans les décennies suivantes : assimilation d'un nombre d'observations toujours plus conséquent, des résolutions de plus en plus fines ou encore des méthodes de perturbation plus efficientes. D'autres sources d'erreurs dans la prévision sont en particulier explorées comme les erreurs commises par le modèle lui-même. Buizza (2019) retrace ces évolutions au CEPMMT. Pour compléter cette introduction à la prévision d'ensemble, notons que des prévisions d'ensemble ont aussi été construites en agrégeant des prévisions de modèles déterministes lancées à des heures différentes et/ou issus de multiples centres météorologiques, on parle ici d'ensemble « du pauvre »

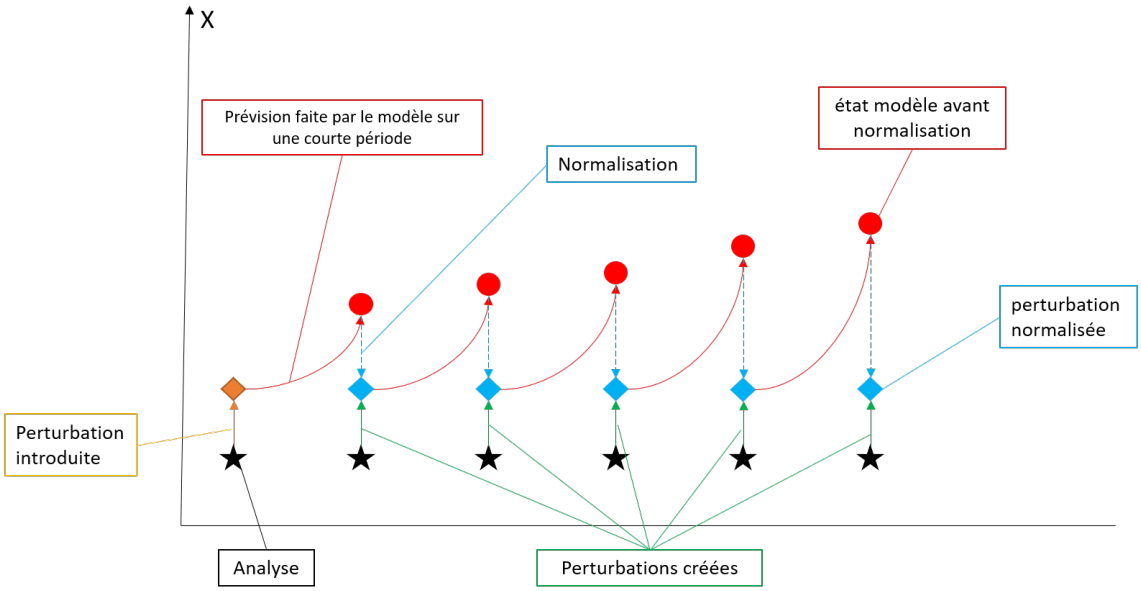


FIGURE 1.11 – *Représentation schématique de la construction des perturbations de l'état initial (fléches vertes) en partant d'une perturbation aléatoire (flèche orange) à l'aide d'un breeding*

car plus simple à construire que les ensembles mentionnés plus hauts. C'était en particulier une méthode très utilisée lorsque les systèmes de prévision ensembliste n'existaient pas en condition opérationnelle. Cette méthode basée sur la comparaison de plusieurs déterministes reste un moyen utilisé par les prévisionnistes pour évaluer l'incertitude qui pèse sur leur prévision, en complément des prévisions d'ensemble. Une agrégation de plusieurs prévisions d'ensemble peut aussi être utilisée, un ensemble d'ensembles en d'autres termes, comme l'illustre le projet THORPEX Interactive Grand Global Ensemble (TIGGE, Bougeault et al. 2010).
Pour conclure sur cet historique, le développement des systèmes de prévision a été poussé d'une part par le raffinement des systèmes déterministes, d'autre part par le développement des systèmes ensemblistes. A l'aube des années 2020, ces deux voies se rejoignent pour aboutir à des systèmes ensemblistes haute résolution. L'augmentation de résolution rapproche cependant de plus en plus les systèmes déterministes de la limite de prévisibilité (Lorenz, 1969). Il est assez probable, par exemple, que les modèles les plus performants aient atteint la limite de prévisibilité pour les trajectoires des cyclones tropicaux (Landsea and Cangialosi, 2018; Plu, 2011). Pour bien d'autres phénomènes météorologiques cependant (intensité des cyclones tropicaux, brouillard, etc), la limite n'est pas atteinte et des progrès en modélisation restent à faire. Cette augmentation de résolution rend donc nécessaire une approche ensembliste y compris à de courtes échéances.

1.3.5 L'arrivée en condition opérationnelle des modèles méso-échelle (années 2000-2020)

Les modèles de grande échelle que nous venons de décrire permettent de fournir des prévisions à l'horizon d'une quinzaine de jours sur l'ensemble du globe et avec des mailles horizontales de l'ordre de 10km. Cette maille ne permet cependant

pas d'appréhender correctement les phénomènes appartenant à la méso-échelle ou à l'échelle aérologique (en particulier la convection) car ils appartiennent à la catégorie des phénomènes sous-mailles pour ces modèles. Mais certains de ces phénomènes sont des enjeux majeurs pour la sécurité des personnes et des biens. Les orages par exemple peuvent détruire des récoltes à cause du vent ou de la grêle ou causer d'importantes inondations. De même le brouillard peut impacter fortement la circulation ou le trafic aérien. Pour correctement appréhender ces phénomènes, il est nécessaire de créer de nouveaux modèles avec une maille plus fine et couvrant des échéances plus courtes car l'horizon de prévisibilité couvre des échéances plus courtes. Pour compenser l'augmentation du coût de calcul, plusieurs centres météorologiques ont développé des modèles à aire limitée qui ne couvrent qu'une petite partie du globe. Ces modèles ont des mailles de l'ordre du kilomètre et peuvent donc résoudre explicitement la convection. Ils peuvent également mieux représenter certains phénomènes de basses couches comme le brouillard. Il faut cependant coupler ces modèles avec d'autres pour leur fournir les conditions aux bords de leurs domaines. C'est ainsi qu'est né le modèle AROME à Météo-France (Seity et al., 2011). Avec une maille horizontale de 1.3 km et 90 niveaux verticaux, il est couplé au modèle grand échelle ARPEGE de Météo-France pour ses conditions au bord. L'état initial du modèle AROME est généré à l'aide d'un schéma 3D-Var, les observations à plus ou moins 30 minutes sont prises en compte. La représentation du cycle de l'eau est plus complexe que dans ARPEGE avec plus de catégories d'hydrométéores ce qui permet de développer un diagnostic de grêle en sortie du modèle. Contrairement à ARPEGE, les précipitations convectives sont résolues et non paramétrisées. AROME dispose également de champs plus réalistes comme le champ de réflectivités simulées (figure 1.12) qui permet une comparaison plus directe avec les données radars que un champ de cumul de pluie (Caumont et al., 2006). AROME a permis entre autres des progrès dans la prévision des épisodes précipitants méditerranéens, phénomènes météorologiques majeurs sur le Sud-Est de la France pour la sécurité des personnes et des biens.

1.3.6 Prévisibilité de l'échelle convective

L'arrivée de ces modèles plus fines échelles pose les mêmes questions que pour les modèles grandes échelles concernant leur horizon de prévisibilité et la façon de traiter les incertitudes inhérentes à la prévision. Ces modèles s'intéressant à l'échelle convective, c'est donc celle-ci qui va être étudiée de près. Les travaux en particulier de Hohenegger and Schar (2007a) ont rapidement montré que les approches retenues pour traiter les incertitudes à l'échelle globale ne peuvent pas être appliquées pour traiter l'échelle convective qui est fondamentalement différente. Par exemple, l'approximation linéaire tangente, nécessaire en pratique pour le calcul des vecteurs singuliers et qui est valable jusqu'à environ +50h pour un modèle global n'est valable que 1,5h pour un modèle à échelle convective (Hohenegger and Schar, 2007a). Ces travaux et d'autres (Hohenegger and Schar, 2007b; Wu et al., 2013) ont montré que la convection humide joue un rôle majeur dans la prévisibilité à ces échelles en introduisant d'importantes non-linéarités ce qui explique en partie pourquoi l'approximation linéaire tangente devient rapidement fausse.

Cependant, la prévisibilité généralement mauvaise en présence de convection

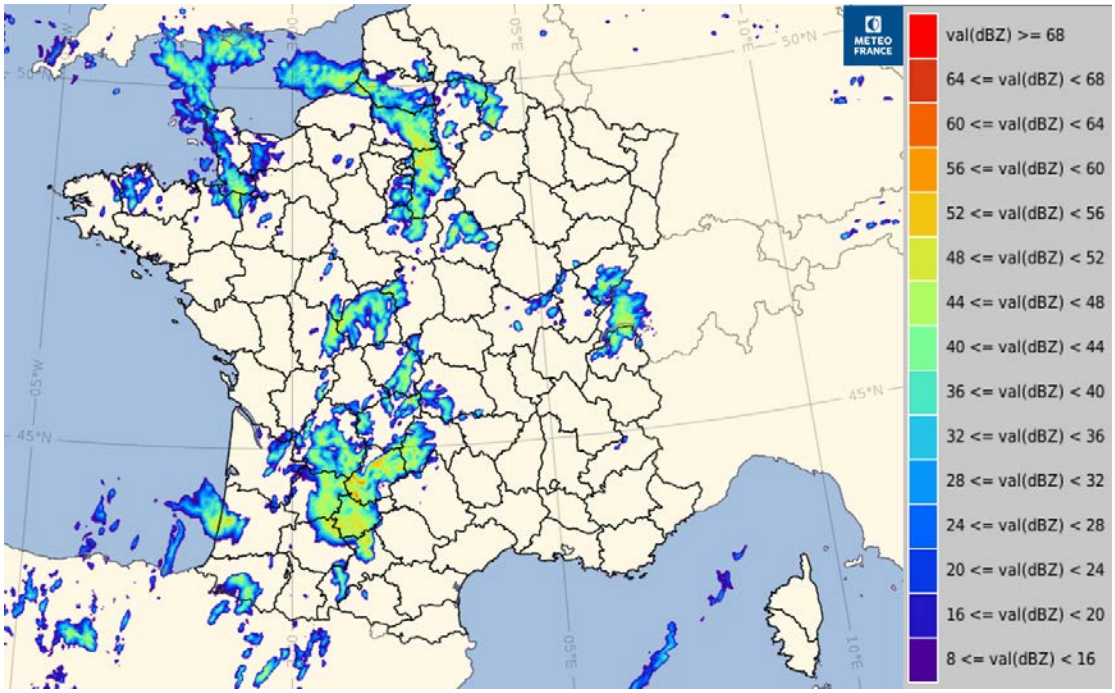


FIGURE 1.12 – *Champ de réflectivités simulées en dBz du modèle AROME du 3 juin 2022 à 12h UTC pour le 4 juin 2022 à 18h UTC. Plus les couleurs sont chaudes, plus les réflectivités sont intenses (Source : Météo-France, outil Synopsis).*

humide peut varier sensiblement d'un cas à l'autre selon les connexions entre cette convection et la plus grande échelle (Hohenegger et al., 2006). Le principe de la cascade de Richardson (1922) où les erreurs de la petite échelle peuvent contaminer la grande et inversement est en effet à prendre en compte (figure 1.4). Les travaux récents sur des lignes orageuses (Durran and Weyn, 2016) ou sur des situations convectives de façon plus générale (Surcel et al., 2015) illustrent parfaitement ces interactions entre les échelles et la nécessité d'une approche ensembliste pour traiter ces problématiques. Cependant, les travaux de Surcel et al. (2015) montrent que même en s'appuyant sur une prévision d'ensemble, l'information aux échelles inférieures à 100km en cas de situations orageuses est très vite perdue. Les membres sont en effet totalement décorrélés après seulement 2h d'échéance. Mais les scores utilisés sont sensibles aux erreurs de position et on ne peut donc pas en déduire que toute information à des échelles plus fines est inutilisable. En particulier, Surcel et al. (2015) rappellent que les caractéristiques des orages simulés sur une zone peuvent être une information intéressante en contexte opérationnel. Ce point sera de nouveau abordé dans la suite de ce manuscrit.

Dans ce contexte où l'horizon de prévisibilité à l'échelle convective peut ne pas excéder quelques heures, des prévisions d'ensemble à échelle convective et à aire limitée ont été développées avec des méthodes de perturbation qui diffèrent des prévisions d'ensemble à l'échelle globale. Nous allons plus particulièrement détailler dans la sous-section suivante les caractéristiques de cet ensemble à Météo-France, nommé Prévision d'Ensemble AROME (PE-AROME).

1.4 La Prévision d'Ensemble AROME (PE-AROME)

1.4.1 Méthodes de perturbation

Le développement de la PE-AROME (Bouttier et al., 2016) a commencé rapidement après celui du modèle AROME. Les membres PE-AROME emploient le même modèle AROME mais des perturbations différentes sont ajoutées pour chaque membre. Ces perturbations introduites dans les membres PE-AROME sont présentées dans cette sous-section. Différentes méthodes de perturbation sont appliquées, on peut les diviser en quatre groupes : perturbation de l'état initial, de la surface, la prise en compte de l'erreur modèle et enfin les conditions aux bords du domaine.

Perturbation de l'état initial

L'état initial de l'atmosphère n'étant pas connu parfaitement, il constitue une source d'incertitude. Pour estimer cette incertitude, une solution possible consiste à générer plusieurs états initiaux légèrement perturbés à l'aide d'un ensemble d'assimilations (EDA). C'est le rôle de l'Assimilation d'Ensemble AROME (AEARO).

L'AEARO est un ensemble de 25 membres avec une résolution horizontale de 3,25 km et 90 niveaux verticaux. Un ensemble d'analyses perturbées est obtenu à travers des perturbations implicites des ébauches (Raynaud and Bouttier, 2016; Montmerle et al., 2018) et des perturbations explicites des observations selon une distribution gaussienne (Brousseau et al., 2011). Chaque ébauche perturbée est couplée aux bords aux prévisions de l'Assimilation d'Ensemble ARPEGE (Berre et al., 2007, AEARP). Un ensemble de N états complets ($\tilde{x}_{a,i}, i \in [1, N]$) est ainsi obtenu. Ces états sont alors recentrés autour de leur moyenne ($\bar{x}_a$) et ajoutés à l'analyse 3D-Var AROME (x_a^{ARO}) pour obtenir les perturbations de l'AEARO ($x_{a,i}^{AEARO}$) :

$$x_{a,i}^{AEARO} = x_a^{ARO} + \tilde{x}_{a,i} - \bar{x}_a \tag{1.7}$$

Ce recentrage est fait pour tirer profit de l'analyse AROME qui est elle à 1.3 km. La PE-AROME disposant de moins de membres que l'AEARO, les perturbations utilisées par la PE-AROME sont tirées aléatoirement. Pour les premières versions de la PE-AROME, en l'absence d'une AEARO, des perturbations issues de la PEARP étaient interpolées sur le domaine à aire limitée.

Perturbation de la surface

La paramétrisation modèle de la surface et de ses échanges avec l'atmosphère joue un rôle important à l'échelle convective. On peut citer par exemple le rôle primordial de l'humidité des sols dans la prévision des précipitations convectives (Aligo et al., 2007). L'incertitude de la surface est représentée en perturbant certains paramètres du modèle SURFEX (indice de végétation, albedo, SST, flux de surface, humidité des sols, ...). Bouttier et al. (2016) détaillent ces perturbations.

Erreurs modèle

Une autre source d'incertitude réside dans le modèle lui-même car la résolution des équations ou la représentation des phénomènes sous-mailles (trop fins pour la taille des points de grille du modèle) restent imparfaites. Le choix est fait à Météo-France pour la PE-AROME de perturber les tendances physiques dans le modèle pour le vent, la température et l'humidité. On parle de la méthode « Stochastically Perturbed Parameterization Tendencies » (Palmer et al., 2009; Vié et al., 2012; Bouttier et al., 2012; Wastl et al., 2019, SPPT). D'autres méthodes proposent de perturber directement certains paramètres des paramétrisations physiques comme les méthodes Perturbed Parameter (Reynolds et al., 2011, PP), Random Parameter (Baker et al., 2014, RP) ou Stochastic Parameter Perturbation (Jankov et al., 2019, SPP). Ces méthodes sont d'ailleurs explorées pour la PE-AROME (Wimmer, 2021). Une autre possibilité est d'utiliser plusieurs physiques différentes dans les membres en changeant complètement le schéma de représentation de certains phénomènes comme la convection. Ce type d'approche peut cependant être assez coûteux à entretenir car nécessite de maintenir plusieurs schémas différents. Cette méthode est appliquée pour la PEARP à Météo-France (Descamps et al., 2015).

Conditions aux bords

Les modèles PE-AROME et AROME sont des modèles à aire limitée. Il est donc nécessaire de leur associer un modèle coupleur de plus grande échelle pour renseigner les conditions aux bords du domaine. Le choix du modèle coupleur est alors une source d'incertitude et a un fort impact sur le modèle à aire limitée en particulier après les premières échéances. La méthode retenue pour la PE-AROME consiste à coupler chaque membre avec un de ceux de la PEARP.

La Prevision d'Ensemble ARPEGE (Descamps et al., 2015, PEARP) est un modèle de prévision d'ensemble grande échelle avec 35 membres dont un de contrôle qui n'est pas perturbé. La PEARP couvre tout le globe mais utilise une maille étirée dont la taille varie selon la position. Elle est d'environ 5 km sur la France mais grimpe à 24 km aux antipodes, avec 105 niveaux verticaux. Les perturbations sur les conditions initiales sont créées à l'aide de l'AEARP (Berre et al., 2007) composée de 50 membres. En plus des perturbations issues de l'AEARP, des vecteurs singuliers sont également appliqués. Ces vecteurs sont calculés sur plusieurs domaines comme l'Europe, l'Atlantique Nord ou la bande intertropicale. L'erreur de modélisation est représentée depuis 2022 à l'aide d'une approche multiphysique avec 2 schémas de convection et d'une méthode RP.

Comme la PEARP possède plus de membres que la PE-AROME, il est nécessaire d'établir une méthode de sélection des membres PEARP. Une méthode de classification est utilisée à partir des champs de vents horizontaux, du géopotentiel et de l'humidité relative. Ces paramètres sont pris en compte à 3 niveaux (500,700,850 hPa) et à plusieurs échéances (+12,18,24,30,36 et 42h). La méthode est détaillée dans l'article de Bouttier and Raynaud (2018).

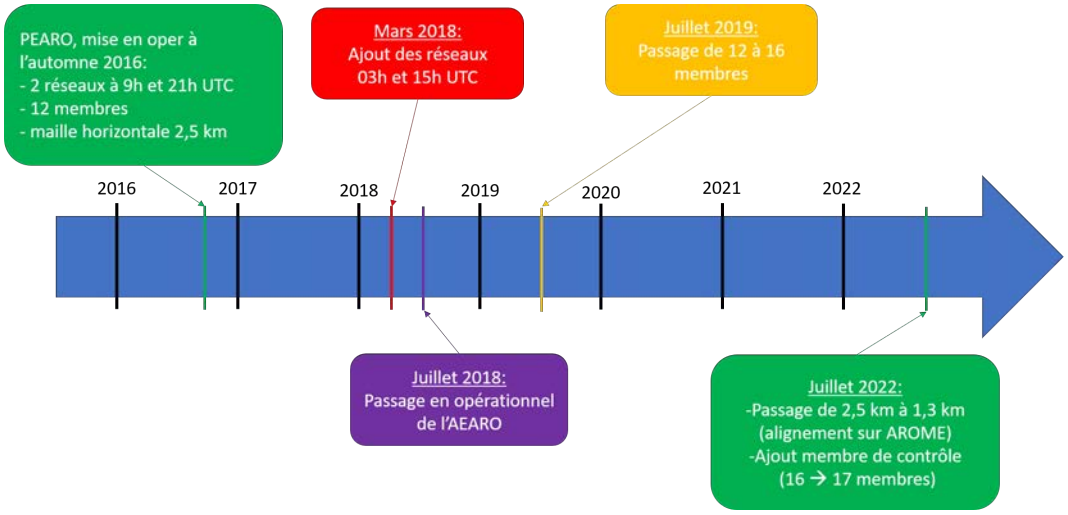


FIGURE 1.13 – *Principales évolutions du système ensembliste PE-AROME depuis sa mise en opérationnel à l'automne 2016 jusqu'à l'été 2022. Cette liste n'est pas exhaustive et n'intègre pas en particulier le changement de méthode de classification pour choisir les coupleurs PEARP ou l'assimilation de nouvelles données pour l'AEARO.*

1.4.2 Évolutions de la PE-AROME de 2016 à 2022

La PE-AROME est devenue opérationnelle à l'automne 2016. Elle se composait alors de 12 membres, la maille horizontale était de 2.5 km (contre 1,3 km pour AROME) avec 90 niveaux verticaux. Ce système ensembliste était lancé deux fois par jour à 09h et 21h UTC. Depuis 2016, la PE-AROME a bénéficié de nombreuses améliorations qui sont représentées sur la figure 1.13. Ces améliorations se poursuivent aujourd'hui avec en particulier le développement d'une PE-AROME haute résolution à 1.3 km, s'alignant ainsi sur AROME, ce qui conduit à le considérer maintenant comme le membre non perturbé de la PE-AROME (aussi appelé membre de contrôle). Le domaine couvert par AROME et la PE-AROME correspond à la grille EURW1S40 (figure 1.14). Cette nouvelle version est devenue opérationnelle en juin 2022.

Bibliographie

E. A. Aligo, W. A. Gallus, and M. Segal. Summer Rainfall Forecast Spread in an Ensemble Initialized with Different Soil Moisture Analyses. 22 :299–314, 2007.

L. H. Baker, A. C. Rudd, S. Migliorini, and R. N. Bannister. Representation of model error in a convective-scale ensemble prediction system. *Nonlinear Processes in Geophysics*, 21(1) :19–39, 2014. doi : 10.5194/npg-21-19-2014.

P. Bauer, A. Thorpe, and G. Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525 :47–55, 2015. doi : 10.1038/nature14956.

L. Berre, O. Pannekoucke, G. Desroziers, S. E. Stefanescu, B. Chapnik, and L. Raynaud. A variational assimilation ensemble and the spatial filtering of its error covariances : increase of sample size by local spatial averaging. In *ECMWF*

FIGURE 1.14 – *Domaine couvert par la PE-AROME actuelle (domaine EURW1S40). Ce domaine est le même pour AROME. Le domaine s'étend en latitude de 37.5°N à 55,4°N et de 12° W à 16°E pour une grille de sortie de 717 par 1121 points de grille à 2,5km de résolution.*

Workshop on Flow-dependent aspects of data assimilation, 11-13 June 2007, pages 151–168, 2007.

P. Bougeault, Z. Toth, C. Bishop, B. Brown, D. Burridge, D. H. Chen, B. Ebert, M. Fuentes, T. M. Hamill, K. Mylne, J. Nicolau, T. Paccagnella, Y. Park, D. Parsons, B. Raoult, D. Schuster, P. D. Dias, R. Swinbank, Y. Takeuchi, W. Tennant, L. Wilson, and S. Worley. The THORPEX Interactive Grand Global Ensemble. *Bull. Amer. Meteor. Soc.*, 91(8) :1059–1072, 2010. doi : 10.1175/2010BAMS2853.1.

F. Bouttier and P. Courtier. Data Assimilation Concepts and Methods, 2002. URL `https://www.ecmwf.int/node/16928`.

F. Bouttier and L. Raynaud. Clustering and selection of boundary conditions for limited-area ensemble prediction. *Quart. J. Roy. Meteor. Soc.*, 144 :2381–2391, 2018. doi : 10.1002/qj.3304.

F. Bouttier, B. Vié, O. Nuissier, and L. Raynaud. Impact of stochastic physics in a convection-permitting ensemble. *Mon. Wea. Rev.*, 140 :3706–3721, 2012. doi : 10.1175/MWR-D-12-00031.1.

F. Bouttier, L. Raynaud, O. Nuissier, and B. Ménétrier. Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. *Quart. J. Roy. Meteor. Soc.*, 142 :390–403, 2016. doi : 10.1002/qj.2622.

F. P. Bretherton. Momentum transport by gravity waves. *Quart. J. Roy. Meteor. Soc.*, 95(404) :213–243, 1969. doi : https://doi.org/10.1002/qj.49709540402.

P. Brousseau, L. Berre, F. Bouttier, and G. Desroziers. Background-Error Covariances for a Convective-Scale Data-Assimilation System : AROME–France 3D-Var. *Quart. J. Roy. Meteor. Soc.*, 137 :409–422, 2011. doi : 10.1002/qj.750.

R. Buizza. Introduction to the special issue on "25 years of ensemble forecasting". *Quart. J. Roy. Meteor. Soc.*, 145(S1) :1–11, 2019. doi : https://doi.org/10.1002/qj.3370.

R. Buizza, P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei. A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev.*, 133(5) :1076–1097, 2005. doi : 10.1175/MWR2905.1.

C. Calas. *Concepts et méthodes pour le météorologiste - Tome 1, les savoirs : modèles conceptuels et données disponibles.* Météo-France, 2013.

A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen. Data assimilation in the geosciences : An overview of methods, issues, and perspectives. *WIREs Climate Change*, 9(5) :e535, 2018. doi : https://doi.org/10.1002/wcc.535.

O. Caumont, V. Ducrocq, G. Delrieu, M. Gosset, J. Pinty, J. Parent du Châtelet, H. Andrieu, Y. Lemaître, and G. Scialom. A Radar Simulator for High-Resolution Nonhydrostatic Models. *J. Atmos. Oceanic Technol.*, 23 :1049–1067, 2006. doi : 10.1175/JTECH1905.1.

J. Charney. On the scale of atmospheric motions. *Geofys. Publ.*, (2) :17, 1948.

M. Dameris. Tropopause. In *Encyclopedia of Atmospheric Sciences*, pages 2345–2348. 2003. doi : https://doi.org/10.1016/B0-12-227090-8/00418-8.

L. Descamps, C. Labadie, A. Joly, E. Bazile, P. Arbogast, and P. Cébron. PEARP, the Météo-France short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, 141(690) :1671–1685, 2015. doi : https://doi.org/10.1002/qj.2469.

G. M. Dunnavan and J. W. Diercks. An Analysis of Super Typhoon Tip (October 1979). *Mon. Wea. Rev.*, 108(11) :1915–1923, 1980. doi : 10.1175/1520-0493(1980)108<1915:AAOSTT>2.0.CO;2.

D. R. Durran and J. A. Weyn. Thunderstorms Do Not Get Butterflies. *Bull. Amer. Meteor. Soc.*, 97(2) :237–243, 2016. doi : 10.1175/BAMS-D-15-00070.1.

E. T. Eady. *The Quantitative Theory of Cyclone Development*, pages 464–469. 1951. doi : 10.1007/978-1-940033-70-9_39.

E. S. Epstein. Stochastic dynamic prediction. *Tellus*, 21(6) :739–759, 1969. doi : https://doi.org/10.1111/j.2153-3490.1969.tb00483.x.

J. R. Eyre, S. J. English, and M. Forsythe. Assimilation of satellite data in numerical weather prediction. Part I : The early years. *Quart. J. Roy. Meteor. Soc.*, 146(726) :49–68, 2020. doi : https://doi.org/10.1002/qj.3654.

J. R. Eyre, W. Bell, J. Cotton, S. J. English, M. Forsythe, S. B. Healy, and E. G. Pavelin. Assimilation of satellite data in numerical weather prediction. Part II : Recent years. *Quart. J. Roy. Meteor. Soc.*, 148(743) :521–556, 2022. doi : https://doi.org/10.1002/qj.4228.

C. L. Fefferman. *Existence and Smoothness of the Navier–Stokes equation*, pages 57–62. 2006.

R. D. Fletcher. The general circulation of the tropical and equatorial atmosphere. *J. Atmos. Sci.*, 2(3) :167–174, 1945. doi : 10.1175/1520-0469(1945)002<0167:TGCOTT>2.0.CO;2.

G. Hadley. Concerning the cause of the general trade-winds. *Philosophical Transactions of the Royal Society of London*, 39(437) :58–62, 1735. doi : 10.1098/rstl.1735.0014.

C. Hohenegger and C. Schar. Atmospheric Predictability at Synoptic Versus Cloud-Resolving Scales. *Bull. Amer. Meteor. Soc.*, 88(11) :1783–1794, 2007a. doi : 10.1175/BAMS-88-11-1783.

C. Hohenegger and C. Schar. Predictability and Error Growth Dynamics in Cloud-Resolving Models. *J. Atmos. Sci.*, 64(12) :4467–4478, 2007b. doi : 10.1175/2007JAS2143.1.

C. Hohenegger, D. Lüthi, and C. Schär. Predictability Mysteries in Cloud-Resolving Models. *Mon. Wea. Rev.*, 134(8) :2095–2107, 2006. doi : 10.1175/MWR3176.1.

I. Jankov, J. Beck, J. Wolff, m. Harrold, J. B. Olson, T. Smirnova, C. Alexander, and J. Berner. Stochastically Perturbed Parameterizations in an HRRR-Based Ensemble. *Mon. Wea. Rev.*, 147(1) :153–173, 2019. doi : 10.1175/MWR-D-18-0092.1.

J. Lacarra and O. Talagrand. Short range evolution of small perturbation in a barotropic model. *Tellus*, 40A :81–95, 1988. doi : 10.1111/j.1600-0870.1988.tb00408.x.

C. W. Landsea and J. P. Cangialosi. Have We Reached the Limits of Predictability for Tropical Cyclone Track Forecasting ? *Bull. Amer. Meteor. Soc.*, 99(11) : 2237–2243, 2018. doi : 10.1175/BAMS-D-17-0136.1.

C. E. Leith. Theoretical Skill of Monte Carlo Forecasts. *Mon. Wea. Rev.*, 102(6) : 409–418, 1974. doi : 10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2.

J. Lewis. Roots of ensemble forecasting. *Mon. Wea. Rev.*, 133(7) :1865 – 1885, 2005. doi : 10.1175/MWR2949.1.

O. Lorentin. ANTILOPE : hourly rainfall analysis over france merging radar and rain gauges data. In *11th International Precipitation Conference*, 2013.

E. N. Lorenz. Deterministic Nonperiodic Flow. *J. Atmos. Sci.*, 20(2) :130–141, 1963. doi : 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

E. N. Lorenz. The predictability of a flow which possesses many scales of motion. *Tellus*, 21(3) :289–307, 1969. doi : https://doi.org/10.1111/j.2153-3490.1969.tb00444.x.

E. N. Lorenz. Atmospheric predictability experiments with a large numerical model. *Tellus*, 34 :505–513, 1982. doi : 10.3402/tellusa.v34i6.10836.

J. Lu, G. A. Vecchi, and T. Reichler. Expansion of the Hadley cell under global warming. *Geophys. Res. Lett.*, 34(6), 2007. doi : https://doi.org/10.1029/2006GL028443.

P. Lynch and G. Moor. Les origines de la prévision numérique du temps et de la modélisation climatique. *La Météorologie*, 8(63) :14–24, 2008. doi : 10.4267/2042/21887.

S. Malardel. *Fondamentaux de météorologie : à l'école du temps.* Cepadues, 2009.

J. Marshall and W. Palmer. The distribution of raindrops with size. *J. Atmos. Sci.*, 5 :165–166, 1948. doi : 10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2.

F. Molteni, R. Buizza, T. N. Palmer, and T. Petroliagis. The ECMWF ensemble prediction system : Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, 122(529) :73–119, 1996.

T. Montmerle, Y. Michel, E. Arbogast, B. Ménétrier, and P. Brousseau. A 3D ensemble variational data assimilation scheme for the limited-area AROME model : Formulation and preliminary results. *Quart. J. Roy. Meteor. Soc.*, 144 : 2196–2215, 2018. doi : 10.1002/qj.3334.

E. Novikov. On the problem of predictability of synoptic processes. *Izv. Acad. Sci. USSR*, (11) :1209–1211, 1959.

I. Orlanski. A Rational Subdivision of Scales for Atmospheric Processes. *Bull. Amer. Meteor. Soc.*, 56(5) :527–530, 1975.

T. N. Palmer, R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer. Stochastic parametrization and model uncertainty. 2009.

M. B. Pereira and L. Berre. The Use of an Ensemble Approach to Study the Background Error Covariances in a Global NWP Model. *Mon. Wea. Rev.*, 134 (9) :2466–2489, 2006. doi : 10.1175/MWR3189.1.

M. Plu. A New Assessment of the Predictability of Tropical Cyclone Tracks. *Mon. Wea. Rev.*, 139(11) :3600–3608, 2011. doi : 10.1175/2011MWR3627.1.

L. Raynaud and F. Bouttier. Comparison of initial perturbation methods for ensemble prediction at convective scale. *Quart. J. Roy. Meteor. Soc.*, 142(695) : 854–866, 2016. doi : https://doi.org/10.1002/qj.2686.

C. A. Reynolds, J. A. Ridout, and J. G. Mclay. Examination of parameter variations in the U. S. navy global ensemble. *Tellus A : Dynamic Meteorology and Oceanography*, 63(5) :841–857, 2011. doi : 10.1111/j.1600-0870.2011.00532.x.

L. F. Richardson. Weather prediction by numerical process. *Quart. J. Roy. Meteor. Soc.*, 48(203) :282–284, 1922. doi : https://doi.org/10.1002/qj.49704820311.

C.-G. Rossby. Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action. *J. Mar. Res.*, 2 :38–55, 1939.

D. Ruelle and F. Takens. On the nature of turbulence. *Communications in Mathematical Physics*, 20 :167–192, 1971. doi : 10.1007/BF01646553.

E. Saltikoff, K. Friedrich, J. Soderholm, K. Lengfeld, B. Nelson, A. Becker, R. Hollmann, B. Urban, M. Heistermann, and C. Tassone. An Overview of Using Weather Radar for Climatological Studies : Successes, Challenges, and Potential. *Bull. Amer. Meteor. Soc.*, 100(9) :1739–1752, 2019. doi : 10.1175/BAMS-D-18-0166.1.

Y. Seity, P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson. The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, 139 :976–991, 2011. doi : 10.1175/2010MWR3425.1.

D. J. Stensrud. *Parameterization schemes : keys to understanding numerical weather prediction models*. Cambridge University Press, 2009.

M. Surcel, I. Zawadzki, and M. K. Yau. A Study on the Scale Dependence of the Predictability of Precipitation Patterns. *J. Atmos. Sci.*, 72(1) :216–235, 2015. doi : 10.1175/JAS-D-14-0071.1.

P. D. Thompson. Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. *Tellus*, 9(3) :275–295, 1957. doi : https://doi.org/10.1111/j.2153-3490.1957.tb01885.x.

S. Tracton and E. Kalnay. Operational Ensemble Prediction at the National Meteorological Center : Practical Aspects. 8(3) :379–398, 1993. doi : 10.1175/1520-0434(1993)008<0379:OEPATN>2.0.CO;2.

R. Trapp, S. Tessendorf, E. Godfrey, and H. Brooks. Tornadoes from squall lines and bow echoes. part I : Climatological distribution. *Wea. Forecasting*, 20 : 23–34, 2005. doi : 10.1175/WAF-835.1.

B. Vié, G. Molinié, O. Nuissier, B. Vincendon, V. Ducrocq, F. Bouttier, and E. Richard. Hydro-meteorological evaluation of a convection-permitting ensemble prediction system for Mediterranean heavy precipitating events. *Natural hazards and earth system sciences*, 12 :2631–2645, 2012. doi : 10.5194/nhess-12-2631-2012.

C. Wastl, Y. Wang, A. Atencia, and C. Wittmann. A Hybrid Stochastically Perturbed Parametrization Scheme in a Convection-Permitting Ensemble. *Mon. Wea. Rev.*, 147(6) :2217–2230, 2019. doi : 10.1175/MWR-D-18-0415.1.

M. Wimmer. *Représentation de l'erreur de modélisation dans le système de prévision d'ensemble régional PEARO*. PhD thesis, Université Paul Sabatier - Toulouse III, 2021.

D. Wu, Z. Meng, and D. Yan. The predictability of a squall line in South China on 23 April 2007. *Adv. Atmos. Sci.*, 30(2) :485–502, 2013. doi : 10.1007/s00376-012-2076-x.

Chapitre 2

Méthodes de visualisation des systèmes ensemblistes

Le chapitre précédent a montré la nécessité d'adopter une vision ensembliste à toutes les échelles afin d'explorer au mieux l'espace des possibles pour aboutir à une prévision robuste qui renseigne sur ses incertitudes associées. La vision ensembliste avec non pas une mais plusieurs prévisions offre une quantité d'information plus riche mais rend son utilisation plus complexe. Un effort de synthétisation est en effet nécessaire pour extraire l'information pertinente et établir des prévisions. Pour réaliser cette synthèse, l'approche la plus simple est la vision "vignette" où chaque membre est visualisé un à un et l'effort de synthèse est porté par l'expert ou l'utilisateur. Cette représentation sous forme de vignettes est illustrée figure 2.1 sur une situation avec un risque d'orage pouvant provoquer d'importantes inondations. Le département du Gard concentre le principal risque d'orage sévère avec des couleurs rouges vives pour la plupart des membres (membres 2,4,6,7,8,9,10,12). D'autres membres centrent cette convection sur les Bouches du Rhône (membres 1,3,5,11,13,16) ou d'autres ne prévoient pas de convection sévère (membres 14 et 15). Ce travail de synthèse a été fait ici rapidement sur une échéance et un paramètre. Or pour appréhender une situation il est nécessaire de considérer de nombreux paramètres et leur évolution temporelle. Cet effort de synthèse peut alors vite devenir chronophage, fastidieux et impossible à faire en temps contraint comme c'est le cas en condition opérationnelle. Par ailleurs, les évolutions futures des prévisions d'ensemble conduisent à envisager un nombre croissant de membres dans les années à venir, ce qui rendra l'exploitation directe des membres encore plus difficile.

Pour ces raisons mais également pour rendre ces prévisions plus simples d'utilisation, plusieurs méthodes de visualisation ont été développées et sont présentées dans ce chapitre. Ces méthodes de visualisation ont pour objectif de synthétiser une information qui était déjà à 4 dimensions avec l'espace et le temps et multiparamètre en vision déterministe. La vision ensembliste ajoute une dimension supplémentaire avec plusieurs membres ce qui ajoute un niveau de complexité supplémentaire avec plusieurs problèmes à résoudre : Comment représenter cette complexité en 2D ou 3D ? Et comment représenter l'incertitude, apport majeur de la vision ensembliste ? La complexité de ce problème explique pourquoi cette étape de mise en forme de l'information occupe une place importante dans la chaine de

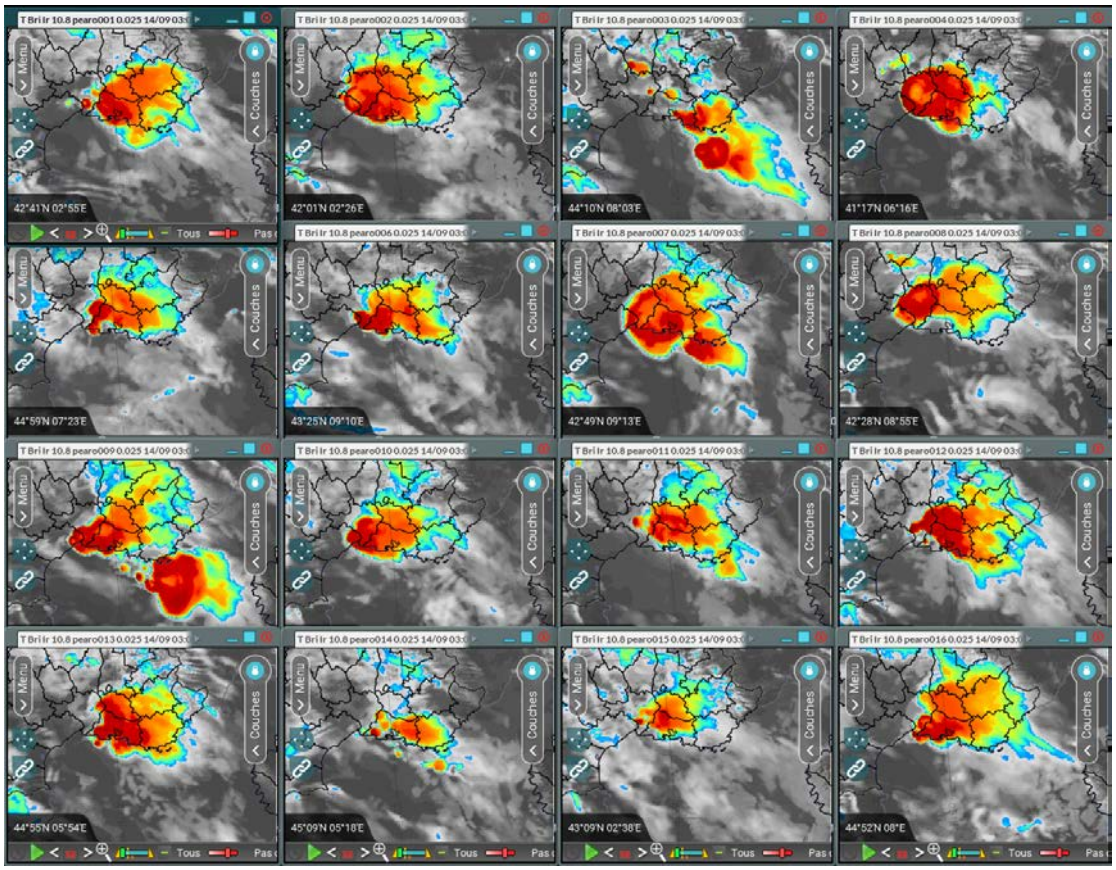


FIGURE 2.1 – *Représentation sous forme de vignettes des 16 membres de la PE-AROME à un instant donné. Les membres sont classés de haut en bas et de gauche à droite (membre 1 en haut à gauche et membre 16 en bas à droite). Les vignettes représentent la prévision de la température de brillance qui permet d'estimer la hauteur des sommets convectifs et donc l'intensité de la convection. Plus la température est froide, plus la couleur tend vers le rouge et plus le sommet est haut dans la troposphère. Prévision du 14 septembre 2021 à 03h UTC pour 08h UTC sur le Sud-Est de la France (Source : Météo-France, outil Synopsis).*

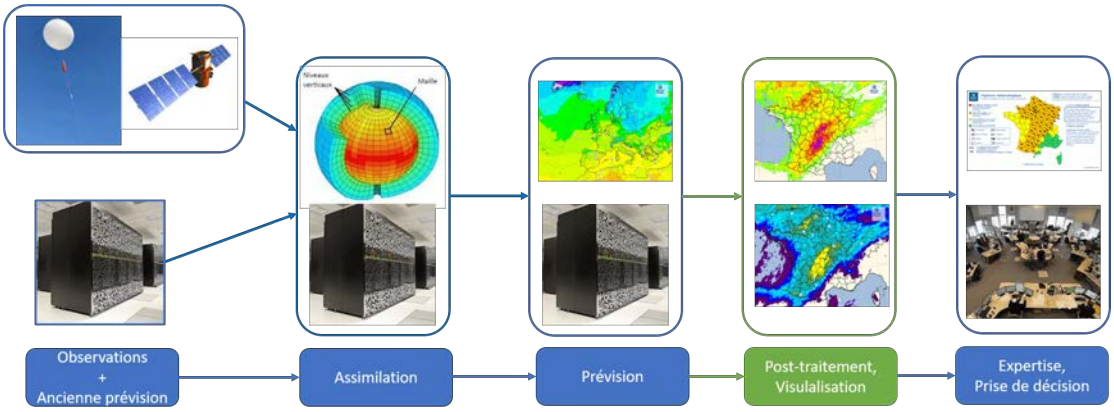


FIGURE 2.2 – *Représentation schématique de la chaîne de prévision. Le chapitre 2 s'intéresse à la partie en vert dédiée aux post-traitements des modèles (représentés ici par deux cartes qui seront détaillées dans ce chapitre).*

prévision (figure 2.2). Ce chapitre traite des méthodes de visualisation et non de l'ensemble des méthodes de post-traitement. C'est pourquoi les méthodes pour corriger les sorties modèles comme les Adaptations Statistiques (AS) ne sont pas présentées dans ce chapitre, même si celles-ci occupent également une place importante dans la chaine de prévision. En effet, ces AS sont souvent utilisées à la place des données brutes en sortie des modèles déterministes et ensemblistes (Vannitsem et al., 2021). Ce chapitre ne traitera pas non plus les post-traitements dédiés à des modèles applicatifs en aval pour différents domaines météo-sensibles comme l'agriculture (Aleksovska et al., 2021). Une première partie détaillera dans ce chapitre les méthodes de visualisation à l'échelle locale, en un point précis à la surface du globe. Une deuxième partie s'intéressera aux synthèses spatialisées pour des échelles régionale ou globale, et une dernière partie présentera des méthodes de visualisation pour des phénomènes particuliers comme les cyclones tropicaux.

2.1 Synthétisation en un point

Cette forme de synthèse est utilisée lorsque la prévision a pour objectif d'être établie de façon très localisée (un parc éolien, un aéroport, une ville par exemple). La façon la plus courante pour représenter l'information probabiliste est alors sous forme de panaches (figure 2.3, haut). Sur cet exemple pour l'aéroport de Toulouse-Blagnac, le panache permet de visualiser l'augmentation de l'incertitude avec les jours de prévision. La prévisibilité semble très bonne jusqu'au vendredi 10 juin avec des prévisions très similaires. L'incertitude augmente jusqu'au lundi 20 juin pour la température maximale qui est prévue entre 20 et 40°C si on considère les membres les plus extrêmes de la distribution pour cette journée. Si certains seuils particuliers sont scrutés, comme les seuils de canicule en été, l'information issue de la PE peut être représentée sous forme de probabilité de dépasser (ou de ne pas atteindre) ce seuil (figure 2.3, bas). Il est par exemple probable à 86% dans cet exemple que la température maximale dépasse les 30°C le 11 puis le 12 juin.

De manière similaire, les boîtes à moustaches sont aussi utilisées pour représenter l'information en un point (figure 2.4). Pour chaque jour, ces boîtes à moustaches représentent la médiane des prévisions (trait central) le 1er et 3e quartile

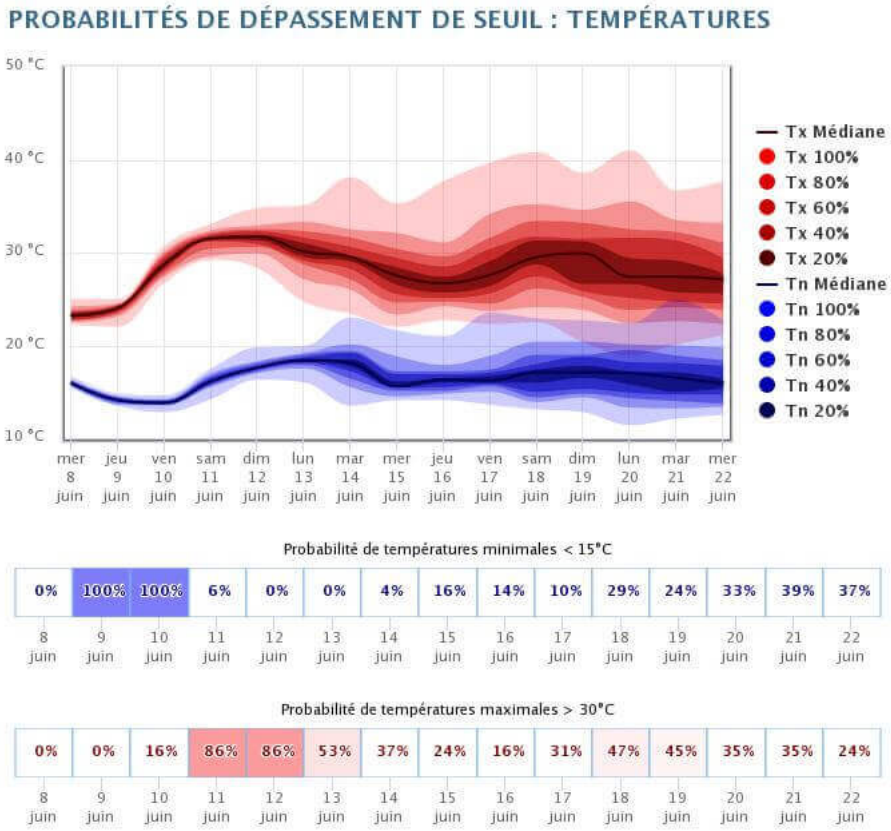


FIGURE 2.3 – *Représentation sous forme de panaches des prévisions à 15 jours de température minimales (bleu) et maximales (rouge) pour l'aéroport Toulouse-Blagnac (prévision du 8 au 22 juin 2022). Le trait continu plus épais représente la médiane des prévisions. Les enveloppes de teintes de couleurs différentes représentent les prévisions faites par les 20% des membres les plus proches de la médiane pour la couleur la plus foncée jusqu'à 100% des membres pour l'enveloppe la plus claire qui correspond aux prévisions les plus éloignés de la médiane des deux côtés de la distribution. Plus bas, représentation du risque pour chaque jour que la température minimale soit inférieure à 15 degrés et pour terminer du risque que la température maximale dépasse les 30 degrés. Prévision faite à partir de l'ensemble du CEPMMT. (Source : Extranet Météo-France)*

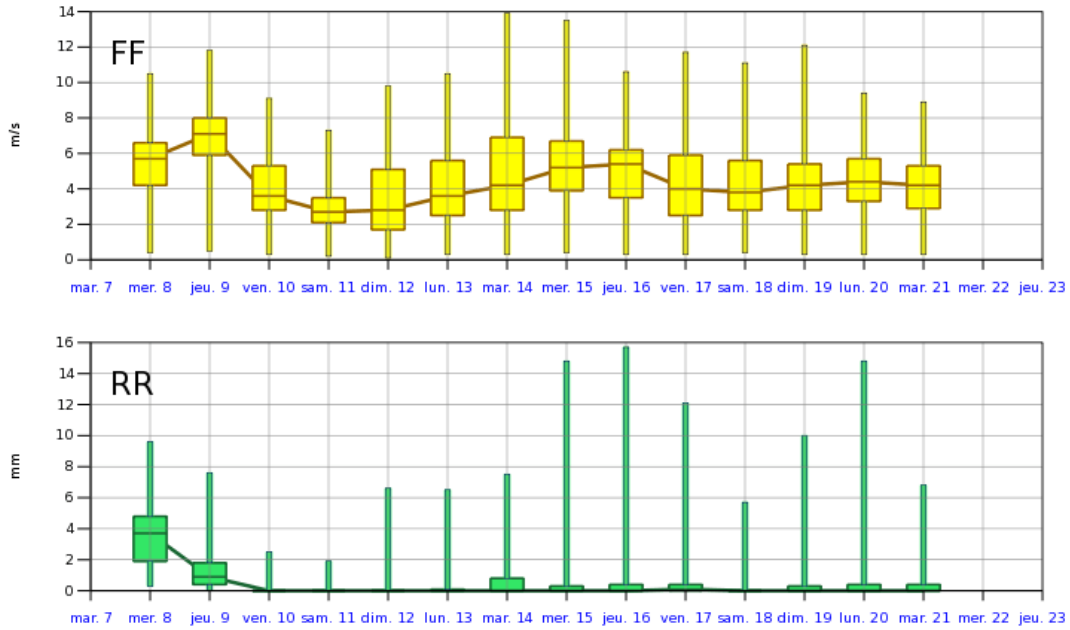


FIGURE 2.4 – *Représentation sous forme de boîtes à moustaches des prévisions ensemblistes du modèle européen (CEPMMT) pour l'aéroport de Toulouse-Blagnac du 8 au 22 juin 2022 pour la force du vent (jaune) en* $m \cdot s^{-1}$ *et les cumuls de pluie sur 24h (vert) en mm. (Source :* `previsibilite.meteo.fr`*, Météo-France)*

de la distribution (rectangle) ainsi que les valeurs extrêmes (extrémités des traits). On peut alors déduire de la figure 2.4 un risque faible de pluie pour les journées du 11 et 12 juin. A l'inverse, le risque de pluie pour la journée du 8 juin est très fort mais avec des quantités modestes ne dépassant pas les 10 mm. Le risque de pluie pour le 15 juin est plus faible mais les quantités pourraient être plus importante, jusqu'à 16 mm d'après l'ensemble du CEPMMT.

Ces représentations ne s'intéressent cependant qu'à un point de grille précis sur la surface. Il est souvent intéressant de regarder aussi l'ensemble de la colonne atmosphérique au dessus de la surface en particulier en situation convective. Le CEPMMT, dont l'ensemble sera nommé Ensemble Prediction System (EPS) ou ensemble européen dans la suite de ce chapitre, a développé pour cela un produit dont une partie est représentée figure 2.5 (Ihasz and Tajti, 2011). Il est fréquent de représenter la température et la température du point de rosée [1] selon les niveaux de pression afin de repérer un environnement instable (de façon simpliste les profils penchent vers la gauche dans ce cas) ou également si certains niveaux sont plus ou moins humides ce qui s'évalue avec la proximité des 2 profils (une température égale à la température du point de rosée indique un taux d'humidité de 100%). Cette différence entre les 2 profils est adaptée à l'échelle d'une PE à l'aide de panaches sur la verticale. On peut remarquer sur cet exemple que la prévision déterministe est beaucoup plus humide entre 400 et 300 hPa que la plupart des membres de la PE. Le profil est ici instable avec un cisaillement de vent marqué qu'on peut repérer à l'aide des barbules du déterministe en les comparant sur la verticale : le vent est très faible en dessous de 800 hPa (vent de 5 à 10 nœuds) alors que le vent est déjà beaucoup plus fort vers 500 hPa avec un vent

1. température en dessous de laquelle la particule d'air humide se condense par saturation

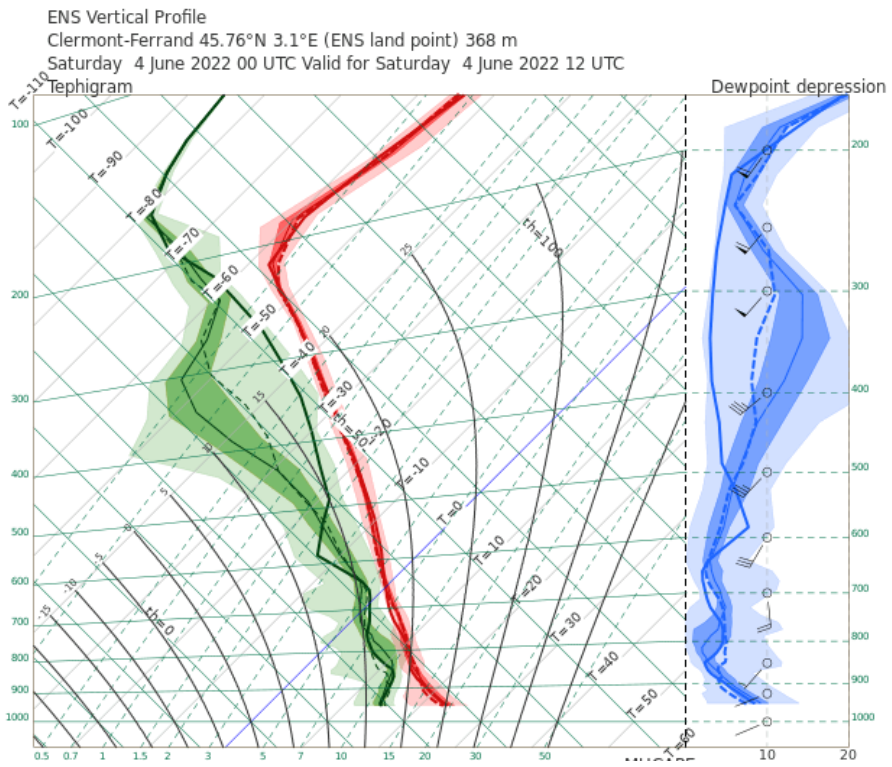


FIGURE 2.5 – *Partie supérieure de la synthèse des principales informations utiles en situation convective à partir de l'ensemble du CEPMMT. Exemple de la ville de Clermont-Ferrand (Centre de la France) pour le 4 juin 2022 à 12h UTC. En haut à gauche, représentation sous forme de panaches des différents profils verticaux (en rouge le profil de température selon les niveaux pressions, en vert la température du point de rosé). Les courbes plus épaisses correspondent au modèle déterministe du CEPMMT. La différence de ces deux courbes permet de tracer le panache bleu en haut à droite. Plus la différence entre la courbe rouge et bleu est faible, plus l'humidité et le risque de nuage sont importants. Le vent du modèle déterministe du CEPMMT selon les niveaux pressions est aussi représenté sur ce même graphe (barbules noires). (Source : https://www.ecmwf.int/en/newsletter/156/meteorology/using-ecmwfs-new-ensemble-vertical-profiles, CEPMMT)*

autour de 35 nœuds. Ce fort cisaillement de vent induit un risque accru de convection sévère (Markowski and Richardson, 2010). D'autres produits sont proposés en complément de cette figure par le CEPMMT comme un hodographe [2] en vision ensembliste mais ils ne seront pas détaillés.

2.2 Synthétisation spatialisée

Les synthèses à l'échelle d'un point sont utiles mais elles ne permettent pas d'avoir une vision globale de la situation. Elle n'est donc pas pertinente pour un prévisionniste qui a sous sa responsabilité des zones à l'échelle régionale, nationale ou mondiale. Pour traiter ces échelles d'autres synthèses de l'information ensembliste existent. Une première partie détaillera des synthèses spatialisées qui tentent

2. Il permet de repérer et quantifier le cisaillement de vent et sa rotation ou non avec la verticale qui est un point important pour la prévision des orages.

de représenter la confiance et/ou les incertitudes qui existent sur la prévision, une deuxième partie détaillera des méthodes pour identifier un risque d'évènement extrême et une dernière partie s'intéressera à des synthèses particulières qui ont pour objectif de regrouper les membres entre eux. Certaines formes de synthèse sont plus couramment utilisées à certaines échelles spatiales ou temporelles. Ce point sera précisé dans le titre du paragraphe si c'est le cas.

2.2.1 Représentation de l'incertitude

Cartes de probabilité

Une des approches les plus naturelles est de généraliser sur une carte en deux dimensions certaines des idées présentées dans la section 2.1. En particulier, il est possible de proposer des cartes de probabilités pour une variable et un seuil donné comme illustré sur la figure 2.6. La carte de gauche correspond à la probabilité de cumuler plus de 10 mm de pluie en 24h selon la PE-AROME. Cette carte montre que ce risque est très élevé sur le Massif Central avec un risque supérieur à 70%. Ce risque est aussi assez élevé sur la Normandie (40-50%) mais à l'inverse, ce risque est très faible voire nul autour de la Méditerranée. Ce mode de représentation permet de repérer rapidement les zones à risque pour des enjeux particuliers comme les canicules, les inondations ou les tempêtes.

Ce calcul point de grille par point de grille des probabilités présente cependant un écueil important pour certains évènements comme les orages. En effet, un même évènement de petite taille mais modélisé à quelques kilomètres de distance sans intersection par 2 membres différents conduira à une probabilité faible de l'évènement sur la zone. Or cette précision à l'échelle du point de grille n'est pas pertinente car si on reprend l'exemple de l'orage, il est impossible de connaître avec précision la position de celui-ci plusieurs heures à l'avance, de même pour l'heure exacte. Cette difficulté se retrouve aussi pour évaluer les modèles, les scores subissent alors le phénomène de la « double peine » mais ce point n'est pas l'objet de ce chapitre et sera détaillé dans le chapitre 3. Concernant les cartes de probabilités, la connaissance du risque en englobant une zone plus large est plus pertinente. Pour cela il est classique de calculer des probabilités en ajoutant une tolérance spatiale mais aussi temporelle, c'est que l'on appelle la méthode des voisins. Contrairement à la première carte de probabilité figure 2.6, celles-ci sont calculées en prenant en compte le point de grille de départ mais aussi ses voisins avec une distance limite à déterminer. De même l'instant t n'est pas le seul à être pris en compte mais est remplacé par une fenêtre temporelle. Par exemple sur une situation convective comme sur le cas de la figure 2.6, cette méthode des voisins appliquée à la figure 2.6 de gauche avec un voisinage de 40 km et une fenêtre temporelle de plus ou moins 1h aboutit à la carte de droite. Le risque de cumul dépassant les 10 mm apparaît toujours nettement sur le Massif Central. La Normandie se dégage aussi très nettement avec des probabilités d'un niveau comparable au Massif Central. Cette information induit qu'il très probable que les 10 mm soient dépassés quelque part sur la Normandie et le Massif Central. Mais ce risque reste néanmoins localisé sur la Normandie par rapport au Massif Central car les probabilités sans voisinage sont beaucoup plus faibles. De façon générale, la prise en compte du voisinage tend à augmenter les probabilités pour un évènement donné et est un bon indicateur

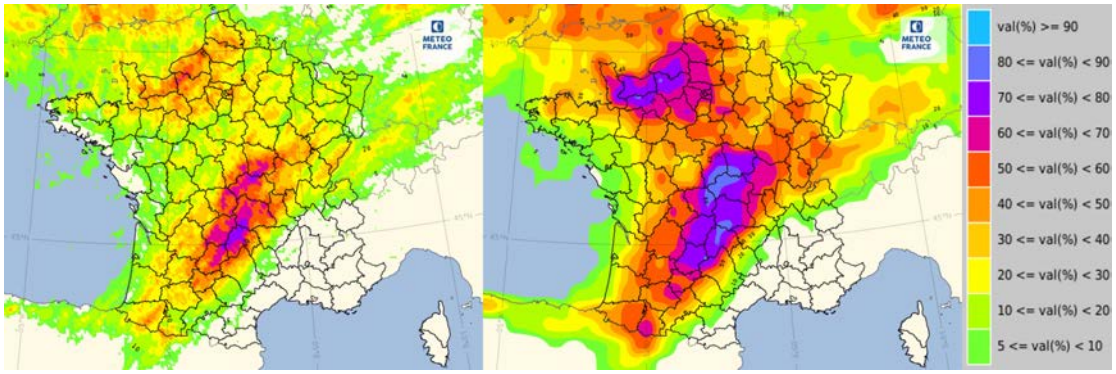


FIGURE 2.6 – *Cartes de probabilités concernant le risque de dépasser les 10 mm en cumul de pluie sur 24h pour la journée du 4 juin 2022 selon le modèle PE-AROME (run du 3 juin 2022 à 03h UTC). La carte de gauche représente les prévisions probabilistes brutes. Concernant la carte de droite, une méthode des voisins spatio-temporelle est appliquée avec un voisinage de 40 km pour la tolérance spatiale et plus ou moins 1h pour la tolérance temporelle (Source : Météo-France, outil Synopsis).*

pour évaluer le risque d'apparition de cet évènement sur une zone large. Mais il est un très mauvais indicateur de l'étendue spatiale du phénomène car surévalué par le voisinage. Une limite importante réside dans les paramètres du voisinage. Ces paramètres sont fixes alors que la tolérance nécessaire peut varier grandement selon la situation (la prévisibilité étant plus faible pour la convection, elle doit a priori être plus grande que pour des fronts synoptiques). Cette tolérance ne prend pas non plus en compte le relief alors que celui-ci peut jouer un rôle majeur sur les précipitations (une zone de Foehn en montagne sur le versant opposé aux précipitations ne sera pas concernée par des précipitations mais la méthode des voisins pourrait englober cette zone). Une autre limite concernant les probabilités de façon générale est qu'elles étudient la distribution à travers une multitude de seuils. En reprenant l'exemple de la figure 2.6, on peut affirmer qu'il est très probable d'observer plus de 10 mm sur le Massif Central mais il n'est pas possible de donner une évaluation quantitative. Peut-on observer seulement 10-15 mm ? ou jusqu'à 30 mm ? Cette carte ne permet pas seule d'y répondre. Il est nécessaire de regarder plusieurs seuils et donc plusieurs cartes. Une dernière limite à signaler est l'incapacité à reconnaître la structure spatiale des phénomènes à partir de ce type de cartes alors que c'est pourtant un point essentiel dans l'expertise que doit apporter le prévisionniste. Ce point sera détaillé dans le chapitre 3.

Cartes de quantiles

Les cartes de quantiles sont de bonnes approches pour résoudre certains des problèmes soulevés par les cartes de probabilités. Le quantile Q d'une variable se définit à partir d'un certain seuil s_q compris entre 0 et 1 (qu'on exprime souvent sous forme de pourcentage) et de la fonction de répartition F construite à l'aide des membres de l'ensemble :

$$Q(s_q) = inf\,\{x; F(x) \geqslant s_q\} \tag{2.1}$$

Le quantile $Q(s_q)$ correspond à la plus petite valeur de la variable telle que s_q % des membres soient inférieurs à cette valeur [3]. Les quantiles les plus fréquemment utilisés sont le quantile 0 qu'on nomme Q_{min} car il correspond à la valeur la plus petite simulée parmi les membres, le quantile 10% (noté Q10), le Q25, le Q50 qui correspond à la médiane, le Q75, le Q90 et le Q_{max} (plutôt que le Q100). La figure 2.7 correspond aux Q75 et aux Q_{max} pour la même variable et le même jour que la figure 2.6. Le Q75 à gauche indique des valeurs importantes de pluie sur le Massif Central avec 75% des membres ayant des prévisions inférieures à 20 mm, mais surtout 25% qui ont des valeurs supérieures. Le Q75 est plus faible sur la Normandie en général. Concernant le Q_{max} à droite, les valeurs sont très homogènes sur le Massif Central et comprises entre 30 et 50 mm. Mais concernant la Normandie, les extrêmes sont localement très importants avec des maximums dépassant les 70 mm. On peut donc conclure de ces cartes que les pluies devraient être fortes et assez homogènes sur le Massif Central. Sur la Normandie, les pluies devraient être beaucoup plus hétérogènes avec des cumuls en général plus faibles mais localement des pluies très importantes pouvant occasionner des inondations. Ces cartes de probabilités et de quantiles sont très utilisées car elles s'adaptent à un très grand nombre de variables météorologiques (pluie, vent, température, ...). Elle présente cependant une limite importante et similaire à celle des cartes de probabilités : ces cartes n'ont aucune cohérence physique. Il est presque certains que la carte du Q_{max} ne sera pas observé car les valeurs pour chaque point de grille sont calculées de façon indépendante et peuvent correspondre à des membres différents. On peut par exemple avoir des orages provoquant des maximums à 30-40mm sur le Massif Central mais on ne peut pas avoir en pratique ces cumuls simultanément sur une très large zone qui dépasse largement la taille caractéristique des orages comme on peut le voir sur cette carte de Q_{max}.

Moyenne/Ecart-type (pour l'échelle synoptique)

Les cartes de moyenne et d'écart-type sont peut-être les plus évidentes pour traiter l'information d'une distribution de valeurs. Mais en pratique, elles ne sont pas forcément pertinentes pour les petites échelles car elles lissent trop l'information. Elles sont par contre utiles pour traiter une information plus grande échelle. Ces cartes sont représentées figure 2.8 avec un traceur intéressant pour les fortes chaleurs qui est la température à 850 hPa. La prévision est ici faite à une semaine. Dans cet exemple, on constate que la moyenne de l'ensemble est assez élevée sur le Sud de la France et dépasse les 20 degrés. Ces valeurs en altitude impliquent au sol des températures très probablement supérieures à 30 degrés. L'écart-type sur l'image de droite indique cependant que celui-ci est assez élevé sur la France en particulier sur le Sud-Ouest où il dépasse les 4 degrés. Cette information nous indique donc que l'incertitude est encore forte sur la masse d'air à 850 hPa qui se positionnera sur le Sud-Ouest de la France. A l'inverse, les températures très chaudes sur le Sud-Est de la France et le Sud de l'Espagne semblent nettement plus fiables.

3. Dans le cas discret des PEs ou si la fonction de répartition n'est pas strictement croissante, il n'y a pas unicité de cette valeur d'où la nécessité de préciser la plus petite valeur.

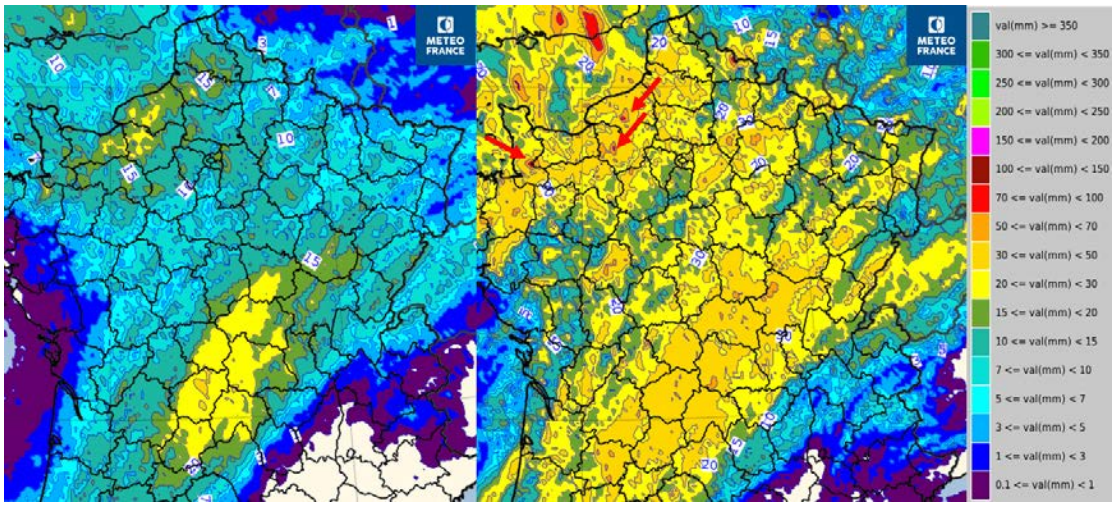


FIGURE 2.7 – *Cartes de quantiles concernant les cumuls de pluie sur 24h (en mm) pour la journée du 4 juin 2022 selon le modèle PE-AROME (run du 3 juin 2022 à 03h UTC). Ces cartes sont zoomées sur le Centre et le Nord de la France. La carte de gauche représente le Q75 et la carte de droite le Q_{max}. Les flèches rouges indiquent des extremas locaux en Normandie dépassant les 70 mm (Source : Météo-France, outil Synopsis).*

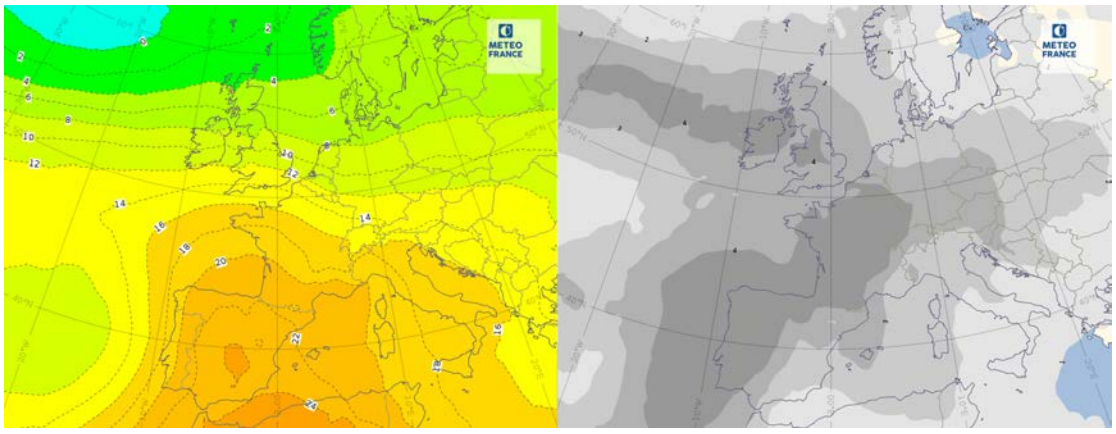


FIGURE 2.8 – *Représentation de la moyenne (gauche) et de l'écart-type (droite) de la température à 850 hPa de l'EPS pour le 17 juin 2022 à 0h UTC (run du 10 juin 2022 à 0h UTC). Les couleurs chaudes sur la moyenne indiquent des températures élevées et les couleurs sombres sur la carte de droite des écart-types importants (Source : Météo-France, outil Synopsis).*

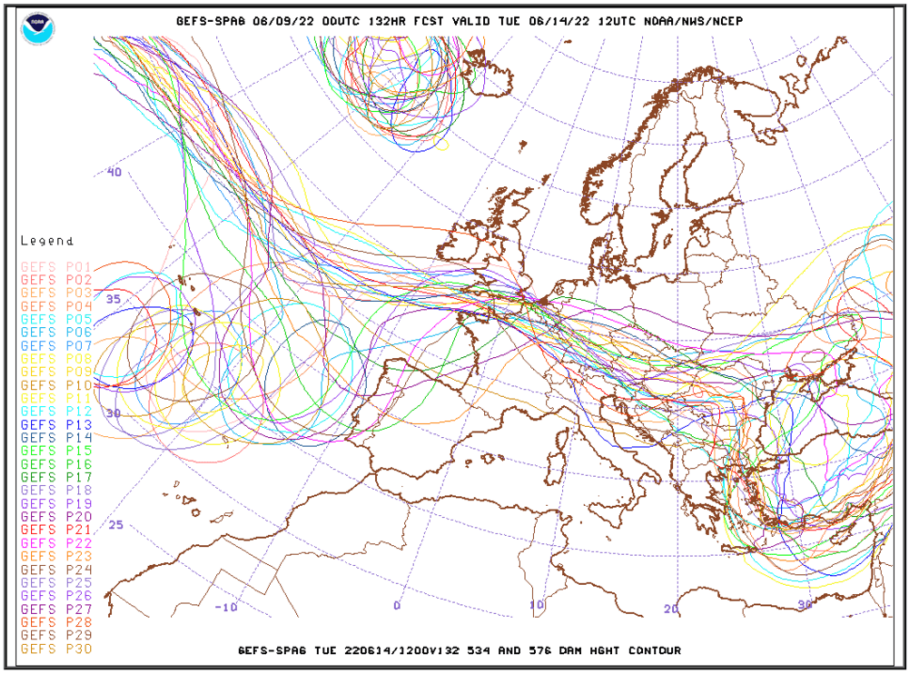


FIGURE 2.9 – *Représentation des prévisions d'ensemble du modèle américain GEFS sous forme de spaghettis. Les contours du géopotentiel à 500 hPa à deux seuils (534 et 576 dam) sont tracés avec une couleur pour chaque membre. Prévision faite le 9 juin 2022 à 0h UTC pour le 14 juin 2022 à 12h UTC (Source :* `https://mag.ncep.noaa.gov/model-guidance-model-parameter.php?group=Model%20Guidance&model=GEFS-SPAG&area=EUROPE&ps=model`*, NOAA)*

Les spaghettis (pour l'échelle synoptique)

Une autre forme de synthèse couramment utilisée pour la grande échelle est le plot sous forme de spaghettis (figure 2.9). Comme pour les « paintball plots », une couleur est attribuée à chaque membre. Chaque contour correspond à une valeur particulière atteinte par la variable étudiée. Sur la figure 2.9, le champ étudié est le géopotentiel à 500 hPa avec ici deux seuils à 534 (contours vers l'Islande) et à 576 dam [4] (contours sur l'Atlantique, l'Europe et au large de la Péninsule Ibérique). Cette carte est intéressante ici car elle permet de repérer dans les membres de l'ensemble américain ce qu'on appelle une goutte froide au large du Portugal qui est matérialisée par les contours fermés. Or, on peut remarquer que beaucoup de membres proposent cette goutte froide mais sa position exacte est incertaine. Cette position influençant directement le risque de vague de chaleurs ou de dégradation orageuse sur la France, c'est un point essentiel à surveiller. Ce graphique présente cependant la limite de devoir définir un seuil qui peut changer l'interprétation de la situation. En effet, la goutte froide pourrait ne pas être perceptible si le seuil est mal choisi par exemple. Ces spaghettis sont utilisés pour scruter le géopotentiel et la P_{mer}.

4. décamètre géopotentiel

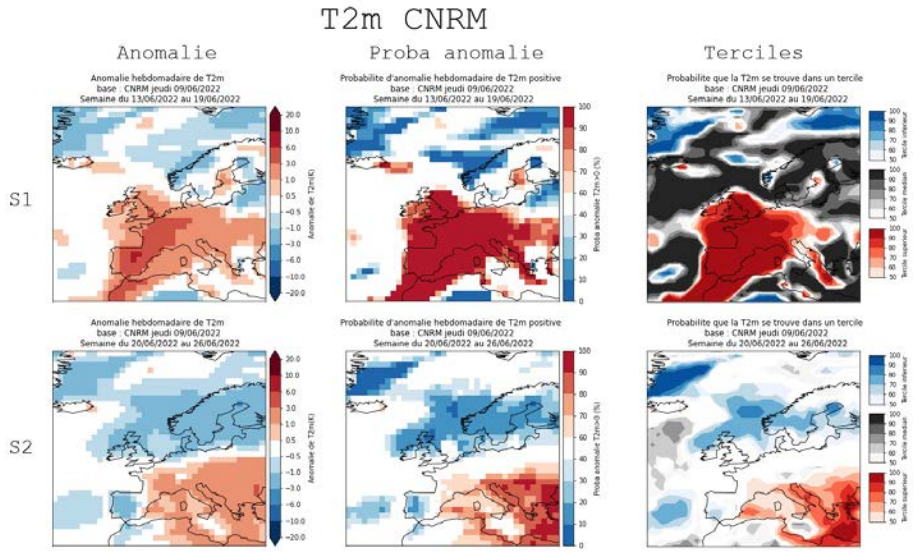


FIGURE 2.10 – *Extrait des prévisions mensuelles du modèle du CNRM pour la température à 2 mètres. Prévision du 9 juin 2022 pour les semaines du 13 au 19 juin (S1) et du 20 au 26 juin 2022 (S2). Ces prévisions sont exprimées à l'aide d'anomalies de températures (gauche), de probabilités d'observer une anomalie positive (centre) et de terciles (droite, Source :* `http://intra.cnrm.meteo.fr/previsionS2S/page.html`*, site interne de Météo-France)*

Les anomalies et terciles (pour les prévisions mensuelles ou au-delà)

En s'intéressant à des échéances encore plus lointaines, la prévisibilité de la grande échelle diminuant, les valeurs des variables deviennent de moins en moins pertinentes. On s'intéresse plutôt à des anomalies par rapport à la climatologie. C'est en particulier sous cette forme que les prévisions mensuelles, saisonnières ou climatiques sont souvent synthétisées comme sur la figure 2.10 pour la prévision mensuelle faite par le modèle du Centre National de Recherches Météorologiques (CNRM). On observe pour la semaine suivante (S1) une anomalie de température positive qui se dégage sur l'Europe de l'Ouest mais qui a tendance à régresser vers l'Est du bassin Méditerranéen la semaine suivante. Le raisonnement en probabilités sur la colonne centrale de la figure 2.10 semble le confirmer. Ce calcul d'anomalie pose cependant un problème. On ne peut pas savoir lorsqu'une anomalie est proche de 0 si elle signifie que l'ensemble des membres est autour de 0 ou si ces membres sont éclatés entre des anomalies positives et négatives et aboutissent à une moyenne proche de 0. La colonne de droite permet d'y répondre avec des terciles qui divisent les membres en 3 groupes (anomalie positive, proche des normales ou négative). La couleur grise très dominante sur la S1 signifie que les zones blanches sur les cartes précédentes correspondaient à des membres proches de 0 et donc une forte probabilité d'être dans les normales de saison. Mais la couleur blanche sur les terciles en deuxième semaine (S2) qui est très dominante signifie ici que l'incertitude est grande et qu'aucun tercile ne se dégage vraiment.

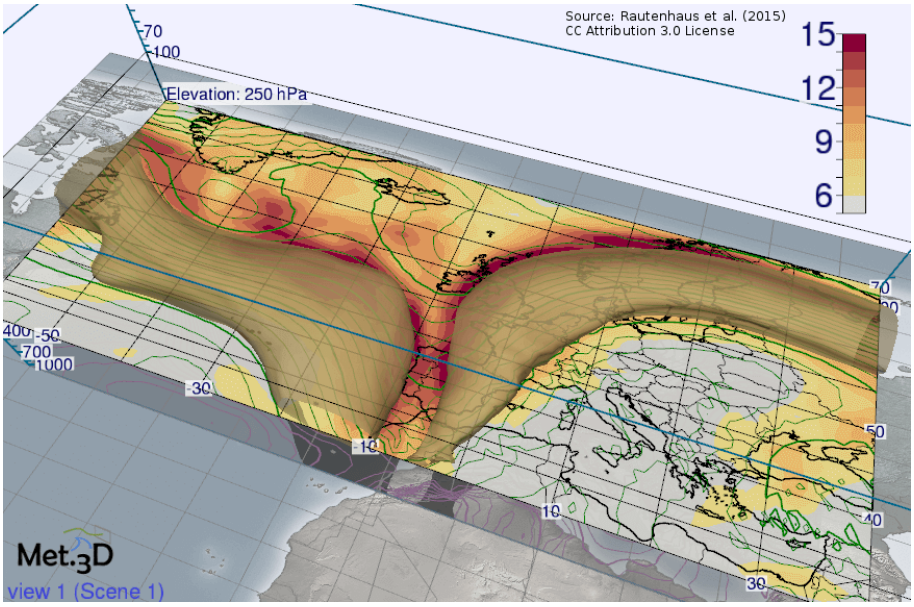


FIGURE 2.11 – *Représentation du champ de vent moyen en haute troposphère (250 hPa) à l'aide de l'outil Met3D sur l'Atlantique et l'Europe. Les formes marrons/jaunes transparentes correspondent aux zones où la moyenne de l'ensemble du CEPMMT dépasse les 30m · s^{-1}. L'écart-type est superposé avec des couleurs jaunes à rouges. (Source : `https://met3d.wavestoweather.de/gallery.html`, © Met.3D contributors 2017. CC Attribution 4.0 International License. met3d.wavestoweather.de)*

Visualisation en 3D

Une autre méthode développée vise à synthétiser l'information en 3 dimensions. Cette visualisation se fait à l'aide de l'outil Met3D (Rautenhaus et al., 2015b). Un exemple de représentation possible est présenté figure 2.11 pour le vent à 250 hPa qui correspond à la circulation atmosphérique dans la haute troposphère. Cette figure montre un courant jet qui circule sur l'Atlantique et qui plonge vers le Maroc avant de remonter sur l'Europe de l'Ouest puis vers la Russie. On remarque également que l'incertitude sur la position de cette zone de vent fort est importante sur la Péninsule Ibérique. La position exacte du plongement du jet vers le Maroc est donc incertaine. Ce type de représentation est développé pour des cas variés avec des développements spécifiques pour représenter certains objets météorologiques à l'instar de la bande transporteuse pour des campagnes de mesure (Rautenhaus et al., 2015a).

2.2.2 Risque d'évènements extrêmes

Cette partie montre deux exemples de production dont l'objectif est orienté vers l'anticipation d'évènements extrêmes. Ces approches sont généralisables à plusieurs variables météorologiques (pluie, température, vent,...).

« paintball plots » (pour l'échelle convective)

Concernant en particulier les situations convectives, il est possible de proposer des « paintball plots » (Demuth et al., 2020). La figure 2.12 donne un exemple sur

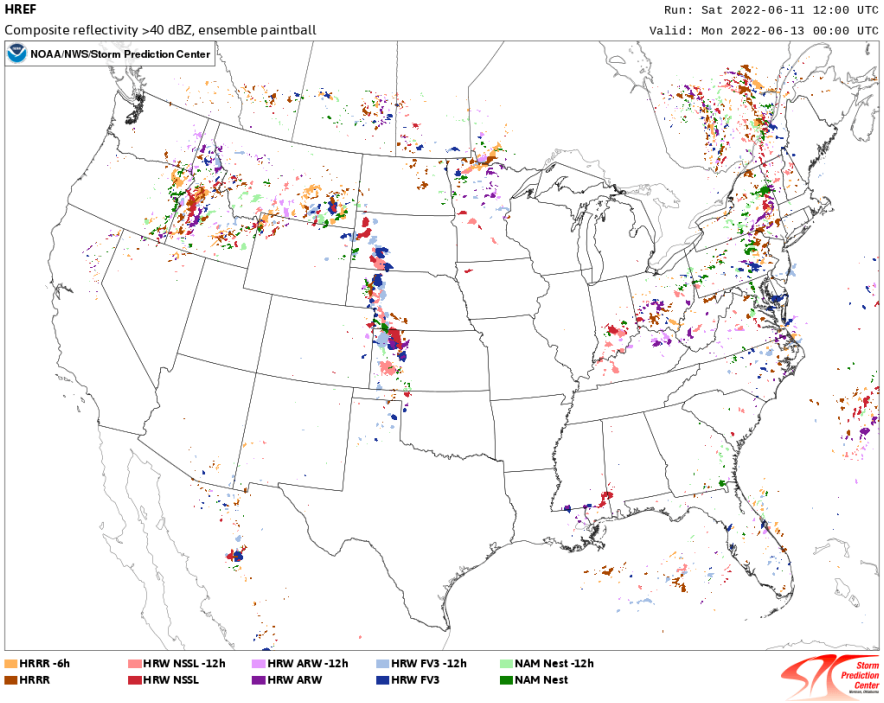


FIGURE 2.12 – *Exemple de « paintball plot ». Représentation des zones de réflectivités simulées dépassant le seuil de 40 dBz avec une couleur pour chaque modèle. Prévision du 11 juin 2022 à 12h UTC pour le 13 juin à 0h UTC (Source :* `https://www.spc.noaa.gov/exper/href/?model=href&product=cref_pb40&sector=conus`*, Storm Prediction Center, NOAA).*

les États-Unis avec une couleur pour chaque modèle et des zones qui correspondent à des réflectivités simulées dépassant les 40 dBz, donc des pluies relativement soutenues et probablement orageuses. Ce mode de représentation permet de savoir rapidement où des zones sont à risque de connaître ce phénomène et intuitivement la probabilité d'apparition est fournie par le nombre de couleurs différentes mais avec l'avantage, par rapport aux cartes de probabilités, de conserver une information sur la structure spatiale des pluies dans chaque modèle ou membre. Cette représentation peut être pertinente pour de petits objets. D'autres variables sont aussi présentées à travers ces plots par la NOAA comme les zones de vents forts ou les zones possédant de fortes ascendances et des tourbillons verticaux car elles représentent un risque élevé d'orages violents (tornades et grosses grêles).

EFI/SOT

L'EFI pour Extreme Forecast Index et le SOT pour Shift Of Tails sont des indices qui comparent une prévision d'ensemble à la climatologie de celle-ci pour mesurer le caractère extrême de la prévision faite en un point et une date précise (Lalaurette, 2003). Cet indice peut se calculer pour différentes échelles de temps, des variables différentes et des modèles aussi bien grande échelle qu'à échelle convective (Boisserie et al., 2015; Raynaud et al., 2018). Dans cette approche, la distribution de la PE du jour est comparée à la climatologie de l'ensemble construite en ce

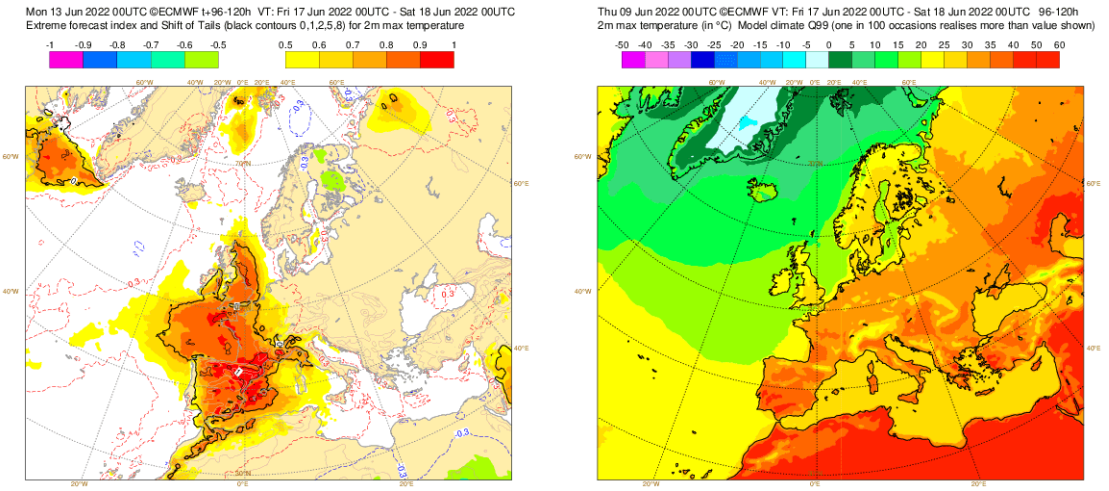


FIGURE 2.13 – *La carte de gauche représente les indices EFI (plage de couleurs) et SOT (contour noir) de la température maximale du 17 juin 2022 (prévision de l'EPS du 13 juin 2022). La carte de droite représente le quantile 99 du climat du modèle pour ce paramètre. (Source :* `https://apps.ecmwf.int/webapps/opencharts/products/efi2web_2tmax?area=Europe&base_time=202206140000&day=1&quantile=99`*, CEPMMT)*

même point et sur les années précédentes [5]. L'indice EFI est compris entre -1 et 1. En particulier, si celui-ci vaut 1 cela signifie que l'ensemble des membres ont une prévision qui dépasse la valeur maximale de la climatologie et que donc cet évènement prévu a un caractère exceptionnel. Cependant, l'EFI ne discrimine pas des prévisions qui seraient égales aux records et des prévisions qui les dépasseraient, l'EFI serait égal à 1 dans les deux cas. Cette discrimination est faite à l'aide de l'indice SOT qui s'intéresse particulièrement aux queues de distribution. Plus l'indice SOT est élevé plus l'évènement dépasse significativement la climatologie. Cette synthèse est souvent faite sous forme de cartes comme sur la figure 2.13. Les prévisions de températures maximales dans l'EPS indiquent un évènement notable de l'Espagne vers l'Angleterre avec des températures très chaudes comparées à la climatologie en particulier sur la Catalogne (zone rouge foncée). Ces indices sont construits à partir de la climatologie des modèles et évitent ainsi une comparaison entre les observations et le modèle qui peut être difficile à faire selon les biais existants ce qui en fait un atout. Cependant la construction de cette climatologie peut être lourde et difficile à réaliser (nécessité de disposer de plusieurs années de prévisions ensemblistes). Cette mesure du caractère extrême n'est également pas toujours corrélée avec l'impact que peut avoir l'évènement. Par exemple, de faibles précipitations dans une zone aride entraineront un EFI élevé mais pas de conséquence notable a priori.

2.2.3 Regroupement de membres

Cette section s'intéresse particulièrement à des méthodes développées pour traiter l'information ensembliste en regroupant les membres entre eux. Cette ap-

5. En pratique, une tolérance spatiale et temporelle peut être faite pour construire cette climatologie.

proche se base sur des méthodes de classification. On peut les diviser en deux approches différentes : la méthode de classification peut être appliquée directement sur les membres en les classant sans a priori ou la méthode de classification peut être utilisée pour définir des structures climatologiques récurrentes. Les membres sont dans un deuxième temps projetés sur ces structures climatologiques et le regroupement des membres se fait par similarité aux structures climatologiques (2 membres proches d'une même structure climatologique sont déduits comme ressemblant). Ces deux approches possibles sont présentées. Ces méthodes s'appliquent surtout à des modèles comme l'EPS ou GEFS et sont donc plutôt destinées à traiter la grande échelle à des échéances supérieures à une semaine.

Regroupement des membres sans a priori

Les regroupements de membres sans a priori ont pour objectif de former des groupes de membres (des classes) et ainsi synthétiser l'information à l'aide de quelques membres qui seraient les représentants de chaque classe. Cette méthode est utilisée sur l'EPS et l'ensemble américain GEFS (non montré) pour traiter les sorties ensemblistes sur quelques jours consécutifs comme sur la figure 2.14. La variable concernée est ici le géopotentiel à 500 hPa. La méthode de classification utilisée est un KMeans (pour l'EPS) ou une méthode hiérarchique (pour GEFS). Ces méthodes seront détaillées plus loin dans ce manuscrit. Les scénarios obtenus sont classés sur cette figure entre les 4 régimes de temps connus en météorologie (Michelangeli et al., 1995; Cassou et al., 2011). Des encadrés en couleur selon ces régimes permettent de repérer rapidement les scénarios qui divergent notoirement à l'échelle de l'Europe comme c'est le cas à la dernière échéance proposée ici (colonne de droite) à 10 jours. On remarque en particulier que le temps en France serait très différent selon le scénario 1 qui est sec et chaud (première ligne) et le dernier plus frais et humide (dernière ligne) en se fiant aux champs de géopotentiel. Dans ce cas présent, les trois scénarios possèdent à peu près le même nombre de membres, ce nombre étant visible dans le titre de chaque imagette : 19 membres pour le 1er, 16 pour le 2e et 16 pour le dernier. Il est donc difficile d'établir un scénario majoritaire. La méthode utilisée pour construire ces scénarios est décrite dans Ferranti and Corti (2011).

D'autres travaux s'intéressent également à représenter l'information sous cette forme. Ferstl et al. (2016b) s'intéressent aux contours (spaghettis) comme ceux présentés figure 2.9 en les plongeant dans un espace réduit, défini à l'aide d'une Analyse en Composantes Principales (ACP). La classification est faite dans cet espace à l'aide d'un algorithme d'esperance-maximisation (EM). Le centroïde de la classe est alors reconstruit dans l'espace physique pour représenter la classe (le centroïde n'étant a priori pas un membre, l'objet obtenu peut ne pas être physiquement cohérent). Un panache autour de ce contour reconstruit est tracé en prenant en compte la dispersion à l'intérieur de la classe, les délimitations de ce panache ne sont cependant pas physiques non plus a priori. La synthèse se retrouve alors sous la forme de quelques spaghettis reconstruits (un par classe) et d'une zone d'incertitude (figure 2.15). Cette synthèse n'intègre pas une composante temporelle ce que les travaux de Ferstl et al. (2016a) apportent en sélectionnant une fenêtre d'intérêt où la classification est faite, généralement aux échéances lointaines, et celle-ci est rétropropagée vers les échéances plus courtes en fusionnant au fur et

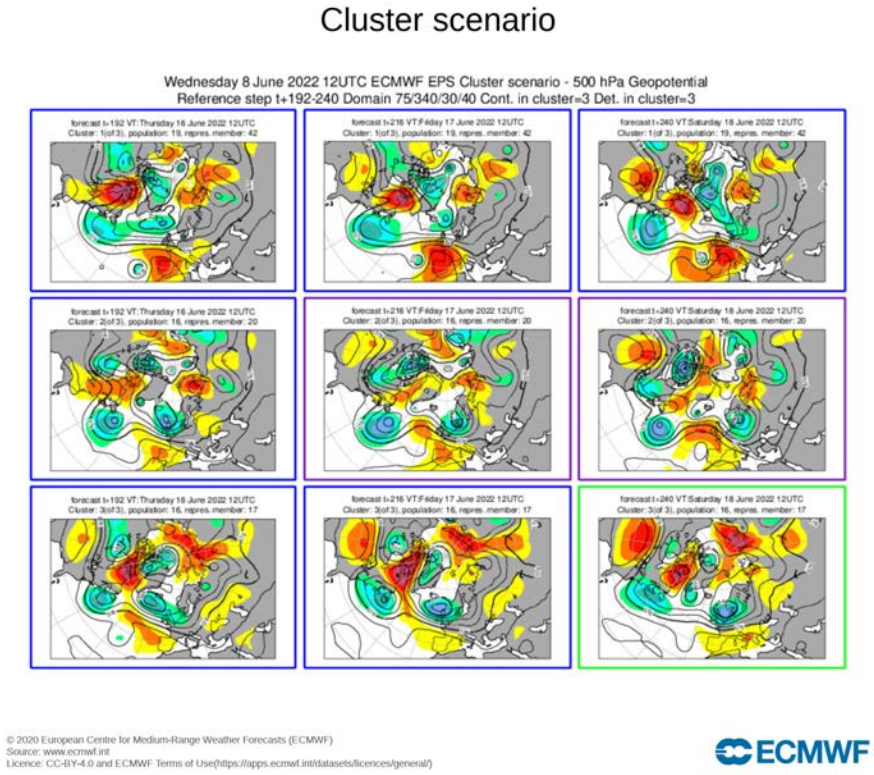


FIGURE 2.14 – *Partitionnement des membres de l'EPS en différents scénarios avec le champ de géopotentiel à 500 hPa. Les lignes correspondent à des scénarios différents et les colonnes correspondent à des jours différents (prévision à 8,9 et 10 jours faite le 8 juin 2022). Chacun de ces scénarios est représenté par un membre. En plus du champ de géopotentiel en noir, les anomalies par rapport à la climatologie sont représentées en couleur. Les cadres de différentes couleurs correspondent à l'un des 4 régimes de temps connus en météorologie sur l'Europe et l'Atlantique Nord. Le membre du scénario 2 à 10 jours encadré en violet appartient par exemple au régime de temps « dorsale atlantique » (Source :* `https://apps.ecmwf.int/webapps/opencharts/products/cluster_plot_legA?base_time=202206130000&cluster=192_240&parameter=500`, *CEPMMT).*

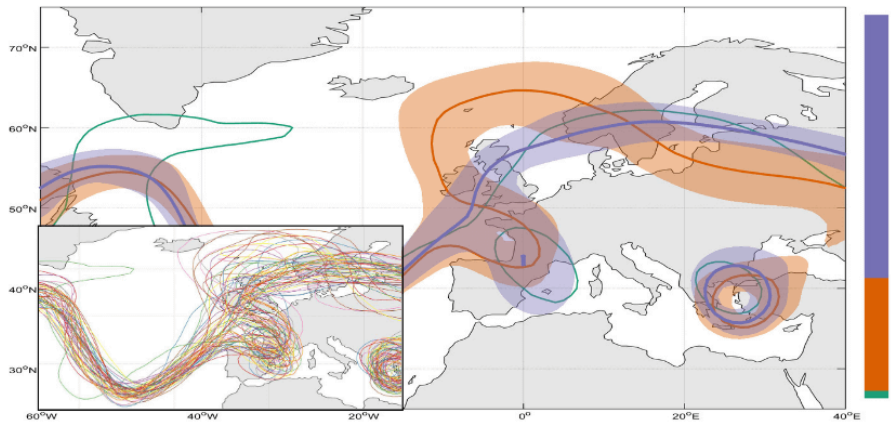


FIGURE 2.15 – *Représentation des spaghettis (en bas à gauche) sous forme de quelques classes représentés par leur contour et leur zone d'incertitude. D'après Ferstl et al. (2016b).*

à mesure les classes quand la distance entre deux classes devient trop faible pour juger qu'il soit effectivement différent. On suppose que en prenant des échéances de plus en plus proches, le nombre de classes différents tend à diminuer [6]. Les travaux de Kumpf et al. (2018) s'intéressent pour leur part à la robustesse des membres à l'intérieur de différentes classes. En effet, certains membres peuvent se trouver à la frontière de deux classes, cette information ne sera pas possible à retranscrire si le membre est assigné à un scénario sans plus d'information.

Structures météorologiques : les régimes de temps

Les régimes de temps, qui ont déjà été évoqués pour la figure 2.14, sont les structures météorologiques les plus connus pour l'Atlantique Nord et l'Europe. Ces régimes ont été définis avec des méthodes de classification sur de longues périodes et à grandes échelles à partir en général du géopotentiel à 500 hPa (Michelangeli et al., 1995; Cassou et al., 2011). Des travaux ont montré des liens entre les anomalies de température et de précipitation relevées localement et ces régimes de temps à grande échelle (Cassou, 2010). Ces régimes de temps sont utilisés pour tenter d'anticiper la circulation générale à grande échelle au niveau de l'Europe et pour en déduire des premiers éléments sur le temps à l'échelle de la France. La synthèse d'une prévision d'ensemble peut être faite en mesurant la distance de chaque membre aux régimes de temps comme sur la figure 2.16. Ces courbes à deux semaines semblent indiquer que le régime NAO+ (figure 2.16, bas à droite) est plutôt improbable sur la période alors que le régime de dorsale atlantique (figure 2.16, bas à gauche) est probable à l'horizon de 7 jours indiquant des conditions plutôt fraîches en flux de Nord sur la France.

La figure 2.17 s'intéresse aussi aux régimes de temps mais la méthode de visualisation est différente. La figure précédente mesurait une distance entre le membre et ces 4 régimes de temps. La méthode repose ici sur l'assignation d'un membre à un des régimes pour chaque jour de prévision et une couleur grise si le

6. Cette vision est acceptable pour des champs grande échelle et sur des périodes longues mais serait plus contestable sur des échelles plus petites comme l'échelle convective où le cycle diurne peut jouer un rôle crucial sur la dispersion.

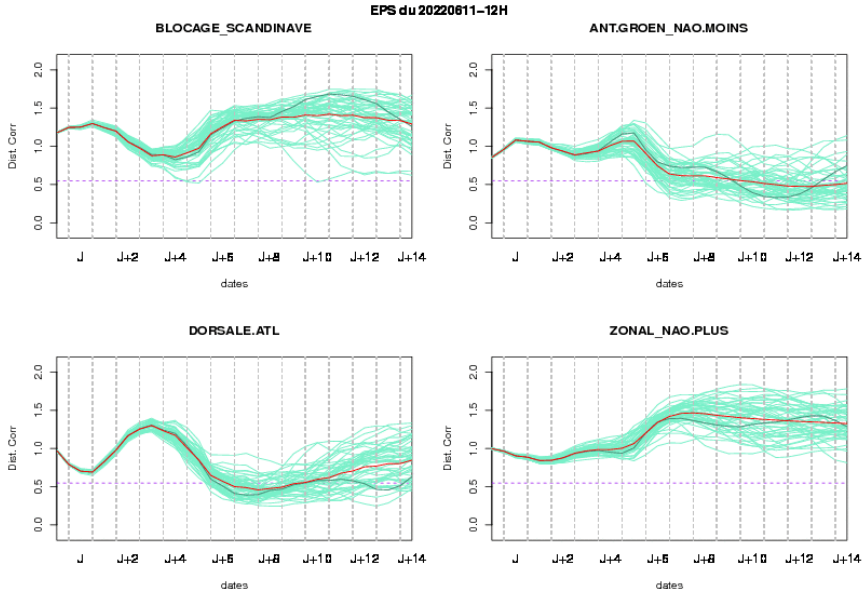


FIGURE 2.16 – *Mesure de la proximité de chaque membre de l'ensemble du CEPMMT aux 4 régimes de temps. Cette mesure est faite avec la distance de corrélation spatiale. La courbe plus foncée est le membre de contrôle de l'ensemble, la courbe rouge la moyenne des membres (Source : `previsibilite.meteo.fr`, site interne de Météo-France).*

membre est trop éloigné des 4 régimes. On remarque alors sur cet exemple que le régime NAO+ (bleu) semble presque certain les premiers jours puis une transition s'opère vers le régime de dorsale (violet) autour du 19 juin. Par la suite l'ensemble semble assez indécis avec cependant le régime de dorsale qui semble peu probable. On peut remarquer que l'assignation de la moyenne de l'ensemble à une classe (carré hachuré en bas du diagramme) ne semble pas toujours cohérent avec la classe majoritaire. Par exemple, la moyenne est assignée au régime NAO+ autour du 17 juillet à droite de la figure alors que le régime de blocage (rouge) semble majoritaire ce qui incite à se méfier de la moyenne de l'ensemble lorsque celui-ci est dispersé. Cette différence peut s'expliquer par les membres non classés qui rentrent dans le calcul de la moyenne.

Autres structures météorologiques

Cette méthode à partir des régimes de temps peut être reproduite à différentes échelles et pour des besoins plus spécifiques. Par exemple la figure 2.18 présente un exemple où des patterns climatologiques en P_{mer} ont été définis afin d'évaluer plus finement le temps au Royaume-Uni. L'exemple ici est fait de 8 patterns mais une autre décomposition est également proposée en 30 patterns différents (Neal et al., 2016). On peut alors assigner chaque membre à un des patterns et regarder comment ils se répartissent entre ces différents patterns. Ce type d'approche peut aussi être proposé afin de traiter un enjeu particulier comme les épisodes fortement précipitant en Méditerranée (Nuissier et al., 2011). A partir de ces classifications de Nuissier et al. (2011), un indice HPE (Heavy Precipitation Event) a été développé pour la PEARP afin d'anticiper ce risque de fortes pluies (non montré).

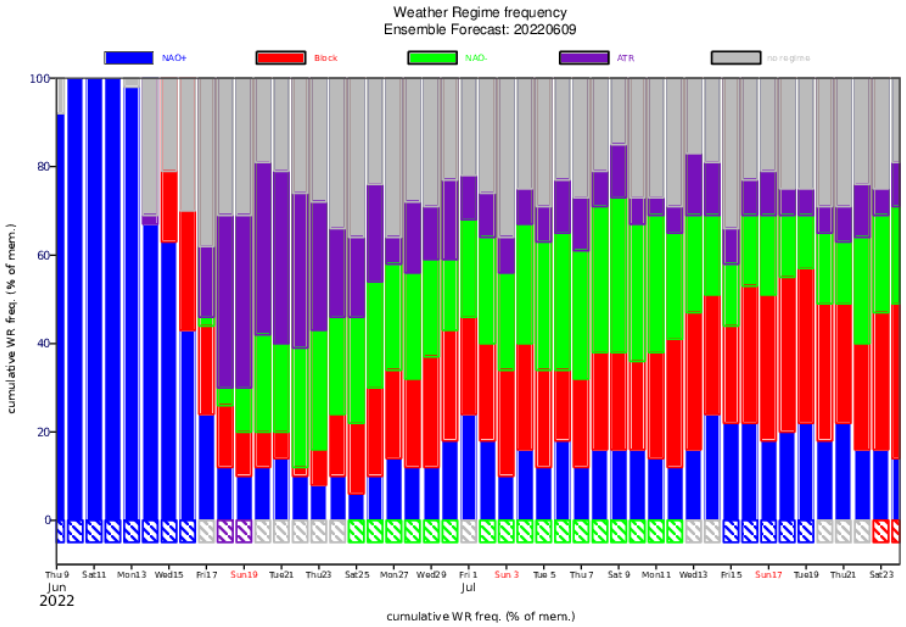


FIGURE 2.17 – *Représentation des régimes de temps assignés aux membres de l'ensemble européen de façon quotidienne. Chaque couleur correspond à un régime et la dernière couleur en gris correspond à des membres qui n'ont pas pu être attribués car trop éloignés des 4 régimes de temps. Le carré hachuré en bas du diagramme correspond à l'attribution de la moyenne de l'ensemble à l'une des 5 catégories. Prévision du 9 juin jusqu'au 23 juin 2022 (Source :* `https://apps.ecmwf.int/webapps/opencharts/products/extended-regime-probabilities?forecast_from=latest`*, CEPMMT).*

	Thu 1 Jan	Fri 2 Jan	Sat 3 Jan	Sun 4 Jan	Mon 5 Jan	Tue 6 Jan	Wed 7 Jan	Thu 8 Jan	Fri 9 Jan	Sat 10 Jan	Sun 11 Jan	Mon 12 Jan	Tue 13 Jan	Wed 14 Jan	Thu 15 Jan
Pattern 1															
Pattern 2	100	100	100		2	18	57	92	96	96	82	78	69	69	67
Pattern 3				2						2	10	8	10	4	18
Pattern 4					24	67	41	4	2	2	6	8	20	18	8
Pattern 5					2									2	
Pattern 6				98	73	16	2	2			2	4	2	6	2
Pattern 7												2			4
Pattern 8								2	2					2	2

FIGURE 2.18 – *Répartition des membres de l'ensemble européen à travers 8 patterns climatologiques pré-définis en P_{mer} sur l'Europe et l'Atlantique Nord. Prévision du 1er Janvier 2015. Le cercle noir représente la position attribuée à l'analyse a posteriori. D'après Neal et al. (2016).*

2.3 Synthétisation pour des événements particuliers

Les précédentes parties de ce chapitre ont montré différentes approches qui permettent de synthétiser la PE mais sans s'intéresser à un objet météorologique particulier et bien identifié. Nous allons voir dans cette partie à travers trois exemples que des synthèses qui suivent certains objets ou schémas conceptuels ont également été développés.

Les cyclones

Les cyclones font l'objet d'une attention particulière et leur suivi à l'aide d'algorithmes de suivi automatique est effectué depuis de nombreuses années avec plusieurs avancées qui sont résumées dans la review de Heming et al. (2019). Un tel suivi peut être fait dans une PE et ainsi proposer un ensemble de trajectoires possibles pour un cyclone qu'on représente souvent sous la forme d'un cône d'incertitude comme sur la figure 2.19 avec ce panache de différentes couleurs selon les probabilités associées d'observer au voisinage de la zone le centre du cyclone. L'avantage de cette synthèse est d'exprimer cette incertitude sur la prévision à travers un objet météorologique connu des prévisionnistes et même du grand public ce qui permet une appropriation plus simple de l'incertitude qui pèse sur la prévision.

Les fronts

Le suivi d'objets météorologiques a aussi été proposé aux moyennes latitudes pour les tempêtes (non montré) ou les fronts (figure 2.20). Sur cette figure chaque trait représente un front détecté par l'algorithme de Hewson and Titley (2010) dans un membre de l'EPS. Les fronts délimitent des masses d'air et sont souvent associés à des bandes pluvieuses, les fronts froids en bleu étant souvent plus actifs. Ces cartes peuvent permettre en particulier en les animant sur plusieurs échéances de repérer le déplacement des masses d'air et l'échéance à partir de laquelle ce suivi ne peut plus être fait car la dispersion dans l'ensemble sera devenue trop importante.

La MJO

Le suivi de la Madden-Julian Oscillation (MJO) est un peu différent car il s'agit du suivi d'ondes équatoriales qui se propagent entre l'Afrique de l'Est et l'Ouest du Pacifique en quelques semaines. C'est donc un élément plus grande échelle que les cyclones ou les fronts. On traduit sous la forme d'un indice l'activité de la MJO ce qui permet de repérer rapidement la position de la zone active (cadran 1 à 8 sur la figure 2.21). Ces oscillations intéressent évidemment les zones directement concernées mais ces oscillations ont des impacts également au niveau du globe. Il existe par exemple des liens entre les régimes de temps en Europe et les phases de la MJO (Cassou, 2008). La figure 2.21 représente les prévisions d'évolution de la MJO selon les membres de l'EPS. Il semble très probable que la MJO ne soit pas active 5 jours après la prévision mais qu'elle deviendra active à partir de 10 jours

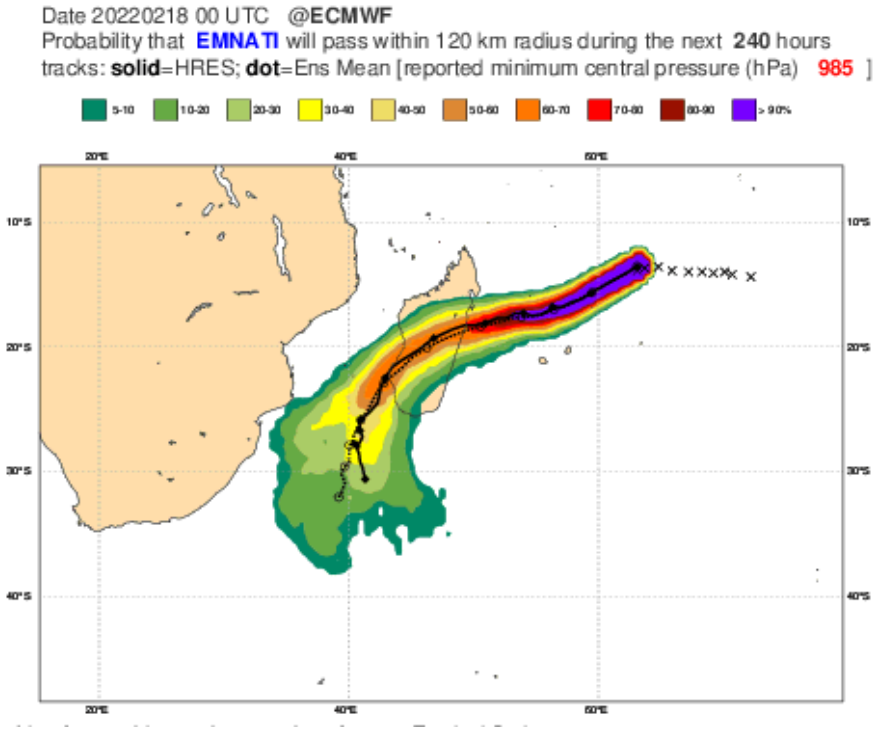


FIGURE 2.19 – *Prévision de la trajectoire du cyclone EMNATI par l'EPS du 18 février 2022 jusqu'à 10 jours. La courbe noire en trait plein représente la trajectoire prévu par le modèle déterministe, la courbe en tiret la trajectoire moyenne de l'ensemble et le panache les trajectoires possibles avec une probabilité associée. Une faible probabilité signifie que peu de membres proposent une trajectoire dans la zone considérée (Source : https://www.ecmwf.int/en/forecasts/charts/tcyclone/tc_strike_probability?facets=undefined&time=2022021800,0,20220218000&unique_id=05S_EMNATI_2022, CEPMMT).*

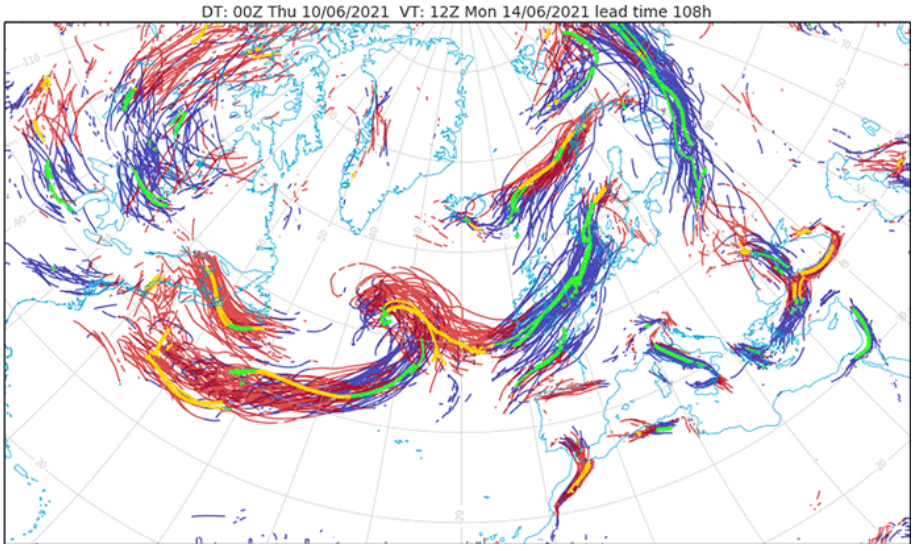


FIGURE 2.20 – *Détections de fronts chaud (rouge) et froids (bleu) avec l'algorithme de Hewson and Titley (2010) dans chaque membre de la prévision d'ensemble du CEPMMT. Prévision du 10 juin 2021 à 0h UTC pour le 14 juin 2021 à 12h UTC (Source : https://www.ecmwf.int/en/about/media-centre/focus/2021/ecmwf-synoptic-style-weather-charts-set-tools-help-forecast-adverse, CEPMMT)*

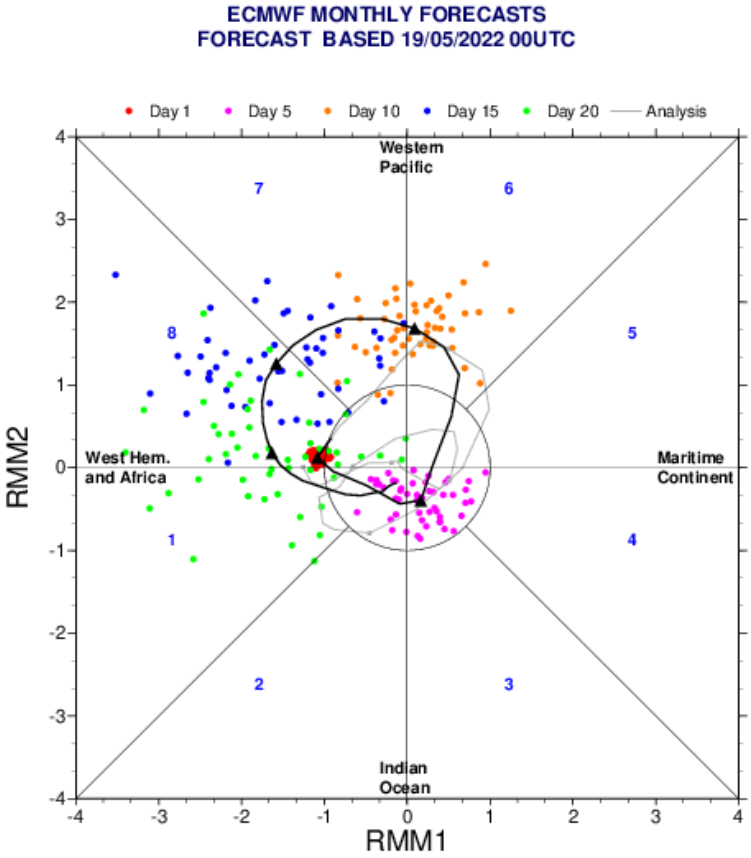


FIGURE 2.21 – *Évolution de l'indice MJO selon les différents membres de l'ensemble européen. Les points à l'intérieur du rond central indiquent une MJO peu active. L'extérieur de ce cercle indique une MJO active classée selon 8 phases (chiffres bleus). Les points sont colorés selon les échéances de 1 jour (rouge) à 20 jours (vert). La courbe noire et les triangles représentent la moyenne de l'ensemble, la courbe grise les valeurs de l'indice sur les 30 derniers jours à partir des analyses. Prévision du 19 mai 2022 de l'EPS (Source : https://www.ecmwf.int/en/forecasts/charts/catalogue/mofc_multi_mjo_family_index?facets=Range,Extended%20(42%20days)&time=2022061300,0,2022061300, CEPMMT).*

et dans le cadran 6 sur la zone Pacifique. La MJO devrait par la suite se propager vers les cadrans 7 et 8 avec une incertitude de plus en plus importante concernant l'intensité et la chronologie exacte liée à sa vitesse de propagation.

On peut conclure ce chapitre en soulignant que la généralisation des prévisions d'ensemble doit encourager à la recherche de nouveaux modes de représentation pour espérer une utilisation la plus complète possible de l'information issue des PEs. Cet effort doit particulièrement être porté sur les PEs à échelle convective pour les prévisions à courtes échéances car elles couvrent des décisions importantes à prendre (carte de vigilance). Ces décisions doivent être par ailleurs prises en un temps très contraint pour des phénomènes où la prévisibilité peut être très limitée. De multiples enjeux sur ces modes de visualisation existent, il faut en effet communiquer efficacement sur le risque d'occurrence d'un phénomène, sa position, son intensité, sa durée et sur l'incertitude qui pèse sur ces paramètres.

Bibliographie

I. Aleksovska, L. Raynaud, R. Faivre, F. Brun, and M. Raynal. Design and Evaluation of Calibrated and Seamless Ensemble Weather Forecasts for Crop Protection Applications. 36(4) :1329–1342, 2021. doi : 10.1175/WAF-D-20-0128.1.

M. Boisserie, L. Descamps, and P. Arbogast. Calibrated forecasts of extreme windstorms using Extreme Forecast Index (EFI) and Shift Of Tails (SOT). 31, 2015. doi : 10.1175/WAF-D-15-0027.1.

C. Cassou. On the interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. *Nature*, 455 :523–527, 2008. doi : 10.1038/nature07286.

C. Cassou. Euro-Atlantic regimes and their teleconnections. pages 6–9, 2010.

C. Cassou, M. Minvielle, L. Terray, and C. Périgaud. A statistical-dynamical scheme for reconstructing ocean forcing in the Atlantic. Part I : weather regimes as predictors for ocean surface variables. *Climate Dynamics*, 36(1-2) :19–39, 2011. doi : 10.1007/s00382-010-0781-7.

J. Demuth, R. Morss, I. Jankov, T. Alcott, C. Alexander, D. Nietfeld, T. Jensen, D. Novak, and S. Benjamin. Recommendations for developing useful and usable convection-allowing model ensemble information for nws forecasters. *Wea. Forecasting*, 35 :1381–1406, 2020. doi : 10.1175/WAF-D-19-0108.1.

L. Ferranti and S. Corti. New Clustering Products. *ECMWF Newsletter*, (127) : 6–11, 2011.

F. Ferstl, M. Kanzler, M. Rautenhaus, and R. Westermann. Time-hierarchical clustering and visualization of weather forecast ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 23, 2016a. doi : 10.1109/TVCG.2016.2598868.

F. Ferstl, M. Kanzler, M. Rautenhaus, and R. Westermann. Visual Analysis of Spatial Variability and Global Correlations in Ensembles of Iso-Contours. *Computer Graphics Forum*, 35(3) :221–230, 2016b. doi : https://doi.org/10.1111/cgf.12898.

J. Heming, F. Prates, M. Bender, R. Bowyer, J. Cangialosi, P. Caroff, T. Coleman, J. Doyle, A. Dube, G. Faure, J. Fraser, B. Howell, Y. Igarashi, R. McTaggart-Cowan, M. Mohapatra, J. Moskaitis, J. Murtha, R. Rivett, M. Sharma, and Y. Xiao. Review of Recent Progress in Tropical Cyclone Track Forecasting and Expression of Uncertainties. *Tropical Cyclone Research and Review*, 8 :181–218, 2019. doi : 10.1016/j.tcrr.2020.01.001.

T. D. Hewson and H. A. Titley. Objective identification, typing and tracking of the complete life-cycles of cyclonic features at high spatial resolution. *Meteor. Appl.*, 17(3) :355–381, 2010. doi : https://doi.org/10.1002/met.204.

I. Ihasz and D. Tajti. Use of ECMWF's ensemble vertical profiles at the Hungarian Meteorological Service. pages 25–29, 2011. doi : 10.21957/tzs8ekmh.

A. Kumpf, B. Tost, M. Baumgart, M. Riemer, R. Westermann, and M. Rautenhaus. Visualizing confidence in cluster-based ensemble weather forecast analyses. *IEEE Transactions on Visualization and Computer Graphics*, 24(1) : 109–119, 2018. doi : 10.1109/TVCG.2017.2745178.

F. Lalaurette. Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *Quart. J. Roy. Meteor. Soc.*, 129(594) :3037–3057, 2003. doi : https://doi.org/10.1256/qj.02.152.

P. Markowski and Y. Richardson. *Mesoscale Meteorology in Midlatitudes*. John Wiley and Sons, Ltd, 2010. doi : 10.1002/9780470682104.ch9.

P. Michelangeli, R. Vautard, and B. Legras. Weather regimes : Recurrence and quasi stationarity. *J. Atmos. Sci.*, 52(8) :1237–1256, 1995.

R. Neal, D. Fereday, R. Crocker, and R. E. Comer. A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteor. Appl.*, 23(3) :389–400, 2016. doi : https://doi.org/10.1002/met.1563.

O. Nuissier, B. Joly, A. Joly, V. Ducrocq, and P. Arbogast. A statistical downscaling to identify the large-scale circulation patterns associated with heavy precipitation events over southern France. *Quart. J. Roy. Meteor. Soc.*, 137(660) : 1812–1827, 2011. doi : https://doi.org/10.1002/qj.866.

M. Rautenhaus, C. M. Grams, A. Schäfler, and R. Westermann. Three-dimensional visualization of ensemble weather forecasts – Part 2 : Forecasting warm conveyor belt situations for aircraft-based field campaigns. *Geoscientific Model Development*, 8(7) :2355–2377, 2015a. doi : 10.5194/gmd-8-2355-2015.

M. Rautenhaus, M. Kern, A. Schäfler, and R. Westermann. Three-dimensional visualization of ensemble weather forecasts – Part 1 : The visualization tool met.3d (version 1.0). *Geoscientific Model Development*, 8(7) :2329–2353, 2015b. doi : 10.5194/gmd-8-2329-2015.

L. Raynaud, B. Touzé, and P. Arbogast. Detection of Severe Weather Events in a High-Resolution Ensemble Prediction System Using the Extreme Forecast Index (EFI) and Shift of Tails (SOT). 33(4) :901–908, 2018. doi : 10.1175/WAF-D-17-0183.1.

S. Vannitsem, J. B. Bremnes, J. Demaeyer, G. R. Evans, J. Flowerdew, S. Hemri, S. Lerch, N. Roberts, S. Theis, A. Atencia, Z. B. Bouallègue, J. Bhend, M. Dabernig, L. De Cruz, L. Hieta, O. Mestre, L. Moret, I. O. Plenković, M. Schmeits, M. Taillardat, J. Van den Bergh, B. Van Schaeybroeck, K. Whan, and J. Ylhaisi. Statistical Postprocessing for Weather Forecasts : Review, Challenges, and Avenues in a Big Data World. *Bull. Amer. Meteor. Soc.*, 102(3) :681–699, 2021. doi : 10.1175/bams-d-19-0308.1.

Chapitre 3

Synthétiser les ensembles pour la prise de décision

Le chapitre 1 a permis de montrer comment la prévisibilité limitée de l'atmosphère, quelles que soient les échelles spatio-temporelles, rend l'utilisation de prévisions d'ensemble nécessaire. Le chapitre 2 a récapitulé les différentes méthodes existantes pour synthétiser l'information présente dans les membres des ensembles. Ce chapitre va s'intéresser aux objectifs de cette thèse en décrivant de nouvelles méthodes pour extraire l'information de la PE-AROME à destination des prévisionnistes. Cette thèse cherche à mieux prendre en compte des structures cohérentes atmosphériques pour extraire l'information issue de la PE-AROME. Les structures cohérentes sont à la base de notre compréhension de l'atmosphère et de l'expertise des prévisionnistes. Pour un expert, reconnaître ces structures dans un modèle déterministe est possible car le modèle peut être étudié en détail, il est cependant beaucoup plus difficile de le faire dans une prévision d'ensemble par manque de temps et plus fondamentalement en raison de l'inadéquation des méthodes de visualisation existantes qui sont rarement physiquement cohérentes comme expliqué dans le chapitre 2. La première partie de ce chapitre s'intéressera à l'expertise des modèles par les prévisionnistes basée sur ces structures cohérentes et les décisions qui en découlent. La partie suivante s'intéressera à l'approche « objet » qui peut servir de lien entre ces structures et les sorties numériques. La troisième partie s'intéressera à la structure cohérente qui fait l'objet d'une partie des travaux de thèse : les échos arqués, une organisation orageuse potentiellement dangereuse. La structure des échos arqués sera présentée ainsi que la méthode proposée durant cette thèse pour mieux extraire le risque d'échos arqués des membres PE-AROME. Cette détection d'objets est intéressante pour évaluer le risque d'un évènement particulier mais n'est pas pertinente dans le cas général où aucun phénomène précis n'est recherché. La quatrième partie de ce chapitre présentera alors une autre méthode développée durant cette thèse pour extraire l'information de la PE-AROME : la synthèse par scénarios. S'appuyant sur les méthodologies existantes décrites dans la section 2.2.3, l'objectif est de regrouper les membres entre eux pour représenter la PE-AROME sous la forme de quelques scénarios désignant des groupes de membres.

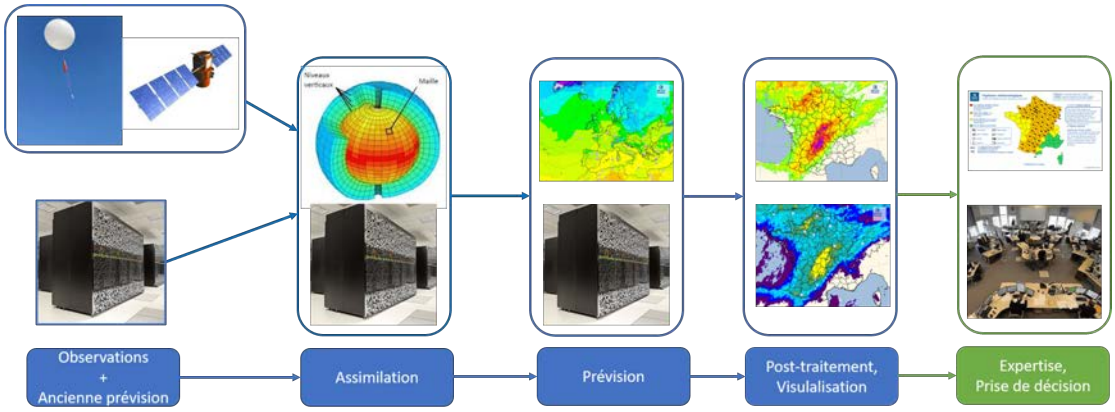


FIGURE 3.1 – *Représentation schématique de la chaîne de prévision. Cette partie s'intéresse à l'expertise faite par les prévisionnistes et les prises de décisions qui en découlent, représentées par une carte de vigilance et une image du Centre National de Prévision (CNP).*

3.1 Expertise du prévisionniste

Pour réaliser une prévision, les données des modèles arrivent sur le poste du prévisionniste sous différentes formes (dont celles décrites dans le chapitre 2). Cet ensemble de données n'est pas simple à appréhender, c'est le rôle du prévisionniste de les analyser pour en tirer le meilleur parti. L'objectif de cette thèse étant de développer de nouvelles méthodes pour faciliter l'utilisation des ensembles en opérationnel, il est nécessaire de s'intéresser aux moyens employés par les prévisionnistes pour expertiser les modèles. Mieux comprendre ces moyens aidera à concevoir des outils de synthèse en adéquation avec ceux-ci, et permettant l'utilisation des ensembles en se focalisant sur l'information utile pour ces experts. Cette partie s'intéresse donc à l'un des derniers maillons de la chaîne de prévision à Météo-France, la partie dédiée à l'expertise des modèles par les prévisionnistes et les prises de décision qui en découlent pour les bulletins météo à destination du grand public, la carte de vigilance et bien d'autres applications (figure 3.1).

3.1.1 Les structures cohérentes

L'expertise d'un modèle météorologique et les prévisions qui en découlent s'appuient très largement sur des « structures cohérentes » qu'on observe à différentes échelles et dont on a déjà parlé dans les chapitres précédents comme les ondes de Rossby, les dépressions, les anticyclones, les fronts ou encore les lignes de grains. Ces structures cohérentes se définissent comme une combinaison de variables météorologiques avec une emprise spatiale pouvant aller de quelques kilomètres pour certains orages à l'ensemble de la zone intertropicale pour les cellules de Hadley. L'étude de ces structures cohérentes a permis d'en déduire des cycles de vie couvrant la naissance, la phase de maturité et la disparition des phénomènes. En décrivant ces cycles de vie, on a alors construit des schémas conceptuels, comme pour le creusement des dépressions aux moyennes latitudes (ou cyclogénèse, figure 3.2). Ces structures cohérentes étaient pour certaines connues bien avant l'existence de modèles météorologiques, c'est le cas par exemple du modèle norvégien

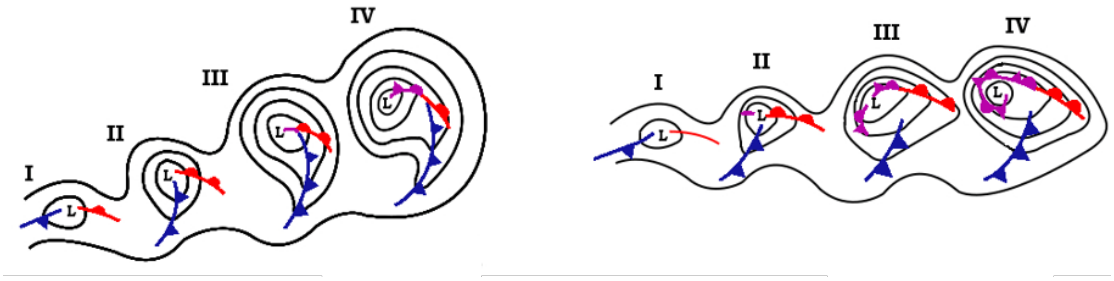


FIGURE 3.2 – *Schémas conceptuels du cycle de vie (de I à IV) d'une dépression aux moyennes latitudes selon le modèle conceptuel norvégien à gauche (Bjerknes and Solberg, 1922) et le modèle conceptuel de Shapiro and Keyser (1990) à droite. Les contours correspondent aux isobares du champ de P_{mer} dont le centre dépressionnaire est représenté par un « L ». Les fronts sont représentés par les objets rouges, bleus et violets dont la signification est détaillée en annexe figure A.1. (Source : `http://tempetes.meteo.fr/spip.php?article231`)*

pour la cyclogénèse. Cette connaissance des schémas conceptuels et la compréhension des processus physiques en jeu a permis de faire des prévisions bien avant l'existence des modèles numériques. En partant des observations, des prévisions pouvaient être faites dès la fin du XIX^{e} siècle.

Les progrès réalisés depuis cette époque et l'arrivée des modèles numériques n'ont absolument pas remis en cause cette méthodologie (Calas, 2013). Ces progrès ont plutôt permis de compléter les schémas conceptuels. La compréhension de la dynamique d'altitude près de la tropopause a permis des progrès majeurs concernant entre autres la cyclogénèse et la convection. La dynamique d'altitude fut alors intégrée aux ANAlyses SYnoptiques Graphiques (ANASYG) qui sont des cartes techniques avec les différentes structures cohérentes sur l'Atlantique Nord et l'Europe (Santurette and Joly, 2002). Les structures cohérentes liées à la cyclogénèse se retrouvent toujours sur l'ANASYG produit par les prévisionnistes sur la figure 3.3 en 2017 avec une dépression sur la France décrivant le schéma de Shapiro and Keyser et une autre vers le Canada répondant plutôt au schéma norvégien décrit par Bjerknes and Solberg presque un siècle auparavant. Ces structures cohérentes sont comme une grille de lecture pour les prévisionnistes qui leur permettent de comprendre les processus observés dans les modèles. La prévision est ensuite présentée en décrivant l'évolution des structures dans ces modèles. L'incertitude de la prévision est souvent exprimée avec un vocabulaire propre aux structures (le modèle X prévoit le cœur dépressionnaire 100 km plus au sud que le modèle Y, le front du modèle X est en retard par rapport au modèle Y, ...). En plus de permettre l'analyse et la critique des modèles, les structures cohérentes sont associées à un temps sensible. Elles permettent de traduire avec des termes simples (pluie, averse, ensoleillé, ...) la prévision. Un front froid actif est par exemple associé à des pluies continues, suivi souvent à l'arrière par un ciel de traîne avec de nombreuses averses. Ces éléments permettent aux prévisionnistes de tracer différentes zones soumises à un type de temps homogène comme pour cette production figure 3.4. Sur cette figure, l'Est et le Sud-Est de la France bénéficient d'un temps ensoleillé, une large bande centrale est soumise à un risque d'averses localement orageuses sur les Pyrénées. La façade Ouest est elle soumise à un temps plus nuageux mais sec. Les structures cohérentes permettent aux prévisionnistes d'établir une prévi-

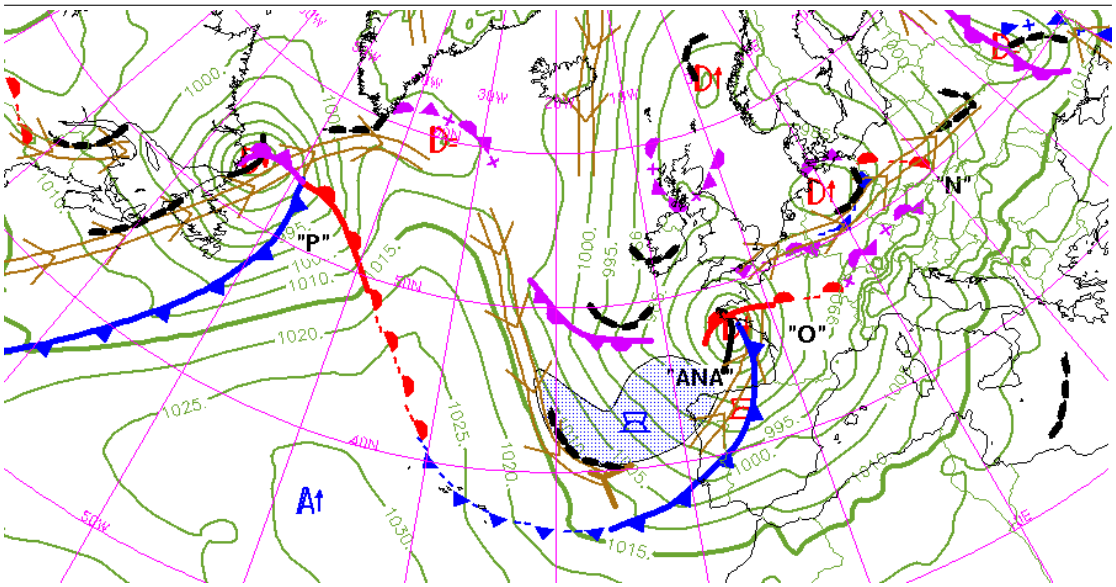


FIGURE 3.3 – *Analyse technique de la situation météorologique du 11 décembre 2017 à 00h UTC sur l'Europe et l'Atlantique Nord à travers l'ANAlyses SYnoptiques Graphiques (ANASYG) produite par le CNP. La signification des différents symboles est disponible en annexe (figure A.1).*

sion qui est cohérente dans l'espace et le temps ce qui est un point essentiel pour délivrer un message clair et compréhensible vers les interlocuteurs extérieurs. Le besoin d'un discours cohérent rend également nécessaire à travers ces structures de chercher de la cohérence dans les membres de l'ensemble. Il faut retenir de cette partie que les structures cohérentes jouent un rôle clé que ce soit dans l'expertise d'un modèle par un prévisionniste ou la transmission de l'information météorologique. **Pour utiliser pleinement les prévisions d'ensemble, il semble donc souhaitable de construire des synthèses se basant sur ces structures cohérentes.**.

3.1.2 La prise de décision en contexte incertain

Même si ce manuscrit ne traite pas des problématiques liées à la prise de décision ni des théories sur ce sujet, il est important de rappeler comment l'information météorologique est retranscrite vers des usagers extérieurs. Les prévisionnistes produisent aujourd'hui une information qui est essentiellement déterministe. Par exemple le bulletin grand public diffusé par les médias ne transmet que peu d'informations sur l'incertitude d'une prévision surtout aux courtes échéances où une température précise est annoncée et non un intervalle. Le temps est également défini comme ensoleillé ou pluvieux avec peu de place laissé à l'incertitude. Des messages nuancés mentionnant des scénarios minoritaires sont ainsi impossibles à transmettre ou presque. Ce constat rend la communication concernant des phénomènes peu probables et peu prévisibles mais à fort impact difficile à faire. **La principale conséquence de cette façon de présenter l'information est la nécessité pour les experts de ne pas seulement analyser les sorties des modèles mais aussi de prendre des décisions en privilégiant certains modèles par rapport à d'autres.**

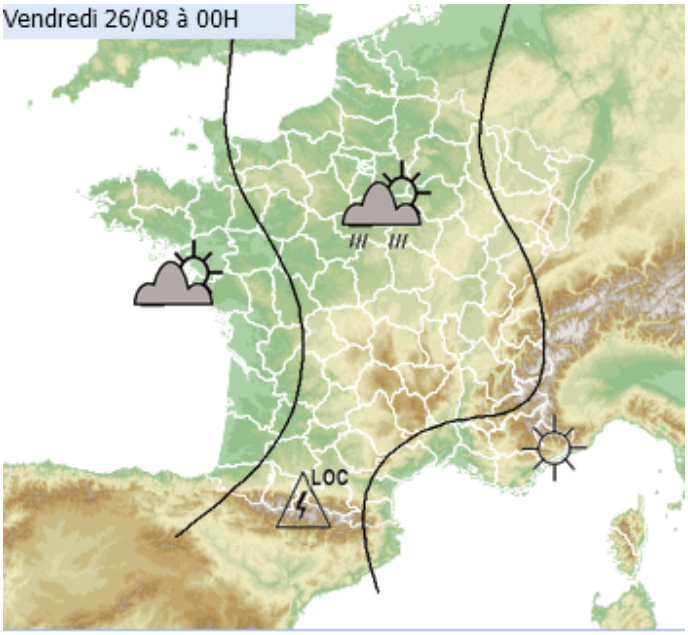


FIGURE 3.4 – *Représentation graphique « grand public » de la situation météorologique prévue le 26 août 2022 à minuit. Prévision faite le 21 août 2022 (Source : `previ-gene.meteo.fr`, site interne de Météo-France)*

La prise de décision peut devenir très critique dans le cadre de la vigilance. Les acteurs de la sécurité civile attendent par exemple de savoir si des mesures préventives sont nécessaires en lien avec un aléa météorologique. Mais que faire si cet aléa a 5%, 10% ou 20% de risque de se produire ? Cette question est primordiale pour prendre une décision avec une vision probabiliste. La connaissance de la vulnérabilité de l'usager devient alors cruciale : dans quelle mesure celui-ci accepte-t-il de prendre le risque de ne pas être prévenu d'un évènement ? La réponse à cette question est en général très difficile à obtenir. Elle est même très complexe dans le cadre de la vigilance météorologique car d'autres facteurs sont également à prendre en compte comme l'acclimatation des habitants d'une région à un aléa météorologique (la neige par exemple dans l'Est de la France par rapport à l'Ouest), l'heure de l'aléa (est-il de nuit ou en pleine journée ?) ou la date de l'aléa (est-on en période de vacances avec de nombreux touristes ?). Pour la vigilance météorologique, on peut être tenté de répondre qu'il ne faut prendre aucun risque. Cette stratégie peut cependant aboutir à de très nombreuses fausses alarmes qui risquent de décrédibiliser le système de vigilance. On voit donc que la prise de décision est un équilibre entre la prévision météorologique alliée à son incertitude et la vulnérabilité de la société à l'aléa. Le prévisionniste a donc besoin de connaître le mieux possible l'ensemble des systèmes qu'il utilise pour être en capacité de justifier ses décisions. **Il est donc primordial, pour faciliter l'utilisation d'outils de synthèse de prévisions d'ensemble, d'être le plus exhaustif possible sur leurs performances et de connaître les limitations de chaque approche pour les communiquer aux prévisionnistes**. L'objectif est donc de déterminer une forme de « guide d'utilisation » pour chaque produit en collaborant entre les différents services. En plus de son expérience, le prévisionniste utilisera chaque outil en connaissance de cause et il sera en capacité d'appuyer ses décisions sur des bases statistiquement et empiriquement éprouvées. Ce besoin de

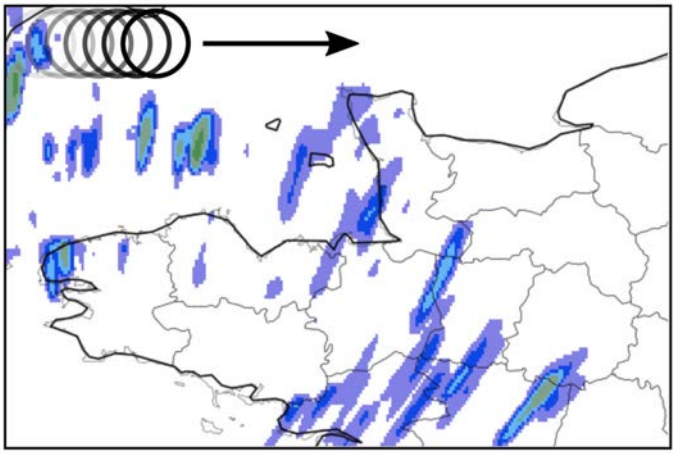


FIGURE 3.5 – *Représentation du balayage d'une méthode objet en calculant la distribution du voisinage de plusieurs points de grille, espacés de façon régulière. Les voisinages sont représentés par les cercles noirs. (Source : `http://intra.cnrm.meteo.fr/recyf/Appl/Objets//aide.php`, site interne de Météo-France)*

connaître les limitations d'un outil en condition opérationnelle est d'ailleurs l'un des points clés d'après un sondage fait auprès de plusieurs prévisionnistes américains (Harrison et al., 2021). Ce besoin d'expliciter au mieux les méthodes sera pris en compte dans ce manuscrit.

En lien avec les remarques faites dans le paragraphe précédent, la construction d'outils de synthèse pour les prévisionnistes passe par de nombreux échanges pour définir les problématiques à traiter ou encore pour critiquer et améliorer les outils de synthèse (Stuart et al., 2022). Ces échanges permettent également une appropriation facilitée des futurs outils en suivant le processus de recherche. **Plusieurs échanges avec les prévisionnistes ont donc eu lieu au cours de la thèse et seront précisés explicitement**.

3.2 Approche objet

3.2.1 Représentation des structures cohérentes

La section précédente a montré la nécessité d'intégrer les structures cohérentes pour faciliter l'utilisation des prévisions d'ensemble. La sortie d'un modèle numérique à un instant donné se représente comme un tableau en 3 dimensions (ou plus si on considère plusieurs variables météorologiques). Il faut se demander comment faire émerger les structures cohérentes de ces tableaux de valeurs. Avec cet objectif, plusieurs algorithmes ont été développés, ceux-ci cherchant à repérer des zones avec un temps homogène en traçant automatiquement des contours pour définir des objets. Pour des objets de pluies, une première méthode (Davis et al., 2006) s'appuie sur le lissage du champ de pluies puis sur l'application d'un seuil sur ce champ lissé. D'autres méthodes définissent des objets en regroupant les points possédant un voisinage similaire (Arbogast et al., 2016; Rottner et al., 2019). Ces méthodes mesurent la similarité des voisinages de plusieurs points à l'aide des

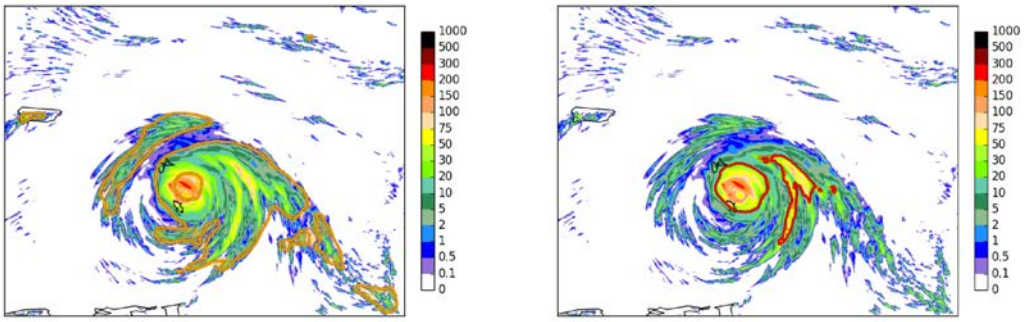


FIGURE 3.6 – *Représentation du cyclone Maria avec le cumul de pluie sur 6h. Les objets de pluies modérées sont tracés avec les contours orange à gauche et les objets de pluies fortes à droite avec des contours rouges. Ces objets sont tracés avec la méthode décrite dans Rottner et al. (2019) (Source :* `http://intra.cnrm.meteo.fr/recyf/Appl/Objets//aide.php`*, site interne Météo-France)*

distributions spatiales issues de leurs environnements proches (figure 3.5). Ces distributions sont comparées à des distributions de référence qui permettent de décider si le point de grille appartient à un objet ou pas. Ces méthodes ont été appliquées pour définir des objets précipitants pour différents niveaux d'intensité (figure 3.6). Des objets pour évaluer la texture des champs de pluies (pluies continues ou intermittentes) ont également été développés à l'aide de méthodes d'apprentissage automatique appelées forêts aléatoires (Hamidi et al., 2020). Les approches objets permettent de ne plus regarder les sorties des modèles comme une suite de valeurs mais plutôt sous la forme de différentes structures dont on peut faire un suivi au fil du temps en décrivant leurs déplacements, leurs intensifications ou leurs disparitions. **L'approche objet semble être la meilleure façon d'appréhender les structures météorologiques, elle sera donc très utilisée tout au long de ce manuscrit.**

3.2.2 Solution à la problématique de la double peine

L'identification d'un objet permet d'en extraire des caractéristiques ou attributs particuliers. On peut penser à la taille de l'objet ou encore la position de son centre de masse par exemple. Appliquer la méthode objet à la fois sur les sorties des modèles et les observations permet de définir des scores objets en comparant ces attributs (Raynaud et al., 2019; Gilleland et al., 2020). Ces scores objets présentent l'avantage d'être souvent plus compréhensibles pour une partie des utilisateurs car ils s'appuient sur des objets qu'ils connaissent et dont on évalue l'erreur de position moyenne par exemple. Des scores en point de grille sont plus difficiles à interpréter physiquement. Un autre avantage majeur de ces scores objets par rapport à des scores plus classiques est qu'ils ne sont pas soumis à la problématique de la double peine (Rossa et al., 2008; Ebert, 2008). La double peine s'observe particulièrement pour des phénomènes intenses et très localisés comme des orages ; les modèles à haute résolution sont donc notamment concer-

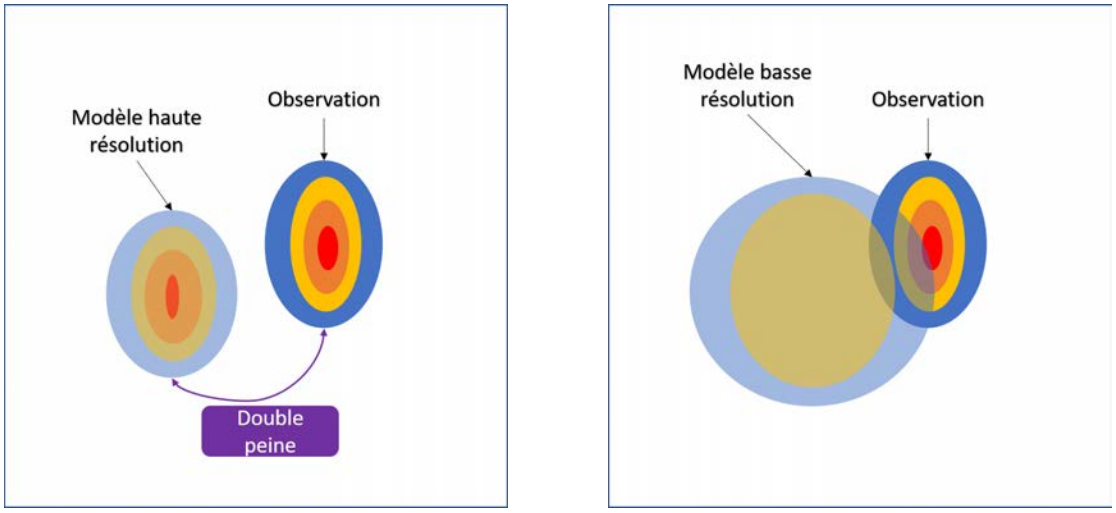


FIGURE 3.7 – *Illustration de la double peine. Les champs idéalisés sont de la pluie avec les intensités les plus fortes en rouge. Les deux cartes représentent des champs de pluies de modèle à haute et basse résolution qui sont transparents pour les différencier des champs de l'observation complètement opaques.*

nés. La double peine est schématisée sur la figure 3.7. Sur l'image de gauche, un score basé sur les points de grille considérera que le modèle n'a pas prévu l'objet observé car ils ne se superposent pas, mais le score va aussi pénaliser la prévision à tort de fortes pluies quelques kilomètres plus loin (d'où le terme de double peine). Les scores vont ainsi être plutôt mauvais. Ils seront notamment moins bons que ceux d'un modèle qui n'a pas du tout prévu de pluie. Ils seront aussi moins bons que ceux d'un modèle basse résolution qui produit en général des champs moins forts et plus grands recouvrant partiellement l'observation comme sur la figure de droite. Pourtant, le modèle haute résolution a bien mieux modélisé le phénomène que le modèle basse résolution. **L'approche objet permet de mesurer ces différences de tailles, d'intensités ou de positions et ainsi s'affranchit du problème de la double peine ce qui la rend très complémentaire des scores « point de grille ».**

3.3 Échos arqués

Les deux premières parties de ce chapitre ont montré l'importance des structures cohérentes en météorologie, l'approche objet étant un moyen de parvenir à les repérer automatiquement dans les modèles. Un des premiers objectifs de la thèse est d'utiliser cette approche objet pour détecter des structures cohérentes automatiquement dans les membres PE-AROME et obtenir une synthèse de ces détections automatiques en quelques graphiques. Le choix de la structure cohérente à détecter a été fait en concertation avec les services de prévision et s'est porté sur une organisation orageuse particulière, les échos arqués. Ce choix est motivé par les conséquences majeures que peut avoir cette organisation orageuse, la capacité de la PE-AROME à modéliser ces structures et la grande difficulté à les repérer avec les méthodes de synthèse des ensembles actuelles. L'essor des méthodes d'apprentissage profond rend aujourd'hui cette détection automatique

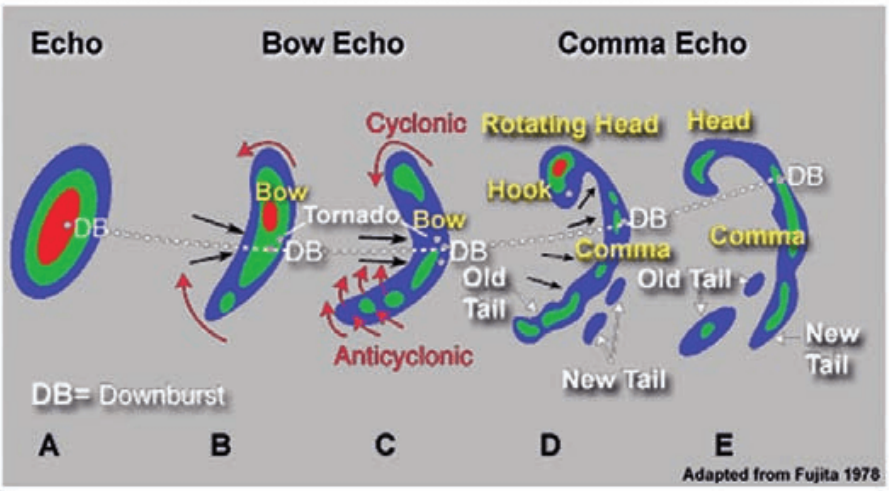


FIGURE 3.8 – *Cycle de vie d'un écho arqué d'après Fujita (1978)*

possible, comme les chapitres 4 et 5 le montreront. Cette partie décrit les caractéristiques générales des échos arqués, comment ils se forment et enfin comment cette thèse va chercher à construire un diagnostic pour mieux les anticiper.

3.3.1 Caractéristiques générales

L'écho arqué est un système convectif de méso-échelle (MCS pour l'acronyme anglais). Il se reconnaît principalement à sa forme en arc dans le champ de réflectivité (cf section 1.1.1 pour le détail de ce champ). Le cycle de vie complet est détaillé sur la figure 3.8. Pour former un écho arqué, un système orageux doit pré-exister, il peut avoir cependant des caractéristiques très diverses (Klimowski et al., 2004). Ce système orageux pré-existant (stade « Echo ») tend à se courber à cause de forts vents dont l'origine sera expliquée dans la sous-partie suivante. Le système prend alors une forme en arc (stade « Bow Echo »). Selon le modèle conceptuel de Markowski and Richardson (2010), un motif de P_{mer} spécifique est observé autour d'un écho arqué, consistant en un minimum de pression en amont de l'arc, un maximum de pression sous l'arc, et un autre minimum de pression sous les pluies stratiformes en aval. De plus, deux tourbillons sont générés sur les bords de l'écho arqué, le bord gauche ayant un sens de rotation cyclonique (anti-horaire) et le bord droit un sens de rotation anticyclonique (horaire). Pour des raisons qu'on ne détaillera pas, l'environnement dans lequel évolue le système convectif tend à renforcer le tourbillon cyclonique au détriment du tourbillon anticyclonique. Le renforcement du tourbillon gauche tend à changer la forme de l'écho arqué qui prend une forme plutôt de virgule (stade « Comma Echo ») avant de se dissiper. La taille caractéristique de cette structure varie de 50 à 200 km en général et la durée de vie de quelques dizaines de minutes à plusieurs heures (Goulet, 2015). Ce système se déplaçant très rapidement, les cumuls de pluies restent modérées. Le principal risque associé aux échos arqués sont les fortes rafales de vent. La zone où le risque est maximal est au centre de l'arc (Przybylinski, 1995). C'est malheureusement ce type de système qui a balayé la Corse le 18 août 2022 où des rafales à plus de 200 $km \cdot h^{-1}$ ont été relevées. Localement des tornades peuvent être observées. Les échos arqués s'observent aussi bien en Amérique du Nord qu'en

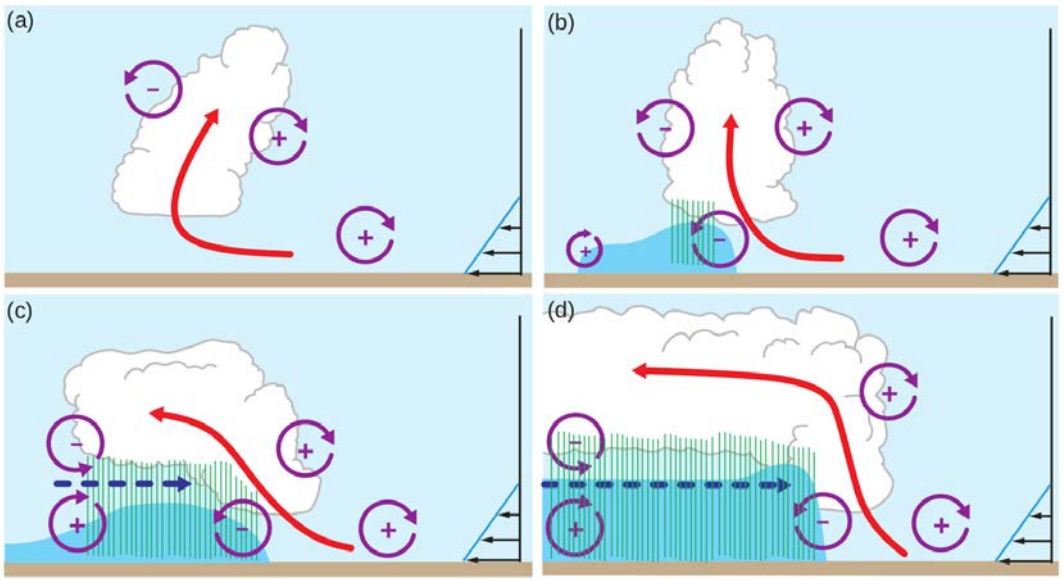


FIGURE 3.9 – *Les différents stades idéalisés pour former un écho arqué d'après Markowski and Richardson (2010). Les stades de (a) à (d) sont détaillés dans la section 3.3.2.*

Europe tout au long de l'année même si les échos arqués qui se forment en été n'auront pas exactement les mêmes caractéristiques que ceux observés en hiver (Goulet, 2015). Ce manuscrit se concentrera sur les échos arqués en été qu'on observe généralement entre mai et octobre et qui sont plus fréquents en France métropolitaine. Les échos arqués sont des organisations orageuses assez rares en France avec en moyenne 5 échos arqués par an entre 2006 et 2020.

3.3.2 Formation

Cette sous-partie va permettre de décrire comment un écho arqué se développe pour mieux comprendre sa forme en arc et pourquoi le risque de fortes rafales de vent est important. Les différentes étapes sont représentées figure 3.9. Un écho arqué a besoin d'un environnement fortement instable [1]. Un fort cisaillement de vent unidirectionnel est aussi nécessaire sur la verticale (surtout entre 0 et 3 km d'altitude), matérialisé par les flèches noires en bas à droite de chaque image de la figure 3.9. Cet environnement cisaillé génère un tourbillon représenté près du sol avec un "+". Dans cet environnement favorable à la convection, des nuages se forment avec des ascendances représentées par une flèche rouge (stade (a)). Cette ascendance se courbe dans le sens du tourbillon de l'environnement. L'apparition de ces nuages génère lui-même la création de nouveaux tourbillons représentés plus haut (toujours stade (a)). Ces nuages, en donnant des précipitations qui peuvent être soutenues, génèrent un courant de densité et une plage froide près du sol matérialisée par la zone bleue au stade (b). Sans rentrer dans les détails, les bords de la plage froide génèrent eux-mêmes de nouveaux tourbillons favorisant les ascendances autour de cette plage froide. La principale conséquence ici est de redresser les ascendances en contre-balançant le tourbillon dû à l'environnement

1. Cette caractéristique est moins indispensable en hiver.

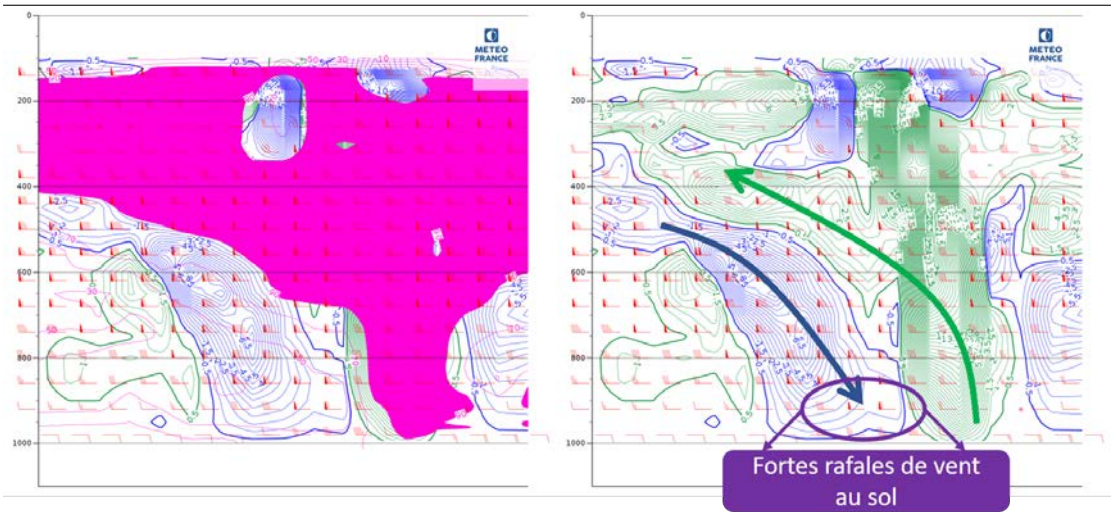


FIGURE 3.10 – *Coupe verticale faite dans un écho arqué simulé par AROME. L'axe vertical représente les niveaux pressions du modèle en hPa et l'axe horizontal différentes positions qui traversent l'écho arqué (l'avant de l'écho arqué est à droite des graphiques). La carte de gauche superpose l'humidité (elle dépasse 90% dans la zone fuchsia), les vitesses verticales par les isocontours bleus et verts (en vert les vents ascendants et en bleu les vents subsidents) et le vent moyen tangent à la coupe verticale par les barbules rouges (un triangle rouge indique des vents à plus de 90 $km \cdot h^{-1}$). La carte de droite ne représente plus l'humidité pour mieux mettre en évidence la circulation des vents.*

(toujours image (b)). En continuant son développement, le tourbillon lié à la plage froide prend le dessus sur celui de l'environnement ce qui rejette les ascendances vers l'arrière et génère des pluies dites stratiformes à l'arrière du système (stade (c)). Ces pluies contribuent au développement et à l'étalement de la plage froide. A l'arrière du système, le tourbillon lié aux nuages de convection et celui lié à la plage froide forment ou renforcent un couloir de vent qu'on appelle le jet arrière (ou RIJ pour Rear Inflow Jet en anglais), matérialisé par la flèche bleue en pointillé. Avec cette organisation, les ascendances et subsidences de l'orage sont bien séparées ce qui lui permet de continuer son développement jusqu'au stade (d) avec des tourbillons renforcés et un jet arrière plus fort. C'est ce jet arrière qui provoque une accélération locale de la vitesse de déplacement des orages qui explique la forme en arc que prend le système (Przybylinski, 1995; French and Parker, 2014). Ce jet arrière se forme en altitude mais les subsidences sont de plus en plus fortes en se rapprochant du front de l'orage par évaporation des précipitations. Ces subsidences peuvent faire plonger le jet arrière jusqu'au sol et produire de très fortes rafales de vent (Fujita, 1978; Atkins and Laurent, 2009; Markowski and Richardson, 2010).

Les modèles haute résolution d'aujourd'hui comme AROME sont capables de simuler ces structures et la circulation des vents associée comme le montre la figure 3.10. L'humidité sur la carte de gauche montre des nuages qui descendent jusque dans les basses couches près du sol, puis qui suivent les ascendances en partant vers l'arrière du système (flèche verte sur la carte de droite). Sous ces nuages et ces ascendances, une zone de vent fort beaucoup plus sèche circule. Les fortes subsidences (flèche bleu) entraînent alors cette zone de vent fort vers les

basses couches qui se matérialise par des barbules dépassant fréquemment les 90 $km \cdot h^{-1}$. On observe alors près du sol de fortes rafales de vent comme expliqué dans le paragraphe précédent.

3.3.3 Outil de synthèse adapté

Les parties précédentes ont montré l'intérêt de repérer les échos arqués à cause principalement des fortes rafales de vent associées. Ayant vu également que les modèles d'aujourd'hui à haute résolution sont capables de les simuler, ce point va présenter la méthodologie générale de cette thèse pour mieux extraire l'information issue de la PE-AROME concernant les échos arqués. La méthodologie peut se résumer en 2 étapes clés décrites sur la figure 3.11 et qui ressemblent à celles présentées pour la détection des fronts dans le chapitre 2 (figure 2.20). Le premier objectif est de développer un outil automatique pour détecter les échos arqués dans les sorties des modèles. A notre connaissance, aucun outil n'est aujourd'hui utilisé pour le faire dans des services météorologiques. Pour les détecter, la partie précédente a montré qu'un écho arqué se caractérise principalement par des réflectivités en forme d'arc. La forme des réflectivités est donc un élément essentiel mais elle ne peut pas être prise en compte efficacement avec les méthodes présentées précédemment dans la section 3.2. L'essor récent des méthodes d'apprentissage profond en météorologie permet de l'envisager car elles sont particulièrement performantes pour réaliser de la reconnaissance de formes comme l'expliquera le chapitre 4. Ces méthodes requièrent cependant de nombreux exemples de détections d'échos arqués pour être performante. En l'absence de méthode pré-existante, ces exemples devront être fournies par un expert comme le décrira le chapitre 5. La méthode d'apprentissage profond retenue pour réaliser la détection automatique des échos arqués est un réseau de neurones convolutif qui sera présenté dans ce même chapitre 5. L'étude approfondie des performances de ce réseau de neurones sera faite dans le chapitre 7. Après avoir détecté les échos arqués dans les membres PE-AROME, il reste à superposer ces détections sur des cartes de synthèse comme sur celle de la figure 3.11. La construction de ces graphiques de synthèse avec l'aide de plusieurs prévisionnistes est présentée dans le chapitre 7 également. Ces graphiques permettent de synthétiser le risque d'échos arqués dans PE-AROME, ce qui aurait nécessité l'analyse de plusieurs centaines d'images si cette synthèse était construite manuellement. Comme le montre les contours noirs du graphique de synthèse sur la figure 3.11, ce travail de détection automatique est également réalisé dans AROME déterministe et superposé sur le même graphique. Les performances de PE-AROME et AROME peuvent ainsi être comparées. Différentes comparaisons sur les performances de PE-AROME et AROME seront faites dans le chapitre 8. A terme, il serait souhaitable de pouvoir aussi détecter automatiquement les échos arqués dans les données observées pour principalement établir des scores objectifs de PE-AROME et AROME. Des travaux préliminaires en ce sens seront présentés dans le chapitre 8.

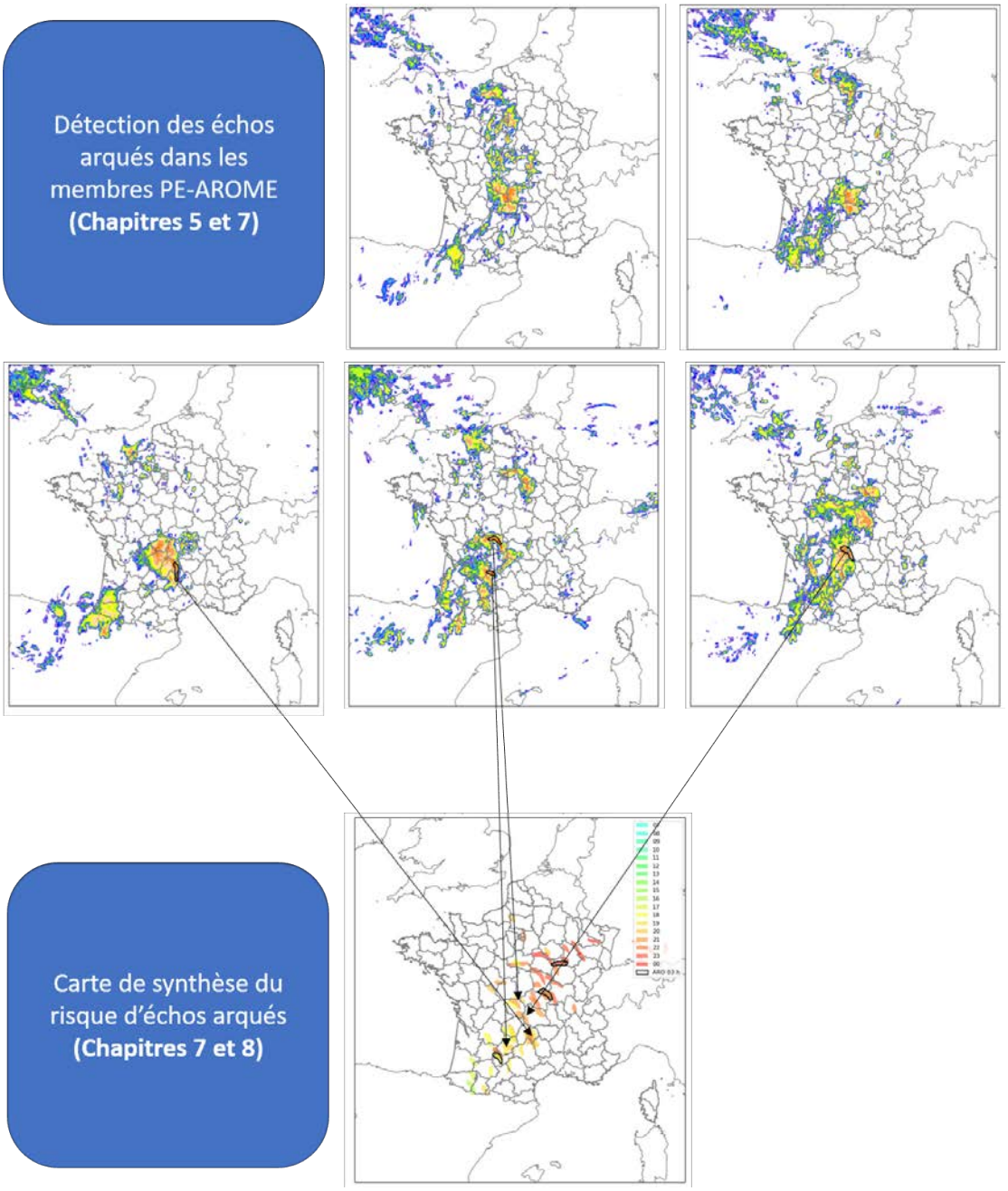


FIGURE 3.11 – *Schéma récapitulant la méthodologie développée pour prendre en compte les échos arqués simulés dans la PE-AROME. Les échos arqués sont d'abord détectés dans chaque membre par un réseau de neurones de façon automatique à l'aide du champ de réflectivités, ces détections sont entourées en noir sur les 3 membres de la deuxième ligne. L'ensemble des détections dans les 16 membres sont superposées sur une carte de synthèse. En plus des détections dans PE-AROME, les détections sont faites dans AROME et sont entourées en noir sur la carte de synthèse pour qu'elles se différencient de celles issues de PE-AROME.*

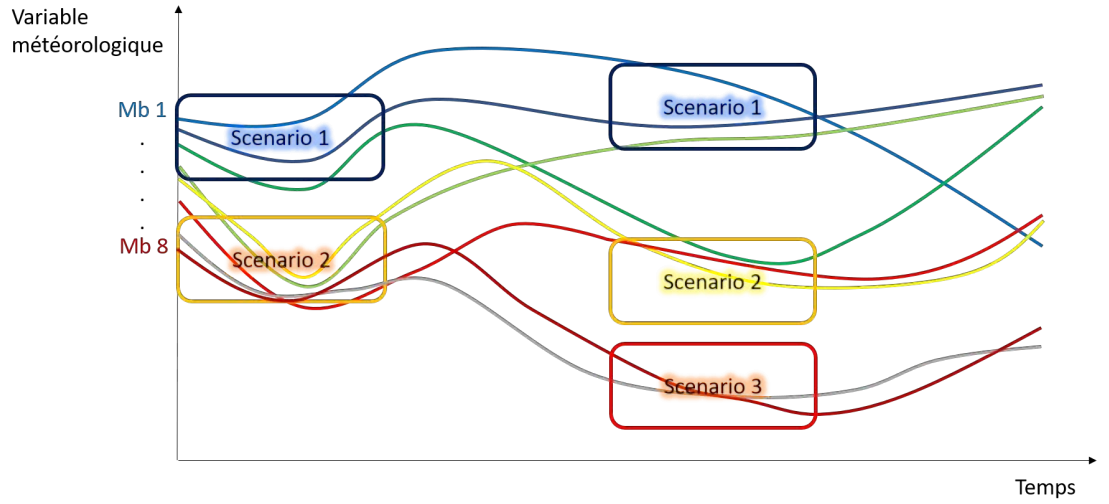


FIGURE 3.12 – *Représentation schématique des scénarios. Chaque membre est représenté par une courbe de couleur différente au cours du temps.*

3.4 Synthèse par scénarios

La méthodologie présentée dans la partie précédente peut être étendue à d'autres types de structures météorologiques, mais elle reste centrée sur la détection d'un type d'évènement particulier (les échos arqués, les fronts, les tempêtes, ...). Bien qu'intéressante pour les évènements à forts enjeux, ce type de synthèse ne permet pas de traiter le cas général et ne permet pas de répondre au besoin concernant l'utilisation de la PE-AROME au quotidien, quel que soit le type de situation. Une autre limitation de ces détections d'objets est le besoin d'une connaissance pointue en météorologie, ce qui est le cas des prévisionnistes mais pas forcément de leurs interlocuteurs. Ces productions sont donc utiles en interne mais peuvent difficilement être exploitées pour communiquer une information vers l'extérieur. Pour répondre à ces limitations, le deuxième objectif structurant de la thèse réside dans le développement d'une nouvelle méthode : la synthèse par scénarios. Elle permet de mettre en perspective une prévision sous une forme qui n'est pas une incertitude qu'on affecte sur une prévision choisie par le prévisionniste mais plutôt en la représentant sous la forme de plusieurs scénarios possibles qui ont une certaine probabilité de se produire. La synthèse par scénarios va être présentée et définie dans une première sous-partie avant de détailler plus précisément la méthode développée.

3.4.1 Définition

La synthèse par scénarios peut se représenter de façon schématique comme sur la figure 3.12. L'objectif est de regrouper les membres PE-AROME entre eux pour former des groupes. La PE-AROME est alors représentée non plus sous la forme de 16 prévisions mais plutôt sous la forme de quelques scénarios. Aux premières échéances, on peut repérer deux scénarios sur la figure 3.12 : le premier avec les membres 1, 2 et 3 et le second avec les membres 4 à 8. Mais selon l'évolution des membres, le nombre de scénarios peut évoluer dans le temps. Toujours sur la figure 3.12, on observe plus tard 3 scénarios au lieu de 2 avec par exemple le premier

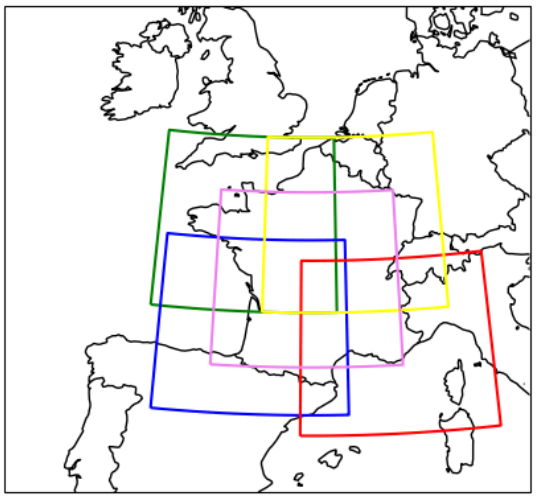


FIGURE 3.13 – *Découpage du domaine AROME en 5 sous-zones régionales : Nord-Ouest (NO) en vert, Nord-Est (NE) en jaune, Sud-Ouest (SO) en bleu, Sud-Est (SE) en rouge et Centre (C) en rose.*

constitué des membres 1, 2 et 4. Cette représentation de l'information a déjà été proposée dans des travaux similaires comme l'a montré le chapitre 2, section 2.2.3 en particulier figure 2.14. L'originalité de la synthèse par scénarios dans cette thèse est de s'intéresser à une prévision d'ensemble à échelle convective comme PE-AROME alors que ce mode de représentation était jusqu'ici développé pour des prévisions d'ensemble à plus grande échelle avec des échelles de temps de l'ordre de la semaine. Un autre aspect innovant vient de la variable considérée : le cumul de pluie sur 1h (RR1h). Le champ de pluie a une grande intermittence spatiale et temporelle, ce qui le rend très sensible à la problématique de la double peine surtout dans un modèle à échelle convective. Ce point nécessitera une attention particulière dans la définition de la méthodologie. Le choix d'étudier des scénarios de cumuls de pluie a été fait à l'issue d'une réunion avec les services de prévision de Météo-France, qui a permis aussi de définir d'autres demandes concernant cette synthèse par scénarios :

- Cette synthèse devra être multi-réseaux, c'est-à-dire synthétiser l'information de plusieurs réseaux successifs de PE-AROME en même temps. L'intérêt du multi-réseaux sera débattu dans le chapitre 11 à la fin de ce manuscrit.
- La synthèse par scénarios doit éviter de proposer plus de 5 scénarios simultanément pour qu'ils puissent être étudiés par les prévisionnistes en un temps limité.
- Cette synthèse par scénarios doit intégrer à la fois les membres PE-AROME et le modèle déterministe AROME pour que les prévisions puissent être comparées par les prévisionnistes.
- La synthèse par scénarios sera à ce stade monovariée, c'est-à-dire construite avec une seule variable météorologique (RR1h).

Un dernier point concernant ces scénarios est leur emprise spatiale. Cherche-t-on à faire des groupes de membres sur l'ensemble du domaine AROME, à l'échelle nationale ou à une échelle plus locale ? Le choix a été fait de construire une synthèse par

scénarios qui est régionale (de l'ordre d'un quart de France) pour plusieurs raisons. Tout d'abord la représentation de la variabilité des précipitations par seulement 16 membres (ou 32 membres en multi-réseaux) est insuffisante pour couvrir les trop nombreux degrés de libertés qu'implique la prise en compte de l'ensemble du territoire. Il est donc préférable de s'intéresser à des zones plus petites. La deuxième raison est la difficulté à traiter plusieurs évènements météorologiques simultanément. Il est par exemple fréquent qu'un front froid aborde l'Ouest du pays alors qu'un épisode méditerranéen est en cours dans le Sud-Est de la France. Une synthèse par scénarios risquerait de mettre en exergue un comportement moyen entre le front froid et l'épisode méditerranéen alors que les deux structures sont totalement différentes ce qui donnerait peu de sens à la synthèse par scénarios (le problème est largement renforcé par le faible nombre de membres). Pour ces raisons, la synthèse par scénarios est construite sur 5 sous-zones régionales, représentées sur la figure 3.13. Ce découpage a été co-construit avec des prévisionnistes.

3.4.2 Méthode développée

La construction de ces scénarios pose de nombreuses questions. Pour créer des groupes de membres, il est nécessaire de réaliser une étape de classification. Et avant de classer des membres, il est nécessaire de définir une métrique capable de mesurer la similarité entre deux membres. Trois principales interrogations se posent :

- Les groupes de membres sont-ils faits échéance par échéance ou sont-ils faits sur une période de temps de plusieurs heures ? Et si la deuxième option est retenue, comment prendre en compte l'aspect temporel ? Quelle que soit la solution retenue, la notion de scénario implique une cohérence temporelle. On ne peut pas accepter de fortes variations du nombre total de scénarios selon les échéances, où par exemple on observerait 5 scénarios à 15h, 2 scénarios à 16h puis de nouveau 5 scénarios à 17h.

- Une autre interrogation concerne la métrique pour mesurer la similarité de champs de RR1h. La problématique de la double peine rend la plupart des métriques classiques inefficaces. Les méthodes objets, qui permettent de s'affranchir de cette problématique, posent d'autres questions comme : quels types d'objets ? Comment prendre en compte plusieurs objets simultanément ? ou Comment les comparer ? Ces questions nécessitent la définition arbitraire de nombreux paramètres pour construire la métrique. La métrique devient alors très complexe et peu lisible. Il devient également impossible d'exporter facilement la méthode d'une variable météorologique à une autre. La métrique définie pour les pluies sera sûrement totalement à retravailler pour le vent par exemple.

- Après avoir défini la métrique, comment classer les membres ? Il existe en effet de nombreuses méthodes de classification. De plus le chapitre 2 a montré l'existence de deux types d'approche : sans *a priori* ou à partir de structures météorologiques définies en amont (section 2.2.3). Il faudra choisir entre les deux approches.

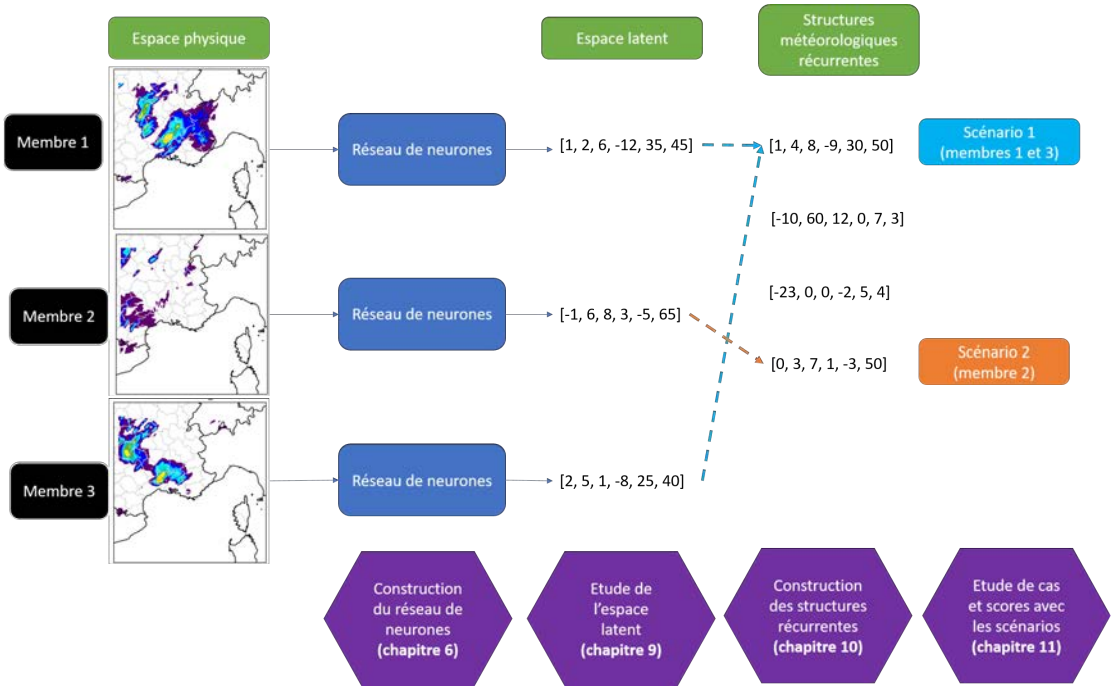


FIGURE 3.14 – *Représentation schématique de la méthode retenue pour construire les scénarios de pluies (exemple pris pour la zone Sud-Est). L'ensemble des valeurs représentées dans l'espace latent et les structures météorologiques récurrentes sont fictives et ne correspondent pas à la réalité. Les différentes étapes sont expliquées dans la section 3.4.2.*

La méthode retenue a été construite en répondant à ces questions et est présentée figure 3.14. Concernant le premier point, il a été choisi de faire des groupes de membres échéance par échéance. La deuxième option complexifie l'approche car elle nécessite entre autres de définir une fenêtre temporelle dont la durée optimale peut largement dépendre de la situation météorologique (on peut s'autoriser des intervalles plus longs pour les situations avec une bonne prévisibilité). Si la construction des groupes de membres échéance par échéance pose des problèmes de continuité temporelle, ce choix pourra être reconsidéré ultérieurement.

L'interrogation sur la métrique montre la difficulté à la définir dans l'espace physique. Un moyen de contourner cette difficulté est de plonger les membres dans un autre espace avec un nombre de dimensions plus restreint où la définition de la métrique sera plus aisée mais en contre-partie l'espace sera plus abstrait. Cette compression de l'information est déjà faite par certains travaux évoqués dans la section 2.2.3 avec une ACP. Cette méthode est cependant connue en météorologie pour être peu efficace sur les champs de precipitation. Une autre méthode émerge des travaux de Karim et al. (2020), qui consiste à projeter les prévisions dans l'espace latent d'un réseau de neurones de type autoencodeur afin de réaliser la classification. Concrètement, l'autoencodeur est un réseau de neurones qui projette le champ de pluie dans un espace latent avant, à partir de cet espace latent, de reconstruire le plus fidèlement possible l'image de départ. L'espace latent contient alors les informations les plus essentielles pour parvenir au mieux à reconstruire le champ de départ. Cette approche nécessite de faire moins d'hypothèses sur le champ de départ qu'avec une ACP, ce qui la rend plus généralisable que cette dernière. Un réseau de neurones de type autoencodeur est donc construit au cours

de cette thèse pour compresser l'information des cumuls de pluies sur 1h. Il fera l'objet du chapitre 6. Chaque champ de pluie est alors représenté dans un espace latent à quelques dizaines de dimensions (vecteurs au centre de la figure 3.14). Cet espace étant abstrait, il est nécessaire de l'étudier pour mieux comprendre les étapes qui suivront. Cette étude de l'espace latent est le sujet traité dans le chapitre 9.

Une fois les membres projetés dans l'espace latent, il faut les classer. Concernant le choix entre regroupement sans *a priori* ou en assignant à une classification pré-existante, le regroupement de membres sans *a priori* impose de faire une classification des 16 membres (ou 32 membres) à chaque échéance. Ce nombre de membres est très faible pour des méthodes de classification et rend son utilisation peu robuste. En particulier, les premiers tests ont montré une grande instabilité temporelle concernant le nombre optimal de scénarios (non montré). Ce problème incite à utiliser plutôt la deuxième option qui présente par ailleurs plusieurs avantages. Avec une méthode similaire à celle développée dans Neal et al. (2016), cf fin de la section 2.2.3, la classification est faite une seule fois sur un très grand nombre de données ce qui rend les méthodes de classification robustes. On définit avec les résultats de la classification des structures météorologiques récurrentes. Comme cette classification est construite en amont sur une climatologie, elle peut être discutée longuement avec les prévisionnistes et ainsi être ajustée selon les besoins. Les structures météorologiques récurrentes définies peuvent également correspondre à des schémas récurrents connus des prévisionnistes. Cette connaissance peut les aider à mieux se représenter le type de temps attendu comme pour les régimes de temps (on sait que le régime de temps NAO- est souvent synonyme de temps humide sur la France par exemple). La méthode utilisée par Neal et al. (2016) est donc choisie pour classer les membres, mais contrairement à ces travaux la classification est faite dans l'espace latent d'un autoencodeur et pas directement dans l'espace physique. Un autre avantage ici de l'espace réduit avec quelques dizaines de dimensions est de permettre une utilisation d'une base de données plus grande pour faire la classification puisque chaque membre est représenté par quelques dizaines de valeurs au lieu de plusieurs milliers. Le chapitre 10 traite de la construction de ces structures météorologiques récurrentes dans l'espace latent qui sont représentées par leurs centroïdes sur la figure 3.14. La question concernant la méthode de classification retenue (K-Means, hiérarchique, ...) sera traitée dans ce même chapitre 10. Maintenant que les structures météorologiques récurrentes sont définies, chaque membre de la PE-AROME du jour est assigné à la structure récurrente qui lui est le plus proche (au sens de la distance euclidienne dans l'espace latent). Les membres qui sont assignés à la même structure météorologique récurrente forment alors un scénario, c'est le cas sur l'exemple de la figure 3.14 des membres 1 et 3. La représentation graphique de la synthèse par scénarios, les études de cas et les scores avec cette approche sont traités dans le chapitre 11, dernier chapitre de ce manuscrit.

Bibliographie

P. Arbogast, O. Pannekoucke, L. Raynaud, R. Lalanne, and E. Mémin. Object-oriented processing of CRM precipitation forecasts by stochastic filtering. *Quart.*

J. Roy. Meteor. Soc., 142(700) :2827–2838, 2016. doi : https://doi.org/10.1002/qj.2871.

N. Atkins and M. S. Laurent. Bow echo mesovortices. part i : Processes that influence their damaging potential. *Mon. Wea. Rev.*, 137 :1497–1513, 2009. doi : 10.1175/2008MWR2649.1.

J. Bjerknes and H. Solberg. Life cycle of cyclones and the polar front theory of atmospheric circulation. *Geophysisks Publikationer*, 3 :1–18, 1922.

C. Calas. *Concepts et méthodes pour le météorologiste - Tome 1, les savoirs : modèles conceptuels et données disponibles*. Météo-France, 2013.

C. Davis, B. Brown, and R. Bullock. Object-based verification of precipitation forecasts. part i : Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, 134 :1772–1784, 2006. doi : 10.1175/MWR3145.1.

E. Ebert. Fuzzy verification of high ?resolution gridded forecasts : a review and proposed framework. *Meteor. Appl.*, 15 :51–64, 2008. doi : 10.1002/met.25.

A. French and M. Parker. Numerical simulations of bow echo formation following a squall line supercell merger. *Mon. Wea. Rev.*, 142 :4791–4822, 2014. doi : 10.1175/MWR-D-13-00356.1.

T. Fujita. *Manual of downburst identification for project NIMROD*. Satellite and Mesometeorology Research Paper 156, 1978.

E. Gilleland, G. Skok, B. G. Brown, B. Casati, M. Dorninger, M. P. Mittermaier, N. Roberts, and L. J. Wilson. A Novel Set of Geometric Verification Test Fields with Application to Distance Measures. *Mon. Wea. Rev.*, 148(4) :1653–1673, 2020. doi : 10.1175/MWR-D-19-0256.1.

L. Goulet. Bow echoes : Conceptual schemes and European relevance. *The European Forecaster*, 2015. URL `http://www.euroforecaster.org/newsletter20/meteofr2`.

Y. Hamidi, L. Raynaud, L. Rottner, and P. Arbogast. Texture-based classification of high-resolution precipitation forecasts with machine-learning methods. *Quart. J. Roy. Meteor. Soc.*, 146(732) :3014–3028, 2020. doi : https://doi.org/10.1002/qj.3823.

D. Harrison, A. McGovern, C. Karstens, J. Demuth, A. Bostrom, I. Jirak, and P. Marsh. Developing Machine Learning for an Operational Environment, 2021. URL `https://drive.google.com/file/d/16RVQCeFX4m0oJaUsq_GgAhrZmlcQvD1S/view?usp=sharing`.

M. Karim, O. Beyan, A. Zappa, I. Costa, D. Rebholz-Schuhmann, M. Cochez, and S. Decker. Deep learning-based clustering approaches for bioinformatics. *Briefings in Bioinformatics*, 22(1) :393–415, 2020. doi : 10.1093/bib/bbz170.

B. Klimowski, M. Hjelmfelt, and M. Bunkers. Radar observations of the early evolution of bow echoes. *Wea. Forecasting*, 19 :727–734, 2004. doi : 10.1175/1520-0434(2004)019<0727:ROOTEE>2.0.CO;2.

P. Markowski and Y. Richardson. *Mesoscale Meteorology in Midlatitudes*. John Wiley and Sons, Ltd, 2010. doi : 10.1002/9780470682104.ch9.

R. Neal, D. Fereday, R. Crocker, and R. E. Comer. A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteor. Appl.*, 23(3) :389–400, 2016. doi : https://doi.org/10.1002/met.1563.

R. Przybylinski. The bow echo : Observations, numerical simulations, and severe weather detection methods. *Wea. Forecasting*, 10 :203–218, 1995. doi : 10.1175/1520-0434(1995)010<0203:TBEONS>2.0.CO;2.

L. Raynaud, I. Pechin, P. Arbogast, L. Rottner, and M. Destouches. Object-based verification metrics applied to the evaluation and weighting of convective-scale precipitation forecasts. *Quart. J. Roy. Meteor. Soc.*, 145 :1992–2008, 2019. doi : 10.1002/qj.3540.

A. Rossa, P. Nurmi, and E. Ebert. Overview of methods for the verification of quantitative precipitation forecasts. 2008. doi : 10.1007/978-3-540-77655-0_16.

L. Rottner, P. Arbogast, M. Destouches, Y. Hamidi, and L. Raynaud. The similarity-based method : a new object detection method for deterministic and ensemble weather forecasts. *Advances in Science and Research*, 16 :209–213, 2019. doi : 10.5194/asr-16-209-2019.

P. Santurette and A. Joly. ANASYG/PRESYG, Météo-Francés new graphical summary of the synoptic situation. *Meteorological Applications*, 9(2) :129–154, 2002. doi : https://doi.org/10.1017/S1350482702002013.

M. A. Shapiro and D. Keyser. *Fronts, Jet Streams and the Tropopause*, pages 167–191. 1990. ISBN 978-1-944970-33-8. doi : 10.1007/978-1-944970-33-8_10.

N. A. Stuart, G. Hartfield, D. M. Schultz, K. Wilson, G. West, R. Hoffman, G. Lackmann, H. Brooks, P. Roebber, T. Bals-Elsholz, H. Obermeier, F. Judt, P. Market, D. Nietfeld, B. Telfeyan, D. DePodwin, J. Fries, E. Abrams, and J. Shields. The Evolving Role of Humans in Weather Prediction and Communication. *Bull. Amer. Meteor. Soc.*, 103(8) :1720–1746, 2022. doi : 10.1175/BAMS-D-20-0326.1.

Deuxième partie

Deep learning : des outils pertinents pour traiter les données atmosphériques

Chapitre 4

Deep learning : concept et applications

Les masses de données manipulées en météorologie ne cessent de croître grâce à des machines plus puissantes, des observations plus nombreuses ou encore grâce à des prévisions à de plus en plus haute résolution et le développement des prévisions d'ensemble. Cette augmentation des quantités de données à manipuler conduit à un intérêt grandissant pour des méthodes de machine learning (ML) et de deep learning [1] (DL) de plus en plus complexes afin de les traiter efficacement. Dans ce chapitre les notions d'intelligence artificielle, de machine learning et de deep learning sont d'abord définies dans une première partie ainsi que les principes généraux. Les points suivants se concentreront sur les réseaux de neurones et les différents modules qui les composent. Plusieurs hyperparamètres seront définis et soulignés en noir dans ce chapitre. Enfin, nous finirons par décrire comment des méthodes de DL sont aujourd'hui utilisées ou en cours de développement en météorologie en allant des observations aux post-traitements des prévisions numériques.

4.1 Définitions et principes

4.1.1 Définitions des concepts d'IA, machine learning et deep learning

Les concepts de machine learning et deep learning appartiennent au domaine plus vaste de l'Intelligence Artificielle (IA). L'IA est née dans les années 1950 et peut se définir selon Chollet (2021) comme *la volonté d'automatiser des tâches intellectuelles normalement réalisées par des humains* [2]. Cette automatisation est souvent faite aujourd'hui à l'aide d'ordinateurs ou de supercalculateurs qui servent à faire des calculs plus rapidement qu'un humain. L'IA peut être schématiquement séparée en 2 grands types de méthode : l'algorithmie classique d'une part qu'on

1. Ces termes se traduisent par apprentissage automatique et apprentissage profond. Mais le vocabulaire anglais étant plus courant, il sera utilisé dans la suite de ce manuscrit.
2. Bien qu'aucune définition ne fasse véritablement consensus à ce jour.

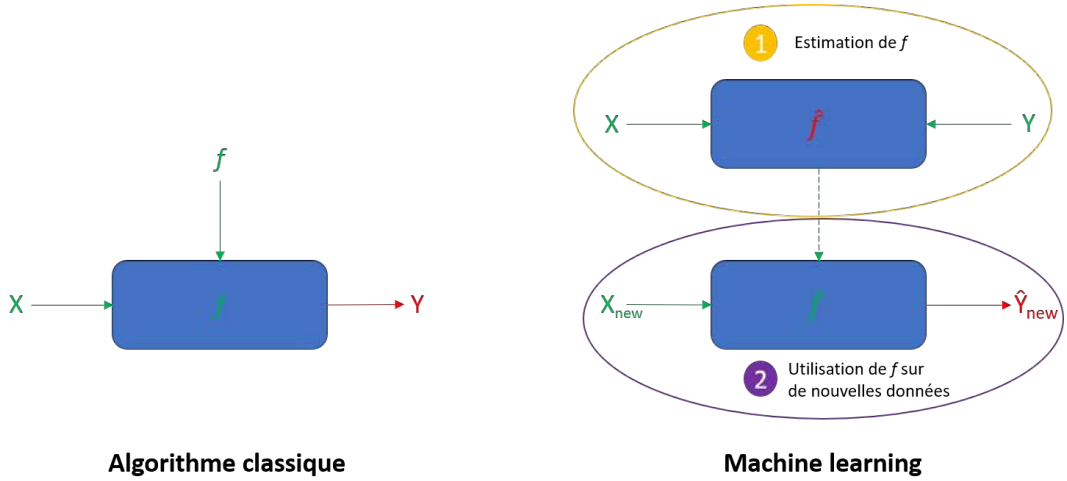


FIGURE 4.1 – *Comparaison des objectifs d'un algorithme classique et d'un algorithme de machine learning. Les données en entrée des algorithmes sont en vert, les données en sortie en rouge.*

nomme parfois IA symbolique et les méthodes d'apprentissage machine d'autre part (ML/DL), lesquelles font l'objet de ce chapitre. On peut décrire tout d'abord l'algorithmie classique par un jeu de données X et des règles qu'on peut représenter à l'aide d'une fonction f qui sont fournies à la machine. La machine applique alors f sur les données X pour former les sorties $Y = f(X)$ (algorithme classique, figure 4.1). L'apprentissage machine s'appuie sur des concepts différents où l'objectif est d'estimer la meilleure fonction $\hat{f}$ possible pour lier les données d'entrées X et les données de sortie Y, qui deviennent donc aussi des données à fournir à notre algorithme de machine learning (étape 1, figure 4.1). Il existe en effet de nombreuses applications où les données X, Y sont disponibles mais où la fonction f est trop complexe pour être estimée directement à l'aide de nos connaissances (reconnaissance d'objet sur une image par exemple). L'objectif est à terme de prédire à partir d'un nouveau jeu de données X_{new} et de la fonction $\hat{f}$ préalablement construite, des nouvelles valeurs $\hat{Y}_{new}$ (étape 2, figure 4.1). Pour compenser cette méconnaissance totale ou partielle de f, ces méthodes requièrent la plupart du temps de gros jeux de données X, Y et nécessitent une puissance de calcul importante. De façon générale, le machine learning ou apprentissage machine désigne *l'étude des algorithmes capables de s'améliorer de façon autonome en apprenant à partir d'un jeu de données* (Mitchell, 1997). Il existe un grand nombre de méthodes de ML comme la régression linéaire, la régression logistique, les Support Vector Machines (SVMs), ou les forêts aléatoires (Hastie et al., 2009). Les méthodes de deep learning qui sont aussi des méthodes de ML seront les seules à être introduites dans la suite de ce chapitre.

Pourquoi une approche deep learning ?

D'après LeCun et al. (2015), *ces algorithmes sont composés de plusieurs couches pour représenter les données à plusieurs niveaux d'abstraction.* Le passage d'une couche à l'autre se fait par des modules non linéaires, le premier niveau désignant les données d'entrée X. Cette construction sur plusieurs niveaux permet aux al-

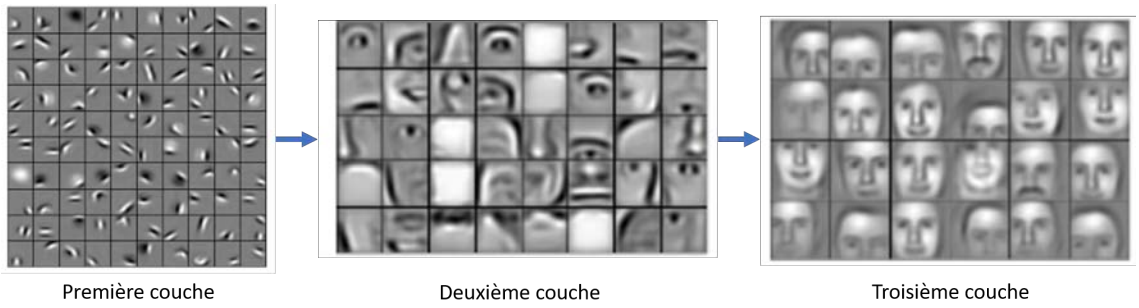


FIGURE 4.2 – *Représentation de l'information à différentes couches d'un algorithme de DL pour la reconnaissance de visage. D'après Lee et al. (2009).*

gorithmes de DL d'extraire des informations pertinentes à différentes échelles à partir des données d'entrée X. Ces méthodes sont ainsi performantes et populaires dans de nombreux domaines comme le traitement d'image, du son ou de la langue (suggestion de mots, traduction). La figure 4.2 illustre ces différents niveaux avec des premières couches qui recherchent des détails simples, principalement des gradients. L'information dans les couches suivantes se complexifie pour rechercher des yeux ou la bouche par exemple et enfin la dernière couche a une vue d'ensemble avec des visages complets.

4.1.2 Deep learning : principes de fonctionnement

Les différentes couches des algorithmes de DL entrent dans la catégorie des réseaux de neurones pour l'immense majorité des cas (Chollet, 2021). C'est donc le principe de fonctionnement des réseaux de neurones qui sera détaillé dans ce paragraphe. Un algorithme de DL cherche à estimer une fonction f comme toute méthode de ML. Cette fonction est paramétrée à l'aide de poids p qui définissent le comportement de chaque couche du réseau de neurones (on les appelle aussi paramètres entraînables). Le rôle de ces poids est différent selon les couches et sera détaillé dans la partie suivante. Ce sont donc ces poids p qu'un algorithme de DL cherche à optimiser.

Afin de rendre plus concret les explications qui vont suivre sur la méthode d'optimisation des poids, un exemple très simple est utilisé. Le réseau de neurones tente de reproduire la fonction $f(x) = 2x+1$. Les deux poids du réseau de neurones seront (a, b) et modéliseront la fonction $f(x) = ax + b$. Le réseau de neurones dispose des données d'entrée $X = [1, 3]$ et des sorties $Y = [2 \times 1 + 1, 2 \times 3 + 1] = [3, 7]$. Concrètement les poids p sont souvent initialisés aléatoirement, avec ici $a = -1$ et $b = 2$ (étape 1, figure 4.3). Une première prédiction $\hat{Y}$ est alors faite à partir des données X, ce qui donne dans notre exemple $\hat{Y} = [-1\times1+2, -1\times3+2] = [1, -1]$ (étape 2, figure 4.3). L'algorithme a alors besoin d'évaluer la pertinence de ses poids p pour établir de bonnes prédictions. Pour cela, les prédictions $\hat{Y}$ sont comparées avec les vraies sorties Y à l'aide d'une fonction coût L [3] (étape 3, figure 4.3). Plus la valeur de L est petite, plus la sortie $\hat{Y}$ ressemble à Y. Cette fonction coût est également à fournir par l'utilisateur, tout comme le couple (X, Y). Dans

3. on la nomme aussi fonction de perte ou loss en anglais d'où la notation L

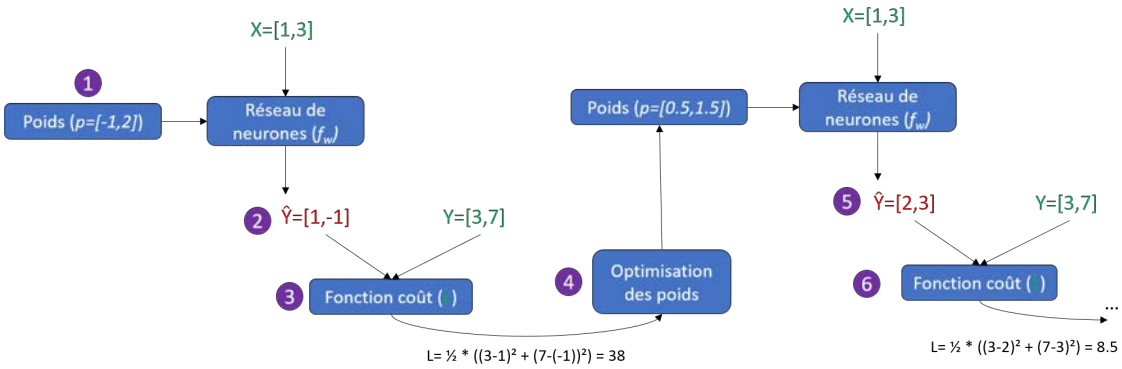


FIGURE 4.3 – *Phase d'entraînement d'un réseau de neurones à partir des données X, Y et d'une fonction coût L (en vert). Les sorties du réseau de neurones $\hat{Y}$ sont en rouge.*

notre exemple, la fonction coût sera la Mean Squared Error (MSE). En partant de poids aléatoire, il est très probable que les prédictions faites par le réseau de neurones soient très mauvaises. Dans l'exemple présenté, la MSE des premières prédictions est égale à 38. Mais l'évaluation des performances par la fonction coût L va permettre à l'aide d'une méthode d'optimisation de mettre à jour les poids p dans le but d'améliorer la sortie $\hat{Y}$ pour minimiser la fonction coût L (étape 4, figure 4.3). Cette optimisation des poids se fait à l'aide d'une descente de gradient [4] qui sera détaillée dans la section 4.1.3. Dans l'exemple de la figure 4.3, $a = 0.5$ et $b = 1.5$ après optimisation. Les étapes 2 à 4 sont répétées plusieurs fois (étapes 6 et 7, figure 4.3) pour aboutir à un modèle de DL qui fait de bonnes prédictions $\hat{Y}$ en partant de X. La mise à jour des poids dans l'exemple a permis de réduire la MSE de 38 à 8.5 et peut être réitérée pour aboutir à une MSE proche de 0. On obtient alors un réseau de neurones satisfaisant qui sait correctement reproduire les sorties Y à partir des données X. Cette phase de prédiction, évaluation par L et mise à jour des poids est appelée la phase d'apprentissage. On dit à la fin de cette étape que le réseau de neurones a été entraîné.

4.1.3 Phase d'apprentissage : notion de batch, d'epoch et de learning rate

Durant l'apprentissage, les données X, Y sont donc utilisées pour entraîner le réseau de neurones. Cependant, les données sont préalablement découpées en paquets qu'on nomme batch (leur taille est un hyperparamètre noté B_{size}). La taille de ces paquets (le batch size) est alors un paramètre à fixer. Ces batchs sont utilisés de façon successive pour réaliser la descente de gradient et ajuster les poids. On dit qu'une epoch est terminée lorsque l'ensemble des batchs a été utilisé pour l'entrainement. Le processus est alors répété. Plusieurs epochs sont en effet nécessaires en général pour réaliser l'entraînement complet d'un réseau de neurones. Le nombre maximal d'epoch est un hyperparamètre à fixer noté E_{max}. Enfin on appelle itération le nombre de batchs nécessaires pour finir une epoch.

4. La descente de gradient est un algorithme d'optimisation qui permet de converger vers le minimum d'une fonction convexe de façon itérative.

Par exemple, si une base de données compte 3 840 images et qu'un batch size de 32 images est fixé alors 120 itérations sont nécessaires.

L'évaluation de la fonction coût L sur ces batchs permet de mettre à jour les poids grâce à une descente de gradient. La mise à jour d'un poids p_i à la i-ème itération peut se schématiser sous la forme :

$$p_i = p_{i-1} - \alpha \cdot \frac{\partial L}{\partial p} \tag{4.1}$$

Avec α le learning rate qui est un paramètre qui contrôle donc la vitesse de convergence de l'algorithme. Cette descente de gradient peut se schématiser comme sur la figure 4.4. Plusieurs méthodes existent comme la Batch Gradient Descent (BGD) où la mise à jour des poids est faite avec l'ensemble du jeu de données en une fois. Cette méthode tend cependant à lisser l'information des extrêmes du jeu de données avec des temps de calcul pouvant être longs. Une méthode alternative est la Stochastic Gradient Descent (SGD) où la mise à jour des poids est faite avec un seul échantillon de la base de données ce qui aboutit à une convergence plus bruitée comme sur l'exemple de la figure 4.4. Enfin une méthode intermédiaire est la Mini Batch Gradient Descent (MBGD) qui utilise un groupe d'échantillons. Ces représentations restent schématiques et ne correspondent pas à celle qui est implémentée par exemple dans la méthode d'optimisation Adam qui sera utilisée dans la suite de ce manuscrit. Celle-ci en particulier utilise un algorithme de convergence qui garde en mémoire la mise à jour des poids à l'étape précédente pour faire la suivante afin de lisser cette convergence par rapport à une méthode SGD (méthode du moment). L'optimiseur est détaillé dans l'article de Kingma and Ba (2014). La descente de gradient ne sera pas plus détaillée dans la suite de ce manuscrit.

4.1.4 Base d'apprentissage, de validation et de test

L'estimation de $\hat{f}$ par une descente de gradient sur les données X, Y n'est cependant pas sans risque. En effet, le paragraphe dédié au ML a précisé que $\hat{f}$ était réutilisée sur de nouvelles données X_{new}. Mais il n'est pas possible de savoir en procédant ainsi si les prédictions de $\hat{f}$ sont aussi bonnes sur les nouvelles données X_{new} que sur les données X. Il existe en effet un risque que la fonction $\hat{f}$ ne soit pas utilisable sur de nouvelles données car elle a appris des caractéristiques qui dépendent trop fortement du jeu de départ X, Y et qui ne sont donc pas généralisables. On parle de surapprentissage ou *overfitting*. Pour éviter cet écueil, plusieurs jeux de données indépendants sont créés.

Le premier est utilisé pour réaliser l'entrainement de f, on parle de **base d'entraînement ou d'apprentissage**. Avant de lancer cet entrainement, cette base est elle-même divisée en deux aléatoirement dans des proportions autour de 75%/25%. Les 75% du jeu de données servent à faire l'entraînement comme décrit dans la section 4.1.2. On calcule alors différents scores comme la valeur de la fonction coût selon les epochs pour tracer *la courbe d'apprentissage*. Ces mêmes scores sont calculés sur les 25% restants, ces 25 % dessinent alors *la courbe de validation*. L'objectif est alors de tracer les deux courbes ainsi obtenues (figure 4.5). Par construction, la courbe d'apprentissage ne peut que décroître avec les

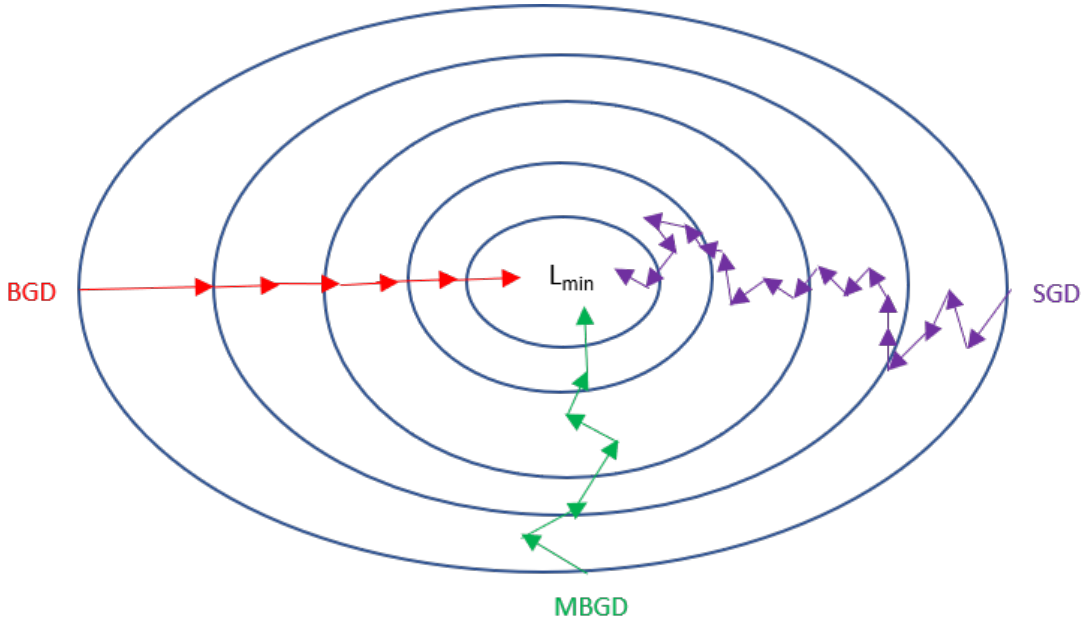


FIGURE 4.4 – *Représentation schématique de la descente de gradient vers le minimum de la fonction coût L_{min} par 3 méthodes différentes : par batch (BGD en rouge), minibatch (MBGD en vert) ou stochastique (SGD en violet). Chaque flèche représente une itération.*

epochs (descente de gradient). Ce n'est cependant pas le cas pour la courbe de la base de validation et les poids obtenus à l'epoch correspondant au minimum de la courbe sont considérés comme optimaux.

La base d'apprentissage permet de réaliser l'entrainement mais la sélection des hyperparamètres du réseau de neurones comme le nombre de couches ne peut pas se faire avec la précédente base car la subdivision est faite de façon aléatoire et donc la sélection des hyperparamètres se ferait sur des jeux de données différents. Pour cela, une autre base indépendente de la précédente est créée, la **base de validation pour les hyperparamètres**. Elle permet une comparaison des performances selon différentes combinaisons d'hyperparamètres sur un jeu de données commun mais indépendant de celui utilisé pour l'apprentissage.

Cependant les bases d'apprentissage et de validation ne suffisent pas en pratique. En effet si la sélection des hyperparamètres est faite à l'aide des scores sur la base décrite précédemment alors ces scores sont biaisés. Il est nécessaire de créer un troisième et dernier jeu de données indépendant, **la base test**, qui servira à calculer les scores du réseau de neurones dont la combinaison d'hyperparamètres a été retenue. La représentativité de chaque base est un point important pour assurer une bonne évaluation des réseaux de neurones.

4.2 Couches d'un réseau de neurones et fonction coût

Cette section détaille les différents modules utilisés par les réseaux de neurones dans les parties suivantes de ce manuscrit en s'intéressant séparément aux

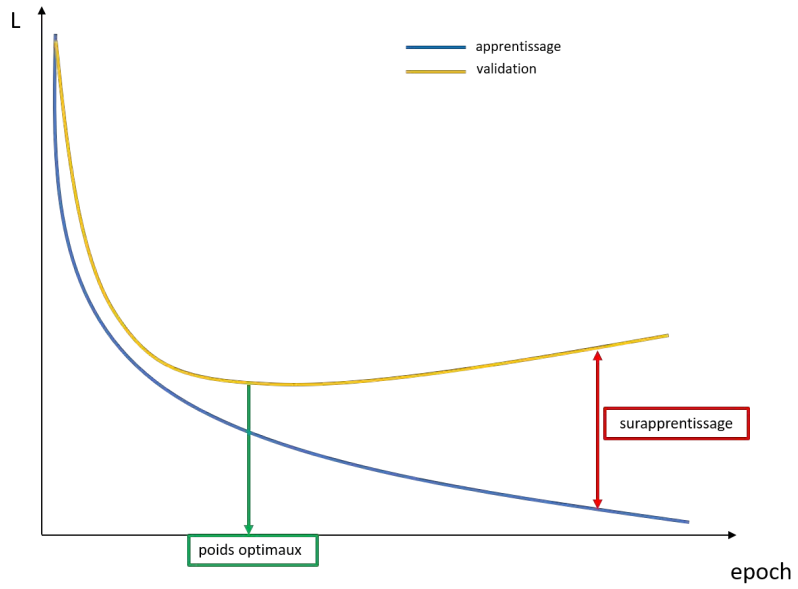


FIGURE 4.5 – *Représentation schématique de la fonction coût selon les epochs pour la courbe d'apprentissage (bleu) et la courbe de validation (orange).*

couches du réseau de neurones et aux fonctions coût. Ce sont donc les combinaisons de ces « briques » qui formeront un réseau de neurones complet. Les différents hyperparamètres à fixer seront récapitulés à la fin de celle-ci.

4.2.1 Couches d'un réseau de neurones

Couche dense et fonction d'activation

Une des couches les plus simples afin de traiter des données pour un réseau de neurones est la couche dense (figure 4.6). Ces couches peuvent servir par exemple à établir un lien en un point entre des variables météorologiques (humidité, vent, température) et l'occurrence d'un évènement (chute de grêle, brouillard). Cette couche peut se représenter comme plusieurs neurones n_j dont le nombre m est un hyperparamètre à fixer. Chaque neurone utilise en entrée l'ensemble des données x_i (désignant les données d'entrée du réseau de neurones ou la couche précédente), pondérées par des poids p_{ji} puis sommés. Un terme de biais b_j est généralement ajouté à cette somme pondérée. Une fonction d'activation σ est ensuite appliquée pour introduire des non linéarités dans la couche. L'opération peut alors s'écrire sous la forme :

$$Y = \sigma(P \cdot X + B) \tag{4.2}$$

Avec X les données d'entrée avec une dimension n, P la matrice des poids de dimension $m \times n$, B le vecteur des biais de dimension m et enfin Y la sortie de dimension m. Les poids des matrices P et B sont modifiés durant la phase d'entraînement. Concernant σ, plusieurs fonctions d'activation existent. On retrouve parmi les plus courantes la fonction ReLU pour Rectified Linear Unit (figure 4.7) qui sera utilisée dans la suite de ce manuscrit :

$$ReLU(x) = max(0, x) \tag{4.3}$$

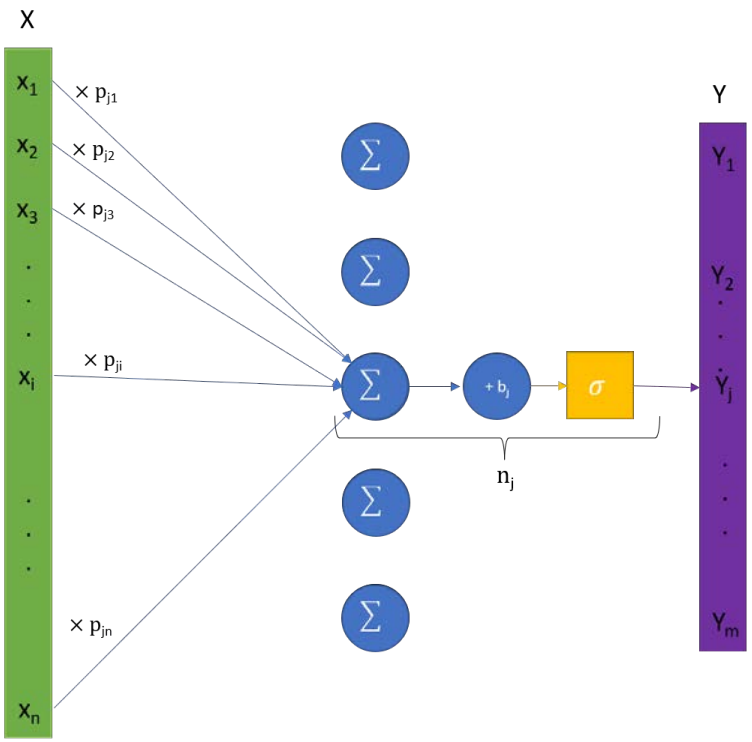


FIGURE 4.6 – *Représentation schématique d'une couche dense entre les données d'entrée de la couche X et les données de sortie de la couche Y.*

Couche de convolution

La couche dense n'est cependant pas efficace pour traiter les données spatiales pour plusieurs raisons. Le premier problème vient de la taille de la matrice P qui croit très rapidement si on augmente la taille de l'entrée ce qui augmente le temps de calcul et le risque de surapparentissage. La couche dense n'est de plus pas efficace pour apprendre des corrélations spatiales entre un point de grille et ses voisins les plus proches car chaque neurone prend en compte des informations issues de la totalité des entrées. En d'autres termes, la couche dense ne permet pas de rechercher des motifs particuliers dans une image ce qui est pourtant essentiel pour traiter ce type de données. C'est le rôle de la couche de convolution (LeCun and Bengio, 1995) qui va réaliser une convolution entre les données d'entrée X et une matrice de coefficient C qui est plus petite que les données d'entrée et qu'on appelle souvent un filtre (figure 4.8). La taille du filtre $h_{conv} \times w_{conv}$ qui s'applique sur l'input de taille $H \times W$ est un hyperparamètre à fixer préalablement. Ce filtre va alors être appliqué localement sur l'image avec une multiplication terme à terme (figure 4.8). L'opération est réitérée à intervalle régulier, cet intervalle est le stride s_{conv} et est aussi un hyperparamètre à fixer en amont. Enfin, l'application de ces filtres pose des problèmes pour la gestion des bords de l'image et peut aboutir à une utilisation tronquée de l'image de départ. Ce problème peut-être compensé en ajoutant artificiellement des valeurs sur le bord, on parle de *padding* (figure 4.9). Un 0-padding sera utilisé, c'est-à-dire qu'on ajoutera des zéros sur les bords de l'image en entrée. D'autres méthodes existent comme une reproduction en miroir

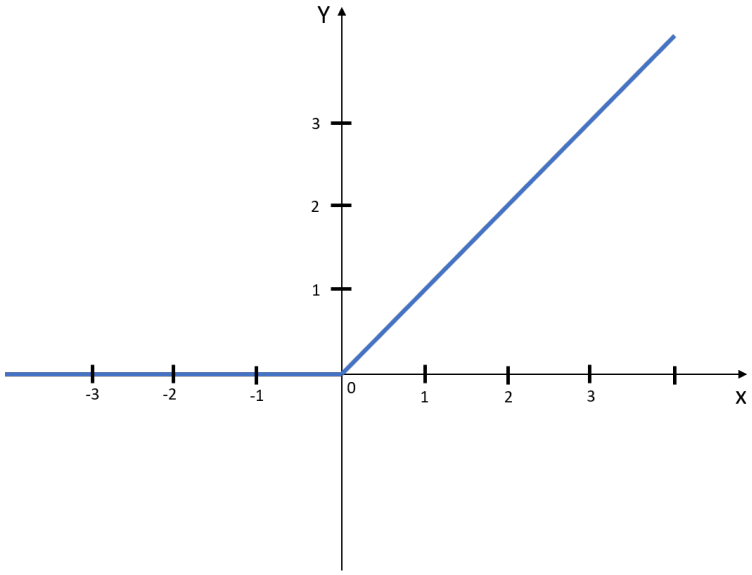


FIGURE 4.7 – *Représentation de la fonction ReLU*

des valeurs aux bords de l'image mais ne seront présentées dans ce manuscrit. Le nombre de colonnes ou lignes à rajouter sur les bords est noté p. La hauteur de l'image en sortie peut alors se déduire de la formule suivante :

$$H_s = \frac{H - h_{conv} + 2p}{s_{conv}} + 1 \tag{4.4}$$

La formule pour la largeur W_s est très similaire à la formule pour H_s. On fixe en général p pour que H_s soit égal à la partie supérieure de H/s (de même pour W_s). Un terme de biais est également ajouté comme pour la couche dense ainsi qu'une fonction d'activation. La formulation mathématique est alors très proche de celle de la couche dense :

$$Y = \sigma(C * X + B) \tag{4.5}$$

$*$ désignant la convolution entre les données d'entrée et le filtre. Les matrices C et B sont les parties entraînables de la couche. Une couche de convolution se compose non pas d'un mais de plusieurs filtres qui recherchent des motifs différents comme sur la figure 4.2 (gradients horizontaux, verticaux, maximum local ...). Le nombre de filtres N_{conv} dans une couche de convolution doit aussi être fixé préalablement. Le problème est ici présenté en deux dimensions mais il se généralise en trois, si on a plusieurs variables météorologiques par exemple, en ajoutant aussi une dimension au filtre C. Ces couches de convolutions combinées à une ou des couches denses sont très utilisées dans les réseaux de neurones qui font de l'annotation d'image (détecter si une entrée est une image de chat, de chien, de voiture,...).

Couche de pooling

La couche de convolution ne permet pas toujours de réduire la taille de l'image d'entrée (en particulier si le stride est fixé à 1 avec un padding), ce qui est pourtant primordial si on souhaite compresser l'information pour aboutir à des décisions

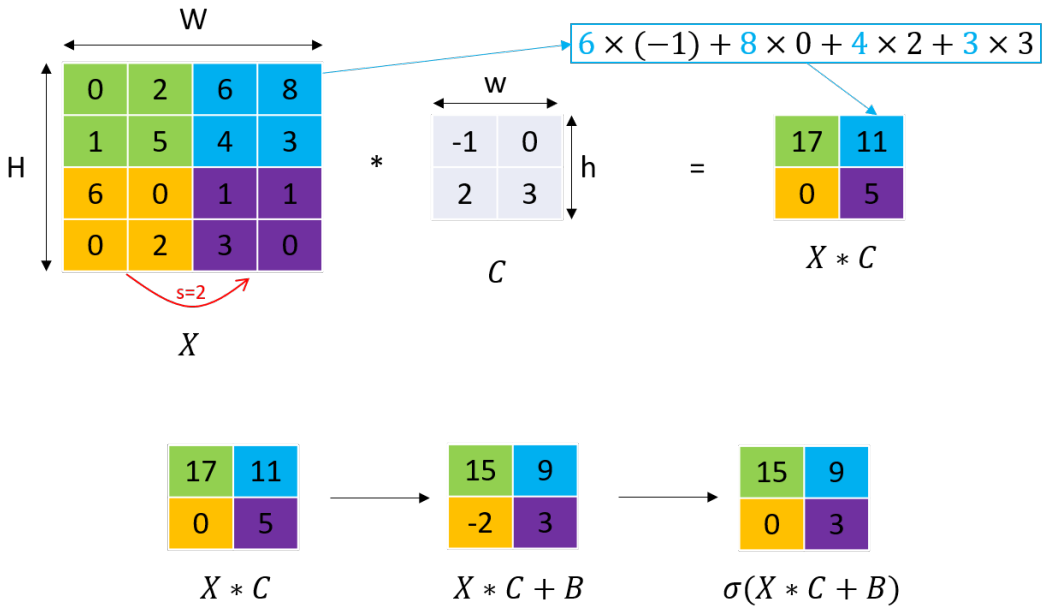


FIGURE 4.8 – *Représentation schématique d'une couche de convolution pour une entrée X de taille 4×4 avec un filtre C de taille 2×2 et un stride de 2. Le terme de biais est ici un scalaire égal à -2 (qui s'applique sur chaque point de grille) et la fonction d'activation σ est la fonction ReLU.*

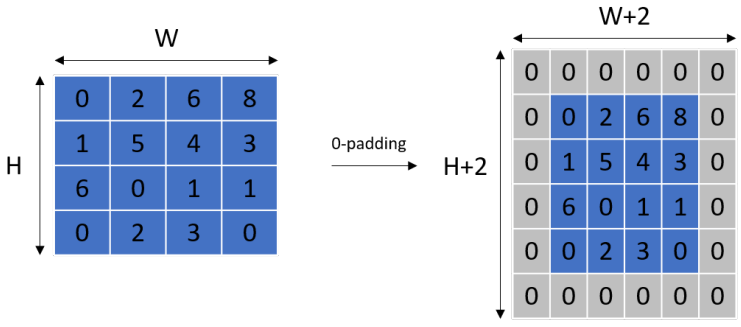


FIGURE 4.9 – *Représentation schématique d'un 0-padding avec ajout d'une colonne et d'une ligne sur les bords (cas $p = 1$).*

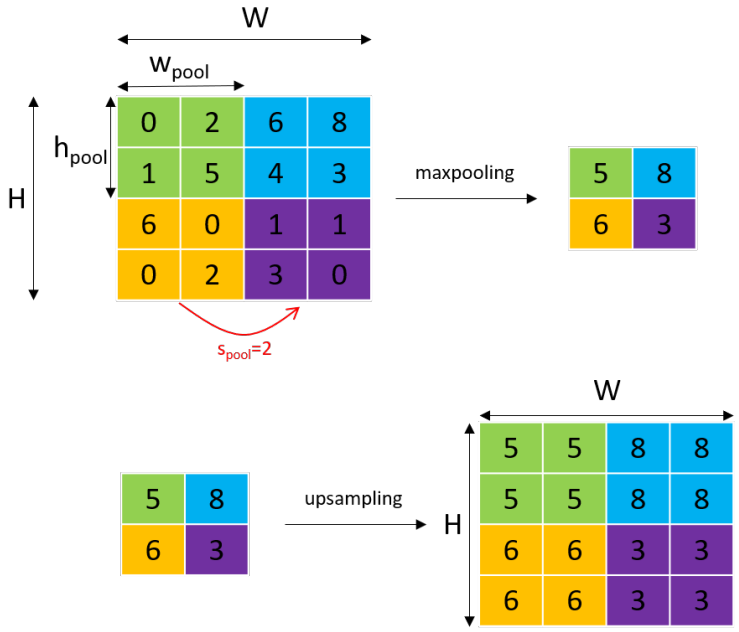


FIGURE 4.10 – *Représentation schématique de la couche de maxpooling pour une taille* 2×2 *et un stride de 2 (haut). Représentation schématique de la couche de upsampling (bas).*

simples comme pour l'exemple précédent (image de chat oui/non ?). C'est le rôle de la couche de pooling (figure 4.10, haut). Cette couche a un comportement similaire à la couche de convolution. Le pooling s'applique aussi localement sur une portion de l'image de taille $h_{pool} \times w_{pool}$, l'opération est répétée avec un stride s_{pool}. Différents types de pooling existent en réalisant la moyenne de l'entrée ou en sélectionnant seulement le maximum par exemple. Cette deuxième option est celle retenue par la suite dans ce manuscrit, on parle alors de maxpooling. Cette opération se fait avec un stride plus grand que 1 et aboutit donc à une diminution de la taille de la sortie par rapport à l'entrée. Un padding est souvent appliqué également car les mêmes problèmes se posent que pour la couche de convolution. Le padding permet alors de fixer la sortie avec une taille $H_{pool} \times W_{pool}$:

$$H_{pool} = \lfloor \frac{H - h_{pool}}{s_{pool}} \rfloor + 1 \tag{4.6}$$

La couche de pooling a principalement pour objectif de concentrer l'information. L'application d'autres couches comme des couches de convolution après une couche de maxpooling permettra alors d'extraire de l'information plus grande échelle. En effet, le champ récepteur [5] sera alors plus grand car deux points de grille voisins coderont des informations issues de sous-domaines de plus grande taille sur l'entrée X. Par exemple appliquer un filtre 2×2 après maxpooling sur la figure 4.10 permet de prendre en compte simultanément de l'information issue de l'ensemble de la grille de départ 4×4. Le même effet peut être obtenu en appliquant successivement des couches de convolution avec un stride différent de 1.

5. le champ récepteur désigne les données d'entrées capables d'induire une réponse différente en sortie de la couche pour un neurone donné

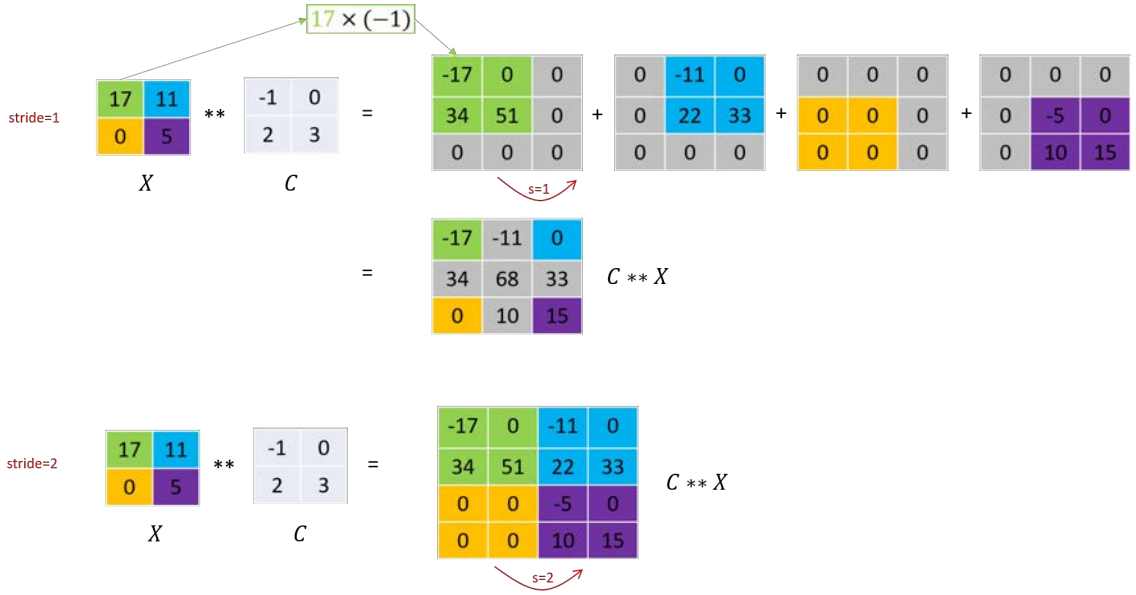


FIGURE 4.11 – *Représentation schématique de la couche de déconvolution. Exemple avec un stride égal à 1 (haut) et un exemple avec un stride égal à 2 (bas).*

Couche de suréchantillonage ou upsampling

Les couches précédemment décrites permettent de faire de la classification d'image (trouve-t-on un chat oui ou non ?) mais ne permettent pas à elles seules de faire de la segmentation d'image (Dessiner un contour sur l'image de départ pour repérer la position du chat). Pour faire cela, il est en effet nécessaire après une phase de compression ou codage de l'information de la décoder en créant une image de sortie de la taille de l'image d'entrée. Cette opération peut se faire à l'aide de couches de suréchantillonage et de déconvolutions.

La couche de suréchantillonage joue le rôle inverse de la couche de maxpooling (figure 4.10, bas). Elle cherche à obtenir une sortie avec une taille plus grande que l'entrée. Pour cela, chaque valeur de la grille de départ est répétée un nombre de fois fixé (n_{rep}) dans toutes les directions (1 fois sur l'exemple de la figure 4.10 permettant de doubler la taille du champ).

Couche de déconvolution

La couche de déconvolution ou de convolution transposée joue un rôle similaire à la couche de suréchantillonage (figure 4.11). Elle peut se penser comme l'opération opposée à la convolution. L'objectif est ici d'obtenir une sortie Y de taille supérieure ou égale à l'entrée X. Un filtre est défini comme pour une couche de convolution. Chaque valeur en entrée de la grille est dupliquée pour appliquer le filtre C. Les grilles ainsi obtenues sont sommées pour obtenir la sortie Y (figure 4.11, haut). Un terme de biais et une fonction d'activation peuvent aussi être appliqués comme pour une couche de convolution. Ce terme de biais et le ou les filtres C sont les parties entrainables de la couche. La taille du filtre C (h_{dec}/w_{dec}), le nombre de filtres (N_{dec}) ainsi que le stride s_{dec} (représenté sur la figure 4.11) sont des hyperparamètres à fixer. Un padding peut aussi être appliqué pour fixer la taille de sortie Y égale au stride multiplié par la taille d'entrée.

Méthode de régularisation : le dropout

La couche de dropout peut s'appliquer sur les couches précédemment citées. Elle a pour objectif de combattre l'overfitting durant l'apprentissage en désactivant de façon aléatoire les sorties de la couche sélectionnée (Srivastava et al., 2014). La probabilité de désactiver une sortie est un hyperparamètre noté $\underline{p}$ (la valeur de sortie du neurone est remplacée par un 0). Ces neurones désactivés ne sont pas toujours les mêmes à chaque epoch. L'objectif du dropout est ici d'introduire délibérément du bruit pour éviter au réseau d'apprendre des relations entre les données qui ne seraient pas significatives. Tous les neurones sont réactivés durant la phase de test, mais chaque sortie est pondérée par $1-p$ car le nombre de neurones activés n'est plus le même que durant la phase d'apprentissage. En pratique, la pondération est plutôt faite durant la phase d'entraînement pour justement ne pas avoir à la faire plus tard en phase de test (Chollet, 2021).

Softmax

Une couche softmax peut être utilisée en fin de réseau afin de transformer la sortie du réseau en probabilité. Ce qui est très utile en particulier pour les problèmes de classification (probabilité d'observer un chat sur l'image). Elle transforme un vecteur d'entrée X (x_i, $i \in [1, N]$) en une sortie $s(X)$ tel que :

$$s(x_i) = \frac{e^{x_i}}{\sum_{k=1}^{N} e^{x_k}} \tag{4.7}$$

D'autres couches existent pour construire des réseaux de neurones. On peut penser aux couches d'attention (Vaswani et al., 2017) ou aux couches de Batch Norm (Ioffe and Szegedy, 2015) par exemple. Elles ne sont pas présentées ici car elles ne seront pas utilisées dans la suite de ce manuscrit.

4.2.2 Fonction coût

Comme pour les couches d'un réseau de neurones, il existe également de nombreuses fonctions coût selon les applications. Cependant, les mêmes fonctions ne sont pas utilisées si la variable de sortie peut prendre un ensemble fini de valeurs (on parle de classification ou de problème catégoriel, les détections d'objets appartiennent à cette catégorie) ou si la variable de sortie varie de façon continue (on parle de régression, c'est le cas si on cherche à prévoir une température par exemple). Seulement deux fonctions coût seront utilisées dans ce manuscrit et présentées dans cette partie.

Classification

Pour les problèmes de classification, l'entropie croisée pondérée sera utilisée. Elle convient en particulier lorsque les effectifs des classes sont très déséquilibrés, ce qui sera le cas de notre application. Pour un problème binaire (classe 0 et 1) et pour un ensemble de données $(x_i)_{i\in[1,N]}$, le réseau de neurone associe les probabilités $(p_i)_{i\in[1,N]}$ d'appartenir à la classe 1. Des labels Y_i sont fournis pour

TABLE 4.1 – *Hyperparamètres à régler selon 3 catégories : phase d'entrainement/apprentissage du réseau, architecture du réseau et fonction coût. Les hyperparamètres pour l'architecture sont détaillés couche par couche.*

Entraînement du réseau	
taille des batchs (B_{size}), nombre maximal d'epochs (E_{max}), learning rate (α)	
Architecture du réseau	
Couche	hyperparamètre
Dense	nombre de neurones (m)
Convolution	nombre de filtres (N_{conv}), taille des filtres (h_{conv}/w_{conv}), stride (s_{conv})
Pooling	taille des filtres (h_{pool}/w_{pool}), stride (s_{pool})
Suréchantillonage	nombre de répétitions (n_{rep})
Déconvolution	nombre de filtres (N_{dec}), taille des filtres (h_{dec}/w_{dec}), stride (s_{dec})
Dropout	probabilité d'éteindre un neurone (p)
Fonction coût	
Pour la classification, poids des classes (w_0 et w_1)	

désigner les bonnes classes d'appartenances des données ($Y_i = 1$ si la donnée appartient à la classe 1, 0 sinon). Les classes sont pondérées par des poids w_0 et w_1. La fonction L s'écrit alors :

$$L = -\frac{1}{N}\sum_{k=1}^{N} w_0\,(1 - Y_k)\,log(1 - p_k) + w_1\,Y_k\,log(p_k) \tag{4.8}$$

Les poids w_0 et w_1 sont des hyperparamètres à fixer.

Régression

Pour les problèmes de régression, la MSE (Mean Squared Error) sera utilisée entre les données de sortie du réseau de neurones $\hat{y}_i, i \in [1, N_{gp}]$ et les valeurs de référence y_i :

$$L = \frac{1}{N_{gp}}\sum_{k=1}^{N_{gp}} (\hat{y}_i - y_i)^2 \tag{4.9}$$

Il n'y a donc ici aucun hyperparamètre à fixer.

4.2.3 Récapitulatif des hyperparamètres

Les différents hyperparamètres à fixer pour entraîner un réseau de neurones sont rappelés dans le tableau 4.1. Cette liste n'est pas tout à fait exhaustive car le choix de la fonction d'activation peut aussi être considéré comme un hyperparamètre. Quelques simplifications ont aussi été apportées. Par exemple, un stride unique est présenté alors que celui-ci peut être différent selon la hauteur et la largeur mais ce ne sera jamais le cas dans la suite de ce manuscrit.

4.3 Applications en météorologie

Les parties précédentes de ce chapitre ont détaillé les principes de fonctionnement et plusieurs couches existantes afin de construire un réseau de neurones. Avant de détailler les réseaux de neurones utilisés durant ces travaux dans les chapitres suivants, cette partie va rapidement présenter les principales applications ou recherches en cours en DL dans le domaine de la météorologie. Cette partie sera scindée en deux en commençant par présenter les apports de ces méthodes que ce soit sur le système d'observation, l'assimilation, le modèle ou le post-traitement de celui-ci. La deuxième partie s'intéressera aux limites et défis à relever avec l'arrivée de ces méthodes de DL.

4.3.1 Recherches en cours sur la chaîne observation, assimilation, prévision

Les apports des méthodes de DL sont listés sous forme de différents points selon leur domaine de l'observation jusqu'au post-traitement des modèles :

Concernant les observations,

Les méthodes de DL sont explorées afin de corriger les données satellitaires en particulier pour les débiaiser (He et al., 2016; Lee et al., 2019; Zhou and Grassotti, 2020). Des méthodes de DL sont aussi développées sur ce type de données pour détecter plusieurs types d'objet ou évènement comme des inondations ou de la glace (Wang et al., 2017) ou encore pour estimer l'intensité de cyclones tropicaux grâce aux micro-ondes (Wimmers et al., 2019). Des réseaux de neurones sont également développés pour évaluer la qualité des observations et détecter les observations aberrantes dans les stations au sol sur des données pluviométriques (Sha et al., 2021). Une autre application possible qui concerne les radars météorologiques est la suppression du bruit dans les échos lié aux bâtiments, oiseaux, insectes, ... (Lepetit et al., 2022).

Concernant l'assimilation,

De nombreux travaux s'intéressent à des approches hybrides mixant ces nouvelles approches de DL (ou ML) avec les modèles existants. Cette thématique se développe particulièrement en assimilation de données. Il existe en effet une grande proximité entre le machine/deep learning et l'assimilation de données qui cherchent à résoudre dans les deux cas des problèmes inverses et qui peuvent être unifiés dans un cadre bayésien (Bocquet et al., 2020; Geer, 2020). Des assimilations de données hybrides mixant modèle et deep learning sont donc développées pour tirer le meilleur parti de chacun (Farchi et al., 2021). Certaines méthodes commencent à être en opérationnel au CEPMMT depuis quelques années (Bonavita and Laloyaux, 2020). Le remplacement de l'opérateur d'observation est par exemple étudié lorsque celui-ci est difficile ou coûteux à estimer, on peut alors entraîner une méthode DL directement à l'aide des observations et de l'analyse (Rodríguez-Fernández et al., 2019). D'autres travaux tentent d'émuler des analyses avec seulement des réseaux de neurones pour un GCM simplifié (Cintra and de Campos Velho, 2018). L'intérêt réside ici dans la recherche de produire une analyse en un temps beaucoup plus restreint qu'avec une approche classique (Boukabara et al., 2021). Une dernière piste explorée s'intéresse à des méthodes

d'assimilation dans un espace latent [6] (Peyron et al., 2021)

Concernant la modélisation de l'atmopshère,

Beaucoup de travaux se concentrent sur les paramétrisations physiques. Un des objectifs possibles est de reproduire à l'aide de réseaux de neurones des schémas pré-existants comme le rayonnement, la couche limite ou les ondes de gravité orographiques mais en réduisant leurs coûts de calcul (Wang et al., 2017; Chantry et al., 2021). L'objectif peut aussi être une amélioration de ces paramétrisations lorsque celles-ci sont très complexes et que les lois physiques sous-jacentes sont encore mal comprises (Brenowitz and Bretherton, 2018). Enfin, certains chercheurs tentent de recréer certaines paramétrisations à l'aide de méthodes de ML et DL en repartant directement de certaines observations et donc en ne s'appuyant plus sur le schéma de paramétrisation pré-existant (Boukabara et al., 2021).

Plusieurs travaux tentent également de remplacer complètement le modèle dans une démarche « tout-IA » en repartant des analyses pour établir des prévisions à grande échelle et à plusieurs jours (Dueben and Bauer, 2018; Keisler, 2022; Pathak et al., 2022). Ces travaux s'appuient sur l'émergence de nouvelles architectures de réseaux de neurones comme les GraphNets (Pfaff et al., 2020) ou les Transformers (Guibas et al., 2021). Cependant ce type d'approche ne rivalise pas pour le moment avec les modèles opérationnels existants contrairement à certaines IA pour la Prévision Immédiate (PI). La plupart de ces méthodes de DL en PI s'appuient directement sur les données d'observation pour leurs entraînements. Plusieurs travaux s'intéressent à la PI avec pour objectif d'anticiper des phénomènes extrêmes comme les tornades (Lagerquist et al., 2020) ou la grêle (Gagne II et al., 2019). Sans se concentrer sur des phénomènes particuliers, d'autres réseaux de neurones sont développés pour prévoir l'évolution de la nébulosité en PI sur la France métropolitaine (Berthomier et al., 2020) ou encore des réflectivités radars au Royaume-Uni (Ravuri et al., 2021).

Concernant le post-traitement des modèles,

Des post-traitements pour corriger les modèles en aval sont explorés sur les sorties des prévisions d'ensemble pour améliorer des scores globaux comme le Continuous Rank Probability Score (Rasp and Lerch, 2018; Grönquist et al., 2021; Finn, 2021). Ces travaux sont également menés sur les modèles déterministes pour post-traiter par exemple la couverture nuageuse (Dupuy et al., 2020). Les méthodes DL sont explorées pour faire de la descente d'échelle et ainsi obtenir des sorties haute résolution à des coûts plus faibles par rapport à un modèle complet de PNT plus haute résolution (Sha et al., 2020a,b). La reconnaissance d'objets est aussi largement explorée afin d'extraire de l'information de gros jeux de données. Les enjeux sont multiples avec la volonté d'extraire l'information pertinente rapidement afin de faciliter certaines prises de décision dans le contexte de la prévision opérationnelle. Ce type d'approche peut également être utilisé pour observer l'impact d'un changement de version de modèle sur ces objets ou leurs évolutions dans les projections climatiques. Des travaux s'intéressent par exemple aux cyclones tropicaux (Kurth et al., 2018), aux fronts (Niebler et al., 2021) ou à la bande transporteuse (Quinting and Grams, 2022; Quinting et al., 2022).

6. espace abstrait d'un réseau de neurones entre l'encodeur et le decodeur. Exemple d'espace latent dans le chapitre 6, figure 6.3.

4.3.2 Limites et challenges liés aux méthodes de deep learning

Ces évolutions en cours ou attendues dans la chaîne observation, assimilation, prévision ne sont pas sans poser quelques défis et challenges nouveaux à relever :

- Un challenge important avec les algorithmes de DL (et ML) à résoudre reste la prise en compte des connaissances que nous avons a priori sur le système durant l'apprentissage (Dueben and Bauer, 2018). Une IA peut en effet ne pas respecter certaines lois physiques ce qui peut induire une dérive des systèmes de PNT si on souhaite par exemple remplacer un schéma de paramétrisation (Balogh et al., 2022). La prise en compte explicite de contraintes physiques dans les réseaux de neurones est donc étudiée (Daw et al., 2017; Pannekoucke and Fablet, 2020; Beucler et al., 2021). Yang et al. (2020) suggère également que la prise en compte de ces contraintes permet de réduire la taille du jeu de données nécessaire lors de l'entrainement.

- Une autre limite importante est l'opacité des algorithmes de DL, on parle souvent de « boîte noire ». Beaucoup de travaux cherchent donc à rendre ces méthodes IA plus explicables (McGovern et al., 2019). Ce point semble essentiel pour faciliter l'utilisation en condition opérationnelle de ces outils. Gagne II et al. (2019) montre par exemple que le réseau de neurones qui prévoit un risque de grêle relie ce risque à des structures de la convection qui correspondent aux schémas conceptuels connus. Relier le comportement du réseau de neurones à des schémas conceptuels connus en météorologie est un bon moyen de parvenir à rendre ces méthodes explicables. Les challenges liés à l'ajout de contraintes physiques et rendre ces approches plus explicables semblent donc étroitement liés (Boukabara et al., 2021).

- Les méthodes de DL nécessitent de grosses bases de données mais surtout des exemples de données de sortie (Y dans les parties précédentes). Or ces données ne sont pas toujours simples à obtenir (Boukabara et al., 2021). On peut penser en particulier aux problèmes en détection d'objet qui peuvent nécessiter une labellisation manuelle en amont. Cette labellisation sur des milliers de cas peut vite s'avérer chronophage. Des jeux de données commencent à émerger et être mis à disposition comme c'est le cas par exemple pour ClimateNet où des cyclones tropicaux et des rivières atmosphériques ont été étiquetés à la main (Prabhat et al., 2021).

Bibliographie

B. Balogh, D. Saint-Martin, and A. Ribes. How to Calibrate a Dynamical System With Neural Network Based Physics ? *Geophys. Res. Lett.*, 49(8), 2022. doi : https://doi.org/10.1029/2022GL097872.

L. Berthomier, B. Pradel, and L. Perez. Cloud Cover Nowcasting with Deep Learning. 2020. doi : 10.1109/ipta50016.2020.9286606.

T. Beucler, M. Pritchard, S. Rasp, J. Ott, P. Baldi, and P. Gentine. Enforcing Analytic Constraints in Neural Networks Emulating Physical Systems. *Phys. Rev. Lett.*, 126, 2021. doi : 10.1103/PhysRevLett.126.098302.

M. Bocquet, J. Brajard, A. Carrassi, and L. Bertino. Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization. *Foundations of Data Science*, 2(1) :55–80, 2020. doi : 10.3934/fods.2020004.

M. Bonavita and P. Laloyaux. Machine Learning for Model Error Inference and Correction. *Journal of Advances in Modeling Earth Systems*, 12(12), 2020. doi : https://doi.org/10.1029/2020MS002232.

S. A. Boukabara, V. Krasnopolsky, S. G. Penny, J. Q. Stewart, A. McGovern, D. Hall, J. E. Ten Hoeve, J. Hickey, H. L. Allen Huang, J. K. Williams, K. Ide, P. Tissot, S. E. Haupt, K. S. Casey, N. Oza, A. J. Geer, E. S. Maddy, and R. N. Hoffman. Outlook for Exploiting Artificial Intelligence in the Earth and Environmental Sciences. *Bull. Amer. Meteor. Soc.*, 102(5) :1016–1032, 2021. doi : 10.1175/BAMS-D-20-0031.1.

N. D. Brenowitz and C. S. Bretherton. Prognostic Validation of a Neural Network Unified Physics Parameterization. *Geophys. Res. Lett.*, 45(12) :6289–6298, 2018. doi : https://doi.org/10.1029/2018GL078510.

M. Chantry, S. Hatfield, P. Dueben, I. Polichtchouk, and T. Palmer. Machine Learning Emulation of Gravity Wave Drag in Numerical Weather Forecasting. *Journal of Advances in Modeling Earth Systems*, 13(7), 2021. doi : https://doi.org/10.1029/2021MS002477.

F. Chollet. *Deep learning with Python*. Simon and Schuster, 2021.

R. S. Cintra and H. F. de Campos Velho. Data assimilation by artificial neural networks for an atmospheric general circulation model. *Advanced applications for artificial neural networks*, 265, 2018. doi : 10.5772/intechopen.70791.

A. Daw, A. Karpatne, W. Watkins, J. Read, and V. Kumar. Physics-guided Neural Networks (PGNN) : An Application in Lake Temperature Modeling, 2017.

P. D. Dueben and P. Bauer. Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, 11(10) :3999–4009, 2018. doi : 10.5194/gmd-11-3999-2018.

F. Dupuy, O. Mestre, M. Serrurier, V. Burdá, M. Zamo, N. Cabrera-Gutiérrez, M. C. Bakkay, M. Mader, G. Oller, and J.-C. Jouhaud. ARPEGE Cloud Cover Forecast Postprocessing with Convolutional Neural Network. 36, 2020. doi : 10.1175/WAF-D-20-0093.1.

A. Farchi, P. Laloyaux, M. Bonavita, and M. Bocquet. Using machine learning to correct model error in data assimilation and forecast applications. *Quart. J. Roy. Meteor. Soc.*, 147(739) :3067–3084, 2021. doi : 10.1002/qj.4116.

T. S. Finn. Self-Attentive Ensemble Transformer : Representing Ensemble Interactions in Neural Networks for Earth System Models, 2021.

D. Gagne II, S. Haupt, D. Nychka, and G. Thompson. Interpretable deep learning for spatial analysis of severe hailstorms. *Mon. Wea. Rev.*, 147 :2827–2845, 2019. doi : 10.1175/MWR-D-18-0316.1.

A. J. Geer. Learning earth system models from observations : machine learning or data assimilation ? (863), 2020. doi : 10.21957/7fyj2811r.

P. Grönquist, C. Yao, T. Ben-Nun, N. Dryden, P. Dueben, S. Li, and T. Hoefler. Deep learning for post-processing ensemble weather forecasts. *Philosophical Transactions of the Royal Society A : Mathematical, Physical and Engineering Sciences*, 379(2194), 2021. doi : 10.1098/rsta.2020.0092.

J. Guibas, M. Mardani, Z. Li, A. Tao, A. Anandkumar, and B. Catanzaro. Adaptive Fourier Neural Operators : Efficient Token Mixers for Transformers, 2021.

T. Hastie, R. Tibshirani, and J. . Friedman. *The elements of statistical learning : data mining, inference, and prediction*, volume 2. Springer, 2009.

Q. He, Z. Wang, and J. He. Bias Correction for Retrieval of Atmospheric Parameters from the Microwave Humidity and Temperature Sounder Onboard the Fengyun-3C Satellite. *Atmos.*, 7(12), 2016. ISSN 2073-4433. doi : 10.3390/atmos7120156.

S. Ioffe and C. Szegedy. Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift, 2015.

R. Keisler. Forecasting Global Weather with Graph Neural Networks, 2022.

D. Kingma and J. Ba. Adam : A Method for Stochastic Optimization. *International Conference on Learning Representations*, 2014.

T. Kurth, S. Treichler, J. Romero, M. Mudigonda, N. Luehr, E. Phillips, A. Mahesh, M. Matheson, J. Deslippe, and M. Fatica. Exascale deep learning for climate analytics. pages 649–660, 2018.

R. Lagerquist, A. McGovern, C. Homeyer, D. Gagne II, and T. Smith. Deep learning on three-dimensional multiscale data for next-hour tornado prediction. *Mon. Wea. Rev.*, 148 :2837–2861, 2020. doi : 10.1175/MWR-D-19-0372.1.

Y. LeCun and Y. Bengio. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361 :1995, 1995.

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521 :436–44, 2015. doi : 10.1038/nature14539.

H. Lee, G. Roger, R. Ranganath, and N. Andrew. Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations. In *Proceedings of the 26th Annual International Conference on Machine Learning*, page 609–616, 2009. doi : 10.1145/1553374.1553453.

Y. Lee, D. Han, M. Ahn, J. Im, and S. J. Lee. Retrieval of Total Precipitable Water from Himawari-8 AHI Data : A Comparison of Random Forest, Extreme Gradient Boosting, and Deep Neural Network. *Remote Sens.*, 11(15), 2019. ISSN 2072-4292. doi : 10.3390/rs11151741.

P. Lepetit, C. Ly, L. Barthès, C. Mallet, N. Viltard, Y. Lemaitre, and L. Rottner. Using Deep Learning for Restoration of Precipitation Echoes in Radar Data. *IEEE Transactions on Geoscience and Remote Sensing*, 60 :1–14, 2022. doi : 10.1109/TGRS.2021.3052582.

A. McGovern, R. Lagerquist, D. J. Gagne II, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith. Making the Black Box More Transparent : Understanding the Physical Implications of Machine Learning. *Bull. Amer. Meteor. Soc.*, 100(11) :2175–2199, 2019. doi : 10.1175/BAMS-D-18-0195.1.

T. M. Mitchell. *Machine Learning*. McGraw-Hill, 1997. ISBN 978-0-07-042807-2.

S. Niebler, A. Miltenberger, B. Schmidt, and P. Spichtinger. Automated detection and classification of synoptic scale fronts from atmospheric data grids. *Weather Clim. Dynam.*, 2021. doi : 10.5194/wcd-2021-27.

O. Pannekoucke and R. Fablet. PDE-NetGen 1.0 : from symbolic partial differential equation (PDE) representations of physical processes to trainable neural network representations. *Geoscientific Model Development*, 13(7) :3373–3382, 2020. doi : 10.5194/gmd-13-3373-2020.

J. Pathak, S. Subramanian, P. Harrington, S. Raja, A. Chattopadhyay, M. Mardani, T. Kurth, D. Hall, Z. Li, K. Azizzadenesheli, P. Hassanzadeh, K. Kashinath, and A. Anandkumar. FourCastNet : A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators, 2022.

M. Peyron, A. Fillion, S. Gürol, V. Marchais, S. Gratton, P. Boudier, and G. Goret. Latent space data assimilation by using deep learning. *Quart. J. Roy. Meteor. Soc.*, 147(740) :3759–3777, 2021. doi : https://doi.org/10.1002/qj.4153.

T. Pfaff, M. Fortunato, A. Sanchez-Gonzalez, and P. W. Battaglia. Learning Mesh-Based Simulation with Graph Networks. 2020. doi : 10.48550/ARXIV.2010.03409.

Prabhat, K. Kashinath, M. Mudigonda, S. Kim, L. Kapp-Schwoerer, A. Graubner, E. Karaismailoglu, L. von Kleist, T. Kurth, A. Greiner, A. Mahesh, K. Yang, C. Lewis, J. Chen, A. Lou, S. Chandran, B. Toms, W. Chapman, K. Dagon, C. A. Shields, T. O'Brien, M. Wehner, and W. Collins. Climatenet : an expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather. *Geoscientific Model Development*, 14(1) :107–124, 2021. doi : 10.5194/gmd-14-107-2021.

J. F. Quinting and C. M. Grams. EuLerian Identification of ascending AirStreams (ELIAS 2.0) in numerical weather prediction and climate models – Part 1 : Development of deep learning model. *Geosci. Model Dev.*, 15 :715–730, 2022. doi : 10.5194/gmd-15-715-2022.

J. F. Quinting, C. M. Grams, A. Oertel, and M. Pickl. EuLerian Identification of ascending AirStreams (ELIAS 2.0) in numerical weather prediction and climate models – Part 2 : Model application to different datasets. *Geosci. Model Dev.*, 15 :731–744, 2022. doi : 10.5194/gmd-15-731-2022.

S. Rasp and S. Lerch. Neural networks for postprocessing ensemble weather forecasts. *Mon. Wea. Rev.*, 146(11) :3885–3900, 2018. doi : 10.1175/mwr-d-18-0187.1.

S. Ravuri, K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge, R. Prudden, A. Mandhane, A. Clark, A. Brock, K. Simonyan, R. Hadsell, N. Robinson, E. Clancy, A. Arribas, and S. Mohamed. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878) :672–677, 2021. doi : 10.1038/s41586-021-03854-z.

N. Rodríguez-Fernández, P. de Rosnay, C. Albergel, P. Richaume, F. Aires, C. Prigent, and Y. Kerr. SMOS Neural Network Soil Moisture Data Assimilation in a Land Surface Model and Atmospheric Impact. *Remote Sensing*, 11 (11), 2019. ISSN 2072-4292. doi : 10.3390/rs11111334.

Y. Sha, D. J. Gagne II, G. West, and R. Stull. Deep-Learning-Based Gridded Downscaling of Surface Meteorological Variables in Complex Terrain. Part I : Daily Maximum and Minimum 2-m Temperature. *J. Appl. Meteor. Climatol.*, 59(12) :2057–2073, 2020a. doi : 10.1175/JAMC-D-20-0057.1.

Y. Sha, D. J. Gagne II, G. West, and R. Stull. Deep-Learning-Based Gridded Downscaling of Surface Meteorological Variables in Complex Terrain. Part II : Daily Precipitation. *J. Appl. Meteor. Climatol.*, 59(12) :2075–2092, 2020b. doi : 10.1175/JAMC-D-20-0058.1.

Y. Sha, D. J. Gagne II, G. West, and R. Stull. Deep-Learning-Based Precipitation Observation Quality Control. *Journal of Atmospheric and Oceanic Technology*, 38(5) :1075–1091, 2021. doi : 10.1175/JTECH-D-20-0081.1.

N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout : A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(56) :1929–1958, 2014. URL `http://jmlr.org/papers/v15/srivastava14a.html`.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention Is All You Need, 2017.

L. Wang, K. Scott, and D. Clausi. Sea Ice Concentration Estimation during Freeze-up from SAR Imagery Using a Convolutional Neural Network. *Remote Sensing*, 9, 2017. doi : 10.3390/rs9050408.

A. Wimmers, C. Velden, and J. H. Cossuth. Using deep learning to estimate tropical cyclone intensity from satellite passive microwave imagery. *Mon. Wea. Rev.*, 147(6) :2261–2282, 2019. doi : 10.1175/MWR-D-18-0391.1.

L. Yang, D. Zhang, and G. E. Karniadakis. Physics-Informed Generative Adversarial Networks for Stochastic Differential Equations. *SIAM Journal on Scientific Computing*, 42(1) :292–317, 2020. doi : 10.1137/18M1225409.

Y. Zhou and C. Grassotti. Development of a Machine Learning-Based Radiometric Bias Correction for NOAA's Microwave Integrated Retrieval System (MiRS). *Remote Sens.*, 12(19), 2020. ISSN 2072-4292. doi : 10.3390/rs12193160.

Chapitre 5

Détection des échos arqués avec un U-Net

Le chapitre précédent a montré en quoi les réseaux de neurones convolutifs sont des méthodes statistiques intéressantes pour le traitement d'images et comment des réseaux de neurones fonctionnent. Le chapitre 3 a montré l'intérêt de détecter les échos arqués dans les sorties des modèles de prévision et en quoi leurs détections nécessitent de reconnaître une forme en arc dans les champs de réflectivités. Ce chapitre s'intéresse aux moyens mis en place pour détecter correctement les échos arqués avec un réseau de neurones convolutif dont l'architecture se nomme « U-Net ». La première partie de ce chapitre s'intéressera à la construction des jeux de données et en particulier la labellisation manuelle des échos arqués sur des milliers d'images. La deuxième partie présentera l'architecture du U-Net, la troisième partie l'entraînement du réseau de neurones qui détaillera les étapes de pré-traitement nécessaires sur la base d'apprentissage. La quatrième partie expliquera concrètement comment à partir d'un champ de réflectivité PE-AROME, un champ binaire est créé en sortie pour dessiner les contours des échos arqués. Enfin la dernière partie présentera les scores objets utilisés pour évaluer les performances du U-Net dans le chapitre 7. Ce chapitre présente des travaux qui sont également repris dans l'article publié au cours de cette thèse et qui fera l'objet du chapitre 7.

5.1 Construction des jeux de données

5.1.1 Labellisation des échos arqués

Les données d'entraînement et de validation pour le réseau de neurones sont construites en utilisant les prévisions de 'pseudo-réflectivité' des membres de PE-AROME. Le champ de réflectivité disponible dans les sorties des membres PE-AROME est calculé avec un simulateur radar (Caumont et al., 2006) en $mm^6 \cdot m^{-3}$. La relation Z-R de Marshall and Palmer (1948) est appliquée pour convertir les champs de réflectivité en champs de pseudo-réflectivité en millimètres par heure ($mm \cdot h^{-1}$). Le choix des unités $mm \cdot h^{-1}$ plutôt que dBZ (comme pour la figure 1.2) pour les sorties des modèles est un choix historique datant de l'époque où le

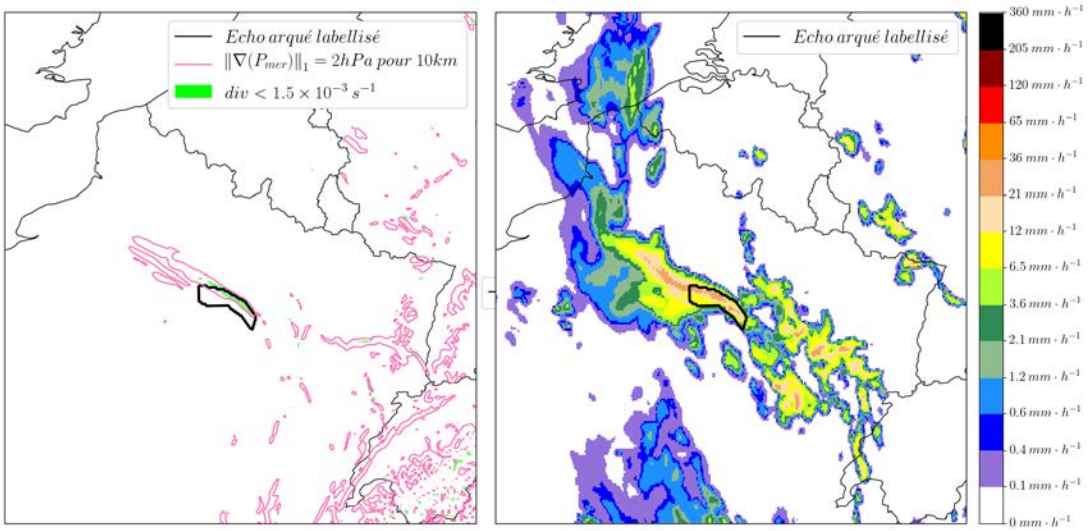


FIGURE 5.1 – *Exemple de champs utilisés lors du processus de labellisation. Les champs sont zoomés sur le Nord-Est de la France. L'image de gauche correspond aux contours de* $\|\nabla(MSLP)\|_1$ *égal à 2 hPa pour 10 km sont en rose, les zones où la divergence du vent horizontal (div dans la légende) est inférieure à* $1.5 \times 10^{-3}\ s^{-1}$ *sont en vert, et le contour de l'écho arqué labellisé est en noir. (droite) Le champ de pseudo-réflectivité correspondant.*

cumul de précipitation était mis à jour toutes les heures. Même si les champs en $mm \cdot h^{-1}$ et dBZ sont maintenant produits tous les deux en opérationnel, seulement les champs de pseudo-réflectivité sont disponibles pour les dates les plus anciennes des bases de données d'entraînement et de validation. Ce champ de pseudo-réflectivité se calcule sur plusieurs niveaux mais c'est la valeur maximale dans la colonne atmosphérique qui est utilisée comme prédicteur pour les réseaux de neurones. Pour entraîner les réseaux de neurones et calculer les scores de la base de données de validation, un ensemble de données étiquetées est produit indiquant la présence ou non d'un écho arqué. Cet ensemble de données est construit à partir des contours des échos arqués tracés manuellement par un expert à l'aide du logiciel Visual Geometry Group (VGG) Image Annotator (VIA) (Dutta and Zisserman, 2019). Après une étape de post-traitement, la grille issue des labellisations a la même taille que l'entrée des pseudo-réflectivités, avec une valeur de 1 pour chaque point de grille à l'intérieur d'un écho arqué et 0 en dehors (il s'agit donc d'une grille binaire). Le processus de labellisation utilise trois variables : la pseudo-réflectivité maximale, la pression au niveau de la mer (P_{mer}) et la vitesse du vent à 10 m. À partir de ces trois prédicteurs, l'expert utilise trois conditions pour étiqueter manuellement un écho arqué. La première est d'observer une forme d'arc dans le champ de pseudo-réflectivité. La seconde est de remarquer un gradient dans le champ de P_{mer} correspondant à la forme de l'arc. Pour rappel d'après Markowski and Richardson (2010), un motif de P_{mer} spécifique est observé autour d'un écho arqué (cf section 3.3.1). La dernière condition est d'observer la forme d'arc dans le champ de vitesse du vent et surtout la forte augmentation de la vitesse du vent avec l'arrivée de l'écho arqué.

En pratique, le processus de labellisation prend beaucoup de temps. Pour faciliter l'analyse des trois conditions à remplir, seulement deux figures sont fournies

à l'expert (figure 5.1). La première est le champ de pseudo-réflectivité maximale (droite figure 5.1) et l'autre les contours de $\|\nabla(MSLP)\|_1$ égal à 2 hPa pour 10 km superposés aux zones où la divergence du vent horizontal est inférieure à $1.5 \times 10^{-3}\ s^{-1}$ (gauche figure 5.1). Les contours des échos arqués labellisés à la main sont tracés avec ces deux champs (contours noirs figure 5.1).

5.1.2 Bases d'apprentissage et de validation

Le processus de labellisation est appliqué à des cas spécifiques où des échos arqués sont observés (dans les données radar) en France. L'étude de chaque réseau de PE-AROME sur une longue période prendrait trop de temps, on suppose donc que les échos arqués simulés peuvent être identifiés préférentiellement parmi les 16 membres sur ces cas spécifiques. Une liste de ces échos arqués observés en 2017, 2018 et 2019 est créée avec l'aide de 10 prévisionnistes pour être sûr qu'aucun écho arqué observé n'est omis. Les quatre réseaux PE-AROME avant l'apparition des échos arqués observés sont utilisés et seulement les échéances autour des échos arqués observés sont retenues pour gagner du temps lors du processus de labellisation. D'autres cas de convection sévère (supercellules, lignes de grains et systèmes convectifs quasi-stationnaires) sont également intégrés dans la base de validation. La capacité des réseaux de neurones à différencier les formes de pseudo-réflectivités appartenant à d'autres orages et pas à des échos arqués doit en effet être vérifiée. Environ 10 échos arqués ont été tracés dans ces autres cas de convection sévère. En effet, certains membres peuvent simuler des échos arqués dans le domaine AROME même si un système convectif quasi-stationnaire est finalement observé (par exemple). Concernant la base d'apprentissage, d'autres formes de convection sévère (en particulier les supercellules) sont indirectement intégrées car elles sont simulées sur les mêmes cas que les échos arqués dans des environnements similaires par certains membres PE-AROME. Il n'a pas été jugé nécessaire d'ajouter spécifiquement d'autres cas pour la base d'apprentissage.

La base d'apprentissage comprend un total de 11 cas d'échos arqués observés (18 mai 2017, 27 juin 2017, 10 juin 2018, 8 août 2018, 6 juillet 2019, 6-11 août 2019, 21 septembre 2019 et 14-23 octobre 2019), pour un total de 6206 champs de pseudo-réflectivité et 556 échos arqués labellisés manuellement. Lors de la conception de la base d'apprentissage, la moitié des échantillons PE-AROME correspondant à un cas d'écho arqué spécifique (21 septembre 2021) est sélectionnée au hasard. Ce sous-échantillonnage est nécessaire car tous les échos arqués simulés dans les membres PE-AROME sont très similaires avec les mêmes intensités de pseudo-réflectivité et les mêmes tailles pour ce cas (tous les contours des échos arqués sont disponibles dans le supplément l'article de Mounier et al. (2022), section 7.3, Fig. S2). Si ce sous-échantillonnage n'était pas effectué, ce cas spécifique pourrait créer un grand nombre d'échantillons très similaires dans la base d'apprentissage et par conséquent dégrader le processus d'entraînement. Concernant la base de validation, 4 cas d'échos arqués observés différents (8 août 2017, 26 mai 2018, et 4-5 juin 2019) ainsi que trois autres cas convectifs (lignes de grains : 2 janvier 2018 et 3 décembre 2018, système convectif quasi-stationnaire : 5 août 2019) sont intégrés, ce qui conduit à 2440 champs de pseudo-réflectivité et 264 échos arqués labellisés à la main. Cette répartition correspond à environ 70 % des données utilisées pour l'entraînement et 30 % pour le jeu de validation. Les bases d'apprentissage et de

validation sont composées de cas indépendants afin de limiter au maximum la corrélation entre les deux bases de données.

5.1.3 Choix du prédicteur

La sélection du ou des prédicteurs est cruciale et affecte la capacité du réseau de neurones à détecter correctement les échos arqués. En raison de la puissance de calcul limitée, un seul prédicteur peut être choisi comme première étape vers la construction d'un système d'identification des échos arqués. Les trois variables utilisées pendant le processus de labellisation (pseudo-réflectivité maximale, P_{mer}, et force du vent à 10 m) sont examinées. Le champ de P_{mer} est trop bruité autour des zones montagneuses (notamment les Alpes et les Pyrénées) avec un bruit d'une magnitude similaire aux perturbations induites par les échos arqués. Le coin inférieur droit de la figure 5.1 est un bon exemple pour les Alpes avec de nombreux contours roses mais sans précipitation. Le champ de force du vent a parfois un comportement irréaliste dans les modèles AROME et est rejeté à ce stade. Ce comportement est observé pour des situations convectives et sèches ou dans des averses à l'arrière de fronts froids. Les vitesses moyennes de vent peuvent être supérieures à 120 $km \cdot h^{-1}$ (rafales supérieures à 210 $km \cdot h^{-1}$). Ces valeurs ne peuvent être atteintes que très exceptionnellement en France mais ce signal est beaucoup trop fréquent dans les modèles AROME pour être réaliste. Par ailleurs, les modèles AROME peuvent également créer de forts courants de densité dans des cas de convection avec pas ou peu de précipitations associées. Même si ce comportement est parfois réaliste (en cas de microrafales sèches), les prévisionnistes ont tendance à utiliser la vitesse du vent avec précaution dans les cas de convection (la même conclusion s'applique au paramètre des rafales de vent). Pour cette raison, les réseaux de neurones seront des modèles univariés avec seulement les champs de pseudo-réflectivité en entrée. Cependant, la vitesse du vent et surtout la vitesse des rafales de vent à 10m (FFgust) restent des paramètres cruciaux pour les échos arqués. Les fortes rafales de vent sous les échos arqués constituent le principal risque et les prévisionnistes doivent donc connaître l'ampleur de ces rafales pour transmettre des informations. Dans le chapitre 7, les rafales de vent à l'intérieur des échos arqués seront prises en compte lors de l'évaluation orientée objet.

5.2 Architecture du réseau de neurones

L'architecture du réseau de neurones sélectionné doit être adaptée aux petites bases de données car les échos arqués sont des événements rares (dans l'espace et dans le temps). Une architecture de type U-Net (Ronneberger et al., 2015) a été choisie car des études antérieures ont montré qu'elle donne des résultats satisfaisants lorsque peu de données sont disponibles. Cette architecture U-Net (figure 5.2) est divisée en deux parties avec une partie encodage (gauche du « U ») et une partie décodage (droite du « U »). La partie encodage se compose de 2 couches de convolution (32 filtres 3×3) suivies d'une fonction d'activation ReLU et d'un dropout de 0,2. Une opération de maxpooling 2×2 est appliquée pour diviser la taille de l'imagette par 2. La même séquence est répétée deux fois avec respectivement 64 et 128 filtres. La partie décodage effectue d'abord une opé-

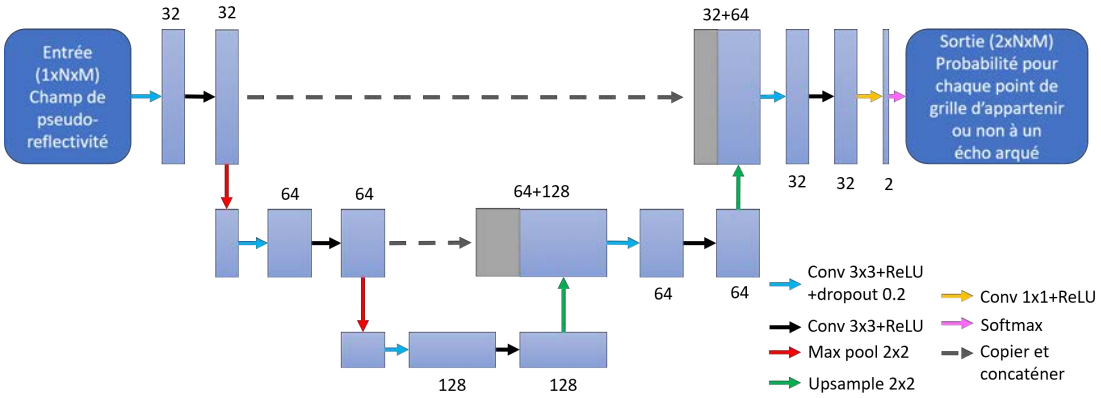


FIGURE 5.2 – *Architecture du U-Net*

ration de sur-échantillonnage 2 × 2. Les imagettes après sur-échantillonage sont ensuite concaténées avec les imagettes correspondantes dans la partie encodage et 2 couches de convolution suivies d'une activation ReLU sont appliquées (avec le même nombre de filtres que dans la couche correspondante de la partie encodage). Le sur-échantillonnage, la concaténation et les convolutions sont répétés une autre fois. À ce stade, les champs en sortie ont la même taille que ceux en entrée. Une dernière couche de convolution avec un filtre 1 × 1 est ajoutée pour obtenir le nombre de classes souhaité (deux classes ici). Pour obtenir une probabilité d'appartenir ou non à un écho arqué en sortie, une fonction softmax est appliquée. L'architecture U-Net est présentée ici avec une taille d'entrée de N × M points de grille. Plusieurs tailles d'entrées seront testées dans le chapitre 7. La taille des imagettes (**input_size**[1]) en entrée fait donc partie des hyperparamètres testés ultérieurement, ils seront notés **en gras** dans ce chapitre. A l'inverse, d'autres hyperparamètres de l'architecture U-Net, tels que le nombre et la taille des filtres, le dropout et le choix de l'activation ReLU ne sont pas testés dans la suite de ces travaux.

5.3 Entraînement du réseau de neurones

5.3.1 Pré-traitement de la base d'apprentissage

La puissance de calcul limitée ne permet pas de prendre l'intégralité du domaine AROME comme entrée du réseau de neurones. Pour entraîner les U-Nets, des imagettes de N × M points de grille sont extraites des champs de pseudo-réflectivité (717 × 1121 points de grille). En outre comme la taille d'un écho arqué est beaucoup plus petite que celle du domaine, il existe un fort déséquilibre entre les classes écho arqué et sans écho arqué dans les bases de données. L'extraction d'imagettes plus petites permet de concevoir des bases de données plus équilibrées en sélectionnant préférentiellement certaines imagettes. Cette extraction nécessite la définition de certains hyperparamètres qui seront discutés dans le chapitre 7. La figure 5.3 illustre les différentes étapes de l'extraction des imagettes (de 1 à 6) :

1. Les hyperparamètres ne sont pas traduits de l'anglais vers le français pour correspondre à ceux du chapitre 7

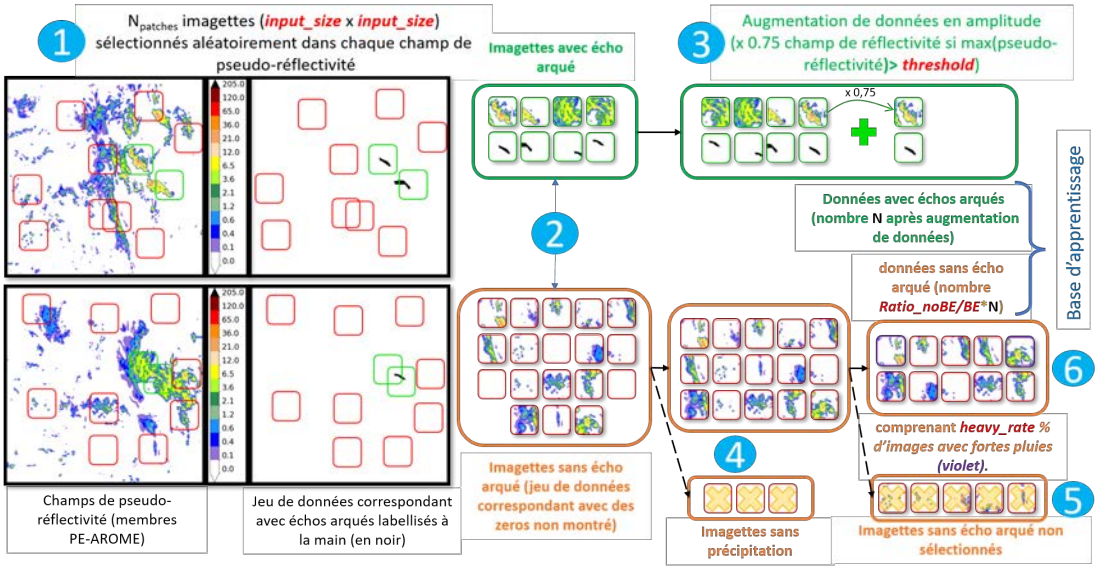


FIGURE 5.3 – *Construction de la base d'apprentissage. Les étapes 1 à 6 (cercles bleus) sont décrites dans la section 5.3.1. Les hyper-paramètres discutés dans le chapitre 7 apparaissent en rouge.*

- Étape 1 : les imagettes des champs de pseudo-réflectivité sont extraites de manière aléatoire. Un nombre total de N_{patch} sont extraits de chaque champ (valeurs dans le tableau 5.1). Les imagettes correspondantes du predictand sont extraites simultanément. Pour rappel le predictand est ici le champ binaire avec des 1 et des 0 selon la présence ou non du point de grille dans un écho arqué (issu des labellisations manuelles).

- Étape 2 : les paires d'imagettes (pseudo-réflectivité/predictand) sont divisées en deux groupes : les imagettes avec des échos arqués où au moins un point de grille est étiqueté comme dans un écho arqué et les imagettes sans écho arqué. Cette définition à partir d'un seul point de grille pourra être revue dans les travaux futurs. La sélection d'imagettes avec seulement quelques points de grille dans un écho arqué ne permet pas de le représenter correctement.

- Étape 3 : le nombre d'imagettes avec écho arqué est très limité par rapport aux imagettes sans écho arqué (1 imagette avec écho arqué pour 1600 imagettes sans). Durant les premiers tests, les échos arqués avec des pseudo-réflectivités modérées n'étaient pas correctement détectés car ils étaient sous-représentés dans la base de données. Une augmentation de données en amplitude a été ajoutée pour résoudre ce problème : dans une imagette donnée, les pseudo-réflectivités sont multipliées par un coefficient de 0.75 si un écho arqué se trouve sur cette imagette et si la pseudo-réflectivité maximale est supérieure à un seuil donné (**Data augmentation threshold**). Cette nouvelle imagette est ajoutée à celles qui comportent des échos arqués (étape 2). Le champ de pseudo-réflectivité après augmentation de données doit rester physiquement cohérent autant que possible. C'est pourquoi les coefficients inférieurs à 0.75 ne sont pas étudiés et seulement les imagettes avec de fortes pseudo-réflectivités sont concernées par cette augmentation de données.

TABLE 5.1 – *Valeurs de N_{stride} et N_{patch} selon la taille de l'imagette en entrée*

taille entrée	N_{stride}	N_{patch}
$24 \cdot 24$	7	4000
$48 \cdot 48$	15	1000
$96 \cdot 96$	30	250

- Étape 4 : pour réduire le nombre d'imagettes sans écho arqué, les imagettes sans précipitation (pseudo-réflectivité maximale < 0,1 mm h-1) sont supprimées de la base de données.

- Étape 5 : même après la suppression de nombreuses imagettes à l'étape 4, le nombre d'imagettes sans écho arqué reste élevé (1 imagette avec écho arqué pour 400 imagettes sans écho arqué). Pour réduire encore le nombre d'imagettes sans écho arqué, le rapport entre les imagettes avec et sans écho arqué (**Ratio_noBE/BE**) est fixé à une valeur inférieure. Les imagettes retenues sont sélectionnées aléatoirement et les imagettes sans écho arqué non retenues sont supprimées.

- Étape 6 : au cours des premiers tests, toutes les structures pluvieuses présentant de fortes pseudo-réflectivités ont été détectées comme des échos arqués et par conséquent le nombre de fausses alarmes était très élevé dans la base de validation. Les fortes pseudo-réflectivités sont rares dans l'espace et dans le temps et la majorité des imagettes sans écho arqué ne contiennent pas ou peu de précipitations, alors que les échos arqués sont fréquemment associés à de fortes précipitations. Seulement 0,7 % des imagettes sans écho arqué sont associées à de fortes précipitations (c'est-à-dire supérieures à 60 $mm \cdot h^{-1}$), alors qu'après l'augmentation de données en amplitude (étape 3), environ 55 % des imagettes avec écho arqué sont associées à de fortes précipitations. Dans ce cas, le réseau de neurones se fie trop à la valeur des pseudo-réflectivités pour détecter les échos arqués dans les champs de pseudo-réflectivités et pas assez à la forme. Pour résoudre ce problème, les imagettes présentant des pseudo-réflectivités correspondant à de fortes précipitations mais sans écho arqué sont forcées dans la base d'apprentissage. Pour cela, le taux d'imagettes avec fortes précipitations et sans écho arqué par rapport au nombre total d'imagettes sans écho arqué (**heavy pseudoreflectivity patch percentage**) est fixé. Une imagette est considérée comme ayant de fortes précipitations si le maximum est supérieur à 60 $mm \cdot h^{-1}$.

5.3.2 Mise en place de l'entraînement

Pour la mise en œuvre de l'entraînement, le logiciel TensorFlow/Keras est utilisé (Abadi et al., 2015). En utilisant uniquement les champs de pseudo-réflectivités comme entrée, le U-Net est un modèle univarié. Les bases d'apprentissage et de validation n'ont donc pas besoin d'être normalisées. Après quelques tests et vérifications des courbes de perte, le nombre d'epochs maximal est fixé à 50. Différentes

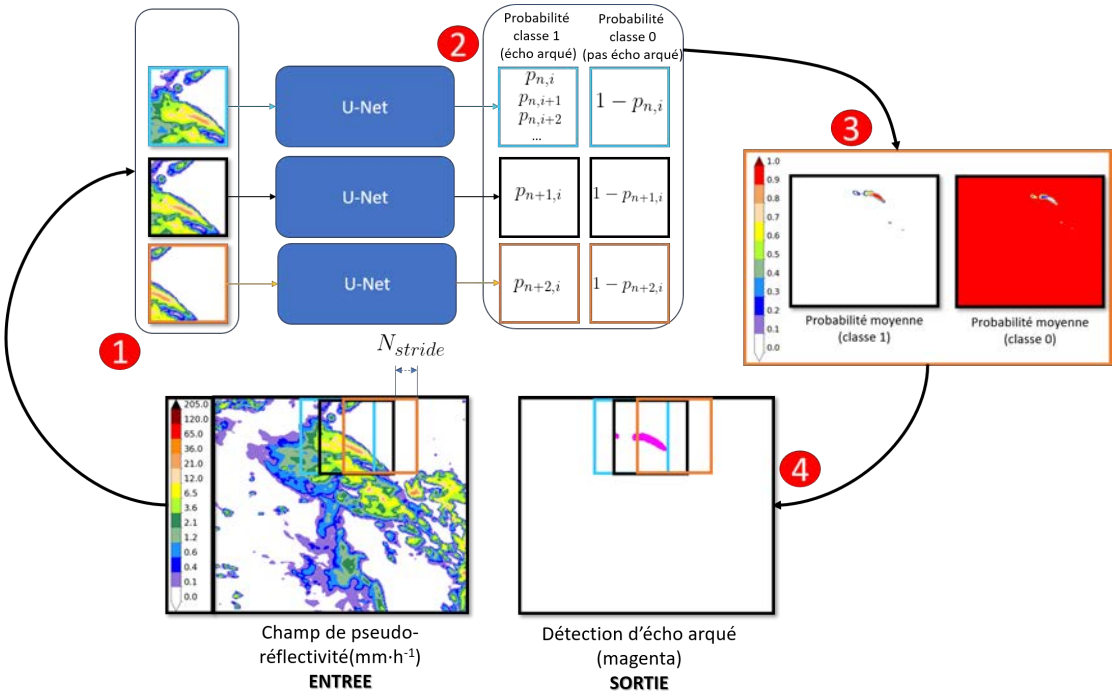


FIGURE 5.4 – *Description schématique du processus de prediction complet en partant du champ de pseudo-réflectivité complet (ENTRÉE) jusqu'à la prédiction d'un champ binaire avec un 1 pour les points de grille dans un écho arqué et 0 sinon (SORTIE). Les étapes 1 à 4 (cercles rouges) sont décrites dans la section 5.4.*

tailles de batchs ont été testées (16, 32 ou 64) : les résultats sont similaires avec 16 et 32, mais avec 64 on observe du sur-apprentissage. La taille des batchs est fixée à 32. Le learning rate par défaut (10^{-3}) est appliqué. La fonction coût utilisée est l'entropie croisée présentée dans le chapitre précédent, section 4.2.2. Cette fonction coût nécessite la définition des poids de chaque classe (w_0 et $\mathbf{w_1}$). Contrairement aux poids du U-Net, les poids de la fonction coût sont fixes. Cette fonction coût pondérée a été choisie car les échos arqués sont des événements rares et les réseaux de neurones ont alors tendance à surestimer la classe 0 (sans écho arqué) avec une fonction coût non pondérée. Pour les entraînements, le poids de la classe 0 (w_0) est fixé à 1. Différentes valeurs de $\mathbf{w_1}$ seront testées dans le chapitre 7.

5.4 Prédiction sur la grille complète

Les entrées et sorties du U-Net sont des imagettes de taille N × M points de grille. Comme cette taille ne correspond pas à la grille complète du domaine AROME, il est nécessaire de définir une méthode qui à partir de ce U-Net réalise une prédiction sur cette grille complète. Le même U-Net est appliqué à des imagettes qui se chevauchent, comme le montre la figure 5.4. La grille entière est divisée en imagettes ordonnées avec un pas de N_{stride} points de grille le long des longitudes et des latitudes (figure 5.4, étape 1 : carrés bleus, noirs et orange). Les valeurs de N_{stride} dépendent de la taille des imagettes en entrée du U-Net et sont données dans le tableau 5.1. Une prédiction pour chaque imagette est calculée. Les résultats de la prédiction sont des imagettes représentant la probabilité de chaque

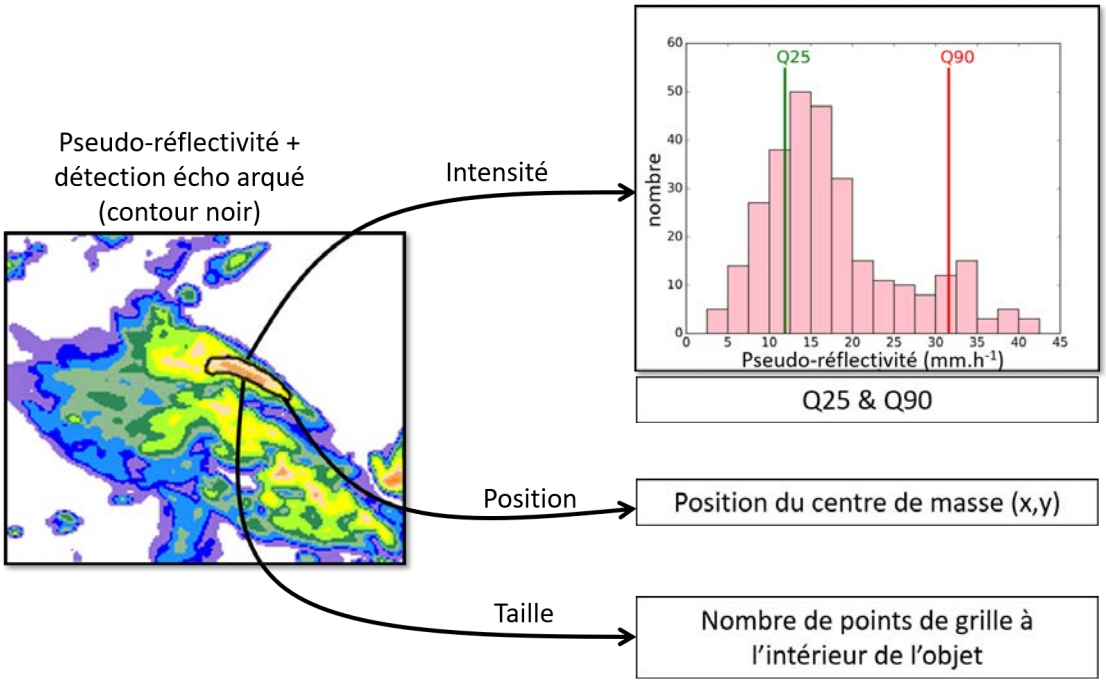


FIGURE 5.5 – *Attributs des objets. Un écho arqué est caractérisé sur cette figure par 3 attributs : intensité, position et taille. L'intensité de l'écho arqué est extraite avec les quantiles 25 et 90 de pseudo-reflectivités à l'intérieur de l'objet (histogramme), la position est extraite avec la position du centre de masse et la taille avec le nombre de points de grille qui constitue l'objet.*

classe (figure 5.4, étape 2 : deux imagettes après l'architecture U-Net pour chaque imagette de pseudo-réflectivités, $p_{n,i}$ est la probabilité du $i$ème point de grille pour la $n$ième imagette). Pour chaque point de grille, la probabilité moyenne pour chaque classe est calculée (figure 5.4, étape 3) en considérant toutes les imagettes où le point de grille est inclus. Enfin, les probabilités sont utilisées (on retient la classe la plus probable entre les deux) pour définir un champ binaire avec les échos arqués détectés (figure 5.4, étape 4). Le champ en sortie couvre ainsi l'ensemble du domaine AROME et associe à chaque point de grille une valeur de 1 ou 0 selon l'appartenance à un écho arqué.

5.5 Scores orientés objets

5.5.1 Attributs des objets échos arqués

La capacité des U-Nets à détecter les échos arqués est évaluée par une approche orientée objet. Une approche par points de grille n'est pas adaptée à l'évaluation des détections des échos arqués pour les raisons évoquées dans le chapitre 3, section 3.2 dédiée à l'approche objet. Les attributs des objets utilisés dans ces travaux sont les quantiles 25 (Q25) et 90 (Q90) du champ de pseudo-réflectivités au sein de l'objet écho arqué, la position du centre de masse de l'objet, la taille de l'objet et le maximum des rafales de vent au sein de l'objet. Les attributs sont présentés sur la figure 5.5, sauf le dernier. Les quantiles Q25 et Q90 permettent de différencier

TABLE 5.2 – *Tableau de contingence. Les bonnes détections (a) correspondent au nombre d'images avec au moins un écho arqué labellisé et un écho arqué détecté par le U-Net. Les fausses alarmes (b) sont le nombre d'images avec un ou des échos arqués détectés dans le U-Net mais aucun labellisé. Les non détections (c) correspondent au nombre d'images avec un ou des échos arqués labellisés mais aucun détecté. Les bonnes non détections (d) sont les images avec aucun écho arqué détecté et labellisé.*

Tableau de contingence		écho arqué labellisé	
		OUI	NON
écho arqué détecté	OUI	a	b
	NON	c	d

les échos arqués forts des échos arqués plus modérés. Une autre utilisation des Q90 et Q25 est de séparer la partie la plus active des échos arqués (centre) de la partie la moins active (bord). La taille de l'objet est utilisée pour distinguer les grands et les petits échos arqués. Le maximum des rafales de vent est un indicateur de l'intensité du vent dans l'écho arqué, qui est un point important pour évaluer l'intensité de ce dernier. La distribution des attributs des objets est calculée en utilisant tous les objets labellisés à la main et tous ceux détectés par les U-Nets. Une comparaison des attributs deux à deux est également effectuée pour les objets qui peuvent être appariés entre labellisation manuelle et détection par U-Net. Des objets sont appariés si la distance entre les centres de masse est inférieure à un seuil donné qui sera explicité dans le chapitre 7.

5.5.2 Scores

Des scores basés sur des tableaux de contingence sont utilisés pour quantifier la capacité des différents U-Nets à détecter correctement l'occurrence ou la non-occurrence des échos arqués. Ces scores sont complémentaires de la comparaison des attributs des objets labellisés et détectés qui a été évoquée dans la partie précédente. Des tableaux de contingence sont calculés sur la base de validation comme décrits dans le tableau 5.2. Le risque d'avoir plusieurs échos arqués sur le même champ est très faible car les échos arqués sont des événements rares. Le tableau de contingence est donc basé sur la détection d'au moins un écho arqué labellisé ou détecté sur l'ensemble de la grille pour chaque champ (sans se poser la question du nombre d'écho arqué). A partir de ce tableau de contingence, le taux de bonnes détections (TD ou HR en anglais, Eq 5.1) et le taux de fausses alarmes (TFA ou FAR en anglais, Eq 5.2) sont utilisés pour évaluer les performances du U-Net :

$$TD = \frac{a}{a+c} \tag{5.1}$$

$$TFA = \frac{b}{a+b} \tag{5.2}$$

Le score de Jaccard (ou CSI en anglais, Eq 5.3) sera utilisé pour combiner les fausses alarmes et les non détections :

$$Jac = \frac{a}{a+b+c} \tag{5.3}$$

Ce chapitre a permis d'introduire l'ensemble les éléments qui seront utilisés dans la suite de ces travaux pour entraîner des U-Nets puis évaluer leurs performances. La suite de ces travaux sur les échos arqués sera dans le chapitre 7 et commencera par des tests de sensibilité concernant les différents hyperparamètres introduits dans ce chapitre afin de mieux comprendre leurs rôles et choisir la meilleure configuration pour détecter des échos arqués.

Bibliographie

M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow : Large-Scale Machine Learning on Heterogeneous Systems, 2015. URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

O. Caumont, V. Ducrocq, G. Delrieu, M. Gosset, J. Pinty, J. Parent du Châtelet, H. Andrieu, Y. Lemaître, and G. Scialom. A Radar Simulator for High-Resolution Nonhydrostatic Models. *J. Atmos. Oceanic Technol.*, 23 :1049–1067, 2006. doi : 10.1175/JTECH1905.1.

A. Dutta and A. Zisserman. The VGG image annotator (VIA). *ArXiv*, abs/1904.10699, 2019.

P. Markowski and Y. Richardson. *Mesoscale Meteorology in Midlatitudes*. John Wiley and Sons, Ltd, 2010. doi : 10.1002/9780470682104.ch9.

J. Marshall and W. Palmer. The Distribution of Raindrops with Size. *J. Atmos. Sci.*, 5 :165–166, 1948. doi : 10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2.

A. Mounier, L. Raynaud, L. Rottner, M. Plu, P. Arbogast, M. Kreitz, L. Mignan, and B. Touzé. Detection of Bow Echoes in Kilometer-Scale Forecasts Using a Convolutional Neural Network. *Artificial Intelligence for the Earth Systems*, 1 (2) :e210010, 2022. doi : 10.1175/AIES-D-21-0010.1.

O. Ronneberger, P. Fischer, and T. Brox. U-Net : Convolutional Networks for Biomedical Image Segmentation. *LNCS*, 9351 :234–241, 2015. doi : 10.1007/978-3-319-24574-4_28.

Chapitre 6

Réduction de dimension avec un autoencodeur

Dans le chapitre précédent, nous avons vu l'utilisation d'un réseau de neurones convolutif (U-Net) afin de détecter des échos arqués automatiquement dans les sorties des modèles à partir de labellisations faites à la main. Ce réseau de neurones est efficace pour traiter les problèmes de segmentation d'image. Le U-Net n'est cependant pas la bonne architecture pour réaliser une réduction de dimension, ce qui constitue la première étape pour définir des scénarios comme décrit dans le chapitre 3 section 3.4.2 (figure 3.14). Pour réaliser cette compression de l'information, un autoencodeur sera utilisé. L'objectif d'un autoencodeur est de reproduire en sortie aussi fidèlement que possible les données d'entrée (Goodfellow et al., 2016). Mais pour cela l'autoencodeur devra compresser l'information (partie encodage), transiter par un espace latent, puis reconstruire l'image de départ à partir de cet espace latent (partie décodage). C'est donc à terme l'espace latent où la donnée est compressée qui nous intéressera. Ce chapitre va s'intéresser à la construction des autoencodeurs afin de compresser l'information issue du champ de cumul de pluie sur 1h (RR1h) qui est pour rappel la variable sélectionnée pour développer cette approche par scénario. La première partie s'intéressera à la construction des jeux de données (apprentissage et validation). Contrairement au chapitre précédent sur le U-Net, la principale difficulté réside dans la grande quantité de données disponible pour réaliser l'entraînement et la nécessité d'en sélectionner seulement une partie. Cette première partie détaillera donc les méthodes utilisées pour réaliser cette sélection. La deuxième partie détaillera l'architecture retenue pour l'autoencodeur et les différents hyperparamètres qui seront étudiés. La troisième partie présentera les scores orientés objets qui seront utilisés pour évaluer les performances des différents autoencodeurs. La sélection des hyperparamètres selon ces scores sera détaillée dans la section 4. La dernière partie présentera une comparaison des performances des autoencodeurs avec deux autres méthodes : l'Analyse en Composantes Principales (ACP) et les ondelettes.

6.1 Construction des jeux de données

6.1.1 Base d'apprentissage

Pour obtenir un autoencodeur le plus robuste possible, il est souhaitable d'obtenir la base de données la plus diversifiée et représentative possible. Pour obtenir cette diversité, il est possible de prendre en compte le plus de dates possibles mais aussi plusieurs membres différents. Ces deux aspects sont pris en compte pour créer les bases d'apprentissage et de validation.

La base d'apprentissage utilisera les prévisions de la PE-AROME sur 4 ans de données du 1er janvier 2017 au 31 décembre 2020. Pour chaque jour, deux réseaux différents de PE-AROME seront utilisés (09h/21h ou 03h/15h selon les disponibilités, cf figure 1.13). Seulement les membres 1 à 8 seront extraits car la base de données serait trop importante avec des temps de désarchivage très long et une occupation en mémoire trop importante si l'ensemble des membres étaient exploités. Les échéances de +2h [1] à +45h seront utilisées. Enfin, les champs extraits sont découpés en 5 sous-zones régionales (figure 3.13), l'objectif étant de définir des scénarios régionaux et non nationaux (cf chapitre 3, section 3.4.1). Ces zones régionales ont été définies en collaboration avec des prévisionnistes. Elles sont de tailles rigoureusement identiques (256×384 points de grille) afin d'utiliser le même autoencodeur pour traiter chaque sous-domaine et non cinq différents ce qui serait long à mettre en place. Les performances des autoencodeurs seront cependant regardées zone par zone dans la section 6.4.6 pour s'assurer que les réseaux entraînés proposent une qualité de compression de l'information équivalente. Au total, la base ainsi obtenue contient 4 ans $\times$ 2 réseaux $\times$ 8 membres $\times$ 44 échéances $\times$ 5 zones $\approx$ 5 millions d'échantillons de champs de RR1h, chacun étant une grille de 256×384 points (figure 6.1, gauche).

Ce jeu de données est trop important pour être utilisé en l'état car il nécessiterait encore une place en mémoire et une puissance de calcul trop importante, il doit être réduit. Comme la variable est ici le champ de RR1h, il y a un grand nombre d'échantillons nuls ou presque nuls sur l'intégralité d'un sous-domaine correspondant à l'ensemble des journées où il ne pleut pas. Ces données nombreuses et extrêmement redondantes ne présentent que peu d'intérêt pour l'apprentissage de l'autoencodeur et peuvent être supprimées (Sha et al., 2020). Dans la suite, un échantillon est considéré comme avec peu ou pas de pluie si 99% des points du domaine n'atteignent pas 0.4 mm (ou autrement dit si le quantile spatial à 99% ne dépasse pas 0.4 mm). Ce seuil et ce quantile ont été choisis parmi d'autres car ils permettent de réduire notoirement le jeu de données en le divisant par deux (figure 6.1, centre) tout en préservant les données les plus intéressantes avec les zones de pluies les plus grandes et les plus intenses. On aboutit alors à un jeu de données d'environ 2,5 millions d'échantillons.

1. Les premières échéances ne sont pas considérées car peuvent être soumises aux problèmes de mise en équilibre du modèle (spin-up).

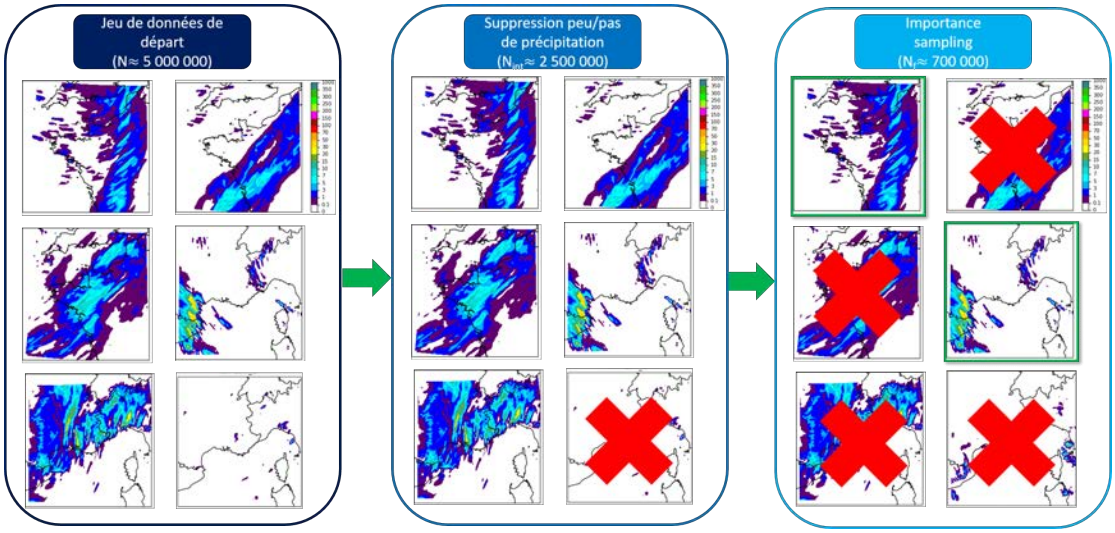


FIGURE 6.1 – *Construction du jeu de données. Le jeu de données est constitué de champs de RR1h sur les 5 zones régionales (même si seulement les zones Nord-Ouest et Sud-Est sont représentées ici). Un jeu de données de plus de $N = 5$ millions d'échantillons est constitué au départ (gauche). Une première sélection est faite pour supprimer les échantillons dont le Q99 n'atteint pas les 0.4mm permettant de diviser environ par deux le jeu de données (centre, N_{int}). Enfin une méthode d'importance sampling est appliquée sur ces données pour ne sélectionner que environ 700 000 échantillons (N_f). Les échantillons restants sont symboliquement encadrés en vert.*

6.1.2 Sélection par échantillonnage préférentiel (importance sampling)

Le jeu de données ainsi réduit à 2.5 millions d'échantillons est cependant toujours trop imposant pour être utilisé lors de l'entraînement. Il est nécessaire de le réduire encore. Une méthode sous la forme d'échantillonnage préférentiel ou importance sampling [2] sera utilisée, inspirée par les travaux de Ravuri et al. (2021). Cette méthode permet de réduire le jeu de données en évitant les effets de seuil comme pour les méthodes présentées dans le chapitre 5 section 5.3.1, tout en sélectionnant préférentiellement les échantillons qui sont les plus intéressants pour l'entraînement (selon des critères à définir). Cela rend également cette méthode plus intéressante qu'un simple tirage aléatoire.

L'importance sampling s'appuie sur la définition d'une valeur d'intérêt pour chaque échantillon a, qu'on peut noter $I(a)$. La formule utilisée par Ravuri et al. (2021) sur des réflectivités est reprise ici pour les RR1h :

$$I(a) = \frac{1}{n}\sum_{i=1}^{n}(1 - e^{-\frac{x_{a,i}}{s_{rr}}}) \tag{6.1}$$

Avec $x_{a,i}$ le cumul de pluie sur 1h du i-ème point de grille de l'échantillon a, s_{rr} un hyperparamètre à fixer et n le nombre total de points de grille (ici 256×384). L'hyperparamètre s_{rr} est homogène à la variable étudiée et peut s'interpréter

2. Le terme anglais sera utilisé dans la suite de ce manuscrit.

comme le seuil approximatif à partir duquel l'expert souhaite prêter une attention particulière. L'intérêt $I(a)$ varie de 0 [3] à 1, une grille remplie de zéros correspondant à un intérêt nul alors qu'une grille où une grande majorité des points dépasserait s_{rr} aurait un intérêt proche de 1. Cette variable d'intérêt $I(a)$ permet de définir une probabilité $P(a)$ de conserver l'échantillon a dans le jeu de données final :

$$P(a) = min\left\{1, P_{min} + m \times I(a)\right\} \tag{6.2}$$

Avec P_{min} et m deux autres hyperparamètres à fixer. P_{min} correspond à la probabilité minimale de conserver une image (dans le cas où $I(a)$ est nul). L'hyperparamètre m est un facteur multiplicatif qui permet de mieux distinguer les images entre elles. En effet en pratique $I(a)$ n'est pas très élevé et ses variations sont faibles car même en cas de fortes pluies, celles-ci ne couvrent jamais l'ensemble de la grille au même moment. Cette quantité $P_{min} + m \times I(a)$ est majorée à 1 pour que $P(a)$ varie de 0 à 1. Pour finir, l'échantillon est alors conservé avec une probabilité $P(a)$ dans le jeu de données final. Cette sélection est appliquée sur chaque échantillon du jeu de départ de façon indépendante.

Plusieurs triplets (P_{min}, m, s_{rr}) peuvent être testés et influencent grandement, en particulier, la taille du jeu de données retenue. Plusieurs valeurs pour les hyperparamètres m et s_{rr} seront testées pour entraîner l'autoencodeur. Concernant P_{min}, la valeur a été fixée à un seuil haut égal à 0.1 car les images les moins intéressantes ont déjà été supprimées à l'étape précédente (séléction par le Q99). Pour évaluer la réduction de la taille du jeu de données par l'importance sampling, le pourcentage du jeu de données de départ qui est conservé a été calculé selon m et s_{rr} et est présenté dans le tableau 6.1. Ces valeurs permettent de conserver au maximum un tiers des données de départ, ce qui rend le coût de calcul pour l'entrainement du futur autoencodeur supportable par nos infrastructures. On obtient alors une base d'entrainement qui possède entre approximativement 350 000 échantillons à 880 000 pour le triplet aboutissant à la plus grosse base de données (figure 6.1, droite). On observe qu'une augmentation de s_{rr} induit une diminution de la taille du jeu de données car pour un échantillon donné cette augmentation conduit à une diminution de $I(a)$ et donc de $P(a)$. A l'inverse une augmentation de m induit une augmentation de $P(a)$ et donc augmente la taille du jeu de données. L'importance sampling permet de réduire la taille du jeu de données mais en sélectionnant préférentiellement certains échantillons. Il a également un impact fort sur la variance de ce jeu de données. Pour illustrer cet impact, la carte de variance du jeu de départ et celle obtenue après importance sampling pour le triplet $(P_{min} = 0.1, m = 1.5, s_{rr} = 1)$ sont représentées figure 6.2. On remarque que l'importance sampling tend à fortement augmenter la variance en sélectionnant les images avec les plus fortes pluies. De plus la variance n'est pas homogène sur l'ensemble de la grille à cause de différentes zones de relief mais elles ne sont pas toujours simples à distinguer car la base de données mélange 5 sous-zones régionales différentes. On peut cependant reconnaître les Alpes dans la partie en haut à droite ou encore le relief italien des Apennins pour les zones avec la plus forte variance. La zone à forte variance plus à l'Ouest en forme de ligne correspond aux Cévennes. La carte de variance fait donc largement ressortir des zones

3. Comme on s'intéresse aux pluies, $x_{a,i}$ ne peut pas prendre de valeurs négatives.

TABLE 6.1 – *Pourcentage du jeu de données de départ conservé après importance sampling selon 2 valeurs de m (1.5 et 3) et 4 valeurs de s_{rr} (1mm, 3mm, 5mm et 7mm). P_{min} est fixé à 0.1.*

m \ s_{rr}	1mm	3mm	5mm	7mm
1.5	21.3%	15.6%	13.8%	12.9%
3	32.5%	21.2 %	17.6 %	15.8 %

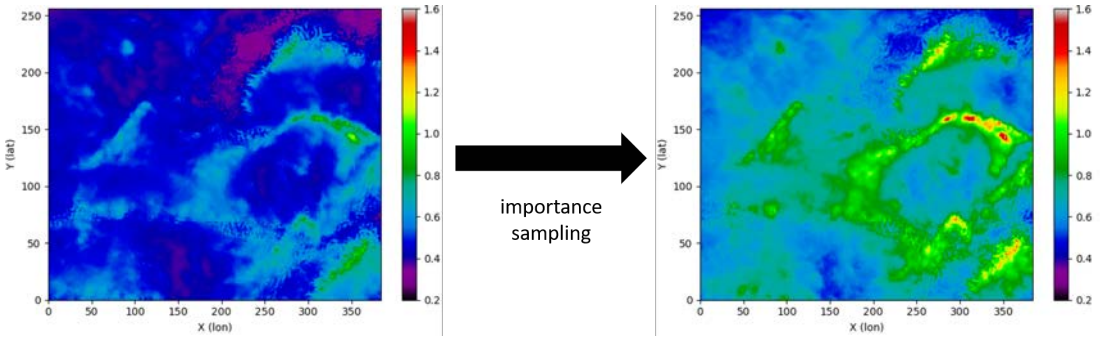


FIGURE 6.2 – *Illustration de l'influence de l'importance sampling sur la variance. Variance du jeu de données avec 2.5 millions d'échantillons à gauche (étape intermédiaire de la figure 6.1). Variance après importance sampling pour les hyperparamètres suivants : P_{min}=0.1, m=1.5 et s_{rr}=1mm à droite. L'échelle de couleur est identique pour les deux cartes.*

propres au Sud-Est de la France qui possède un climat marqué par de nombreux reliefs et d'importants épisodes pluvieux surtout à l'automne. Cette augmentation de variance est un effet souhaité de l'importance sampling qui permet de cibler les échantillons intéressants en évitant les effets de seuils. Les performances des différents autoencodeurs entraînés à partir de ces bases de données citées dans le tableau 6.1 seront comparées dans la section 6.4.

6.1.3 Base de validation pour les hyperparamètres

Comme évoqué dans la section 4.1.4 sur les bases de données, un autre jeu de données indépendant de la base d'apprentissage est construit afin de tester plusieurs valeurs d'hyperparamètres. La base d'apprentissage s'appuyant sur les années 2017 à 2020, cette base de validation sera construite avec l'année 2021. Elle comprendra le réseau PE-AROME de 3h pour chaque journée de l'année 2021, les membres 1 à 4, les échéances de +2 à +45h et avec également les 5 zones régionales ce qui fait un total d'un peu plus de 320 000 échantillons. Cette base est construite avec un seul réseau par jour et avec 4 membres pour limiter le temps de calcul des scores. Mais en conservant l'ensemble des dates d'une année et l'ensemble des échéances cette base reste représentative des RR1h possibles et sur chacune des 5 sous-zones.

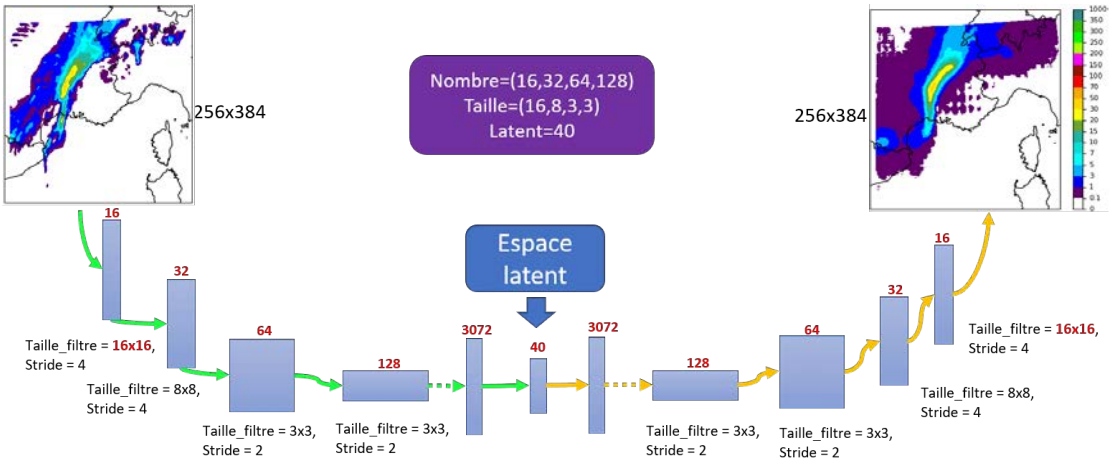


FIGURE 6.3 – *Architecture de l'autoencodeur convolutif. L'autoencodeur utilise en entrée une image de RR1h (en haut à gauche). L'information est compressée avec la partie encodeur (vert) pour aboutir à un espace latent. Puis l'information est décompressée à l'aide de la partie décodeur (orange) pour revenir à une sortie qui a la même taille que l'image d'entrée (en haut à droite). Les flèches en pointillées désignent des étapes de réorganisation des données. Les hyperparamètres du réseau qui seront discutés dans la section 6.4 apparaissent en rouge. L'architecture ici présentée correspond à une suite de filtres 16,32,64,128 de tailles 16,8,3,3 avec un espace latent à 40 dimensions. Les autres architectures testées sont présentées tableau 6.2.*

6.2 Construction de l'autoencodeur

6.2.1 Architecture

Les bases de données présentées dans la partie précédente serviront à entraîner et valider un autoencodeur dont l'architecture est représentée figure 6.3. Cette architecture est inspirée des travaux de Guo et al. (2017) qui ont construit un autoencodeur convolutif afin de faire de la compression d'information. L'autoencodeur se décompose en deux parties entre une partie encodeur qui compresse l'information de l'espace physique jusqu'à l'espace latent et une partie décodeur pour décompresser l'information de l'espace latent vers l'espace physique. Le passage de l'espace physique vers l'espace latent se fait en plusieurs étapes à l'aide de 4 couches de convolution successives avec un stride plus grand que 1 (dans l'ordre 4,4,2 et 2) permettant ainsi de réduire la taille de l'entrée de 256×384 points de grille à un ensemble de 128 grilles 4×6. Plusieurs nombres et tailles de filtres seront testées. Cet ensemble de 128×4×6 données est ensuite réarrangé sous la forme d'un vecteur à 3072 dimensions pour la figure 6.3. Pour finir sur la partie encodage, une couche dense est ajoutée pour passer de ce vecteur à un espace latent à quelques dizaines de dimensions. A ce stade, l'information a donc été compressée au maximum en passant d'environ 98 000 points de grille à un espace à moins de 50 dimensions. Plusieurs tailles pour cet espace latent seront testées. La partie décodage est elle construite de façon symétrique avec une couche dense qui permet de passer de l'espace latent à un vecteur de taille égale à la partie encodeur (ici 3072). Ces données sont réorganisées pour reformer un tableau de 128×4×6 et 4 couches de convolution transposée sont appliquées pour revenir à

une grille de 256×384 points. Contrairement au U-Net, aucune couche de dropout n'est utilisée dans ce réseau car les premiers tests ont montré une dégradation de l'apprentissage que ces couches de dropout soient ajoutées dans les couches les plus profondes ou seulement dans les couches externes de ce réseau.

6.2.2 Entraînement

L'autoencodeur est construit et entraîné à l'aide de la librairie TensorFlow/Keras (Abadi et al., 2015). Il est entraîné à partir de la base d'apprentissage qui est divisée en deux avant l'entraînement (80%/20%) comme expliqué dans la section 4.1.4 afin de détecter le surapprentissage et sélectionner le réseau à l'issue de la meilleure epoch sur les 20% du jeu de données mis de côté. La taille des batchs utilisés lors de l'apprentissage est un hyperparamètre et plusieurs valeurs seront testées. Le nombre maximal d'epoch est fixé à 80, les premiers tests ayant montré qu'il était inutile de continuer plus loin l'apprentissage. Les données servant à mettre à jour les poids (80%) sont remélangées entre chaque epoch. Le learning rate est fixé à 10^{-3}, valeur usuelle et implémentée par défaut car des modifications de celui-ci n'ont pas montré d'apports ou de dégradations significatives. Enfin, la fonction coût utilisée est la Mean Squared Error (MSE) avec pour objectif de reproduire en sortie le champ qui est en entrée du réseau.

6.2.3 Récapitulatif des hyperparamètres

Plusieurs valeurs d'hyperparamètres seront testées pour tenter de trouver l'autoencodeur avec la meilleure capacité à compresser l'information. Un premier hyperparamètre est la taille des batchs (B_{size}) qui concerne le processus d'entrainement de l'autoencodeur (section 4.1.3). Deux autres hyperparamètres concernent l'importance sampling (m et s_{rr}, section 6.1.2). Les trois derniers hyperparamètres sont en lien avec l'architecture du réseau de neurones dont deux concernent les couches de convolutions (cf section 4.2.1) à travers le nombre de filtres ($N_{filtres}$) et la taille du filtre pour la première couche (S_{filtre}). Le dernier hyperparamètre concerne la taille de l'espace latent au centre de l'autoencodeur (section 6.2.1. Les valeurs testées sont résumées dans le tableau 6.2. La taille de l'espace latent est cependant un hyperparamètre particulier car il a une influence bien supérieure sur la qualité de reconstruction de l'image comparé aux autres hyperparamètres. L'objectif sera alors de définir plutôt trois autoencodeurs optimisés avec pour chacun une taille d'espace latent différente. Ces trois autoencodeurs seront étudiés dans la suite des travaux sur les scénarios pour choisir la taille d'espace latent la plus adaptée (partie 4).

6.3 Scores orientés objets

Pour comparer les performances des autoencodeurs, les scores seront calculés en comparant les entrées et les sorties du réseau de neurones sur la base de validation présentée section 6.1.3. Même si l'objectif est ici d'utiliser l'espace latent et non la sortie de l'autoencodeur, celui possédant les sorties les plus ressemblantes aux données d'entrées est aussi celui possédant l'espace latent avec les capacités de

TABLE 6.2 – *Liste des hyperparamètres liés à l'entraînement de l'autoencodeur qui seront testés (colonne de gauche) et leurs différentes valeurs (colonne de droite).*

Entraînement	
Taille des batchs (B_{size})	64, 128, 256, 512 et 1024
Importance sampling	
coefficient multiplicateur (m)	1.5 et 3
seuil d'intérêt (s_{rr})	1,3,5 et 7mm
Architecture (Pour les suites, de la couche la plus externe à la plus profonde)	
nombre de filtres de convolution ($N_{filtres}$)	(16,32,64,128) et (32,64,128,256)
taille du premier filtre (S_{filtre})	16×16 et 8×8
taille de l'espace latent	10,20 et 40

compression les plus intéressantes. Pour faire cette comparaison entre les entrées et les sorties, des scores orientés objets sont calculés. Trois types d'objets sont définis avec des pluies plus ou moins fortes afin d'évaluer séparément sur ces trois catégories les autoencodeurs. La méthode pour détecter ces objets sera d'abord brièvement décrite avant de présenter les scores utilisés pour comparer les réseaux de neurones.

6.3.1 Méthode de détection des objets

Les 3 types d'objets qui sont détectés correspondent à des objets de pluies totales, des objets de pluies modérées et des objets de pluies fortes. L'objectif est d'évaluer la capacité de l'autoencodeur à reproduire ces 3 types d'objets lorsqu'ils sont détectés dans les champs de RR1h en entrée de l'autoencodeur. Un autre objectif est d'évaluer les différences qui pourraient exister concernant la localisation des pluies. Des caractéristiques plus détaillées concernant l'étendue spatiale des pluies à travers la taille des objets et leurs intensités avec le Q90 à l'intérieur des objets seront détaillées plus tard dans le chapitre 9. L'extraction des attributs sera faite comme dans le chapitre 5, section 5.5.1 pour les échos arqués. Ce découpage en trois objets (total,modéré,fort) est issu des travaux de Rottner et al. (2019). Cependant la méthode présentée dans cet article est trop coûteuse pour être appliquée sur de longues séries de données. C'est pourquoi une méthode plus proche de Davis et al. (2006) est privilégiée. Pour définir un des 3 types d'objets, elle se décompose en deux étapes (figure 6.4) : la première étape consiste à lisser le champ à l'aide d'un filtre gaussien, le champ est alors plus ou moins lissé selon la valeur de l'écart-type σ choisi pour définir la gaussienne. La deuxième étape est un seuillage s_{obj} qui est appliqué sur le champ. Les points de grille dépassant ce seuil sont alors considérés dans l'objet en question et les autres en dehors. Les valeurs de σ et s_{obj} pour chacun des 3 objets sont présentées dans le tableau 6.3. Le seuil s_{obj} pour les pluies totales est bas pour englober l'ensemble des précipitations. Les seuils s_{obj} pour les pluies modérées et fortes correspondent approximativement à ceux fournis par Météo-France : `http://pluiesextremes.meteo.fr/france-metropole/Intensite-de-precipitations.html`. Concernant les valeurs de σ, elles ont été

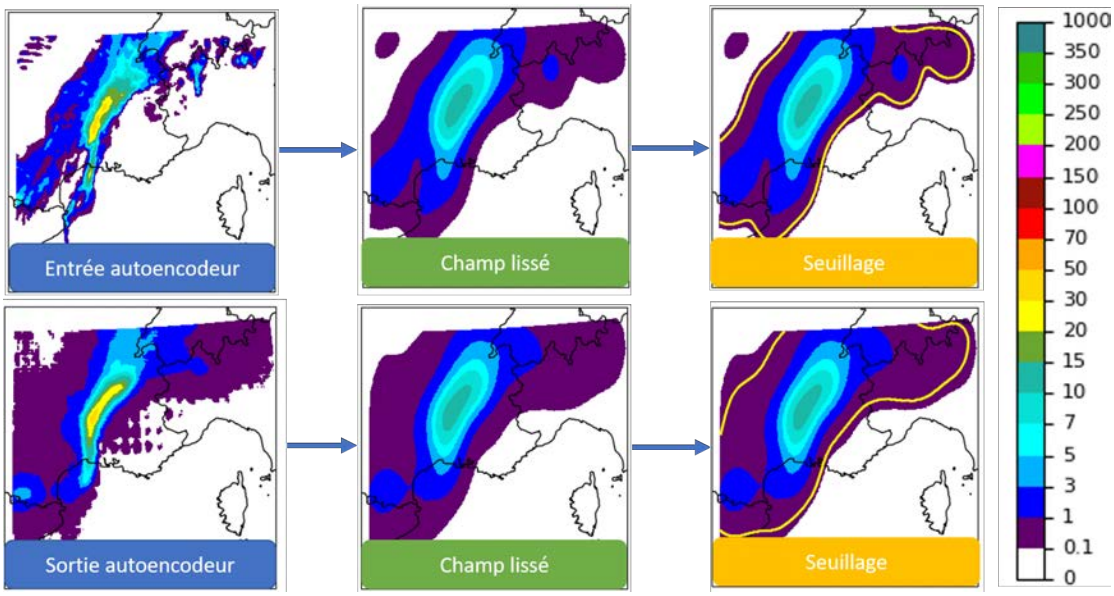


FIGURE 6.4 – *Méthode de détection des objets illustrée avec les objets de pluies totales. Exemple sur les mêmes images que pour la figure 6.3. Les entrée et sortie de l'autoencodeur sont lissées avec un filtre gaussien (ici d'écart-type égal à 15 points de grille, centre). Puis un seuillage à 0.2 mm est appliqué pour définir les objets de pluies totales (contours jaunes, droite).*

TABLE 6.3 – *Valeurs de σ en km ainsi que s_{obj} en mm pour les trois objets : pluie totale, modérée et forte.*

Objets	σ	s_{obj}
Pluie totale	37.5	0.2
Pluie modérée	30	3
Pluie forte	10 25	7

sélectionnées après des tests visuels sur quelques images. L'intensité du lissage diminue avec l'augmentation du seuil car les zones de pluies étudiées sont de plus en plus petites et un lissage trop marqué engendrerait par exemple la disparition des zones de pluie forte. Cette méthode objet est appliquée à la fois sur les champs de RR1h en entrée de l'autoencodeur et ceux reconstruits en sortie.

6.3.2 Scores à partir des attributs

A partir des objets détectés, plusieurs caractéristiques sont extraites qui permettent de définir 3 scores :

- *Taux de Détection (TD)* : D'une façon similaire à celle présentée dans le chapitre 5 (section 5.5.2), un tableau de contingence est construit (Table 5.2). Ce tableau de contingence était construit en vérifiant la présence ou non d'un écho arqué entre les données labellisées et la prediction par le réseau de neurones. Ici, le tableau de contingence est construit selon la présence ou non d'un objet sur le champ en entrée et/ou en sortie de l'autoencodeur.

Ce tableau de contingence est fait de façon indépendante pour les objets de pluies totales, modérées et fortes (3 tableaux donc au total). Pour un type d'objet donné, le TD correspond au nombre de champs où un objet est détecté à la fois en entrée et en sortie de l'autoencodeur divisé par le nombre de champs où un objet est détecté en entrée. Un TD égal à 1 indique donc que l'autoencodeur reconstruit systématiquement l'objet présent en entrée.

- *Taux de Fausses Alarmes (TFA)* : Le TFA se construit de façon similaire avec le même tableau de contingence que le TD. Il correspond au nombre de champs où un objet est détecté en sortie mais où aucun n'est détecté en entrée de l'autoencodeur divisé par le nombre total de champs où un objet est détecté en sortie de l'autoencodeur. Un TFA nul indique donc que l'autoencodeur ne produit jamais un objet si celui-ci n'existe pas en entrée. Comme dans la section 5.5.2 du chapitre 5 le TFA n'est pas défini en divisant par le nombre de champs où aucun objet n'est détecté en sortie de l'autoencodeur, qui est une définition plus classique, car la pluie est un évènement trop rare. Le nombre de champs où aucun objet en sortie de l'autoencodeur n'est détecté est important à cause de ses nombreux champs sans pluie, ce qui implique des TFA proche de 0 quelque soit l'autoencodeur considéré ce qui ne permet pas de les discriminer avec cette définition du TFA.

- $\overline{dist}$: La dernière grandeur peut s'apparenter à une erreur de position moyenne. Elle repose sur les centres de masses du ou des objets détectés s'ils sont plusieurs sur la même image (position X, Y pour chaque objet). Les distances entre les centres de masses en entrée et en sortie sont calculées. Un objet en sortie de l'autoencodeur est associé à l'objet en entrée qui a le centre de masse le plus proche de lui. Les associations sont faites des objets en sortie vers les objets en entrée et non l'inverse car les objets en sortie sont généralement moins nombreux, l'autoencodeur ayant tendance à lisser le champ. Les distances d entre les objets sont moyennées sur l'ensemble des cas dans la base de validation pour former la troisième grandeur étudiée $\overline{dist}$ qui s'exprime en kilomètre :

$$\overline{dist} = \frac{1}{T} \sum_{k=1}^{K} \sum_{i=1}^{N_k^{sortie}} d(i, match_k(i)) \tag{6.3}$$

Avec K désignant l'ensemble des dates, échéances et membres présents dans la base de validation, i l'index des objets en sortie de l'autoencodeur pour un nombre total de N_k^{sortie} objets. $match_k(i)$ désigne l'indice de l'objet d'entrée dans l'échantillon k qui a été associé à l'objet i de sortie. La quantité T désigne le nombre d'objets en sortie de l'autoencodeur sur l'ensemble de la base de validation ET où simultanément des objets en entrée sont bien détectés. La section suivante montrera que le TFA n'est pas toujours nul et que donc il peut arriver, même si ces cas de figure sont rares, qu'un objet en sortie soit détecté alors que ce n'était pas le cas en entrée. Dans ces cas, tenter de calculer une distance n'a pas de sens et c'est pourquoi ils sont retirés du calcul.

Ces trois grandeurs (TD, TFA et $\overline{dist}$) sont scrutées pour chaque type d'objets (total, modéré, fort). Ce qui fait donc 9 valeurs pour évaluer les autoencodeurs dans la section suivante.

6.4 Sélection des hyperparamètres

Cette partie s'intéresse à la sélection des valeurs pour les hyperparamètres qui sont présentés dans le tableau 6.2 à l'aide des 9 grandeurs définies dans la section précédente. La base de validation repose sur un très grand nombre d'objets avec un total de 413 654 objets de pluies totales, 46 684 de pluies modérées et 8 062 de pluies fortes [4]. Cependant, calculer ces scores sur un seul entraînement peut être problématique car la séparation aléatoire en deux de la base d'apprentissage, l'initialisation aléatoire des poids de l'autoencodeur ou encore le réarrangement de la base d'entraînement entre chaque epoch peuvent induire des résultats différents pour des autoencodeurs avec pourtant les mêmes valeurs d'hyperparamètres. C'est pourquoi chaque combinaison sera testée 3 fois à l'identique et non une seule fois pour tenter d'établir des résultats plus robustes. Les temps de calcul ne permettent pas de faire plus que 3 tentatives. Pour les mêmes raisons, toutes les combinaisons ne pourront malheureusement pas être testées. La taille des batchs et le jeu de données après importance sampling vont ainsi être sélectionnés avant de s'intéresser aux hyperparamètres liés à l'architecture du réseau.

6.4.1 Taille des batchs

La taille du batch va être sélectionnée avec une architecture fixée et un jeu de données fixé également. L'architecture utilisée comporte 16, 32, 64, 128, 256 filtres de taille respective 16, 8, 3 et 3 avec un espace latent à 20 dimensions (cf tableau 6.2). Les hyperparamètres pour l'importance sampling sont (P_{min}=0.1,m=1.5,s_{rr} = $3mm$). Les différentes expériences sont rassemblées dans le tableau 6.4. On peut tout d'abord remarquer que les TD et TFA ont tendance à se dégrader quand on regarde les objets les plus intenses par rapport aux objets de pluies totales mais que l'erreur de position diminue. Cependant, les objets de pluies totales étant beaucoup plus gros que les objets de pluies fortes (environ 10 fois plus gros en moyenne), ce gain sur l'erreur de position est à relativiser vis à vis de la taille des objets. Ces moins bons résultats sur les objets de pluies fortes s'expliquent majoritairement par leurs plus petites tailles et leurs plus fortes intensités qui les rendent plus difficile à reproduire fidèlement dans un objectif de réduction de dimension. On note que les résultats restent malgré tout très bons avec des TD à plus de 88% même pour les objets de pluies fortes. Puisque l'autoencodeur tend à lisser les objets, il peut sembler étonnant d'obtenir des TFA non nuls. L'étude des fausses alarmes révèlent que ce sont exclusivement des objets très petits (non montrée). Il est possible que le problème vienne de la méthode objet. En effet, en présence d'objets particulièrement petits et intenses, comme une situation orageuse peut en produire, un lissage trop important peut « éliminer » complètement une possible détection. Mais le champ de pluie en sortie de l'autoencodeur, plus faible mais plus

4. Le décompte est fait sur les données d'entrée car pas sensible à l'autoencodeur.

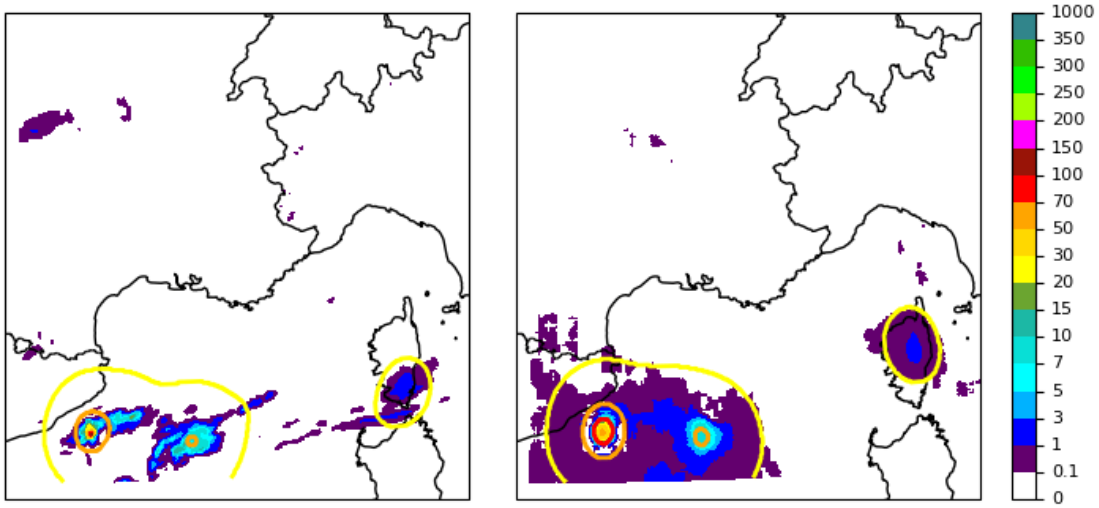


FIGURE 6.5 – *Exemple de fausse alarme pour un objet de pluies fortes. Le champ de gauche correspond à l'entrée de l'autoencodeur et à droite le champ reconstruit. Les contours jaunes représentent les objets de pluies totales, les contours oranges les objets de pluies modérées et le contour rouge sur le champ de droite un objet de pluies fortes.*

étalé spatialement, peut être détectée. C'est par exemple ce qu'il se produit sur l'exemple figure 6.5 pour un objet de pluies fortes. On peut remarquer en parallèle sur la figure 6.5 de larges plages de pluies faibles sur le champ reconstruit. Ces plages ne sont pas très importantes car l'espace latent de l'autoencodeur est le coeur de ces travaux et non l'image reconstruite, elles ne seront donc pas étudiées par la suite [5].

En se recentrant sur la taille des batchs, on note que l'augmentation de taille du batch de 64 à 256 tend à stabiliser les résultats d'une expérience à l'autre avec des scores qui semblent aussi en légère progression. Cependant les scores se dégradent en particulier pour les pluies modérées et fortes en continuant d'augmenter la taille des batchs. Une trop forte augmentation de la taille des batchs peut en effet dégrader les scores (Keskar et al., 2016). Pour la suite des expériences, la taille du batch sera donc fixée à **256**. La significativité des différences, qui restent relativement faibles, n'a pas été testée et pourra être étudiée dans des travaux ultérieurs.

6.4.2 Importance sampling

Comme pour la taille des batchs, la sélection des hyperparamètres m et s_{rr} pour l'importance sampling se fera à architecture constante qui est toujours la même : (16,32,64,128,256) filtres de tailles (16,8,3,3) avec un espace latent à 20 dimensions. La taille des batchs est donc fixée à 256. Les résultats sont présentés dans le tableau 6.5. La comparaison pour les deux valeurs de m semble montrer que, à s_{rr} fixé, les erreurs de position sont réduites lorsque m est plus grand en particulier pour les objets de pluies modérées et fortes (aucun gain significatif concernant les TD et TFA). L'augmentation de s_{rr} ne semble pas changer les scores mais la stabilité entre les n=1,2,3 est moins bonne. Cette perte de stabilité

5. Des figures en annexe de ce chapitre montreront que cette plage de faible pluie diminue en augmentant la taille de l'espace latent.

TABLE 6.4 – *Performances des autoencodeurs avec des tailles de batchs (B_{size}) différentes. n représente les 3 entraînements différents pour les mêmes hyperparamètres. Les scores sont divisés en trois catégories pour chaque type d'objet. Les TD et TFA sont arrondis au centième près et $\overline{dist}$ au dixième de kilomètre. La taille de batch retenue apparaît en gras.*

B_{size}	n	Pluie totale			Pluie modérée			Pluie forte		
		TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$
64	1	0.95	0	40.2	0.91	0.03	16.4	0.88	0.06	12.8
	2	0.96	0	37.9	0.84	0.01	15.0	0.79	0.02	10.9
	3	0.97	0	38.9	0.87	0.01	15.8	0.85	0.03	11.9
128	1	0.95	0	38.3	0.88	0.02	16.1	0.86	0.04	12.0
	2	0.97	0	38.4	0.87	0.01	15.2	0.82	0.02	11.0
	3	0.96	0	38.5	0.85	0.01	14.7	0.82	0.02	10.4
256	**1**	**0.95**	**0**	**37.9**	**0.89**	**0.02**	**16.1**	**0.88**	**0.04**	**12.2**
	2	**0.96**	**0**	**37.4**	**0.91**	**0.02**	**16.6**	**0.87**	**0.04**	**12.0**
	3	**0.96**	**0**	**37.6**	**0.9**	**0.02**	**16.3**	**0.88**	**0.04**	**12.2**
512	1	0.96	0	36.8	0.85	0.01	15.6	0.81	0.02	11.2
	2	0.97	0	37.9	0.9	0.02	16.3	0.87	0.04	12.2
	3	0.96	0	37.0	0.87	0.01	16.3	0.82	0.02	12.1
1024	1	0.95	0	37.8	0.9	0.02	16.5	0.88	0.05	12.1
	2	0.96	0	37.5	0.88	0.01	16.0	0.84	0.03	11.4
	3	0.97	0	38.0	0.9	0.02	16.5	0.87	0.05	12.4

peut s'attribuer en partie à la diminution de la taille du jeu de données pour l'apprentissage. La configuration finale retenue est la plus stable mais celle aussi qui présente les TD et TFA les plus hauts pour les objets de pluies modérées et fortes. Il s'agit du couple $m|s_{rr}$= **1.5|1mm**.

6.4.3 Espace latent à 10 dimensions

Les étapes précédentes ayant permis de sélectionner les hyperparamètres liés à la taille des batchs et à l'importance sampling, la dernière étape consiste à sélectionner ceux qui concernent l'architecture du réseau. L'objectif est de ne retenir que 3 autoencodeurs avec chacun une taille d'espace latent différente. Les différentes architectures seront testées de nouveau pour chaque taille d'espace latent. Les résultats des différents entraînements pour l'espace latent à 10 dimensions sont présentés tableau 6.6. On peut remarquer en premier lieu que les scores ont baissé par rapport aux expériences précédentes. En effet, l'espace latent étant plus petit ici, la quantité d'information qui peut transiter par celui-ci est aussi plus faible et donc la qualité de la reconstruction se dégrade. Les variations du nombre de filtres ou de la taille du premier filtre n'ont pas un impact majeur sur les scores. On remarque que la stabilité de l'apprentissage entre les différents n est généralement dégradée par rapport aux expériences précédentes. Le choix final pour l'autoencodeur avec 10 dimensions se porte sur celui possédant l'un des meilleurs TD pour les fortes pluies, c'est-à-dire l'autoencodeur avec une suite de

TABLE 6.5 – *Tableau similaire au tableau 6.4 pour la sélection des paramètres d'importance sampling m et s_{rr}. La configuration retenue apparaît en gras.*

$m\|s_{rr}$	n	Pluie totale			Pluie modérée			Pluie forte		
		TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$
1.5|1	**1**	**0.95**	**0**	**37.7**	**0.93**	**0.04**	**15.8**	**0.9**	**0.05**	**11.1**
	2	**0.95**	**0**	**38.4**	**0.92**	**0.03**	**16.0**	**0.88**	**0.04**	**11.1**
	3	**0.95**	**0**	**37.6**	**0.9**	**0.02**	**15.8**	**0.88**	**0.05**	**11.1**
1.5|3	1	0.95	0	38.0	0.89	0.02	16.1	0.88	0.04	12.2
	2	0.96	0	37.4	0.91	0.02	16.6	0.87	0.04	12.0
	3	0.96	0	37.6	0.9	0.02	16.3	0.88	0.04	12.2
1.5|5	1	0.96	0	38.7	0.89	0.02	16.4	0.88	0.05	12.0
	2	0.95	0	37.5	0.88	0.01	15.9	0.84	0.03	12.2
	3	0.96	0	37.8	0.89	0.02	16.3	0.85	0.03	11.1
1.5|7	1	0.96	0	39.1	0.9	0.02	17.1	0.88	0.05	13.0
	2	0.96	0	38.5	0.87	0.01	16.2	0.82	0.02	11.0
	3	0.96	0	37.4	0.86	0.01	15.5	0.81	0.03	11.0
3|1	1	0.94	0	37.7	0.91	0.02	14.6	0.88	0.03	10.3
	2	0.95	0	37.9	0.91	0.02	15.1	0.87	0.03	10.2
	3	0.95	0	37.4	0.87	0.01	14.0	0.83	0.02	9.1
3|3	1	0.96	0	37.4	0.91	0.02	15.7	0.88	0.04	11.3
	2	0.96	0	37.7	0.84	0	13.9	0.79	0.01	9.2
	3	0.96	0	38.0	0.91	0.03	15.6	0.89	0.04	10.8
3|5	1	0.95	0	38.1	0.87	0.01	15.2	0.84	0.02	11.0
	2	0.98	0.01	36.8	0.88	0.01	14.9	0.83	0.02	10.4
	3	0.96	0	38.0	0.89	0.02	14.7	0.86	0.03	10.8
3|7	1	0.97	0	37.2	0.92	0.02	16.7	0.89	0.05	12.5
	2	0.96	0	38.4	0.88	0.01	15.9	0.86	0.03	11.3
	3	0.94	0	38.0	0.85	0.01	14.7	0.81	0.02	10.6

filtres **(16,32,64,128)**, un premier filtre de taille $\mathbf{16 \times 16}$ et d'index $\mathbf{n = 3}$. La configuration (32,64,128,256) et 16×16 avec $n = 3$ est proche mais ses résultats moins bons sur les pluies totales incitent à retenir l'autre configuration.

6.4.4 Espace latent à 20 dimensions

La même étude est proposée pour sélectionner l'espace latent à 20 dimensions (tableau 6.7). Les trois premières lignes sont les mêmes que les trois premières du tableau 6.5 car elles correspondent aux trois mêmes expériences. Les changements d'architectures semblent un peu plus impacter les scores que pour l'espace latent à 10 dimensions. En particulier la diminution de la taille du premier filtre ou l'augmentation du nombre de filtres réduisent légèrement les erreurs de position moyennes pour les pluies modérées et fortes ainsi que le TFA. Mais cette amélioration se fait au détriment du TD. Avec une stratégie similaire à celle utilisée pour l'espace latent à 10 dimensions, le réseau retenu est celui possédant le TD le plus

TABLE 6.6 – *Tableau similaire au tableau 6.4 pour la sélection des hyperparamètres concernant l'architecture de l'autoencodeur pour un espace latent à 10 dimensions. Le nombre de filtres $N_{filtres}$ sera désigné sous la forme (16...128) ou (32...256) pour les suites respectivement (16,32,64,128) et (32,64,128,256). La taille du premier filtre S_{filtre} sera désignée avec le nombre 16 (pour 16×16) ou 8 (pour 8×8). L'autoencodeur sélectionné apparaît en gras.*

$N_{filtres}\|S_{filtre}$	n	Pluie totale			Pluie modérée			Pluie forte		
		TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$
(16...128)|16	1	0.88	0	55.8	0.79	0.01	17.3	0.76	0.03	12.2
	2	0.87	0	54.1	0.77	0.01	16.5	0.67	0.02	11.5
	3	**0.91**	**0**	**51.9**	**0.85**	**0.03**	**17.8**	**0.8**	**0.05**	**12.8**
(16...128)|8	1	0.87	0	53.1	0.79	0.01	17.1	0.74	0.02	12.6
	2	0.88	0	54.1	0.83	0.03	18	0.75	0.04	13.4
	3	0.89	0	54.4	0.79	0.01	17.2	0.75	0.03	12.8
(32...256)|16	1	0.87	0	54.2	0.81	0.02	17.2	0.73	0.02	12.4
	2	0.85	0	53.6	0.83	0.02	16.6	0.78	0.04	11.6
	3	0.89	0	53.7	0.84	0.02	17.6	0.8	0.03	12.7
(32...256)|8	1	0.89	0	53.1	0.81	0.01	17.2	0.76	0.03	12.5
	2	0.89	0	53.7	0.79	0.01	17.5	0.71	0.02	12.6
	3	0.89	0	52.3	0.8	0.01	17.3	0.76	0.02	13.1

élevé pour les pluies modérées et fortes, c'est-à-dire l'autoencodeur avec une suite de filtres **(16,32,64,128)**, un premier filtre de taille $\mathbf{16 \times 16}$ et d'index $\mathbf{n = 1}$.

6.4.5 Espace latent à 40 dimensions

Comme pour les deux autres tailles d'espace latent, les différentes expériences pour l'espace latent à 40 dimensions sont présentées (tableau 6.8). Les scores sont en nette progression avec un TD pour les pluies totales proche ou égal à 1. On note cependant l'apparition plus systématique de fausses alarmes pour les pluies totales ce qui n'était pas le cas précédemment. Les erreurs de position sont notoirement réduites. Concernant les écarts entre les différents entraînements, ils peuvent être localement importants mais difficile de tirer des conclusions concernant une tendance par rapport à l'augmentation du nombre de filtres ou la taille du premier filtre. La difficulté à observer des tendances claires à cause de ces variations importantes d'une expérience à l'autre pourrait venir des sélections précédentes. La taille du batch et le jeu de données en entrée ont en effet été optimisés pour une architecture avec 20 dimensions. Il faudrait à l'avenir refaire ces sélections pour des espaces latents avec 10 ou 40 dimensions pour le vérifier. Pour la suite de ce manuscrit, le réseau retenu est toujours celui possédant le TD le plus élevé pour les pluies modérées et fortes, c'est-à-dire l'autoencodeur avec une suite de filtres **(16,32,64,128)**, un premier filtre de taille $\mathbf{16 \times 16}$ et d'index $\mathbf{n = 3}$. La configuration avec une suite de filtres (32,64,128,256), un premier filtre de taille 8×8 et d'index $n = 1$ possédait aussi des performances similaires mais elle n'est

TABLE 6.7 – *Tableau similaire au tableau 6.6 pour la sélection des hyperparamètres concernant l'architecture de l'autoencodeur pour un espace latent à 20 dimensions. L'autoencodeur sélectionné apparaît en gras.*

$N_{filtres}\|S_{filtre}$	n	Pluie totale			Pluie modérée			Pluie forte		
		TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$
(16...128)|16	**1**	**0.95**	**0**	**37.7**	**0.93**	**0.04**	**15.8**	**0.9**	**0.05**	**11.1**
	2	0.95	0	38.4	0.92	0.03	16.0	0.88	0.04	11.1
	3	0.95	0	37.6	0.9	0.02	15.8	0.88	0.05	11.1
(16...128)|8	1	0.97	0.01	38.9	0.92	0.03	15.8	0.89	0.05	12.2
	2	0.96	0	37.1	0.91	0.02	15.2	0.88	0.04	10.8
	3	0.95	0	37.2	0.85	0.01	14.4	0.78	0.01	10.1
(32...256)|16	1	0.95	0	38.2	0.88	0.01	15.2	0.83	0.02	10.8
	2	0.96	0	37.6	0.89	0.02	15.8	0.85	0.02	10.8
	3	0.94	0	37.8	0.88	0.01	15.5	0.83	0.02	11.4
(32...256)|8	1	0.95	0	38.2	0.88	0.01	15.8	0.85	0.03	12.1
	2	0.95	0	37.6	0.88	0.01	15.1	0.85	0.03	11.7
	3	0.95	0	38.7	0.87	0.01	15.0	0.82	0.02	10.3

pas retenue à cause de quelques points de moins en TD.

A l'issue de ces étapes, un autoencodeur pour chaque taille d'espace latent a été retenu. Seulement ces trois autoencodeurs seront étudiés plus minutieusement dans la suite de ce manuscrit. Des exemples d'images reconstruites pour la zone NO (Figure B.1) et la zone SE (Figure B.2) sont disponibles en annexe pour ces trois autoencodeurs.

6.4.6 Scores région par région

Les scores ont été précédemment calculés en prenant en compte les 5 zones régionales. Comme évoqué dans la section 6.1.1, il faut s'assurer a posteriori que les performances des autoencodeurs sont semblables sur les 5 zones prises indépendamment pour vérifier que la compression d'information est efficace partout. Le détail zone par zone pour les 3 autoencodeurs sélectionnés est présenté tableau 6.9. Ce tableau ne montre pas de différence significative de performance entre les différentes régions. On peut cependant noter que la zone SE présente pour les espaces à 20 et 40 dimensions l'erreur de position moyenne la plus faible (2-3 km de moins). Mais il faut souligner que entre les régions, les tailles des objets ne sont pas les mêmes. En moyenne, 6185 points de grille pour les objets de pluies totales sur la zone SE qui possède les objets les plus petits contre 8253 points de grille pour la zone NE qui possède les plus gros objets. La zone NE possède d'ailleurs les plus grosses erreurs de position moyenne. Ces différences sur les erreurs de position ne doivent donc pas être surinterprétées. On peut également noter que les pluies fortes sur la zone NO sont moins bien reconstruites que sur les autres zones pour l'espace à 10 dimensions avec un TD plus bas et des erreurs de position

TABLE 6.8 – *Tableau similaire au tableau 6.6 pour la sélection des hyperparamètres concernant l'architecture de l'autoencodeur pour un espace latent à 40 dimensions. L'autoencodeur sélectionné apparaît en gras.*

$N_{filtres}\|S_{filtre}$	n	Pluie totale			Pluie modérée			Pluie forte		
		TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$
(16...128)|16	1	0.99	0.01	29.7	0.92	0.01	13.3	0.89	0.02	6.8
	2	1	0.02	30.7	0.93	0.02	13.5	0.9	0.03	7.5
	3	**1**	**0.02**	**30.2**	**0.96**	**0.04**	**16.3**	**0.95**	**0.06**	**9.0**
(16...128)|8	1	1	0.02	29.8	0.92	0.02	13.7	0.89	0.03	7.6
	2	0.99	0.02	28.9	0.92	0.02	13.2	0.88	0.02	7.6
	3	0.99	0.01	29.1	0.93	0.03	14.2	0.89	0.03	7.8
(32...256)|16	1	1	0.02	31.0	0.93	0.02	13.6	0.89	0.02	7.6
	2	1	0.02	30.9	0.95	0.03	15.5	0.92	0.04	8.8
	3	0.99	0.01	30.2	0.94	0.03	14.8	0.92	0.04	8.1
(32...256)|8	1	0.99	0.01	30.2	0.94	0.02	15.0	0.91	0.04	8.2
	2	1	0.02	30.6	0.95	0.03	15.6	0.93	0.04	8.8
	3	1	0.02	30.5	0.95	0.03	15.8	0.92	0.04	8.8

plus grandes mais ces scores ne peuvent pas être expliqués par des tailles d'objets différentes. On peut donc retenir que les autoencodeurs à 20 ou 40 dimensions semblent avoir des comportements similaires sur les 5 zones. C'est également globalement vrai pour l'autoencodeur à 10 dimensions avec cependant un bémol pour la zone NO.

6.5 Comparaison avec des méthodes plus classiques

La réduction de dimension par autoencodeurs va être comparée dans cette partie avec deux approches plus classiques que sont l'Analyse en Composantes Principales (Hotelling, 1933) et les ondelettes (Morlet et al., 1982). Il n'est en effet a priori pas nécessaire d'utiliser des autoencodeurs si des approches plus classiques en météorologie proposent des performances similaires sur ce jeu de données.

6.5.1 Principes généraux de l'ACP et des ondelettes

ACP

L'ACP peut se définir comme un changement de base qui permet de transiter de l'espace de départ à un nouvel espace qui optimise au mieux l'explication de la variance du jeu de données. Le repère de ce nouvel espace se construit pas à pas avec un premier vecteur qu'on appelle la première composante principale tel que la variance des données projetées dans cette direction soit maximale. Puis un

TABLE 6.9 – *Détail des scores région par région (figure 3.13) pour les 3 autoencodeurs sélectionnés précédemment.*

Espace latent	Region	Pluie totale			Pluie modérée			Pluie forte		
		TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$
10	SE	0.89	0	50.2	0.88	0.03	16.6	0.82	0.06	11.9
	NO	0.91	0	52.0	0.84	0.04	19.4	0.77	0.06	16.1
	SO	0.9	0	49.3	0.86	0.04	16.9	0.8	0.06	11.3
	NE	0.92	0	53.5	0.83	0.03	18.8	0.81	0.06	13.4
	C	0.91	0	54.5	0.85	0.03	17.8	0.8	0.04	12.2
20	SE	0.95	0	34.4	0.95	0.04	15.1	0.9	0.06	11.2
	NO	0.95	0	38.1	0.91	0.04	16.6	0.88	0.05	10.7
	SO	0.95	0	36.5	0.93	0.04	15.5	0.9	0.06	10.6
	NE	0.96	0	40.2	0.91	0.03	16.4	0.89	0.04	12.2
	C	0.95	0	39.2	0.92	0.03	15.7	0.91	0.04	10.7
40	SE	1	0.02	26.7	0.97	0.04	16.2	0.96	0.06	8.0
	NO	1	0.02	30.4	0.96	0.04	16.0	0.95	0.06	9.3
	SO	1	0.02	29.7	0.96	0.04	14.7	0.95	0.07	10.2
	NE	1	0.02	32.7	0.95	0.03	18.1	0.95	0.05	9.1
	C	1	0.02	31.4	0.96	0.04	16.0	0.96	0.07	9.3

deuxième vecteur orthogonal au premier est construit pour expliquer au mieux la variance résiduelle (celle qui n'est pas expliquée par la première composante). On forme ainsi la deuxième composante. Les vecteurs suivants pour définir le repère de ce nouvel espace sont construits sur le même principe (orthogonaux aux vecteurs précédents et permettent d'expliquer au mieux la variance résiduelle). Conserver seulement les premières composantes permet ainsi de faire de la réduction de dimension tout en préservant au mieux la variance du jeu de données. L'ACP est une méthode courante en météorologie pour faire de la réduction de dimension (Anwender et al., 2008; Ferranti and Corti, 2011; Kumpf et al., 2018). Une ACP est ici faite sur le même jeu de données qui a servi à entraîner l'autoencodeur à l'aide du package Python Scikit-Learn (Pedregosa et al., 2011). Les champs reconstruits avec des autoencodeurs à 10, 20 et 40 dimensions seront comparés respectivement aux champs reconstruits en conservant seulement les 10, 20 ou 40 premières composantes principales. Cette comparaison se fera sur la base de validation de la section précédente (année 2021) à l 'aide des mêmes scores objets.

Ondelettes

La décomposition en ondelettes repose sur une approche très différente. Le formalisme est proche de celui des décompositions de Fourier. Cette décomposition ne se fait cependant pas sur une base sinusoïdale mais sur une famille de fonctions construite par translation et dilatation d'une fonction de base, appelée « ondelette mère ». Il existe des ondelettes mère qui permettent ainsi de construire une base orthogonale. Cette décomposition permet de passer de l'espace physique à

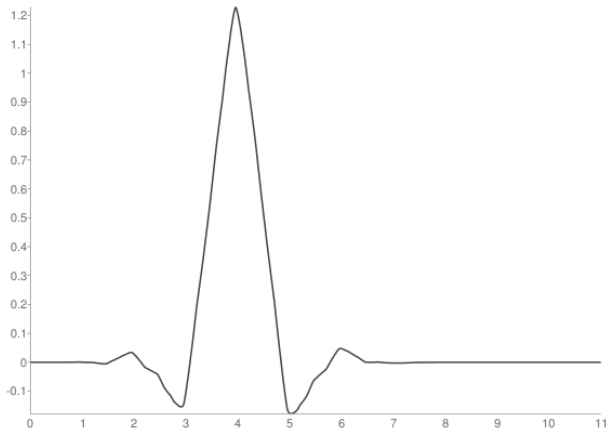


FIGURE 6.6 – *Représentation de l'ondelette mère coiflet à deux moments nuls en 1D (Source : `http://wavelets.pybytes.com/wavelet/coif2/`).*

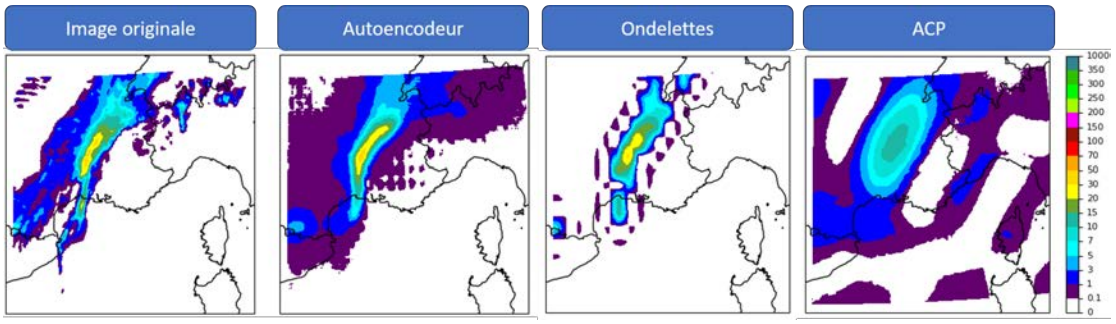


FIGURE 6.7 – *Comparaison des champs reconstruits après compression de l'information dans 40 dimensions à l 'aide d'un autoencodeur, d'une approche par ondelettes ou d'une ACP. L'ondelette mère utilisée ici est coiflet 2.*

un ensemble de coefficients dont on peut conserver seulement les plus grandes valeurs pour faire également de la réduction de dimension. A partir des coefficients restants, il est alors possible de reconstruire le signal ou champ d'origine avec une perte d'information plus ou moins importante selon le nombre de coefficients conservé. La décomposition d'un signal à l'aide d'ondelettes est utilisée en météorologie pour faire de la compression d'information (Yano et al., 2004) mais aussi pour extraire des structures particulières d'un champ météorologique que ce soit pour des tempêtes (Plu et al., 2008) ou pour l'étude de la convection (Klein et al., 2018). On constate que la décomposition du signal ou du champ est plus ou moins performante selon l'adéquation entre les données et l'ondelette mère choisie. Cette approche par ondelettes est testée à l'aide du package Python PyWavelets (Lee et al., 2019) toujours sur la même base de données (année 2021) et avec les mêmes scores objets. Plusieurs ondelettes mères ont été testées sur ce jeu de données. La meilleure ondelette et la seule dont les résultats seront présentés est l'ondelette coiflet [6] à deux moments nuls qui est représentée sur la figure 6.6.

6.5.2 Comparaison des méthodes

Une comparaison subjective des images reconstruites (Figure 6.7) semble en faveur de l'autoencodeur qui arrive mieux à reconstruire à la fois la zone de fortes pluies mais aussi l'enveloppe général des Pyrénées et de la Vallée du Rhône vers les Alpes et le Jura. Les ondelettes proposent une reconstruction de la zone de fortes pluies qui est correcte mais au détriment de l'enveloppe générale. L'ACP quant à elle lisse beaucoup trop le signal et propose des zones de pluies très éloignées de l'image originale vers la Corse par exemple. Cette comparaison est objectivisée à l'aide du tableau 6.10. L'ACP propose des performances bien en deçà des deux autres méthodes avec un TD pour les fortes pluies de seulement 0.16 avec 40 dimensions contre 0.95 et 0.89 pour l'autoencodeur et les ondelettes. L'ACP n'est a priori pas la bonne approche pour traiter les champs de pluies car ceux-ci ont une distribution qui n'est absolument pas gaussienne avec des gradients très importants (passage rapide de 0 à plus de 50 mm en quelques points de grille possible). Les transformations non linéaires permises par l'autoencodeur et la recherche de gradients particuliers ou de formes particulières dans les couches de convolution rendent l'autoencodeur plus performant que l'ACP qui est plus souvent utilisée sur des champs lisses comme le géopotentiel à 500 hPa. Les ondelettes ont des scores intéressants sur les pluies modérées et fortes avec seulement quelques points de TD en moins par rapport à l'autoencodeur et des erreurs de position très inférieures (seulement 3 km pour les pluies fortes avec 40 dimensions). Mais ces performances se font au détriment des pluies totales avec des erreurs de position très importantes. Un des problèmes des ondelettes réside dans son indépendance vis-à-vis du jeu de données. En effet cette décomposition en ondelettes se fait directement sur le jeu de données mais elle n'apprend rien au préalable des caractéristiques de ce jeu contrairement à l'autoencodeur ou à l'ACP. En particulier l'ondelette mère n'est pas particulièrement adaptée ou créée spécifiquement pour ce jeu de données. On peut conclure de cette partie que dans le cas général, l'autoencodeur est une approche plus satisfaisante que les ondelettes ou l'ACP pour compresser les cumuls de pluies sur 1h des membres PE-AROME.

Ce chapitre a présenté une méthode pour faire de la réduction de dimension sur les cumuls de pluies sur 1h. Cette étape est la première pour définir les scénarios et permet de s'extraire de l'espace physique pour se plonger dans un espace latent beaucoup plus petit à 10, 20 ou 40 dimensions. Après avoir démontré que l'approche par autoencodeur était plus performante que les ondelettes ou l'ACP, la suite de ces travaux sur les scénarios sera présentée dans la partie 4 avec un premier chapitre qui sera dédié à une étude plus approfondie de ces espaces latents qui sont à ce stade assez abstraits.

6. associée à une fermeture symétrique sur les bords du champs de pluies.

7. Le TD n'est pas rigoureusement nul (mais inférieur à 0.005). C'est pourquoi une erreur de position peut être définie.

TABLE 6.10 – *Comparaison des scores objets après une compression de l'information par autoencodeur, ondelettes ou ACP. Cette comparaison est faite pour des espaces à 10,20 et 40 dimensions*

Espace latent	Méthode	Pluie totale			Pluie modérée			Pluie forte		
		TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$	TD	TFA	$\overline{dist}$
10	Autoencodeur	0.91	0	51.9	0.85	0.03	17.8	0.8	0.05	12.7
	Ondelettes	0.86	0	87.6	0.62	0	20.9	0.67	0	8.4
	ACP	0.74	0	84.1	0.09	0	38.9	0	0	46.6[7]
20	Autoencodeur	0.95	0	37.7	0.93	0.04	15.8	0.9	0.05	11.1
	Ondelettes	0.93	0	83.5	0.77	0	16.4	0.81	0	5.5
	ACP	0.83	0	74.0	0.47	0	26.6	0.04	0	23.3
40	Autoencodeur	1	0.02	30.2	0.96	0.04	16.3	0.95	0.06	9.0
	Ondelettes	0.97	0	75.2	0.88	0	10.7	0.89	0	3.1
	ACP	0.91	0	53.1	0.47	0	16.8	0.16	0	15.6

Bibliographie

M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow : Large-Scale Machine Learning on Heterogeneous Systems, 2015. URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

D. Anwender, P. A. Harr, and S. C. Jones. Predictability Associated with the Downstream Impacts of the Extratropical Transition of Tropical Cyclones : Case Studies. *Mon. Wea. Rev.*, 136(9) :3226–3247, 2008. doi : 10.1175/2008MWR2249.1.

C. Davis, B. Brown, and R. Bullock. Object-based verification of precipitation forecasts. part i : Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, 134 :1772–1784, 2006. doi : 10.1175/MWR3145.1.

L. Ferranti and S. Corti. New Clustering Products. *ECMWF Newsletter*, (127) : 6–11, 2011.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

X. Guo, X. Liu, E. Zhu, and J. Yin. Deep clustering with convolutional autoencoders. In *Neural Information Processing*, pages 373–382, 2017.

H. Hotelling. Analysis of a complex of statistical variables into principal components, 1933.

N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang. On Large-Batch Training for Deep Learning : Generalization Gap and Sharp Minima, 2016.

C. Klein, D. Belušić, and C. M. Taylor. Wavelet Scale Analysis of Mesoscale Convective Systems for Detecting Deep Convection from Infrared Imagery. *J. Geophys. Res.*, 123(6) :3035–3050, 2018. doi : https://doi.org/10.1002/2017JD027432.

A. Kumpf, B. Tost, M. Baumgart, M. Riemer, R. Westermann, and M. Rautenhaus. Visualizing confidence in cluster-based ensemble weather forecast analyses. *IEEE Transactions on Visualization and Computer Graphics*, 24(1) : 109–119, 2018. doi : 10.1109/TVCG.2017.2745178.

G. R. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary. PyWavelets : a Python package for wavelet analysis. *Journal of Open Source Software*, 4(36) :1237, 2019. doi : 10.21105/joss.01237.

J. Morlet, G. Arens, E. Fourgeau, and D. Glard. Wave propagation and sampling theory—Part I : Complex signal and scattering in multilayered media. *GEOPHYSICS*, 47(2) :203–221, 1982. doi : 10.1190/1.1441328.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn : Machine Learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011.

M. Plu, P. Arbogast, and A. Joly. A Wavelet Representation of Synoptic-Scale Coherent Structures. *J. Atmos. Sci.*, 65(10) :3116–3138, 2008. doi : 10.1175/2008JAS2618.1.

S. Ravuri, K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge, R. Prudden, A. Mandhane, A. Clark, A. Brock, K. Simonyan, R. Hadsell, N. Robinson, E. Clancy, A. Arribas, and S. Mohamed. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878) :672–677, 2021. doi : 10.1038/s41586-021-03854-z.

L. Rottner, P. Arbogast, M. Destouches, Y. Hamidi, and L. Raynaud. The similarity-based method : a new object detection method for deterministic and ensemble weather forecasts. *Advances in Science and Research*, 16 :209–213, 2019. doi : 10.5194/asr-16-209-2019.

Y. Sha, D. J. Gagne II, G. West, and R. Stull. Deep-Learning-Based Gridded Downscaling of Surface Meteorological Variables in Complex Terrain. Part II : Daily Precipitation. *J. Appl. Meteor. Climatol.*, 59(12) :2075–2092, 2020. doi : 10.1175/JAMC-D-20-0058.1.

J.-I. Yano, P. Bechtold, J.-L. Redelsperger, and F. Guichard. Wavelet-compressed representation of deep moist convection. *Mon. Wea. Rev.*, 132(6) :1472–1486, 2004. doi : 10.1175/1520-0493(2004)132<1472:WRODMC>2.0.CO;2.

Troisième partie

Prévision du risque d'échos arqués

Chapitre 7

Synthèse des détections dans PE-AROME et AROME

Cette partie dédiée à la détection des échos arqués dans les modèles PE-AROME et AROME est dans la continuité des chapitres 3 et 5. Le chapitre 3 a permis d'expliquer le schéma conceptuel de l'écho arqué et l'intérêt de chercher à détecter ce type de structure dans les prévisions. Le chapitre 5 a montré comment une base de données a été construite pour permettre à un réseau de neurones U-Net d'apprendre à détecter les échos arqués. Après avoir présenté l'architecture du U-Net et les scores qui seront utilisés pour évaluer ses performances dans ce même chapitre 5, ce chapitre va s'intéresser aux différentes étapes allant du choix des hyperparamètres du U-Net jusqu'à la construction de représentations graphiques pour résumer le risque d'échos arqués dans la PE-AROME. Toutes ces étapes sont décrites dans un article publié dans la revue américaine Artificial Intelligence for the Earth Systems (AIES). L'article est disponible à l'adresse suivante : `https://journals.ametsoc.org/view/journals/aies/1/2/AIES-D-21-0010.1.xml`. L'ensemble des figures ont été faites par mes soins ainsi que la totalité de la rédaction, aidé par les corrections et conseils de mes co-auteurs. Le contenu des sections 1 et 2 de ce papier a déjà été présenté dans les chapitres précédents de ce manuscrit. Ce chapitre traite donc des éléments présentés à partir de la section 3 jusqu'à la fin de cet article.

7.1 Résumé de l'article

L'article intitulé *Detection of bow echoes in kilometer-scale forecasts using a convolutional neural network* montre comment un réseau de neurones a été entraîné à détecter des échos arqués dans les membres PE-AROME puis utilisé sur les sorties du modèle déterministe AROME après transfert d'apprentissage pour construire une synthèse globale du risque d'échos arqués. Après une introduction rappelant ce qu'est un écho arqué et les limites des méthodes actuelles pour retranscrire ce risque à partir des prévisions d'ensemble, la section 2 introduit les éléments déjà présentés dans le chapitre 5 : la création de bases de données à l'aide d'une labellisation manuelle de milliers de champs pour tracer les contours des échos arqués, le pré-traitement de ces données pour ne pas obtenir des classes

à apprendre trop déséquilibrées, la description de l'architecture du U-Net utilisée, l'approche objet et les scores associés pour évaluer les sorties des U-Net. Plusieurs de ces étapes nécessitent la définition d'hyperparamètres.

L'objectif de la section 3 est de comprendre comment les hyperparamètres influencent les performances des U-Nets et de sélectionner la meilleure configuration qui sera utilisée par la suite. En préambule, la section montre qu'il est nécessaire de filtrer les détections les plus petites du U-Net pour améliorer les scores, le seuil est fixé à 100 points de grille. On montre par la suite que la taille de l'imagette joue un rôle prépondérant sur les scores tout comme le poids de la classe écho arqué dans la fonction coût ainsi que le ratio entre imagettes avec ou sans écho arqué. Parmi les 405 U-Nets entraînés avec différentes combinaisons d'hyperparamètres, le réseau de neurones retenu est celui possédant à la fois un score de Jaccard très élevé et le taux de bonnes détections le plus élevé. Ce réseau de neurones a un taux de bonnes détections de 86 % et un taux de fausses alarmes de 39 %.

Ce réseau de neurones considéré comme optimal est ensuite étudié à l'aide de différents scores objets dans la section 4. Cette section montre que le réseau de neurones tend à sous-estimer la taille des échos arqués. Les paramètres liés à l'intensité des pluies et des rafales à l'intérieur de l'écho arqué sont eux correctement détectés. Cette analyse tend à montrer que le réseau de neurones détecte correctement la partie la plus active du système orageux mais que les zones plus éloignées du cœur convectif sont elles moins bien détectées. Cette partie a permis également de montrer que les fausses alarmes sont souvent des objets petits et de faibles intensités. Les caractéristiques des non détections sont plus difficiles à établir. La section 4 se conclut en vérifiant que le U-Net ne confond pas les échos arqués avec une autre organisation orageuse : les supercellules.

La section 5 s'intéresse à l'extension du U-Net sur les sorties du modèle déterministe AROME par transfert d'apprentissage. Cette section montre avec des graphiques similaires à la section 4 que les performances du U-Net sur ce modèle sont comparables à celles obtenues sur les membres PE-AROME. La section conclut qu'il est possible d'utiliser ce réseau de neurones sur les sorties de ce modèle.

Maintenant que les performances du U-Net sont connues sur les membres PE-AROME et AROME, la section 6 s'intéresse aux moyens de synthétiser toutes les détections obtenues dans ces prévisions. Plusieurs cartes sont proposées avec un code couleur selon les membres, selon les heures de détection ou encore une carte de probabilité avec une tolérance spatio-temporelle. Pour conclure cette partie, les retours subjectifs d'un groupe de 10 prévisionnistes sur les performances du U-Net et les cartes de synthèse sont rassemblés. La méthodologie qui décrit comment les prévisionnistes ont participé à ces travaux est disponible dans le supplément de ce papier (section 7.3). Ces retours confirment les scores objectifs des sections précédentes. La dernière section de l'article identifie plusieurs perspectives comme l'ajout de la variable rafale de vent comme prédicteur, la détection des échos arqués dans les données observées ou encore l'extension de ces détections à d'autres objets convectifs comme les supercellules. Les cartes de synthèse du risque d'échos arqués sont produites en temps réel en mode recherche depuis mai 2021. Les temps de calcul restent suffisamment courts (quelques dizaines de minutes) pour avoir un intérêt en condition opérationnelle.

7.2 Article

Generated using the official AMS LaTeX template v5.0 two-column layout. This work has been submitted for publication.

Detection of bow echoes in kilometer-scale forecasts using a convolutional neural network

Arnaud MOUNIER *, Laure RAYNAUD, Lucie ROTTNER, Matthieu PLU

CNRM, University of Toulouse, Météo-France, CNRS, Toulouse, France

Philippe ARBOGAST

Forecasting department, Météo-France, Toulouse, France

Michaël KREITZ

National meteorological school, Météo-France, Toulouse, France

Léo MIGNAN and Benoît TOUZÉ

Forecasting department, Météo-France, Rennes and Toulouse, France

ABSTRACT

Bow echoes (BEs) are bow-shaped lines of convective cells that are often associated with swaths of damaging straight-line winds and small tornadoes. This paper describes a convolutional neural network (CNN) able to detect BEs directly from French kilometer-scale model outputs in order to facilitate and accelerate the operational forecasting of BEs. The detections are only based on the maximum pseudo-reflectivity field predictor (pseudo because expressed in $mm \cdot h^{-1}$ and not in dBZ). A pre-processing of the training database is carried out in order to reduce imbalance issues between the two classes (inside or outside bow echoes). A CNN sensitivity analysis against a set of hyperparameters is done. The selected CNN configuration has a hit rate of 86% and a false alarm rate of 39%. The strengths and weaknesses of this CNN are then emphasized with an object-oriented evaluation. The BE largest pseudo-reflectivities are correctly detected by the CNN which tends to underestimate the size of BEs. Detected BE objects have wind gusts similar to the hand labeled BE. Most of the time, false alarm objects and missed objects are rather small (e.g., < 1500 km^2). Based on a cooperation with forecasters, synthesis plots are proposed that summarize the BE detections in French kilometer-scale models. A subjective evaluation of the CNN performances is also reported. The overall positive feedback from forecasters is in good agreement with the object-oriented evaluation. Forecasters perceive these products as relevant and potentially useful to handle the large amount of available data from numerical weather prediction models.

1. Introduction

Mesoscale convective systems (MCSs, Houze Jr. 2004) are organized thunderstorms with a linear or round shape lasting several hours, whereas single-cell thunderstorms last 20-30 minutes on average. MCSs can cause considerable wind, hail or flood damage. In an operational context, particular care must be taken to closely monitor the different types of MCSs such as squall lines (Trapp et al. 2005) or mesoscale convective complexes (Laing and Fritsch 1997). Among MCSs, bow echoes (BEs) can occur all year round in North America as well as in Europe (Goulet 2015). Even though they are relatively rare in Europe (around 5 BEs per year in France between 2006 and 2020), BEs can have significant meteorological impact because they can generate strong wind gusts and tornadoes. The size and genesis of BEs can differ from one case to another but a general common feature is a bow shape in reflectivity fields. This aspect may be explained by a rear inflow jet in the mid levels of the atmosphere (Przybylinski 1995; French and Parker 2014). This mid-level jet causes a strong downdraft and it is also responsible for strong winds under bow echoes (Fujita 1978; Atkins and Laurent 2009; Markowski and Richardson 2010).

Since the beginning of the 21st century, Numerical Weather Prediction (NWP) models at convection-permitting scale have been developed (Done et al. 2004; Seity et al. 2011). These models were some of the first ones to explicitly simulate MCSs. Due to the limited predictability at such scales (Hohenegger and Schär 2007), convection-permitting ensemble prediction systems (EPSs) are necessary. Systems such as

* *Corresponding author*: Arnaud MOUNIER, arnaud.mounier@meteo.fr

SREF[1] (Du et al. 2003), MOGREPS[2] (Bowler et al. 2008), COSMO-DE-EPS[3] (Peralta et al. 2012) or AROME-EPS[4] (Bouttier et al. 2016) have been developed by several national weather services to supplement deterministic forecasts.

However, leveraging the huge information provided by EPSs to forecast occurrences and trajectories of MCSs, or more generally, of meteorological objects, still remains a challenge. A visual examination of each member is time consuming, and usual products such as point-based probabilities or percentiles are not appropriate to recognize the different MCSs simulated in EPS members. A possible way to overcome these limitations is to automatically detect MCSs in NWP outputs and to develop probabilistic diagnostics from detected objects. For operational forecasting, Updraft Helicity (UH, Kain et al. 2008) has been developed to detect potential severe convective storms such as supercells (Moller et al. 1994). UH has been used to define severe weather indices (Sobash et al. 2016; Gallo et al. 2019; Sobash et al. 2020) and can be combined with reflectivity fields to recognize severe thunderstorms in a "member viewer" approach (Roberts et al. 2019). MCS detection algorithms have been also used on observed data (Patil et al. 2019). Concerning BEs, an automated detection based on computer vision with skeletonization and shape matching approaches has been developed by Kamani et al. (2016). Existing detection algorithms mostly rely on a threshold of predictor fields that dictates BE identification in model outputs. Such approaches require fine tuning and they are rarely designed to recognize specific shapes, which is a key aspect for the detection of objects such as BEs.

In meteorology, machine learning (ML) and deep learning (DL) methods have recently proved their ability to detect patterns and objects in observational and modeling datasets. One of the first DL method applications was to solve classification problems with the aim of predicting a label (e.g., sunny or cloudy) given an image from a large and varied dataset (Elhoseiny et al. 2015). Liu et al. (2016) were among the first to detect features in NWP outputs using Convolutional Neural Networks (CNNs, LeCun and Bengio 1995) in order to track down tropical cyclones, atmospheric rivers and fronts using classification systems. Then, ML and DL methods have been used for segmentation problems in order to detect object contours in forecast outputs. Segmentation algorithms have been applied to extend the previous work on tropical cyclones and atmospheric rivers (Kurth et al. 2018) and to detect fronts (Biard and Kunkel 2019; Matsuoka et al. 2019; Lagerquist et al. 2019).

Regarding the applications to convection, McGovern et al. (2017) and Gagne II et al. (2019) have shown that CNNs can discriminate severe hailstorms according to spatial structures of storms. These CNNs can differentiate a BE from a supercell and a pulse storm to quantify the risk of hail. Using radar data and radio-soundings, Jergensen et al. (2020) applied ML methods to classify convective storms in three categories, viz disorganised, quasi-linear convective system (QLCS) or supercell. ML and DL have also been used for short-term predictions of strong convective wind gusts (Lagerquist et al. 2017) or tornado occurrences (Lagerquist et al. 2020).

In previously mentioned works, CNN input data were manually pre-selected or obtained from heuristic algorithms for training and prediction processes, and ML/DL methods were used to classify these sets of inputs. In this work, we present a segmentation model that can directly detect and identify BEs in deterministic and ensemble model outputs to provide a comprehensive daily production of BE risk. The purpose of the present study is to train a CNN-based detection algorithm for BEs with a database consisting of forecasts from the French kilometer-scale ensemble prediction system, hereafter called AROME-EPS. A transfer learning approach is tested to apply the CNN to the French deterministic AROME model. This CNN is first assessed with object-oriented scores. The advantages of this CNN-based detection for operational purposes are evaluated by ten forecasters. To assess the practical benefit of CNN detections, methods for synthesizing information are also discussed with forecasters, following the suggestions by Demuth et al. (2020).

The CNN architecture, AROME-EPS databases, and scores used to evaluate CNN skills are detailed in section 2. In section 3, a hyperparameter search is presented, highlighting the roles of some important parameters in the CNN setting and in the training dataset design. The optimal CNN configuration derived from section 3 is further detailed and evaluated in section 4 to highlight its strengths and weaknesses. Section 5 discusses the application of the CNN to the deterministic AROME model without re-training. In a dedicated end-user section 6, different visualization products are proposed to summarize the BE risk in AROME-EPS and AROME models. In the last part of section 6 forecasters' feedback on CNN skills and

[1]Short Range Ensemble Forecast

[2]Met Office Global & Regional Ensemble Prediction System

[3]COnsortium for Small scale MOdelling-DEutschland-Ensemble Prediction System

[4]Application of Research to Operations at MEsoscale-Ensemble Prediction System

advantages in an operational context are gathered. Conclusions and perspectives are given in section 7.

2. Methods and Data

a. AROME and AROME-EPS

AROME is the non-hydrostatic high resolution model of Météo-France, operational since 2008 (Seity et al. 2011; Brousseau et al. 2016). The model covers mostly Western Europe (12°W-16°E and 37.5°N-55.4°N, with a size of 2000×2000km approximately, cf Fig.S1 in the online supplement for a figure of the domain). The current operational AROME model has a horizontal grid spacing of 1.3 km and 90 vertical levels, some statistics concerning the simulated convective cells in AROME are available in the section 3 of Brousseau et al. (2016). AROME is initialized five times a day at 00, 03, 06, 12 and 1800UTC with lead times up to 48h. Operational at Météo-France since 2016, AROME-EPS is the convection-permitting ensemble prediction system based on the non-hydrostatic AROME model. Its domain is similar to the one from the deterministic AROME model. AROME-EPS is a 16-member ensemble (since summer 2019) with a horizontal grid spacing of 2.5 km and 90 vertical levels. It is perturbed with four different sources of uncertainties : lateral boundary conditions (Bouttier and Raynaud 2018), surface conditions (Bouttier et al. 2016), initial conditions (Montmerle et al. 2018; Raynaud and Bouttier 2017) and model errors (Bouttier et al. 2012). AROME-EPS is initialized four times a day at 03, 09, 15 and 2100UTC with lead times up to 51h. AROME-EPS has been developed to improve the prediction of high-impact phenomena such as convective systems.

b. Input data

1) Bow Echo (BE) labeling

The training and validation datasets for the CNN segmentation model are built using pseudo-reflectivity forecasts from AROME-EPS members. The reflectivity field available in AROME-EPS members outputs is calculated with a radar simulator (Caumont et al. 2006) in $mm^6 \cdot m^{-3}$. The Marshall-Palmer Z-R relationship (Marshall and Palmer 1948) is applied to convert the reflectivity fields into pseudo-reflectivity fields in $mm \cdot h^{-1}$. The choice of $mm \cdot h^{-1}$ rather dBZ is a historic choice when the rainfall accumulation was updated hourly. Even if the fields in $mm \cdot h^{-1}$ and dBZ are now produced in operations, only pseudo-reflectivity fields are available for the oldest dates of the training and validation databases. The maximum value in the grid column of pseudo-reflectivities is used as an input for the CNNs. To train the CNNs and compute the validation database scores, a corresponding ground-truth hand-labeled dataset is produced. This dataset is constructed from the contours of hand-labeled BEs plotted by one expert and using the VIA software (VGG Image Annotator, Dutta and Zisserman 2019). After a postprocessing step, each hand-labeled field has the same size as the pseudo-reflectivity input, with a value of 1 for every grid point inside a BE and 0 outside. The labeling process utilizes three variables: maximum pseudo-reflectivity, mean sea level pressure (MSLP), and wind speed at 10m. From these three predictors, the expert uses three conditions to hand-label a BE. The first one is to observe a bow shape in the pseudo-reflectivity field. The second one is to notice a gradient in the MSLP field corresponding to the bow shape. According to Markowski and Richardson's (2010) conceptual model, a specific pressure pattern is observed with BEs, consisting of a minimum in pressure leading a BE, a maximum in pressure beneath a BE, and another relative pressure minimum in the trailing stratiform rain region. The last condition is to observe the bow shape in the wind speed field and especially the strong increase of wind speed in front of the BE.

In practice, the labeling process is highly time-consuming. In order to facilitate the analysis of the predictor fields, only two figures are provided to the expert (Fig. 1), showing the maximum pseudo-reflectivity field, and the contours of $\|\nabla(MSLP)\|_1$ equal to 2 hPa for 10km overlaid with the areas where the divergence of horizontal wind is below $1.5 \times 10^{-3} s^{-1}$. The contours of hand-labeled BEs are plotted over the pseudo-reflectivity fields. An example of fields used for the labeling process is shown in Fig. 1 where a BE is labeled. Note that the wind gust field is not considered because it is not instantaneous, only the maximum over the last hour is computed.

2) BE datasets

The labeling process is applied to specific cases where BEs are observed (in radar data) in France. Studying every AROME-EPS run over a long period would be too time-consuming and simulated BEs can be preferentially identified among the 16 members on these specific cases. A list of these observed BEs in 2017, 2018 and 2019 is subjectively created with the help of ten forecasters to be sure that no observed BE is omitted. The previous four AROME-EPS initializations before the beginning of BE events are used and only the lead times around BE events are retained to save time during the labeling process. Some other severe convective cases (supercells,

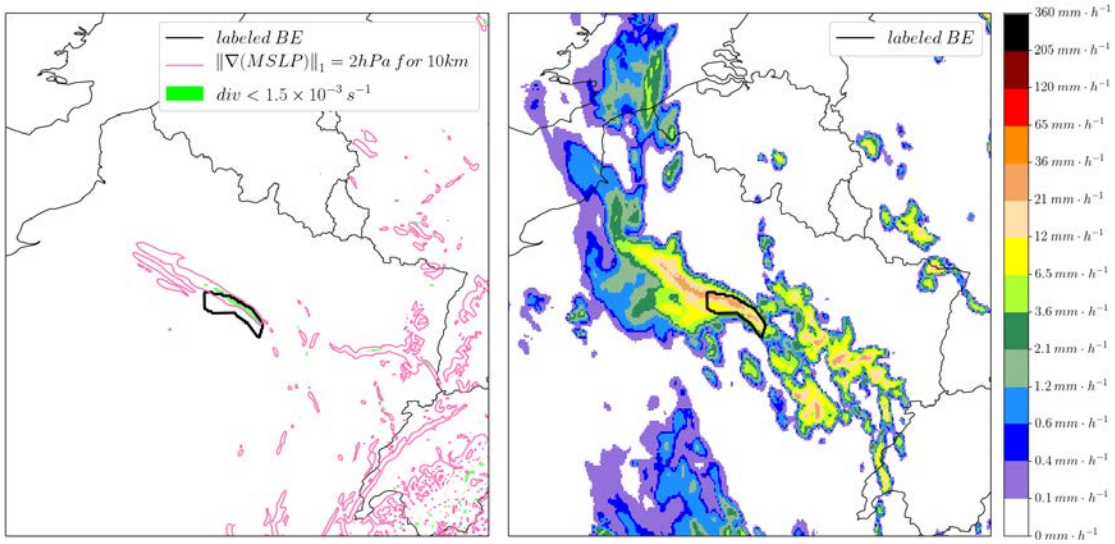


FIG. 1. Example of fields used during the labeling process. The fields are zoomed over the North-East of France. On the left, the contours of $\|\nabla(MSLP)\|_1$ equal to 2 hPa for 10km are in pink, the areas where the divergence of horizontal wind (div in legend) is below $1.5 \times 10^{-3}\ s^{-1}$ in green, and the contour of the labeled BE in black. On the right, the corresponding pseudo-reflectivity field.

squall lines and quasi-stationary convective systems) are also integrated in the validation dataset. The ability of CNNs to learn pseudo-reflectivity signatures not associated with BEs needs to be verified. Some (~ 10) hand-labeled BEs were plotted in those other severe convective cases. Indeed, some members can simulate BEs in the AROME domain even if a quasi-stationary convective system is observed for instance. Concerning the training dataset, other organized thunderstorms (especially supercells) are indirectly integrated because they are simultaneously simulated over the same areas in similar supporting environments.

Finally, the training dataset is based on a total of 11 observed BE cases (18 May 2017, 27 June 2017, 10 June 2018, 8 August 2018, 6 July 2019, 6-9-11 August 2019, 21 September 2019, and 14-23 October 2019), leading to 6206 pseudo-reflectivity fields and 556 hand-labeled BEs. During the design of the training dataset, a half of the AROME-EPS samples corresponding to one specific BE case (21 september 2021) is randomly selected. This subsampling is necessary because all the simulated BEs in the AROME-EPS members are very similar with same pseudo-reflectivity intensities and same sizes for this case (all BE contours are available in the online supplement to this paper, figure S2). If this subsampling were not made, this BE case could create a large number of very similar training samples and consequently could degrade the training process. Concerning the validation dataset, 4 different observed BE cases (8 August 2017, 26 May 2018, 4 and 5 June 2019) and 3 other convective cases (squall lines : 2 January 2018 and 3 December 2018, quasi-stationary convective system : 5 August 2019) are integrated, leading to 2440 pseudo-reflectivity fields and 264 hand-labeled BEs. This corresponds to approximately 70% of data used for the training and 30% for the validation dataset. The training and validation databases are composed of independent cases to limit as far as possible correlation between the two databases.

3) BE PREDICTORS

The selection of CNN inputs is crucial and affects the CNN ability to correctly detect BEs. Due to limited computing power, we can only chose one predictor as a first step toward building a BE identification system. The three variables utilized during the labeling process (maximum pseudo-reflectivity, MSLP and wind speed at 10m) are examined. The MSLP field is too noisy around mountainous areas (especially the Alps and the Pyrenees) with noise similar in magnitude to BE-induced perturbations. An example around the Alps at the bottom right corner of Fig. 1 is shown with pink contours but no precipitation. The wind speed field has sometimes an unrealistic behavior in the AROME models and is rejected at this stage. This behavior is observed in convective and dry conditions or in showers at the rear of cold fronts. The mean wind speeds can be above 120 $km \cdot h^{-1}$ (wind gusts above 210 $km \cdot h^{-1}$) which are unreasonable values in France. The AROME models can also create strong downdrafts in convective

cases without associated rainfall. Even if this behavior is occasionally realistic (i.e dry microbursts), forecasters tend to use wind speed with care in convective cases (same conclusion applies to the wind gust parameter). The CNNs in this paper are univariate models with maximum pseudo-reflectivity forecasts as input for that reason. However, wind speed and especially wind gust speed at 10m (FFgust) remain crucial parameters for BEs. The main threat is strong wind gusts under BEs and therefore forecasters need to know magnitudes of these wind gusts to convey information on hazards. In this paper, the FFgusts inside BEs will be considered during the object-oriented evaluation in section 4 and 5 to evaluate the selected CNN in the section 3. Using also wind speed as another predictor could be tested in future work.

c. CNN architecture

For BE detection, CNNs are used as segmentation models. The selected CNN should be suitable for small training datasets because BEs are rare events (in space and time). A U-Net architecture (Ronneberger et al. 2015) has been chosen because past studies have shown that it gives satisfactory results when few data are available. This U-Net architecture (Fig. 2) is divided into two parts which correspond to the contracting and expanding paths. The contracting path consists of 2 convolutional layers (32 3×3 filters) followed by a rectified linear unit (ReLU) activation function and a dropout rate of 0.2. A 2×2 maxpooling operation is applied to divide the patch size by 2. The same sequence is repeated two more times with respectively 64 and 128 filters. The expanding path first performs a 2×2 upsampling ("upconvolution") operation. The upsampled maps are then concatenated with the corresponding maps in the contracting path, and 2 convolutional layers followed by a ReLU activation are applied (with the same number of filters as in the corresponding contracting path layer). Upsampling, concatenation and convolution are repeated another time. At this stage, the outputs have the same size as the inputs. A final convolutional layer with a 1x1 filter is added to obtain the desired number of classes (two classes here). In order to get a class probability as an output, a softmax function is applied. The U-Net architecture is presented here with an input size of $N \times M$ grid points. Several sizes of inputs are tested in the next section. Other classical parameters for the U-Net architecture such as the number and size of filters, the dropout rate and the choice of the ReLU activation are not discussed in the following sections.

d. CNN training

Limited computing power does not allow to take the entire AROME domain as a CNN input. To train the U-Nets, patches of $N \times M$ grid points are extracted from the original pseudo-reflectivity fields (717×1121 grid points). In addition, because the size of a BE is much smaller than the domain size, there is a strong imbalance between the BE and no BE classes in the training dataset. Extracting smaller patches allows us to design a more balanced training database by selecting the most informative patches. Fig. 3 shows the different steps undertaken to extract patches (from 1 to 6):

- (step 1) Patches from the pseudo-reflectivity fields are randomly selected. $N_{patches}$ are extracted from each field (values in Table 1). The associated groundtruth patches are also extracted.

- (step 2) The pairs of patches (pseudo-reflectivity/groundtruth) are split into two groups : the patches with BEs in which at least one grid point is labeled as a BE and the patches without any BE.

- (step 3) The number of patches with BEs is very limited compared to patches without any BE (1 patch with a BE for every 1600 patches without a BE). Initially, BEs with moderate pseudo-reflectivities were not correctly predicted because they were under-represented in the dataset. A data augmentation technique is proposed to solve this problem: in a given patch, the pseudo-reflectivities are multiplied by a coefficient of 0.75 if a BE is within this patch and the maximum pseudo-reflectivity is above a given threshold (mentioned in the next section). This new patch is added in the ones with BEs (step2). The pseudo-reflectivity field must remain physically consistent as much as possible. That is why lower coefficients are not investigated and only patches with large magnitude in pseudo-reflectivities are taken into account.

- (step 4) To reduce the number of patches without a BE, the patches without precipitation (pseudo-reflectivity maximum $< 0.1\ mm \cdot h^{-1}$) are deleted.

- (step 5) Even after the filtering procedure of step 4, the number of patches without a BE remains high (1 patch with a BE for every 400 patches without a BE). To limit the number of patches without a BE, the ratio between the patches with and without a BE (Ratio_noBE/BE) is

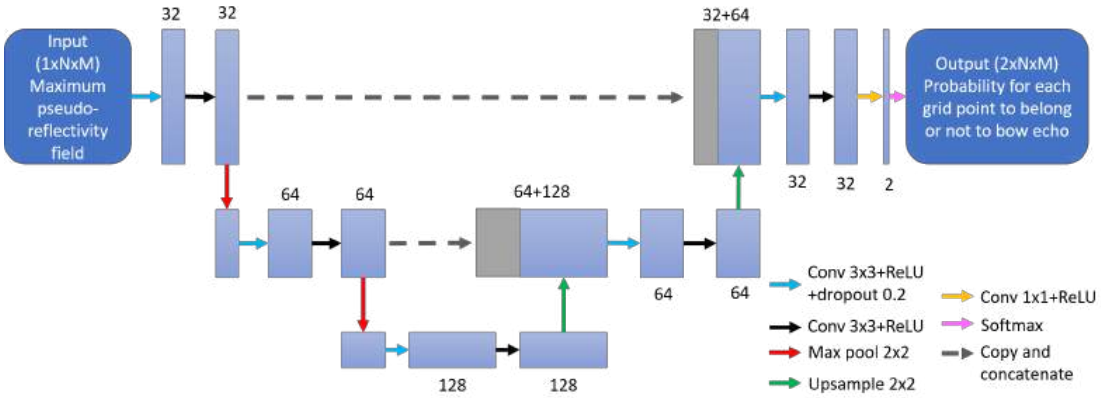


FIG. 2. U-Net architecture for bow echo detection

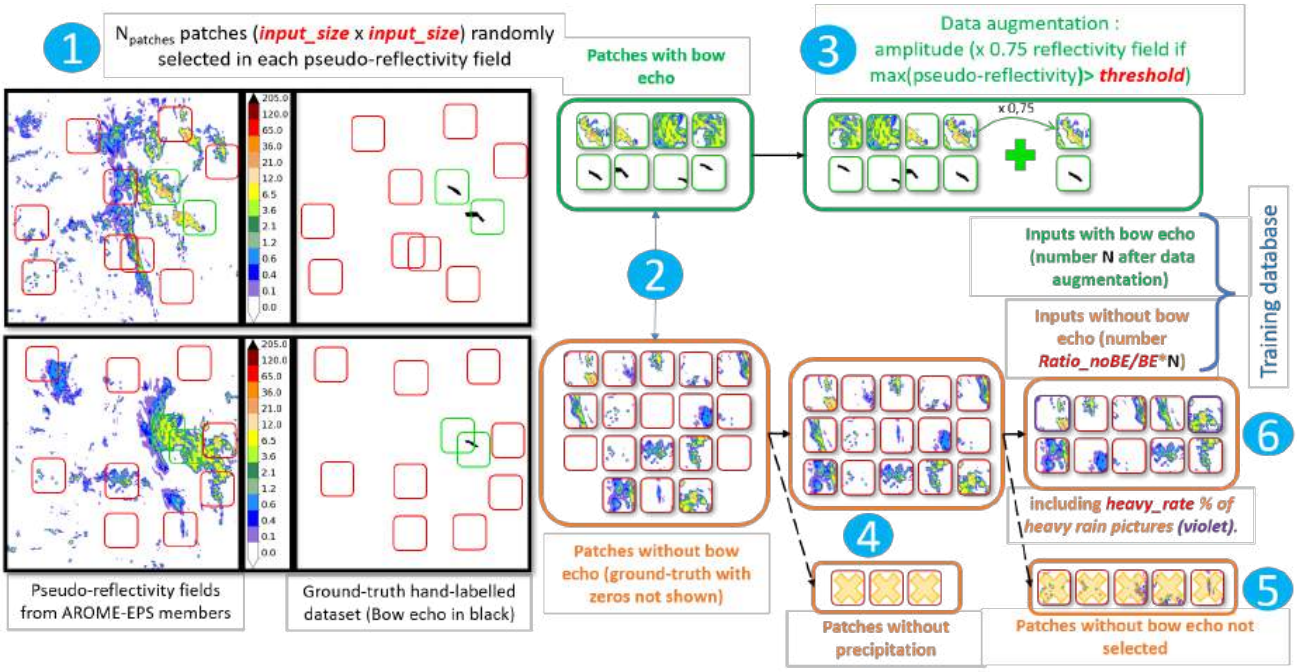


FIG. 3. Training database design. Steps 1-6 are denoted by blue circles and are described in the section 2.d.

fixed to a lower value. The patches retained are randomly selected and the unnecessary patches without a BE are deleted. This ratio is discussed in the next section.

- (step 6) During the first tests, all patterns with strong pseudo-reflectivities were detected as BEs and consequently the number of false alarms was very high in the validation database. Strong pseudo-reflectivities are rare in space and time and the majority of patches without a BE contains no or weak precipitation whereas BEs are frequently associated with heavy precipitation. Only 0.7% of patches without BE are associated with heavy precipitation (i.e above 60 $mm \cdot h^{-1}$) whereas, after the data augmentation in amplitude (step 3), around 55% of patches with BEs are associated with heavy precipitation. In this case, the pseudo-reflectivity magnitude is relied on too heavily to detect the BEs in the pseudo-reflectivity fields. Patches with large magnitude pseudo-reflectivities but without a BE are forced in the training database to solve this problem and the rate of large magnitude pseudo-reflectivity patches without a BE in the total number of patches without a BE is defined (heavy_rate). A patch is considered with large magnitude pseudo-reflectivities if the maximum is above 60 $mm \cdot h^{-1}$. Another way to address

TABLE 1. Values of N_{stride} and N_{patch} according to input size

Input size	N_{stride}	N_{patch}
$24 \cdot 24$	7	4000
$48 \cdot 48$	15	1000
$96 \cdot 96$	30	250

this problem could be to add more input predictors.

e. CNN prediction

The U-Net inputs and outputs are composed of $N \times M$ grid point patches. The aim of this study is to predict the BE risk over the entire original AROME domain. The same U-Net is applied on overlapping patches as depicted in Fig. 4. The entire grid is divided in ordered patches with a stride of N_{stride} grid points along the longitude and latitude axes (Fig. 4 step 1 : blue, black, and orange squares). The values of N_{stride} depend on the input size of the U-Net and are given in Table 1. A prediction for each patch is computed. The prediction results are patches representing the probability of each class (Fig. 4 step 2 : two patches after the U-Net architecture for each pseudo-reflectivity patch, $p_{n,i}$ is the probability of the ith grid point according to the nth patch). The mean probability for each class is computed (Fig. 4 step 3) considering all the patches where the grid point is included. Finally, the probabilities are used (>50% defines a categorical prediction) to define a detected BE (Fig. 4 step 4). Following the training, Fig. 4 depicts the U-Net application in the prediction process from the input pseudo-reflectivity field of 717×1121 grid points to the segmentation mask (same shape of 717×1121).

f. CNN implementation

For the implementation of this CNN, the TensorFlow/Keras software is used (Abadi et al. 2015). Using only pseudo-reflectivity fields as inputs, the U-Net is a univariate model. The training and validation data do not need to be normalized. After some tests and loss curve verifications, the number of epochs is fixed at 50. Different batch sizes have been tested (16, 32 or 64): results are similar with 16 and 32, but with 64 an overfitting is noticed. The batch size is fixed at 32. The weights of neural networks are randomly initialized and updated using the Adam optimizer (Kingma and Ba 2014), a default learning rate value of 10^{-3} is applied. The weights are iteratively optimized during the training to minimize a weighted cross-entropy loss function (L, Eq. 1). Contrary to the U-Net weights, the loss function weights are fixed. This weighted loss function has been chosen because BEs are rare events and neural networks tend to overestimate the "no bow echo" class with a non-weighted loss function.

$$L = -\frac{1}{N_{gp}} \sum_{k=1}^{N_{gp}} w_0 \, (1-Y_k) \, log(1-p_k) + w_1 \, Y_k \, log(p_k). \quad (1)$$

In Eq. 1, N_{gp} is the number of grid points in the training database, p_k the BE probability and Y_k the true label for the k^{th} grid point. w_0 and w_1 represent respectively the weights of class 0 (no BE) and class 1 (BE). The value of w_0 is fixed to 1. The value of w_1 is discussed in the next section with a hyperparameter selection. Finally, Table 1 shows the values of N_{stride} and N_{patch} presented in the above subsections.

g. Object-oriented evaluation

The capacity of this CNN to detect BEs is evaluated with an object-oriented approach. A grid point approach is not suitable for assessing the BE detections because BEs are small objects and rare events. As a consequence, a small shift between groundtruth and predicted labels may be responsible for bad scores, which is known in the literature as double penalty problem (Davis et al. 2006; Rossa et al. 2008; Ebert 2008). An object-oriented approach has already been used to evaluate AROME-EPS precipitation forecasts (Raynaud et al. 2019). Object attributes used in this study are respectively the 0.25-th (Q25) and the 0.9-th (Q90) percentiles of the pseudo-reflectivity field within bow echo objects, the position of the object mass center, the object area and the FFgust maximum within the object. The attributes are presented in Fig. 5 except for the last one. The Q25 and Q90 percentiles allow differentiation of strong from moderate bow echoes. Another use of Q90 and Q25 is to separate the most active part of BEs (center) from the least active part (edge). The object area is used to distinguish large and small BEs. The maximum FFgust reveals the intensity of wind around the BE, which is a complementary signature of a BE. The distribution of object attributes is computed using all labeled and detected objects. An attribute comparison for matching objects is also performed. The labeled and detected objects are matched if the distance between the centers of their mass is lower than a given threshold in the remainder of the paper.

h. Scores

Contingency-based scores are used to quantify the ability of different U-Net configurations to correctly

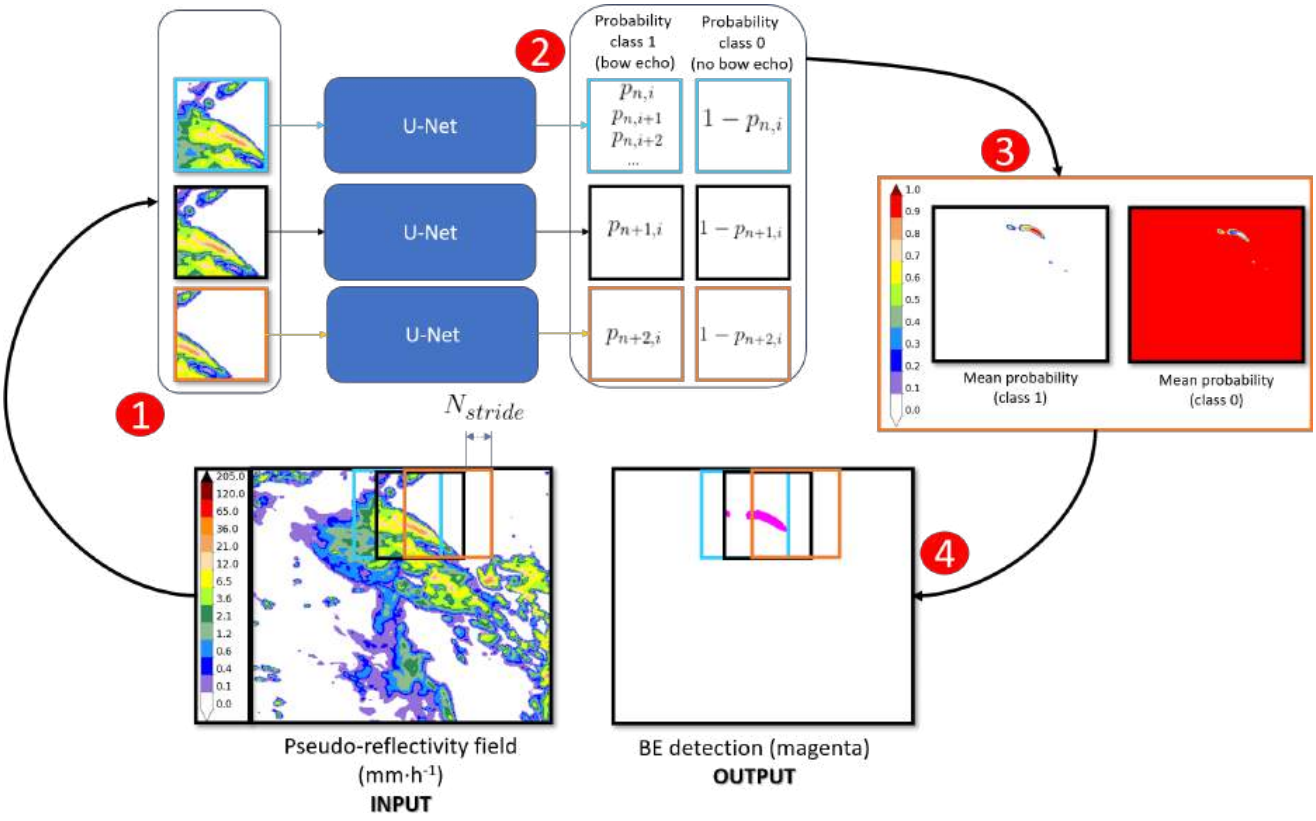


FIG. 4. The whole **prediction** process from pseudo-reflectivity field (INPUT) to BE segmentation mask (OUTPUT) is represented. Steps 1-4 are denoted by red circles and are described in the section 2.e.

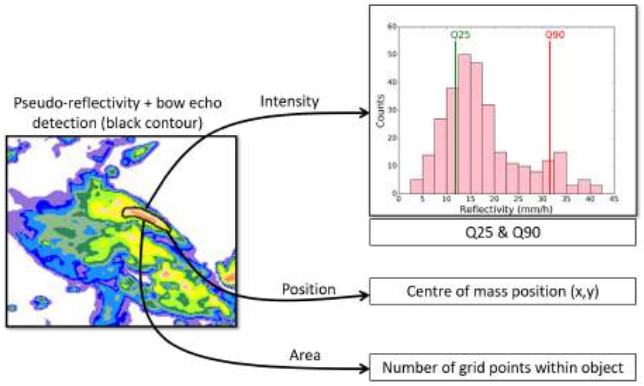


FIG. 5. Object attributes. A bow echo is characterized by 3 attributes: intensity, position and area. Bow echo intensity is described with Q25 and Q90 of pseudo-reflectivities within the object (pink histogram), position with the center of mass and area is the number of grid points within object.

detect the occurrence or non-occurrence of BEs in addition to the comparison of labeled and detected object attributes. A contingency table is computed for the validation database as described in Table 2. The risk of having two or more BEs in the same field is very low because BEs are rare events. The

TABLE 2. Contingency table. Hits (a) correspond to the number of pictures with at least one labeled bow echo and one detected bow echo. False alarms (b) are the number of pictures with detected bow echo(es) but not labeled. Misses (c) are the number of pictures with labeled bow echo(es) but not detected. Correct negatives (d) are pictures with no labeled and detected bow echo.

Contingency Table		BE label(s)	
		YES	NO
BE detection on picture	YES	a	b
	NO	c	d

contingency table is based on the detection of at least one labeled or detected BE over the entire grid for each field. From this contingency table, the classical hit rate (HR, Eq. 2) and false alarm rate (FAR, Eq. 3) are used to evaluate the U-Net skill:

$$HR = \frac{a}{a+c} \tag{2}$$

$$FAR = \frac{b}{a+b} \tag{3}$$

We use the Critical Success Index (or Jaccard Index, Eq. 4) to combine the false alarms and missed

TABLE 3. Values tested for the five considered hyperparameters.

Parameters	Values
input size	24×24,48×48,96×96
data augmentation threshold	30,40,60 $mm \cdot h^{-1}$
ratio noBE/BE patch	2,3,4 noBE for one BE
heavy pseudo-reflectivity patch percentage	15%,25%,35%
class 1 weight (w_1)	2.5,3,3.5,4,4.5

detections:

$$CSI = \frac{a}{a+b+c} \tag{4}$$

3. Hyperparameter Search

a. *Hyperparameter configuration*

The aim of this part is to identify the most influential hyperparameters and to find the optimal combination for the U-Net. We apply the tuning to the following parameters (hyperparameters): input patch size, data augmentation threshold, ratio noBE/BE, heavy pseudo-reflectivity patch percentage and w_1 in the loss function. The selection is based on the scores presented in the previous section and computed on the validation database. The number of values that can be tested is restricted because of limited computing resources. The hyperparameter values tested are presented in Table 3. The 405 combinations are tested. Only the results for input size, ratio noBE/BE and w_1 are presented in the following paragraphs because the sensitivity to data augmentation and heavy pseudo-reflectivity patch percentage turned out to be weak.

b. *Filtering threshold*

The UNets occasionally predict BEs whose size is only a few grid points, on account of the prediction process described in Fig. 4 and especially the choice of a specific probability threshold to split the U-Net outputs into two classes. These detections were mostly false detections which affected the object-oriented scores in a significant way. In order to remove these very small objetcs, the detected BEs with an area below a certain threshold are converted to null events(i.e the class was switched from 1 to 0). The hand-labeled BEs are not affected and they are all incorporated to calculate the object-oriented scores. The optimal threshold is selected with the aid of these object-oriented scores. This threshold is set from the scores of the 405 configurations (Fig. 6). The configurations with HRs and FARs equal to 0 are not taken into account in this figure because they are

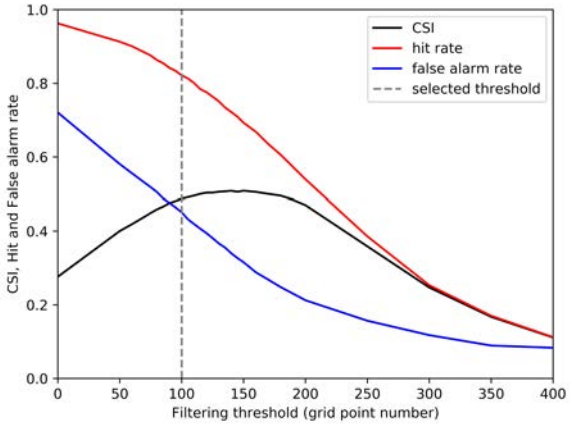


FIG. 6. Scores as a function of filtering threshold. Scores are computed for thresholds from 0 to 400 grid points. Considering the 405 configurations, the CSI, HR and FAR median are plotted in respectively black, red and blue.

associated with unskillful U-Nets (cf section 3.c.1). The hit rate naturally decreases when the filtering threshold increases. The same behavior is observed for the false alarm rate. According to the CSI, the optimum threshold is around 150 grid points. With a 150-grid-point threshold, the HR is around 70%. However, forecasters prefer a higher HR even if the FAR increases simultaneously (see section 6.c, the second item). Hence, a filtering threshold of 100 grid points is preferred, since it leads to a HR median above 80% and a CSI close to the optimal value. Moreover, this threshold approximately corresponds to the size of the smallest hand-labeled bow echoes.

c. *Results*

The sensitivity test results are divided into two parts. The first part focuses on the U-Nets that always predict the same probability value for the class 1 in each grid point. This probability value is equal to 0 or 0.5, depending on the U-Net. This issue is well known for cases of unbalanced datasets (Chawla et al. 2004). For these U-Nets, the loss function remains unchanged according to the epochs. They converge to an unskillful solution where the class 1 is completely missing, or unable to differentiate the class 0 and 1 when the probabilities are equal to 0.5 in every grid point. These U-Nets are called unskillful U-Nets hereafter. The second part examines the performances of other skillful U-Nets able to predict bow echoes.

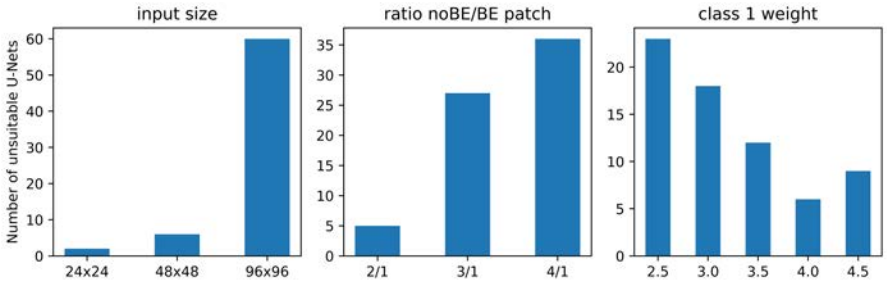


FIG. 7. The number of unskillful U-Nets is counted according to each parameter: input size (left), ratio_noBE/BE patch (center), weight of class 1 in loss function (right).

1) UNSKILLFUL U-NETS

BEs are rare events in space and in time as previously mentioned. The training database must be carefully set up so that a U-Net can properly detect bow echoes. Fig. 7 shows the number of unskillful U-Nets as a function of the tested parameters. Sixty-eight among the 405 U-Nets are found to be unskillful (17%). The number of unskillful U-Nets increases with input size, especially from 48×48 grid points to 96×96 grid points. This result can be explained by the typical length of BEs which is around 100km and corresponds to 40 grid points. Bow echoes occupy relatively less space in a patch of 96×96 grid points than in one of 48×48 grid points. The classes 0 and 1 are more unbalanced in the first case than in the second one and the risk of unskillful U-Nets is higher. Risk of an unskillful U-Net is less frequent when the ratio noBE/BE patch is 2/1 than when it is 3/1 because the classes are less unbalanced. Finally when the weight of class 1 increases, the number of unskillful U-Nets decreases. The main role of a weighted loss function is to avoid unbalanced data issues (Kurth et al. 2018). Those unskillful U-Nets with a HR and FAR equal to 0 are removed in the next sub-subsection.

The intuition behind why a smaller patch size and larger w1 weight produces more skillfull Unets is because the contribution of BE pixels to the overall DL loss function are increased for a smaller patch and a larger w1 weight. In other words, when considering a smaller patch the ratio of pixels labeled 1 (i.e., BE) to pixels labeled 0 (i.e., noBE) is larger. Thus the influence of BE pixels on the loss (i.e., what the DL model learns) is larger. Similarly, the larger w1 weight for the BE class accomplishes the same increased influence.

2) CONFIGURATION SCORES

HR, FAR and CSI are analyzed for each hyperparameter. Only the skillful U-Nets are considered in this sub-subsection. The major results are presented in Fig. 8. The size of patches is the main parameter that explains the fluctuation of scores. HR and FAR tend to decrease with an increasing input size. The CSI is lower for 24×24 input because of a high FAR. These results can be explained by the small input size (24×24) compared to the typical length of BEs (40 grid points). The size of U-Net inputs should be able to get all information of bow echo objects, otherwise all BEs in the preprocessing step (Fig. 3, step 1) are split into different CNN inputs and U-Nets can not properly learn the BE features. Focusing on false alarms, they mainly correspond to objects with moderate pseudo-reflectivities (Q90 maximum around 30 $mm \cdot h^{-1}$). The false alarm distribution slightly changes with input size, with a shift towards weaker pseudo-reflectivities for small input sizes. This result is also consistent with the previous remark concerning the BEs split into different CNN inputs. Considering bow echo class weight (w_1), HR tends to increase with higher weights (same tendency for FAR). Concerning CSI, optimum weights depend on input size, but no trend is found for 48×48 and 96×96 inputs. However, for 24×24 input, CSI is higher for the lowest weights thanks to much lower FAR. Ratio noBE/BE does not have any significant influence on the different scores contrary to what has been mentioned in the above paragraph about unskillful U-Nets (not shown). The role of the different parameters can be summarized as follows:

- Input size must be carefully selected because of its influence on unskillful U-Nets (many more with large-size input) and scores (HR and FAR decrease with input size).

- Bow echo class weight (w_1) is also an important parameter which influences unskillful U-Nets and global scores (higher HR and FAR with higher w_1). Optimum weights depend also on input size.

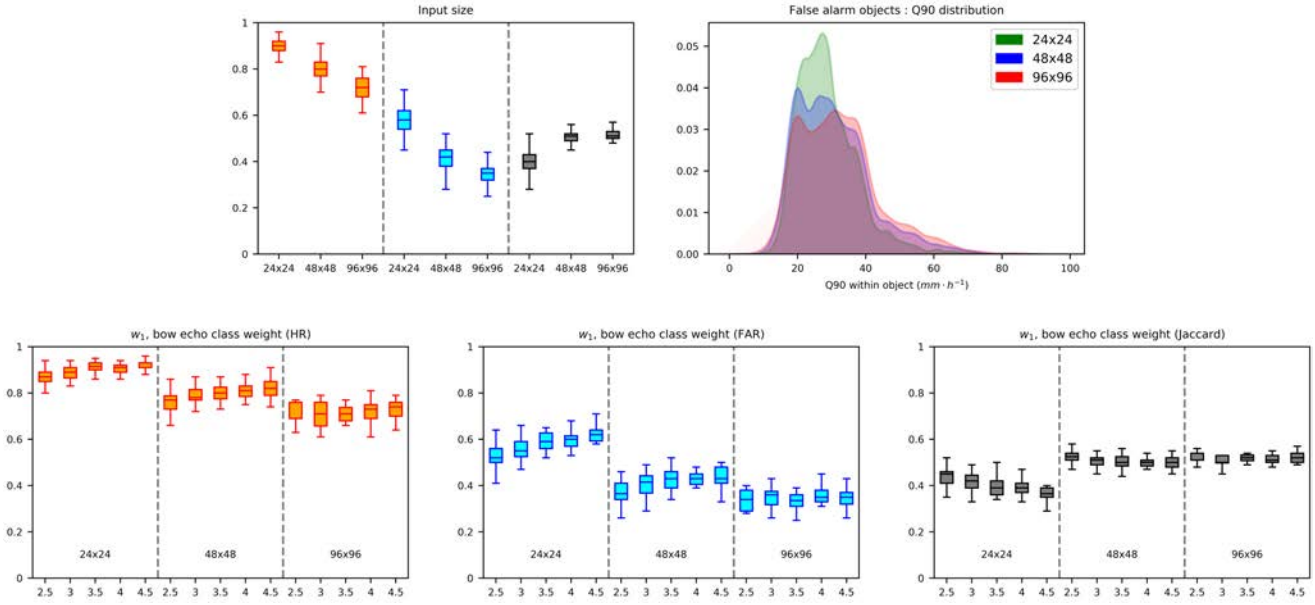


FIG. 8. Performance scores for skillful U-Nets. On the first line, boxplots for HR (orange), FAR (blue) and CSI (black) are represented according to input size (left). False alarm distributions for Q90 attribute in $mm \cdot h^{-1}$ are plotted for the three input sizes (right). On the last line, a focus on bow echo class weight (w_1) and links to input size is proposed. For HR (left), FAR (center) and CSI (right), scores are plotted according to weights and input sizes separated by dashed gray lines.

- Though ratio noBE/BE has an impact on the number of unskillful U-Nets, it does not influence HR and FAR.

d. Optimal configuration

In an operational context, we have to chose a configuration among the 405 ones presented in the section 3.a. To make a choice, the fifteen best configurations according to CSI are presented in Table 4. Only input sizes of 48×48 or 96×96 are included in these fifteen configurations. Even if the CSIs of these configurations are very close, different "strategies" are possible. The HR/FAR pairs for each configuration can have high (8^{th} : 0.8/0.36) or low (3^{rd} : 0.71/0.25) values and consequently the total number of detections varies greatly from one U-Net to another. A way to decide which U-Net is optimal is to consider how this CNN will be used. Since BEs are severe but rare events, misses should be avoided. A U-Net with a high HR is preferable. The optimal configuration on this specific point is the 10^{th} setup with a HR of 0.86. This configuration is now considered the optimal one and will be described in more details in the next section.

TABLE 4. Fifteen best configurations according to the CSI. The values of the five tested hyperparameters are indicated in the following order : input size, data augmentation threshold, ratio noBE/BE, heavy pseudo-reflectivity patch percentage and bow echo class weight. HR and FAR are also shown for each configuration. The selected configuration is in bold.

U-Net configurations					HR	FAR	CSI
96x96	30	3/1	15 %	4	0.73	0.25	0.58
48x48	60	4/1	35 %	2.5	0.76	0.29	0.58
96x96	60	2/1	35 %	3.5	0.71	0.25	0.57
96x96	60	2/1	25 %	4.5	0.78	0.32	0.56
96x96	40	3/1	25 %	4.5	0.77	0.32	0.56
96x96	40	2/1	15 %	2.5	0.77	0.34	0.56
48x48	60	3/1	35 %	2.5	0.73	0.29	0.56
48x48	40	3/1	35 %	3.5	0.8	0.36	0.56
48x48	40	3/1	25 %	4	0.77	0.33	0.56
48x48	**40**	**2/1**	**15 %**	**2.5**	**0.86**	**0.39**	**0.56**
96x96	60	4/1	25 %	4.5	0.72	0.29	0.55
96x96	60	2/1	35 %	4.5	0.78	0.35	0.55
96x96	30	2/1	15 %	4	0.75	0.33	0.55
48x48	60	4/1	35 %	4.5	0.82	0.38	0.55
48x48	40	4/1	25 %	3	0.72	0.29	0.55

4. Optimal U-Net : an object-oriented evaluation

Global scores are not sufficient to precisely evaluate the U-Net skills. To go further, attributes of BE objects must be analyzed to support the HR, FAR and CSI scores. Those scores described in the section 2.h provide no information about the detection quality but only about the BE detection and the BE labeling simultaneously in the same field. A com-

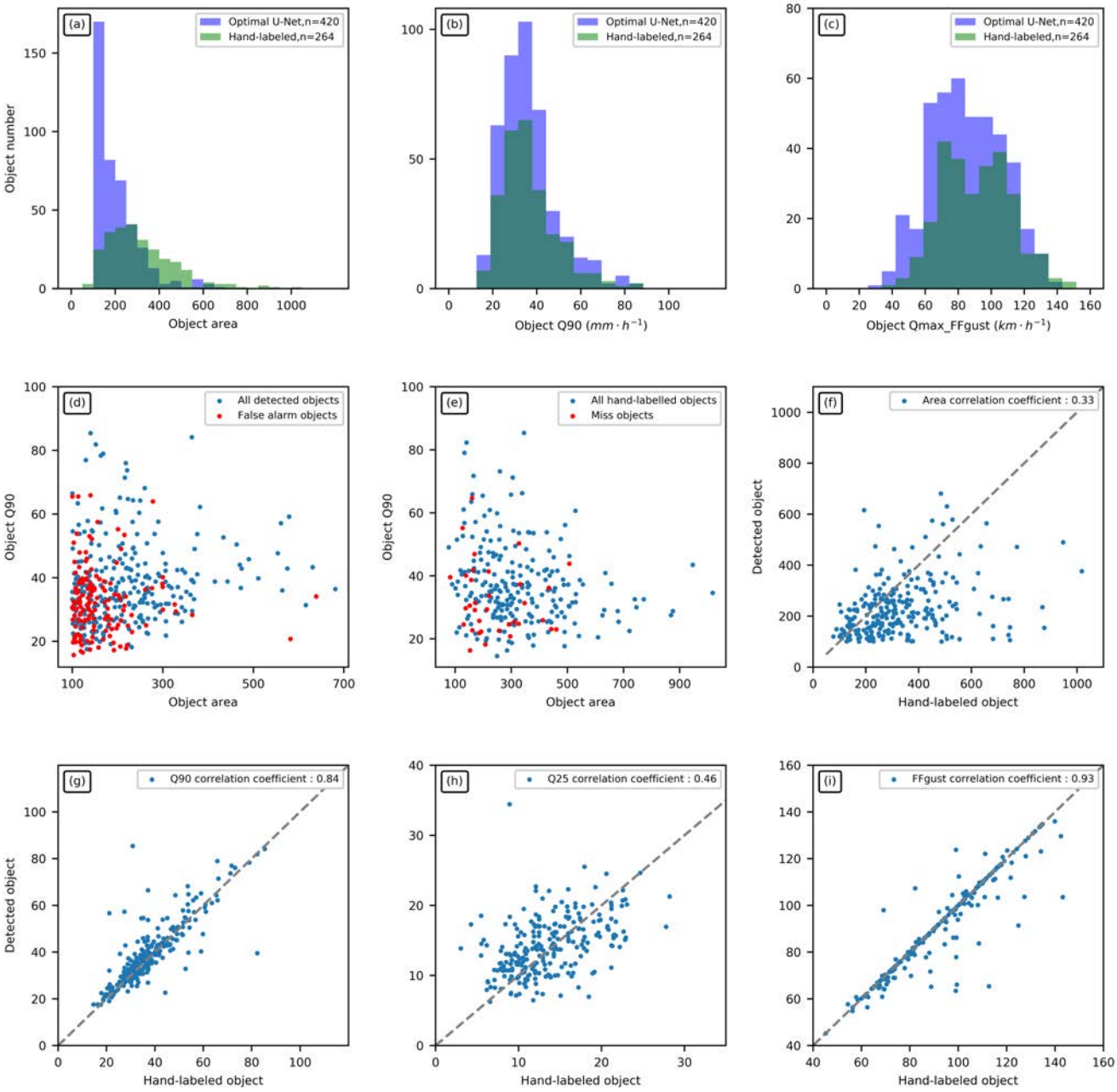


FIG. 9. Evaluation of the U-Net optimal configuration with an object-oriented approach. At the top left, area histograms for the optimal configuration (blue) and the hand-labeled dataset (green) are compared (a). The total number of BEs is visible in the legend. The same is done for Q90 attribute (b) and maximum FFgust (c). Then, the characteristics of false alarm (d) and miss (e) features are presented. All detected objects are plotted according to area and Q90 in blue, the false alarms are in red. The same is done for the hand-labeled dataset (blue) and the misses (red). The correlation between matching pairs of hand-labeled objects (x-axis) and detected objects (y-axis) is studied for four attributes : area (f), Q90 (g), Q25 (h) and Qmax_FFgust (i). The gray dashed line represents a perfect match (y=x). The correlation coefficient is also mentioned in the legend.

prehensive object-oriented evaluation is done on the validation dataset. Results are presented in Fig. 9. Examples of correct detections (Fig.S3), false alarms (Fig.S4) and misses (Fig.S5) are available in the online supplement to this paper.

a. Attribute histograms

The global distributions of bow echoes areas and Q90 are first studied (Fig. 9a and b). The area distribution is different between detected and hand-labeled datasets. The total number of detected ob-

jects is rather high with a hit rate of 0.86 and a false alarm rate of 0.39. No detected object has an area below 100 grid points thanks to the filtering threshold. Detected objects have mostly an area below 250 grid points with several between 100 and 150 grid points. The number of detected objects decreases exponentially with area size. This behavior is not observed for the hand-labeled dataset, with a maximum around 300 grid points and objects more normally distributed. Detected objects are more frequent but they tend to be smaller than hand-labeled objects. The distributions of pseudo-reflectivity Q90 for detected and hand-labeled objects are quite similar even if the total number of detected objects is higher. The maximum of the Q90 distribution is around 35 $mm{\cdot}h^{-1}$ for both detected or hand-labeled datasets.

The hand-labeled distribution for maximum FFgust (Fig. 9c) has a range from 50 $km \cdot h^{-1}$ to 150 $km \cdot h^{-1}$ with two maximums around 70-80 $km \cdot h^{-1}$ and 110 $km \cdot h^{-1}$. The hand-labeled BEs associated with weak FFgusts are rare. They are examined one by one and they are mainly BEs at the beginning of their life cycle with weaker wind gusts. The distribution of detected objects is very similar with a single maximum around 80 $km{\cdot}h^{-1}$. Even if FFgust is not a predictor of U-Nets, the comparison of detected and hand-labeled objects shows that the detected BEs are associated with the same FFgust distribution as the hand-labeled BEs. The wide range of FFgust advocates the addition of a severity attribute for each BE detection to differentiate 'moderate' BEs without significant damage from BEs with potentially significant damages.

b. False alarm and missed attributes

To extract false alarm and missed features, each object is represented in area/Q90 graphs (Fig. 9d and e). BEs with high (respectively low) Q90 are usually associated with small (respectively large) areas at first glance. This characteristic is physically consistent with the life cycle of BEs and the hand-labeled objects (strong and small at the beginning and then weaker and larger (Goulet 2015)). This is a reassuring aspect concerning the general behavior of the U-Net. The false alarms mainly correspond to small and weak objects. On the other hand, the large or strong objects are usually correctly detected. In an operational context, this information may prove useful when discussing the relevance of object detections. The number of misses is limited but their distribution is more homogeneous. They are slightly more concentrated on small objects but this result is not as obvious as for false alarms. Two characteristics are nevertheless common : the areas of missed objects are always below 500 grid points and the maximum wind gusts are constantly below 110 $km \cdot h^{-1}$. The prediction of BEs is more difficult for small and weak objects. This may be related to the more difficult recognition of these BEs by the expert, leading to a less precise labeling. In such cases, the U-Net is not very accurate.

c. Attribute correlations

This section defines a pair of matching objects (detected/hand-labeled) in order to compute attribute correlations. A pair is formed when the distance between two mass centers is less than 100km. Object attributes are compared and a correlation coefficient is computed (Fig. 9f,g,h and i). The correlation for the areas is weak. The U-Net tends to underestimate the size of objects, especially for the large bow echoes. This conclusion is consistent with the comparison of the area histograms. Regarding Q90, the correlation is higher and points are located around the dashed gray line (y=x). Q90 is representative of the most active part (i.e heaviest pseudo-reflectivities) located in the center of BE objects. Q25 is more representative of the BE borders with lower pseudo-reflectivities, and the correlation is worse than that of Q90. These findings point to the ability of the U-Net to properly retrieve the most relevant part of BEs, whereas the BEs borders are less well estimated. The strong correlation of FFgust maximum (as well as Q90) supports the conclusion of the previous sentence since the strongest FFgusts are in the most relevant part. In subsection 6.c, the results about the false alarms, misses and correlation attributes will be further discussed and compared with comments from forecasters.

d. BE and SC confusion

The capacity of the U-Net to differentiate BEs from other types of severe convective storms is also verified. The main issue could be to mix up BE with other convective storms as intensity can be similar. Initially, the CNNs detected all patterns with strong pseudo-reflectivities as BEs (section 2.d, last item). To remove any doubt, it is verified that BEs are not mixed up with isolated supercells (SCs, most common severe convective storm in France) when the confusion between the two convective events is very unlikely for a human. Eleven convective situations that occurred in 2019 and in 2020 are examined (Table 5). These situations are covered by 5066 pseudo-reflectivity forecast fields from AROME-EPS, in

TABLE 5. BE and supercell (SC) overlaps. For each date, the number of fields with hand-labeled supercells (second column) and predicted bow echoes (third column) is reported. The last column, overlaps, shows the number of fields where predicted BE and hand-labeled SC objects have at least one grid point in common.

Date	SCs	BEs	Overlaps
2019-06-15	67	94	3
2019-06-18	37	27	0
2019-06-19	18	21	0
2019-06-24	47	88	6
2019-07-06	36	6	1
2019-07-15	172	12	1
2019-07-26	34	52	0
2019-08-09	90	38	1
2019-08-18	54	38	1
2020-04-17	25	17	1
2020-05-09	48	37	1
Total	628	430	15

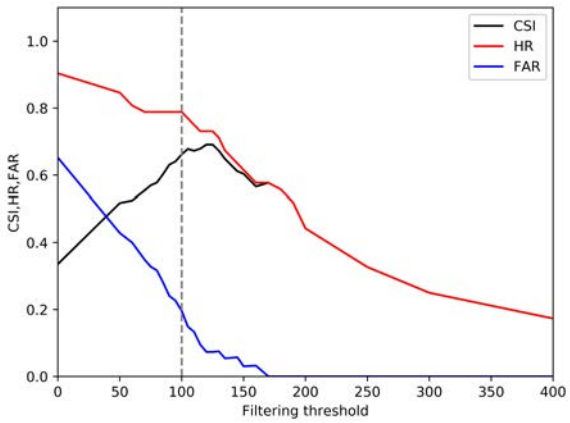


FIG. 10. Similar plot to Fig. 6, but only the optimal U-Net is considered and not all U-Net configurations.

which a same expert hand-labeled 628 SCs. The labeling of SCs rely on pseudo-reflectivity fields and UHmax between 800 and 500 hPa. We keep the maximum in absolute value of UH to correctly detect both right and left moving SCs. The specific contour of 50 $m^2 \cdot s^{-2}$ or (-50 $m^2 \cdot s^{-2}$ for negative values) is plotted over the pseudo-reflectivity fields to help the labeling process. Discrete and embedded SCs in lines of convection are hand-labeled. High reflectivties ($>$ 50 $mm \cdot h^{-1}$) and high UHmax values are required to label a SC. In this analysis, predicted BEs from the U-Net are compared with hand-labeled SC objects. The confusions between SCs and BEs are rare with only 15 occurrences for over 628 SCs and 430 BEs. As SCs can be embedded in quasi-linear convective systems, these overlaps can be understandable. A SC can also sometimes evolve into a BE (Klimowski et al. 2004) and during the transition phase, an overlap between BE and SC labels is possible. These 15 overlapping occurrences have been manually analyzed to check whether the overlaps occurred in one of the previously mentioned two cases. Only 3 of the 15 overlapping occurrences do not occur in one of the two cases and can be considered as abnormal overlaps after examination. To conclude, the risk of confusion between BEs and SCs is very limited with only 0.5% of SCs wrongly detected as BEs.

5. Extension to the AROME deterministic model

At this stage, the U-Net has been trained and tested using AROME-EPS forecasts. The possibility to detect BEs in the deterministic AROME forecasts, which has a higher resolution (1.3 km instead of 2.5 km), is also important for forecasters. Transfer learning from AROME-EPS to AROME is tested because data from deterministic models are limited and setting up another training database would be very time-consuming. We test the optimal configuration on the AROME deterministic model without re-training after encouraging results with AROME-EPS. Pseudo-reflectivity fields from AROME are available on the same grid as AROME-EPS (2.5 km) thanks to a quadratic interpolation method. The cases in the AROME and AROME-EPS validation databases should be the same to have comparable results. There are only 348 pseudo-reflectivity fields and 62 hand-labeled bow echoes in its validation database because deterministic AROME is a single forecast, which can limit the significance of results.

a. Filtering threshold and global scores

Even if the pseudo-reflectivity fields have the same resolution (2.5km) as those of AROME-EPS, the native resolution of AROME is smaller (1.3km) and the pseudo-reflectivity fields in the AROME deterministic model are more realistic with stronger pseudo-reflectivity gradients and consequently we re-examined the filtering threshold applied to AROME data. A similar graph to Fig. 6 is plotted in Fig. 10. With only one setup, the curves are noisier. The maximum CSI is around 100-150 grid points which is in favor of retaining the same filtering threshold of 100 grid points as for AROME-EPS (dashed grey line). With this threshold, the HR (respectively FAR) is equal to 0.79 (respectively 0.2). The FAR is better than the one for the AROME-EPS outputs

(0.39) while the HR is slightly lower. There are several possible explanations besides the small size of the database. Bow echo recognition is in this case easier for experts with less ambiguous cases because pseudo-reflectivity fields are more realistic. A hand-labeled database with higher quality and stronger pseudo-reflectivity gradients can both explain the lower FAR.

b. Results

The same figures as in the previous section (Fig. 11 from a to i) are presented to study more specifically the U-Net behavior in AROME forecast outputs. The same U-Net behavior is observed despite the small database. Even if the number of hand-labeled and detected bow echoes is more even, undersized objects are still predicted with a correct Q90 distribution. The FFgust distributions are also comparable even if some U-Net detections, which are false alarm objects, are associated with weak FFgusts. The false alarm and missed attributes are also very similar to that of the AROME-EPS database. The false alarms correspond to weak and small objects. The misses concern mostly small and weak BEs, but a small and strong BE (Q90 around 85 $mm \cdot h^{-1}$) is also missed. The attribute correlation results are also very similar to AROME-EPS with a high correlation coefficient for the Q90 and FFgust attributes but a worse correlation for area and Q25. No significant change can be noticed when the comparison is made with the AROME-EPS database except a better FAR. We conclude that the same U-Net can be used to detect bow echoes in AROME deterministic model. The extension to AROME will be further discussed in subsection 6.c based on the feedback from the subjective comparison by forecasters. The possibility to extend U-Nets from EPS to deterministic models is an advantage of both DL and ML methods. EPS outputs are indeed well adapted to these methods which require large datasets (Schumacher et al. 2021). This is not always the case for deterministic models. However, this extension is realistic if both EPS and deterministic models are sufficiently similar. For instance, the grid of EPS and deterministic model outputs should be close to assume that spatial features learned during the training process of the EPS database are still valid for the deterministic models. Otherwise, interpolating model outputs could be tested.

6. Utilization of BE detections by end-users : synthesis plots and feedback

The satisfactory evaluation of the U-Net performances for both AROME-EPS and AROME outputs motivates the development of forecasting products about BE detections. For that purpose, synthesis plots are presented in this section. They have been designed in collaboration with forecasters, inspired by Demuth et al. (2020). Three synthesis plots are presented: trajectory and paintball plots in the first subsection and a probability map of BE occurrences in the second subsection. To facilitate the use of U-Net outputs, each AROME-EPS run is divided into three different time periods (Day 1, Night Day 1/2 and Day 2). Thanks to these time periods, the entire life cycle of a BE can be followed over only one of them. The three synthesis plots are computed for the three time periods and consequently, for each AROME-EPS run, 3×3 synthesis plots are produced. The three time periods are based on UTC hour to facilitate comparisons between AROME-EPS runs. The last subsection is a summary of forecasters' feedback concerning the U-Net performance and the synthesis plots.

a. Trajectory and paintball plots

One way to summarize U-Net outputs is to plot the overlapping of the detections of every member for a time period (Day 1, Night day1/2 or Day2). The trajectory plot is very useful to visualize the temporal evolution of a BE in AROME forecasts with a different color for each UTC hour. Fig. 12a shows the BE northward trajectory over the North-East of France. It gives useful information about the BE risk period which is during the afternoon and early evening in this case (14UTC and 18UTC from green to yellow).

The paintball plot (Fig. 12b) helps evaluate the number of different members that predict a BE with a different color for each member. Focusing on a specific BE object, we can assess the member and the UTC time of the detection by combining both the trajectory and paintball plots. The U-Net outputs can be quickly identified thanks to this information.

b. Probabilistic approach

Trajectory and paintball plots are very useful to evaluate BE risks. However, the interpretation of a large number of detections over a small area as shown in Figs. 12a and b can remain troublesome. A probabilistic approach is useful to quantify the risk of BEs in AROME-EPS.

1) SPACE-TIME TOLERANCE

A probabilistic synthesis should allow for some space and time tolerances because of the rather low number of AROME-EPS members and the small size

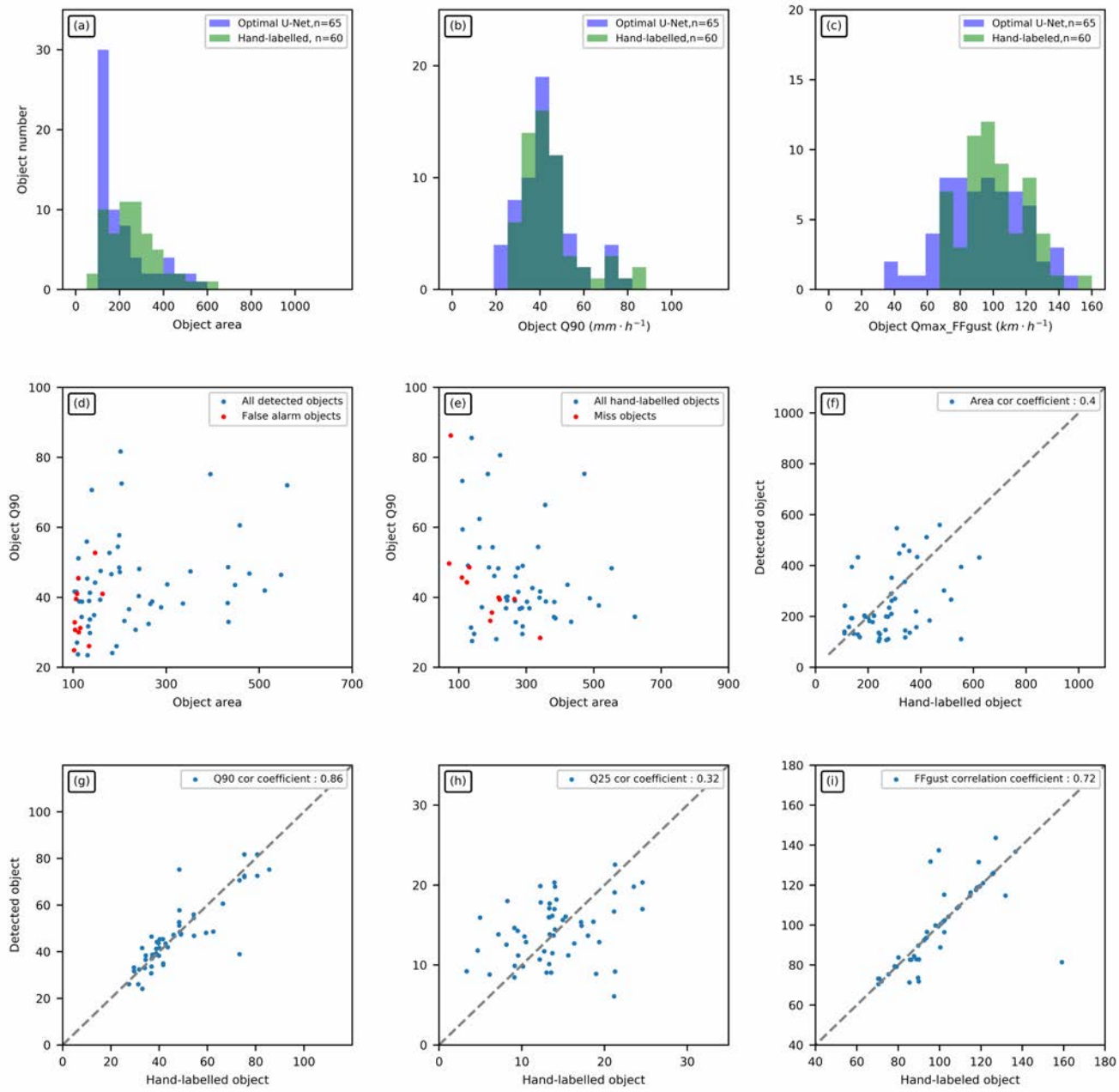


FIG. 11. Evaluation of the U-Net on the AROME deterministic model database. The subplots are the same as in Fig. 9.

of BEs. For that purpose, a neighborhood disc of radius εkm, denoted d_ε, is defined around each grid point x:

$$d_\varepsilon(x) = \{y \mid \|x - y\|_2 \leq \varepsilon\} \tag{5}$$

The value of ε will be discussed in the next paragraph. A neighborhood probability of BE ($P_{\varepsilon,l}$) is defined by :

$$\mathbb{P}^m_{\varepsilon,l}(x) = \begin{cases} 1 & if \ \sum_{y \in d_\varepsilon(x)} \mathbb{1}_{BE}(y) \ \geq l \\ 0 & otherwise \end{cases} \tag{6}$$

$\mathbb{1}_{BE}(y)$ corresponds to the U-Net output for the grid point y (1 in a BE, 0 out). A grid point x for a member m is associated with a probability of 1 if at least l grid points in the neighborhood of x correspond to a BE detection. To include a time tolerance, the probability at time t is computed using forecasts valid at t, $t - 1h$ and $t + 1h$:

$$\mathbb{P}_{\varepsilon,l}(x,t) = \frac{1}{3 \times N_{mb}} \sum_{i=-1}^{1} \sum_{m=1}^{N_{mb}} \mathbb{P}^m_{\varepsilon,l}(x, t+i), \tag{7}$$

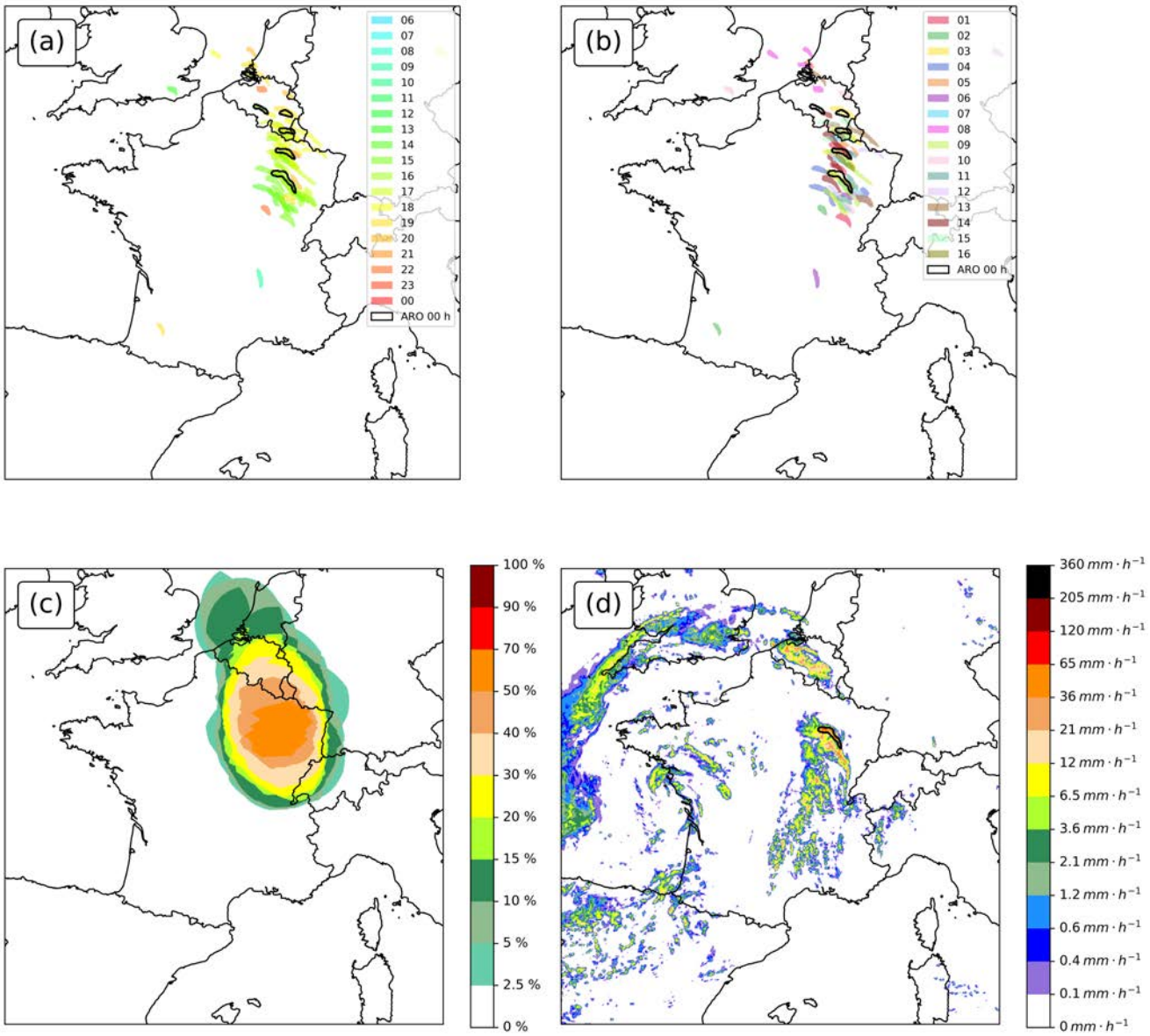


FIG. 12. Synthesis plots of U-Net detections. The forecasts from AROME-EPS launched on 12 June 2020 at 03UTC and from the deterministic AROME model launched on 12 June 2020 at 00UTC are taken as an example. The considered period of forecasts is Day 1, ranging from 12 June 2020 at 06UTC to 13 June 2020 at 00UTC. Trajectory plot is on the top left (a), paintball plot on the top right (b) and probabilistic plot (c) on the bottom left. A different color is used for each UTC hour of the BE detections for the trajectory plot (a). The paintball plot (b) represents the same detections as the trajectory ones, but a different color is used for each member (from 1 to 16) instead of UTC time. Black contours are added around deterministic AROME objects. On the bottom right (d), the predicted pseudo-reflectivity field and the corresponding BE detection (black contour) are overlaid, using the AROME outputs valid on 12 June 2020 at 16UTC. Forecasters examine this kind of superposition (d) in section 6.c for their feedback.

where N_{mb} is the total number of members in AROME-EPS and $\mathbb{P}^{m}_{\varepsilon,l}(x, t+i)$ indicates the BE probability of a grid point x at time $t+i$ for a member m. A probability is defined for each grid point at a specific time t by taking into account equally all AROME-EPS members. Over a given time period (Day 1, Night Day 1/2 or Day 2), only the maximum probability at each grid point is kept (Eq. 8).

$$\mathbb{P}^{max}_{\varepsilon,l}(x) = max\left\{t \in period,\ \mathbb{P}_{\varepsilon,l}(x,t)\right\} \qquad (8)$$

This maximum probability plot is presented in Fig. 12c. The value of the radius ε is equal to 150km and will be discussed hereafter. The value of l is fixed to 10 hereafter and will not be discussed in aid

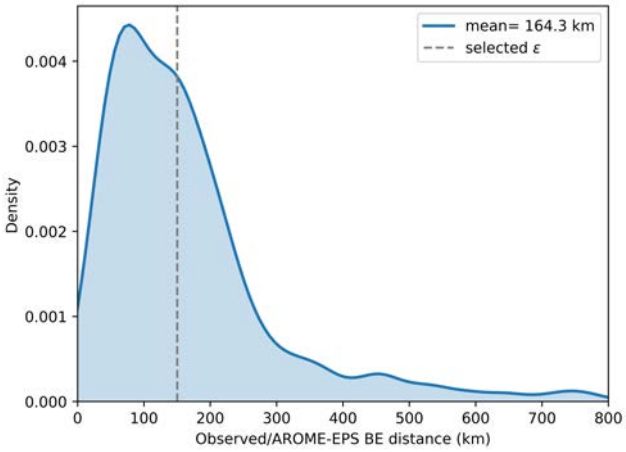


FIG. 13. Density plot of distances between observed BEs and AROME-EPS detections. The average distance is computed in legend (excluding errors above 500km, see text for explanation).

of the value of the radius ε. The maximum probability shown on Fig. 12c is above 50 % (high probability for a BE risk) over Northeastern France, which is consistent with the other synthesis plots. Moreover, isolated detections like the ones over the Southwest of France are filtered (risk below 2,5 %) because of time averaging.

2) NEIGHBORHOOD DISTANCE

An optimal value of the radius ε used for the probability plot should be defined. Position errors between observed BEs and AROME-EPS detections are examined as this ε value depends on AROME-EPS skill. Observed BEs have been hand-labeled on radar data, in the same way as for AROME training or validation databases. However, radar data grid covers only mainland France and near the borders. Only AROME-EPS detections over this smaller grid are kept. A density plot of position errors is presented in Fig. 13. The majority of position errors are under 300km. The average error is around 180km but this value seems to be overestimated because of some unrealistic distances of several hundred kilometers. In that case the detected object should be considered as a false alarm and the observed object as a miss. If distance values above 500km are ignored, the average error is around 150km. The value of ε is set to 150km based on this result.

c. Forecasters' feedback

In this study, a U-Net has been trained and tested to detect bow echoes in AROME-EPS and AROME forecasts. Only one expert has hand-labeled the training and validation databases. All findings and scores are based on these hand-labeled contours of bow echoes. In some cases the detection of BEs is highly subjective and there is a need for more experts to assess the U-Net performance. A subjective evaluation from future users was considered interesting in addition to object-oriented scores. Ten Météo-France forecasters contribute to this study. Each expert is given a superposition of pseudo-reflectivity field and U-Net detection such as that of Fig. 12d to subjectively evaluate the U-Net. The examined cases come from the validation AROME-EPS database and the AROME model. Since the number of pictures is very large (around 3000), experts are divided into three groups, and each group examines a third of the pictures. Each forecaster uses an evaluation form for their evaluation. Each forecaster counts the number of false alarms, misses or correct detections for each date and each member using the same criteria as in section 2.h. A detection rate and false alarm rate are calculated and compared with the HR and FAR of sections 4 and 5. This comparison is feasible because the HR and FAR calculated with the complete validation database are similar to HR and FAR computed with the three subsets one by one (more or less 3%). A last section named 'general comments' in the evaluation form allows for expressing an opinion concerning U-Net behavior. The results of the U-Net object-oriented evaluation are not communicated to the experts to avoid influencing their judgement. Concerning the general comment section, some guidelines are given to focus on miss, false alarm features (more frequent in case of large/small area or strong/weak intensity ?), comparison of U-Net skills between AROME-EPS and AROME but also concerning confusion with other MCSs. Each expert examines, independently of the others, the U-Net outputs. Feedback and conclusions have been gathered hereafter (more details concerning the forecasters' feedback are available in the online supplement to this paper). Based on a broad consensus, the main findings are reported in the following section :

- The feedback is overall positive. The pair of HR/FAR according to the experts is around 70%/20%. Most of the forecasters consider that these detections and this work can facilitate the use of EPS in an operational context. The detection of large BEs is acceptable (still based on the general comments). This point is consistent with findings of section 4.
- Only a few cases are identified as false alarms (mean FAR of 20%). This conclusion is very promising for an operational application, but it contradicts the higher false alarm rate of 39 %

obtained in section 4. Even if most forecasters have found less false alarms, some of them comply with the conclusions of sections 4 and 5 (higher false alarm rate that corresponds to small objects). Some experts consider small objects as cell bow echoes (Klimowski et al. 2004), which can explain these different judgments. This result highlights the subjective point of view of different experts concerning the exact BE definition and the fact that they only had the pseudo-reflectivity field at their disposal to assess it in the model, as the U-Net did. Forecasters tend to favor detection algorithm with high FAR because all detections are carefully examined. On the contrary, a synthesis plot without detection does not draw attention. This conclusion supports the selection of the optimal configuration as described in section 4, a high HR and high FAR are preferred.

- Some misses are also noticed (mean HR of 70%). There are three main reasons to explain these misses. First in case of weak pseudo-reflectivities, which is consistent with the object-oriented evaluation. In case of fragmented BE, the detection can be missed or incomplete. The last one highlights a lack of temporal robustness/stability of the detections. A BE is sometimes not detected at each lead time: for instance a BE is detected at the time t, not at $t+1h$ and detected once again at $t+2h$. In that case, misses are less problematic because they mixed together all the others in the synthesis plots.

- BE contours are most of the time too small and focus on the strongest part of BEs. This observation is consistent with results in sections 4 and 5.

- The detection quality in AROME-EPS and AROME is approximately the same according to the majority of experts (consistent with section 5). The risk of confusion between SCs and BEs in AROME is slightly more important according to some of them.

- A last comment concerns the poor U-Net behavior for a specific weather situation in 2018. In this case, the BE had an unusual propagation axis from South-East to North-West in comparison to BE climatology in France (South-West/North-East or West/East). Both the original curve and the fragmented characteristics of this BE can explain these very bad results of the U-Net. Fig. 14 is a good example with a BE correctly detected at 12 and 13UTC but not at 14UTC. Unusual cases are always difficult to handle for neural networks when these cases are not representative in the training samples.

- Trajectory plot (Fig. 12a) is the preferred visualization for the majority of the expert group, even though this conclusion should be confirmed in an operational context.

In summary, forecasters' feedback is overall positive and consistent with the object-oriented evaluation. This experiment is also a first step toward introducing AI-based products in the forecasters tools. The cooperation to design the synthesis plots in section 6 was strongly appreciated.

7. Conclusion and future work

In this paper a convolutional neural network inspired from a U-Net architecture has been developed to detect Bow Echoes (BEs), specific Mesoscale Convective Sytems (MCSs), in the outputs of the French kilometer-scale AROME and AROME-EPS models. This U-Net is trained and evaluated using a hand-labeled dataset of more than 9000 training samples. Pseudo-reflectivity (pseudo because expressed in $mm \cdot h^{-1}$ and not in dBZ) is the unique input of the neural network. Pseudo-reflectivity fields are cut into smaller patches and several pre-processing steps are implemented in order to design a balanced training database. During the prediction process, pseudo-reflectivity fields are also cut into smaller overlapping patches and fields are reconstructed to detect bow echoes over the whole AROME domain.

A filtering threshold has been defined and fixed to 100 grid points to optimize object-oriented scores. Different configurations are tested to understand the role of several hyperparameters involved in the training dataset design and in the U-Net. The input patch size is the most influential parameter. The size should be large enough to extract information from the whole BE and its surroundings. The proportions of training patches with and without BEs, and the weights in the loss function are also influential hyperparameters. The optimal configuration is determined from the CSI and the hit rate.

This optimal configuration was examined in details with an object-oriented approach. The false alarms and misses mainly correspond to small and weak objects. The U-Net underestimates the size of BEs but properly captures the most active part of BEs. Even if the wind gust speed at 10m (FFgust) is not a predictor for the U-Net, the detected BEs are associated with the same FFgust distribution as the hand-labeled BEs. An extension of the U-Net

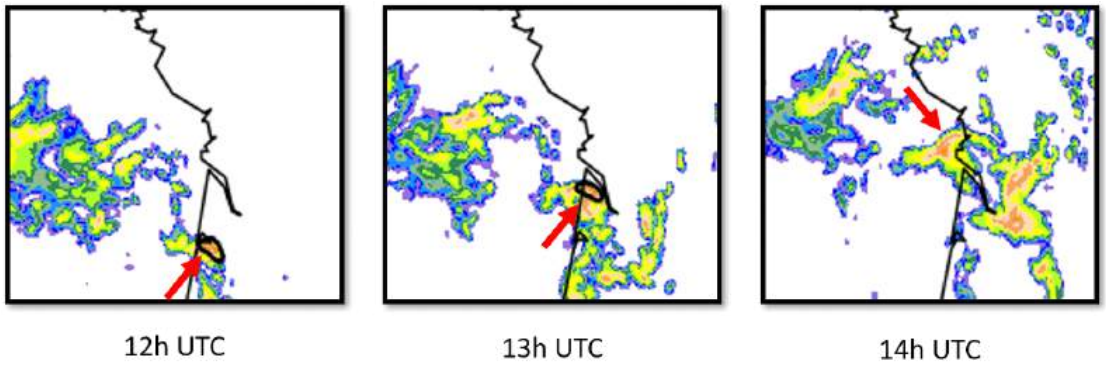


FIG. 14. Forecasts from AROME-EPS initialized on 25 May 2018 at 21UTC, zoomed over western France. Pseudo-reflectivity fields and BE detections (black contours) from member 2 between 26 May 2018 at 12UTC (left) to 14UTC (right) are overlaid. The red arrow indicates the BE.

to AROME output without re-training provided satisfactory results. The quality of the detection, as measured by the object-oriented evaluations, is similar for AROME-EPS and AROME. The success of transfer learning is very promising for future NWP model evolutions, suggesting that re-training the U-Net each time the AROME model changes may not be necessary.

Forecasters have been involved to develop three kinds of synthesis plots to summarize BE detections in AROME-EPS and AROME. They sum up information of more than 700 fields and thus facilitate forecasters' expertise. These plots include a trajectory visualization that presents all detections as a function of their validity time, a paintball visualization with colors according to the member of AROME-EPS and a probability plot including space-time tolerances. To subjectively assess the quality of BE detections, forecasters examine several hundreds of pictures. Their feedback mainly supports the results of the previous sections but more formal evaluations are necessary to validate these findings. To continue the evaluation in a real-time context, a daily production in a research mode is deployed since summer 2021. This daily production is enabled thanks to a short computation time in operations mode. The U-Net predictions (for all AROME-EPS members) take around 10 minutes with a GPU[5].

Future work will explore the addition of FFgust, FF or other parameters inspired by Lagerquist et al. (2017, 2020) as predictors of a new U-Net. This new U-Net could be compared with the current optimal U-Net configuration. Adding a severity attribute based on FFgust in BE objects will be also a priority. Another research direction will explore how to improve the temporal stability of U-Net predictions. Adding pseudo-reflectivity fields at time $t-1h$ and $t+1h$ as inputs of the U-Net could be a possible solution. This study is largely driven by an expert's skill. More objectivity is needed and consequently a panel of experts could hand-label the same images. A probability map for the target value instead of a binary map (0 or 1) could be tested. A U-Net could also be tested and tuned to apply on radar data in order to improve the nowcasting of MCSs. These observed BEs could be used for evaluating the quality of AROME-EPS and AROME BE forecasts. Finally, the U-Net approach could be extended to detect other kinds of severe convective storms. Even if the methodology remains similar, the number of cases should be sufficient to train AI methods. The input size of the U-Net could be modified according to the size of the convective storms. Some additional predictors may be deemed necessary (Updraft Helicity for supercell detection for instance).

Acknowledgments. The authors thank Sébastien Baille, Romaric Cinotti, Pierre-Emmanuel Gallerand, Jean Imbert, Julien Léger and Martin Paccaud for their feedback about U-Net that used parts of their spare time. The authors thank also Fabrice Guillemot for his help to selecting relevant parameters for this study. The authors acknowledge Jean-Antoine Maziejewski and Gregory Roux for proofreading and for improving English in the manuscript.

Data availability statement. The three datasets (training, validation and deterministic AROME databases) and the weights of the optimal U-Net are available online in open access (Mounier (2021) , doi : `10.5281/zenodo.5534445`, license : Etalab Open License 2.0)

[5]NVIDIA Corporation GP102 (GeForce GTX 1080 Ti)

References

Abadi, M., and Coauthors, 2015: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. URL https://www.tensorflow.org/, software available from tensorflow.org.

Atkins, N., and M. S. Laurent, 2009: Bow echo mesovortices. part i: Processes that influence their damaging potential. *Mon. Wea. Rev.*, **137**, 1497–1513, doi:10.1175/2008MWR2649.1.

Biard, J., and K. Kunkel, 2019: Automated detection of weather fronts using a deep learning neural network. *Advances in Statistical Climatology, Meteorology and Oceanography*, **5**, 147–160, doi:10.5194/ascmo-5-147-2019.

Bouttier, F., and L. Raynaud, 2018: Clustering and selection of boundary conditions for limited-area ensemble prediction. *Quart. J. Roy. Meteor. Soc.*, **144**, 2381–2391, doi:10.1002/qj.3304.

Bouttier, F., L. Raynaud, O. Nuissier, and B. Ménétrier, 2016: Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. *Quart. J. Roy. Meteor. Soc.*, **142**, 390–403, doi:10.1002/qj.2622.

Bouttier, F., B. Vié, O. Nuissier, and L. Raynaud, 2012: Impact of stochastic physics in a convection-permitting ensemble. *Mon. Wea. Rev.*, **140**, 3706–3721, doi:10.1175/MWR-D-12-00031.1.

Bowler, N., A. Arribas, K. Mylne, K. Robertson, and S. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134**, 703–722, doi:10.1002/qj.234.

Brousseau, P., Y. Seity, D. Ricard, and J. Léger, 2016: Improvement of the forecast of convective activity from the arome-france system. *Quart. J. Roy. Meteor. Soc.*, **142**, 2231–2243, doi:10.1002/qj.2822.

Caumont, O., and Coauthors, 2006: A radar simulator for high-resolution nonhydrostatic models. *J. Atmos. Oceanic Technol.*, **23**, 1049–1067, doi:10.1175/JTECH1905.1.

Chawla, N., N. Japkowicz, and A. Kotcz, 2004: Editorial: Special issue on learning from imbalanced data sets. *SIGKDD Explorations Newsletter*, **6**, 1–6, doi:10.1145/1007730.1007733.

Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. part i: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, doi:10.1175/MWR3145.1.

Demuth, J., and Coauthors, 2020: Recommendations for developing useful and usable convection-allowing model ensemble information for nws forecasters. *Wea. Forecasting*, **35**, 1381–1406, doi:10.1175/WAF-D-19-0108.1.

Done, J., C. Davis, and M. Weisman, 2004: The next generation of NWP: explicit forecasts of convection using the weather research and forecasting (wrf) model. *Atmos. Sc. Lett.*, **5**, 110–117, doi:10.1002/asl.72.

Du, J., G. DiMego, M. Tracton, and B. Zhou, 2003: NCEP short-range ensemble forecasting (SREF) system: multi-IC, multi-model and multi-physics approach. *Research activities in atmospheric and oceanic modelling*, **33**.

Dutta, A., and A. Zisserman, 2019: The VGG image annotator (VIA). *ArXiv*, **abs/1904.10699**.

Ebert, E., 2008: Fuzzy verification of high?resolution gridded forecasts: a review and proposed framework. *Meteor. Appl.*, **15**, 51–64, doi:10.1002/met.25.

Elhoseiny, M., S. Huang, and A. Elgammal, 2015: Weather classification with deep convolutional neural networks. 3349–3353, doi:10.1109/ICIP.2015.7351424.

French, A., and M. Parker, 2014: Numerical simulations of bow echo formation following a squall line supercell merger. *Mon. Wea. Rev.*, **142**, 4791–4822, doi:10.1175/MWR-D-13-00356.1.

Fujita, T., 1978: *Manual of downburst identification for project NIMROD*. Satellite and Mesometeorology Research Paper 156, 104 pp.

Gagne II, D., S. Haupt, D. Nychka, and G. Thompson, 2019: Interpretable deep learning for spatial analysis of severe hailstorms. *Mon. Wea. Rev.*, **147**, 2827–2845, doi:10.1175/MWR-D-18-0316.1.

Gallo, B., A. Clark, B. Smith, R. Thompson, I. Jirak, and S. Dembek, 2019: Incorporating UH Occurrence Time to Ensemble-Derived Tornado Probabilities. *Wea. Forecasting*, **34**, 151–164, doi:10.1175/WAF-D-18-0108.1.

Goulet, L., 2015: Bow echoes : Conceptual schemes and European relevance. *The European Forecaster*, URL http://www.euroforecaster.org/newsletter20/meteofr2.

Hohenegger, C., and C. Schär, 2007: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteor. Soc.*, **88**, 1783–1794, doi:10.1175/BAMS-88-11-1783.

Houze Jr., R., 2004: Mesoscale convective systems. *Reviews of Geophysics*, **42**, doi:10.1029/2004RG000150.

Jergensen, G., A. McGovern, R. Lagerquist, and T. Smith, 2020: Classifying convective storms using machine learning. *Wea. Forecasting*, **35**, 537–559, doi:10.1175/WAF-D-19-0170.1.

Kain, J., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952, doi:10.1175/WAF2007106.1.

Kamani, M., F. Farhat, S. Wistar, and J. Wang, 2016: Shape matching using skeleton context for automated bow echo detection. doi:10.1109/BigData.2016.7840685.

Kingma, D., and J. Ba, 2014: Adam: A method for stochastic optimization. *International Conference on Learning Representations*.

Klimowski, B., M. Hjelmfelt, and M. Bunkers, 2004: Radar observations of the early evolution of bow echoes. *Wea. Forecasting*, **19**, 727–734, doi:10.1175/1520-0434(2004)019⟨0727:ROOTEE⟩2.0.CO;2.

Kurth, T., and Coauthors, 2018: Exascale deep learning for climate analytics. 649–660.

Lagerquist, R., A. McGovern, and D. Gagne II, 2019: Deep learning for spatially explicit prediction of synoptic-scale fronts. *Wea. Forecasting*, **34**, 1137–1160, doi:10.1175/WAF-D-18-0183.1.

Lagerquist, R., A. McGovern, C. Homeyer, D. Gagne II, and T. Smith, 2020: Deep learning on three-dimensional multiscale data for next-hour tornado prediction. *Mon. Wea. Rev.*, **148**, 2837–2861, doi:10.1175/MWR-D-19-0372.1.

Lagerquist, R., A. McGovern, and T. Smith, 2017: Machine learning for real-time prediction of damaging straight-line convective wind. *Wea. Forecasting*, **32**, 2175–2193, doi:10.1175/WAF-D-17-0038.1.

Laing, A., and M. Fritsch, 1997: The global population of mesoscale convective complexes. *Quart. J. Roy. Meteor. Soc.*, **123**, 389–405, doi:10.1002/qj.49712353807.

LeCun, Y., and Y. Bengio, 1995: Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, **3361**, 1995.

Liu, Y., E. Racah, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, and W. Collins, 2016: Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *arXiv*.

Markowski, P., and Y. Richardson, 2010: *Mesoscale Meteorology in Midlatitudes*. John Wiley and Sons, Ltd, 249-265 pp., doi:10.1002/9780470682104.ch9.

Marshall, J., and W. Palmer, 1948: The distribution of raindrops with size. *J. Atmos. Sci.*, **5**, 165–166, doi:10.1175/1520-0469(1948)005⟨0165:TDORWS⟩2.0.CO;2.

Matsuoka, D., S. Sugimoto, Y. Nakagawa, S. Kawahara, F. Araki, Y. Onoue, M. Iiyama, and K. Koyamada, 2019: Automatic detection of stationary fronts around Japan using a deep convolutional neural network. *SOLA*, **15**, doi:10.2151/sola.2019-028.

McGovern, A., K. Elmore, D. Gagne II, S. Haupt, C. Karstens, R. Lagerquist, T. Smith, and J. Williams, 2017: Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Amer. Meteor. Soc.*, **98**, 2073–2090, doi:10.1175/BAMS-D-16-0123.1.

Moller, A., C. Doswell III, M. Foster, and G. Woodall, 1994: The operational recognition of supercell thunderstorm environments and storm structures. *Wea. Forecasting*, **9**, 327–347, doi:10.1175/1520-0434(1994)009⟨0327:TOROST⟩2.0.CO;2.

Montmerle, T., Y. Michel, E. Arbogast, B. Ménétrier, and P. Brousseau, 2018: A 3D ensemble variational data assimilation scheme for the limited-area AROME model: Formulation and preliminary results. *Quart. J. Roy. Meteor. Soc.*, **144**, 2196–2215, doi:10.1002/qj.3334.

Mounier, A., 2021: Detection of bow echoes in French kilometre-scale models (AROME-EPS & AROME models of Météo-France). Zenodo, doi:10.5281/zenodo.5534445.

Patil, V., S. Das, and A. Phadke, 2019: Methods for mesoscale convective systems detection and tracking: a survey. 1–7, doi:10.1109/ICCCNT45670.2019.8944656.

Peralta, C., Z. Ben Bouallègue, S. Theis, C. Gebhardt, and M. Buchhold, 2012: Accounting for initial condition uncertainties in COSMO-DE-EPS. *J. Geophys. Res.*, **117**, 7108–, doi:10.1029/2011JD016581.

Przybylinski, R., 1995: The bow echo: Observations, numerical simulations, and severe weather detection methods. *Wea. Forecasting*, **10**, 203–218, doi:10.1175/1520-0434(1995)010⟨0203:TBEONS⟩2.0.CO;2.

Raynaud, L., and F. Bouttier, 2017: The impact of horizontal resolution and ensemble size for convective-scale probabilistic forecasts. *Quart. J. Roy. Meteor. Soc.*, **143**, 3037–3047, doi:https://doi.org/10.1002/qj.3159.

Raynaud, L., I. Pechin, P. Arbogast, L. Rottner, and M. Destouches, 2019: Object-based verification metrics applied to the evaluation and weighting of convective-scale precipitation forecasts. *Quart. J. Roy. Meteor. Soc.*, **145**, 1992–2008, doi:10.1002/qj.3540.

Roberts, B., I. Jirak, A. Clark, S. Weiss, and J. Kain, 2019: Postprocessing and visualization techniques for convection-allowing ensembles. *Bull. Amer. Meteor. Soc.*, **100**, 1245–1258, doi:10.1175/BAMS-D-18-0041.1.

Ronneberger, O., P. Fischer, and T. Brox, 2015: U-Net: Convolutional networks for biomedical image segmentation. *LNCS*, **9351**, 234–241, doi:10.1007/978-3-319-24574-4_28.

Rossa, A., P. Nurmi, and E. Ebert, 2008: Overview of methods for the verification of quantitative precipitation forecasts. doi:10.1007/978-3-540-77655-0_16.

Schumacher, R., A. Hill, M. Klein, J. Nelson, M. Erickson, S. Trojniak, and G. Herman, 2021: From random forests to flood forecasts: A research to operations success story. *Bull. Amer. Meteor. Soc.*, **102**, 1742–1755, doi:10.1175/BAMS-D-20-0186.1.

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, **139**, 976–991, doi:10.1175/2010MWR3425.1.

Sobash, R., G. Romine, and C. Schwartz, 2020: A Comparison of Neural-Network and Surrogate-Severe Probabilistic Convective Hazard Guidance Derived from a Convection-Allowing Model. *Wea. Forecasting*, **35**, 1981–2000, doi:10.1175/WAF-D-20-0036.1.

Sobash, R., C. Schwartz, G. Romine, K. Fossell, and M. Weisman, 2016: Severe weather prediction using storm surrogates from an ensemble forecasting system. *Wea. Forecasting*, **31**, 255–271, doi:10.1175/WAF-D-15-0138.1.

Trapp, R., S. Tessendorf, E. Godfrey, and H. Brooks, 2005: Tornadoes from squall lines and bow echoes. part I: Climatological distribution. *Wea. Forecasting*, **20**, 23–34, doi:10.1175/WAF-835.1.

7.3 Supplément de l'article

Generated using the official AMS LaTeX template v5.0

Detection of bow echoes in kilometer-scale forecasts using a convolutional neural network (Supplemental Material)

Arnaud MOUNIER *, Laure RAYNAUD, Lucie ROTTNER, Matthieu PLU

CNRM, University of Toulouse, Météo-France, CNRS, Toulouse, France

Philippe ARBOGAST

Forecasting department, Météo-France, Toulouse, France

Michaël KREITZ

National meteorological school, Météo-France, Toulouse, France

Léo MIGNAN and Benoît TOUZÉ

Forecasting department, Météo-France, Rennes and Toulouse, France

* *Corresponding author*: Arnaud MOUNIER, arnaud.mounier@meteo.fr

A1. Procedure to collect forecasters' feedback

As explained in section 6.c, a group of ten forecasters has examined several meteorological cases with BEs from the validation AROME-EPS dataset:

- BE cases: 8 August 2017, 26 May 2018, 4 June 2019, 5 June 2019
- squall lines: 2 January 2018
- quasi-stationary convective systems: 5 August 2019

and cases from the deterministic AROME model:

- BE cases: 8 August 2017, 26 May 2018, 4 June 2019, 5 June 2019
- squall lines: 2 January 2018
- quasi-stationary convective systems: 5-6 August 2019
- supercells: 15 June 2019, 26 July 2019, 8-9 August 2019, 18 August 2019

A single case has been removed compared with the validation dataset described in section 2.b.2: the squall line case of 3 December 2018. On this case, the U-Net does not detect any BE and consequently this case has no impact on the HR and FAR defined in the section 2.h (i.e. this case is not relevant for the evaluation). Removing this case limits also the number of pictures to examine. Some cases have been added (supercell cases) in the deterministic AROME model dataset compared with the validation dataset in section 5 because the number of pictures was too limited after the split of the still small deterministic AROME dataset into three smaller datasets. Consequently, the scores of section 5 are not comparable with the ones of this section and are not mentioned in the section 6.c. The group of ten forecasters is also divided into three subgroups (group 1: 4 forecasters, group

2: 3 forecasters, and group 3: 3 forecasters). Regarding AROME-EPS dataset, the 12[1] or 16 members of AROME-EPS are split into three groups to assign a third of AROME-EPS runs to each group:

- group 1: member 1 to 4 (or 1-5 if 16 members)
- group 2: member 5 to 8 (or 6-10 if 16 members)
- group 3: member 9 to 12 (or 11-16 if 16 members)

This method for separating the AROME-EPS members is basic because the AROME-EPS members are interchangeable and independent. Concerning the deterministic AROME dataset, the previous five initializations before the event are available and divided between the three expert groups (2 initializations, 2 initializations and 1 initialization). Only one initialization is assigned to one of the three groups, but this group change according to the date. For instance in the evaluation form below, only one AROME initialization is available for the BE case in 2018 (26 May 2018 at 0h UTC), but two initializations in 2017 (7 August 2017 at 0h and 3h UTC).

Each forecaster is asked to fill in an evaluation form. The form assigned to group 1 is added below for instance. Each raw corresponds to one case, one runtime, one member (or the deterministic AROME model), and multiple lead times. The forecaster should write the number of pictures with correct detections of BEs, the number of pictures with false alarms, and the number of pictures with misses based on the pseudo-reflectivity fields and the BE detections of the U-Net. The forecaster can add in the last column 'Comments' the exact lead times for the correct detections, false alarms, misses or any other relevant comment. At the end of the evaluation, a HR, FAR and CSI are computed for each forecaster (Table.S1).

[1]The AROME-EPS is a 16-member ensemble only since July 2019

The HR and FAR vary greatly from one forecaster to another (FAR of '1-C' forecaster is particularly high). However, the CSI is around 0.5-0.6 according to the majority of forecasters because the HR and FAR are correlated (i.e. high HR are most of the time associated with high FAR and low HR with low FAR). Some explanations concerning this behavior are proposed in the section 6.c (second item). However, the 'general comments' reported at the end of the evaluation form are similar with the same conclusions. These conclusions are reported in the section 6.c.

Evaluation grid : Group 1 (Mb 1 to 4) - AROME-EPS

CD : Correct detections of BE, FA : False alarms, M : Misses

Date	Run	MB	Nb_images with CD	Nb_images with FA	Nb_images with M	Comments
07/08/2017	09h	1				
07/08/2017	09h	2				
07/08/2017	09h	3				
07/08/2017	09h	4				
07/08/2017	21h	1				
07/08/2017	21h	2				
07/08/2017	21h	3				
07/08/2017	21h	4				
25/05/2018	15h	1				
25/05/2018	15h	2				
25/05/2018	15h	3				
25/05/2018	15h	4				
25/05/2018	21h	1				
25/05/2018	21h	2				
25/05/2018	21h	3				
25/05/2018	21h	4				
26/05/2018	03h	1				
26/05/2018	03h	2				
26/05/2018	03h	3				
26/05/2018	03h	4				
26/05/2018	09h	1				
26/05/2018	09h	2				
26/05/2018	09h	3				
26/05/2018	09h	4				
03/06/2019	15h	1				
03/06/2019	15h	2				
03/06/2019	15h	3				
03/06/2019	15h	4				
03/06/2019	21h	1				
03/06/2019	21h	2				
03/06/2019	21h	3				
03/06/2019	21h	4				
04/06/2019	03h	1				
04/06/2019	03h	2				
04/06/2019	03h	3				
04/06/2019	03h	4				

Date	Run	MB	Nb_images with CD	Nb_images with FA	Nb_images with M	Comments
04/06/2019	09h	1				
04/06/2019	09h	2				
04/06/2019	09h	3				
04/06/2019	09h	4				
04/06/2019	21h	1				
04/06/2019	21h	2				
04/06/2019	21h	3				
04/06/2019	21h	4				
05/06/2019	03h	1				
05/06/2019	03h	2				
05/06/2019	03h	3				
05/06/2019	03h	4				
05/06/2019	09h	1				
05/06/2019	09h	2				
05/06/2019	09h	3				
05/06/2019	09h	4				
05/06/2019	15h	1				
05/06/2019	15h	2				
05/06/2019	15h	3				
05/06/2019	15h	4				
02/01/2018	09h	1				
02/01/2018	09h	2				
02/01/2018	09h	3				
02/01/2018	09h	4				
05/08/2019	15h	1				
05/08/2019	15h	2				
05/08/2019	15h	3				
05/08/2019	15h	4				
05/08/2019	15h	5				
Total AROME-EPS			0	0	0	

HR	#DIV/0!
FAR	#DIV/0!
CSI	#DIV/0!

Evaluation grid : Group 1 (Mb 1 to 4) - determinsitic AROME

CD : Correct detections of BE, FA : False alarms, M : Misses

64

Date	Run	MB	Nb_images with CD	Nb_images with FA	Nb_images with M	Comments
07/08/2017	0h	ARO				
07/08/2017	3h	ARO				
26/05/2018	0h	ARO				
04/06/2019	3h	ARO				
04/06/2019	6h	ARO				
04/06/2019	18h	ARO				
05/06/2019	0h	ARO				
02/01/2018	12h	ARO				
02/01/2018	18h	ARO				
06/08/2019	6h	ARO				
06/08/2019	12h	ARO				
15/06/2019	0h	ARO				
26/07/2019	3h	ARO				
26/07/2019	6h	ARO				
08/08/2019	18h	ARO				
09/08/2019	0h	ARO				
18/08/2019	0h	ARO				
Total Arome			0	0	0	

HR	#DIV/0!

FAR	#DIV/0!

CSI	#DIV/0!

General comments

LIST OF TABLES

Table S1. Subjective scores of ten forecasters for the AROME-EPS dataset (EPS) and the deterministic AROME dataset (ARO). The first column is an ID for each forecaster (group number and letter from 'A' to 'C' or 'D'). The HR, FAR and CSI for every dataset are reported. . . 9

TABLE S1. Subjective scores of ten forecasters for the AROME-EPS dataset (EPS) and the deterministic AROME dataset (ARO). The first column is an ID for each forecaster (group number and letter from 'A' to 'C' or 'D'). The HR, FAR and CSI for every dataset are reported.

Forecaster	HR (EPS)	FAR (EPS)	CSI (EPS)	HR (ARO)	FAR (ARO)	CSI (ARO)
1-A	0.56	0.03	0.55	0.64	0.06	0.61
1-B	0.83	0.37	0.56	0.80	0.06	0.76
1-C	0.61	0.40	0.43	0.77	0.53	0.41
1-D	0.87	0.27	0.66	0.85	0.27	0.65
2-A	0.61	0.13	0.56	0.51	0.15	0.47
2-B	0.68	0.11	0.63	0.73	0.16	0.64
2-C	0.44	0.05	0.43	0.54	0.16	0.49
3-A	0.82	0.21	0.67	0.60	0.17	0.53
3-B	0.56	0.08	0.53	0.62	0.14	0.56
3-C	0.72	0.21	0.60	0.45	0.14	0.42

LIST OF FIGURES

Fig. S1. The entire domain of AROME and AROME-EPS 11

Fig. S2. All hand-labeled BEs corresponding to 21/09/2019. 12

Fig. S3. Example of correct detections in the validation dataset. BE contours and pseudo-reflectivity fields are overlaid. Hand-labeled objects are magenta contours. U-Net objects are blue contours. 13

Fig. S4. Example of false alarms (blue contours) in the validation dataset. BE contours and pseudo-reflectivity fields are overlaid. 14

Fig. S5. Example of misses (magenta contours) in the validation dataset. BE contours and pseudo-reflectivity fields are overlaid. 15

FIG. S1. The entire domain of AROME and AROME-EPS

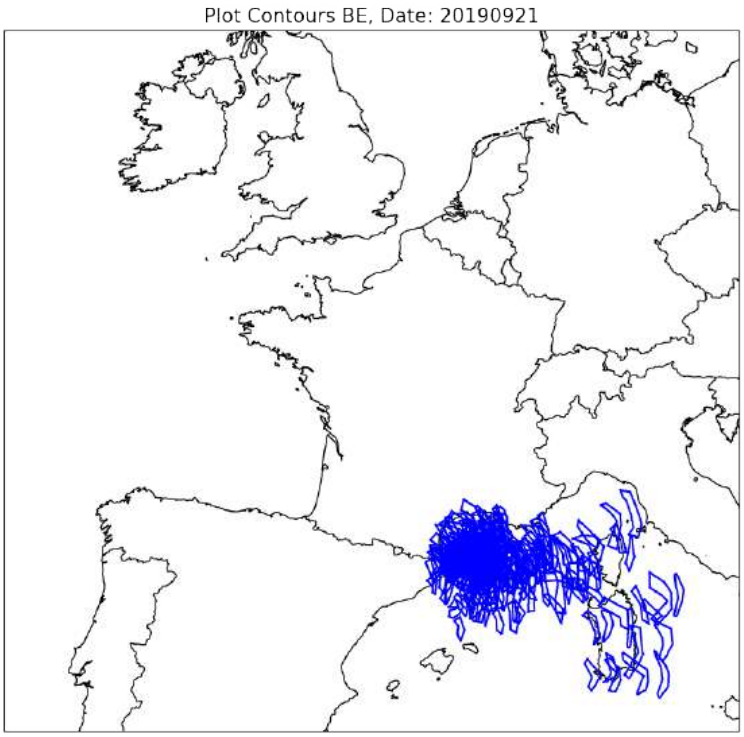


FIG. S2. All hand-labeled BEs corresponding to 21/09/2019.

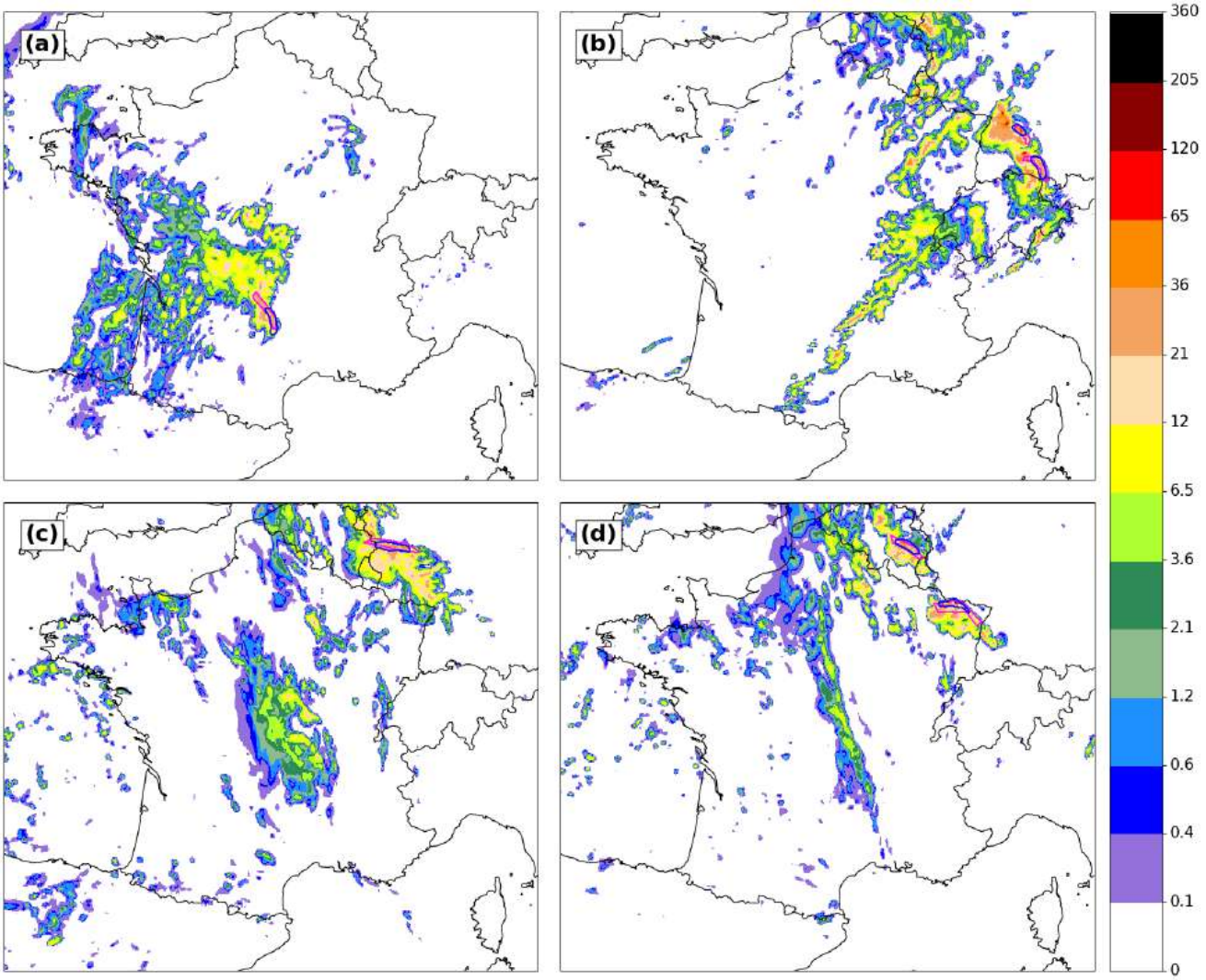
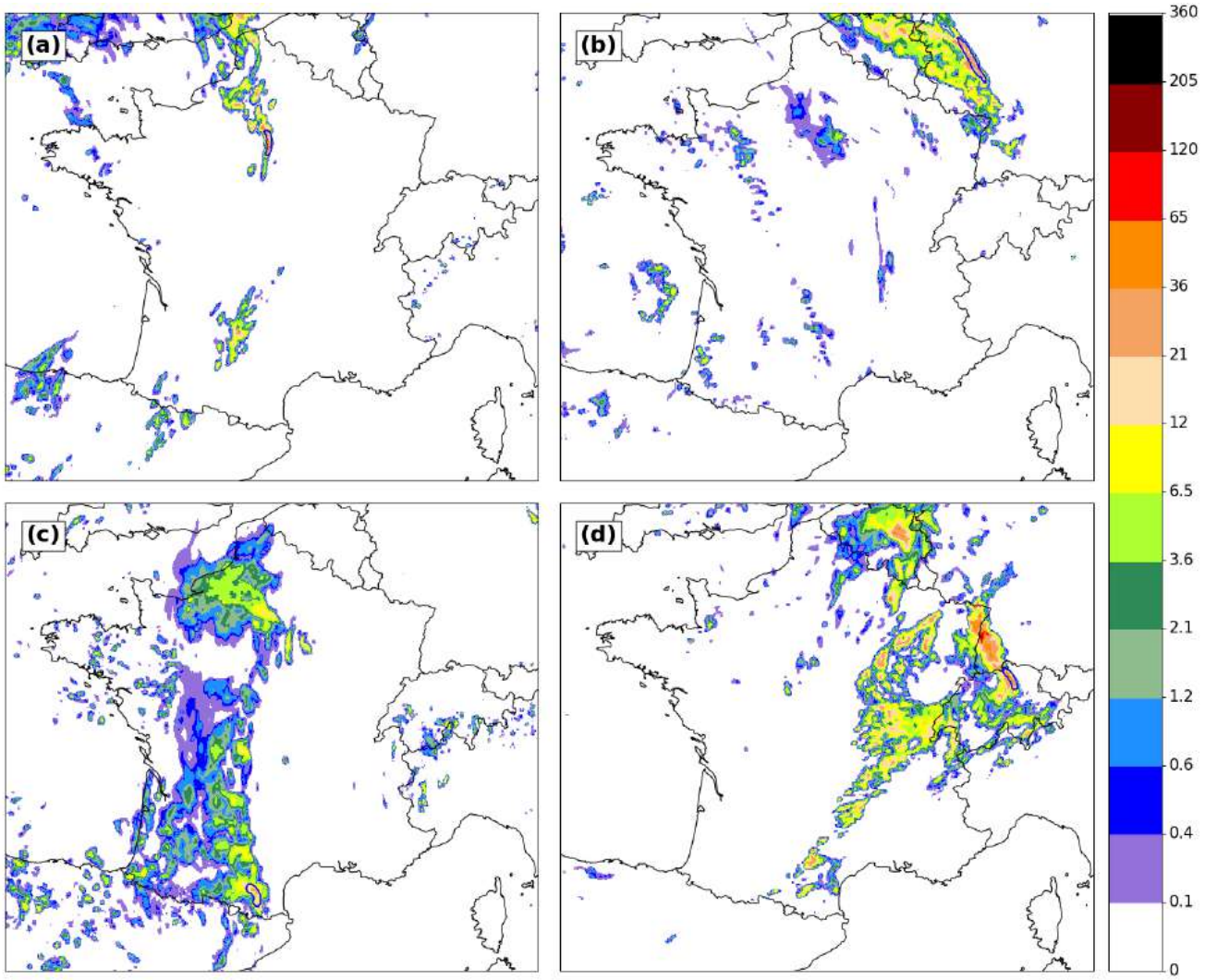


FIG. S3. Example of correct detections in the validation dataset. BE contours and pseudo-reflectivity fields are overlaid. Hand-labeled objects are magenta contours. U-Net objects are blue contours.



86 FIG. S4. Example of false alarms (blue contours) in the validation dataset. BE contours and
87 pseudo-reflectivity fields are overlaid.

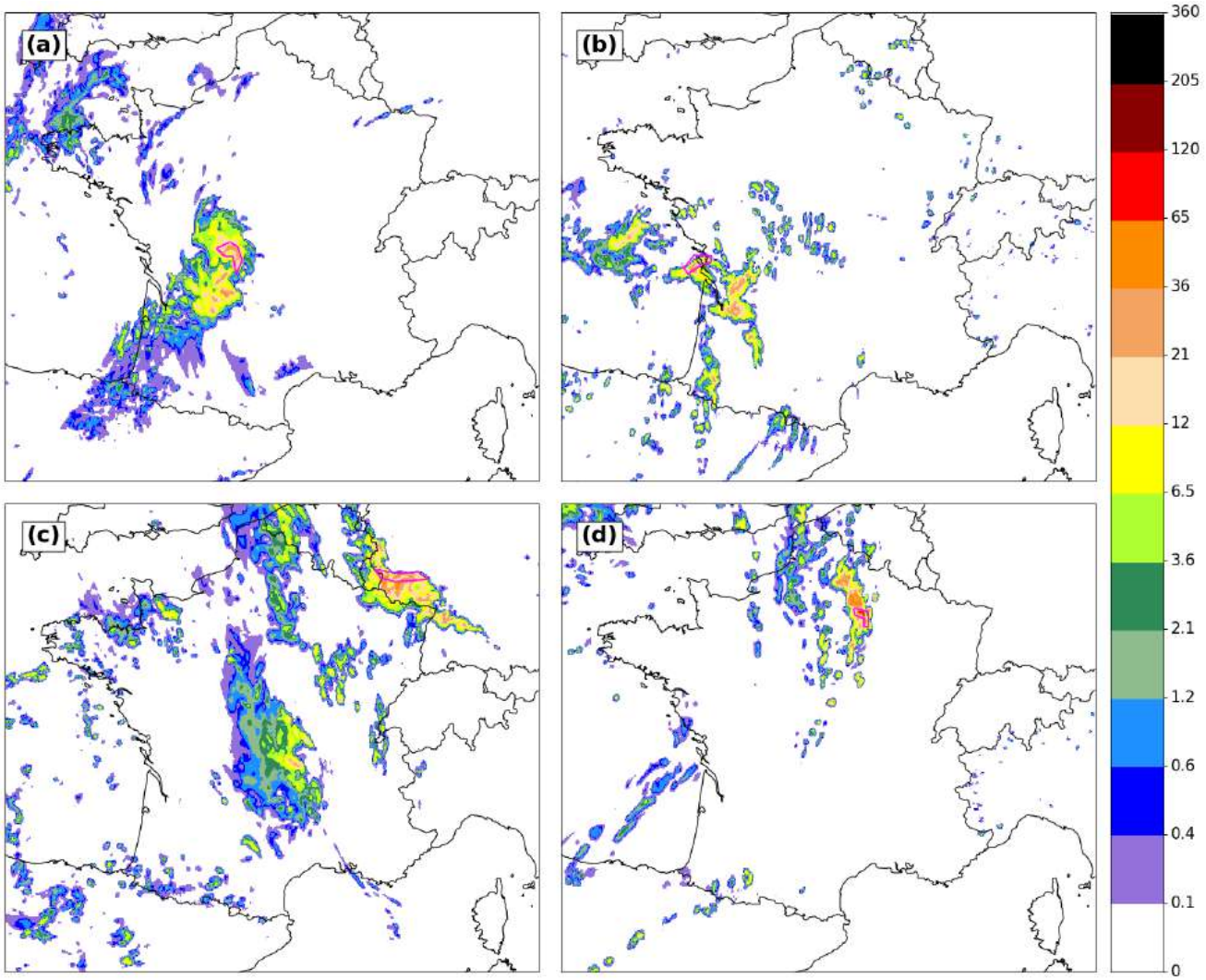


FIG. S5. Example of misses (magenta contours) in the validation dataset. BE contours and pseudo-reflectivity fields are overlaid.

Chapitre 8

Performance des modèles

Le chapitre précédent a présenté un U-Net qui détecte automatiquement les échos arqués simulés dans PE-AROME et AROME puis la création de quelques graphiques de synthèse pour donner un aperçu global du risque d'échos arqués selon les modèles. L'évaluation a porté sur les performances des réseaux de neurones et leur capacité à discriminer correctement les échos arqués, mais la question de la performance des modèles de prévision (simulent-ils des échos arqués sur la bonne zone et au bon moment ?) n'a pas été posée et fait l'objet de ce chapitre. La première partie de ce chapitre est dédiée aux premiers travaux pour développer un U-Net qui détecte les échos arqués dans les données observées, point essentiel pour évaluer les performances des modèles de prévision. Ce travail a été réalisé par Sylvain Galliau lors de son stage de fin d'études que j'ai coencadré avec ma codirectrice de thèse Lucie Rottner. En parallèle de ce stage, des scores objectifs ont été calculés à partir des sorties modèles et des labellisations de Sylvain Galliau dans les données radars, ils font l'objet de la deuxième partie de ce chapitre. Comme évoqué dans la conclusion de l'article du chapitre précédent, une production expérimentale tourne de façon quotidienne depuis mai 2021. Cette production a permis une analyse en temps réel des sorties modèles durant l'été 2021 et peut être utilisée pour répondre en partie à ces interrogations. Le bilan de la production expérimentale sur l'été 2021 fait l'objet de la troisième partie. La quatrième partie de ce chapitre présente l'ajout d'un nouveau graphique qui résume le risque de fortes rafales de vent sous les échos arqués, en réponse aux retours des prévisionnistes sur le bilan de l'été 2021. La dernière partie est consacrée à la comparaison des prévisions faites avec l'ancienne et la nouvelle version des modèles, utilisée en opérationnel depuis l'été 2022.

8.1 Détection des échos arqués dans les observations

L'évaluation de la performance des modèles AROME et PE-AROME nécessite de repérer les échos arqués dans les données observées. Les détections des échos arqués peuvent se faire à l'aide de la mosaïque de réflectivités SERVAL (figure 1.2) issue des radars météorologiques. Le chapitre précédent a montré que le U-Net entraîné sur la PE-AROME pouvait être utilisé sur le modèle déterministe

AROME par transfert d'apprentissage. Ce même transfert a été testé sur la mosaïque SERVAL, après interpolation sur la grille AROME [1]. Plusieurs méthodes d'interpolation ont été tentées mais aucune ne donne des résultats concluants. En particulier, le U-Net produit une grande quantité de fausses alarmes. Malgré des progrès constants en modélisation, les données observées présentent des textures de pluies toujours très différentes des données modèles ce qui peut expliquer cette difficulté dans le transfert d'apprentissage.

Un stage a alors été proposé pour développer un autre U-Net capable de détecter les échos arqués dans les données SERVAL, sur le même principe que le U-Net développé pour PE-AROME et AROME. Le stage de fin d'études réalisé par Sylvain Galliau a commencé par une labellisation manuelle sur des cas d'échos arqués observés en 2018 et 2019. Les cas avant le 1er janvier 2018 ne sont pas étudiés car la mosaïque SERVAL n'existait pas. Ce travail conséquent de labellisation a permis une première évaluation de la performance des modèles PE-AROME et AROME sur ces cas (cf section 8.2). A partir de ces labellisations, Sylvain a employé des méthodes similaires à celles présentées dans le chapitre 5 section 5.3.1, que ce soit pour le pré-traitement des données, l'entraînement ou les scores des U-Nets. Il a commencé par entraîner des U-Nets en repartant des poids du U-Net optimal avec des tailles de vignettes 48×48, puis en utilisant des U-Nets avec une taille 96×96. Il a tenté de ré-entraîner seulement les dernières couches ou le réseau de neurones complet. Toutes ces expériences n'ont pas apporté satisfaction avec toujours de trop nombreuses fausses alarmes.

De nouveaux U-Nets ont donc été entraînés en partant de poids aléatoires. Sylvain a modifié la méthode d'augmentation de données en amplitude en multipliant chaque vignette par 0.9 et 1.1 au lieu de 0.75 dans les U-Nets du chapitre précédent **(Expérience 1)**. Le meilleur réseau de neurones a alors un très haut taux de bonnes détections (0.97), mais un taux de fausses alarmes élevé (0.3). Sylvain a ensuite modifié la taille des vignettes en essayant des entrées 128×128, il a également ajouté une couche supplémentaire aux U-Nets pour les rendre plus profonds **(Expérience 2)**. Ces modifications ont permis de faire chuter le taux de fausses alarmes (0.17) mais le taux de détection a également baissé (0.73). Les dernières modifications apportées concernent l'ajout d'une augmentation de données par rotation (à 90°, 180° et 270°, **Expérience 3**), mais la rotation à 180° proposant des orientations d'échos arqués très peu probables en France métropolitaine (du Nord vers le Sud ou de l'Est vers l'Ouest), cette rotation sera supprimée dans les dernières expériences **(Expérience 4)**. Ces rotations provoquent une augmentation du taux de bonnes détections (0.95) et de fausses alarmes (0.33) par rapport aux dernières expériences citées.

Les résultats des 4 expériences sont illustrés figure 8.1. Toutes les expériences sont capables de détecter l'écho arqué mais la forme diffère sensiblement. Les expériences 1 et 4 sont sur ce cas les plus satisfaisantes car elles englobent bien la partie Ouest de l'arc et ne présentent pas de détection discontinue comme la numéro 3. Il est en tout cas indéniable que les résultats de ces expériences sont visuellement bien plus satisfaisants que les premiers tests de transfert par apprentissage (non montré). Ces résultats sont encourageants mais on observe malheureusement encore trop de fausses alarmes pour envisager une utilisation en condition opérationnelle en particulier à cause de confusions avec des supercellules. Même si le

1. Celle-ci a une grille de 1 km contre 2.5 km pour AROME

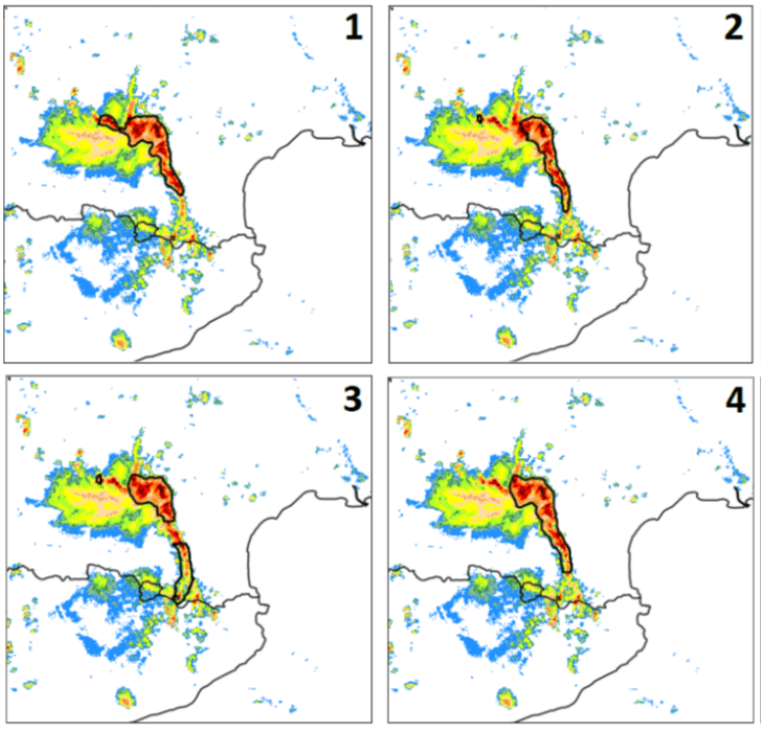


FIGURE 8.1 – *Mosaïque SERVAL du 17 juin 2021 à 15h UTC zoomée sur la région Occitanie. Les contours noirs représentent les détections de l'écho arqué par 4 expériences différentes dont les configurations sont précisées dans la section 8.1. Ces images sont extraites du rapport de stage de Sylvain Galliau.*

taux de fausses alarmes est du même ordre de grandeur que celui dans les modèles de prévision, cet outil doit servir de référence pour évaluer le modèle et n'a à sa disposition qu'une seule image pour décider si oui ou non on observe un écho arqué. A l'inverse, pour la même heure, 16 membres PE-AROME proposent une prévision. Quelques fausses alarmes ou non détections peuvent un peu modifier les cartes de probabilité mais l'impact de ces erreurs est beaucoup plus faible. C'est donc seulement une exigence plus forte sur les données observées qui nécessite de futurs travaux pour encore améliorer ces résultats avant d'envisager de construire des scores pour PE-AROME ou AROME à partir de cet outil.

8.2 Scores objectifs PE-AROME et AROME

8.2.1 Méthodologie

Grâce aux labellisations réalisées sur la mosaïque SERVAL, il est possible d'évaluer objectivement les performances de PE-AROME et AROME pour prévenir le risque d'échos arqués. Pour rappel, l'objectif est d'évaluer la capacité des modèles PE-AROME et AROME à anticiper correctement les échos arqués et non la capacité du réseau de neurones à correctement reconnaître un écho arqué dans ces mêmes modèles. Ce point était l'objet du chapitre 7, section 4 et 5 de l'article. La base de données pour établir ces scores est constituée de 15 dates :

- *Avec échos arqués :* 26 Mai 2018, 9 août 2018, 4 juin 2019, 5 juin 2019, 15

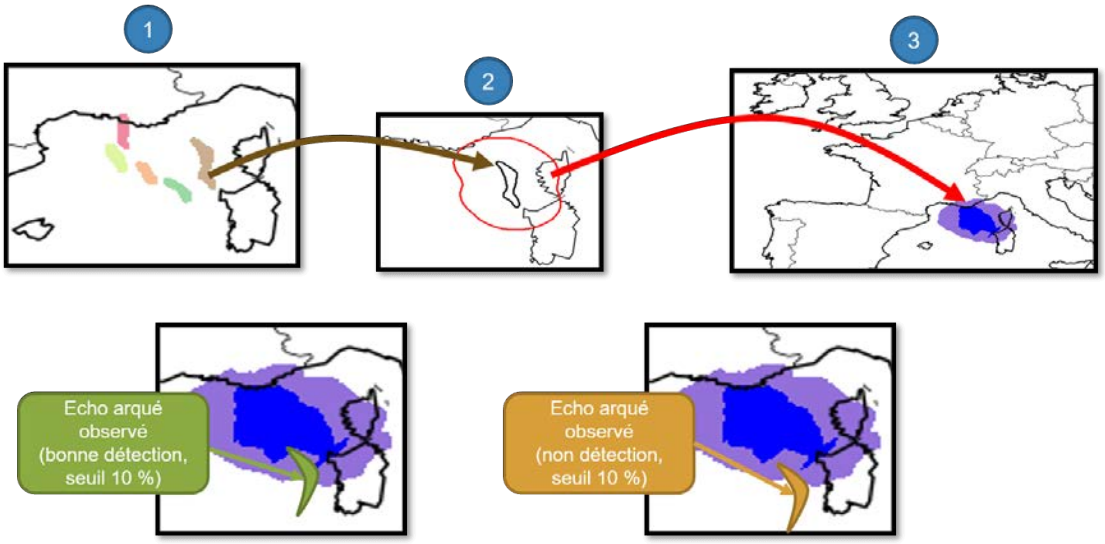


FIGURE 8.2 – *(première ligne) Représentation schématique de la construction des cartes de probabilités du risque d'échos arqués. L'étape 1 consiste à détecter les différents échos arqués dans les membres PE-AROME (une couleur par membre). L'étape 2 consiste pour chaque objet à calculer son voisinage (contour rouge). Les voisinages de chaque objet sont alors superposés pour définir la carte de probabilité à l'étape 3 construite avec une tolérance spatiale (en plus d'une tolérance temporelle de plus ou moins 1h). Plus de détails dans la section 6 de l'article qui constitue la chapitre précédent.*
(deuxième ligne) Représentation schématique de la vérification des performances de PE-AROME. A gauche, si on prend le seuil de 10% (zone bleu foncé), l'écho arqué intersecte cette zone et est donc comptabilisée comme une bonne détection. Sur l'image de droite, il n'intersecte pas cette zone et est donc considéré comme non détecté.

juin 2019, 6 juillet 2019, 11 août 2019, 21 septembre 2019, 14 octobre 2019, 23 octobre 2019

- *Avec convection sévère mais pas d'écho arqué :* 15 juillet 2019, 26 juillet 2019, 6 août 2019, 9 août 2019, 18 août 2019

Les 4 derniers réseaux PE-AROME et AROME couvrant ces dates sont pris en considération pour établir ces scores. Pour que ces scores évaluent l'ensemble PE-AROME et non chacun de ses membres, les échos arqués observés sont comparés à la carte de probabilité construite à l'aide des 3 grandes étapes rappelées sur la figure 8.2. La tolérance spatiale ϵ (cercle rouge à l'étape 2) a été fixée à 150 km dans le chapitre précédent, mais plusieurs valeurs seront testées dans cette partie pour évaluer la sensibilité des scores à la tolérance spatiale. Pour chaque prévision, les scores utilisent donc un champ de probabilité $\mathbb{P}(l)$ et un champ binaire $\mathbb{1}_{SERVAL}$ constitué de 0 et de 1 (1 correspondant aux points de grille dans un écho arqué observé). Le domaine SERVAL (figure 1.2) étant bien plus restreint que le domaine AROME (figure 1.13), le champ $\mathbb{P}(l)$ est masqué en dehors du domaine SERVAL pour permettre la comparaison des deux champs. Un tableau de contingence est calculé comme dans le chapitre 5 (section 5.5.2) à l'aide de ces champs. Le champ de probabilité $\mathbb{P}(l)$ pour un seuil de s_{proba} est compté comme une bonne détection si la zone dépassant s_{proba} intersecte pour au

moins un point de grille les échos arqués du champ $\mathbb{1}_{SERVAL}$ [2]. En prenant un seuil de 10% correspondant à la zone bleu foncé de l'étape 3 figure 8.2, une bonne détection est comptabilisée si au moins un point de grille d'un écho arqué observé est dans cette zone (figure 8.2, en bas à gauche). Si un écho arqué n'est pas observé, une fausse alarme est comptabilisée. Inversement si un écho arqué est observé mais que la zone de probabilité comprenant l'écho arqué ne dépasse pas ce seuil de 10%, une non détection est comptabilisée. En cas de non détection et de fausse alarme sur la même image, les deux sont comptabilisés. C'est par exemple le cas en bas à droite figure 8.2 pour un seuil de 10%. Dans les très rares cas où plusieurs échos arqués sont observés simultanément et si un écho arqué est bien prévu et pas l'autre, l'image est simplement comptée comme une bonne détection. Un tableau de contingence est ainsi calculé pour différentes valeurs de s_{proba}, permettant de calculer des TD, TFA et des indices de Jaccard comme dans le chapitre 5 (section 5.5.2). Pour le modèle déterministe AROME, les scores sont faits en comparant la détection avec un voisinage ϵ et le champ $\mathbb{1}_{SERVAL}$. La détection avec un voisinage ϵ se représente comme la zone cerclée en rouge à l'étape 2 de la figure 8.2. AROME a réalisé une bonne détection si un écho arqué est bien observé dans cette zone rouge. En pratique pour permettre la comparaison avec PE-AROME qui a une tolérance temporelle, la zone cerclée en rouge à l'instant t est sommée avec celles calculées aux instants $t-1h$ et $t+1h$. C'est plutôt l'union de ces zones qui est alors comparée à $\mathbb{1}_{SERVAL}$. Les scores sont calculés de façon séparée pour les échéances courtes (inférieures ou égales à 15h, 8.3) et longues (supérieures à 15h, 8.4).

8.2.2 Échéances courtes ($\leqslant$ 15h)

Les courbes de la figure 8.3 montrent une amélioration des scores avec le voisinage ϵ qui varie de 50 à 300 km. L'augmentation de ϵ permet d'élargir les zones concernées par le risque écho arqué et peut aussi se traduire par une augmentation des valeurs des probabilités. En effet, l'augmentation de ϵ se traduit par un élargissement du cercle rouge si on repart de l'étape 2 de la figure 8.2. Les chances d'intersecter d'autres cercles issus d'autres détections augmentent et donc la probabilité d'observer un écho arqué augmente dans ces conditions. L'augmentation de ϵ permet d'augmenter les détections, de réduire les fausses alarmes et ainsi d'augmenter l'indice de Jaccard. Les scores semblent cependant stagner à partir de 150/200 km. L'augmentation de la tolérance spatiale a effectivement ses limites, elle permet de spatialiser le risque associé à une détection mais elle ne permet pas de compenser une absence de détection dans les membres. La saturation des scores indique qu'une valeur de ϵ qui dépasserait les 150/200 km n'est pas utile car elle ne permet aucun gain alors que les cartes de probabilité seront moins précises avec un risque d'échos arqués qui sera beaucoup plus étalé spatialement. **On retrouve donc une tolérance spatiale optimale d'environ 150 km, comme dans le chapitre précédent**.

Si on se concentre sur les courbes de la PE-AROME, le taux de détection décroit lorsque le seuil de probabilité augmente. Il est d'environ 85% avec un voisinage de 150km pour le seuil à 5%, contre environ 40% pour le seuil à 20%. Les taux de fausses alarmes sont eux plus homogènes autour de 40 à 50%. Un point

2. Ce choix d'un seul point de grille pourra être revu au profit de la position du centre de masse à l'intérieur ou non de la zone dépassant s_{proba} dans des travaux futurs.

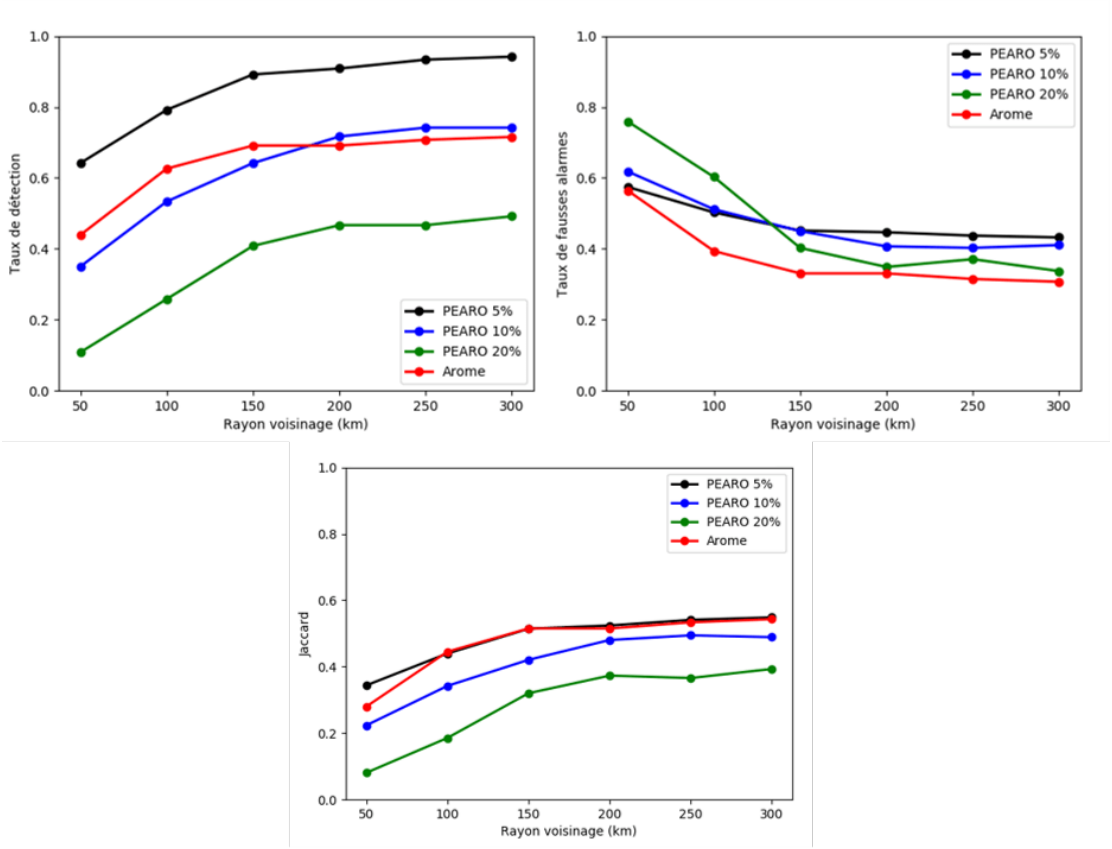


FIGURE 8.3 – *Scores pour les échéances courtes (inférieures ou égales à +15h). Les courbes noires, bleues et vertes représentent les seuils de probabilités à 5%, 10% et 20% pour la PE-AROME. La courbe rouge représente les performances d'AROME. Le graphique en haut à gauche correspond aux TD, le graphique en haut à droite aux TFA et le graphique en bas aux indices de Jaccard. Les courbes sont tracées en fonction de la tolérance spatiale ϵ.*

plus spécifique sera fait concernant ces fausses alarmes dans la partie 8.2.4. Le taux de fausses alarmes reste très élevé même avec un seuil de probabilité élevé car les zones où ce seuil est dépassé sont alors de petites tailles, il est ainsi peu probable que l'écho arqué soit bien observé dans cette zone et cela conduit à une fausse alarme (et une non détection si l'écho arqué est un peu plus loin). **Il ne faut donc pas se focaliser sur les petites zones où le risque est maximal et dépasse les 20% mais bien regarder l'ensemble de l'enveloppe proposée par les membres de la PE-AROME**. Ce sont les zones avec de faibles probabilités qui ont un meilleur score, ce que traduit le score de Jaccard. Il faut cependant préciser que si on se fie au signal avec un seuil de probabilité à 5%, les zones concernées sont très grandes et l'information spatiale peu précise. Enfin le modèle déterministe AROME propose à la fois des taux de détection au niveau de la PE-AROME avec un seuil à 10% tout en ayant les taux de fausses alarmes les plus bas, **ce qui permet à AROME d'avoir des performances comparables à la PE-AROME pour ces échéances courtes**.

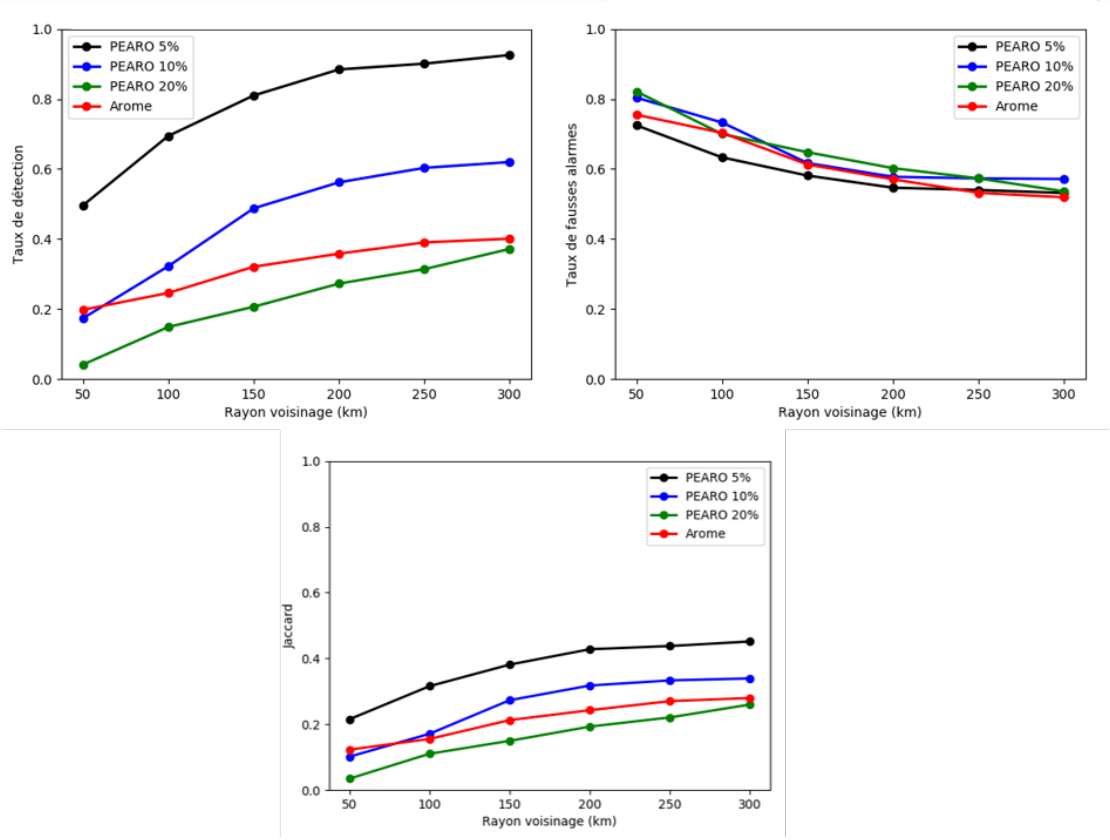


FIGURE 8.4 – *Scores pour les échéances longues (supérieures à +15h). Figure similaire à la figure 8.3.*

8.2.3 Échéances longues (> 15h)

La figure 8.4 pour les échéances longues montre aussi une stagnation des scores et en particulier du taux de fausses alarmes lorsque la tolérance spatiale ϵ augmente. Il semble cependant que le rayon optimal augmente un peu, autour de 200km. Les taux de fausses alarmes sont plus élevés qu'aux courtes échéances. Les taux de détection sont également dégradés, sauf pour le seuil à 5% de la PE-AROME. En lien avec la dégradation de ces performances, le score de Jaccard baisse. Les performances d'AROME se dégradent encore plus fortement que pour la PE-AROME, le taux de détection avec une tolérance de 150km atteint seulement 30% contre plus de 60% aux échéances courtes. Cette dégradation plus forte des performances d'AROME rend la PE-AROME d'autant plus intéressante aux échéances longues. Avec un taux de fausses alarmes semblable, les seuils à 5 et 10% proposent des taux de détections plus importants qu'AROME. **On peut donc conclure que les performances de la PE-AROME et AROME sont comparables aux échéances courtes mais que la PE-AROME propose des performances plus intéressantes aux échéances longues**. Ce résultat traduit bien les limites de la vision déterministe lorsque les échéances augmentent (cf prévisibilité, section 1.3).

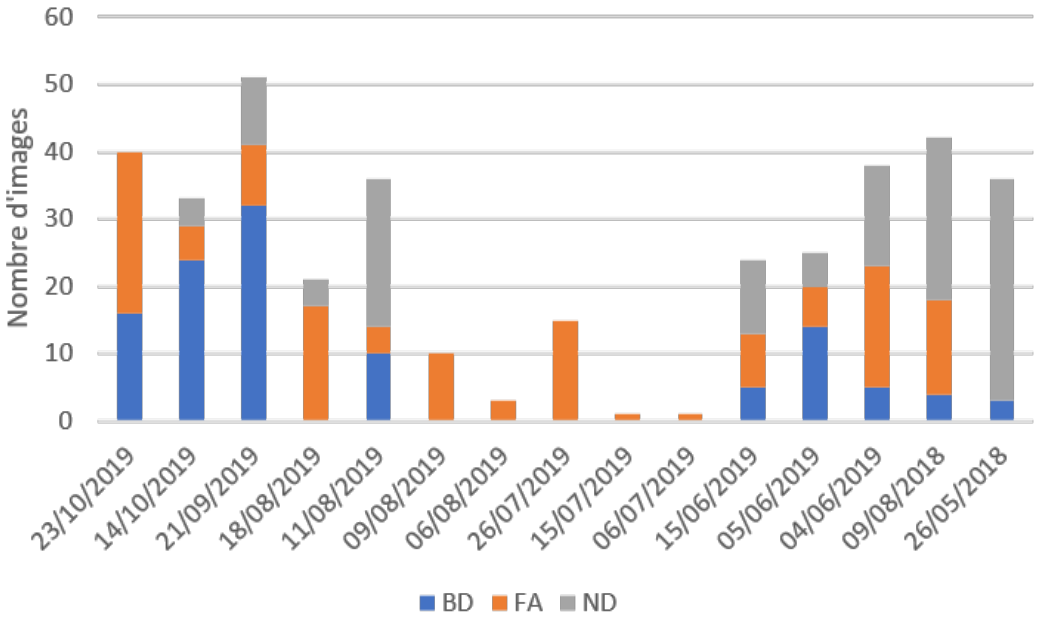


FIGURE 8.5 – *Répartition des bonnes détections (bleu), fausses alarmes (orange) et non détections (gris) selon les 15 dates dans la base de données en prenant un seuil de probabilité à 10% et une tolérance spatiale ϵ de 150 km.*

8.2.4 Étude des fausses alarmes

Le nombre de fausses alarmes est non négligeable que ce soit aux courtes ou aux longues échéances. La figure 8.5 permet d'étudier la répartition de celles-ci selon les 15 dates qui constituent la base de données et sans distinction selon les échéances. La répartition est très inégale et les motifs de ces fausses alarmes variés. Le grand nombre de fausses alarmes pour le cas du 18 août 2019 s'explique par des détections d'échos arqués qui sont principalement en Allemagne. Le voisinage spatial fait alors déborder de fortes probabilités sur le domaine SERVAL qui les prend en considération. En effet, le masque SERVAL a été appliqué sur la carte de probabilité et non sur les détections d'échos arqués, la deuxième proposition donnerait des résultats différents sur ce cas (passage de 17 à 7 fausses alarmes). Pour le 4 juin 2019, le problème est temporel avec une PE-AROME qui est en avance de plusieurs heures par rapport à l'observation. Le problème pour la date du 23 octobre 2019 est encore différent. La PE-AROME propose un écho arqué avec une durée de vie très longue. Un écho arqué sera bien observé comme le montrent les nombreuses bonnes détections mais celui-ci aura une durée de vie plus courte, induisant de nombreuses fausses alarmes. Il n'existe donc pas un unique motif à toutes ces fausses alarmes.

8.3 Bilan de la saison 2021

En parallèle de ces scores calculés sur des situations orageuses en 2018 et 2019, une production expérimentale synthétisant le risque d'échos arqués dans PE-AROME et AROME a été lancée à partir de mai 2021. En temps réel, les

nouveaux réseaux PE-AROME et AROME sont traités et les 3 graphiques de synthèse présentés dans le chapitre précédent sont créés et mis à disposition sur le site de l'équipe LABOratoire de Prévisions (LABO, rattaché à la Prévision Générale). Le site `previlabo.meteo.fr` est un site interne de Météo-France et permet de rendre la production accessible à l'ensemble des prévisionnistes de Météo-France. Un document partagé a permis à plusieurs personnes de faire des retours sur cette production et les performances des modèles sur la saison 2021. L'ensemble des retours de mai à septembre 2021 sont disponibles dans les tableaux C.1 et C.2 en annexe. Les colonnes précisent les dates, les zones concernées par le risque d'échos arqués, la personne qui a rédigé le retour, si après avoir regardé les observations il s'agissait d'une fausse alarme, d'une non détection ou d'une bonne détection, en quoi la PE-AROME apportait une information intéressante ou non et enfin une grande partie « commentaires libres » pour expliquer la situation. En s'appuyant sur ces retours, une réunion avec les différentes personnes qui ont regardé la production expérimentale a été organisée pour dégager des conclusions générales. Il en ressort les points suivants :

- L'été 2021 a été assez riche en situations convectives avec plus de 20 situations reportées entre mai et septembre 2021. Sept contributeurs différents ont participé au document partagé.

- Le nombre de non détections est très faible, mais la méthodologie peut l'expliquer car les retours sont basés sur la réaction des cartes de synthèse à un risque d'écho arqué. Une non détection peut alors plus facilement passer inaperçue qu'une fausses alarme. Ce point avait déjà été soulevé dans les retours des prévisionnistes (section 6 de l'article, chapitre 7).

- Le nombre de fausses alarmes est à peu près équivalent au nombre de bonnes détections. Même si ce point semble cohérent avec la partie précédente, il est cependant difficile de savoir si le problème vient du U-Net qui détecte à tort des échos arqués (taux de fausses alarmes estimé à 39% dans le chapitre précédent) ou si la PE-AROME simule effectivement des échos arqués alors qu'aucun n'est observé. L'ajout d'un nouveau graphique dédié aux rafales de vent sous les échos arqués permettra en partie de différencier les deux et fera l'objet de la section 8.4.

- On observe une situation avec une vigilance orange « orages violents » mais sans aucune réaction du diagnostic. La situation correspondait à un risque de supercellules, cette non réaction était donc logique.

- Si le maximum sur la carte de probabilités n'atteint pas les 10%, le risque semble marginal. A l'inverse, une probabilité maximale de 30 à 40% indique a minima de la convection sévère généralisée.

- La PE-AROME donne en général un signal sans grosse fluctuation d'un réseau à l'autre. Cette stabilité est appréciable pour établir une prévision et semble meilleure que pour le déterministe AROME. Dans certains cas, la synthèse du risque écho arqué donne un balisage pertinent du risque de convection sévère et pas seulement du risque d'échos arqués.

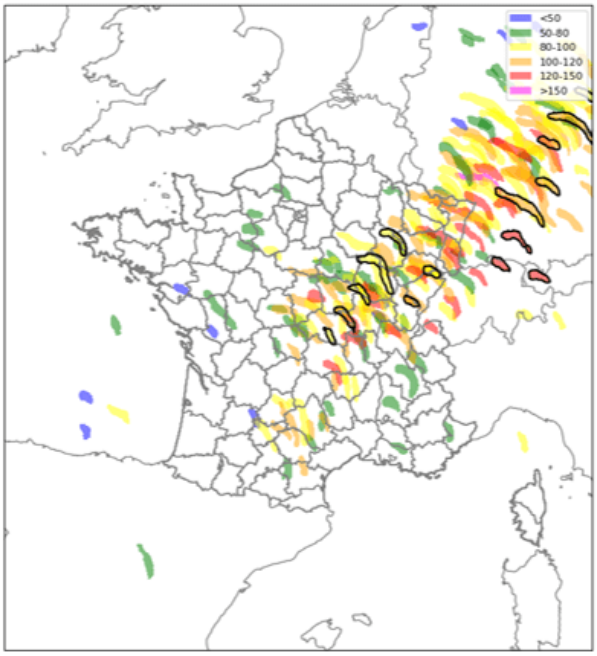


FIGURE 8.6 – *Exemple du nouveau graphique « rafale de vent » pour la journée du 20 juin 2021. Prévision faite par la PE-AROME du 19 juin 2021 à 15h UTC. Les objets entourés en noir correspondent au modèle déterministe AROME. Le code couleur indique les rafales de vent sous les échos arqués en* $km \cdot h^{-1}$.

- Il y a enfin quelques cas atypiques à relever durant cet été. Le diagnostic a fortement réagi sur un seul réseau (donc ici la prévision n'était pas stable) le 25 septembre 2021 à 15h en Méditerranée. Ce fort signal à tort avait disparu au réseau suivant. Un fort signal a également émergé de la PE-AROME très peu de temps avant un évènement convectif, le 5 juillet 2021 à 15h. Ce fort signal était cette fois-ci à raison.

A l'issue de cette réunion, il a été décidé de réitérer l'expérience pour l'été 2022. Celui-ci, beaucoup plus sec et pauvre en écho arqué, s'avèrera cependant moins intéressant pour discuter de ce diagnostic. On relèvera cependant un cas extrême le 18 août 2022 sur la Corse avec des rafales de vent à plus de 200 $km \cdot h^{-1}$.

8.4 Ajout d'un graphique rafale de vent

8.4.1 Construction du graphique

Le bilan de l'année 2021 et les retours sur l'article qui font l'objet du chapitre 7 ont montré un manque dans les graphiques existants. On ne s'intéresse pas au risque majeur lié aux échos arqués : les rafales de vent. En effet, même si le schéma conceptuel de l'écho arqué montre un potentiel important pour de fortes rafales de vent, il est important de différencier les échos arqués avec des rafales à 70-80 $km \cdot h^{-1}$ de ceux avec des rafales plus importantes à 110-130 $km \cdot h^{-1}$. Un nouveau graphique de synthèse est donc proposé (figure 8.6) avec pour objectif de représenter ce risque de fortes rafales de vent. Pour chaque écho arqué détecté, la distribution des rafales de vent à l'intérieur de l'objet simulé est extraite. Pour en

déduire une valeur représentative du risque lié aux rafales de vent sous l'écho arqué, plusieurs quantiles ont été testés. Les fortes rafales restant un évènement localisé, elles ne concernent qu'un nombre limité de points de grille, il faut donc choisir un quantile élevé qui sera le Q98 (le Q95 ne permettait pas de bien distinguer les échos arqués les plus forts des autres et le Qmax était trop sensible à un seul point de grille). Sur la figure 8.6, chaque écho arqué est représenté par une couleur qui dépend du Q98 de la distribution des rafales de vent à l'intérieur de l'objet allant du bleu pour les échos arqués très faibles (Q98 $< 50\ km \cdot h^{-1}$) au violet pour les plus violents (Q98 $> 150\ km \cdot h^{-1}$). Sur cette situation la plupart des échos arqués modélisés entre le Massif-Central et l'Allemagne sont associés à des rafales dépassant les 100 $km \cdot h^{-1}$, cette zone est donc particulièrement à surveiller. Ce graphique peut également servir à repérer des détections erronées en observant une grande concentration d'échos arqués avec des rafales inférieures à 50 $km \cdot h^{-1}$ qui est un niveau d'intensité trop bas pour des échos arqués.

8.4.2 Exemple sur une situation

L'apport de ce nouveau graphique est illustré sur la situation du 17 juin 2021 (figure 8.7). L'analyse de la situation faite par le Centre National de Prévision (CNP) est présentée à travers l'ANAlyses SYnoptiques Graphiques (ANASYG) de 18h UTC. La signification des différents symboles est disponible en annexe (figure A.1). Une dépression remonte dans un flux de Sud-Ouest avec une anomalie de tropopause dynamique favorisant la convection qui est balisée par une zone rouge sur le Sud-Ouest de la France. Météo-France avait émis une vigilance pour des orages violents du Sud-Ouest vers les Ardennes. Un écho arqué s'est formé en fin d'après-midi et a concerné l'Est de la région toulousaine avec des rafales mesurées jusqu'à 110 $km \cdot h^{-1}$. La prévision du risque d'échos arqués par la PE-AROME du 16 juin à 21h UTC est représentée figure 8.8. La carte de probabilités montre que le risque se concentre principalement sur le Sud-Ouest de la France, avec également un risque plus mesuré sur les Ardennes. La carte correspond assez bien à la vigilance émise par Météo-France. La figure au centre montre que le signal sur le Sud-Ouest se décompose en deux salves orageuses différentes. Une première salve remonte d'Espagne en fin d'après-midi (zone cerclée en orange) et une autre salve sur l'Aquitaine est modélisée plus tard en soirée. C'est également en soirée ou début de nuit que sont modélisés les échos arqués sur le Nord de la France (cercles verts). La dernière carte nouvellement introduite à droite permet de se focaliser sur la première salve qui remonte d'Espagne. Ces échos arqués sont en effet associés à des rafales autour de 100 $km \cdot h^{-1}$ alors que les échos arqués sur l'Aquitaine ou le Nord de la France sont beaucoup plus faibles avec des rafales généralement autour de 70 $km \cdot h^{-1}$. La focalisation sur le Sud-Ouest avec ces rafales de vent ne doit pas faire oublier que cette carte ne concerne que le risque de rafales de vent sous des échos arqués, elle ne concerne pas en particulier les supercellules ou d'autres risques comme les chutes de grêle. On ne peut donc pas se fier à cette seule carte pour évaluer l'intensité de la convection et des risques associés.

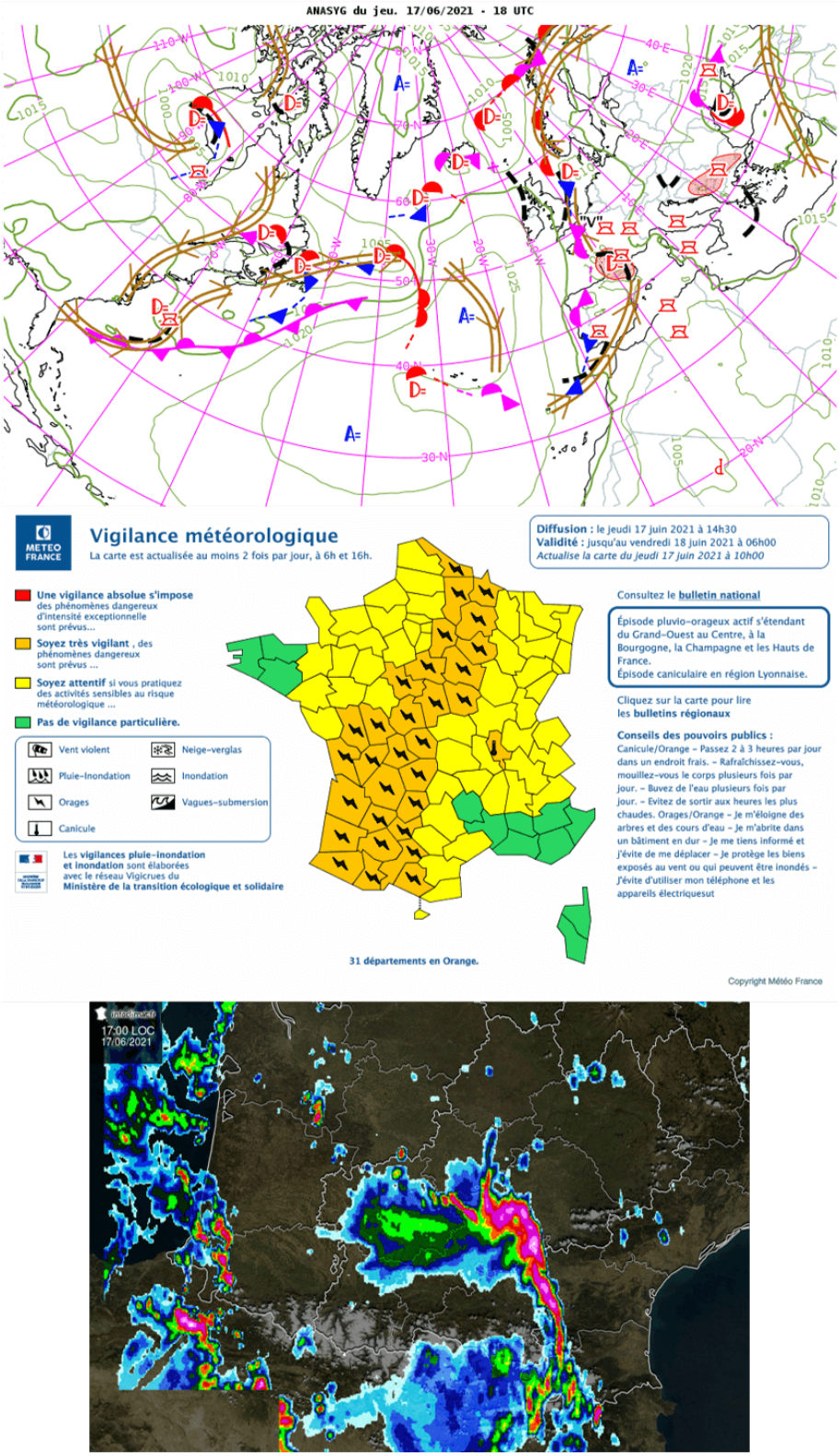


FIGURE 8.7 – *Situation météorologique du 17 juin 2021. La carte du haut représente l'ANASYG produit par le CNP pour le 17 juin 2021 à 18h UTC. La carte au centre représente la carte de vigilance émise par Météo-France le 17 juin 2021 à 14h et la carte du bas la mosaïque radar sur le Sud-Ouest de la France du 17 juin 2021 à 17h (Source : Infoclimat).*

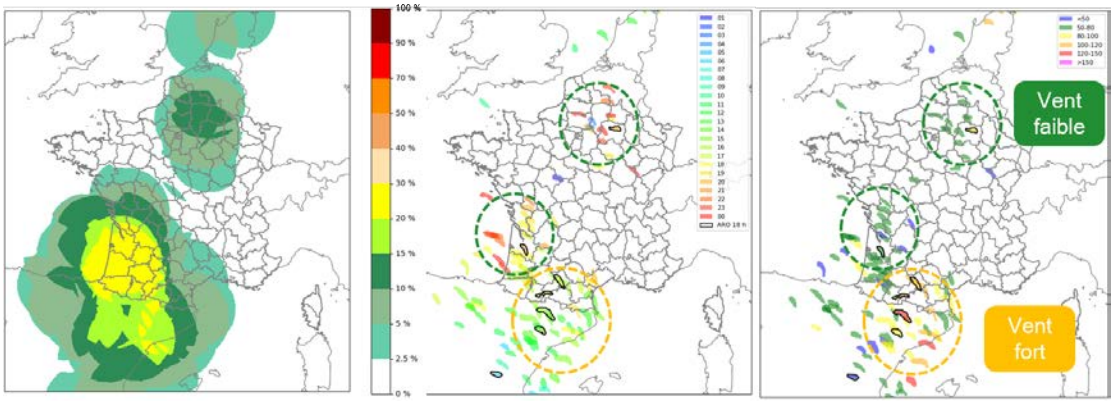


FIGURE 8.8 – *Prévision du risque d'échos arqués pour la journée du 17 juin 2021. Prévision faite à partir du réseau PE-AROME du 16 juin à 21h UTC. La carte de gauche représente le risque sous une forme probabiliste, au centre les échos arqués avec un code couleur selon l'heure de détection et à droite le nouveau graphique avec un code couleur selon les rafales de vent.*

8.5 Comparaison anciennes et nouvelles versions

La dernière partie de ce chapitre concerne la comparaison des prévisions faites par les anciennes versions des modèles PE-AROME et AROME qui étaient opérationnelles en 2021 avec les nouvelles versions utilisées depuis l'été 2022. Une évolution majeure de la PE-AROME est l'augmentation de résolution horizontale à 1.3km qui rejoint la version déterministe AROME. On peut donc s'attendre à une meilleure modélisation des phénomènes convectifs comme les échos arqués. Avec pour objectif de comparer les performances sur des temps longs, les nouvelles versions ont été rejouées sur une longue période débutant en février 2021 jusqu'en 2022 grâce à une coopération entre les services du Groupe de Modélisation et d'Assimilation pour la Prévision (GMAP), de la Direction des Systèmes d'Information (DSI) et de la Direction des Opérations pour la Prévision (DIROP). Les prévisions établies avec les anciennes et les nouvelles versions sont donc disponibles pour l'été 2021 et permettent une comparaison qualitative. Les grilles de sortie des modèles étant inchangées et les membres PE-AROME ressemblant à l'ancienne version d'AROME, il n'y a a priori pas de raison d'observer un net changement des performances du U-Net sur ces nouvelles versions. Cette comparaison reste qualitative et non quantitative car aucun écho arqué observé n'a été labellisé en 2021.

Toutes les situations en 2021 ne vont pas être présentées. Les comparaisons sur deux exemples sont détaillées (figures 8.9 et 8.10). La figure 8.9 correspond à une situation où de la convection sévère sera observée en particulier sur le Nord de la France, prenant la forme d'un écho arqué en se dirigeant vers les frontières belges. Les modèles envisagent particulièrement deux zones à risque qui sont le Sud-Ouest et le Nord-Est de la France vers les Ardennes. La nouvelle version de la PE-AROME cible cependant avec beaucoup plus d'insistance le Nord-Est de la France de façon pertinente. Le graphique dédié aux rafales de vent montre aussi une accentuation des rafales avec plusieurs détections comprenant des rafales de vent entre 120 et 150 $km \cdot h^{-1}$ avec la nouvelle version. Ces valeurs semblent cependant un peu trop élevées, les relevés étant plutôt entre 100 et 120 $km \cdot h^{-1}$ sur

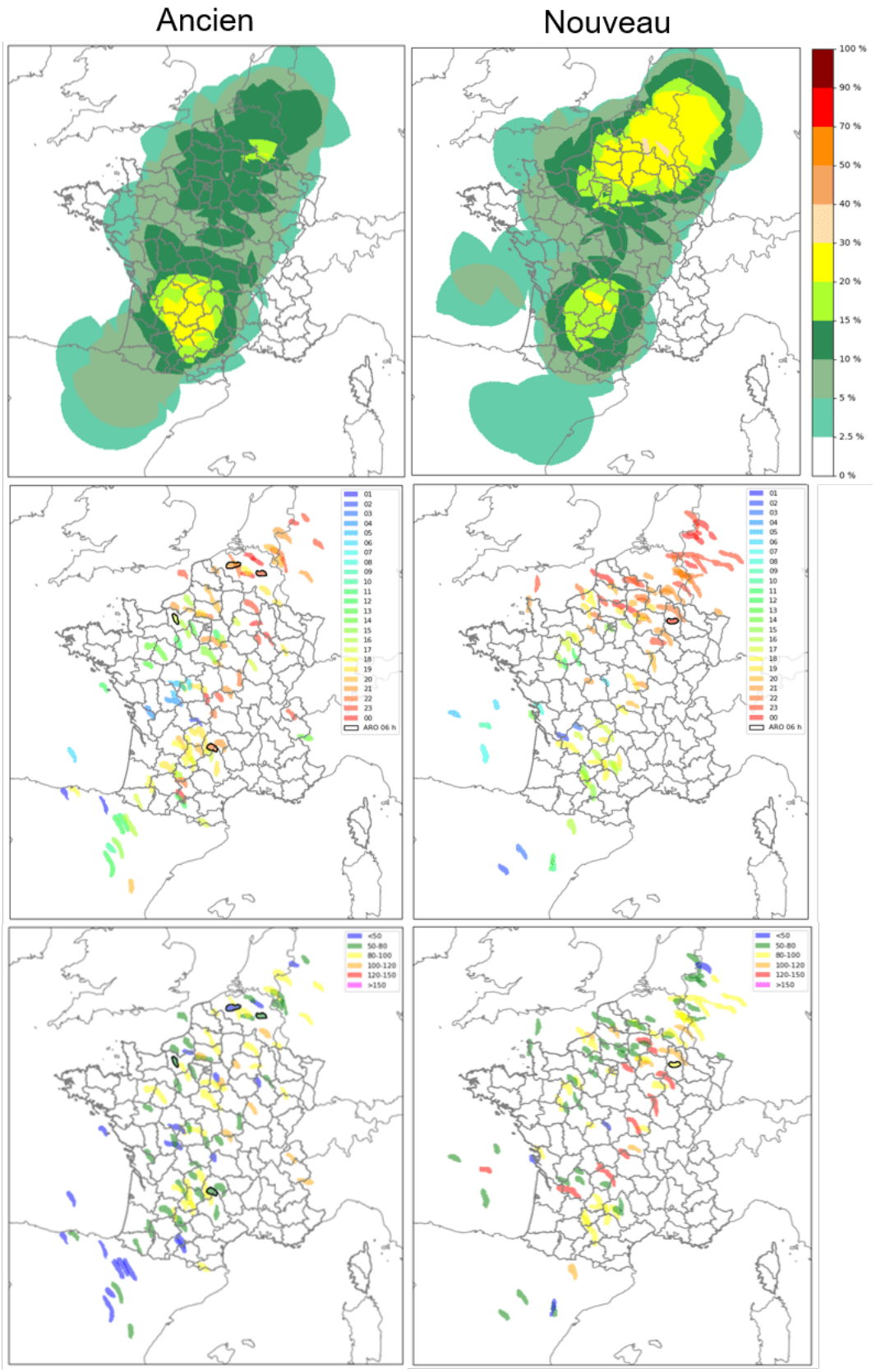


FIGURE 8.9 – *Comparaison des prévisions faites par les anciennes et les nouvelles versions des modèles pour la journée du 19 juin 2021 avec le réseau du 18 juin 2021 à 09h UTC. La première ligne représente les cartes de probabilités, la deuxième ligne les détections avec un code couleur selon l'heure de la journée et la dernière ligne le nouveau graphique avec un code couleur selon les rafales de vent.*

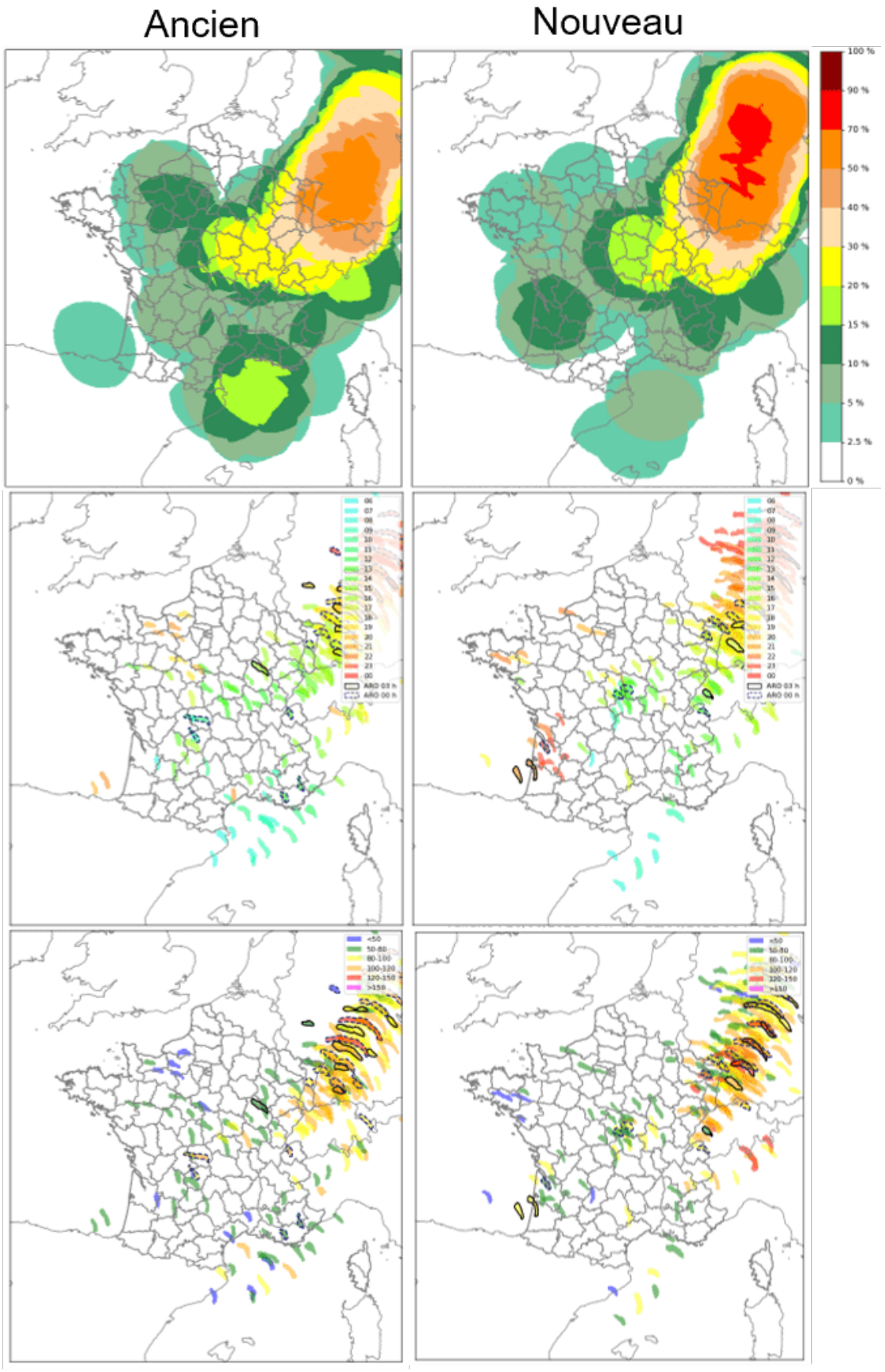


FIGURE 8.10 – *Figure similaire à la figure 8.9. Comparaison des prévisions faites par les anciennes et les nouvelles versions des modèles pour la journée du 20 juin 2021 avec le réseau du 20 juin 2021 à 03h UTC.*

cette situation. Pour la situation du lendemain, la figure 8.10 illustre une situation où la nouvelle version accentue énormément le signal sur l'Allemagne avec des probabilités dépassant les 70% ce qui est très rare. Un écho arqué sera bien observé mais plutôt entre l'Alsace et la Bavière ce qui ne correspond pas exactement à la trajectoire proposée par les deux versions des modèles. La nouvelle version supprime des détections sur la Méditerranée à raison mais en ajoute à tort sur l'Aquitaine. Aucun changement majeur pour les rafales de vent sur cette situation n'est constaté.

De façon plus générale sur l'ensemble des situations de 2021, il a été constaté que la nouvelle version des modèles propose des prévisions qui diffèrent de l'ancienne version. Le nombre de fausses alarmes semble se réduire et plusieurs corrections de localisation sont constatées entre l'ancienne et la nouvelle version. Il y a aussi quelques situations où la prévision semble moins bonne avec la nouvelle version (pour la journée du 17 juin 2021 en particulier). Tous ces éléments devront être confirmés par des scores pour être beaucoup plus robustes et des retours d'expérience sur une longue période faits par des prévisionnistes.

Ce chapitre s'est intéressé à la capacité des modèles AROME et PE-AROME à anticiper correctement la survenue d'échos arqués en France métropolitaine et à ses frontières. Pour répondre aux questions que cela soulève, il est nécessaire de détecter les échos arqués dans les observations. Des travaux en ce sens ont débuté mais ils ne présentent pas encore des résultats suffisamment robustes pour être utilisés sans aucun contrôle par un expert. A partir de labellisations manuelles sur les observations, les scores objectifs calculés pour PE-AROME et AROME ont montré l'intérêt d'utiliser un voisinage spatial à cause d'erreurs de position d'environ 150 km. Ces erreurs de position augmentent légèrement avec les échéances. Ces scores ont aussi montré l'intérêt d'utiliser PE-AROME plutôt que AROME aux échéances supérieures à 15h. Une étude approfondie des fausses alarmes a montré que les origines de celles-ci peuvent être diverses comme une durée de vie de l'écho arqué modélisée comme trop longue par les modèles ou encore un décalage temporel important par rapport à l'observation. Il est cependant difficile d'établir si des biais systématiques existent à cause d'un nombre de cas trop faible. Après la mise en place d'une production quotidienne en 2021 pour évaluer le risque d'échos arqués, il est apparu nécessaire d'ajouter de nouvelles informations sur les rafales de vent sous les échos arqués pour mieux repérer les situations les plus dangereuses pour la sécurité des personnes et des biens. De façon générale, ces graphiques pour évaluer le risque d'échos arqués permettent une prise en compte plus simple et plus rapide de la PE-AROME, apportant un complément intéressant aux informations issues de AROME. Ces graphiques peuvent également être utilisés pour comparer différentes versions de modèle comme la dernière partie de ce chapitre l'a montré. Il reste cependant difficile d'établir des conclusions robustes à cause d'un nombre de cas très faible.

Quatrième partie

Construction des scénarios de pluie

Chapitre 9

Étude approfondie de l'espace latent

Cette partie traite des scénarios de pluie, il n'est donc plus question des échos arqués. Elle est la suite des chapitres 3 et 6 qui avaient introduit respectivement la méthodologie générale des scénarios et les autoencodeurs qui sont utilisés pour faire de la réduction de dimension et ainsi passer d'un espace physique à un espace latent à 10, 20 ou 40 dimensions. La fin du chapitre 4 sur les méthodes de deep learning a montré que l'un des défis lié à ces méthodes est de limiter l'effet « boîte noire » en tentant d'expliciter au mieux comment les réseaux de neurones fonctionnent (McGovern et al., 2019). Cette partie débute par conséquent avec un chapitre consacré à l'étude des espaces latents qui ont été construits par les autoencodeurs. Ces espaces latents ont un rôle clé dans la méthodologie de construction des scénarios, il est donc nécessaire d'approfondir leurs caractéristiques pour la suite des travaux. Ce chapitre permettra ainsi de mieux appréhender certains résultats obtenus dans le chapitre suivant dédié à la classification climatologique dans un espace latent. L'étude de ces espaces latents se fera en premier lieu en approfondissant les résultats du chapitre 6. Une première partie comparera les attributs des objets reconstruits par l'autoencodeur aux objets en entrée de l'autoencodeur pour comprendre les informations qui transitent ou non par les espaces latents. Les deux autres points s'intéresseront plus directement aux espaces latents en abandonnant les images reconstruites. Des structures de pluies idéalisées seront utilisées dans une deuxième partie pour comprendre comment l'information est codée dans les espaces latents pour des objets de tailles, intensités ou positions différentes. La dernière partie sera une étude faite dimension par dimension de ces espaces latents à l'aide d'un important jeu de données sur l'année 2021. Le détail pour chaque espace latent à 10, 20 et 40 dimensions ne pourra pas être fait pour limiter le nombre de figures et rendre le chapitre abordable. Beaucoup de figures seront proposées en annexes. L'espace latent à 40 dimensions sera un peu plus détaillé car il sera retenu pour la synthèse par scénarios à la fin de ce manuscrit, les raisons étant expliquées dans le chapitre suivant.

9.1 Étude comparative des attributs d'objets

Cette partie sur les attributs reprend les trois types d'objets définis dans le chapitre 6 (section 6.3.1) et la même base de données (2021) que celle présentée dans la section 6.1.3. Les attributs des objets sont extraits de la même façon que ceux présentés pour les échos arqués dans le chapitre 5 (section 5.5.1). Les attributs des objets en entrée de l'autoencodeur (nommés dans cette partie objets PE car issus de la PE-AROME) sont comparés aux attributs des objets reconstruits par l'autoencodeur (nommés objets AE). Deux attributs seront étudiés : l'aire des objets (exprimée en nombre de points de grille) et l'intensité grâce au quantile 90 des pluies à l'intérieur de l'objet (exprimée en mm). Les représentations graphiques s'inspirent de celles proposées dans la section 4 de Mounier et al. (2022). Cette comparaison des attributs est divisée en trois parties avec un premier point qui s'intéresse à des caractéristiques générales des objets PE et AE, les deuxième et troisième points se focalisent sur respectivement les non détections et les fausses alarmes mis en évidence à l'aide des scores du chapitre 6.

9.1.1 Caractéristiques générales

Les caractéristiques générales des objets PE et AE sont présentées sur la figure 9.1. Cette figure correspond à l'autoencodeur avec un espace latent de 40 dimensions, les mêmes figures pour les espaces à 10 et 20 dimensions sont disponibles en annexe (figures D.1 et D.2). Les cas avec aucun objet en entrée et en sortie de l'autoencodeur sont exclus pour rendre les graphiques plus lisibles. La première ligne s'intéresse à la somme des aires des objets en entrée (PE) et en sortie de l'autoencodeur (AE). Cette grandeur est préférée à la comparaison objet à objet (non montrée). En effet, la méthode utilisée pour les appariements entre les objets est la même que celle décrite section 6.3.2, chaque objet en sortie est associé à l'objet en entrée dont le centre de masse est le plus proche de lui (sous réserve qu'un tel objet existe). Cette méthode engendre des erreurs en appareillant par moment de petits objets avec des plus gros qui ne décrivent pas les mêmes structures atmosphériques. Ces mauvais appariements rendent les graphiques objet à objet difficiles à lire . Même si ces mauvais appariements sont peu nombreux vis-à-vis des 400 000 objets de pluies totales [1], ils saturent une bonne partie du graphique loin de l'axe $y = x$. Pour ces raisons, les graphiques présentés comparent la somme des aires des objets et non les aires objet à objet. Ce changement est surtout important pour les objets de pluies totales, beaucoup moins pour les objets de pluies fortes car il y a rarement plus de un objet de pluies fortes sur le champ à la fois. Considérer l'objet ou la somme des objets aboutit dans ce cas au même résultat.

La première ligne de la figure 9.1 montre que la somme des aires des objets sur un champ de RR1h est très bien reproduite par l'autoencodeur pour les trois types d'objets avec des droites de régression très proches de la reproduction parfaite ($y = x$) et un coefficient de détermination R^2 très élevé à 0.98. Contrairement à l'espace latent à 40 dimensions, il apparaît que les aires des gros objets de pluies modérées et fortes sont sous-estimées avec un espace à 10 ou 20 dimensions. Ces

1. Ils n'étaient donc pas préjudiciables pour le calcul de l'erreur de position moyenne dans le chapitre 6.

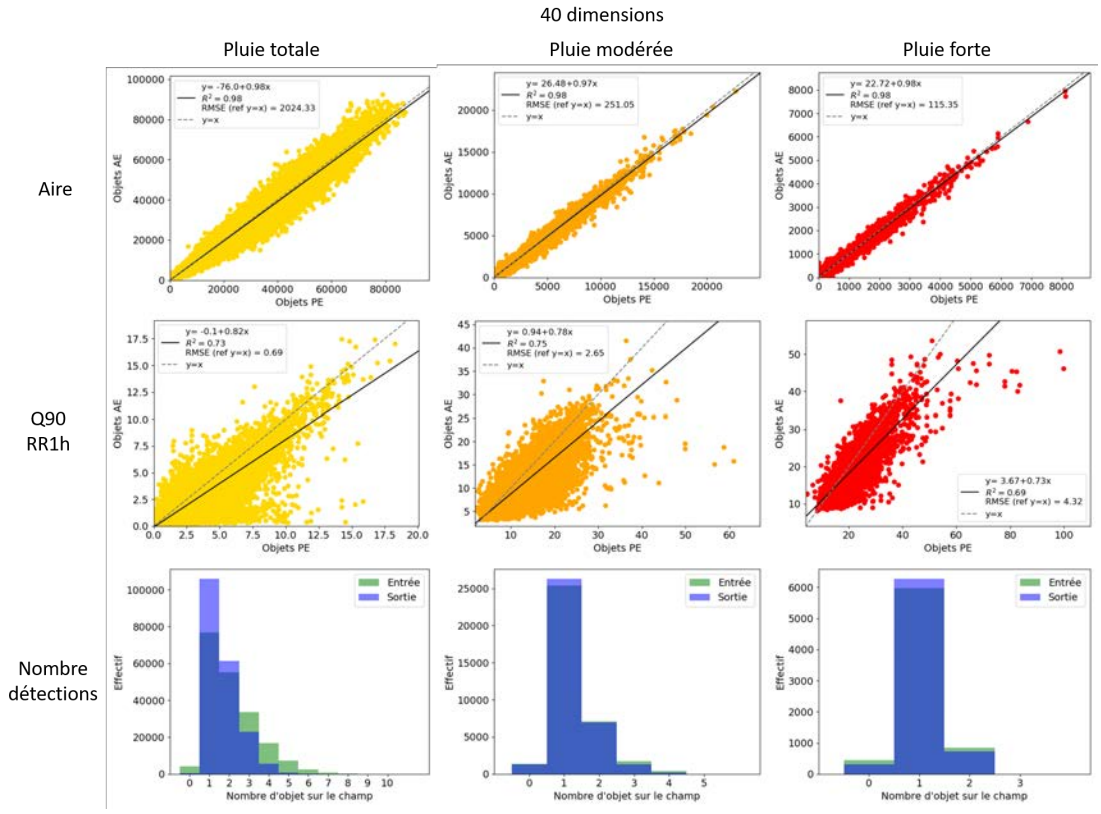


FIGURE 9.1 – *Scores orientés objets pour comparer les objets en entrée et en sortie de l'autoencodeur. L'autoencodeur présenté est celui possédant un espace latent à 40 dimensions. Chaque colonne correspond à un type d'objet (pluie totale, modérée ou forte). La première ligne compare l'aire des objets en entrée et en sortie exprimée en nombre de points de grille, une droite de régression en noir est tracée. L'équation de la droite de régression, le coefficient de détermination (R^2) et la RMSE par rapport à la droite $y = x$ (en gris tireté) sont précisés dans la légende. La deuxième ligne correspond aux mêmes graphiques mais pour le Q90 des RR1h des objets exprimé en mm. La dernière ligne compare les distributions du nombre d'objets détectés sur un champ de RR1h pour les données en entrée (objets PE, vert) et en sortie (objets AE, bleu) de l'autoencodeur.*

moins bonnes performances se mesurent aussi avec la RMSE qui est égale par exemple à environ 670 points de grille pour un espace à 10 dimensions pour les pluies modérées contre 250 points de grille pour un espace à 40 dimensions.

La deuxième ligne s'intéresse à l'intensité des objets (Q90). Sommer ou moyenner des intensités n'est pas possible comme pour des surfaces. On étudiera donc une comparaison objet à objet contrairement à la première ligne, en retenant que certaines associations sont erronées. Par exemple on observe des objets AE de pluies totales avec des intensités inférieures à 2.5 mm associés à des objets PE beaucoup plus forts supérieurs à 10 mm, ce qui semble peu vraisemblable pour des objets comparables. Par rapport à l'aire des objets, l'intensité des objets est plus difficile à reconstruire pour l'autoencodeur avec une tendance pour tous les objets à sous-estimer les intensités, cette sous-estimation s'amplifiant avec les objets les plus intenses [2]. Cette sous-estimation s'accroît avec des espaces à 10 ou 20

2. Ce problème pourrait être résolu avec une MSE pondérée dans des futurs travaux.

dimensions. Par exemple pour les pluies fortes, la RMSE est de 6.6 mm pour 10 dimensions contre 4.3 mm pour 40 dimensions. Les intensités sont donc lissées par l'autoencodeur, cet effet étant fréquemment observé quand on cherche à réduire la dimension d'un problème. On ne peut plus en effet expliquer la totalité de la variance du jeu de données en réduisant le nombre de dimensions (Murphy, 2012).

La dernière ligne de la figure 9.1 s'intéresse au nombre d'objets détectés sur un champ de RR1h. L'histogramme est tracé pour les objets PE (vert) et AE (bleu) . Comme les cas avec aucun objet en entrée et en sortie de l'autoencodeur sont exclus, il est impossible d'observer une situation où on comptabiliserait 0 objet en entrée et en sortie simultanément. On relève cependant l'absence occasionnelle d'objet PE ou AE sur ces histogrammes. Ces cas correspondent aux non détections s'ils appartiennent aux histogrammes bleus [3] et aux fausses alarmes s'il s'agit des histogrammes verts. On remarque que le nombre d'objets de pluies totales est plus important (jusqu'à 6 ou 7 objets possibles pour les objets PE) que le nombre d'objets de pluies modérées (3-4 objets) ou de pluies fortes (2 objets). L'histogramme bleu montre qu'avec 40 dimensions, l'autoencodeur peut reconstruire jusqu'à 4 objets différents sur la même image pour les pluies totales mais qu'il ne peut pas représenter le bon nombre d'objets quand le champ devient trop complexe. Ce maximum de 4 objets est encore plus faible si l'espace est de 10 ou 20 dimensions (3 objets). L'espace à 40 dimensions permet cependant de reproduire correctement le nombre d'objets de pluies modérées ou fortes, ce qui n'est pas le cas avec 10 ou 20 dimensions. Il est en particulier très difficile avec 10 dimensions de reproduire 2 objets de pluies fortes.

9.1.2 Caractéristiques des non détections

La figure 9.2 permet de se focaliser sur les non détections de l'autoencodeur à 40 dimensions. Les figures correspondantes D.3 et D.4 sont disponibles en annexe pour les espaces à 10 et 20 dimensions respectivement. Les graphiques sur la première ligne représentent l'ensemble des objets PE dans un repère Aire/Q90 des objets, les points bleus étant les objets non détectés. Les non détections ayant été tracées après l'ensemble des objets PE, les points bleus ne peuvent pas se situer sous un grand nombre d'autres points qui ne les rendraient pas perceptibles. Quel que soit le type d'objet, les non détections sont des objets petits (gauche des graphiques). Cette observation est confirmée par la deuxième ligne qui est l'histogramme des aires des objets PE où l'histogramme des non détections est superposé. On remarque que les non détections pour les pluies totales avec 40 dimensions ne sont pas perceptibles car trop peu nombreuses par rapport au nombre total d'objets. Les non détections sont aussi préférentiellement des objets petits avec un espace à 10 ou 20 dimensions même si des objets de plus en plus grands sont concernés quand le nombre de dimensions diminue.

Concernant les intensités, les non détections sont principalement des objets faibles en intensité comme le montre la dernière ligne de la figure 9.2 qui représente la distribution des Q90 des objets PE et des non détections en bleu. Les queues de distribution sont cependant de plus en plus épaisses quand on limite la taille de l'espace latent avec donc des objets de pluies de plus en plus intenses qui ne sont

3. On est dans une situation avec aucun objet AE et au moins un objet PE car sinon ce cas ne serait pas représenté. Ce cas correspond alors à une non détection.

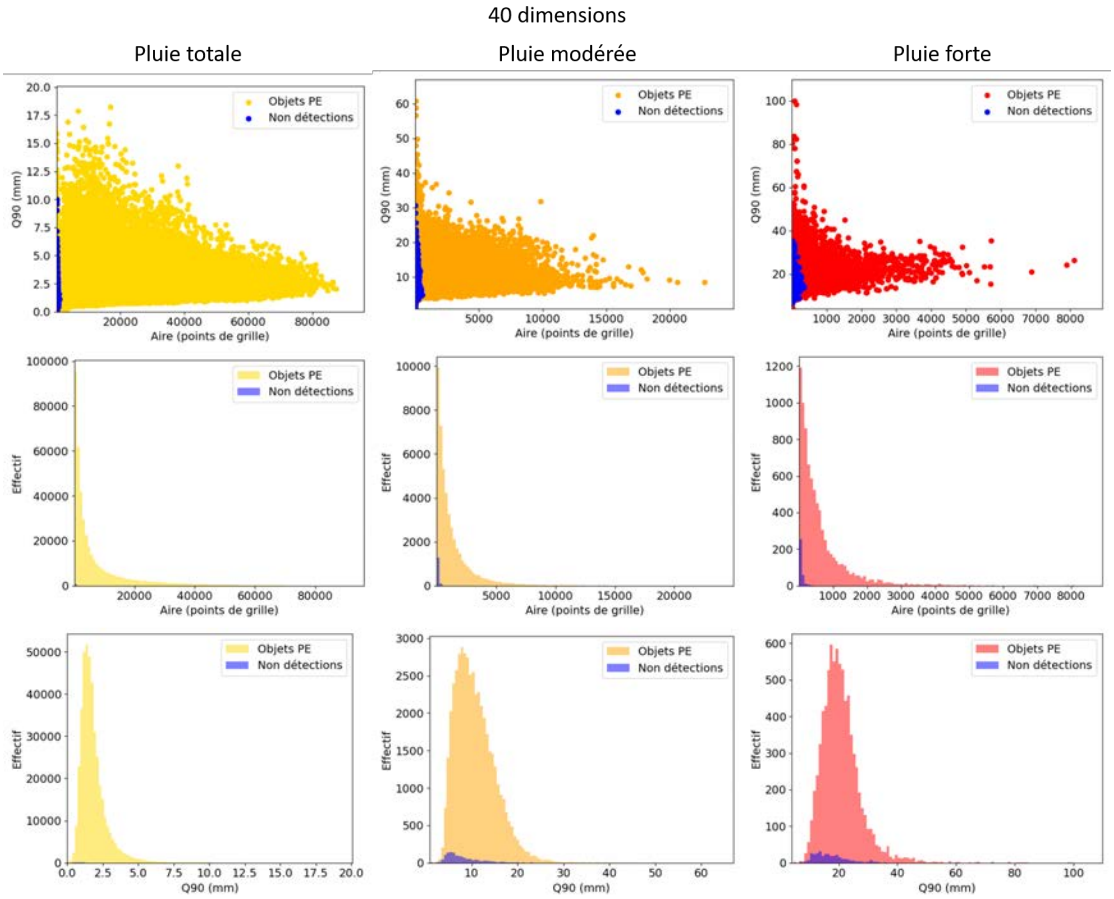


FIGURE 9.2 – *Scores orientés objets pour étudier les non détections de l'autoencodeur avec un espace latent à 40 dimensions. Chaque colonne correspond à un type d'objet (pluie totale, modérée ou forte). La première ligne représente l'ensemble des objets en entrée de l'autoencodeur selon leur aire en nombre de points de grille et selon leur intensité par l'intermédiaire du Q90 en mm. Les objets non détectés en sortie de l'autoencodeur sont ensuite représentés en bleu. La deuxième ligne représente la distribution des aires pour l'ensemble des objets et celle des non détections en bleu. La troisième ligne correspond aux mêmes histogrammes mais pour le Q90 des objets.*

pas reconstruits correctement par l'autoencodeur. La première ligne montre qu'il existe une dépendance entre l'aire et la taille des non détections. Les objets non détectés les plus gros sont des objets avec des pluies peu intenses et à l'inverse les non détections avec des intensités fortes sont des objets petits.

On peut donc conclure que les non détections de l'autoencodeur sont principalement des objets petits avec des intensités qui ont tendance à être plutôt faibles. L'effet de lissage sur les champs que peut avoir un autoencodeur et plus généralement une réduction de dimension comme évoquée dans la section 9.1.1 peut expliquer ce résultat. Cette analyse a permis de s'assurer qu'aucun objet de taille importante et intense n'est oublié par l'autoencodeur. On peut ajouter que les intensités supérieures à 40 mm pour les objets de pluies fortes sont toujours détectées avec l'espace à 40 dimensions, ce qui permet de dire que les objets les plus forts, y compris les plus petits, sont correctement détectés avec cet autoencodeur. C'est un point important en particulier quand on s'intéresse aux fortes pluies sous certaines cellules orageuses stationnaires qui se caractérisent par des cumuls de

pluies importants sur des zones très restreintes.

9.1.3 Caractéristiques des fausses alarmes

Sur le même principe que la figure 9.2, la figure 9.3 permet de se focaliser sur les fausses alarmes de l'autoencodeur à 40 dimensions. Les figures D.5 et D.6 sont disponibles en annexe pour les espaces à 10 et 20 dimensions. L'ensemble des objets AE sont ici considérés et les fausses alarmes apparaissent en bleu. Les conclusions sont similaires à celles obtenues dans le paragraphe précédent, les fausses alarmes sont des objets petits. Il semble cependant pour les objets de pluies modérées ou fortes qu'aucun signal franc ne se dégage sur les intensités qui peuvent prendre une gamme de valeur très étendue. Ces conclusions sont également valables pour les espaces à 10 et 20 dimensions. La présence de fausses alarmes semble en contradiction avec la conclusion précédente qui était que l'autoencodeur tend à lisser les champs en entrée. Elle peut s'expliquer par la méthode objet utilisée pour l'évaluation comme évoqué dans le chapitre 6, section 6.4.1. L'exemple présenté figure 6.5 correspond à la fausse alarme avec le Q90 le plus élevé autour de 50 mm pour l'autoencodeur à 40 dimensions. Il semble donc possible qu'en présence d'objets petits en entrée, l'autoencodeur reconstruise un objet plus faible en intensité mais un peu plus étalé que l'entrée. L'opération de lissage avant le seuillage du champ pour définir les objets peut alors filtrer le potentiel objet en entrée mais pas celui en sortie, aboutissant à une fausse alarme.

Cette étude comparative des attributs des objets en entrée et en sortie a montré que les autoencodeurs semblaient capables de reproduire plutôt fidèlement les tailles des objets mais qu'ils sous-estimaient leurs intensités. Ils ne sont également pas capables de reproduire un grand nombre d'objets différents simultanément. Ces observations ont tendance à s'amplifier quand la taille de l'espace latent se réduit. Les études des non détections et des fausses alarmes ont révélé qu'ils correspondaient à des objets petits (et plutôt faibles en intensité pour les non détections).

9.2 Structures pluvieuses idéalisées

L'étude des objets reconstruits dans la section précédente permet d'identifier les informations qui transitent correctement par l'espace latent mais elle reste une mesure indirecte. L'objectif de cette partie va être d'étudier directement l'espace latent à l'aide de structures pluvieuses idéalisées. Ces structures pluvieuses idéalisées vont être utilisées en entrée de l'autoencodeur et projetées dans l'espace latent sous la forme d'un vecteur à 10, 20 ou 40 dimensions par la partie encodeur du réseau de neurones. L'évolution de ce vecteur pour des objets de tailles, intensités ou positions différentes va être étudiée. Cette partie s'inspire des travaux de Gilleland et al. (2020) qui ont utilisé des structures idéalisées pour tester plusieurs métriques objets.

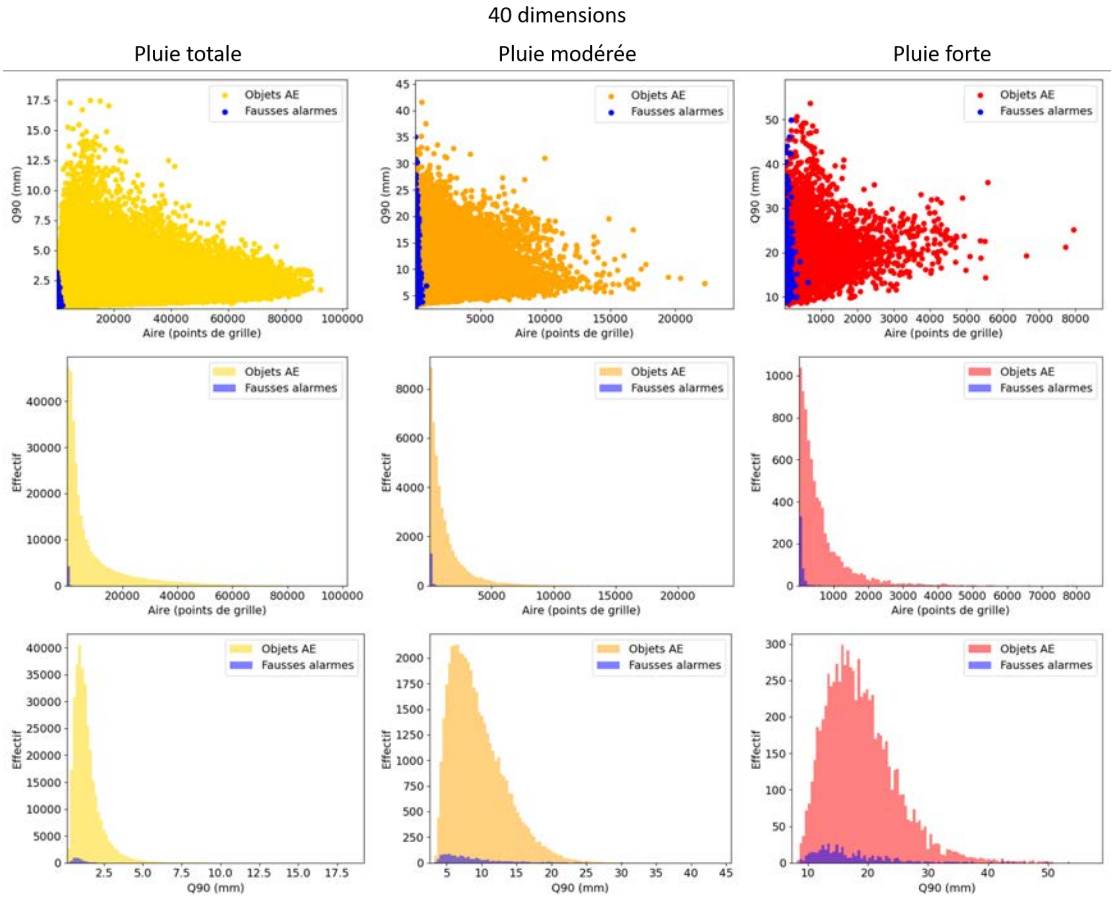


FIGURE 9.3 – *Scores orientés objets pour étudier les fausses alarmes de l'autoencodeur avec un espace latent à 40 dimensions. Les graphiques sont les mêmes que ceux de la figure 9.2 mais en superposant l'ensemble des objets reconstruits par l'autoencodeurs et ceux qui sont des fausses alarmes en bleu.*

9.2.1 Génération des structures

Les structures pluvieuses idéalisées sont définies à l'aide d'une gaussienne en deux dimensions qui permet de représenter en première approximation un objet de pluie. Les structures idéalisées se définissent sous la forme :

$$RR_{ideal}(x,y) = \frac{M}{max(f)} \times f_\sigma(x,y) \tag{9.1}$$

Avec f_σ la fonction de densité d'une gaussienne deux dimensions d'écart-type $\sigma = \sigma_x = \sigma_y$. M est un paramètre à fixer qui détermine le maximum en pluie sur le champ RR_{ideal} et s'exprime donc en mm. Les deux paramètres σ et M permettent ainsi de définir différentes structures idéalisées de tailles et intensités différentes. La taille de la structure, qu'on délimite par l'ensemble des points (x, y) où RR_{ideal} dépasse 0.1 mm, dépend des valeurs des variables mais elle est en général plus petite que le champ global de 256×384 points de grille. On peut donc également choisir sa position sur le champ et la faire varier avec un M et un σ fixé.

9.2.2 Modification de la taille

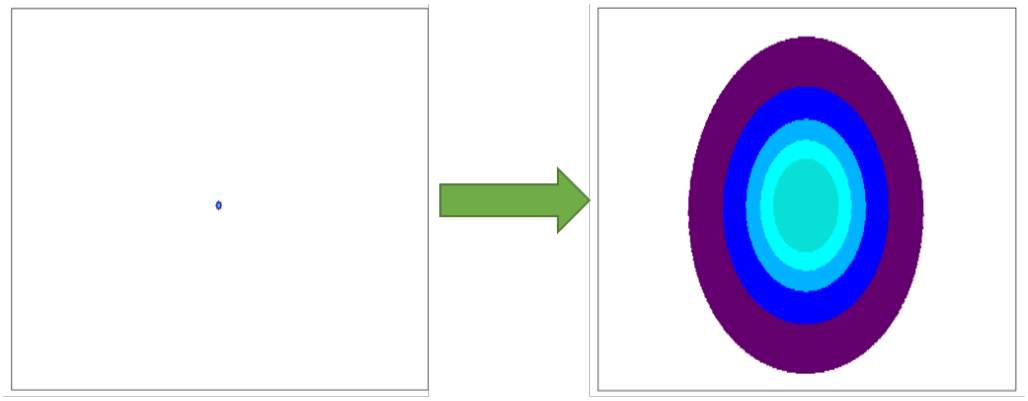
Un premier test est fait en modifiant la taille de l'objet (σ). L'objet est positionné au centre de la grille, M est fixé à 10 mm, correspondant à des pluies plutôt fortes en une heure sans être intense, et σ varie entre 1 point de grille (2.5 km) et 40 points de grille (100 km) avec un pas de un point de grille. Quarante champs différents sont donc générés et projetés dans l'espace latent. L'évolution dimension par dimension ainsi que les normes 1, 2, 3 et infini du vecteur latent sont représentées sur la figure 9.4 pour l'espace latent à 10 dimensions. La figure D.7 représente la même figure pour l'espace à 20 dimensions et les figures D.8 et D.9 pour l'espace à 40 dimensions qui est scindé en deux pour représenter de façon lisible les 40 dimensions. L'espace latent ne semble pas réagir aux premiers champs avec un vecteur qui reste proche du vecteur nul ce qu'on peut observer sur chaque dimension et sur chacune des normes. Ce résultat met en évidence une non linéarité introduite par la partie encodeur de l'autoencodeur et implique que les objets les plus petits sont filtrés, ce qui est cohérent avec les observations faites sur les non détections dans la partie précédente. On remarque que ce phénomène se réduit si la taille de l'espace latent augmente, il est même presque inexistant avec 40 dimensions ce qui peut expliquer aussi pourquoi le nombre de fausses alarmes augmente car permises par un espace latent qui, contrairement aux autres dimensions, ne filtre plus les petits objets. D'autres tests ont montré que l'espace latent pouvait malgré tout réagir à des petits objets même avec 10 dimensions si l'intensité des pluies devient très forte (non montré). Si on omet cet effet de seuil, les normes du vecteur dans l'espace latent semblent réagir de façon linéaire, il en va de même pour la dimension 8 dans l'espace à 10 dimensions. Ce constat n'est pas partagé pour toutes les dimensions, en particulier la dimension 3 par exemple. Ces conclusions s'étendent aux autres espaces latents à 20 et 40 dimensions.

9.2.3 Modification de l'intensité

Avec le même principe que dans le paragraphe précédent, on peut modifier l'intensité de la structure de pluie en fixant σ et en modifiant M. L'objet de pluie est toujours positionné au centre du champ avec un σ fixé à 20 points de grille (50 km) et M qui varie de 1 à 50 mm avec un pas de 1 mm pour un total de 50 champs différents. Les évolutions de l'espace latent à 20 dimensions sont représentées figure 9.5. Les espaces à 10 dimensions (figure D.10) et à 40 dimensions (figures D.11 et D.12) sont en annexe. La réaction de l'espace latent à l'augmentation de M semble linéaire pour l'ensemble des dimensions. On notera qu'il subsiste un léger effet de seuil pour l'espace à 10 dimensions pour les toutes premières structures les plus faibles. Cette réponse linéaire suggère qu'à position fixe, l'intensité des pluies est codée par l'amplitude des dimensions dans l'espace latent.

9.2.4 Modification de la position

Pour cette expérience, σ et M sont fixés à 10 points de grille (25 km) et 10 mm. La position de l'objet est modifiée en partant du bas à gauche de l'image puis elle est déplacée de 10 points de grille vers la droite jusqu'à l'autre extrémité, elle est alors remontée de 10 points de grille puis est déplacée de droite à



10 dimensions

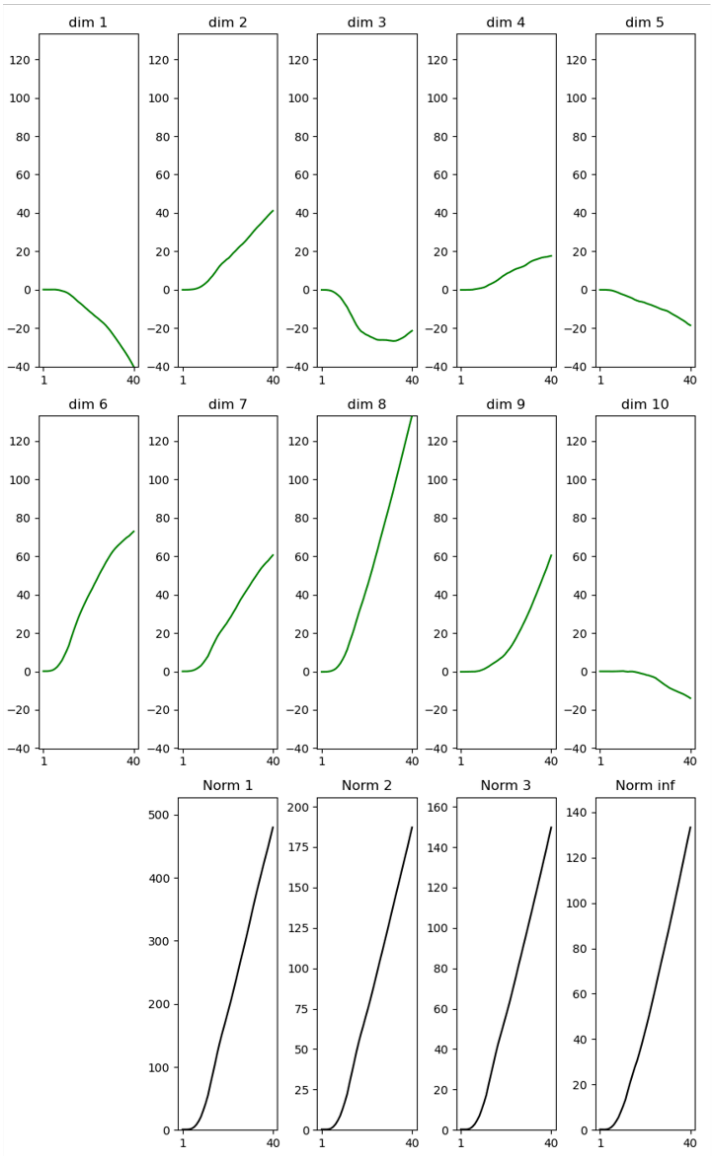


FIGURE 9.4 – *Modification de la taille de la structure, impact sur l'espace latent à 10 dimensions. Le maximum est fixé à 10 mm, σ varie de 2.5 km à 100 km avec un pas de 2.5 km. Les images en haut représentent le premier champ (σ=2.5 km) et le dernier (σ=100 km). Les évolutions dimension par dimension entre ces deux champs de l'espace latent sont représentées en dessous ainsi que l'évolution des normes 1,2,3 et infini du vecteur dans l'espace latent.*

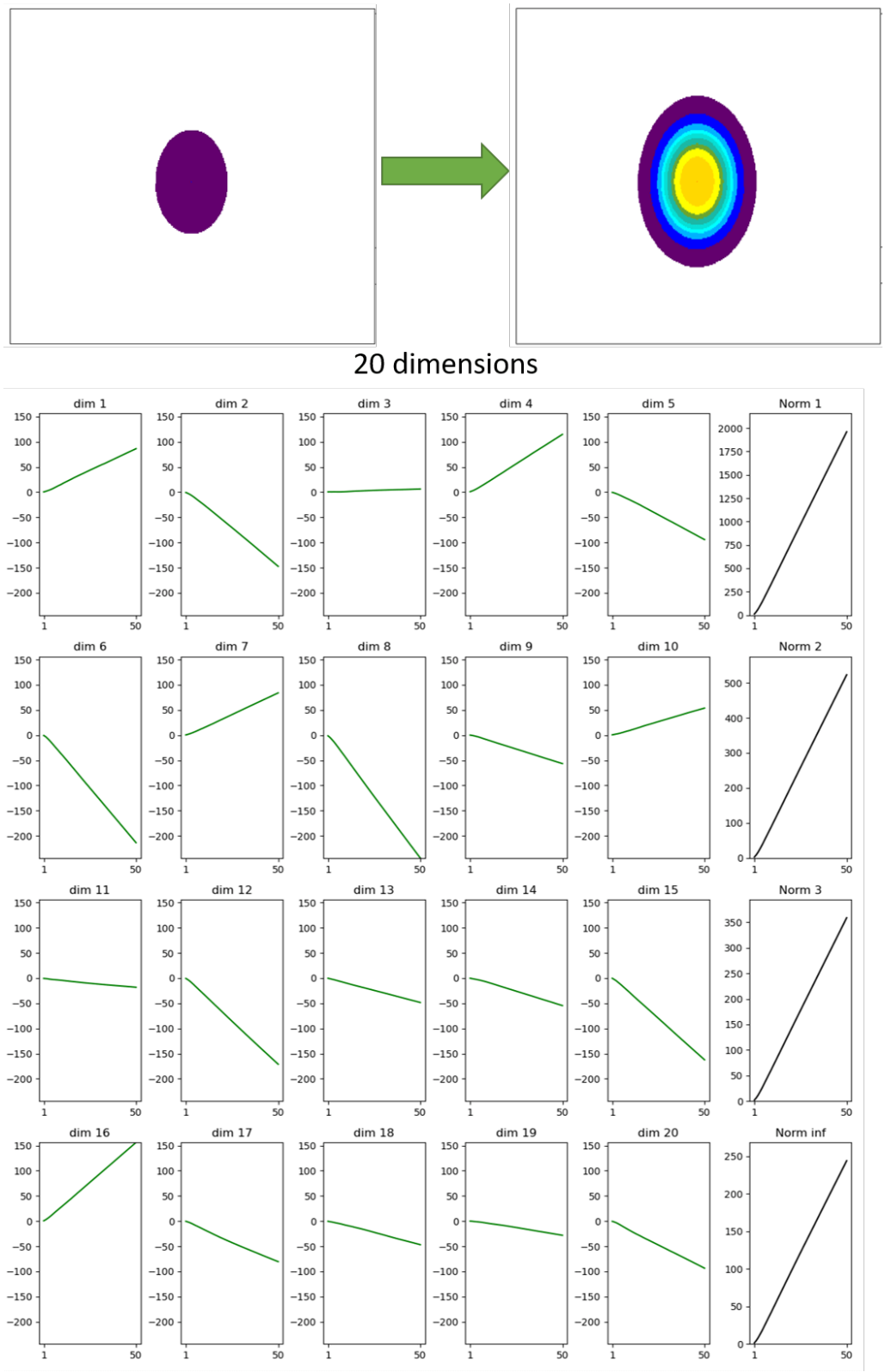


FIGURE 9.5 – *Modification de l'intensité de la structure, impact sur l'espace latent à 20 dimensions. Le σ est fixé à 20 points de grille, M varie de 1 mm à 50 mm avec un pas de 1 mm. Les images en haut représentent le premier champ ($M = 1$ mm) et le dernier ($M = 50$ mm). Les évolutions dimension par dimension entre ces deux champs de l'espace latent sont représentées en dessous ainsi que l'évolution des normes 1,2,3 et infini du vecteur dans l'espace latent.*

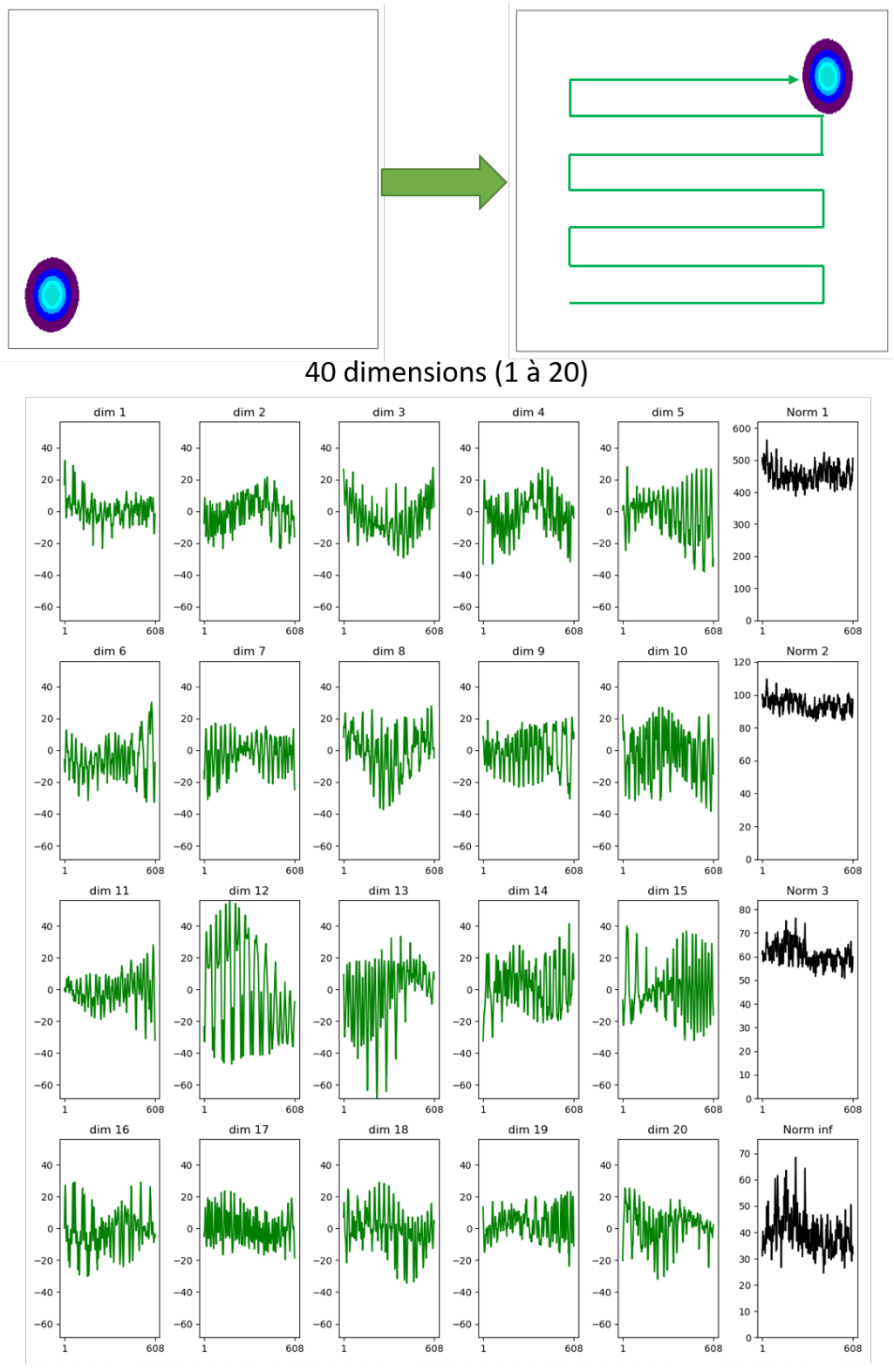


FIGURE 9.6 – *Modification de la position de la structure, impact sur l'espace latent à 40 dimensions (partie 1, dimension 1 à 20). Le σ est fixé à 10 points de grille et M à 10 mm. Les positions varient schématiquement comme sur la flèche verte de l'image de droite avec un pas de 10 points de grille que ce soit en X ou en Y. Les images en haut représentent le premier champ (bas à gauche) et le dernier (haut à droite). Les évolutions dimension par dimension entre ces deux champs de l'espace latent sont représentées en dessous ainsi que l'évolution des normes 1,2,3 et infini du vecteur dans l'espace latent.*

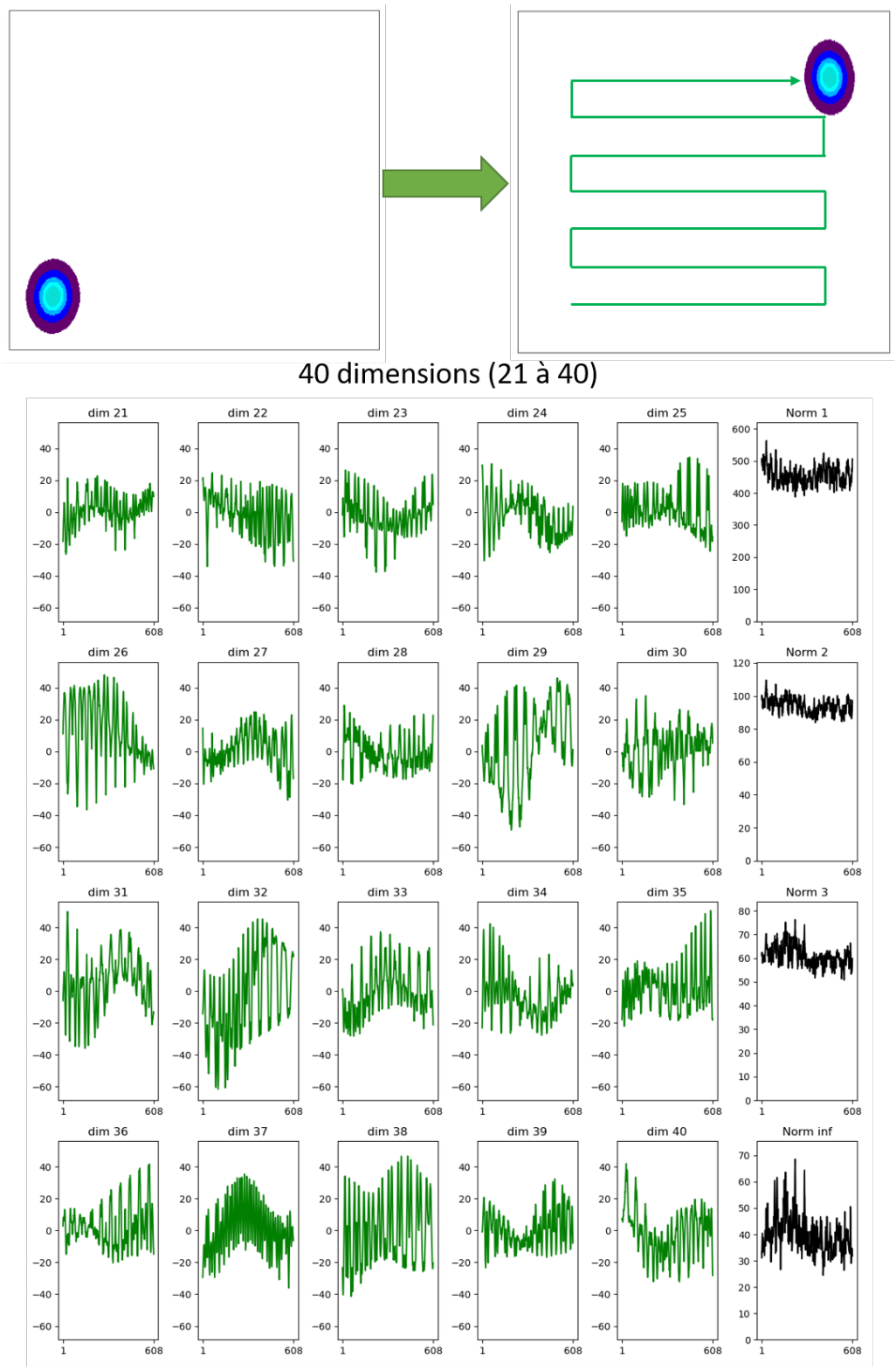


FIGURE 9.7 – *Modification de la position de la structure, impact sur l'espace latent à 40 dimensions (partie 2, dimension 21 à 40), similaire à la figure 9.6.*

gauche et ainsi de suite jusqu'à atteindre le haut de l'image comme indiqué sur les figures 9.6 et 9.7 qui représentent cette expérience pour l'espace à 40 dimensions. La même expérience pour les espaces à 10 et 20 dimensions est disponible en annexe (figures D.13 et D.14). On peut constater que les normes des vecteurs (en particulier la norme 2) fluctuent relativement peu selon la position de l'objet par rapport aux expériences précédentes, on l'observe également pour des valeurs de σ et M différentes (non montrés) ce qui assimile la norme 2 dans l'espace latent à un invariant de position dans l'espace physique[4]. Cette conclusion est valable pour l'espace à 40 dimensions mais devient de moins en moins vraie avec les espaces plus restreints à 10 ou 20 dimensions. Contrairement aux paragraphes précédents, on observe d'importantes fluctuations pour chacune des dimensions qui s'activent plus ou moins selon la position de l'objet. Les dimensions ont donc une signature dans l'espace importante. Par exemple, on peut remarquer que la dimension 12 pour l'espace à 40 dimensions a des oscillations importantes aux premières itérations qui correspondent au bas du champ, celles-ci atteignent un pic environ au premier tiers sud de l'image avant de se réduire considérablement en remontant plus haut ce qui indiquerait que la dimension 12 code principalement une partie de l'information dans la partie sud. Les importantes fluctuations avec un mouvement de l'objet qui va de gauche à droite et de droite à gauche indique aussi une sensibilité importante à la position de l'objet selon cette direction.

Ces tests à l'aide de structures pluvieuses idéalisées ont permis de mettre en évidence plusieurs points concernant les espaces latents. Tout d'abord la partie encodeur de l'autoencodeur filtre les objets trop petits en particulier pour les espaces latents à 10 et 20 dimensions. Si on omet cet effet de seuil, la réaction des espaces latents à des objets plus gros ou plus intenses est globalement linéaire. La modification de la position de l'objet induit une réponse bien différente avec des dimensions qui s'activent plus ou moins. Malgré ces importantes fluctuations dimension par dimension, la norme 2 du vecteur fluctue elle beaucoup moins. Ce mode de fonctionnement pour stocker l'information avec des dimensions qui représentent des zones de l'espace différentes et dont l'amplitude est reliée à l'intensité des pluies est très proche de celui qui est adopté par une ACP en première approche. Même si contrairement à une ACP, l'autoencodeur est capable d'introduire des non linéarités dans le codage ou le décodage de l'espace latent. La partie suivante va également montrer une organisation des dimensions différentes par rapport aux composantes d'une ACP.

9.3 Étude dimension par dimension

La dernière partie de ce chapitre se concentre sur une étude dimension par dimension des espaces latents. La méthode pour y parvenir est résumée par la figure 9.8. La première étape est de constituer une base de données assez large (étape 1) où chaque champ sera projeté dans l'espace latent (étape 2). Cette projection est relativement peu coûteuse numériquement, c'est pourquoi la base utilisée sera plus large que pour la section 9.1 avec une base toujours similaire à celle de la

4. vis à vis des fluctuations constatées dans les parties précédentes.

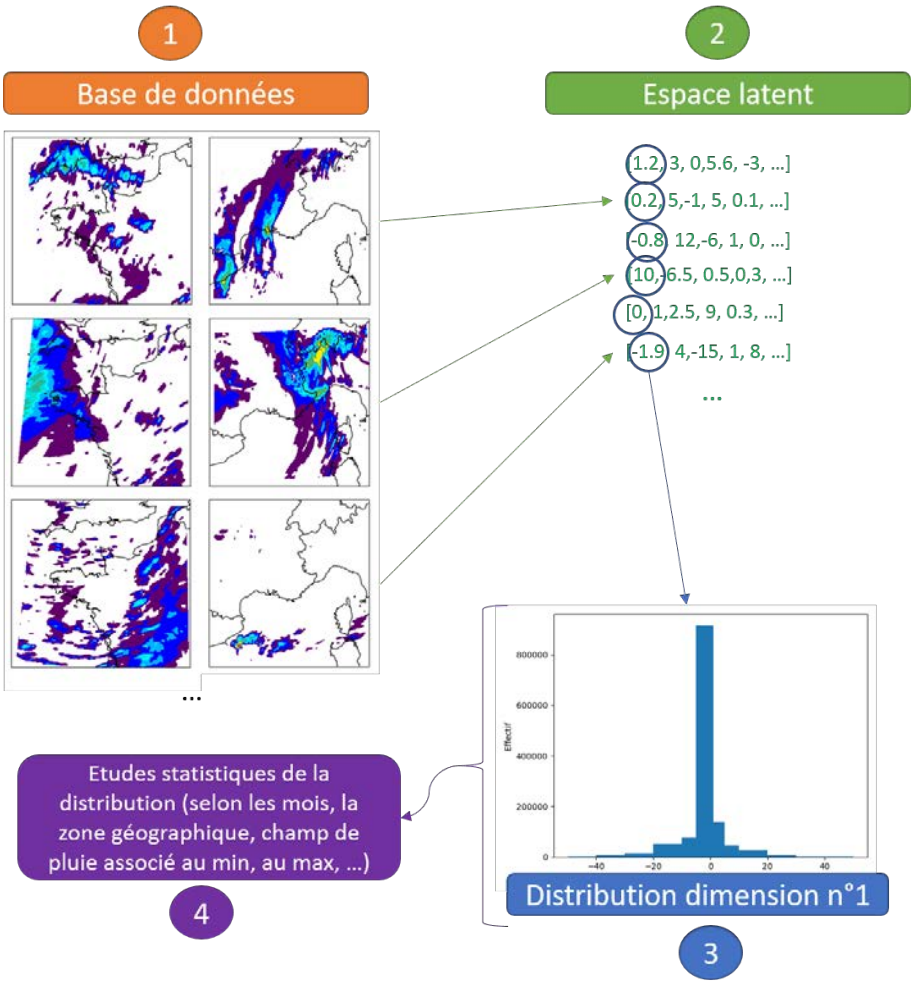


FIGURE 9.8 – *Construction de l'étude dimension par dimension en quatre étapes. Exemple pour la dimension n°1.*

section 6.1.3 (année 2021) mais où la contrainte sur les membres est supprimée, les 16 membres sont ainsi utilisés et non plus seulement les 4 premiers. L'étude n'est pas étendue aux années précédentes pour ne pas réutiliser des cas similaires à ceux de la base d'apprentissage de l'autoencodeur. Après l'étape 2, il est alors possible de tracer la distribution des valeurs prises par chaque dimension une à une (étape 3). L'étude de cette distribution va permettre d'extraire différentes caractéristiques de la dimension étudiée (étape 4). Ces caractéristiques seront présentées une à une dans la suite de cette partie. La présentation de chacune des dimensions serait chronophage et assez répétitive. C'est pourquoi l'étude pour une dimension va être faite : la dimension 12 de l'espace latent à 40 dimensions (figure 9.9) et les conclusions plus générales seront présentées en parallèle. L'équivalent de la figure 9.9 est disponible en annexe pour chacune des dimensions de l'espace à 10 dimensions (figures D.15 à D.24), à 20 dimensions (figures D.25 à D.44) et à 40 dimensions (figures D.45 à D.84).

9.3.1 Répartition selon les échéances

Le premier graphique présenté est celui en haut à gauche de la figure 9.9. La distribution globale (étape 3, figure 9.8) est découpée selon les échéances. On obtient ainsi une distribution pour chaque échéance. Le premier graphique représente alors l'évolution de la moyenne de ces distributions selon les échéances en noir. Les autres courbes représentent l'évolution de quantiles allant de 1% à 99% qui sont explicités dans la légende à droite de la figure 9.9. Il est nécessaire de représenter les queues de distribution car la base de données est constituée en grande majorité de journées non pluvieuses avec des valeurs proches de 0. Il est précisé sur la figure 9.9 que les réseaux de PE-AROME utilisés sont exclusivement ceux de 3h UTC, l'évolution des échéances peut donc aussi se traduire en heure de validité. Par exemple, l'échéance 10 correspond à 13h UTC et donc plutôt au début d'après-midi du jour J de la PE-AROME. L'échéance 34 représente aussi 13h UTC mais pour la journée suivante (J+1). Ce graphique permet tout d'abord d'évaluer approximativement la gamme de valeur prise par la dimension. Si on regarde l'intervalle Q1-Q99, cette dimension varie entre -70 et +70 avec une moyenne proche de 0. Elle est à peu près symétrique autour de 0. On repère également sur ce graphique pour la partie positive des quantiles plus élevés autour des échéances 15 et 40. Ces échéances correspondent à 18h UTC, ce maximum s'explique simplement par le cycle diurne qui impacte fortement les précipitations. Ce cycle diurne impacte surtout les précipitations convectives, moins les précipitations stratiformes mais celles-ci ne sont pas ou peu représentées dans les queues de distribution car les intensités sont plus faibles. La partie négative présente également des fluctuations selon les échéances avec des minimums atteints aussi autour des échéances 15 et 40 à chaque fois précédé par une zone avec des valeurs faiblement négatives autour des échéances 8 et 33 correspondant au minimum d'activité des précipitations liés à ce cycle diurne. Ce minimum d'activité est aussi observable sur la partie positive mais de façon un peu moins nette .

De façon plus générale, ces fluctuations selon les échéances s'observent sur l'ensemble des dimensions que ce soit avec 10, 20 ou 40 dimensions. On remarque cependant qu'elles sont beaucoup plus fortes en général pour l'espace à 10 dimensions. La symétrie autour de 0 avec des gammes de valeurs comparables (ici -70/70) s'observe régulièrement mais ce n'est cependant pas valable pour toutes les dimensions. Par exemple, les dimensions 8 pour l'espace à 10 dimensions (figure D.22), 15 pour l'espace à 20 dimensions (figure D.39) ou 26 pour l'espace à 40 dimensions (figure D.70) sont dissymétriques. Un point commun pouvant expliquer cette dissymétrie sera mis en évidence section 9.3.4.

9.3.2 Répartition selon les zones régionales

Le graphique présenté est ici celui au centre en haut de la figure 9.9. Il est construit sur le même principe que le précédent mais en divisant la distribution de départ selon les zones régionales qui sont dans l'ordre sur le graphique : le Nord-Ouest, le Sud-Ouest, le Sud-Est, le Nord-Est et la zone Centrale (cf figure 3.13 pour le détail du découpage). Les distributions sont moyennées sur toutes les échéances.

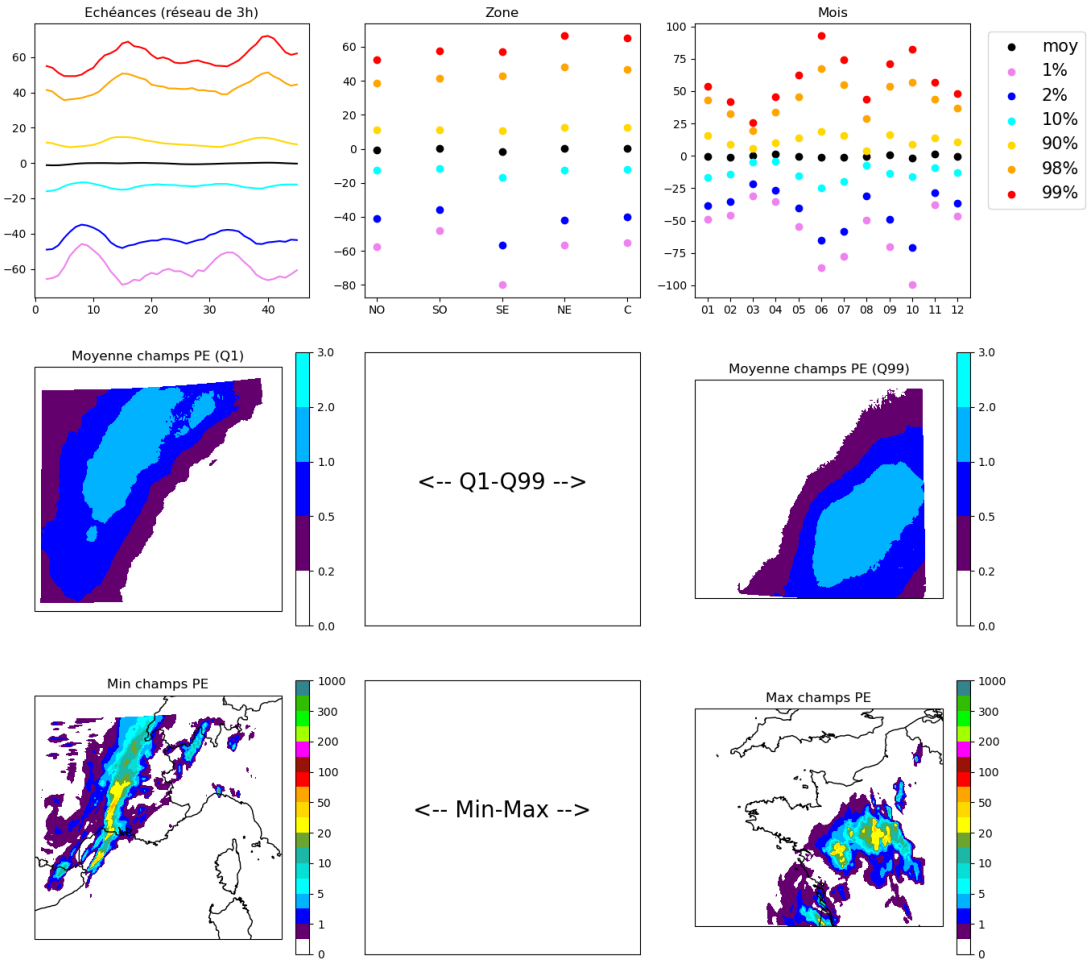


FIGURE 9.9 – *Caractéristiques de la dimension 12 de l'espace latent à 40 dimensions à partir de la distribution calculée comme à l'étape 3 de la figure 9.8. La signification de chaque graphique de cette figure est détaillée dans la section 9.3. La première ligne correspond de gauche à droite à la répartition des valeurs prises par cette dimension selon les échéances (9.3.1), selon les zones régionales (9.3.2) et selon les mois (9.3.3). La deuxième ligne s'intéresse de nouveau à la distribution complète en se focalisant sur le champ moyen des images comprises dans le Q1 à gauche et le Q99 à droite (9.3.4). La dernière ligne s'intéresse aux images correspondants au minimum de la distribution à gauche et au maximum à droite (9.3.5).*

Ce découpage selon les régions permet d'observer des distributions différentes. La zone Sud-Est est associée aux valeurs les plus basses alors que le Nord-Est et la zone Centrale concentrent les valeurs les plus fortes. Cette information seule est difficilement exploitable mais sera complémentaire avec d'autres graphiques ou figures en particulier dans la section 9.3.4. Il est par ailleurs difficile de tirer des conclusions générales pour ce graphique. On notera simplement que la zone Sud-Est semble souvent contenir les valeurs extrêmes des dimensions ce qui peut s'expliquer par la climatologie de cette zone soumise à de plus fortes pluies que le reste de la France, surtout en automne.

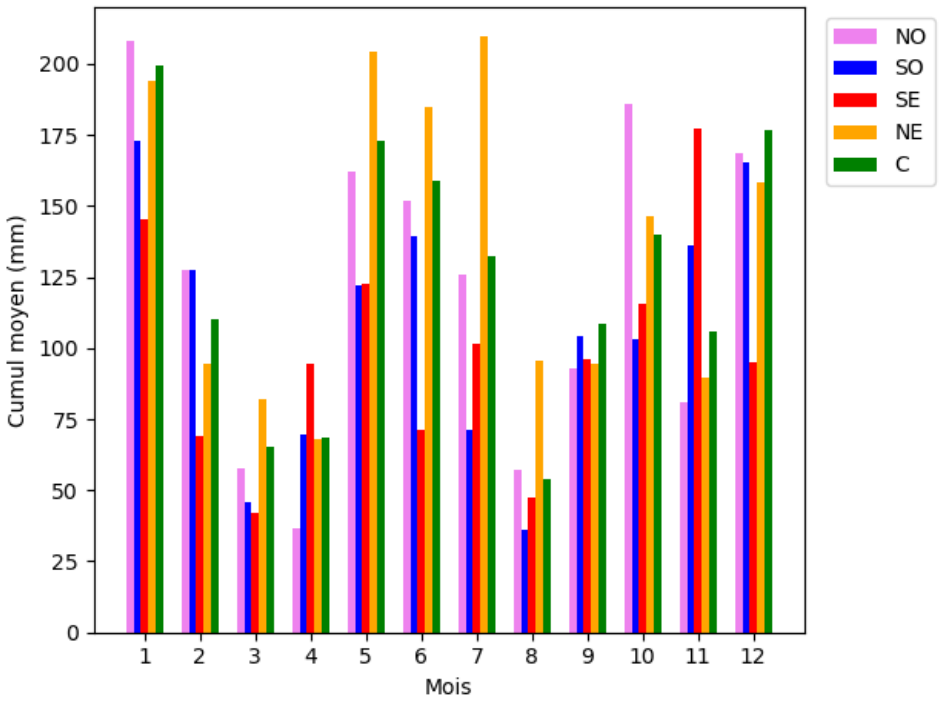


FIGURE 9.10 – *Distribution mensuelle sur l'année 2021 des pluies pour les 5 zones régionales dans les prévisions PE-AROME. Ces cumuls de pluies sont moyennés sur les membres et les points de grille.*

9.3.3 Répartition selon les mois

Le graphique en haut à droite de la figure 9.9 est construit en séparant la distribution selon les mois de l'année 2021. Les mois de mars et août ont des distributions particulièrement resserrées par rapport aux mois voisins. Cette particularité, qui s'observe pour toutes les dimensions, est due à des mois particulièrement sec sur la France (figure 9.10). En l'absence de pluie, il est cohérent d'observer ce signal. A l'inverse, les extrêmes pour cette distribution s'observent en juin et en octobre. En combinant cette information avec celle du paragraphe précédent on peut donc en déduire que les valeurs les plus basses de cette dimension s'observent plutôt sur la zone Sud-Est, et au mois d'octobre ce qui est un signal cohérent si ces pluies sont associées aux épisodes méditerranéens. Les valeurs les plus fortes s'observent généralement sur la zone Nord-Est ou Centrale et au mois de juin qui est aussi un mois avec une activité convective importante à l'échelle nationale et plus particulièrement sur ces zones (figure 9.10). Ce signal est donc également cohérent. Plus généralement, la plupart des dimensions présentent des extrêmes sur les mêmes mois de l'année (fin du printemps, été, automne) au détriment de l'hiver ce qui corrobore également l'observation sur les pluies convectives qui peupleraient plus les queues de distribution que les pluies stratiformes, majoritairement observées en hiver.

9.3.4 Champ moyen des pluies associées au Q1 ou au Q99

Les deux cartes présentées dans cette partie correspondent à la deuxième ligne de la figure 9.9. Ces cartes sont construites en repartant de la distribution générale présentée à l'étape 3 de la figure 9.8. A partir de cette distribution, le Q1 et le Q99 sont calculés. L'objectif est alors de remonter aux champs de pluies qui, après projection dans l'espace latent, se retrouvent en dessous du Q1 pour la carte de gauche ou au dessus du Q99 pour la carte de droite. Les 12 848 champs de pluies qui vérifient la condition sont ensuite moyennés pour former la carte de gauche pour le Q1 et la carte de droite pour le Q99. Les champs tracés sont donc la moyenne des champs de pluies associés aux plus faibles ou fortes valeurs de la dimension étudiée. Pour ces cartes, la palette de couleurs a été modifiée car on ne dépasse que rarement les 3 mm. Les champs sont projetés sur la zone régionale qui correspond à la carte située en dessous sur la troisième ligne (cf section 9.3.5). Ce point explique pourquoi les projections sont par moment un peu différentes comme ici avec la carte de gauche pour le Q1 projeté sur la zone Sud-Est et la carte de droite sur la zone Nord-Ouest. Les cartes pour cette dimension font clairement ressortir deux zones différentes, la moyenne des Q1 correspond à des pluies un peu au-dessus de la diagonale Sud-Ouest/Nord-Est alors que la moyenne des Q99 correspond à des pluies un peu au-dessous de cette même diagonale. Cette observation montre que la dimension ne code pas du tout la même information entre sa partie négative et sa partie positive. Il semble même que les zones concernées soient complémentaires, au sens que la somme (ou union) des deux zones de pluies corresponde à des pluies sur la totalité de l'image alors que l'intersection de ces zones est nulle. Ces cartes donnent également un éclairage important aux graphiques précédents. Les valeurs minimales correspondent majoritairement à la zone Sud-Est au mois d'octobre. Cette enveloppe de pluie sur la carte de gauche, si elle est transposée sur la zone Sud-Est, correspond géographiquement aux Cevennes en remontant vers la Vallée du Rhône. Ces zones sont particulièrement exposées aux remontées de fortes pluies depuis la Méditerranée à l'automne. L'ensemble des informations semble donc cohérent. De même pour la carte de droite qui correspond sur la zone Nord-Est aux Alpes et au Jura ou pour la zone Centrale au Sud-Est du Massif Central. Ces zones de relief sont aussi très exposées à la convection mais plutôt à l'été ou au printemps pour les Alpes et le Jura et donc les valeurs maximales sont plutôt observées en juin pour cette dimension.

Ces cartes permettent également de revenir sur une observation faite section 9.2.2 où on observait que la dimension 8 de l'espace à 10 dimensions réagissait de façon linéaire lorsqu'on augmentait la taille de l'objet au centre de l'image (figure 9.4). L'étude de cette dimension (figure D.22) montre une large zone de pluie qui couvre le centre de l'image pour la partie positive (carte droite, Q99). L'augmentation de la taille de l'objet dans l'expérience de la partie précédente avait comme effet d'avoir un objet de plus en plus imposant mais toujours situé dans cette large zone ce qui a entraîné a priori une augmentation continue de la valeur de la dimension 8. Ces cartes pour le Q1 et le Q99 permettent également de revenir sur une autre observation dans la section 9.3.1 sur des dimensions dissymétriques (figures D.22, D.39, D.70). On peut observer pour ces dimensions que les zones de pluies sont de tailles très différentes entre le Q1 et le Q99 avec des zones très étendues pour l'un des deux quantiles correspondant à la partie de la dimension

avec de fortes amplitudes. Cette forte dissymétrie semble donc s'expliquer par des zones couvertes par les parties positives et négatives de la dimension de tailles très différentes.

De façon plus globale, on observe pour toutes les dimensions cette séparation nette entre partie négative et positive ainsi que cette complémentarité entre les deux cartes. On remarque que les zones de pluies sont de tailles en général plus petites lorsque la taille de l'espace latent tend à augmenter ce qui permet a priori des reconstructions de champs de plus en plus précises. Certaines dimensions deviennent cependant de plus en plus difficiles à analyser en augmentant la taille de l'espace latent comme la dimension 34 (figure D.78).

9.3.5 Champ de pluie associé aux extrêmes de la distribution

La dernière partie à expliciter de cette figure 9.9 est la troisième ligne. Les cartes correspondent aux champs de pluies associés à la valeur minimale (gauche) ou maximale (droite) de la distribution complète (figure 9.8). La projection dépend de la zone géographique qui correspond à ce champ de pluie. C'est pourquoi les cartes de la deuxième ligne ont été alignées sur la même projection pour faciliter les comparaisons Min/Q1 d'un côté et Max/Q99 de l'autre. On remarque une bonne adéquation dans la colonne de gauche avec les remarques dans les parties précédentes car le champ de pluie est sur la zone Sud-Est et correspond au schéma décrit dans la section précédente (pluies des Cevennes vers la Vallée du Rhône). Cette adéquation est moins vraie pour la carte de droite car elle appartient au Nord-Ouest qui pourtant est la zone dont le Q99 est le plus bas pour cette dimension (cf graphique en haut au centre, figure 9.9). Ce champ de pluie est cependant cohérent avec la carte sur la deuxième ligne. Pour les autres dimensions en annexe, on remarque en général que ce minimum ou maximum est associé à de fortes pluies et aucune incohérence en comparant les cartes Q1 et minimum ou Q99 et maximum n'est relevée. Les 5 zones géographiques sont représentées dans ces cartes, même si la zone Sud-Est est plus souvent présente. Par exemple si on considère les minimums et maximums pour chaque dimension de l'espace qui en compte 40 (donc un total de 80 cartes), la zone Sud-Est est représentée 38 fois, le Nord-Est 16 fois, la zone Centrale 12 fois, le Nord-Ouest 9 fois et le Sud-Ouest 5 fois. Le minimum ou le maximum d'une dimension peut, en particulier sur la zone Sud-Est, correspondre à un objet très petit mais très intense comme pour le minimum de la dimension 30 (figure D.74). Cette information accrédite encore l'hypothèse que les objets très intenses se retrouvent dans les queues de distribution, elle est également cohérente avec la figure 9.5 qui montrait une augmentation linéaire des valeurs des dimensions en réponse à une augmentation d'intensité.

Ce chapitre a permis de mieux comprendre les espaces latents et leurs différences à travers une comparaison des attributs des objets en entrée et en sortie, des structures de pluies idéalisées et une étude dimension par dimension à l'aide d'une importante base de données. Plusieurs éléments importants ont pu être mis en évidence comme le filtrage des plus petits objets surtout si le nombre de di-

mensions est limité. Plus généralement, l'autoencodeur tend à lisser les champs de pluies se traduisant par des intensités de pluies en sortie de l'autoencodeur plus faibles. Chaque dimension code l'information de façon locale avec deux structures différentes pour les parties positives et négatives qui sont complémentaires. L'augmentation de la taille ou de l'intensité des pluies sur la zone que représente la partie positive (ou négative) de la dimension tend à augmenter (ou diminuer) la valeur de celle-ci. Ce comportement se rapproche d'une ACP mais contrairement à celle-ci l'autoencodeur peut coder deux informations complémentaires dans les parties positives et négatives de la dimension, ce qui n'est pas le cas d'une ACP. Le comportement non linéaire de l'encodeur et du décodeur est aussi une différence fondamentale par rapport à une ACP. Concernant les ondelettes, celles-ci sont limitées en particulier par l'ondelette mère qui est prédéfinie et qui ne peut pas parfaitement correspondre au champ étudié. L'autoencodeur a d'une certaine façon la possibilité de la définir lui-même. Il peut également varier les motifs dans ces dimensions, ce que peut plus difficilement faire une approche par ondelette.

Bibliographie

E. Gilleland, G. Skok, B. G. Brown, B. Casati, M. Dorninger, M. P. Mittermaier, N. Roberts, and L. J. Wilson. A Novel Set of Geometric Verification Test Fields with Application to Distance Measures. *Mon. Wea. Rev.*, 148(4) :1653–1673, 2020. doi : 10.1175/MWR-D-19-0256.1.

A. McGovern, R. Lagerquist, D. J. Gagne II, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith. Making the Black Box More Transparent : Understanding the Physical Implications of Machine Learning. *Bull. Amer. Meteor. Soc.*, 100(11) :2175–2199, 2019. doi : 10.1175/BAMS-D-18-0195.1.

A. Mounier, L. Raynaud, L. Rottner, M. Plu, P. Arbogast, M. Kreitz, L. Mignan, and B. Touzé. Detection of Bow Echoes in Kilometer-Scale Forecasts Using a Convolutional Neural Network. *Artificial Intelligence for the Earth Systems*, 1 (2) :e210010, 2022. doi : 10.1175/AIES-D-21-0010.1.

K. P. Murphy. *Machine learning : a probabilistic perspective*. MIT press, 2012.

Chapitre 10

Structures météorologiques récurrentes

Le chapitre précédent a permis d'explorer et de mieux comprendre les espaces latents des autoencodeurs. L'étape suivante pour mettre en place la synthèse par scénarios est la construction des structures météorologiques récurrentes sur lesquelles, après projection dans l'espace latent, les membres de la PE-AROME du jour seront assignés pour former les groupes de membres (cf chapitre 3, section 3.4.2). Ce chapitre va s'intéresser à la construction de ces structures météorologiques. Elles seront déduites d'une classification faite sur une base de données regroupant des membres PE-AROME sur plusieurs années. La méthode jusqu'ici identique pour traiter les 5 zones régionales va devenir spécifique à chaque région car les structures météorologiques ne sont pas transposables d'une zone à l'autre. Cette étape doit donc être réalisée dans son intégralité à 5 reprises sur les 5 zones régionales. Par manque de temps et pour permettre de décrire l'intégralité de la méthode dans ce manuscrit, **les deux derniers chapitres se focaliseront exclusivement sur la zone Sud-Est**. Cette zone a été choisie car elle connaît un climat varié entre les zones de montagne, les bords de la Méditerranée et un climat plus continental au Nord. Elle concentre également la plupart des épisodes de pluies les plus intenses en France métropolitaine ce qui rend la variable RR1h particulièrement pertinente pour cette zone. Pour définir ces structure météorologiques récurrentes sur la zone Sud-Est, le premier point de ce chapitre détaille la base de données retenue pour faire cette classification. Le deuxième point présentera rapidement différentes méthodes de classification existantes et comment celle utilisée pour la synthèse par scénarios a été choisie. Le point suivant présente la méthode pour visualiser les classes créées. La sélection de la taille de l'espace latent (10,20 ou 40 dimensions) sera faite et justifiée dans un quatrième point, tout comme le nombre total de structures météorologiques (ou classes) différentes dans un dernier point.

10.1 Base de données

La base de données doit être construite pour représenter au mieux la climatologie des pluies simulées par la PE-AROME sur la zone Sud-Est. Elle doit donc

s'appuyer sur une période de temps la plus longue possible sachant que la PE-AROME est opérationnelle depuis l'automne 2016 (section 1.4.2). La base sélectionnée représente 4 ans de données du 1er septembre 2016 au 31 août 2020. Cette base a été définie de telle sorte à exclure l'automne 2020 car plusieurs études de cas pourraient être intéressantes à mener pour tester la méthode par scénarios et l'année 2021 est préservée pour établir des scores sur une longue période de façon indépendante dans le chapitre suivant. Comme pour la base d'apprentissage dédiée à l'autoencodeur dans le chapitre 6, deux réseaux PE-AROME sont retenus par jour. Seulement les membres 1 à 8 sont extraits pour les mêmes raisons que dans le chapitre 6 (temps d'extraction très long et place mémoire limitée). Toutes les échéances de +2h à +45h sont utilisées. Même en ne se focalisant que sur la zone Sud-Est, la base de données fait alors plus d'un million d'échantillons [1] ce qui est inutilisable à cause des temps de calcul que cela induirait pour manipuler cette base. Il est donc nécessaire de réduire le jeu de données, mais la pluie étant un évènement plutôt rare le jeu de données compte de nombreux champs de RR1h avec peu pas pas de pluie. Il est souhaitable de conserver préférentiellement les champs de RR1h avec de la pluie mais en ne réduisant pas le nombre de dates pour conserver une climatologie sur 4 ans dont la profondeur n'est déjà pas très importante. Il n'est pas souhaitable non plus de réduire le nombre d'échéance car la variable utilisée est le cumul de pluie sur 1h qui est aussi le pas de temps entre deux échéances. Il existerait un risque d'omettre des prévisions de fortes pluies importantes pour la climatologie de la PE-AROME si le nombre d'échéance était arbitrairement réduit. Le choix est donc fait de réduire le nombre de membre, surtout que ceux-ci peuvent être assez ressemblants aux premières échéances et apporter une information redondante. La sélection arbitraire d'un membre ou un tirage aléatoire ne semble pas une bonne solution au risque là aussi d'omettre des cas importants. Pour résoudre ce problème, pour chaque jour, réseau et échéance, la norme 2 dans l'espace physique est calculée pour chacun des 8 membres. Seulement le membre associé à la norme 2 la plus élevée est conservée dans la base de données ce qui permet de diviser sa taille par huit et d'obtenir une base de 128 436 échantillons (ou individus). Cette base est alors projetée dans les espaces latents pour former 128 436 vecteurs de 10, 20 ou 40 dimensions.

10.2 Méthodes de classification

L'emploi d'une méthode de classification est nécessaire pour former différentes classes qui seront les futures structures météorologiques récurrentes et donc un élément essentiel de la méthode par scénarios. Il est nécessaire en amont de s'interroger sur la meilleure méthode pour créer ces classes car il en existe plusieurs. Le problème présenté ici est bien différent de ceux étudiés dans la partie 2 dédiée au deep learning. Un jeu de données X est disponible mais pas de label ou de variable cible Y car il est impossible de définir les classes recherchés en amont. On parle de méthodes non supervisées. Pour créer des classes, il est également nécessaire d'évaluer la similarité de deux individus. Cette mesure de similarité doit permettre à la méthode de classification d'évaluer si ces deux individus seront suffisamment ressemblants ou non pour être dans la même classe. Il est donc

1. Un échantillon étant une grille de 256×384 points

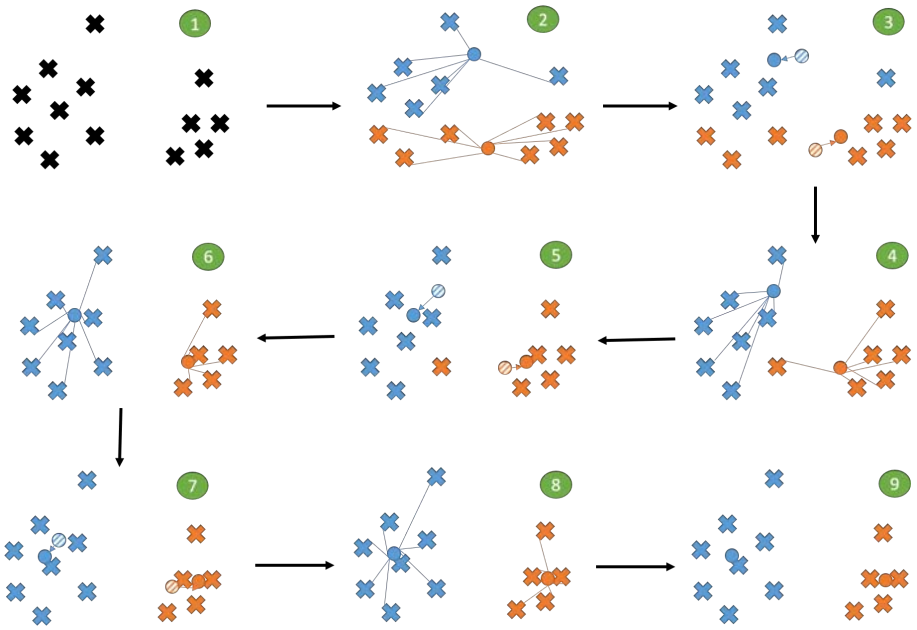


FIGURE 10.1 – *Représentation schématique de l'algorithme des K-Means. Un jeu de données est représenté par des croix à l'étape 1. L'objectif est de créer deux classes (bleu et orange), deux points qui représentent les futurs centroïdes des deux classes sont créés à l'étape 2. L'assignation des croix à l'une des deux classes est représentée par leur couleur. La position des centroïdes changent à l'étape 3 avec la position ancienne qui apparaît sous forme d'un rond hachuré. Les étapes 2 et 3 sont répétées jusqu'à la classification finale étape 9.*

toujours nécessaire avec une méthode de classification de définir une distance (ou métrique). Plusieurs familles de méthode existent et sont présentées succinctement dans les sous-parties suivantes.

10.2.1 Méthodes par partitionnement

Un des principaux algorithmes qui appartient à cette famille est l'algorithme des K-Means ou K-Moyennes en français (Forgy, 1965; Lloyd, 1982). Son principe de fonctionnement est détaillé à travers un exemple figure 10.1. Cet algorithme nécessite la définition au préalable du nombre de classes recherchés, K. Le même nombre, K points qui seront les futurs centroïdes des classes, sont créés et positionnés aléatoirement [2] dans l'espace des données (étape 2, ici $K = 2$ avec les points bleu et orange). K classes sont créées, chaque individu étant assigné au centroïde qui est le plus proche au sens de la distance euclidienne (étape 2, coloration des croix). L'étape suivante consiste alors de repositionner les K centroïdes pour qu'ils correspondent à la moyenne des positions des individus appartenant à leur classe (étape 3). Ce changement de position implique un possible changement dans l'assignation des individus, i.e le centroïde le plus proche n'est peut-être plus le même pour chaque individu. L'étape 2 d'assignation est ainsi répétée, tout comme l'étape 3 (étapes 4 et 5). Ces étapes sont réitérées jusqu'à un critère d'arrêt défini (un maximum d'itérations) ou lorsque plus aucun individu ne change

2. Ce positionnement n'est pas toujours totalement aléatoire. Cet aspect sera abordé plus loin section 10.2.5

de classe comme c'est le cas entre l'étape 6 et 8 où les centroïdes ont changé de position mais où chaque individu n'a pas changé de classe. L'algorithme a alors fini de converger et peut être stoppé (étape 9). K classes sont ainsi créées avec K centroïdes qui correspondent aux barycentres des classes. Plusieurs variantes de cet algorithme existent comme les K-médianes ou les K-Medoids. Dans ce dernier cas, les K centroïdes correspondent à des individus du jeu de données et ne peuvent pas se déplacer librement dans l'espace. L'algorithme des K-Means ne s'utilise qu'avec la distance euclidienne alors que d'autres distances s'utilisent avec les K-Medoids. Ces méthodes présentent l'avantage de créer des classes plutôt homogènes autour des centroïdes et elles sont relativement peu coûteuses par rapport à d'autres méthodes en temps de calcul. Elles présentent cependant l'inconvénient d'être sensibles à l'initialisation des K centroïdes. Les résultats ne sont donc pas toujours reproductibles car l'algorithme peut converger vers différents minimums locaux. Ces méthodes sont également sensibles aux données aberrantes et sont peu performantes pour les jeu de données qui ne sont pas linéairement séparables.

10.2.2 Méthodes hiérarchiques

Contrairement à la méthode K-Means qui effectue un partitionnement de l'espace en minimisant une fonction coût, une classification hiérarchique procède par agrégation (classification ascendante) ou désagrégation (classification descendante) d'individus selon leur similarité. Les classifications hiérarchiques (Johnson, 1967) nécessitent au préalable le calcul de la matrice des distances, i.e les distances deux à deux entre tous les individus du jeu de données. La méthode la plus courante est la classification ascendante hiérarchique. On commence ainsi à rassembler les deux individus les plus proches du jeu de données. La distance entre ce groupe et les autres individus du jeu de données est calculée pour de nouveau établir une matrice des distances, ce groupe remplaçant les deux individus. Cette étape nécessite la définition d'une nouvelle métrique pour définir la liaison/proximité entre des groupes et des individus (ou entre deux groupes). En effet cette liaison peut être définie comme la distance entre un individu et celui qui est le plus proche de lui dans le groupe, ou à l'inverse celui qui est le plus éloigné, ou encore entre l'individu et le centre de gravité du groupe. La liaison dans l'exemple figure 10.2 est indiquée comme complète et correspond à la distance maximale. Après redéfinition de la matrice des distances à l'aide de la liaison choisie, on regroupe alors de nouveau les individus (ou groupes) qui possèdent la distance la plus faible entre eux. Le processus est réitéré jusqu'à regrouper l'intégralité du jeu de données. L'ensemble de ces étapes permet d'établir un arbre de classification appelé dendrogramme (figure 10.2, gauche) qui représente les regroupements créés entre les individus du jeu de données qui sont en abscisse. L'ordonnée est une mesure de dissimilarité, donc plus la liaison entre deux groupes ou individus est basse dans le dendrogramme, plus ils sont ressemblants. La définition des classes repose sur une coupe du dendrogramme à un niveau fixé. Il existe plusieurs méthodes pour fixer ce niveau qui ne sont pas détaillées ici. On obtient dans l'exemple figure 10.2 quatre classes différentes matérialisées par les quatre couleurs. Les méthodes hiérarchiques proposent une structuration générale des données à plusieurs échelles qui peut être un atout dans plusieurs situations. Ces méthodes s'adaptent à toutes les métriques. Elles ont cependant plusieurs inconvénients comme un coût de cal-

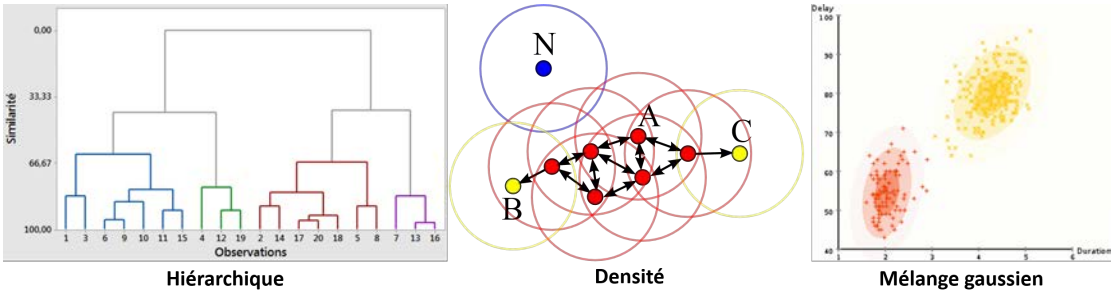


FIGURE 10.2 – *Représentation schématique d'autres méthodes de classification que la famille des K-Means. Les méthodes hiérarchiques sont représentées par un dendrogramme sur la figure de gauche, chaque classe est représentée par une couleur (Source : Minitab). Les méthodes à densité sont représentées par plusieurs points avec leurs voisinages représentés sous la forme de cercles. Les points au cœur de la classe sont en rouge, les points à la frontière de la classe sont en jaune et le point N est un point isolé qui n'appartient pas à la classe (Source :* `https://commons.wikimedia.org/wiki/File:DBSCAN-Illustration.svg`*). Les méthodes basées sur un mélange gaussien, où chaque classe est représentée par une gaussienne dont les paramètres sont à déterminer, sont représentées sur la figure de droite. On suppose dans cet exemple l'existence de deux classes jaune et rouge, la dernière itération est ici représentée avec les densités de chaque gaussienne en arrière plan (Source :* `https://commons.wikimedia.org/wiki/File:EM_Clustering_of_Old_Faithful_data.gif`*).*

cul très important à cause de la matrice des distances. Ces méthodes sont aussi très sensibles à l'ajout ou la suppression d'un seul individu dans le jeu de données.

10.2.3 Méthodes par densité

Une classification par densité procède de façon assez différente des deux types de classification précédentes puisqu'ici il s'agit de détecter comme classes les zones denses, c'est-à-dire contenant beaucoup d'individus, tandis que les zones peu denses seront considérées comme du bruit. Un des principaux algorithmes de classification utilisant la densité des points est l'algorithme DBSCAN (Ester et al., 1996). Cet algorithme nécessite la définition d'un voisinage ϵ autour des individus et un nombre minimal d'individus Min_C à fixer pour former une classe. L'algorithme parcourt les individus du jeu de données jusqu'à en trouver un qui a dans son voisinage ϵ au moins Min_C individus. Une classe est alors formée, elle est complétée en cherchant les individus dans le voisinage de ce premier individu qui vérifient aussi les mêmes conditions sur ϵ et Min_C. Ces individus appartiennent alors à la même classe, et la méthode est ensuite réappliquée sur ces nouveaux individus et ainsi de suite, la classe se construisant de proche en proche. L'ensemble de ces individus sont les points centraux de la classe (points rouges, figure 10.2 au centre). Les individus dans le voisinage de points centraux mais qui ne vérifient pas la condition sur Min_C sont dans la classe mais des points frontières (points jaunes, figure 10.2). Enfin certains individus peuvent n'appartenir à aucune classe, on dit que ces individus sont isolés (point bleu, figure 10.2). D'autres méthodes similaires existent comme l'algorithme OPTICS (Ankerst et al., 1999)

ou HDBSCAN (Campello et al., 2013), ce dernier mixte une approche par densité et hiérarchique. Ces méthodes ont l'avantage de ne pas nécessiter la définition explicite du nombre de classes et ne sont pas sensibles aux données aberrantes car classées comme isolées. Il est cependant souvent difficile d'estimer les paramètres ϵ et Min_C et cette méthode nécessite des classes avec des densités comparables ce qui n'est pas toujours le cas.

10.2.4 Méthodes par mélange gaussien

Les méthodes par mélange gaussien reposent sur la représentation de chaque classe par une gaussienne dont les paramètres sont à déterminer (figure 10.2, droite). Ces paramètres sont le vecteur moyen et la matrice de variance-covariance. Ces méthodes nécessitent, comme pour la famille des K-Means, la définition au préalable du nombre de classes. Les paramètres des gaussiennes sont déterminés de façon itérative en maximisant la vraisemblance de l'échantillon global qui suit la loi d'un mélange gaussien. Cette optimisation se fait à l'aide de l'algorithme d'espérance-maximisation (Dempster et al., 1977). Grâce à la représentation des classes par une gaussienne, un formalisme probabiliste permet ensuite de calculer pour chaque individu la probabilité d'appartenance à chacune des classes. Il est alors assigné à la classe dont la probabilité est la plus élevée. Le formalisme probabiliste est un des principaux avantages de ces méthodes qui permettent de repérer les individus à la frontière de plusieurs classes. Ce formalisme permet aussi de repérer des classes mal formées si par exemple tous les individus qui composent une classe ont des probabilités d'appartenance assez faibles, ce qui induirait que peut-être le nombre de classes n'est pas adapté. Un des principaux inconvénients est l'hypothèse forte qui est faite sur les classes par l'intermédiaire de leur représentation par une gaussienne. Selon le nombre de dimensions dans lequel est représenté le jeu de données et le nombre de classes, le nombre de paramètres à déterminer peut vite croître et rendre la mise en œuvre de la méthode complexe.

D'autres méthodes de classification existent encore comme les Self-Organising Maps (SOM, Kohonen 1982) mais elles ne seront pas présentées dans ce manuscrit.

10.2.5 Méthode de classification retenue

Les différentes méthodes présentées dans les sous-parties précédentes proposent des moyens très variés pour faire une classification d'un jeu de données. On peut alors se demander si une méthode est meilleure qu'une autre. Cette idée est contredite par les travaux de von Luxburg et al. (2012) et de Hennig (2015a,b) qui montrent qu'il n'existe pas dans le cas général une méthode meilleure que les autres. Ils montrent que le partitionnement de la donnée (i.e la classification) est une tâche qui ne peut pas être totalement objectivée. Le choix de la distance utilisée pour mesurer la similarité entre deux individus est aussi un choix qui incombe à l'utilisateur et qui peut avoir une grande influence sur le résultat final. De plus les scores qui existent pour mesurer les performances d'une classification ne mesurent que certaines propriétés des classes. Les scores peuvent ainsi ne pas proposer des conclusions similaires si les propriétés mesurées ne sont pas les mêmes. Ces mêmes travaux de von Luxburg et al. (2012) et de Hennig (2015a,b) suggèrent que c'est

à l'utilisateur en amont de s'interroger sur les propriétés espérées pour ses classes. Ces propriétés permettent alors de mettre en avant certaines méthodes plutôt que d'autres. Il en va de même pour les scores ou les distances entre les individus, il est nécessaire de définir ou trouver les formulations qui correspondent au problème étudié (Hennig, 2019).

Dans le cadre de la synthèse par scénarios, on sait que les structures météorologiques récurrentes vont servir de base pour classer les membres et établir des groupes. Une propriété essentielle des classes semble donc être leur homogénéité pour garantir que les membres à l'intérieur de la même classe sont ressemblants entre eux. Un autre point important est la possibilité d'extraire le plus facilement possible quelques propriétés de la classe qui aideront les utilisateurs à se représenter la classe le plus rapidement possible, un peu comme pour les régimes de temps (cf chapitre 3, section 3.4.2). Par exemple, si des membres sont dans la classe 1 et qu'on est en capacité d'expliquer que cette classe est celle des membres avec des pluies sur les Cevennes, cette information même imprécise rend cette classe moins anonyme et donne un premier aperçu des membres contenus dans cette classe. Cette propriété est intimement liée à l'homogénéité des classes. Il est en effet plus facile a priori de représenter succinctement une classe si celle-ci est homogène. Enfin on peut rappeler que le chapitre 3 a montré que cette étape de classification présentait aussi l'avantage d'être en partie modifiable selon les besoins exprimés par les prévisionnistes. Cette modularité passe en particulier par une discussion sur le nombre total de classes à retenir. Pouvoir choisir le nombre de classes est donc ici un élément favorable à une méthode de classification. Cet ensemble de propriétés pousse à choisir une méthode par partitionnement qui présente plus de garantie sur l'homogénéité des classes que les autres. De plus, les méthodes hiérarchiques sont trop coûteuses numériquement (mémoire et temps de calcul) pour ce jeu de données. Les méthodes par densité ne conviennent pas pour cette application car elles ne permettent pas de choisir le nombre de classes. La présence d'individus isolés qui ne sont pas classés est aussi un problème car elle pose la question de la représentation de ces membres dans le cadre d'une vision par scénarios telle que celle décrite dans le chapitre 3. Enfin concernant les méthodes par mélange gaussien, l'hypothèse de représentation des classes sous la forme d'une gaussienne semble très forte dans notre cas compte tenu de nos connaissances actuelles sur le jeu de données. Pour toutes ces raisons, **l'algorithme de classification retenu sera celui des K-Means** avec une distance euclidienne pour évaluer la similarité de deux membres. Cette distance, qui aurait pu être problématique dans l'espace physique à cause de la double peine, semble convenir au regard des résultats du chapitre précédent sur les espaces latents (linéarité de la réponse à l'augmentation de taille, des intensités, ...).

La mise en œuvre de cet algorithme de classification va être un peu plus détaillé dans ce paragraphe que dans la section 10.2.1. Le positionnement initial des K centroïdes se fera à l'aide de la méthode K-Means++ (Arthur and Vassilvitskii, 2007). Cette méthode consiste à positionner aléatoirement le premier centroïde parmi les individus du jeu de données. Le deuxième centroïde est positionné parmi les autres individus du jeu de données avec une probabilité proportionnelle à $D^2(x)$, la distance de l'individu x au premier centroïde. Puis l'étape est réitérée pour positionner les $K-2$ autres centroïdes, $D(x)$ devenant alors la distance au centroïde le plus proche et pas seulement au dernier positionné. Cette initialisation permet de

mieux répartir les centroïdes sur le jeu de données ce qui peut éviter la convergence vers certaines solutions sous-optimales ou peut permettre une convergence plus rapide. Il subsiste néanmoins l'un des principaux défauts de l'algorithme K-Means qui reste sensible à la position initiale de ces centroïdes. Pour surmonter cette difficulté, l'algorithme des K-Means n'est pas lancé une seule fois mais plusieurs fois. Sur ces N tentatives, on conserve la classification qui, après convergence, a l'inertie intraclasse la plus faible. L'inertie intraclasse est la somme des distances au carré entre les individus et le centroïde le plus proche pour chacun d'eux, il s'agit donc d'un critère qui mesure l'homogénéité au sein des classes. Le jeu de données étant très conséquent, le nombre de tentatives est très important et fixé à $N = 1000$. L'algorithme de classification est lancé avec ces paramètres à l'aide du package Python Scikit-Learn (Pedregosa et al., 2011). Ces classifications seront testées pour les 3 espaces latents à 10, 20 et 40 dimensions et pour un nombre total de classes qui est variable. Il restera à sélectionner une taille d'espace latent et un nombre total de classes dans les parties suivantes.

10.3 Représentation synthétique des classes

Avant de sélectionner la taille de l'espace latent et le nombre de classes, cette partie s'intéresse à la visualisation des classes qui guidera en partie les choix faits dans les parties suivantes. Le jeu de données comptant 128 436 individus (section 10.1), la plupart des classes compteront plus de 1 000 individus et certaines plus de 10 000. Il est nécessaire de trouver des méthodes efficaces pour représenter ces classes. Sur le même principe que la section 9.3 qui présentait différents graphiques ou cartes pour étudier dimension par dimension les espaces latents, plusieurs modes de représentation sont rassemblés sur la figure 10.3. Chaque graphique ou carte est présenté dans une sous-partie.

10.3.1 Répartition selon les mois et années

Le graphique en haut à gauche de la figure 10.3 est une première façon d'analyser une classe en regardant comment les individus à l'intérieur de celle-ci se répartissent selon les mois (abscisse) ou selon les années (couleur différente pour chaque année). Le nombre total d'individu dans la classe est aussi mentionné dans le titre de ce graphique (ici 4446). On remarque que cette classe concentre ses individus plutôt sur la période estivale de mai à août, ce qui signifie que cette classe est particulièrement sensible aux pluies liées à de la convection. On remarque cependant que cette classe peut aussi être très représentée en hiver, dominé plutôt par des pluies stratiformes, en particulier en janvier et décembre 2018 qui sont des mois très pluvieux à l'échelle nationale pour le mois de janvier et vers les Alpes pour le mois de décembre (figure E.1 et E.2 en annexe). Ce graphique permet de repérer les mois privilégiés pour une classe ce qui permet d'en déduire la typologie dominante des pluies (stratiforme ou convective) ou encore la signature des épisodes méditerranéen qui se concentreront à l'automne. La distinction année par année permet de repérer s'il existe une forte variabilité, comme c'est le cas ici pour les mois de janvier et décembre.

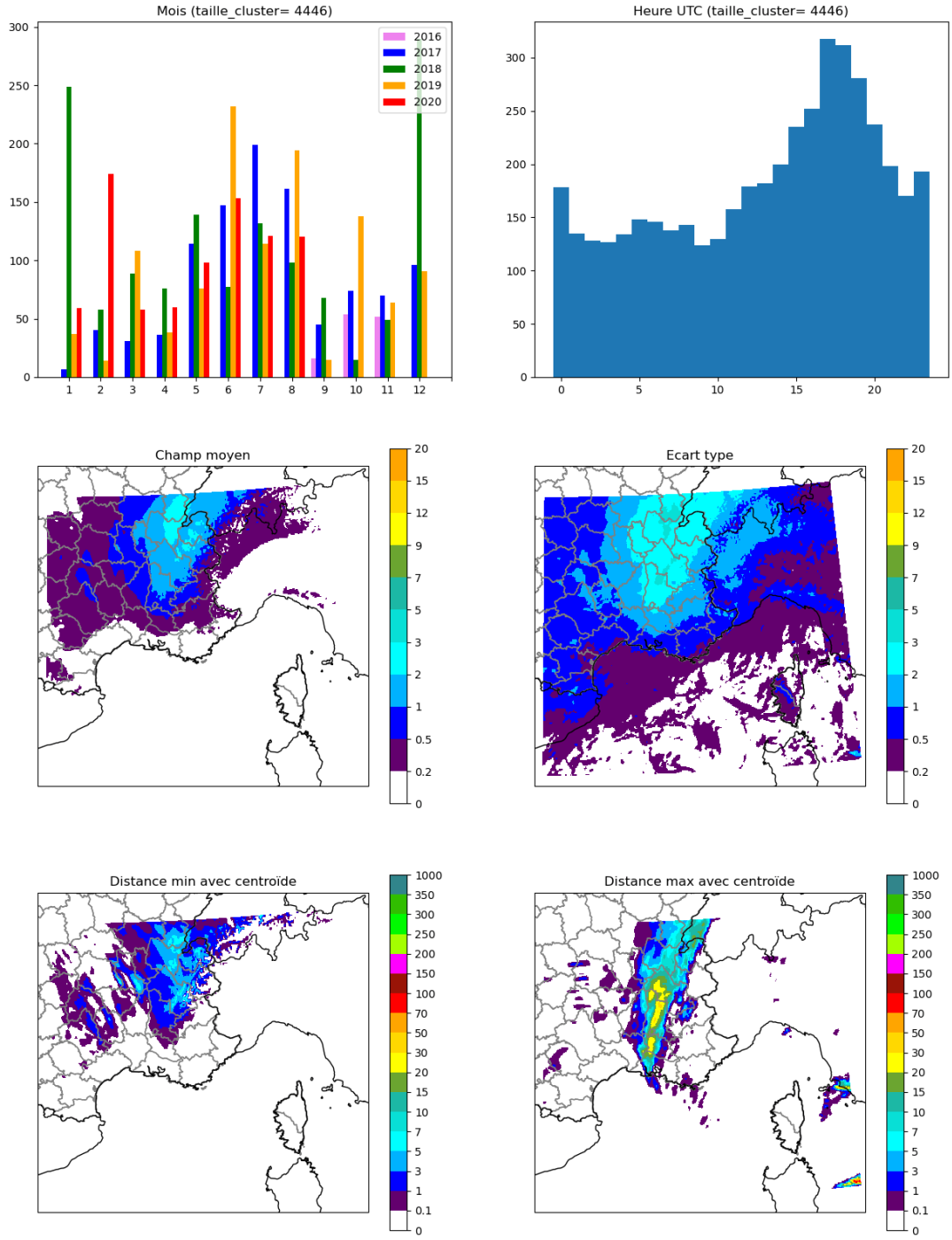


FIGURE 10.3 – *Représentation synthétique de la classe n°20 obtenue à l'aide d'un K-Means avec l'espace latent à 40 dimensions pour un nombre total de classes égal à 20. Chaque graphique ou carte est expliqué dans la section 10.3.*

10.3.2 Répartition selon les heures de validité

Le graphique en haut à droite de la figure 10.3 répartit les individus selon les heures de validité de 0h à 23h UTC. Ce graphique n'est pas exprimé avec les échéances comme dans le chapitre précédent figure 9.9 car la base de données mélange ici plusieurs réseaux. Une échéance de +3h peut correspondre à 6h UTC si le réseau est celui de 3h ou 18h UTC si le réseau est celui de 15h. Chaque échéance et réseau pour chaque individu est donc extrait pour en déduire son heure de validité pour tracer ce graphique. On remarque que le maximum se situe autour de 18h UTC ce qui est concordant avec le cycle diurne et une signature de pluies convectives. Ce graphique conforte les conclusions de la sous-partie précédente. Cette sensibilité au cycle diurne est aussi une différence importante par rapport aux épisodes convectifs intenses qui y sont moins sensibles.

10.3.3 Moyenne/Écart-type

Les cartes sur la deuxième ligne de la figure 10.3 vont permettre de représenter l'emprise spatiale de la classe. Pour cela l'ensemble des champs de pluies correspondant aux individus de la classe, qui eux sont pour rappel exprimés dans l'espace latent, sont extraits. La moyenne de ces champs est tracée sur la carte de gauche et l'écart-type de ces champs de pluies sur la figure de droite. Ces représentations s'expriment en mm, la plage de couleur a cependant été modifiée par rapport aux autres cartes de RR1h dans ce manuscrit pour mieux faire ressortir les contrastes. Le champ moyen montre que la classe concentre les zones de pluies sur le Jura, les Pré-Alpes, les Alpes jusqu'en descendant vers la Drôme. Cette emprise spatiale correspond bien aux descriptions précédentes, cette zone est en effet soumise à la fois à des orages et de l'évolution diurne en été mais aussi à d'importantes précipitations par blocage sur les zones de reliefs en hiver (blocage orographique). Le maximum d'écart-type est aussi sur la même zone géographique mais on remarque que l'écart-type dépasse les 0.5 mm sur une grande partie du domaine. On peut donc en conclure que cette classe regroupe des individus où il pleut préférentiellement sur cette zone du Jura à la Drôme en passant par les Alpes mais qu'il peut aussi pleuvoir sur d'autres zones au même moment.

10.3.4 Champ de pluie le plus proche et le plus éloigné du centroïde

Le dernier point de la sous-partie précédente va être illustré par cette troisième et dernière ligne de la figure 10.3. La méthode des K-Means attribue à chaque classe un centroïde qui est le centre de gravité de la classe. La distance dans l'espace latent entre chaque individu de la classe et le centroïde est calculé. On représente alors sur la gauche le champ de pluie qui correspond à l'individu le plus proche du centroïde et à droite l'individu le plus éloigné. L'individu à gauche correspond bien à la zone délimitée par le champ moyen au-dessus. On peut noter que les pluies sont plutôt faibles et lisses spatialement correspondant à des pluies majoritairement observées en hiver. Cette carte correspond d'ailleurs à une prévision pour le 22 janvier 2018. La carte de droite est au contraire associée à des pluies plus fortes qui recouvrent certes la zone du Jura à la Drôme mais elles descendent aussi beaucoup plus

au sud jusqu'aux Bouches-du-Rhône ce qui explique probablement pourquoi la distance entre le centroïde de la classe et cet individu est grande. Cette carte illustre l'écart-type important sur le domaine montrant des individus avec par moment des pluies sur d'autres zones géographiques que celles correspondant au champ moyen. Ces pluies à caractère convectif sont issues d'une prévision pour le 22 septembre 2021. Cette carte montrant le champ de pluie correspondant au vecteur le plus éloigné du centroïde permet de repérer un des individus « extrêmes » de la classe ce qui peut devenir un des angles pour discuter de la future utilisation de ces classifications sous la forme de question comme : Peut-on accepter que cette prévision appartienne à cette classe ou au contraire, est-il nécessite de faire une autre classification ? Ces graphiques ou cartes pour représenter les classes seront utilisés pour discuter la pertinence de certaines classifications dans les parties suivantes, à commencer par la taille de l'espace latent.

10.4 Sélection de la taille de l'espace latent

10.4.1 Sensibilité de la classification à la taille de l'espace latent

Avant de s'intéresser directement à la sélection d'une taille d'espace latent, cette sous-partie s'intéresse à la sensibilité de la classification à la taille de cet espace. Il est en effet possible qu'en cherchant à classer le même jeu de données au départ dans l'espace physique et avec la même méthode de classification, on aboutisse malgré un changement de taille d'espace latent à des résultats identiques. Dans ce cas, la synthèse par scénarios serait alors peu sensible à la taille d'espace latent retenue.

Une classification associe à chaque individu de la base de données ($x_i, i \in [1, N_{data}]$) une classe ($c_j, j \in [1, N_{classes}]$). Pour mesurer la similarité de deux classifications, plusieurs scores existent pour comparer les associations d'une première classification (x_i^1,c_j^1) et d'une deuxième (x_i^2,c_j^2). Le score utilisé dans ces travaux est l'Adjusted Mutual Information (AMI, Vinh et al. 2010). Ce score est utilisé plutôt que d'autres car il est adapté à des classifications déséquilibrées, i.e avec des tailles de classes très différentes (Romano et al., 2016). Or les classes seront déséquilibrées en particulier à cause d'une classe correspondant aux cas où il ne pleut pas ou presque. Cette classe est ultra majoritaire, sa taille exacte dépendant du nombre total de classes. Concernant l'AMI, la formule de ce score ne sera pas présentée dans ce manuscrit. L'article de Vinh et al. (2010) détaille cette formule si le lecteur souhaite plus d'informations. On retiendra que ce score varie de 0 à 1, 1 correspondant à deux classifications identiques et 0 correspond à l'espérance de ce score si les deux classifications étaient faites de façon aléatoire. Plus le score est élevé, plus les classifications se ressemblent.

Pour tester la sensibilité à la taille de l'espace latent, la classification par les K-Means est réalisée pour un nombre de classes de 2 à 50 et pour les espaces à 10, 20 et 40 dimensions. Pour une taille d'espace latent fixe et pour un nombre de classe fixe, la section 10.2.5 avait précisé que l'algorithme était relancé 1000 fois car on risquait selon les initialisations des centroïdes de converger vers des minimums

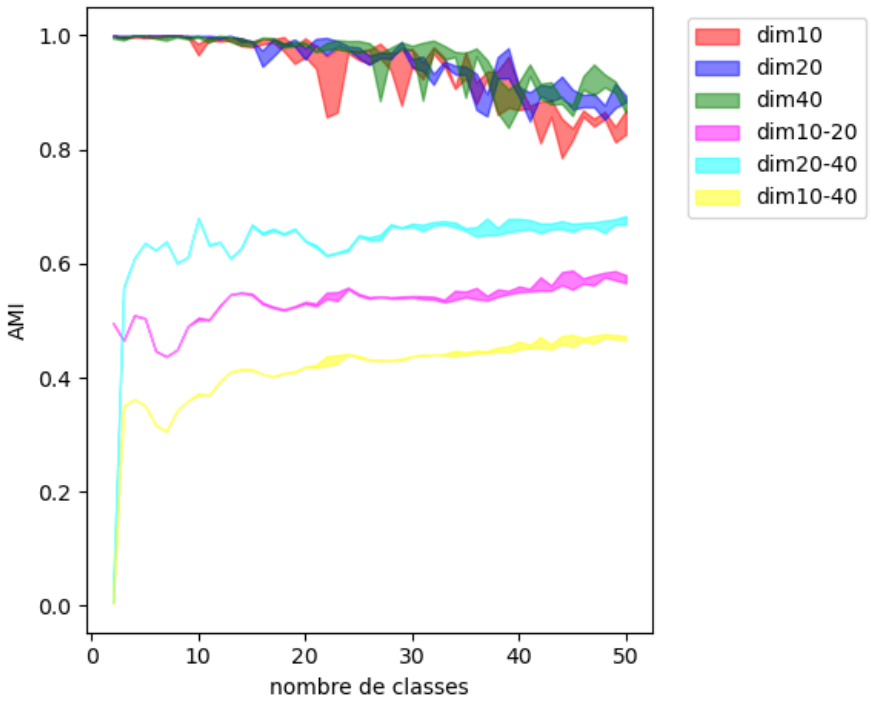


FIGURE 10.4 – *Évolution de l'AMI (ordonnée) selon le nombre de classes (abscisse) et selon les tailles d'espace latent (couleurs différentes). Les panaches représentent l'AMI minimale et maximale en comparant les différentes classifications disponibles (3 tentatives à chaque fois pour un nombre de dimensions et un nombre de classes fixés).*

locaux et qu'un moyen de s'en prémunir était de relancer plusieurs fois l'algorithme. Mais il n'est pas certain à ce stade que 1000 répétitions soient suffisantes et qu'on obtienne en sortie la meilleure classification [3]. Pour le vérifier, ces 1000 répétitions ne sont pas lancées une seule fois mais 3 fois (3×1000). A chaque fois, on garde seulement la meilleure classification (C^1_{best},C^2_{best},C^3_{best}). L'AMI est alors calculée deux à deux entre ces classifications. Si le nombre de 1000 répétitions est suffisant, l'AMI doit alors être proche de 1 pour chaque comparaison deux à deux. Les scores AMI sont donc calculés entre ces 3 classifications pour chaque espace latent (10, 20 et 40) et pour chaque nombre de classes de 2 à 50. Ces scores AMI sont représentés sur la figure 10.4. Les panaches rouge, bleu et vert représentent les comparaisons à taille d'espace latent fixée, avec les limites du panache qui correspondent à l'AMI minimal et maximal calculés pour chaque nombre de classes. Pour un nombre de classes fixé, on peut également calculer l'AMI entre les 3 classifications d'une taille d'espace latent donnée et les 3 classifications d'une autre taille. On obtient alors 9 valeurs différentes de l'AMI. Les classifications d'espace latent avec des tailles différentes sont ainsi comparées également sur la figure 10.4 (panaches cyan, magenta et jaune).

Si on raisonne à taille d'espace latent fixe avec les panaches rouge, bleu et vert, l'AMI est très proche de 1 jusqu'à des classifications à 20 classes, puis elle tend à légèrement décroître tout en restant au dessus de 0.8 jusqu'à 50 classes. En multipliant les classes, on s'intéresse à des classes de plus en plus petites et le par-

3. Pour rappel, la classification associée à la plus petite inertie intraclasse est considérée comme la meilleure.

titionnement des données se complexifie ce qui peut expliquer ce résultat. Un AMI au dessus de 0.8 est néanmoins bien supérieur aux AMI observés lorsqu'on tente de comparer les classifications avec des tailles d'espaces latents différentes (panaches cyan, magenta et jaune). Ces scores se situent entre 0.4 et 0.7 à une exception notable pour 2 classes où l'AMI est proche de 0 quand on compare l'espace à 40 dimensions avec les autres. Pour le cas particulier à 2 classes, la classification tend à proposer une zone avec de la pluie à la frontière de l'Italie ou de l'Espagne et une deuxième classe avec l'ensemble des autres individus. Cette classe minoritaire n'est pas la même (Italie ou Espagne) selon la taille de l'espace latent et explique ce score proche de 0. Mais à partir de 3 classes, la classification propose une classe sur l'Espagne, une sur l'Italie et une autre avec les autres individus quelle que soit la taille d'espace latent ce qui explique la remontée brutale des scores pour les panaches cyan et jaune. La tendance à l'augmentation de l'AMI selon le nombre de classes s'observe par construction du score. Si on pousse jusqu'à un nombre de classes égal aux nombres d'individus, l'AMI sera égal à 1. Cette tendance ne s'observe pas pour les AMI à espace latent fixe car on cherche à vérifier si le nombre de répétition est suffisant. Cette objectif différent ne conduit pas aux observations habituellement faites avec l'AMI. La baisse de l'AMI dans ce cas traduit simplement un nombre de possibilité plus grand quand le nombre de classes augmente et donc une plus grande difficulté à converger de façon reproductible vers le minimum global si le nombre de répétition reste fixe, ce qui est le cas dans ces travaux. Pour conclure, ces courbes montrent que **la taille de l'espace latent influence la classification** et ce quel que soit le nombre de classes. Il est donc important de choisir correctement la taille de l'espace latent.

10.4.2 Taille de l'espace latent retenue

Pour choisir au mieux l'espace latent, il est important de comprendre ce qui diffère dans les classifications d'un espace latent à l'autre. Pour cela, on peut s'intéresser à des tables de contingence pour un nombre de classes fixé et deux tailles d'espaces latents différentes comme celui présenté figure 10.5. Sur cette figure, le nombre total de classes est égal à 3 et les deux espaces comparés sont ceux à 10 et 20 dimensions. Les numéros 1,2 et 3 représentent les numéros des 3 classes. Ce tableau permet de regarder comment les 128 436 individus du jeu de données (section 10.1) sont répartis entre ces classes. Par exemple, 117 134 individus appartiennent à la fois à la classe 1 de l'espace à 10 dimensions et à la classe 1 de l'espace à 20 dimensions, mais ils ne sont que 9 à appartenir à la classe 3 pour l'espace à 10 dimensions et à la classe 2 pour l'espace à 20 dimensions. On remarque avec cette table de contingence que la grande majorité des individus sont dans la classe 1 que ce soit pour l'espace à 10 dimensions ou l'espace à 20 dimensions. Cette classe est celle qui correspond aux cas où il ne pleut pas ou presque avec un champ moyen presque nul (non montré). Cette table montre que les classes 2 et 3 ne se mélangent presque pas (seulement 9 et 11 individus qui passent de l'une à l'autre alors que les classes font plusieurs milliers d'individus). Par contre, il y a beaucoup d'individus pour les classes 2 et 3 de l'espace à 20 dimensions qui étaient dans la classe 1 pour l'espace à 10 dimensions (2185 et 3307 respectivement, soit environ 50% de la taille totale des classes pour l'espace à 20 dimensions). L'inverse est cependant beaucoup moins vrai, il y a très peu d'individus qui étaient dans la

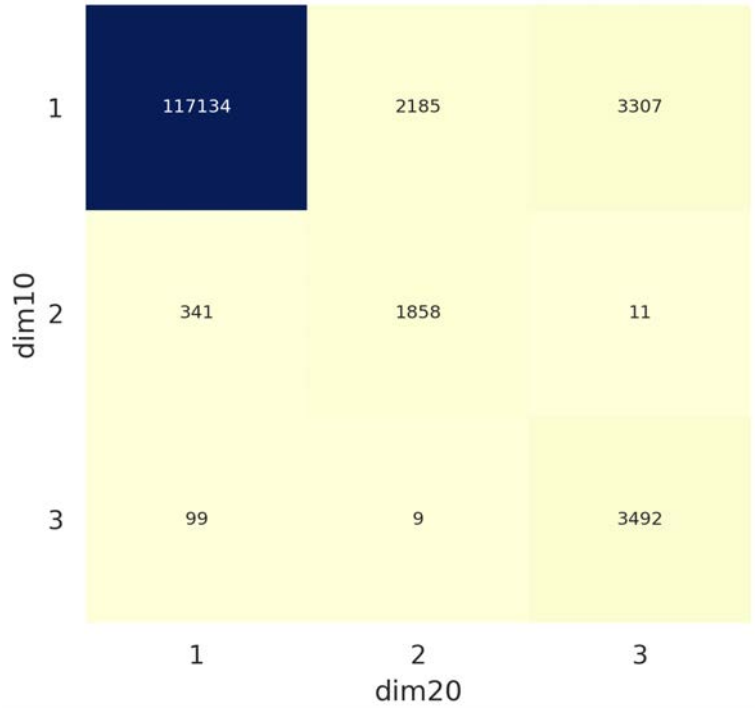


FIGURE 10.5 – *Table de contingence pour un nombre total de 3 classes. Une classification K-Means a été faite sur l'espace à 10 dimensions et une autre sur l'espace à 20 dimensions. On s'intéresse à la répartition des individus parmi ces 3 classes. On peut lire par exemple dans ce tableau que seulement 11 individus sont dans la classe 3 pour l'espace à 20 dimensions et dans la classe 2 pour l'espace à 10 dimensions.*

classe 1 de l'espace à 20 dimensions et qui étaient dans la classe 2 ou 3 de l'espace à 10 dimensions (341 et 99). On peut conclure de cette table de contingence que pour 3 classes, les classifications ne sont pas structurellement si différentes malgré un AMI plutôt faible. En effet, toutes les classes ne se mélangent pas. La différence importante est le transfert conséquent d'individus qui étaient dans la classe 1 (pas ou peu de pluie) avec 10 dimensions vers les classes 2 et 3 (avec pluie) de l'espace à 20 dimensions. Ce transfert concerne des individus avec des pluies trop faibles pour s'extraire de la classe majoritaire avec 10 dimensions mais ces pluies sont suffisamment fortes pour être dans les classes 2 et 3 pour un espace à 20 dimensions. Cette observation semble directement liée aux conclusions du chapitre précédent sur le comportement des espaces latents. On avait montré que plus l'espace latent était petit, plus il avait tendance à filtrer les petits objets de pluie. Il est alors plus difficile voire impossible de différencier les individus avec peu ou pas de pluie s'ils sont codés dans l'espace latent d'une façon similaire. Ce comportement semble expliquer en grande partie la différence entre les espaces à 10 et 20 dimensions concernant les importantes variations du nombre d'individus dans les classes 2 et 3. Cette conclusion qui s'observe ici pour 3 classes se vérifie pour un nombre de classes plus grand (non montré). On observe aussi le même comportement en comparant les espaces à 20 et 40 dimensions (non montré également).

Cette différence dans les individus qui composent la classe majoritaire (correspondant à la classe 1 sur la figure 10.5 avec peu ou pas de pluie) se traduit par des cartes d'écart-type pour cette même classe majoritaire très différentes. Ces cartes pour les 3 tailles d'espaces latents (10, 20 et 40) et pour 3 nombres de

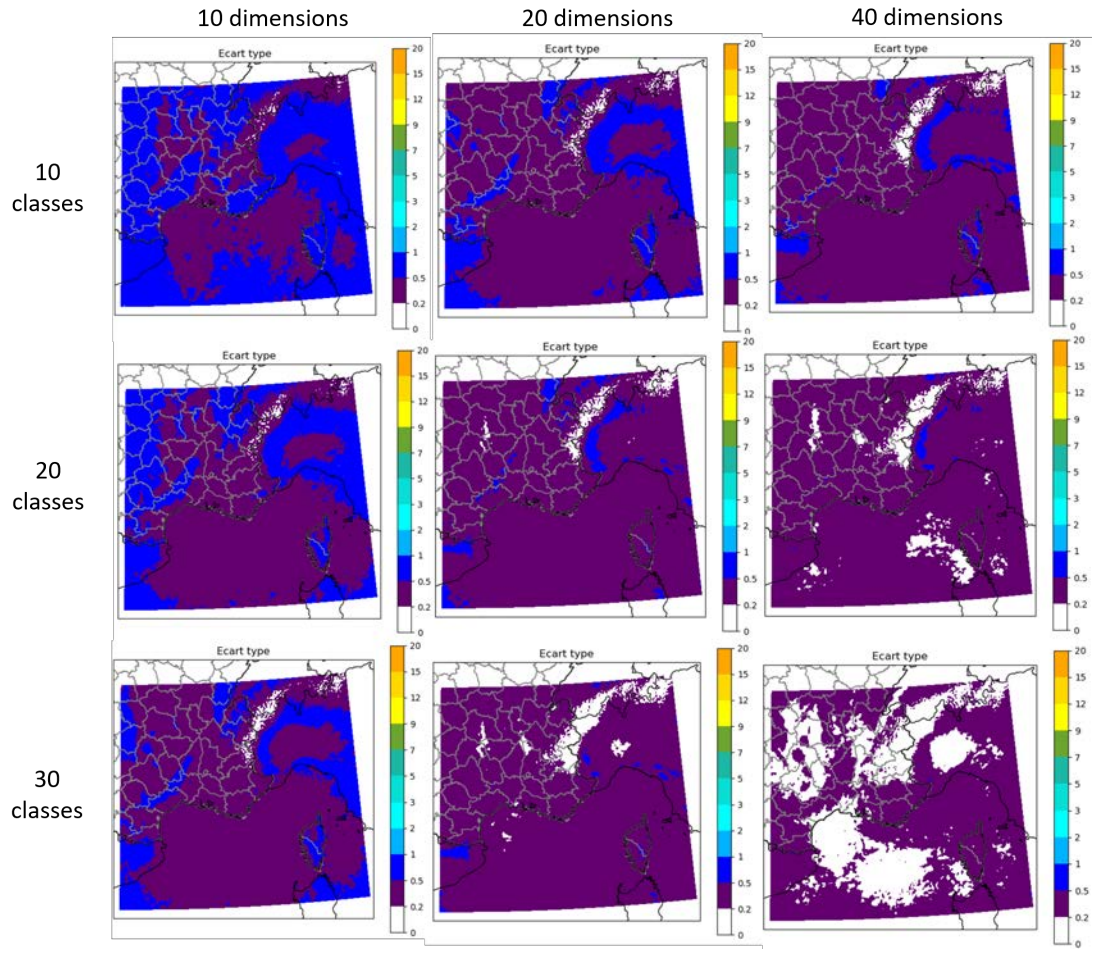


FIGURE 10.6 – *Cartes d'écart-type pour la classe majoritaire selon la taille de l'espace latent (colonne de gauche à droite avec 10,20 ou 40 dimensions) et selon le nombre de classes (ligne de haut en bas avec 10,20 ou 30 classes).*

classes différents (10, 20 et 30) sont présentées figure 10.6. Ces cartes montrent qu'augmenter le nombre de classes permet de réduire l'écart-type et donc d'extraire de la classe majoritaire des champs de pluie non nuls. L'augmentation de la taille de l'espace latent permet aussi de limiter grandement les écart-types et ce quel que soit le nombre total de classes. Ces différences se traduisent également par la taille de la classe majoritaire qui décroît quand on augmente la taille de l'espace latent passant par exemple, pour le cas avec 20 classes au total, de 109 976 individus pour 10 dimensions à 92 820 pour 20 dimensions et 83 342 pour 40 dimensions. Comme il est souhaitable d'avoir pour la synthèse par scénarios une classe majoritaire qui représente au mieux une classe sans pluie ou a minima avec des pluies peu significatives, **ces résultats conduisent à retenir pour la suite l'espace latent à 40 dimensions**. D'autres tailles d'espaces latents plus grandes (60,80 ou 100) ne sont pas testées par manque de temps dans cette thèse.

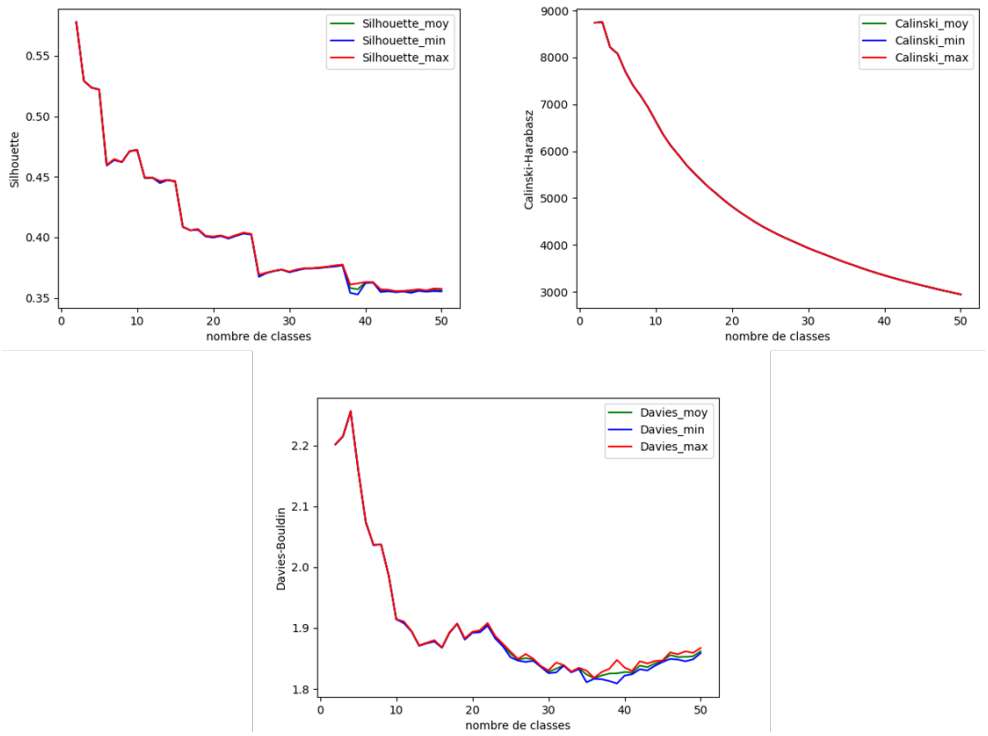


FIGURE 10.7 – *Scores pour établir le nombre de classes avec l'espace latent à 40 dimensions. Les scores de silhouette (haut,gauche), Caliński and Harabasz (haut,droite) et Davies and Bouldin (bas,centre) sont calculés pour les 3 tentatives évoquées section 10.4.1 permettant de calculer le score minimal, moyen ou maximal pour chaque nombre de classes tenté, de 2 à 50.*

10.5 Sélection du nombre de classes

La taille de l'espace latent étant fixée, il reste à fixer le nombre total de classes pour définir la classification qui servira à mettre en place la méthode par scénarios expérimentée dans le dernier chapitre de ce manuscrit. Pour choisir ce nombre de classes, quelques scores classiques ont été testés dans une première partie. La seconde partie présentera l'évolution du nombre de scénarios différents selon le nombre de classes, ce point aidant à déterminer le nombre de classes retenu.

10.5.1 Scores pour fixer le nombre de classes

La section 10.2.5 avait évoqué l'impossibilité de déterminer de façon purement objective la « meilleure » classification. Il en va de même pour la détermination du nombre de classes. Il existe cependant quelques scores pour faire ce choix, ils sont testés dans cette partie pour voir si un nombre de classes émerge. Les trois scores utilisés sont le score de silhouette (Rousseeuw, 1987), l'indice de Caliński and Harabasz (1974) et celui de Davies and Bouldin (1979). Le score de silhouette varie de -1 à 1 avec pour objectif de le maximiser. Il se calcule pour chaque individu de la classification et repose sur la différence entre d'une part la distance moyenne qui sépare cet individu des autres individus de sa classe et d'autre part la distance moyenne qui sépare cet individu des autres individus d'une autre classe (la classe qui est la plus proche de la sienne). Ce score est alors moyenné sur les individus

pour obtenir un score pour la classification complète. Le score de Caliński and Harabasz est le ratio entre la variance inter-classe et la variance intra-classe. Ce score doit donc être maximisé pour minimiser la variance intra-classe et maximiser la variance inter-classe. Enfin l'indice de Davies and Bouldin varie de 0 à $+\infty$ avec l'objectif de minimiser ce score. Ce score se base sur la distance moyenne entre un individu et le centre de son groupe et sur les distances entre les centres des groupes. Ces scores sont détaillés plus amplement dans les références citées précédemment.

Ces scores sont représentés sur la figure 10.7. Le score de silhouette semble indiquer quelques nombres de classes préférentiels autour de 15 classes, 25 classes et 35 classes juste avant de brusques chutes du score. Cette conclusion n'est cependant absolument pas corroborée par les autres scores, le score de Caliński and Harabasz ne propose qu'une décroissance exponentielle du score selon le nombre de classes. Cette décroissance exponentielle s'explique en partie par le choix de la méthode de classification. Les K-Means tendent à minimiser la variance intra-classe par construction qui intervient directement dans le score de Caliński and Harabasz . Le score de Davies and Bouldin propose plutôt un nombre optimal autour de 35-40 classes mais il est plus difficile de fixer un nombre précis comme pour le score de silhouette. Les scores ne semblent donc pas tous concordants car ces scores ne mesurent pas exactement les mêmes propriétés pour évaluer la classification (Hennig, 2015b). De plus, une classification à l'aide d'environ 35 classes semble trop important pour un quart de France, ce nombre ne semble pas cohérent avec l'application à la prévision météorologique. Cette difficulté à fixer objectivement le nombre de classes pour partitionner des champs atmosphériques se retrouvent dans d'autres travaux similaires comme ceux de Fereday et al. (2008) et de Neal et al. (2016). Ces travaux concluent qu'il est impossible de fixer objectivement le nombre de classes et le fixe en fonction de l'application souhaitée. Cette partie montre qu'il n'est a priori pas possible d'obtenir une cohérence entre les différents scores et incite alors à procéder autrement pour fixer le nombre de classes.

10.5.2 Nombre de scénarios différents

Un moyen pour sélectionner un intervalle acceptable du nombre de classes découle directement de la demande exprimée au chapitre 3 de n'avoir pas plus de 5 scénarios simultanément (section 3.4.1). Plus le nombre de classes est grand, plus l'ensemble sera séparé en différents scénarios. En étudiant des sorties PE-AROME sur une longue durée avec une synthèse par scénarios testée avec différents nombres de classes, on peut alors s'intéresser au nombre de scénarios produits et éliminer ainsi les nombres de classes qui séparent l'ensemble en un nombre de scénarios trop grand. Pour réaliser ce calcul, on utilise la base de données de 2021, du 1er janvier au 31 décembre 2021. La synthèse par scénario sera détaillée dans le chapitre suivant mais on retiendra à ce stade qu'on utilise deux PE-AROME successives pour faire du multi-réseaux et obtenir un ensemble de 32 membres au lieu de 16. Le nombre moyen de scénarios différents est alors tracé selon les échéances figure 10.8. Le nombre de scénarios différents correspond au nombre de différentes classes explorées par les membres à une échéance donnée (même si un seul membre sur 32 l'explore). Le nombre moyen de scénarios tend à croître avec les échéances traduisant une augmentation de la dispersion de l'ensemble. Il augmente également quand le nombre de classes augmente mais les valeurs moyennes entre 1.5 et 3

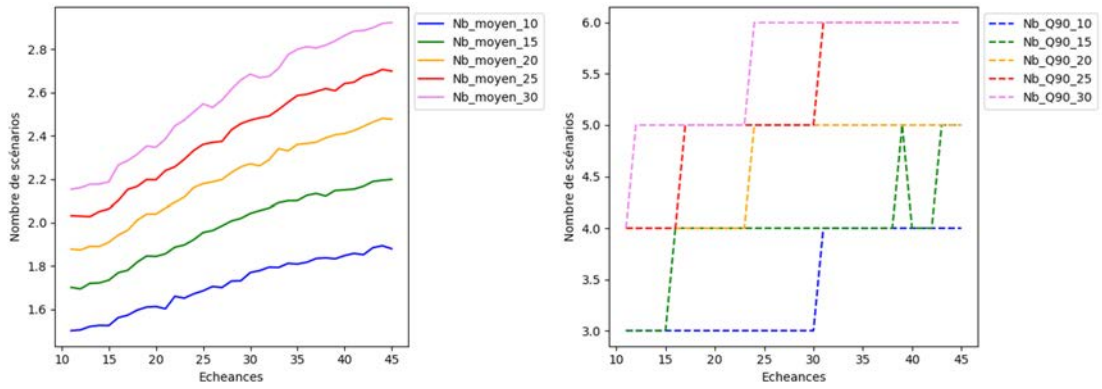


FIGURE 10.8 – *Nombre de scénarios selon les échéances. Deux réseaux PE-AROME sont utilisés avec 6h d'intervalle (PE-AROME 3h + PE-AROME 9h par exemple), l'échéance indiquée correspond à celle du réseau le plus ancien. Chaque courbe représente un nombre de classes différent allant de 10 à 30 avec un pas de 5 classes. Le graphique de gauche correspond à la moyenne pour chaque échéance, le graphique de droite au Q90 de la distribution pour chaque échéance.*

scénarios sont cependant difficilement exploitables pour discriminer le nombre de classes car beaucoup de cas sans pluie avec 1 seul scénario prévu (sans pluie) font chuter cette moyenne. Plutôt que s'intéresser à la moyenne de la distribution selon les dates, le graphique de droite figure 10.8 représente le Q90 compris entre 3 et 6 scénarios. Un nombre de classes fixé à 25 ou 30 semble trop important par rapport aux objectifs du nombre de scénarios inférieur à 5 d'après cette figure. A l'inverse, pour un nombre de classes égal à 10 on semble disposer de trop peu de scénarios et donc peut-être d'une information trop imprécise en ne dépassant pas les 4 scénarios. Ces conclusions incitent donc à chercher le nombre de classes optimal entre 15 et 25. On peut également remarquer que le nombre de classes fixé à 35 comme suggéré par certains scores apparaît trop important (la courbe serait encore supérieure à celle pour 30 classes).

Le dernier problème à résoudre est de faire un choix situé entre donc 15 et 25 classes différentes. Ce choix a été fait comme pour les travaux de Fereday et al. (2008) et de Neal et al. (2016) de façon arbitraire en se basant sur la représentation synthétique des classes (section 10.3). Les connaissances dans le domaine d'étude jouent ici un rôle clé pour arbitrer entre les classifications dont les classes semblent plus ou moins « naturelles ». Cette analyse a fait porter **le choix sur un nombre de classes égal à 20** [4]. Cette classification retenue sera celle utilisée dans le chapitre suivant mais pourra et devra être largement débattue avec des prévisionnistes pour s'assurer que ce découpage correspond bien à leur attente. Chaque classe est représentée en annexe (figures E.3 à E.22). Un nom est ajouté au dessus de chaque classe pour désigner de façon synthétique les zones couvertes par la moyenne du champ. Ces noms seront utilisés pour désigner ces classes (ou structures météorologiques récurrentes) dans la synthèse par scénarios dans le chapitre suivant. Ces 20 classes représentent des positions de pluies différentes mais pas des intensités différentes sur une même position. Des classes représentant la même zone mais avec des intensités différentes apparaissent lorsque le nombre de

4. Ce choix est fait en fonction des besoins des prévisionnistes. Ce nombre pourrait être différent si le domaine d'étude était différent.

classes est plus grand (non montré).

Ce chapitre a permis de faire une classification sur une base de données de 4 ans de membres PE-AROME afin d'établir des structures météorologiques récurrentes qui sont un élément essentiel dans la synthèse par scénarios. Ce chapitre a en particulier montré comment la méthode de classification a été choisie puis comment l'espace latent à 40 dimensions a été sélectionné ainsi que le nombre de classes, fixé à 20 au total. Maintenant que les structures météorologiques sont définies, la synthèse par scénarios peut être utilisée en mode prédiction sur les 32 membres de deux PE-AROME successives. Le dernier chapitre sera consacré à des études de cas et à des scores pour évaluer l'apport de cette nouvelle synthèse afin d'utiliser plus efficacement la PE-AROME.

Bibliographie

M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander. OPTICS : Ordering Points To Identify the Clustering Structure. pages 49–60. ACM Press, 1999.

D. Arthur and S. Vassilvitskii. K-Means++ : The Advantages of Careful Seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, page 1027–1035, 2007. ISBN 9780898716245.

T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics*, 3(1) :1–27, 1974. doi : 10.1080/03610927408827101.

R. J. G. B. Campello, D. Moulavi, and J. Sander. Density-Based Clustering Based on Hierarchical Density Estimates. In *Advances in Knowledge Discovery and Data Mining*, pages 160–172, 2013.

D. L. Davies and D. W. Bouldin. A Cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2) :224–227, 1979. doi : 10.1109/TPAMI.1979.4766909.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society : Series B (Methodological)*, 39(1) :1–22, 1977.

M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.

D. R. Fereday, J. R. Knight, A. Scaife, C. Folland, and A. Philipp. Cluster Analysis of North Atlantic–European Circulation Types and Links with Tropical Pacific Sea Surface Temperatures. *Journal of Climate*, 21 :3687–3703, 2008. doi : 10.1175/2007JCLI1875.1.

E. W. Forgy. Cluster analysis of multivariate data : efficiency versus interpretability of classifications. *Biometrics*, 21 :768–769, 1965.

C. Hennig. Clustering strategy and method selection. 2015a. doi : 10.48550/ARXIV.1503.02059.

C. Hennig. What are the true clusters ? *Pattern Recognition Letters*, 64 :53–62, 2015b. ISSN 0167-8655. doi : https://doi.org/10.1016/j.patrec.2015.04.009.

C. Hennig. *Cluster Validation by Measurement of Clustering Characteristics Relevant to the User*, chapter 1, pages 1–24. 2019. ISBN 9781119597568. doi : https://doi.org/10.1002/9781119597568.ch1.

S. C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3) :241–254, 1967.

T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1) :59–69, 1982.

S. P. Lloyd. Least squares quantization in PCM. *IEEE Trans. Inf. Theory*, 28 : 129–136, 1982.

I. Minitab. Minitab, 2020. URL `https://support.minitab.com/fr-fr/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/cluster-observations/interpret-the-results/all-statistics-and-graphs/dendrogram/`.

R. Neal, D. Fereday, R. Crocker, and R. E. Comer. A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteor. Appl.*, 23(3) :389–400, 2016. doi : https://doi.org/10.1002/met.1563.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn : Machine Learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011.

S. Romano, N. X. Vinh, J. Bailey, and K. Verspoor. Adjusting for chance clustering comparison measures. *The Journal of Machine Learning Research*, 17(1) :4635–4666, 2016.

P. J. Rousseeuw. Silhouettes : a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20 :53–65, 1987. ISSN 0377-0427. doi : https://doi.org/10.1016/0377-0427(87)90125-7.

N. X. Vinh, J. Epps, and J. Bailey. Information theoretic measures for clusterings comparison : Variants, properties, normalization and correction for chance. *The Journal of Machine Learning Research*, 11 :2837–2854, 2010.

U. von Luxburg, R. C. Williamson, and I. Guyon. Clustering : Science or Art ? In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, volume 27, pages 65–79, 2012.

Chapitre 11

Synthèse par scénarios : études de cas et scores

Le chapitre précédent a permis de définir les structures météorologiques qui serviront de base pour classer et projeter les membres de la PE-AROME. A l'aide de l'autoencodeur pour exprimer les membres dans l'espace latent et de ces structures météorologiques, tous les éléments sont réunis pour réaliser la synthèse par scénarios. Pour rappel, chaque membre de la PE-AROME va être exprimé dans l'espace latent à 40 dimensions. Ce vecteur sera ensuite assigné à l'une des 20 classes définies dans le chapitre précédent. La classe choisie est celle dont le centroïde, qui s'exprime également dans ce même espace à 40 dimensions, est le plus proche au sens de la distance euclidienne. Chaque membre PE-AROME est alors, pour chaque échéance, assigné à la classe « Rien signif' » ou encore « Centre Médit' » par exemple (noms donnés en annexe aux 20 classes, figures E.3 à E.22). Ce chapitre va d'une part, détailler comment cette synthèse est représentée graphiquement et d'autre part, évaluer objectivement la pertinence de l'information qu'elle permet d'extraire. Par ailleurs, une des forces de cette synthèse par scénarios est la possibilité de l'étendre au-delà des membres de l'ensemble. Il suffit en effet de disposer d'un champ de RR1h sur la même grille de 256$\times$ 384 points de grille pour pouvoir l'exprimer dans l'espace latent puis l'assigner à une classe. Cette synthèse par scénarios s'adapte en particulier au modèle AROME déterministe, qui pourra donc être directement représenté au côté des membres PE-AROME dans une synthèse commune. L'observation ANTILOPE peut aussi être projetée dans l'espace latent en interpolant au préalable la grille d'1km sur la grille des modèles à 2.5 km, ANTILOPE étant pour rappel un produit de fusion entre les données radar et des pluviomètres pour faire l'estimation la plus juste possible des cumuls de pluie observés (section 1.1.1, figure 1.3).ANTILOPE sert alors de référence et permet d'envisager de calculer des scores pour évaluer les performances des modèles avec cette vision par scénarios. L'utilisation d'ANTILOPE et d'AROME déterministe est techniquement possible mais on ne peut pas garantir que les performances de l'autoencodeur soit comparable. Pour cela, les scores du chapitre 6 (section 6.3.2) sont calculés sur la même période (section 6.1.3) avec ANTILOPE, AROME déterministe et PE-AROME sur la zone Sud-Est. Les scores montrent des erreurs de position similaires entre PE-AROME, AROME et ANTILOPE. Pour des raisons que nous n'avons pas encore eu le temps d'explorer, ANTILOPE a des TD

et TFA légèrement meilleurs que ceux de PE-AROME et AROME. Ces scores ne remettent en tout cas pas en cause l'utilisation de AROME et d'ANTILOPE dans l'approche par scénarios. Ce chapitre commencera par une première partie sur la représentation graphique de la synthèse par scénarios. Des études de cas seront présentées dans une deuxième partie pour montrer comment la synthèse par scénarios peut aider à extraire l'information présente dans les différents modèles disponibles. Enfin, la dernière partie de ce chapitre présentera certains scores objectifs qui donneront quelques éléments pour utiliser au mieux la synthèse par scénarios et permettront une comparaison entre les modèles PE-AROME et AROME opérationnels en 2021 et la nouvelle version de ces modèles utilisée en opérationnel depuis l'été 2022.

11.1 Représentation de la synthèse par scénarios

11.1.1 Synthèse globale

La représentation des scénarios est faite par l'intermédiaire d'un premier graphique qui montre la répartition des membres parmi les 20 classes selon les échéances. La figure 11.1 est un exemple de ce mode de représentation. Il n'y a cependant que 16 classes représentées en ordonnée. En effet, les deux classes sous la dénomination « Espagne (1) » et « Espagne (2) » sont regroupées, tout comme les quatre en Italie sous le nom « Italie ». Ces regroupements ont été faits car ces distinctions n'ont a priori que peu d'intérêt pour la France métropolitaine et permettent de simplifier un peu ce graphique. Cette synthèse regroupe en un seul et même graphique les prévisions issues de deux réseaux PE-AROME et deux réseaux AROME :

- PE-AROME, réseau de n h UTC
- AROME, réseau de $n + 3$ h UTC
- PE-AROME, réseau de $n + 6$ h UTC
- AROME, réseau de $n + 9$ h UTC

La synthèse regroupe donc pour chaque échéance 34 prévisions différentes, les prévisions AROME étant représentées par les cercles rose et vert. La représentation par l'intermédiaire de points bleus et de ces cercles vert et rose montre une cohérence temporelle, bien que la classification soit faite échéance par échéance et de façon indépendante. Cette cohérence est assurée par les membres PE-AROME (et AROME) eux-mêmes qui sont cohérents physiquement et dans le temps, ce que la synthèse par scénarios semble bien retranscrire avec une évolution entre les classes qui n'est pas erratique. Ce graphique permet de conclure rapidement que les 34 prévisions n'envisagent pas de pluie significative entre le 13 septembre 02h UTC et le 13 septembre 15h UTC. Ce scénario sans pluie significative reste majoritaire jusqu'à la dernière heure couverte par ce graphique, le 14 septembre à 12h UTC. Cependant, quelques membres de la PE-AROME envisagent des pluies sur les Pyrénées dans la soirée du 13 septembre, puis sur le Massif-Central en début de nuit. La matinée du 14 septembre est marquée par un scénario alternatif sur la zone Cevennes/Gard/Hérault-Nord puis plus à l'Est vers la zone Var/PACA

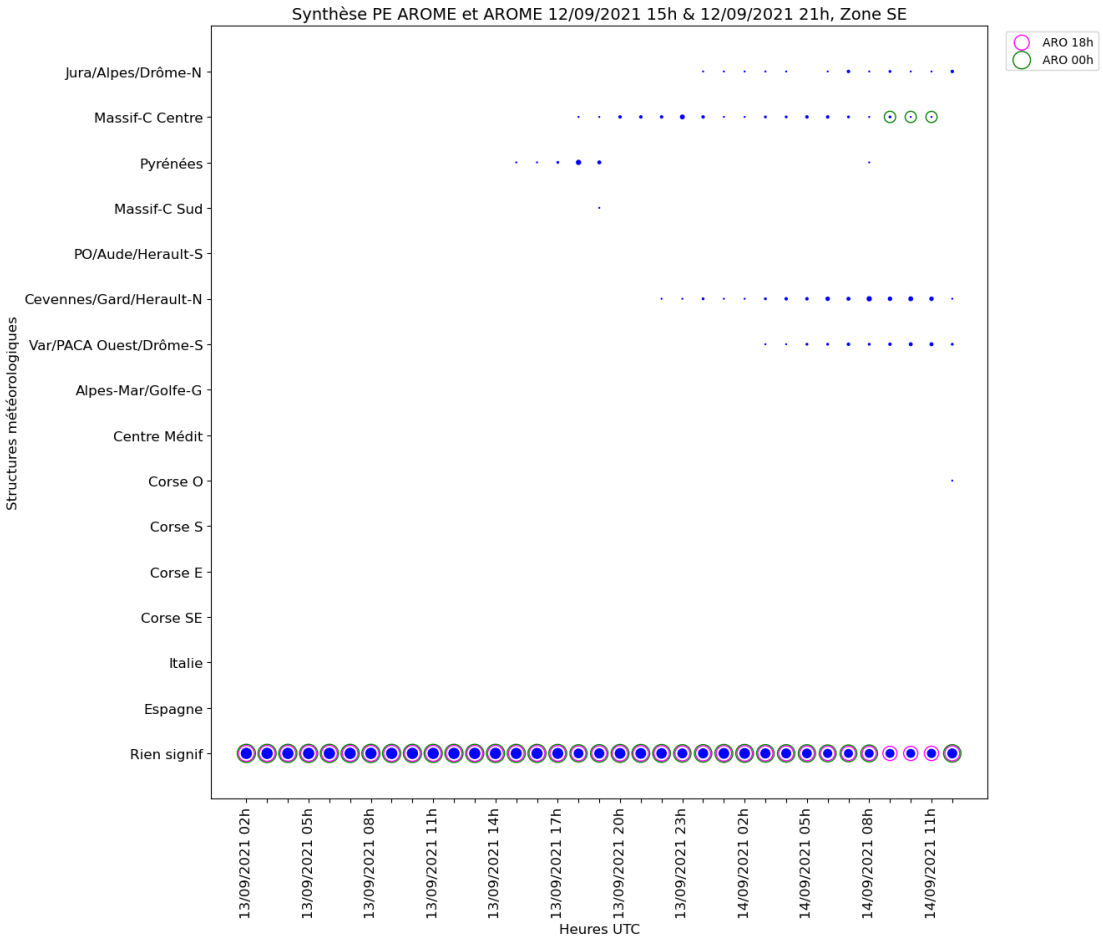


FIGURE 11.1 – *Représentation de la synthèse par scénarios pour la zone Sud-Est sous forme condensée. Les réseaux PE-AROME représentés sont ceux du 12 septembre 2019 à 15h et 21h UTC. Les heures de prévision sont représentées en abscisse, du 13 septembre 02h UTC jusqu'au 14 septembre 12h UTC avec un pas horaire. L'ordonnée représente 16 des 20 classes ou structures météorologiques définies dans le chapitre précédent. Les points bleus représentent la répartition des 32 membres PE-AROME parmi les 16 classes. Plus le point est gros, plus le nombre de membres est important. Cette synthèse représente aussi les prévisions faites par deux réseaux AROME, ceux du 12 septembre à 18h UTC et du 13 septembre à 0h UTC. Ces prévisions sont représentées sous forme d'un cercle rose pour le réseau le plus ancien (18h UTC) et vert pour le plus récent (0h UTC).*

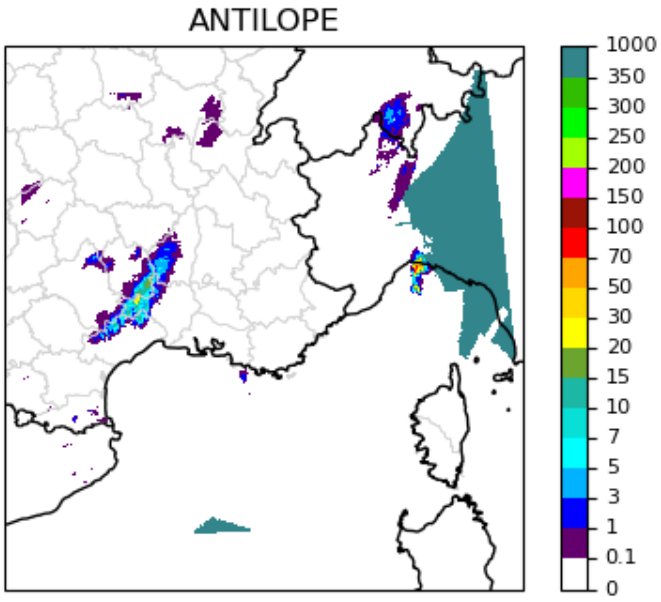


FIGURE 11.2 – *Exemple de carte ANTILOPE sur la zone Sud-Est. Cumul de pluie entre le 3 octobre 2021 à 02h et 03h UTC. Les points de grille dépassant les 350 mm (couleur bleu canard) sur cette carte sont hors domaine pour le produit ANTILOPE. Cette couleur bleu canard correspond donc au masque ANTILOPE (Italie et extrême sud du domaine en Méditerranée).*

Ouest/Drôme-Sud. Quelques membres et la prévision AROME du 13 septembre à 0h envisagent des pluies sur le Massif-Central en fin de matinée. Ce graphique permet de donner un aperçu général sur la dispersion de la PE-AROME, comment AROME se positionne dans l'ensemble PE-AROME avec déjà plusieurs informations intéressantes à retirer comme le montreront les études de cas dans la section 11.2. Les variations de ce graphique à travers les réactualisations des prévisions seront aussi intéressantes à regarder. Dans notre cas présent, on peut se demander si les scénarios alternatifs sur les Pyrénées ou Cevennes/Gard/Herault-Nord vont prendre de l'ampleur ou disparaître avec la prévision de la PE-AROME du 13 septembre à 03h UTC. Il faut cependant souligner que ce graphique reste trop simple pour une analyse fine de la situation. Il est impossible en particulier d'évaluer la dispersion des membres dans une classe. De plus, ces classes représentent une information sur la position des pluies mais pas sur leur intensité. C'est pourquoi des graphiques supplémentaires pour représenter les scénarios sont présentés dans la section suivante 11.1.2.

Ce graphique représente donc une production possible en condition réelle. Mais comme évoqué dans l'introduction, il est également envisageable de représenter l'observation ANTILOPE dans cette synthèse par scénarios en tant que vérification. Il faut cependant prendre quelques précautions car ANTILOPE n'est pas disponible sur tout le domaine Sud-Est (figure 11.2). Pour résoudre ce problème, les zones non couvertes par ANTILOPE sont considérées comme sans pluie (0 mm). Par ailleurs, la qualité d'ANTILOPE se dégrade en absence de pluviomètre ou en s'éloignant des radars. En particulier, les données sont de moins bonne qualité en mer ou en allant vers les pays voisins. Il faudra donc prêter attention lors des études de cas, si des pluies se situent dans ces zones il est possible qu'AN-

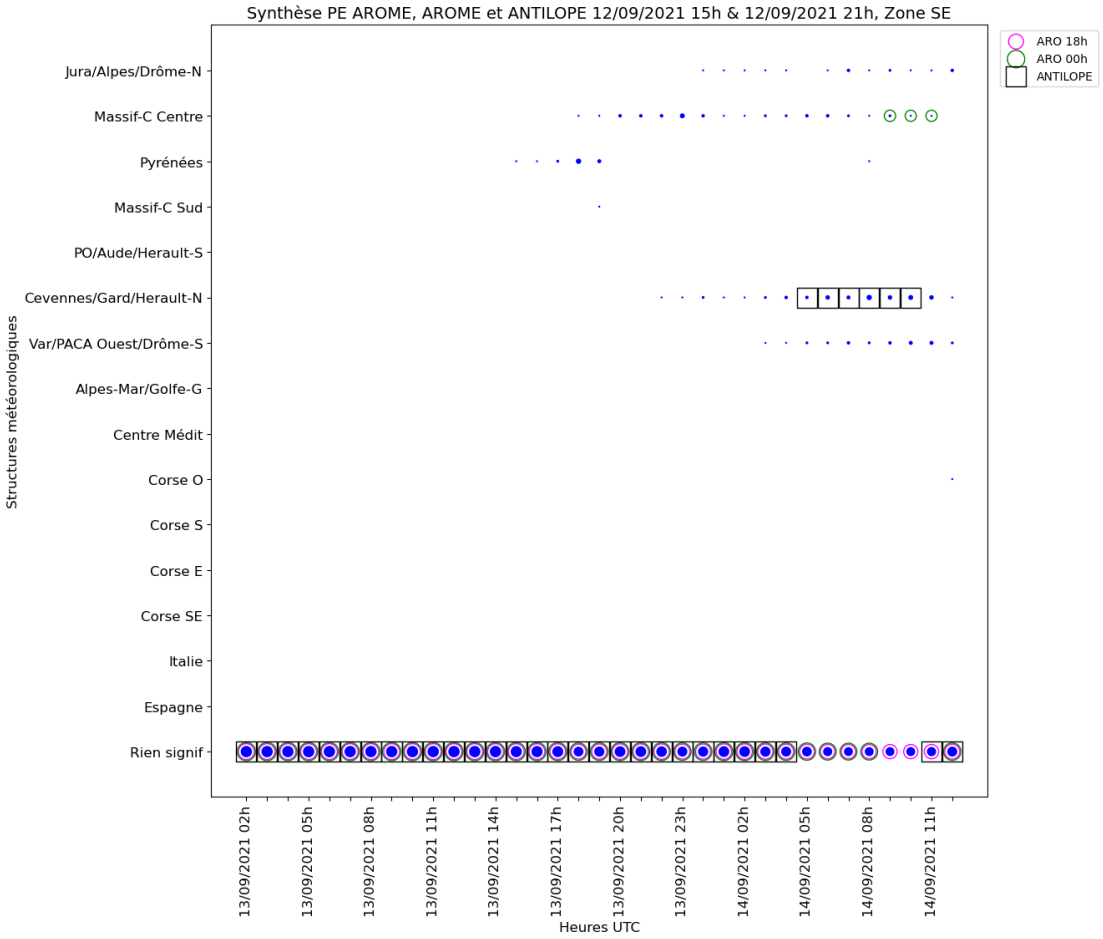


FIGURE 11.3 – *Représentation de la synthèse par scénarios sous forme condensée en ajoutant la vérification ANTILOPE. La représentation est similaire à celle de la figure 11.1 en ajoutant la projection d'ANTILOPE sur les structures météorologiques à l'aide d'un carré noir.*

TILOPE ne soit pas une bonne référence. Pour les scores, les solutions retenues pour surmonter ces problèmes de qualité médiocre ou d'absence de données seront présentées dans la partie dédiée 11.3. Après avoir remplacé les points hors du domaine par des zéros, ANTILOPE est assigné à une des 20 structures météorologiques, il est alors possible de représenter sur le même graphique ANTILOPE et les modèles (figure 11.3). On remarque qu'ANTILOPE est classé sans pluie significative la plupart du temps sauf le 14 septembre entre 5h et 10h UTC où ANTILOPE est assigné à la classe Cevennes/Gard/Herault-N, ce qui donne raison au scénario minoritaire sur cette période, mais pas au scénario minoritaire sur les Pyrénées.ANTILOPE n'étant pas disponible en temps réel, ce graphique ne peut être utilisé que dans un objectif de vérification et d'évaluation à posteriori.

11.1.2 Représentation des scénarios

La manière la plus simple de visualiser un scénario correspondant à un point sur les figures 11.1 et 11.3 est de tracer les cartes de RR1h des membres assignés à ce scénario. La figure 11.4 représente les 3 scénarios proposés par les 34 membres et

ANTILOPE pour le 13 septembre 2021 à 18h UTC en lien avec le graphique de la figure 11.3. Chaque champ de pluie est désigné soit par le réseau de la PE-AROME correspondant et le numéro de membre soit pour AROME par la dénomination « ARO » et le réseau correspondant soit par le terme « ANTILOPE » pour l'observation. La figure 11.4 montre que même dans un scénario, il reste de la dispersion. On remarque par exemple que la plupart des membres dans le scénario « Rien de significatif » ne propose pas de pluie mais quelques uns en proposent entre les Pyrénées et le Massif-Central. Ces membres sont malgré tout dans ce scénario car les pluies sont faibles. Les pluies pour les 7 membres dans le scénario « Pyrénées » sont plus fortes. Ces membres se ressemblent beaucoup même si le membre 3 du réseau de 21h propose aussi une zone de pluie sur le Massif-Central. Il n'est cependant pas classé dans le scénario « Massif-Central Centre » car la zone de pluie la plus active est bien sur les Pyrénées alors que c'est l'inverse pour le membre 10 du réseau de 15h qui est isolé dans son scénario sur le Massif-Central. **Cet exemple montre qu'il est nécessaire de regarder au moins rapidement quelques membres dans les scénarios et de ne pas se contenter de la synthèse globale, au risque de tirer des conclusions hâtives**. Les 20 structures météorologiques permettent de donner quelques informations sur les membres dans leur groupe mais elles restent insuffisantes pour faire une analyse détaillée. Les études de cas le montreront également dans la partie suivante. La représentation des scénarios sous cette forme est cependant fastidieuse, d'autres travaux en complément de ce manuscrit devront suivre pour permettre de rendre compte de la dispersion dans un scénario sans pour autant recourir à une visualisation de tous les membres.

11.2 Études de cas

Les différents graphiques et cartes présentés dans la partie précédente vont être utilisés sur trois études de cas pour montrer les informations que la synthèse par scénarios permet d'extraire des modèles. Les évolutions après chaque nouveau réseau PE-AROME disponible seront montrées pour voir comment cette synthèse évolue dans le temps en se rapprochant de l'évènement. Contrairement à la figure 11.4 qui présentait une situation classique sans gravité, les trois études de cas correspondent à des situations ayant donné lieu à des pluies extrêmes sur la zone Sud-Est mais avec un comportement différent des modèles à chaque fois. La première étude de cas sera dédiée aux fortes pluies très localisées entre Montpellier et Nîmes le 14 septembre 2021, la deuxième aux inondations dans les Alpes-Maritimes et en Italie le 2 octobre 2020 et la dernière étude de cas aux pluies intenses sur les Cévènnes et la région Provence-Alpes-Côte d'Azur (PACA) les 3 et 4 octobre 2021.

11.2.1 14 septembre 2021

Situation météorologique

La situation météorologique est résumée avec la figure 11.5 à l'aide tout d'abord de l'ANAlyses SYnoptiques Graphiques (ANASYG) produit par le Centre National

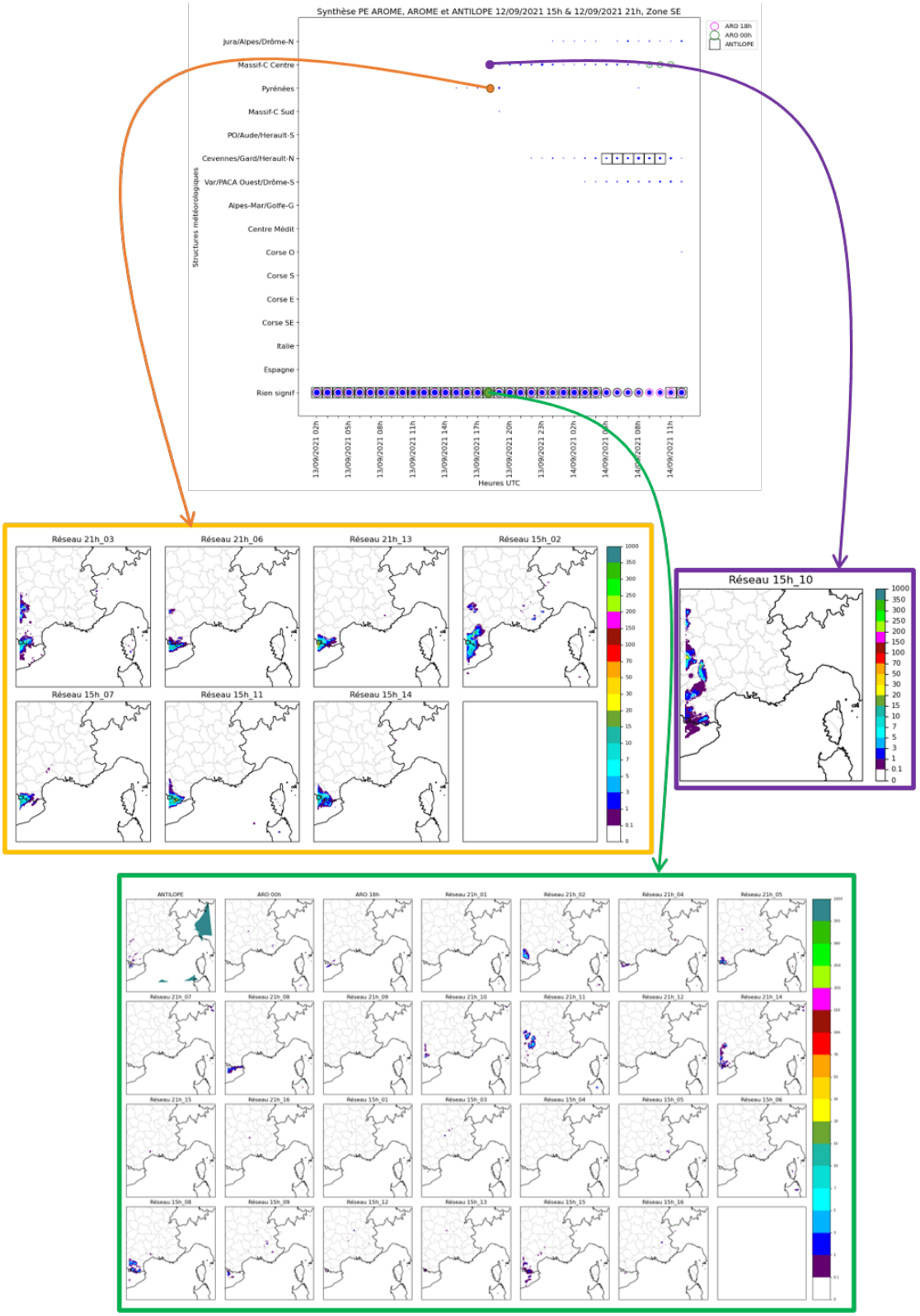


FIGURE 11.4 – *Représentation des scénarios « Rien de significatif » (encadré vert), « Pyrénées » (encadré orange) et « Massif-Central Centre » (encadré violet) de la figure 11.3 pour le 13 septembre 2021 à 18h UTC.*

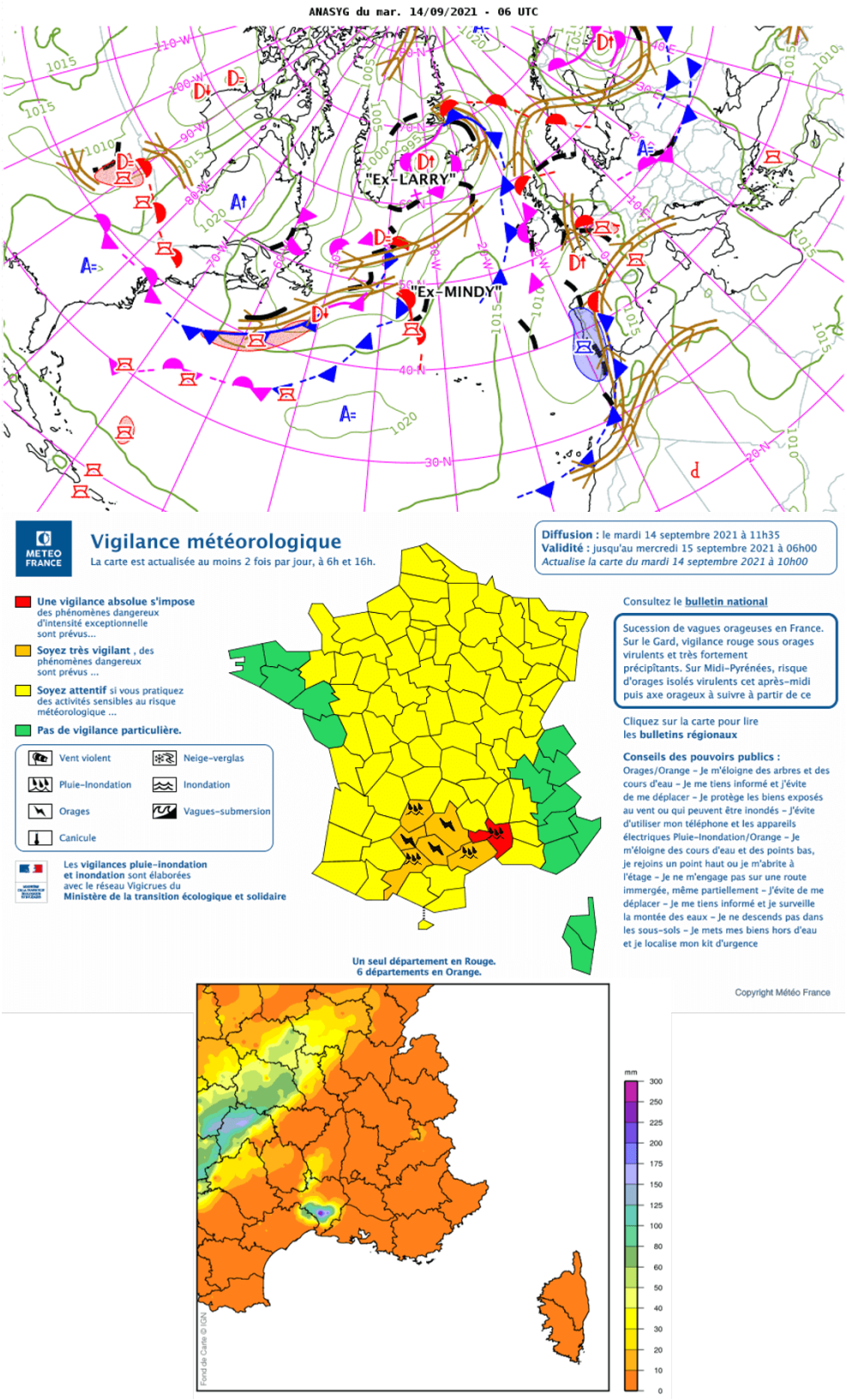


FIGURE 11.5 – *Situation météorologique du 14 septembre 2021. La carte du haut représente l'ANASYG produit par le CNP pour le 14 septembre 2021 à 06h UTC. La carte au centre représente la carte de vigilance émise par Météo-France le 14 septembre 2021 en fin de matinée et la carte du bas le cumul de pluie observé sur la journée du 14 septembre 2021 (Source : Okapi, Météo-France).*

de Prévision (CNP) concernant la situation du 14 septembre 2021 à 06h UTC. La légende correspondant aux différents objets météorologiques représentés sur cette carte est disponible en annexe (figure A.1). L'ANASYG montre la circulation d'un jet à proximité des côtes méditerranéennes française mais sans autre élément synoptique notable. La carte fait également état d'une activité convective dans le Golfe du Lion, représentant les orages qui vont engendrer de forts cumuls dans cette zone. En conséquence, Météo-France avait placé les départements de l'Herault et du Gard en vigilance orange pour orages et pluie-inondation à 06h UTC avec un bulletin annonçant des cumuls dépassant localement les 100 mm. Mais comme le montre la carte des cumuls sur la figure 11.5, les cumuls vont dépasser les 250 mm [1] ce qui provoquera le placement du Gard en vigilance rouge en fin de matinée. Le phénomène responsable de ces cumuls très importants et localisés est un système convectif quasi-stationnaire dont le cycle de vie est présenté dans les papiers de Chappell (1986) et Bresson (2011). Les autres départements en alerte sur la carte de vigilance concernent d'autres orages dans l'après-midi.

Prévisions PE-AROME et AROME selon les réseaux

Les différentes synthèses par scénarios en partant des réseaux PE-AROME du 12 septembre à 15h et 21h UTC (correspondant à la figure 11.1) jusqu'aux réseaux du 13 septembre à 21h et du 14 septembre à 03h sont représentées sur la figure 11.6. Les graphiques se lisent de gauche à droite et de haut en bas. Ces différents graphiques montrent que le scénario majoritaire est celui correspondant à la classe « Rien de significatif » pour les quatre premiers graphiques (Réseaux PE-AROME du 12 septembre à 21h jusqu'au 13 septembre à 15h UTC), avec un scénario alternatif sur la zone « Cevennes/Gard/Herault-Nord » dont la taille reste à peu près stable. Le scénario alternatif pour le deuxième graphique (en haut à droite de la figure 11.6) correspondant à la prévision pour le 14 septembre à 09h UTC est représenté figure 11.7. ANTILOPE est également affiché pour servir de référence. On remarque que dans ce scénario alternatif qui compte 6 membres, plusieurs proposent des pluies intenses dépassant localement les 70 mm en 1h : les membres 5 et 16 pour le réseau de 21h et les membres 3 et 7 pour le réseau de 3h. Les deux autres membres dans ce scénario correspondent à des pluies plus faibles. On peut donc en conclure que quelques membres de la PE-AROME avaient anticipé un risque de cumuls localement importants plus de 24h avant l'évènement. On remarque également qu'aucun AROME n'avait anticipé ce risque, les réseaux restant tous dans la classe « Rien de significatif ». Un changement important se produit avec l'arrivée du réseau du 13 septembre à 21h UTC, le scénario alternatif devient de taille identique au scénario majoritaire. A cet instant, deux scénarios de taille comparable existent alors pour la matinée du 14 septembre. Les réseaux AROME proposent aussi ce changement. Après un réseau AROME du 13 septembre à 18h qui avait déjà changé de classe pour quelques heures, le réseau du 14 septembre à 00h bascule clairement dans ce qui était le scénario alternatif. L'ancien scénario alternatif devient majoritaire quelques heures avant l'évènement sur le dernier graphique, constitué de 21 membres et des 2 réseaux d'AROME pour la même heure de prévision que la figure 11.7, c'est-à-dire le 14 septembre à 09h UTC (figure 11.8). Il faut cependant noter que même si le scénario

1. La station de Saint-Dionizy a relevé 276 mm en 4h, dont 244 mm en 3h.

FIGURE 11.6 – *Évolution de la synthèse par scénarios en partant des réseaux PE-AROME du 12 septembre 2021 à 15h et 21h UTC (figure 11.1), jusqu'aux réseaux du 13 septembre 2021 21h et 14 septembre 2021 03h avec un pas de 6h. L'encadré en rouge représente la période d'intérêt correspondant aux heures où ANTILOPE a été classé dans la zone Cevennes/Gard/Herault-Nord (cf figure 11.3).*

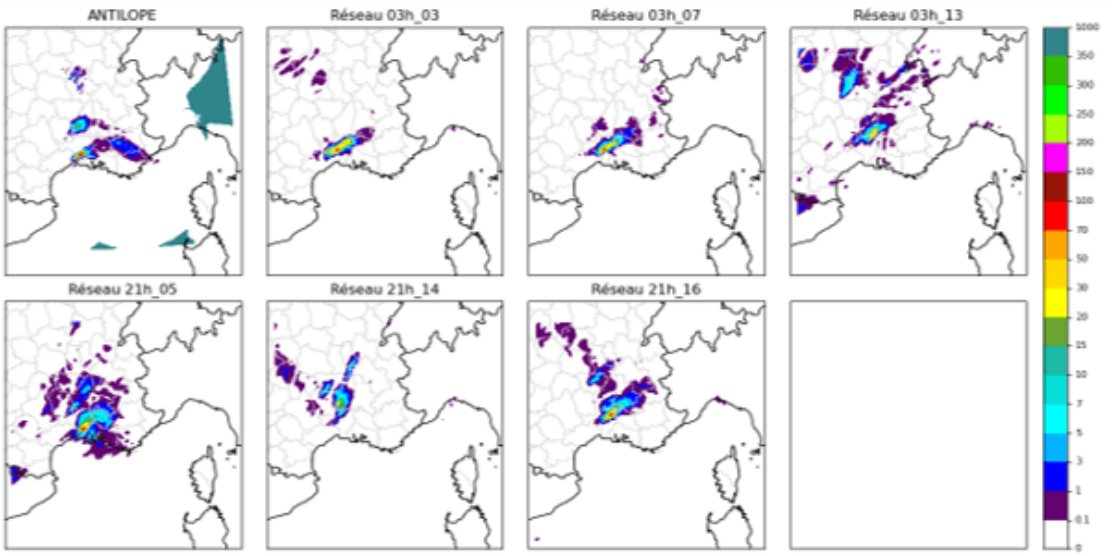


FIGURE 11.7 – *Scénario Cevennes/Gard/Herault-Nord pour les réseaux du 12 septembre 2021 à 21h UTC et du 13 septembre 2021 à 03h UTC, prévision pour le 14 septembre 2021 à 09h UTC.*

« Cevennes/Gard/Herault-Nord » est devenu majoritaire, beaucoup de membres proposent toujours des intensités plus faibles que celles observées.

Cette étude de cas a montré que la PE-AROME a su anticiper avec quelques membres un risque de fortes pluies plus de 24h avant l'évènement se matérialisant par un scénario alternatif qui a fini par devenir majoritaire quelques heures avant l'évènement. On notera aussi que les réseaux AROME ont eux subitement changé de prévision à partir du réseau du 13 septembre à 18h UTC. Concernant la suite de la journée du 14 septembre, on peut noter que des pluies sont envisagées sur le Massif-Central (figure 11.6). On retrouve bien cette activité pluvieuse sur les cumuls observés de la figure 11.5, qui correspond à une remontée d'orages depuis le Sud-Ouest du pays.

11.2.2 2 octobre 2020

Situation météorologique

La situation météorologique du 2 octobre 2020 est bien différente de celle du 14 septembre 2021 (figure 11.9). L'ANASYG montre un dynamisme d'altitude bien plus marqué, piloté par la tempête Alex centrée sur la Bretagne à 12h UTC. Le Sud-Est de la France est soumis à l'avant à un rapide flux de sud qui va engendrer des pluies orageuses intenses en provenance de la Méditerranée. De l'activité orageuse est d'ailleurs matérialisée dans le Golfe de Gênes. Cette situation a poussé les services de Météo-France à émettre une vigilance rouge pluie-innondation sur les Alpes-Maritimes dès 6h du matin où on attend localement 250 à 300 mm sur la journée. Le bulletin national annonce localement entre 100 et 130 mm sur les départements voisins en vigilance orange. Les cumuls observés vont malheureusement dépasser ces estimations avec 500 mm relevé à Saint-Matin-de-Vésubie dans l'arrière pays des Alpes-Maritimes. Des cumuls exceptionnels sont aussi relevés dans la région du Piémont en Italie avec 630 mm en 24h à Sambughetto.

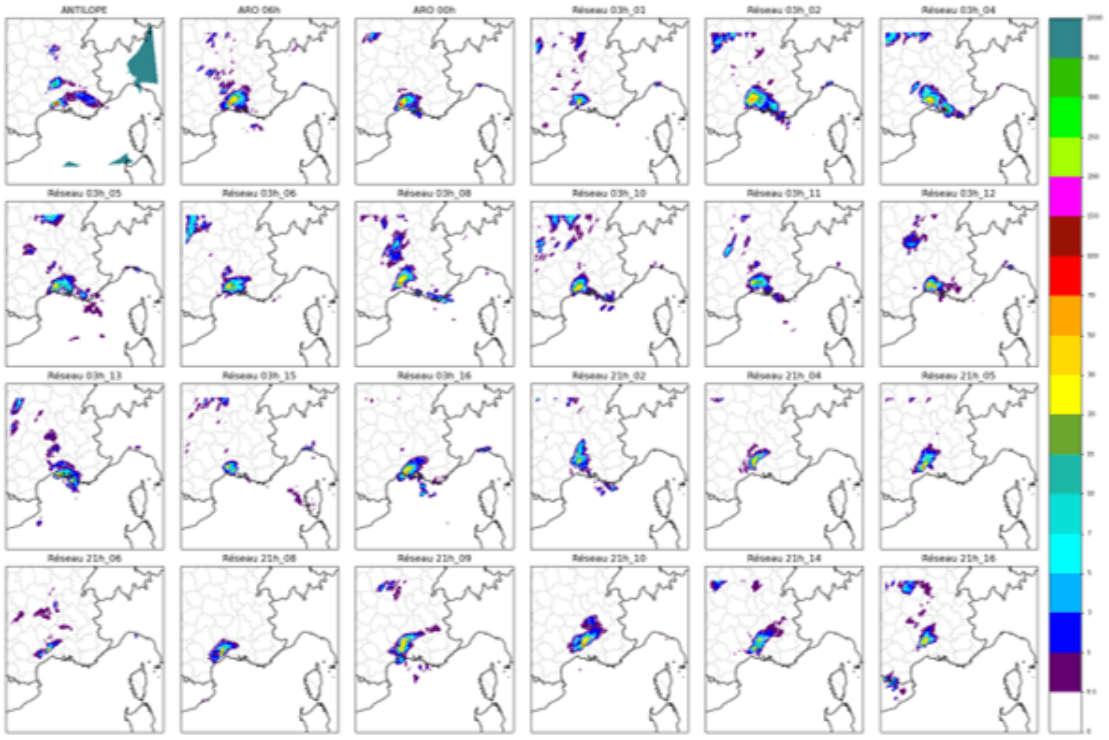


FIGURE 11.8 – *Scénario Cevennes/Gard/Herault-Nord pour les réseaux du 13 septembre 2021 à 21h UTC et du 14 septembre 2021 à 03h UTC, prévision pour le 14 septembre 2021 à 09h UTC.*

Prévisions PE-AROME et AROME selon les réseaux

Dans cette autre situation météorologique, la synthèse par scénario est bien différente de la dernière étude de cas (figure 11.10). Le scénario « Rien de significatif » n'est plus proposé par aucun membre durant la deuxième partie de nuit du 1er au 2 octobre jusqu'au début de journée le 3 octobre (dernier graphique en bas à gauche). On observe sur cette période principalement deux scénarios : « Alpes-Maritimes/Golfe de Gênes » et « Italie » (encadrés rouges). On peut se demander ce qui diffère entre ces deux scénarios. Ils sont représentés pour la synthèse des PE-AROME du 30 septembre à 21h UTC et du 1er octobre à 03h UTC, prévision faite pour le 2 octobre à 15h UTC (figures 11.11 et 11.12). On peut remarquer en premier lieu qu'il pleut sur une bonne partie de la zone Sud-Est sur l'ensemble des membres. La différence réside dans la zone la plus active des pluies. Les membres dans le scénario « Alpes-Maritimes/Golfe de Gênes » proposent les plus fortes pluies sur cette zone alors que pour les autres membres, elles sont plutôt sur le Piémont en Italie. La différence des deux scénarios s'explique donc par la position des pluies les plus fortes. Comme des cumuls très importants sont bien observés sur ces deux zones simultanément, cette répartition des membres est plutôt cohérente. On remarque encore ici pourquoi il est important d'aller voir les membres dans les scénarios et ne pas se contenter des cartes de synthèse. La synthèse par scénarios ne change que très peu selon les réseaux car à l'aide d'un dynamisme à l'échelle synoptique bien plus important, la prévisibilité est probablement bien meilleure que pour la situation précédente. Enfin, on remarque que ANTILOPE est beaucoup plus souvent assigné à la classe Alpes-Maritimes/Golfe de Gênes,

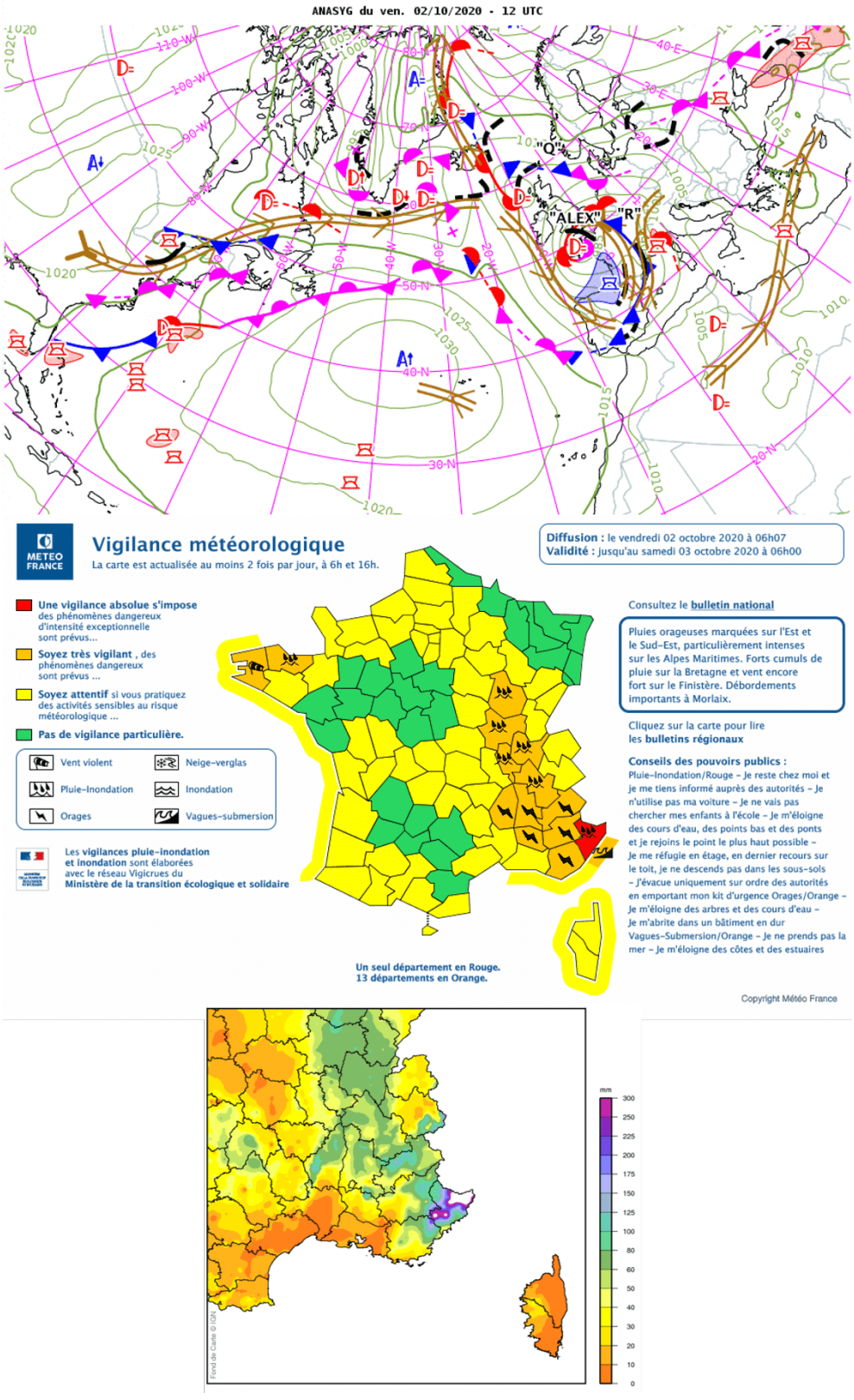


FIGURE 11.9 – *Situation météorologique du 2 octobre 2020. La carte du haut représente l'ANASYG produit par le CNP pour le 2 octobre 2020 à 12h UTC. La carte au centre représente la carte de vigilance émise par Météo-France le 2 octobre 2020 à 6h du matin et la carte du bas le cumul de pluie observé sur la journée du 2 octobre 2020 (Source : Okapi, Météo-France).*

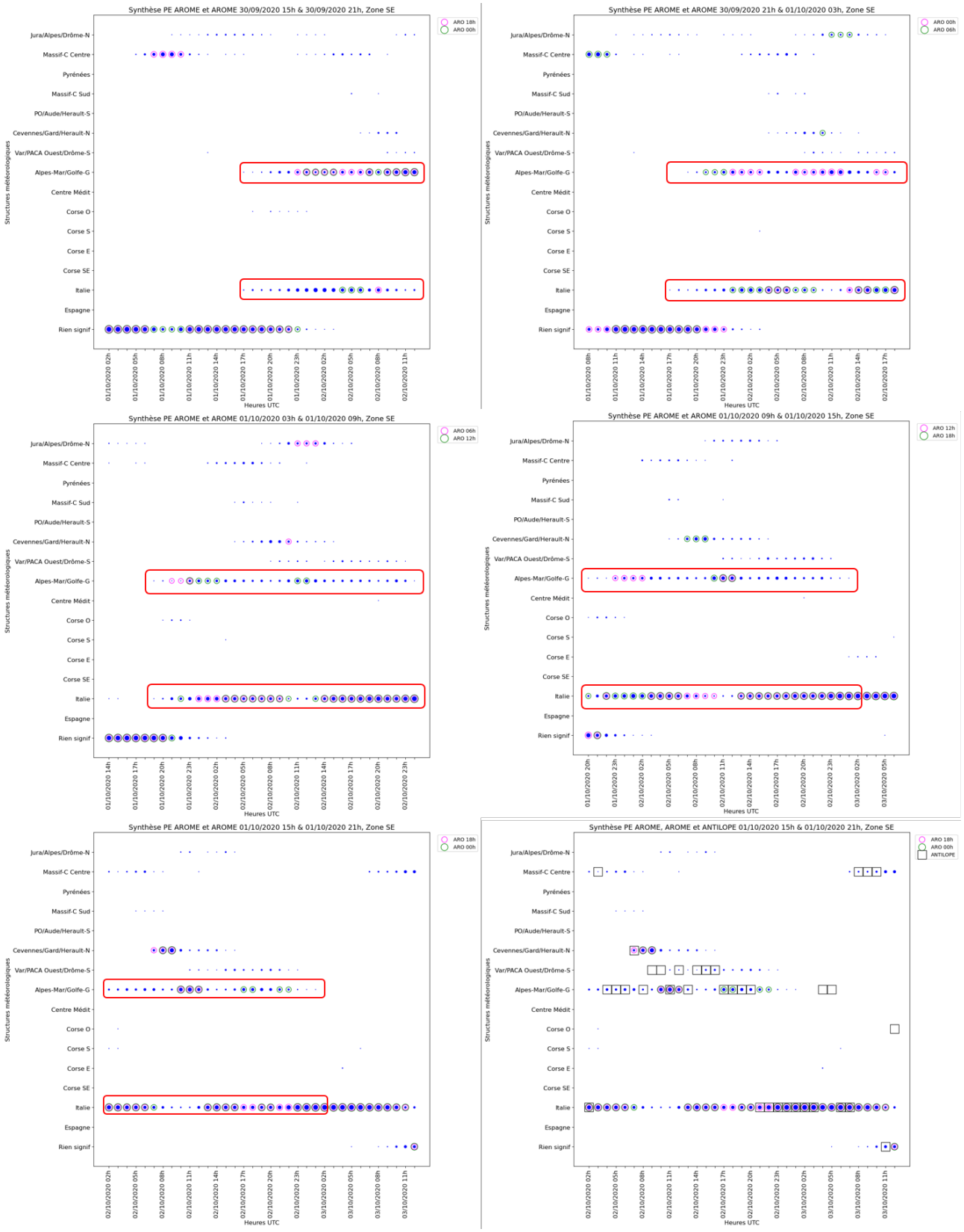


FIGURE 11.10 – *Évolution de la synthèse par scénarios en partant des réseaux PE-AROME du 30 septembre 2020 à 15h et 21h UTC jusqu'aux réseaux du 1er octobre 2020 à 15h et 21h UTC avec un pas de 6h. Le dernier graphique en bas à droite représente de nouveau la synthèse par scénarios pour les réseaux du 1er octobre 2020 à 15h et 21h UTC, mais avec ANTILOPE.*

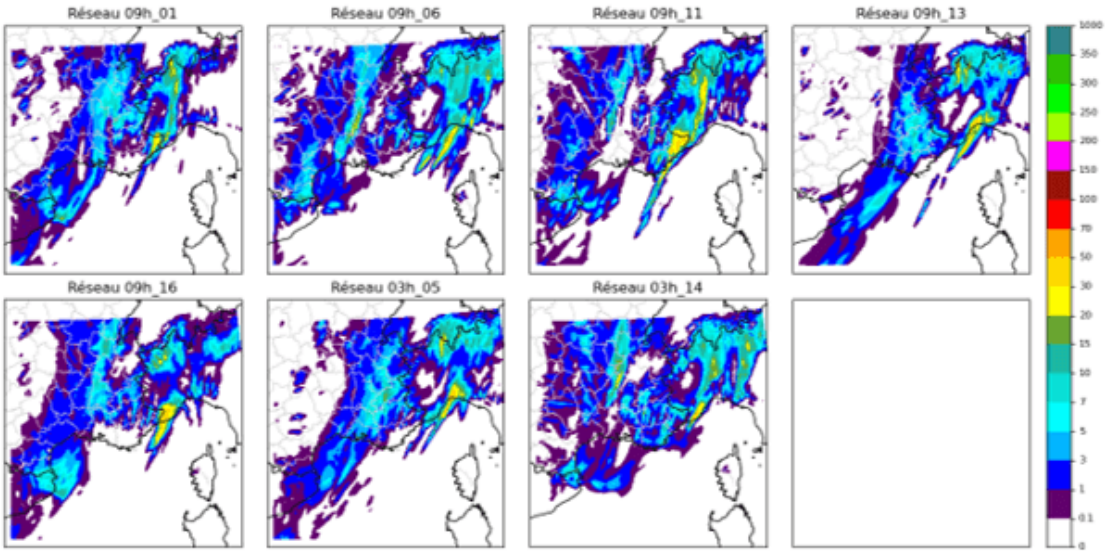


FIGURE 11.11 – *Scénario Alpes-Maritimes/Golfe de Gênes pour les réseaux du 30 septembre 2020 à 21h UTC et du 01 octobre 2020 à 03h UTC, prévision pour le 2 octobre 2020 à 15h UTC.*

mais il faut rappeler la plus mauvaise qualité d'ANTILOPE sur l'Italie qui peut fausser son assignation dans une classe sur cette situation. On remarque aussi qu'ANTILOPE est par moment assigné à la classe Var/PACA Ouest/Drôme-Sud lorsque les plus fortes pluies se positionnent un peu plus à l'Ouest, à la limite du Var et des Alpes-Maritimes (non montré).

11.2.3 3-4 octobre 2021

Situation météorologique

Les journées des 3 et 4 octobre 2021 sont marquées par des pluies intenses sur une bonne partie du Sud-Est de la France (figure 11.13). Après un épisode cevenol dans la journée du 3 octobre, un axe pluvieux quasi-stationnaire se décale des Cevennes vers les Bouches-du-Rhône dans le nuit du 3 au 4 octobre. Il tombe plus de 120 mm sur ce département et en particulier sur l'agglomération de Marseille. En lien avec le minimum dépressionnaire visible sur l'ANASYG de 12h et en prévision de nouvelles pluies attendues dans la journée du 4 octobre, Météo-France place le département des Bouches-du-Rhône en alerte rouge pluie-inondation ainsi que toute la région PACA et la Haute-Corse en alerte orange pour ces mêmes remontées pluvio-orageuses. De nouveaux cumuls importants tombent durant l'après-midi du 4 octobre sur le Var et la Corse. Sur l'ensemble de l'épisode, on relève 468 mm sur la station de Villefort dans les Cevennes, 173 mm à Marseille, 183 mm dans le Var sur la commune de Régusse et 174 mm à Alistro en Corse.

Prévisions PE-AROME et AROME selon les réseaux

La figure 11.14 illustre l'évolution de la synthèse par scénarios selon les réseaux. On repère tout d'abord facilement l'épisode cevenol (premier cadre rouge)

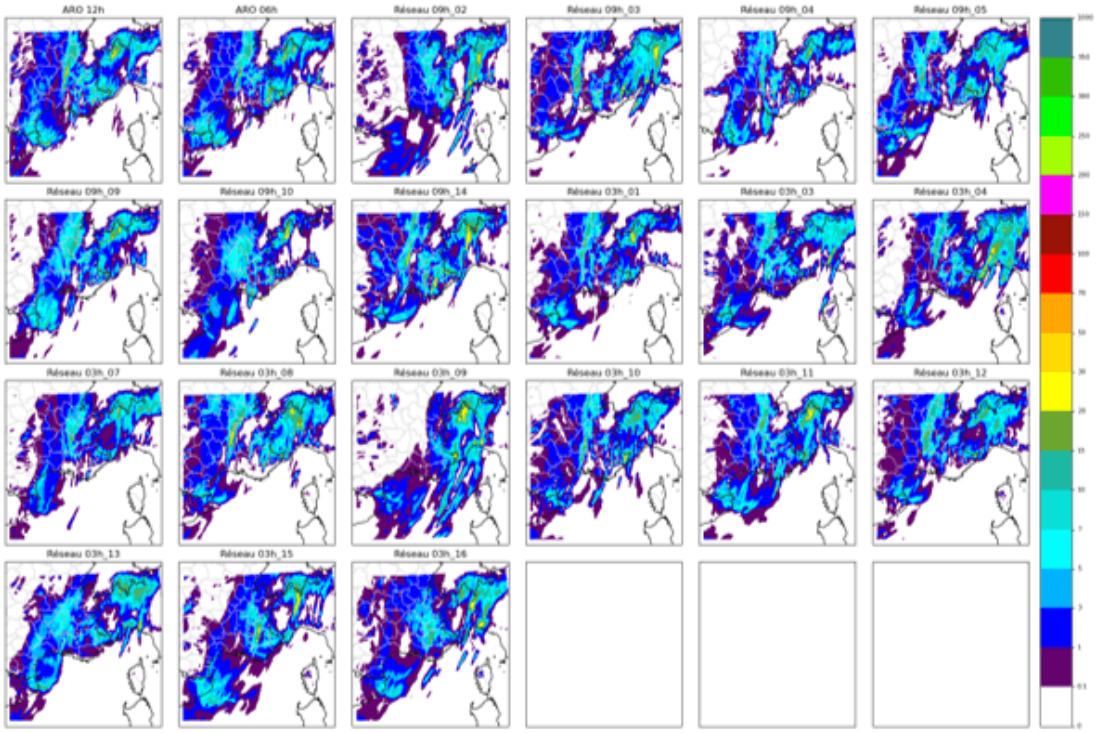


FIGURE 11.12 – *Scénario Italie, mêmes réseaux et même heure de prévision que la figure 11.11.*

avec un grand nombre de membre assigné dans la classe Cevennes/Gard/Herault-Nord pour la journée du 3 octobre. Sur les dernier graphiques, on repère également à partir de la soirée du 4 octobre un scénario majoritaire sur l'Italie (deuxième encadré rouge). Il est en revanche très difficile de trouver un ou même deux scénarios pour la journée du 4 octobre. La dispersion semble très importante dans la PE-AROME. On peut alors se demander si, comme pour l'étude de cas précédente, les différences entre les scénarios s'expliquent par des intensités de pluie plus ou moins fortes. Pour le vérifier, les figures 11.15 et 11.16 représentent le scénario « Espagne » et le scénario « Var/PACA Ouest/Drôme-Sud » pour les réseaux PE-AROME du 3 octobre 15h et 21h (dernier graphique en bas à gauche, figure 11.14). La prévision est faite pour le 4 octobre à 03h, donc à des échéances assez courtes inférieures à +15h. Les prévisions entre les deux scénarios sont ici très différentes. En particulier les membres dans le scénario de la figure 11.15 ne proposent pas ou presque pas de précipitation sur les Bouches-du-Rhône qui est la zone avec les plus fortes pluies sur les membres de l'autre scénario. Ces différences illustrent donc une faible prévisibilité. Les réseaux en se rapprochant de l'évènement n'améliorent pas beaucoup ce point. On peut souligner d'ailleurs que même quelques heures avant les pluies dans la nuit du 3 au 4 octobre sur les Bouches-du-Rhône, la plupart des membres proposent des pluies sur l'Espagne. On note aussi de grandes fluctuations dans le positionnement des réseaux AROME à travers les classes, illustrant là aussi une prévisibilité très limitée. On peut conclure de ces graphiques que la prévisibilité est très bonne pour l'épisode cevenol et le décalage des pluies vers l'Italie dans la soirée du 4 octobre mais qu'elle est très mauvaise pour la journée du 4 octobre. Il est alors particulièrement difficile de choisir un scénario et d'établir une seule et unique prévision dans ces conditions.

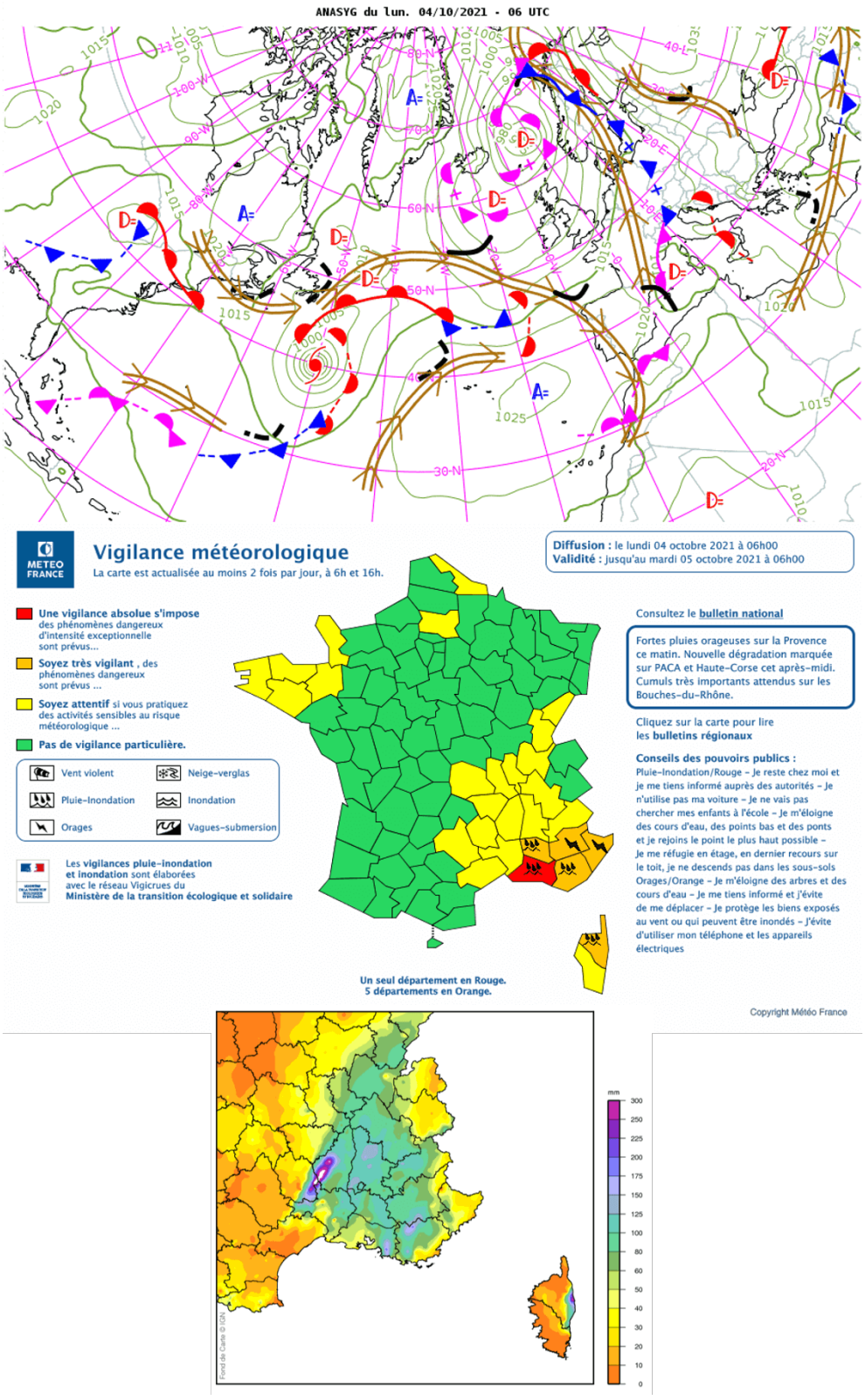


FIGURE 11.13 – *Situation météorologique du 4 octobre 2021. La carte du haut représente l'ANASYG produit par le CNP pour le 4 octobre 2021 à 06h UTC. La carte au centre représente la carte de vigilance émise par Météo-France le 4 octobre 2021 à 6h du matin et la carte du bas le cumul de pluie observé sur les journées du 3 et 4 octobre 2021 (Source : Okapi, Météo-France).*

FIGURE 11.14 – *Évolution de la synthèse par scénarios en partant des réseaux PE-AROME du 2 octobre 2021 à 15h et 21h UTC jusqu'aux réseaux du 3 octobre 2021 à 15h et 21h UTC avec un pas de 6h. Le dernier graphique en bas à droite représente de nouveau la synthèse par scénarios pour les réseaux du 3 octobre 2021 à 15h et 21h UTC, mais avec ANTILOPE.*

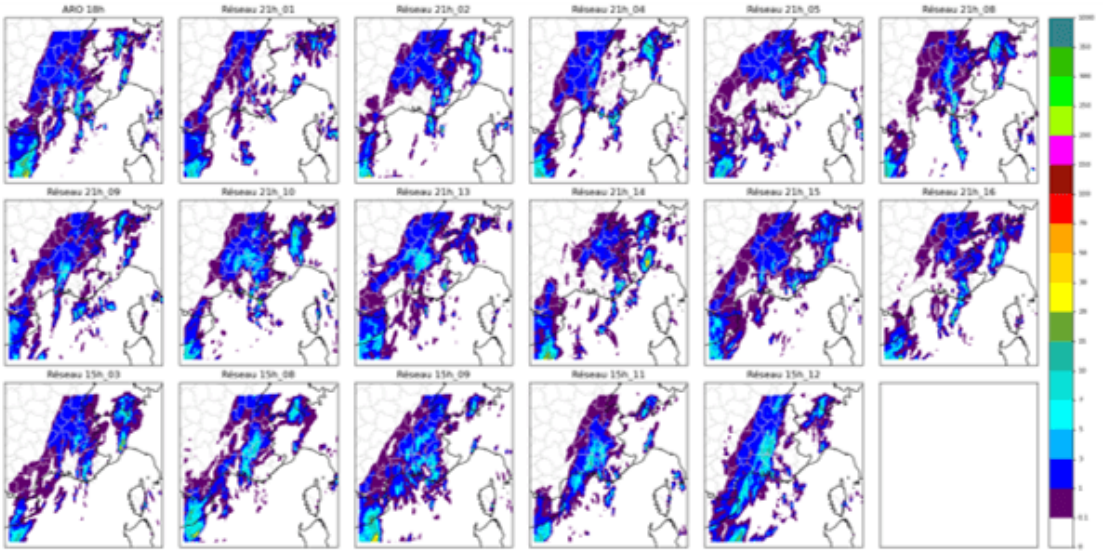


FIGURE 11.15 – *Scénario Espagne pour les réseaux du 3 octobre 2021 à 15h UTC et à 21h UTC, prévision pour le 4 octobre 2021 à 03h UTC.*

Pour conclure sur ces trois études de cas, elles ont montré que la synthèse par scénarios permet d'informer sur la prévisibilité du phénomène ou de détecter précocement des évènements. Elle permet de repérer pour la journée du 14 septembre 2021 un scénario alternatif mais violent sur les Cevennes qui n'était pas proposé par les réseaux AROME plus de 24h avant l'évènement puis de voir ce scénario devenir majoritaire avec l'arrivée de réseaux plus récents. La situation du 2 octobre 2020 montre une prévision beaucoup plus stable avec de fortes pluies sur les Alpes-Maritimes et l'Italie, les différences s'expliquant par la zone qui concentre les intensités les plus fortes selon les échéances. La dernière étude de cas montre que la synthèse par scénario peut aussi traduire une prévisibilité très faible, comme c'était le cas pour la journée du 4 octobre 2021. On peut remarquer sur l'ensemble de ces études de cas qu'aucun élément dans les graphiques de synthèse ne permet de différencier les scénarios avec fortes pluies des autres. Il manque une information sur l'intensité des pluies dans un scénario qui permettrait de surveiller plus spécifiquement certains d'entre eux. Il faut souligner que cette partie reste simplement une présentation de 3 études de cas et que ces conclusions ne peuvent pas être généralisées. En particulier, un scénario secondaire violent n'est pas plus probable qu'un scénario majoritaire « non significatif ». Il faut d'ailleurs rappeler que ces 3 études de cas sont des situations orageuses qui ont nécessité l'émission d'une vigilance par les services de Météo-France. Aucune étude de cas ne correspond à une situation calme où certaines de ces observations sur les scénarios risquent d'être non valables.

11.3 Évaluation de la méthode et des modèles

Après ces études de cas qui ont permis de mieux comprendre comment prend forme la synthèse par scénarios, des scores sont calculés sur une année complète avec deux objectifs : donner des éléments pour utiliser au mieux cette nouvelle

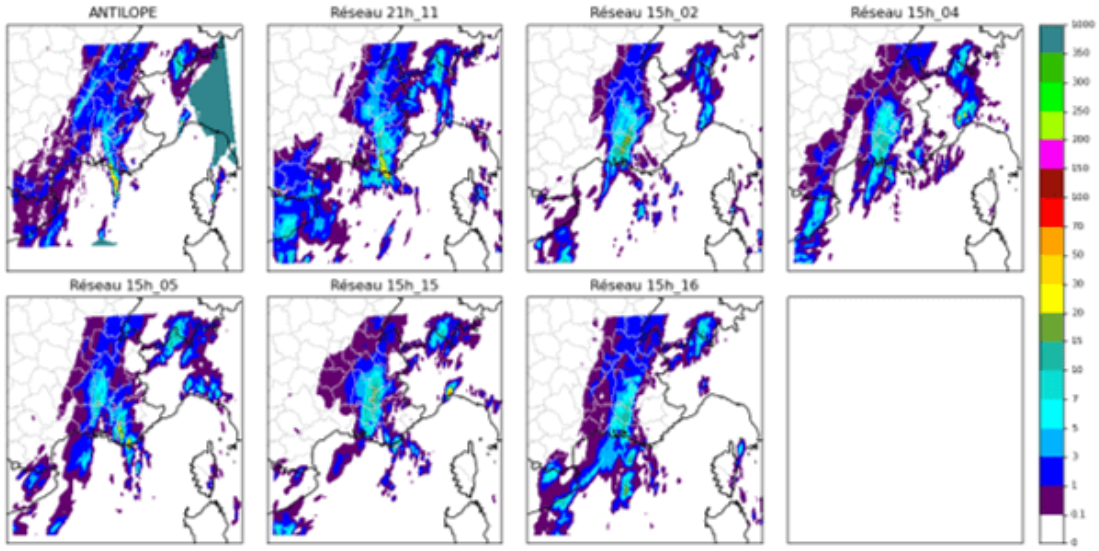


FIGURE 11.16 – *Scénario Var/PACA Ouest/Drôme-Sud, mêmes réseaux et même heure de prévision que la figure 11.15.*

forme de synthèse et évaluer ou comparer les modèles entre eux par le prisme de cette approche. En particulier cette approche va permettre de comparer les anciennes versions d'AROME et PE-AROME avec les nouvelles, utilisées en opérationnel depuis l'été 2022. Un rejeu de ces nouvelles versions qui commence le 1er février 2021 a en effet été mis à disposition pour permettre entre autres ce type de travaux grâce à des équipes du Groupe de Modélisation et d'Assimilation pour la Prévision (GMAP), de la Direction des Systèmes d'Information (DSI) et de la Direction des Opérations pour la Prévision (DIROP). L'année sur laquelle sont calculés ces scores va du 1er février 2021 au 31 janvier 2022 pour correspondre au rejeu des nouvelles versions. Sur cette période, la synthèse par scénarios est calculée quotidiennement pour deux réseaux successifs et avec un pas de 6h. Par exemple, elle est calculée pour les réseaux PE-AROME du 1er février à 03h et 09h, puis pour les réseaux de 09h et 15h, puis 15h et 21h, ... jusqu'au 31 janvier de l'année suivante. La figure 11.2 montrait qu'ANTILOPE est masqué en partie sur le domaine Sud-Est contrairement aux modèles. Cette différence peut engendrer des scores erronés car parmi les 20 structures météorologiques, certaines peuvent être impossibles à atteindre pour ANTILOPE. Contrairement aux études de cas, **les membres PE-AROME et les AROME seront donc masqués comme ANTILOPE pour établir les scores**.

11.3.1 Répartition ANTILOPE parmi les 20 classes

En premier lieu, le début de ce chapitre a évoqué des performances moindres d'ANTILOPE lorsque la zone est loin des radars et en l'absence de pluviomètre, ce qui est le cas sur plusieurs zones dans le Sud-Est de la France. Pour tenter de mettre en exergue ces différences, la figure 11.17 montre la répartition d'ANTILOPE parmi les 20 classes météorologiques, exception faite de la classe « Rien de significatif » car beaucoup plus fréquente que les autres (82,5 % du temps ANTILOPE est assigné à cette classe dans la base de données). La répartition est très inégale entre les classes mais elle cohérente avec la taille variable des

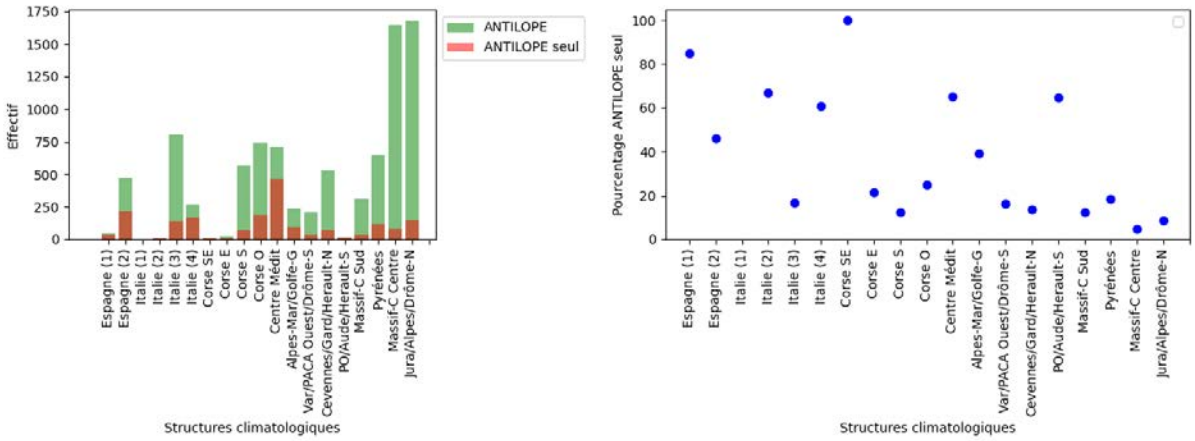


FIGURE 11.17 – *Répartition d'ANTILOPE à travers 19 des 20 structures météorologiques. Les noms correspondent à ceux donnés en annexe, figures E.3 à E.22. La classe « Rien de significatif » n'est pas représentée car elle représente un trop grand nombre de cas et rendrait la distinction des autres classes complexe. Les histogrammes de gauche correspondent en vert au nombre de fois qu'ANTILOPE a été assigné dans la classe et en rouge le nombre de fois où ce fut le cas seul, sans être accompagné par un membre PE-AROME. La carte de droite trace le pourcentage de fois où ce cas de figure se présente (ratio entre les valeurs en rouge et vert).*

classes qui forment les 20 structures météorologiques. Pour mettre en avant des différences de comportement, le nombre de fois où ANTILOPE a été assigné à une classe mais sans qu'aucun membre PE-AROME n'ait été simultanément assigné dans cette même classe est également tracé en rouge sur le graphique de gauche. Le graphique de droite représente le ratio entre les bars rouges et vertes sous forme de pourcentages. On remarque que ce pourcentage fluctue grandement d'une classe à l'autre. ANTILOPE se retrouve fréquemment seul sur les classes en Espagne, 3 des 4 classes en Italie, la classe « Corse SE », la classe « Centre Médit' » et la classe « PO/Aude/Herault-S ». Ce constat peut signifier soit que la PE-AROME a des performances très inégales selon les zones du domaine (puisque la condition porte sur l'absence de membre en accord avec ANTILOPE) ou que les données ANTILOPE ont des performances dégradées. Si on excepte la classe « PO/Aude/Herault-S » qui est sur la France ou proche des côtes, les autres classes sont loin des radars et avec aucun pluviomètre à proximité, ces éléments et ceux fournis par les équipes de la Direction des Systèmes d'Observation (DSO) en charge de ce produit nous conduisent à avoir des doutes sur la qualité d'ANTILOPE sur ces zones. **Pour la suite de ce chapitre, chaque échéance où ANTILOPE est assignée à l'une des classes suivantes ne sera pas pris en considération : Espagne (1), Espagne (2), Italie (1), Italie (2), Italie (4), Corse SE et Centre Médit'**. La classe « PO/Aude/Herault-S » est conservée car les éléments rassemblés sur ANTILOPE ne permettent pas d'aboutir au même raisonnement que pour les autres classes. Cette classe présente cependant la particularité d'être la plus petite des 20 classes et avec un champ moyen avec des valeurs élevées. Une tendance à la sous-estimation par la PE-AROME pourrait expliquer ce résultat et ne justifie pas dans ce cas une exclusion pour les scores à venir. Les cas exclus représentent 2,95 % du jeu de données au total, un peu moins de 17 % si la classe « Rien de significatif » n'est pas comptabilisée.

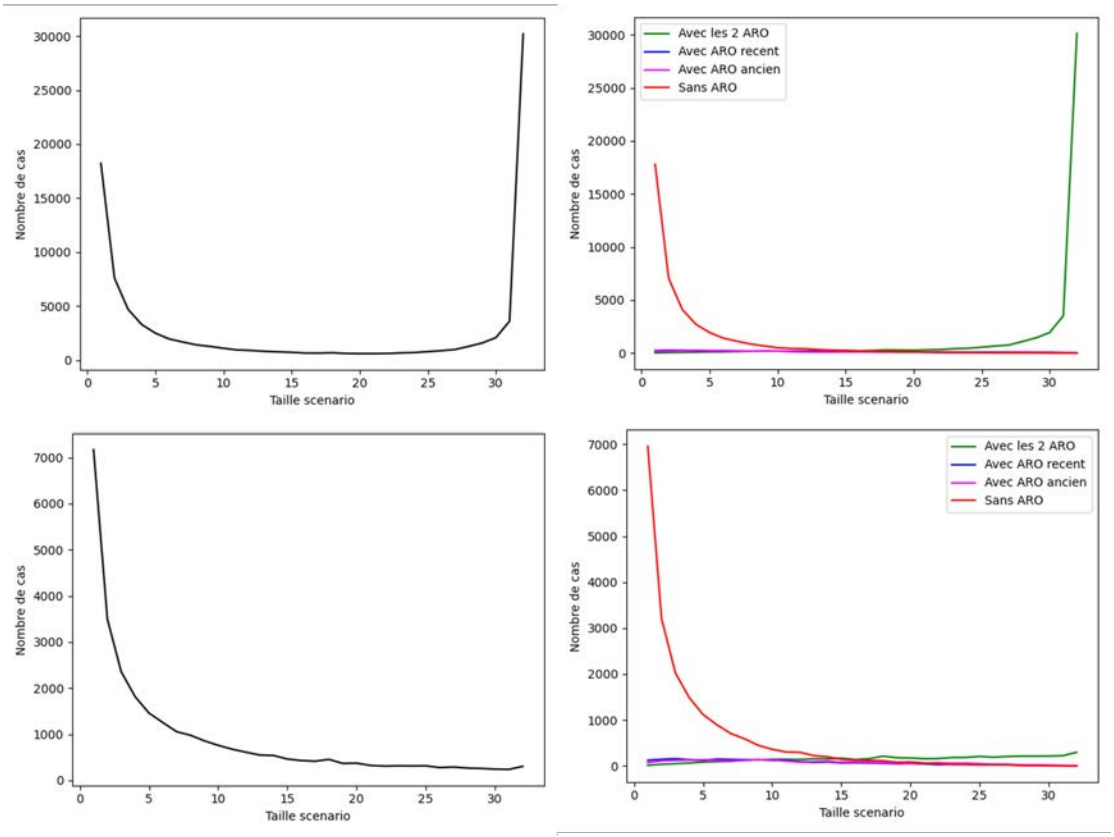


FIGURE 11.18 – *Répartition de la taille des scénarios. Le graphique en haut à gauche représente la distribution complète de la taille des scénarios sur le jeu de données. Le graphique en haut à droite divise cette distribution en sous-catégories : si les deux réseaux AROME appartiennent au scénario (vert), si seulement le réseau AROME le plus récent appartient au scénario (bleu), si seulement le réseau AROME appartient au scénario (violet) et enfin si aucun AROME appartient au scénario (rouge). Les graphiques en bas correspondent aux mêmes types de distribution mais où l'ensemble des cas avec ANTILOPE assigné à la classe « Rien de significatif » a été retirés.*

11.3.2 Taille des scénarios

Ce point est dédié à l'étude de la distribution des tailles des scénarios. On considère comme scénario toute structure météorologique récurrente à laquelle est assignée au moins un membre PE-AROME. Cette sous-partie permettra de mieux comprendre les scores qui suivront. La figure 11.18 (haut à gauche) permet de remarquer que la plupart des scénarios possèdent soit très peu de membres (moins de 5 membres), soit au contraire un nombre très conséquent supérieur à 30 membres sur 32. La figure en haut à droite montre que ces deux extrêmes de la distribution possèdent des propriétés différentes avec des scénarios à 32 membres soutenus par les deux réseaux de AROME mais très peu pour les scénarios plus petits. Ces scénarios à 32 membres se retrouvent en situation calme où il ne pleut pas ou presque. Cette situation, assez fréquente sur la zone Sud-Est, se traduit par une prévision dans la classe « Rien de significatif » pour l'immense majorité des membres, conduisant à ce grand nombre de scénarios à plus de 30 membres. Pour confirmer cette hypothèse, la distribution est retracée en bas à gauche sur la figure 11.18 en supprimant les cas où ANTILOPE a été assigné à la classe « Rien de

significatif ». Les scénarios avec 32 membres ont presque disparu mais on observe toujours beaucoup de scénarios avec peu de membres. La suppression de la classe « Rien de significatif » conduit à se concentrer sur les cas avec de la pluie et donc souvent une prévisibilité plus limitée, ce grand nombre de scénarios avec peu de membres correspond aux différentes possibilités explorées par la PE-AROME. Ce très grand nombre de bonnes prévisions « simples » lorsqu'il ne pleut pas conduira certains scores à être peu informatifs (problèmes liés aux évènements rares). C'est pourquoi la plupart des scores seront tracés en excluant les cas où ANTILOPE est dans la classe « Rien de significatif ». **On dira qu'on se concentre sur les cas pluvieux**.

11.3.3 Intérêt du multi-réseaux

On peut se demander si l'utilisation de plusieurs réseaux de la PE-AROME a un intérêt pour la synthèse par scénarios avec la question suivante : quel est l'apport du réseau le plus ancien par rapport au plus récent ? Pour répondre à cette question, la répartition des membres dans les scénarios corrects selon leur réseau d'origine va être étudiée. **Les scénarios corrects désignent les classes où ANTILOPE est assigné**, on étudie alors les membres qui l'accompagnent. Cette notion de « scénario correct » sera réutilisée à plusieurs reprises dans ces scores à venir. En comptabilisant l'ensemble des membres accompagnant les scénarios corrects sur l'ensemble des cas pluvieux, on calcule que 52,15 % des membres sont issus du réseau PE-AROME le plus récent (et donc 47,85 % du réseau le plus ancien). Le réseau le plus récent est donc légèrement plus représenté mais l'écart est faible avec l'ancien réseau. Ce faible écart accrédite donc l'utilisation de deux réseaux PE-AROME pour établir la synthèse par scénarios. L'apport des réseaux les plus anciens par rapport aux plus récents confirme les travaux de Raynaud et al. (2015, 2019) qui avaient abouti aux mêmes conclusions. Si on s'intéresse à la répartition des réseaux selon la taille de tous les scénarios (et pas seulement des scénarios corrects), on remarque que la PE-AROME la plus ancienne est plus représentée dans les petits scénarios à moins de 3 membres (environ 55 %, non montré). Ce résultat peut s'expliquer par la dispersion dans cette PE-AROME qui est en moyenne un peu plus grande pour une date donnée que pour la PE la plus récente à cause de ce décalage de 6h entre les deux réseaux. La PE-AROME la plus ancienne peut alors explorer des scénarios avec quelques membres qui seront abandonnés avec l'arrivée du réseau le plus récent donnant de petits scénarios où la PE-AROME la plus ancienne est majoritaire (voire uniquement présente).

11.3.4 Fiabilité

Un score classique pour évaluer les ensembles en météorologie est la courbe de fiabilité. Un ensemble est fiable si sur l'ensemble des cas où la probabilité prévue de l'évènement est de X %, l'évènement est bien observé X % du temps. Cette propriété peut se mesurer à l'aide d'une courbe de fiabilité qui est tracée figure 11.19. Cette courbe représente le pourcentage de scénarios corrects en ordonnée selon la taille des scénarios, exprimée en pourcentage des 32 membres. La classe « Rien de significatif » est ici conservée. **Cette courbe montre que l'approche par scénario est globalement fiable (courbe proche de la droite $y = x$)**,

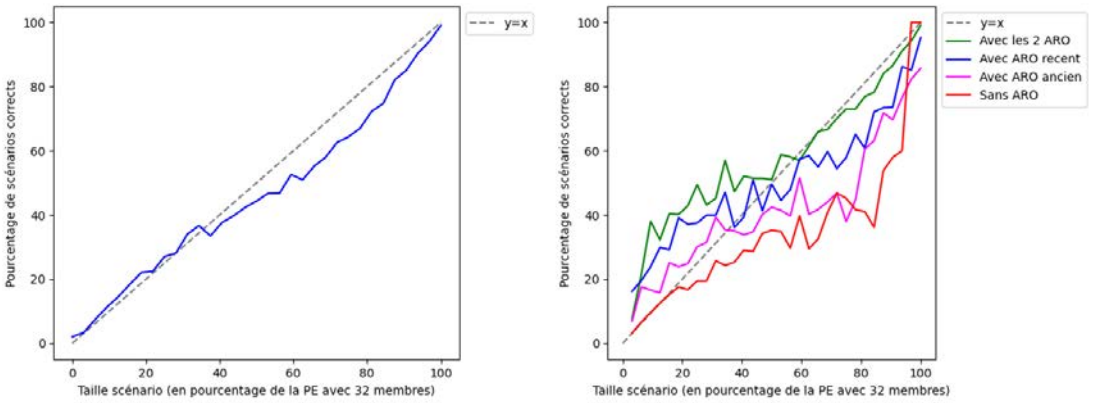


FIGURE 11.19 – *Courbe de fiabilité avec une approche par scénarios (graphique de gauche). Toutes les échéances sont prises en compte. Cette courbe est décomposée en quatre sur le graphique de droite comme ce fut le cas pour la taille des scénarios pour la figure 11.18.*

les scénarios représentant 20 % des membres de l'ensemble sont bien en moyenne associés à ANTILOPE 20 % du temps. La forme de la courbe traduit cependant une légère sous-dispersion de l'ensemble. L'ensemble est « trop confiant » notamment avec les scénarios qui représentent 80 % de l'ensemble où ANTILOPE n'est associé à ceux-ci qu'un peu plus de 60 % du temps. Comme pour la figure 11.18, la courbe de fiabilité peut être tracée selon l'appartenance ou non d'un ou des réseaux AROME au scénario (graphique de droite, figure 11.19). On peut remarquer que les 4 courbes diffèrent par moment grandement avec celles sur le graphique de gauche. Il faut cependant rappeler que le nombre de cas est très différent pour tracer chacune des 4 courbes. Les scénarios avec un seul AROME (représentés par les courbes bleue et rose) sont très peu nombreux (figure 11.18), ils n'ont qu'un poids limité pour reconstruire la courbe de fiabilité (figure 11.19, gauche). La courbe rouge, qui se focalise sur les scénarios sans aucun réseau AROME, a elle un poids très fort pour les petites tailles et la courbe verte (scénarios avec les deux réseaux AROME) pour les grandes tailles (figure 11.18). Ces explications permettent de mieux comprendre ces différences entre la courbe de gauche et ces 4 courbes à droite. Le point important à retenir est que la courbe rouge (scénario sans un réseau AROME) reste généralement en dessous des autres courbes, en particulier la verte rassemblant les scénarios avec les deux AROME. **On peut donc en conclure qu'un scénario issu de la PE-AROME mais sans aucun AROME en accord avec celui-ci a moins de chance de s'avérer correct qu'un scénario de taille similaire mais où au moins un AROME appartient à ce même scénario**. Cette conclusion indique que les réseaux AROME sont de bons moyens pour guider dans le choix entre deux scénarios. Mais il faut aussi rappeler que le nombre de scénarios peut être plus conséquent, le choix entre 3 ou 4 scénarios peut alors devenir plus complexe et ceci reste une conclusion sur une longue période. L'étude de cas du 14 septembre 2021 avait par exemple montré l'émergence d'un scénario minoritaire à raison dans les PE-AROME qui n'a été confirmé que plus tardivement par les réseaux AROME.

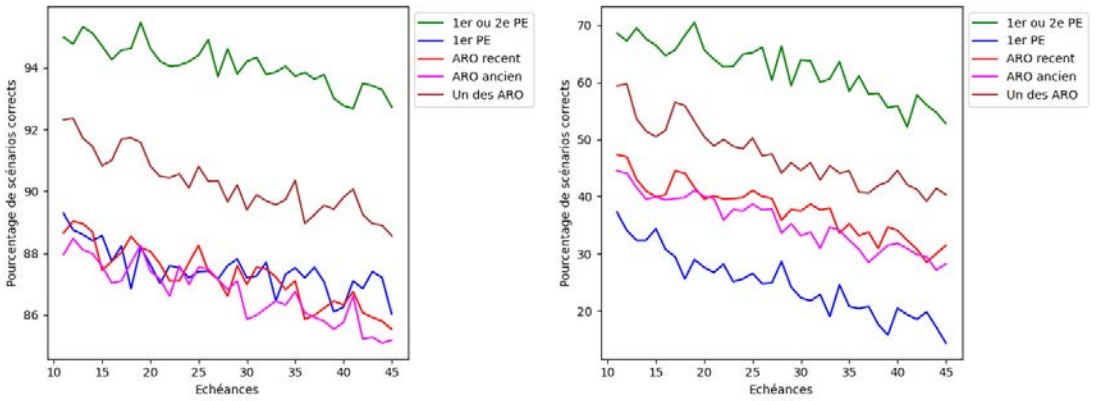


FIGURE 11.20 – *Pourcentage de scénarios corrects selon les scénarios ou modèles choisis. Les échéances en abscisse sont celles de la PE-AROME la plus ancienne (cf section 11.1.1, correspondant au réseau de n h). Pour les scénarios, les courbes bleues désignent le scénario majoritaire, les courbes vertes le scénario majoritaire ou secondaire. Les courbes rouges et roses désignent les deux réseaux AROME et la courbe marron si on considère les deux réseaux AROME. Ces courbes sont tracées dans le cas général (avec la classe « Rien de significatif ») à gauche, seulement sur les cas pluvieux à droite.*

11.3.5 Comment choisir un scénario de prévision ?

Dans cette sous-partie, on se demande dans cette vision par scénarios s'il est plus intéressant de se concentrer sur les sorties d'AROME ou sur les scénarios de la PE-AROME. La figure 11.20 permet de répondre à cette interrogation. Les courbes vertes montrent le pourcentage de fois où le scénario correct correspond soit au scénario majoritaire, soit au deuxième scénario le plus peuplé (scénario secondaire). Sur le même principe, les courbes marrons désignent ce pourcentage pour un scénario correct qui contient au moins l'un des deux réseaux AROME. Dit autrement, il s'agit du pourcentage de scénarios corrects parmi les scénarios contenant l'un des 2 AROME. On peut tout d'abord remarquer que l'ensemble des courbes tendent à décroître avec les échéances, traduisant une difficulté accrue de trouver le bon scénario. Cette décroissance est lié au nombre de scénarios possibles qui tend à croître avec les échéances (cf figure 10.8) car la prévisibilité diminue. Le graphique à gauche montre que la confiance à avoir dans le scénario majoritaire est à peu près identique à celle pour un réseau AROME. Des stratégies reposant sur le scénario majoritaire ou un réseau AROME reviennent à ne considérer qu'un seul scénario, on rejoint une vision déterministe. Les courbes marrons et vertes traduisent des stratégies différentes où on s'autorise à ne pas regarder un seul scénario mais deux. Il est alors nettement plus souhaitable de se concentrer sur les deux scénarios les plus peuplés issus de la PE-AROME que sur les deux réseaux d'AROME. Ce graphique à gauche présente d'excellents scores supérieurs à 85 % grâce aux nombreuses prévisions « simples » d'absence de pluie. Le graphique à droite se concentrant sur les cas pluvieux, les scores chutent. Le scénario majoritaire devient alors moins intéressant à regarder qu'un réseau AROME. Mais les autres conclusions ne changent pas, en particulier **il est préférable de regarder les deux scénarios les plus peuplés plutôt que les deux réseaux d'AROME pour espérer trouver le bon scénario**. L'écart est conséquent

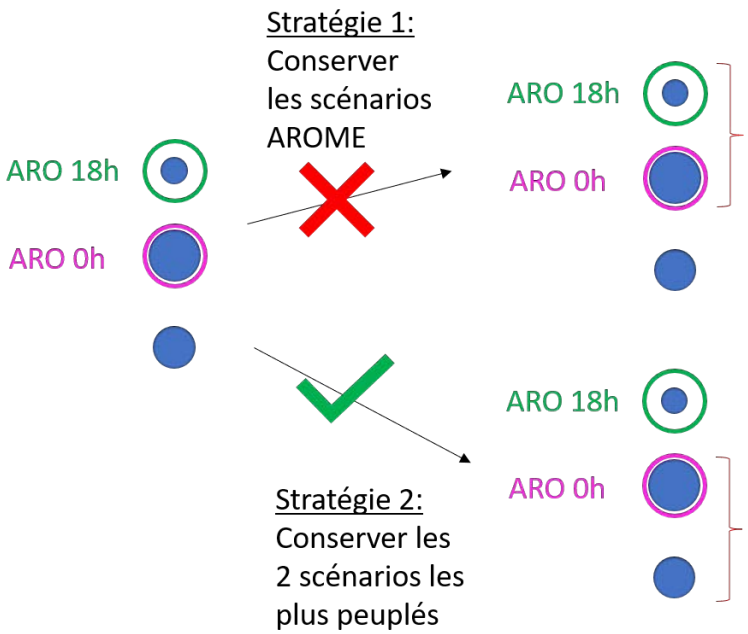


FIGURE 11.21 – *Schéma pour prendre une décision entre trois scénarios (points bleus), dont deux accompagnés par un réseau AROME (cercles vert et rose).*

(environ 10 %) quelles que soient les échéances. Cette information peut se traduire comme dans l'exemple de la figure 11.21. Si on doit choisir deux des trois scénarios, le scénario le moins peuplé doit être éliminé même s'il est accompagné d'un déterministe au profit de l'autre scénario PE-AROME. La stratégie 1 aboutit en moyenne à un taux de succès de 45 % aux longues échéances, la deuxième à un taux de 55%. Mais en lien avec les conclusions de la partie précédente, à choisir entre les deux scénarios restants, il faut conserver celui associé à AROME sauf si l'écart en taille est vraiment très important et en défaveur du scénario avec AROME.

11.3.6 Statistiques sur les scénarios corrects

Cette partie est dédiée spécifiquement à l'étude des scénarios corrects (Pour rappel, ceux vérifiés par ANTILOPE). Ces scénarios corrects seront étudiés à travers leur répartition entre scénarios avec ou sans AROME, leur taille moyenne selon les échéances et enfin leur répartition entre les scénarios majoritaires, secondaires, tertiaires et quaternaires.

Scénarios avec ou sans AROME

Dans la continuité de la partie précédente, on peut se demander comment les scénarios corrects se répartissent entre des scénarios avec les deux réseaux AROME, un seul réseau ou aucun (figure 11.22). Le graphique de gauche montre que dans le cas général le scénario correct est un scénario PE-AROME avec les deux réseaux AROME. Mais ce graphique est peu informatif car trop fortement impacté par les bonnes prévisions d'absence de précipitation. Le graphique de droite est plus intéressant et montre qu'aux premières échéances les scénarios corrects

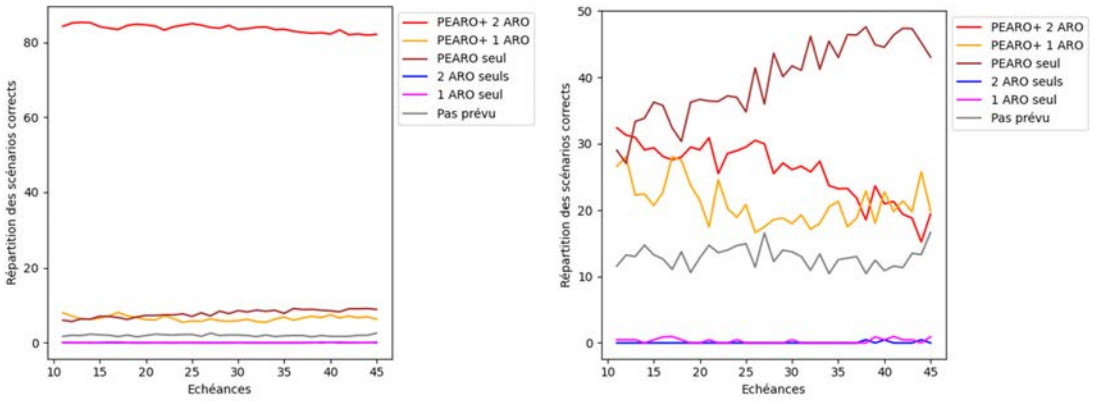


FIGURE 11.22 – *Répartition des scénarios corrects selon la présence ou non d'un réseau AROME dans le scénario. La courbe rouge représente les scénarios avec les deux réseaux AROME, la courbe jaune les scénarios avec strictement un réseau AROME, la courbe marron les scénarios avec seulement des membres PE-AROME, la courbe bleue avec exclusivement les deux réseaux AROME, la courbe rose exclusivement un réseau AROME et la courbe grise représente les scénarios corrects anticipés par aucun membre et aucun réseau AROME. La carte de gauche représente le cas général, la carte de droite les cas pluvieux. Ces courbes sont tracées en fonction des échéances comme pour la figure 11.20.*

se répartissent équitablement entre les scénarios avec les deux réseaux AROME, avec un seul et ceux sans aucun réseau AROME. Il reste 10% du temps où le scénario correct n'a été anticipé par aucun membre ou réseau AROME. Ces proportions évoluent au fil des échéances. Les scénarios avec deux réseaux AROME deviennent moins fréquents, tout comme les scénarios avec un seul réseau. Cette baisse se fait au profit des scénarios sans aucun réseau AROME qui sont en nette augmentation. Aux longues échéances, les scénarios corrects sont environ 45% du temps des scénarios sans aucun réseau AROME, 20 % du temps avec deux réseaux, 20 % avec un seul réseau et 15 % du temps le scénario n'a été anticipé par aucun membre PE-AROME ou réseau AROME. Ce graphique illustre très bien les limites des prévisions déterministes AROME. Elles perdent de plus en plus d'influence lorsque les échéances augmentent, ce qui est cohérent avec les travaux sur la prévisibilité évoqués dans le chapitre 1. **On retiendra qu'il faut prendre de moins en moins en considération la prévision des réseaux AROME lorsque les échéances augmentent. Le scénario correct est en effet plus souvent sans aucun réseau AROME qu'avec les deux ou l'un des deux réseaux aux longues échéances.** On remarquera également que les scénarios corrects ne sont presque jamais accompagnés exclusivement par des réseaux AROME (courbes bleu et rose). On peut donc en conclure qu'il est peu probable qu'une prévision déterministe explore de façon pertinente un scénario qui serait couvert par aucun membre de la PE-AROME. Une dernière remarque concerne la courbe grise et les cas où aucun membre ni réseau AROME n'a anticipé le bon scénario. Ce taux est d'environ 10 % et est globalement stable alors qu'on pourrait s'attendre à ce que ce taux augmente avec les échéances. Les cas qui constituent cette courbe grise ont été décomposés selon les 20 structures météorologiques. Cette décomposition (non montrée) révèle que pour la plupart des classes on observe bien une légère augmentation de cas non couverts par les mo-

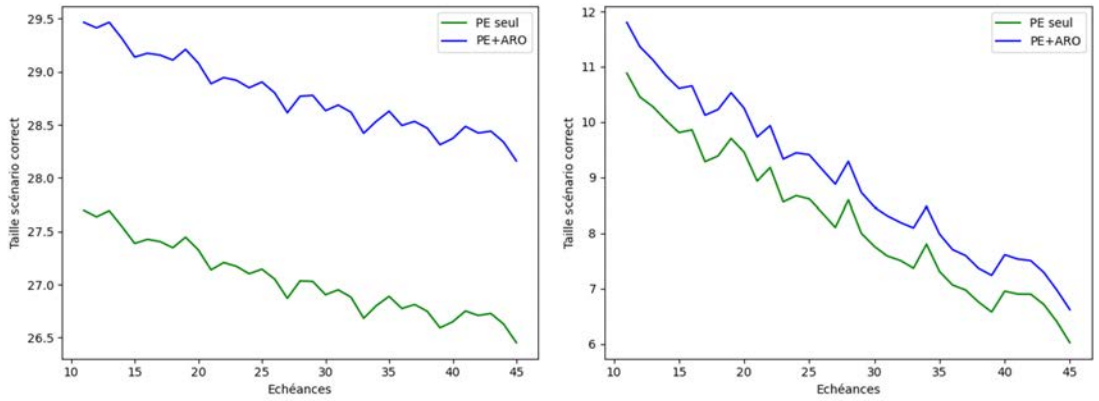


FIGURE 11.23 – *Taille moyenne des scénarios corrects. La courbe verte représente la taille moyenne si on considère uniquement les membres de la PE-AROME. Sur la courbe bleue, les réseaux AROME sont également comptabilisés. La carte de gauche représente le cas général, la carte de droite les cas pluvieux. Ces courbes sont tracées en fonction des échéances comme pour la figure 11.20.*

dèles lorsque les échéances augmentent. Mais quelques classes compensent avec au contraire de moins en moins de cas. C'est particulièrement vrai pour la seule classe « Italie » qui n'a pas été exclue au début de cette partie (section 11.3.1) et la classe « Corse O ». Ces classes ne semblaient pas problématiques sur la figure 11.17 mais elles sont malgré tout, elles aussi, éloignées des pluviomètres et radar de Météo-France. Ce point devra être éclairci dans de futurs travaux mais le comportement de ces classes notoirement différent par rapport aux autres peut aussi poser la question de la qualité d'ANTILOPE dans ces zones.

Taille

Pour évaluer à combien de membres sont associés en moyenne le scénario correct, la taille moyenne de ce scénario est tracée selon les échéances sur la figure 11.23. Comme pour le graphique précédent, le graphique de gauche est peu intéressant avec une taille moyenne très élevée à environ 27 membres sur 32. En se concentrant sur les cas pluvieux, on note que le nombre est beaucoup plus restreint et varie de environ 11 membres sur 32 aux premières échéances à un peu plus de 6 aux dernières échéances. L'ajout des réseaux AROME permet une augmentation d'environ 0.8 points. L'écart entre les courbes verte et bleue est un peu plus important aux premières échéances (+0.9) qu'aux dernières (+0.6) traduisant comme pour les membres PE-AROME une plus grande difficulté à trouver le bon scénario pour les AROME. En moyenne, sans compter les réseaux AROME, le scénario correct a une taille de 8.3 membres sur les cas pluvieux. Mais si on calcule cette moyenne en différenciant les scénarios comme pour la figure 11.22 selon l'accord ou non des ou d'un réseau AROME avec le scénario, cette moyenne fluctue grandement : les scénarios corrects avec deux réseaux AROME comptent 16.6 membres (18.6 si on compte les AROME), 9.5 membres avec un réseau strictement (10.5 si on le comptabilise) et seulement 4.9 membres pour les scénarios corrects lorsqu'aucun AROME n'est d'accord avec celui-ci. **Ces différences montrent que lorsque les membres PE-AROME ont raison alors que les deux réseaux AROME se trompent, ce scénario est petit avec seulement 5 membres**

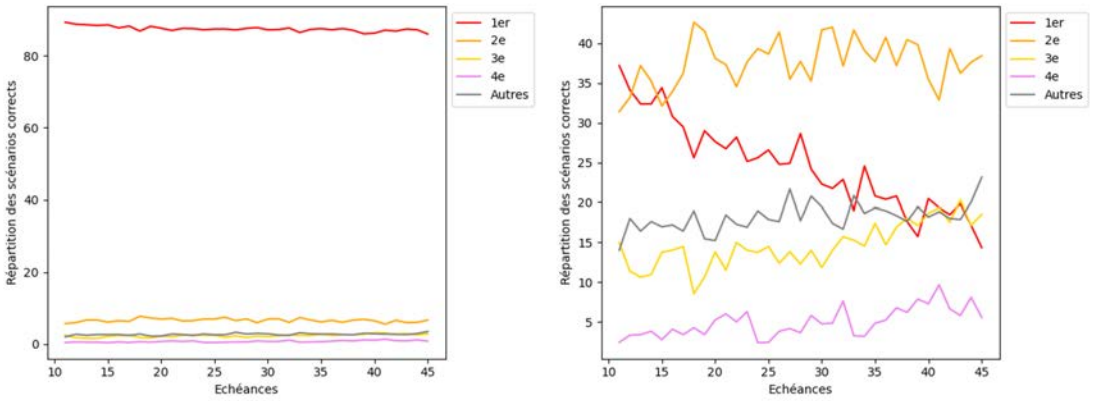


FIGURE 11.24 – *Répartition des scénarios corrects selon l'ordre des scénarios du plus peuplé au moins peuplé. La courbe rouge représente les scénarios majoritaires, la courbe orange les scénarios secondaires, la courbe jaune les scénarios tertiaires, la courbe rose les scénarios quaternaires et la courbe grise les cas autres (au-delà du quatrième scénario ou non couvert par les membres). La carte de gauche représente le cas général, la carte de droite les cas pluvieux. Ces courbes sont tracées en fonction des échéances comme pour la figure 11.20.*

sur 32 en moyenne. Il ne faut donc pas s'attendre à trouver fréquemment de grands scénarios PE-AROME à plus de 16 membres qui auraient raison contre de plus petits soutenus par des réseaux AROME.

Répartition entre les quatre scénarios les plus peuplés

Un dernier aspect concernant les scénarios corrects est leur répartition à travers le scénario majoritaire, secondaire, tertiaire et quaternaire. Une question sous-jacente étant combien de scénarios est-il nécessaire d'examiner pour espérer regarder le bon ? ou encore quel scénario regarder en priorité, en particulier est-ce toujours le majoritaire le plus intéressant ? La figure 11.24 permet de répondre à ces questions. Il n'est plus question dans cette partie des réseaux AROME. Le graphique de gauche montre que dans le cas général le scénario majoritaire est le plus intéressant et de très loin, suivi par le scénario secondaire puis par le scénario tertiaire et les cas non couverts par la PE-AROME ou qui sont au delà du quatrième scénario, ce dernier étant représenté en rose. Cette répartition est cependant comme pour les derniers graphiques relativement peu informative. Le graphique de droite qui se concentre sur les cas pluvieux montre que les scénarios majoritaires et secondaires représentent chacun autour de 35 % des scénarios corrects. Mais la proportion de scénarios majoritaires chute avec les échéances et tombe à environ 15 % aux échéances lointaines. Cette chute est au bénéfice de tous les autres scénarios. Elle s'explique car sur des situations pourtant pluvieuses, le scénario majoritaire est malgré tout associé avec la classe « Rien de significatif » dans des proportions importantes : 51% du temps durant les 5 premières échéances et ce pourcentage grimpe à 69 % durant les 5 dernières échéances. La forte augmentation de ce pourcentage explique la baisse constatée sur le graphique de droite. En conclusion, il ne faut pas faire aveuglement confiance au scénario majoritaire car il sera fréquemment associé à la classe « Rien de significatif » même lorsqu'un cas pluvieux risque de se présenter. **Il est globalement difficile d'établir une règle**

arbitraire pour choisir le bon scénario, les scénarios corrects pouvant tomber aussi bien dans le scénario majoritaire, secondaire que tertiaire pour les longues échéances. Un raisonnement probabiliste est donc nécessaire. Les trois premiers scénarios couvrent 85 % des scénarios corrects aux première échéances mais ce pourcentage tombe à environ 70 % pour les longues échéances. Il faut rappeler que 10 à 15 % des cas ne sont prévus par aucun modèle (figure 11.22) et sont donc impossibles à trouver même en regardant le 5e ou 6e scénario.

11.3.7 Comparaison anciennes et nouvelles versions

Les scores présentés dans les parties précédentes ont été calculés avec les versions de PE-AROME et AROME qui étaient en opérationnel avant l'été 2022. La configuration des membres PE-AROME était alors différente différente de celle d'AROME, cette différence a disparu avec les nouvelles versions (grille à 1.3 km en particulier). Cette dernière partie s'intéresse à la comparaison de ces scores entre les anciennes versions et les nouvelles qui sont utilisées en opérationnel depuis l'été 2022. Pour rappel, cette comparaison est permise par un rejeu des nouvelles versions sur une période commune avec les anciennes du 1er février 2021 au 31 janvier 2022. La figure 11.25 rassemble les derniers scores présentés pour les deux versions. En premier lieu ces graphiques montrent que les conclusions des parties précédentes restent valables sur l'utilisation des scénarios majoritaires et secondaires ou encore sur la considération à porter aux réseaux AROME avec les nouvelles versions. La courbe de fiabilité en haut à gauche montre que la nouvelle version de PE-AROME est plus fiable que l'ancienne dans cette vision par scénarios. La décomposition de la courbe de fiabilité comme pour le graphique de droite de la figure 11.19 montre que ces gains s'observent pour l'ensemble des scénarios avec ou sans AROME (non montré). Le graphique en haut à droite correspondant à la section 11.3.5 montre des gains importants que ce soit pour les scénarios majoritaires et secondaires ou les réseaux AROME aux longues échéances entre +5 et +10 % de taux de réussite. Les graphiques au centre de la figure 11.25 montrent comment la répartition des scénarios corrects a été modifiée dans les nouvelles versions. A gauche, on constate qu'aux longues échéances le nombre de bons scénarios comptant les deux réseaux AROME a progressé, au détriment des réseaux sans AROME et des cas prévus ni par PE-AROME ni par les AROME (courbe grise). Le graphique de droite montre que le scénario majoritaire a progressé tout au long des échéances au détriment des scénarios au-delà du quatrième scénario ou non prévu par les modèles (courbe grise). Enfin les deux derniers graphiques en bas de la figure 11.25 montrent l'évolution de la taille des scénarios corrects en comptabilisant seulement les membres PE-AROME (gauche) ou également les réseaux AROME (droite). La taille moyenne si on ne comptabilise pas les réseaux AROME passe de 8.3 membres à 9 membres, avec des gains plus importants aux longues échéances dépassant +1 membre. On observe la même tendance sur le graphique de droite en comptabilisant les réseaux AROME. **On peut conclure de ces graphiques que la vision par scénarios montre des améliorations notables avec les nouvelles versions des modèles.** Il restera à vérifier que ces améliorations sont significatives statistiquement parlant. Cette partie démontre que la synthèse par scénarios pourrait aussi devenir un moyen d'évaluer une nouvelle version des modèles de façon complémentaire aux scores déjà utilisés comme

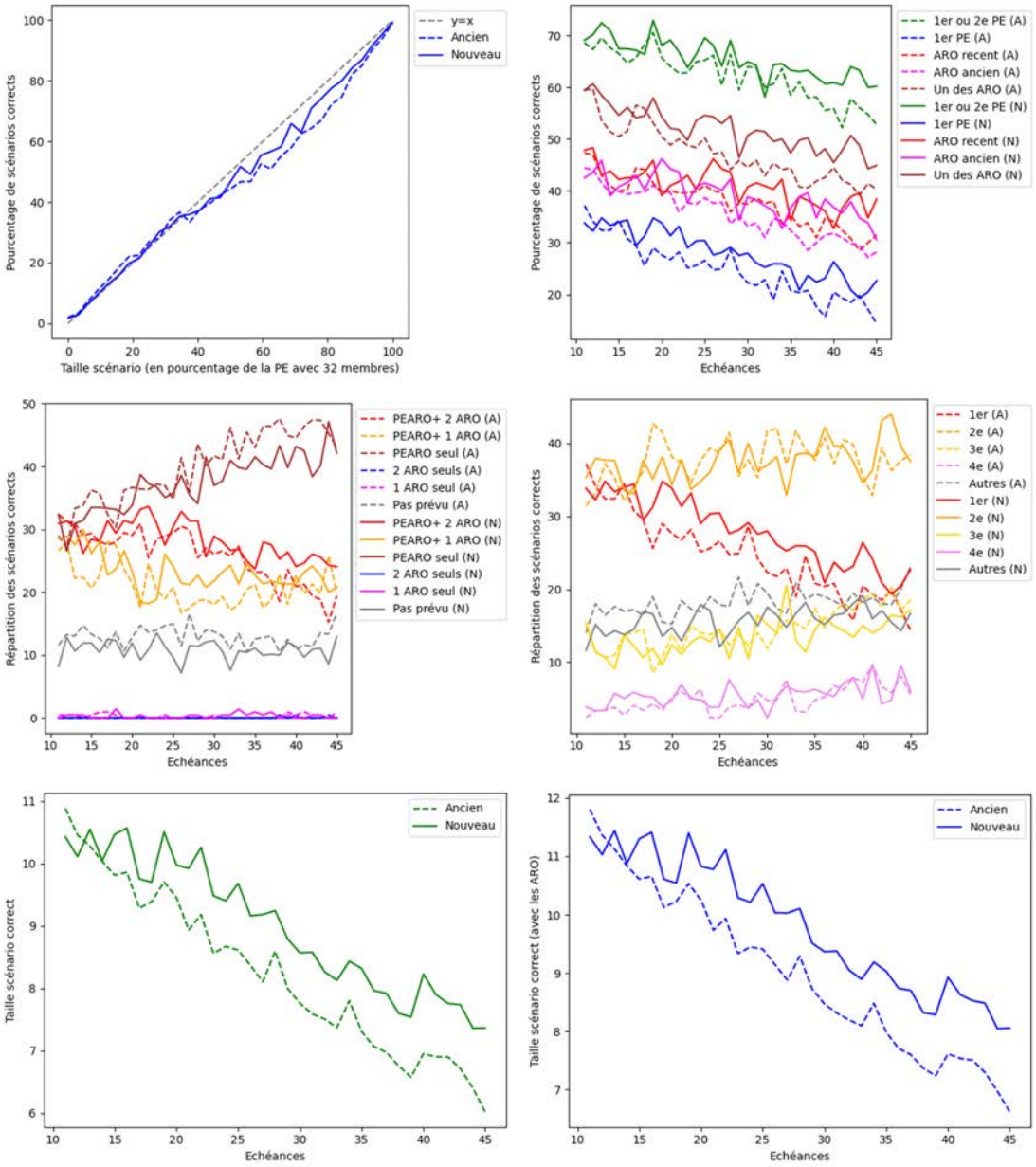


FIGURE 11.25 – *Comparaison des scores des sections 11.3.4 à 11.3.6 entre les anciennes versions des modèles PE-AROME et AROME (courbes tiretés) et les nouvelles versions en opérationnel depuis l'été 2022 (courbes pleines). A l'exception des courbes de fiabilité, les autres graphiques se concentrent uniquement sur les cas pluvieux (le cas général étant jugé moins intéressant).*

le CRPS.

Le dernier chapitre de ce manuscrit a commencé par la présentation graphique de la synthèse par scénarios et la représentation de chaque scénario à l'aide des membres qui le composent. Ces modes de représentation ont été utilisés sur trois études de cas pour montrer comment la synthèse par scénarios peut :

- Repérer l'émergence d'un scénario secondaire dont la taille croit en se rapprochant de l'évènement (14 septembre 2021)
- Retranscrire une situation avec plusieurs zones impactées par de fortes précipitations simultanément (2 octobre 2020)
- Mettre en évidence une situation avec une prévisibilité très faible (4 octobre 2021)

Après ces études de cas, des scores ont été calculés sur un an de données. Ces scores ont permis de mettre en évidence plusieurs points importants pour l'utilisation de la synthèse par scénarios :

- Le multiréseau présente un réel intérêt pour cette approche.
- L'approche par scénarios est fiable statistiquement parlant.
- Le scénario majoritaire issu de la PE-AROME propose des performances comparables ou un peu inférieures par rapport à la prévision déterministe. Mais la PE-AROME propose des scénarios secondaires ou tertiaires qui permettent d'explorer l'espace des possibles de façon plus pertinente qu'en utilisant plusieurs réseaux de la prévision déterministe. En particulier, il est souhaitable si on ne doit conserver que deux scénarios de garder les deux plus gros issus de la PE-AROME et non les deux issus des réseaux AROME.
- Entre deux scénarios de taille comparable, il faut s'intéresser en priorité à celui qui est accompagné par un déterministe mais l'influence du déterministe diminue avec les échéances. Par exemple aux longues échéances, le scénario correspondant à ANTILOPE est proposé environ 45 % du temps par un scénario avec seulement des membres PE-AROME [2], 40% sont des scénarios avec au moins un déterministe AROME et 15% des cas ne sont pas correctement anticipés par des modèles.
- Les scénarios corrects avec seulement des membres PE-AROME ont une taille en moyenne de seulement 5 membres, il faut donc également considérer avec intérêt des scénarios de taille modeste surtout aux longues échéances.
- Il est difficile d'établir un critère pour choisir systématiquement le bon scénario.
- Les performances des nouvelles versions de PE-AROME et AROME ont été comparées avec les anciennes versions pour mettre en évidence des performances meilleures sur beaucoup d'aspects (qui devront être confirmées par des tests de significativité).

2. pour les cas où on a observé des pluies significatives

Pour finir sur ce chapitre, il faut souligner que l'entraînement de l'autoencodeur et la classification sont des étapes très longues mais l'approche par scénarios en contexte opérationnel est très peu coûteuse. La projection des membres dans l'espace latent et leurs assignations aux structures météorologiques récurrentes ne prend que quelques minutes. C'est aussi un point essentiel pour une utilisation future de cette approche.

Bibliographie

E. Bresson. *Mécanismes de formation des systèmes convectifs quasi-stationnaires en Méditerranée nord-occidentale : application au cas du 15 juin 2010 sur le Var*. PhD thesis, Université Paul Sabatier - Toulouse III, 2011.

C. F. Chappell. Quasi-stationary convective events. In *Mesoscale meteorology and forecasting*, pages 289–310. 1986. doi : 10.1007/978-1-935704-20-1_13.

L. Raynaud, O. Pannekoucke, P. Arbogast, and F. Bouttier. Application of a Bayesian weighting for short-range lagged ensemble forecasting at the convective scale. *Quart. J. Roy. Meteor. Soc.*, 141(687) :459–468, 2015. doi : https://doi.org/10.1002/qj.2366.

L. Raynaud, I. Pechin, P. Arbogast, L. Rottner, and M. Destouches. Object-based verification metrics applied to the evaluation and weighting of convective-scale precipitation forecasts. *Quart. J. Roy. Meteor. Soc.*, 145(722) :1992–2008, 2019. doi : https://doi.org/10.1002/qj.3540.

Conclusion

Le chapitre 1 a montré la structuration de l'atmosphère et les difficultés rencontrées afin de modéliser ses états futurs. Pour établir une prévision, des modèles déterministes ont été développés depuis les années 1980 pour la grande échelle, se caractérisant par une prévision unique en chaque point de grille à chaque échéance. Cette approche est problématique pour prévoir les évolutions futures d'un système chaotique et c'est pourquoi des prévisions d'ensemble ont été développées. L'arrivée des modèles à fine échelle au début des années 2000, qui sont capables de simuler explicitement la convection, s'est logiquement suivie du développement de systèmes ensemblistes pour cette échelle, comme PE-AROME à Météo-France.

Analyser les membres des ensembles un à un serait trop chronophage en condition opérationnelle. Plusieurs méthodes de post-traitement ont donc été développées et ont fait l'objet du chapitre 2. Ces multiples états futurs possibles peuvent être synthétisés sous forme de panaches ou de boîtes à moustache. Les approches les plus classiques à deux dimensions reposent sur des cartes de probabilités ou de quantiles pour chaque paramètre météorologique. Ces membres peuvent également être regroupés dans des classes. Concernant les évènements extrêmes, plusieurs approches tentent de détecter et de suivre ces évènements dans les différents membres pour construire un éventail des possibilités futures. Des indices de sévérité (EFI) existent également pour repérer les jours et les zones où une prévision d'ensemble propose une distribution qui diffère nettement de la climatologie du modèle.

La plupart de ces méthodes de post-traitement présentent cependant un écueil important pour les prévisionnistes. Elles ne permettent pas le suivi de façon cohérente des objets structurants de l'atmosphère. En effet, notre compréhension de l'atmosphère s'est développée à travers des structures cohérentes à toutes les échelles. Ce sont ces objets qui permettent aux météorologues de faire une analyse de la situation et de critiquer ou comparer les sorties des modèles météorologiques comme l'a montré le chapitre 3. Une approche purement statistique par points de grille des prévisions d'ensemble est donc un frein important à leur utilisation. Les méthodes objets semblent être des approches pertinentes pour répondre en partie à cet écueil. L'objectif de la thèse était donc de développer de nouvelles méthodes de post-traitement pour la PE-AROME qui prennent mieux en compte ces structures cohérentes.

Cette thèse a tout d'abord proposé une approche dédiée à une structure météorologique particulière : les échos arqués. Ces structures orageuses accompagnées de fortes rafales de vent se distinguent par une forme en arc dans le champ de réflectivité. La première étape consiste à détecter ces structures dans chaque membre PE-AROME. Le chapitre 4 a montré que les réseaux de neurones convolutifs sont des outils intéressants pour détecter cette forme en arc. Le chapitre 5 a alors pré-

senté comment des réseaux de neurones ont été entraînés à reconnaître des échos arqués. Ces réseaux de neurones s'appuient sur des labellisations d'échos arqués faites manuellement sur des milliers d'images. L'écho arqué étant un évènement rare, plusieurs techniques de ré-équilibrage de la base de données utilisant des hyperparamètres ont été nécessaires pour aboutir à des entraînements satisfaisants.

A partir de scores orientés objets, le chapitre 7 a étudié le rôle de ces hyperparamètres. Il a aussi permis de définir le meilleur réseau de neurones pour détecter les échos arqués qui a un taux de détections de 86% et un taux de fausses alarmes de 39%. Les performances de ce réseau de neurones ont été détaillées avec des approches objets pour extraire les points forts et les points faibles de celui-ci. Cette partie a en particulier montré que les fausses alarmes et les non détections étaient principalement des échos arqués de faible intensité et petits qui sont souvent des cas limites où la distinction entre écho arqué et d'autres types de convection est difficile à faire y compris par un expert. Ce réseau de neurones a ensuite été testé avec succès sur le modèle déterministe AROME. Un groupe d'experts a aussi pu évaluer les performances de ce réseau de neurones de façon subjective. Ce groupe a conclu à des performances similaires à celles obtenues en étudiant les scores objets, validant le lancement d'une production expérimentale à partir de mai 2021. Au préalable, des outils de synthèse pour résumer l'ensemble des détections faites dans PE-AROME et AROME ont été développés conjointement avec les services opérationnels. La capacité des modèles PE-AROME et AROME a été en parallèle évaluée et décrite dans le chapitre 8. A l'aide de premiers travaux sur la détection des échos arqués dans les données radar qui ont servi de référence, les scores montrent de meilleurs performances pour la PE-AROME comparée à AROME aux échéances lointaines et des performances comparables aux courtes échéances. La comparaison de la PE-AROME opérationnelle avec la nouvelle version disponible depuis l'été 2022 a permis de montrer plusieurs changements sur quelques cas mais l'absence de score objectif et le faible nombre de cas étudiés ne permettent pas de conclure de façon robuste à une amélioration ou non.

Ces travaux ont permis de progresser sur l'extraction d'informations pour des phénomènes précis, mais ce mode de synthèse doit être complété par d'autres moins spécialisés. C'est le rôle de la synthèse par scénarios qui a été introduite dans le chapitre 3. La synthèse par scénarios vise à regrouper des membres proposant des prévisions similaires. Cette synthèse a été développée pour les cumuls de pluies sur 1h (RR1h) et à l'échelle régionale (un quart de France). L'approche par scénarios a été pleinement étudiée sur le quart Sud-Est de la France au cours de cette thèse. Cette approche a nécessité de développer un réseau de neurones de type autoencodeur pour projeter les membres au sein de son espace latent et non dans l'espace physique afin de s'affranchir de la définition d'une métrique complexe et soumise fortement à la problématique de la double peine. Puis dans cet espace latent, une classification sur un grand nombre de membres a été faite pour définir des structures météorologiques récurrentes. La synthèse par scénarios peut maintenant être utilisée dans le cadre d'une prévision PE-AROME pour un jour donné. Le regroupement des membres se fait après projection dans l'espace latent en assignant chaque membre à l'une de ces structures météorologiques récurrentes. Les membres associés à la même structure météorologique forment alors un scénario.

La première étape pour mettre en place cette méthode consistait à construire

des autoencodeurs, ce qui était l'objet du chapitre 6. Plusieurs tailles d'espace latent ont été testées, trois autoencodeurs ont été conservés avec des tailles de 10, 20 et 40 dimensions. Ce chapitre a montré également que l'autoencodeur compresse l'information d'une façon plus efficace que des méthodes plus classiques comme l'ACP ou les ondelettes. Ces espaces latents ont été étudiés en profondeur dans le chapitre 9. On a montré en particulier que les espaces latents n'ont pas une réponse linéaire à la taille des objets en entrée avec un filtrage complet des objets les plus petits, ce filtrage étant plus prononcé avec les petits espaces latents. On a montré aussi que les dimensions s'organisent avec des parties positives et négatives qui codent des informations spatiales complémentaires. L'amplitude des valeurs dans la dimension d'un espace latent est liée à l'intensité des pluies en entrée. Ce comportement se rapproche sur certains aspects de celui d'une ACP mais en permettant des non linéarités qui sont bénéfiques pour compresser l'information. Après avoir étudié ces espaces latents, les structures climatologiques récurrentes ont été construites dans le chapitre 10. Cette étape nécessitait de s'interroger sur la bonne méthode de classification pour le faire. Il a été choisi de faire cette classification avec la méthode des K-Means. Les résultats de ces classifications ont également abouti à choisir l'espace latent à 40 dimensions avec un nombre total de structures météorologiques récurrentes égal à 20. Après construction de l'autoencodeur et des structures météorologiques récurrentes, le chapitre 11 s'est intéressé à l'utilisation de la synthèse par scénarios sur des études de cas. Ces études de cas ont montré que cette synthèse peut permettre de repérer des situations peu prévisibles où la prise de décision sera difficile. Ce chapitre 11 a aussi montré que la synthèse par scénarios peut dégager quelques scénarios bien distincts et permet de visualiser comment ils évoluent lors de l'actualisation des modèles. Des scores sur un an de synthèse par scénarios ont aussi été présentés. Ils ont montré les apports conjoints de la PE-AROME et de AROME. La PE-AROME permet d'explorer plus efficacement l'espace des possibles qu'AROME. Mais on a montré aussi qu'à taille de scénario comparable, un scénario qui comprend la prévision d'un modèle déterministe AROME a plus de chance de s'avérer correct qu'un autre. Cette influence du déterministe tend à s'effacer lorsque les échéances sont plus longues. Ces scores ont révélé aussi qu'il ne faut pas toujours faire confiance au scénario majoritaire, qu'il est nécessaire de s'intéresser à des scénarios secondaires ou tertiaire ce qui rend le choix automatique d'un scénario impossible à réaliser. Ces scores ont permis de comparer l'ancienne et la nouvelle version des modèles PE-AROME et AROME depuis l'été 2022. Cette comparaison a montré de réelles améliorations avec ces nouvelles versions sous le prisme de la synthèse par scénarios.

Plusieurs perspectives s'ouvrent pour poursuivre ces travaux. Concernant la détection des échos arqués, il est important de continuer à travailler conjointement avec les services opérationnels pour améliorer encore cet outil et l'analyser sur des temps longs pour apprendre à en tirer son plein potentiel. Des améliorations pourraient également être faites concernant le réseau de neurones. On peut penser à l'ajout d'autres prédicteurs ou au changement des méthodes de pré-traitement pour tendre vers des méthodes similaires à l'importance sampling qui a été présenté dans le chapitre 6. La poursuite des travaux sur la détection des échos arqués dans les données observées est aussi un point important pour évaluer les modèles plus facilement à l'avenir. Une dernière perspective porte sur le développement

d'outils similaires pour d'autres types d'objets. On pourrait s'intéresser aux supercellules qui sont responsables de fortes rafales de vent, de chutes de grêle et parfois de tornades. Être capable de détecter les échos arqués et les supercellules permettraient de couvrir une part importante de la convection sévère en France métropolitaine [3].

Concernant la synthèse par scénarios, d'importants échanges sont nécessaires avec les services opérationnels pour expliquer ce nouveau mode de représentation et discuter d'éventuelles modifications. On peut penser par exemple à de possibles ajustements concernant les structures météorologiques récurrentes ou à l'ajout d'un critère de sévérité aux scénarios. Cette synthèse devra également être développée sur les autres quarts de France (Nord-Ouest, Nord-Est, Sud-Ouest et Centre) avec un nombre de structures météorologiques récurrentes qui sera probablement différent. Cette synthèse pourra également être étendue à d'autres variables météorologiques et en multivarié. On peut penser par exemple à des scénarios mixant les cumuls de pluies et la force du vent. Enfin cette approche a été développée ici pour une utilisation par des prévisionnistes de Météo-France mais elle pourrait également convenir à d'autres usagers des prévisions avec une sensibilité un peu différente aux conditions météorologiques. Des travaux en collaboration avec la Direction des Services Météorologiques (DSM) vont ainsi débuter pour développer des synthèses par scénarios qui dépendront des aléas météorologiques mais également de la vulnérabilité de l'usager à ces aléas.

Les perspectives pour la suite de cette thèse s'appuient encore très largement sur le domaine de l'intelligence artificielle et particulièrement du deep learning. Cette thèse a montré que expertise humaine et deep learning peuvent être complémentaires. Loin de remettre en question l'expertise humaine des modèles, les algorithmes de deep learning entraînés à partir de données expertisées par un humain peuvent aider les prévisionnistes à mieux appréhender la masse de données à leur disposition pour établir leur prévision.

3. Il ne faut pas oublier les cellules en V qui peuvent provoquer de forts cumuls de pluies.

Table des figures

1.1 Présentation des différentes couches de l'atmosphère de la troposphère à l'exosphère et profil vertical de température (Source : Météo-France). 8
1.2 Mosaïque de réflectivités SERVAL pour la France métropolitaine. Image du 15 mai 2022 à 22h UTC. Plus les couleurs sont chaudes (vert/jaune), plus les dBZ sont importants et les intensités fortes. L'absence de couleur indique l'absence de précipitation (Source : Météo-France, outil Synopsis). 10
1.3 Cumul de précipitations (produit ANTILOPE) combinant les informations issues des mesures des pluviomètres et des données radars. Cumul de précipitation du 15 mai 2022 de 21h à 22h UTC (Source : Météo-France, outil Synopsis). 11
1.4 Représentation schématique des différentes échelles météorologiques avec un exemple pour chacune des échelles : l'échelle planétaire (circulation générale), l'échelle synoptique (anticyclone, représenté à l'aide d'isobares en P_{mer}), la mésoéchelle (ondes de relief), l'échelle aérologique (cellule orageuse) et la microéchelle (rafale de vent). Inspiré de Malardel (2009). 12
1.5 Représentation schématique de la chaîne de prévision. Le chapitre 1 s'intéresse à la partie en vert à travers : l'observation du temps présent à l'aide de différents instruments et l'utilisation d'anciennes prévisions (à gauche), la prise en compte de ces données par le modèle à l'aide de l'assimilation de données pour créer un état initial (au centre) et enfin la prévision des états futurs à l'aide du modèle (à droite). 14
1.6 Évolution du nombre d'observations utilisées par le modèle français ARPEGE entre 2002 et 2022 par type de données (Source : équipe COMPAS, Météo-France). 15
1.7 Représentation schématique de la construction de l'état analysé (rouge) à partir de l'ébauche (bleu) et des observations (orange) pour une assimilation 4D-Var. 17
1.8 Attracteur de Lorenz (Source : `https://commons.wikimedia.org/w/index.php?curid=3700799`) 20
1.9 Qualité de prévision du géopotentiel à 500 hPa entre les années 1980 et 2010 à 3 jours (bleu), 5 jours (rouge), 7 jours (jaune) et 10 jours (gris). La courbe en trait gras représente les prévisions sur l'héminsphère Nord, en trait plus fin sur l'hémisphère Sud. D'après Bauer et al. (2015). 21

1.10 Représentation schématique d'une prévision d'ensemble (schéma de Carole LABADIE, Météo-France). Chaque trajectoire noire est un membre de la prévision d'ensemble, la courbe verte la prévision déterministe et en rouge l'état observé. 22

1.11 Représentation schématique de la construction des perturbations de l'état initial (fléches vertes) en partant d'une perturbation aléatoire (flèche orange) à l'aide d'un breeding 23

1.12 Champ de réflectivités simulées en dBz du modèle AROME du 3 juin 2022 à 12h UTC pour le 4 juin 2022 à 18h UTC. Plus les couleurs sont chaudes, plus les réflectivités sont intenses (Source : Météo-France, outil Synopsis). 25

1.13 Principales évolutions du système ensembliste PE-AROME depuis sa mise en opérationnel à l'automne 2016 jusqu'à l'été 2022. Cette liste n'est pas exhaustive et n'intègre pas en particulier le changement de méthode de classification pour choisir les coupleurs PEARP ou l'assimilation de nouvelles données pour l'AEARO. 28

1.14 Domaine couvert par la PE-AROME actuelle (domaine EURW1S40). Ce domaine est le même pour AROME. Le domaine s'étend en latitude de 37.5°N à 55,4°N et de 12° W à 16°E pour une grille de sortie de 717 par 1121 points de grille à 2,5km de résolution. 29

2.1 Représentation sous forme de vignettes des 16 membres de la PE-AROME à un instant donné. Les membres sont classés de haut en bas et de gauche à droite (membre 1 en haut à gauche et membre 16 en bas à droite). Les vignettes représentent la prévision de la température de brillance qui permet d'estimer la hauteur des sommets convectifs et donc l'intensité de la convection. Plus la température est froide, plus la couleur tend vers le rouge et plus le sommet est haut dans la troposphère. Prévision du 14 septembre 2021 à 03h UTC pour 08h UTC sur le Sud-Est de la France (Source : Météo-France, outil Synopsis). 36

2.2 Représentation schématique de la chaîne de prévision. Le chapitre 2 s'intéresse à la partie en vert dédiée aux post-traitements des modèles (représentés ici par deux cartes qui seront détaillées dans ce chapitre). 37

2.3 Représentation sous forme de panaches des prévisions à 15 jours de température minimales (bleu) et maximales (rouge) pour l'aéroport Toulouse-Blagnac (prévision du 8 au 22 juin 2022). Le trait continu plus épais représente la médiane des prévisions. Les enveloppes de teintes de couleurs différentes représentent les prévisions faites par les 20% des membres les plus proches de la médiane pour la couleur la plus foncée jusqu'à 100% des membres pour l'enveloppe la plus claire qui correspond aux prévisions les plus éloignés de la médiane des deux côtés de la distribution. Plus bas, représentation du risque pour chaque jour que la température minimale soit inférieure à 15 degrés et pour terminer du risque que la température maximale dépasse les 30 degrés. Prévision faite à partir de l'ensemble du CEPMMT. (Source : Extranet Météo-France) 38

2.4 Représentation sous forme de boîtes à moustaches des prévisions ensemblistes du modèle européen (CEPMMT) pour l'aéroport de Toulouse-Blagnac du 8 au 22 juin 2022 pour la force du vent (jaune) en $m \cdot s^{-1}$ et les cumuls de pluie sur 24h (vert) en mm. (Source : `previsibilite.meteo.fr`, Météo-France) 39

2.5 Partie supérieure de la synthèse des principales informations utiles en situation convective à partir de l'ensemble du CEPMMT. Exemple de la ville de Clermont-Ferrand (Centre de la France) pour le 4 juin 2022 à 12h UTC. En haut à gauche, représentation sous forme de panaches des différents profils verticaux (en rouge le profil de température selon les niveaux pressions, en vert la température du point de rosé). Les courbes plus épaisses correspondent au modèle déterministe du CEPMMT. La différence de ces deux courbes permet de tracer le panache bleu en haut à droite. Plus la différence entre la courbe rouge et bleu est faible, plus l'humidité et le risque de nuage sont importants. Le vent du modèle déterministe du CEPMMT selon les niveaux pressions est aussi représenté sur ce même graphe (barbules noires). (Source : `https://www.ecmwf.int/en/newsletter/156/meteorology/using-ecmwfs-new-ensemble-vertical-profiles`, CEPMMT) . . 40

2.6 Cartes de probabilités concernant le risque de dépasser les 10 mm en cumul de pluie sur 24h pour la journée du 4 juin 2022 selon le modèle PE-AROME (run du 3 juin 2022 à 03h UTC). La carte de gauche représente les prévisions probabilistes brutes. Concernant la carte de droite, une méthode des voisins spatio-temporelle est appliquée avec un voisinage de 40 km pour la tolérance spatiale et plus ou moins 1h pour la tolérance temporelle (Source : Météo-France, outil Synopsis). 42

2.7 Cartes de quantiles concernant les cumuls de pluie sur 24h (en mm) pour la journée du 4 juin 2022 selon le modèle PE-AROME (run du 3 juin 2022 à 03h UTC). Ces cartes sont zoomées sur le Centre et le Nord de la France. La carte de gauche représente le Q75 et la carte de droite le Q_{max}. Les flèches rouges indiquent des extremas locaux en Normandie dépassant les 70 mm (Source : Météo-France, outil Synopsis). 44

2.8 Représentation de la moyenne (gauche) et de l'écart-type (droite) de la température à 850 hPa de l'EPS pour le 17 juin 2022 à 0h UTC (run du 10 juin 2022 à 0h UTC). Les couleurs chaudes sur la moyenne indiquent des températures élevées et les couleurs sombres sur la carte de droite des écart-types importants (Source : Météo-France, outil Synopsis). 44

2.9 Représentation des prévisions d'ensemble du modèle américain GEFS sous forme de spaghettis. Les contours du géopotentiel à 500 hPa à deux seuils (534 et 576 dam) sont tracés avec une couleur pour chaque membre. Prévision faite le 9 juin 2022 à 0h UTC pour le 14 juin 2022 à 12h UTC (Source : `https://mag.ncep.noaa.gov/model-guidance-model-parameter.php?group=Model%20Guidance&model=GEFS-SPAG&area=EUROPE&ps=model`, NOAA) 45

2.10 Extrait des prévisions mensuelles du modèle du CNRM pour la température à 2 mètres. Prévision du 9 juin 2022 pour les semaines du 13 au 19 juin (S1) et du 20 au 26 juin 2022 (S2). Ces prévisions sont exprimées à l'aide d'anomalies de températures (gauche), de probabilités d'observer une anomalie positive (centre) et de terciles (droite, Source : `http://intra.cnrm.meteo.fr/previsionS2S/page.html`, site interne de Météo-France) 46

2.11 Représentation du champ de vent moyen en haute troposphère (250 hPa) à l'aide de l'outil Met3D sur l'Atlantique et l'Europe. Les formes marrons/jaunes transparentes correspondent aux zones où la moyenne de l'ensemble du CEPMMT dépasse les $30m \cdot s^{-1}$. L'écart-type est superposé avec des couleurs jaunes à rouges. (Source : `https://met3d.wavestoweather.de/gallery.html`, © Met.3D contributors 2017. CC Attribution 4.0 International License. met3d.wavestoweather.de) 47

2.12 Exemple de « paintball plot ». Représentation des zones de réflectivités simulées dépassant le seuil de 40 dBz avec une couleur pour chaque modèle. Prévision du 11 juin 2022 à 12h UTC pour le 13 juin à 0h UTC (Source : `https://www.spc.noaa.gov/exper/href/?model=href&product=cref_pb40&sector=conus`, Storm Prediction Center, NOAA). 48

2.13 La carte de gauche représente les indices EFI (plage de couleurs) et SOT (contour noir) de la température maximale du 17 juin 2022 (prévision de l'EPS du 13 juin 2022). La carte de droite représente le quantile 99 du climat du modèle pour ce paramètre. (Source : `https://apps.ecmwf.int/webapps/opencharts/products/efi2web_2tmax?area=Europe&base_time=202206140000&day=1&quantile=99`, CEPMMT) 49

2.14 Partitionnement des membres de l'EPS en différents scénarios avec le champ de géopotentiel à 500 hPa. Les lignes correspondent à des scénarios différents et les colonnes correspondent à des jours différents (prévision à 8,9 et 10 jours faite le 8 juin 2022). Chacun de ces scénarios est représenté par un membre. En plus du champ de géopotentiel en noir, les anomalies par rapport à la climatologie sont représentées en couleur. Les cadres de différentes couleurs correspondent à l'un des 4 régimes de temps connus en météorologie sur l'Europe et l'Atlantique Nord. Le membre du scénario 2 à 10 jours encadré en violet appartient par exemple au régime de temps « dorsale atlantique » (Source : `https://apps.ecmwf.int/webapps/opencharts/products/cluster_plot_legA?base_time=202206130000&cluster=192_240&parameter=500`, CEPMMT). 51

2.15 Représentation des spaghettis (en bas à gauche) sous forme de quelques classes représentés par leur contour et leur zone d'incertitude. D'après Ferstl et al. (2016b). 52

2.16 Mesure de la proximité de chaque membre de l'ensemble du CEPMMT aux 4 régimes de temps. Cette mesure est faite avec la distance de corrélation spatiale. La courbe plus foncée est le membre de contrôle de l'ensemble, la courbe rouge la moyenne des membres (Source : `previsibilite.meteo.fr`, site interne de Météo-France). 53

2.17 Représentation des régimes de temps assignés aux membres de l'ensemble européen de façon quotidienne. Chaque couleur correspond à un régime et la dernière couleur en gris correspond à des membres qui n'ont pas pu être attribués car trop éloignés des 4 régimes de temps. Le carré hachuré en bas du diagramme correspond à l'attribution de la moyenne de l'ensemble à l'une des 5 catégories. Prévision du 9 juin jusqu'au 23 juin 2022 (Source : `https://apps.ecmwf.int/webapps/opencharts/products/extended-regime-probabilities?forecast_from=latest`, CEPMMT). 54

2.18 Répartition des membres de l'ensemble européen à travers 8 patterns climatologiques pré-définis en P_{mer} sur l'Europe et l'Atlantique Nord. Prévision du 1er Janvier 2015. Le cercle noir représente la position attribuée à l'analyse a posteriori. D'après Neal et al. (2016). 54

2.19 Prévision de la trajectoire du cyclone EMNATI par l'EPS du 18 février 2022 jusqu'à 10 jours. La courbe noire en trait plein représente la trajectoire prévu par le modèle déterministe, la courbe en tiret la trajectoire moyenne de l'ensemble et le panache les trajectoires possibles avec une probabilité associée. Une faible probabilité signifie que peu de membres proposent une trajectoire dans la zone considérée (Source : `https://www.ecmwf.int/en/forecasts/charts/tcyclone/tc_strike_probability?facets=undefined&time=2022021800,0,2022021800&unique_id=05S_EMNATI_2022`, CEPMMT). 56

2.20 Détections de fronts chaud (rouge) et froids (bleu) avec l'algorithme de Hewson and Titley (2010) dans chaque membre de la prévision d'ensemble du CEPMMT. Prévision du 10 juin 2021 à 0h UTC pour le 14 juin 2021 à 12h UTC (Source : `https://www.ecmwf.int/en/about/media-centre/focus/2021/ecmwf-synoptic-style-weather-charts-set-tools-help-forecast-adverse`, CEPMMT) 56

2.21 Évolution de l'indice MJO selon les différents membres de l'ensemble européen. Les points à l'intérieur du rond central indiquent une MJO peu active. L'extérieur de ce cercle indique une MJO active classée selon 8 phases (chiffres bleus). Les points sont colorés selon les échéances de 1 jour (rouge) à 20 jours (vert). La courbe noire et les triangles représentent la moyenne de l'ensemble, la courbe grise les valeurs de l'indice sur les 30 derniers jours à partir des analyses. Prévision du 19 mai 2022 de l'EPS (Source : `https://www.ecmwf.int/en/forecasts/charts/catalogue/mofc_multi_mjo_family_index?facets=Range,Extended%20(42%20days)&time=2022061300,0,2022061300`, CEPMMT). 57

3.1 Représentation schématique de la chaîne de prévision. Cette partie s'intéresse à l'expertise faite par les prévisionnistes et les prises de décisions qui en découlent, représentées par une carte de vigilance et une image du Centre National de Prévision (CNP). 62

3.2 Schémas conceptuels du cycle de vie (de I à IV) d'une dépression aux moyennes latitudes selon le modèle conceptuel norvégien à gauche (Bjerknes and Solberg, 1922) et le modèle conceptuel de Shapiro and Keyser (1990) à droite. Les contours correspondent aux isobares du champ de P_{mer} dont le centre dépressionnaire est représenté par un « L ». Les fronts sont représentés par les objets rouges, bleus et violets dont la signification est détaillée en annexe figure A.1. (Source : `http://tempetes.meteo.fr/spip.php?article231`) . . . 63

3.3 Analyse technique de la situation météorologique du 11 décembre 2017 à 00h UTC sur l'Europe et l'Atlantique Nord à travers l'ANAlyses SYnoptiques Graphiques (ANASYG) produite par le CNP. La signification des différents symboles est disponible en annexe (figure A.1). 64

3.4 Représentation graphique « grand public » de la situation météorologique prévue le 26 août 2022 à minuit. Prévision faite le 21 août 2022 (Source : `previ-gene.meteo.fr`, site interne de Météo-France) 65

3.5 Représentation du balayage d'une méthode objet en calculant la distribution du voisinage de plusieurs points de grille, espacés de façon régulière. Les voisinages sont représentés par les cercles noirs. (Source : `http://intra.cnrm.meteo.fr/recyf/Appl/Objets//aide.php`, site interne de Météo-France) 66

3.6 Représentation du cyclone Maria avec le cumul de pluie sur 6h. Les objets de pluies modérées sont tracés avec les contours orange à gauche et les objets de pluies fortes à droite avec des contours rouges. Ces objets sont tracés avec la méthode décrite dans Rottner et al. (2019) (Source : `http://intra.cnrm.meteo.fr/recyf/Appl/Objets//aide.php`, site interne Météo-France) 67

3.7 Illustration de la double peine. Les champs idéalisés sont de la pluie avec les intensités les plus fortes en rouge. Les deux cartes représentent des champs de pluies de modèle à haute et basse résolution qui sont transparents pour les différencier des champs de l'observation complètement opaques. 68

3.8 Cycle de vie d'un écho arqué d'après Fujita (1978) 69

3.9 Les différents stades idéalisés pour former un écho arqué d'après Markowski and Richardson (2010). Les stades de (a) à (d) sont détaillés dans la section 3.3.2. 70

3.10 Coupe verticale faite dans un écho arqué simulé par AROME. L'axe vertical représente les niveaux pressions du modèle en hPa et l'axe horizontal différentes positions qui traversent l'écho arqué (l'avant de l'écho arqué est à droite des graphiques). La carte de gauche superpose l'humidité (elle dépasse 90% dans la zone fuchsia), les vitesses verticales par les isocontours bleus et verts (en vert les vents ascendants et en bleu les vents subsidents) et le vent moyen tangent à la coupe verticale par les barbules rouges (un triangle rouge indique des vents à plus de 90 $km \cdot h^{-1}$). La carte de droite ne représente plus l'humidité pour mieux mettre en évidence la circulation des vents. 71

3.11 Schéma récapitulant la méthodologie développée pour prendre en compte les échos arqués simulés dans la PE-AROME. Les échos arqués sont d'abord détectés dans chaque membre par un réseau de neurones de façon automatique à l'aide du champ de réflectivités, ces détections sont entourées en noir sur les 3 membres de la deuxième ligne. L'ensemble des détections dans les 16 membres sont superposées sur une carte de synthèse. En plus des détections dans PE-AROME, les détections sont faites dans AROME et sont entourées en noir sur la carte de synthèse pour qu'elles se différencient de celles issues de PE-AROME. 73

3.12 Représentation schématique des scénarios. Chaque membre est représenté par une courbe de couleur différente au cours du temps. . . 74

3.13 Découpage du domaine AROME en 5 sous-zones régionales : Nord-Ouest (NO) en vert, Nord-Est (NE) en jaune, Sud-Ouest (SO) en bleu, Sud-Est (SE) en rouge et Centre (C) en rose. 75

3.14 Représentation schématique de la méthode retenue pour construire les scénarios de pluies (exemple pris pour la zone Sud-Est). L'ensemble des valeurs représentées dans l'espace latent et les structures météorologiques récurrentes sont fictives et ne correspondent pas à la réalité. Les différentes étapes sont expliquées dans la section 3.4.2. 77

4.1 Comparaison des objectifs d'un algorithme classique et d'un algorithme de machine learning. Les données en entrée des algorithmes sont en vert, les données en sortie en rouge. 84

4.2 Représentation de l'information à différentes couches d'un algorithme de DL pour la reconnaissance de visage. D'après Lee et al. (2009). 85

4.3 Phase d'entraînement d'un réseau de neurones à partir des données X, Y et d'une fonction coût L (en vert). Les sorties du réseau de neurones $\hat{Y}$ sont en rouge. 86

4.4 Représentation schématique de la descente de gradient vers le minimum de la fonction coût L_{min} par 3 méthodes différentes : par batch (BGD en rouge), minibatch (MBGD en vert) ou stochastique (SGD en violet). Chaque flèche représente une itération. 88

4.5 Représentation schématique de la fonction coût selon les epochs pour la courbe d'apprentissage (bleu) et la courbe de validation (orange). 89

4.6 Représentation schématique d'une couche dense entre les données d'entrée de la couche X et les données de sortie de la couche Y. . . 90

4.7 Représentation de la fonction ReLU 91

4.8 Représentation schématique d'une couche de convolution pour une entrée X de taille 4×4 avec un filtre C de taille 2×2 et un stride de 2. Le terme de biais est ici un scalaire égal à -2 (qui s'applique sur chaque point de grille) et la fonction d'activation σ est la fonction ReLU. 92

4.9 Représentation schématique d'un 0-padding avec ajout d'une colonne et d'une ligne sur les bords (cas $p = 1$). 92

4.10 Représentation schématique de la couche de maxpooling pour une taille 2×2 et un stride de 2 (haut). Représentation schématique de la couche de upsampling (bas). 93

4.11 Représentation schématique de la couche de déconvolution. Exemple avec un stride égal à 1 (haut) et un exemple avec un stride égal à 2 (bas). 94

5.1 Exemple de champs utilisés lors du processus de labellisation. Les champs sont zoomés sur le Nord-Est de la France. L'image de gauche correspond aux contours de $\|\nabla(MSLP)\|_1$ égal à 2 hPa pour 10 km sont en rose, les zones où la divergence du vent horizontal (div dans la légende) est inférieure à $1.5 \times 10^{-3}\, s^{-1}$ sont en vert, et le contour de l'écho arqué labellisé est en noir. (droite) Le champ de pseudo-réflectivité correspondant. 106

5.2 Architecture du U-Net 109

5.3 Construction de la base d'apprentissage. Les étapes 1 à 6 (cercles bleus) sont décrites dans la section 5.3.1. Les hyper-paramètres discutés dans le chapitre 7 apparaissent en rouge. 110

5.4 Description schématique du processus de prediction complet en partant du champ de pseudo-réflectivité complet (ENTRÉE) jusqu'à la prédiction d'un champ binaire avec un 1 pour les points de grille dans un écho arqué et 0 sinon (SORTIE). Les étapes 1 à 4 (cercles rouges) sont décrites dans la section 5.4. 112

5.5 Attributs des objets. Un écho arqué est caractérisé sur cette figure par 3 attributs : intensité, position et taille. L'intensité de l'écho arqué est extraite avec les quantiles 25 et 90 de pseudo-reflectivités à l'intérieur de l'objet (histogramme), la position est extraite avec la position du centre de masse et la taille avec le nombre de points de grille qui constitue l'objet. 113

6.1 Construction du jeu de données. Le jeu de données est constitué de champs de RR1h sur les 5 zones régionales (même si seulement les zones Nord-Ouest et Sud-Est sont représentées ici). Un jeu de données de plus de $N = 5$ millions d'échantillons est constitué au départ (gauche). Une première sélection est faite pour supprimer les échantillons dont le Q99 n'atteint pas les 0.4mm permettant de diviser environ par deux le jeu de données (centre, N_{int}). Enfin une méthode d'importance sampling est appliquée sur ces données pour ne sélectionner que environ 700 000 échantillons (N_f). Les échantillons restants sont symboliquement encadrés en vert. 119

6.2 Illustration de l'influence de l'importance sampling sur la variance. Variance du jeu de données avec 2.5 millions d'échantillons à gauche (étape intermédiaire de la figure 6.1). Variance après importance sampling pour les hyperparamètres suivants : P_{min}=0.1, m=1.5 et s_{rr}=1mm à droite. L'échelle de couleur est identique pour les deux cartes. 121

6.3 Architecture de l'autoencodeur convolutif. L'autoencodeur utilise en entrée une image de RR1h (en haut à gauche). L'information est compressée avec la partie encodeur (vert) pour aboutir à un espace latent. Puis l'information est décompressée à l'aide de la partie décodeur (orange) pour revenir à une sortie qui a la même taille que l'image d'entrée (en haut à droite). Les flèches en pointillées désignent des étapes de réorganisation des données. Les hyperparamètres du réseau qui seront discutés dans la section 6.4 apparaissent en rouge. L'architecture ici présentée correspond à une suite de filtres 16,32,64,128 de tailles 16,8,3,3 avec un espace latent à 40 dimensions. Les autres architectures testées sont présentées tableau 6.2. 122

6.4 Méthode de détection des objets illustrée avec les objets de pluies totales. Exemple sur les mêmes images que pour la figure 6.3. Les entrée et sortie de l'autoencodeur sont lissées avec un filtre gaussien (ici d'écart-type égal à 15 points de grille, centre). Puis un seuillage à 0.2 mm est appliqué pour définir les objets de pluies totales (contours jaunes, droite). 125

6.5 Exemple de fausse alarme pour un objet de pluies fortes. Le champ de gauche correspond à l'entrée de l'autoencodeur et à droite le champ reconstruit. Les contours jaunes représentent les objets de pluies totales, les contours oranges les objets de pluies modérées et le contour rouge sur le champ de droite un objet de pluies fortes. . . 128

6.6 Représentation de l'ondelette mère coiflet à deux moments nuls en 1D (Source : `http://wavelets.pybytes.com/wavelet/coif2/`). . 135

6.7 Comparaison des champs reconstruits après compression de l'information dans 40 dimensions à l 'aide d'un autoencodeur, d'une approche par ondelettes ou d'une ACP. L'ondelette mère utilisée ici est coiflet 2. 135

8.1 Mosaïque SERVAL du 17 juin 2021 à 15h UTC zoomée sur la région Occitanie. Les contours noirs représentent les détections de l'écho arqué par 4 expériences différentes dont les configurations sont précisées dans la section 8.1. Ces images sont extraites du rapport de stage de Sylvain Galliau. 185

8.2 (première ligne) Représentation schématique de la construction des cartes de probabilités du risque d'échos arqués. L'étape 1 consiste à détecter les différents échos arqués dans les membres PE-AROME (une couleur par membre). L'étape 2 consiste pour chaque objet à calculer son voisinage (contour rouge). Les voisinages de chaque objet sont alors superposés pour définir la carte de probabilité à l'étape 3 construite avec une tolérance spatiale (en plus d'une tolérance temporelle de plus ou moins 1h). Plus de détails dans la section 6 de l'article qui constitue la chapitre précédent. (deuxième ligne) Représentation schématique de la vérification des performances de PE-AROME. A gauche, si on prend le seuil de 10% (zone bleu foncé), l'écho arqué intersecte cette zone et est donc comptabilisée comme une bonne détection. Sur l'image de droite, il n'intersecte pas cette zone et est donc considéré comme non détecté. 186

8.3 Scores pour les échéances courtes (inférieures ou égales à +15h). Les courbes noires, bleues et vertes représentent les seuils de probabilités à 5%, 10% et 20% pour la PE-AROME. La courbe rouge représente les performances d'AROME. Le graphique en haut à gauche correspond aux TD, le graphique en haut à droite aux TFA et le graphique en bas aux indices de Jaccard. Les courbes sont tracées en fonction de la tolérance spatiale ϵ. 188

8.4 Scores pour les échéances longues (supérieures à +15h). Figure similaire à la figure 8.3. 189

8.5 Répartition des bonnes détections (bleu), fausses alarmes (orange) et non détections (gris) selon les 15 dates dans la base de données en prenant un seuil de probabilité à 10% et une tolérance spatiale ϵ de 150 km. 190

8.6 Exemple du nouveau graphique « rafale de vent » pour la journée du 20 juin 2021. Prévision faite par la PE-AROME du 19 juin 2021 à 15h UTC. Les objets entourés en noir correspondent au modèle déterministe AROME. Le code couleur indique les rafales de vent sous les échos arqués en $km \cdot h^{-1}$. 192

8.7 Situation météorologique du 17 juin 2021. La carte du haut représente l'ANASYG produit par le CNP pour le 17 juin 2021 à 18h UTC. La carte au centre représente la carte de vigilance émise par Météo-France le 17 juin 2021 à 14h et la carte du bas la mosaïque radar sur le Sud-Ouest de la France du 17 juin 2021 à 17h (Source : Infoclimat). 194

8.8 Prévision du risque d'échos arqués pour la journée du 17 juin 2021. Prévision faite à partir du réseau PE-AROME du 16 juin à 21h UTC. La carte de gauche représente le risque sous une forme probabiliste, au centre les échos arqués avec un code couleur selon l'heure de détection et à droite le nouveau graphique avec un code couleur selon les rafales de vent. 195

8.9 Comparaison des prévisions faites par les anciennes et les nouvelles versions des modèles pour la journée du 19 juin 2021 avec le réseau du 18 juin 2021 à 09h UTC. La première ligne représente les cartes de probabilités, la deuxième ligne les détections avec un code couleur selon l'heure de la journée et la dernière ligne le nouveau graphique avec un code couleur selon les rafales de vent. 196

8.10 Figure similaire à la figure 8.9. Comparaison des prévisions faites par les anciennes et les nouvelles versions des modèles pour la journée du 20 juin 2021 avec le réseau du 20 juin 2021 à 03h UTC. . . . 197

9.1 Scores orientés objets pour comparer les objets en entrée et en sortie de l'autoencodeur. L'autoencodeur présenté est celui possédant un espace latent à 40 dimensions. Chaque colonne correspond à un type d'objet (pluie totale, modérée ou forte). La première ligne compare l'aire des objets en entrée et en sortie exprimée en nombre de points de grille, une droite de régression en noir est tracée. L'équation de la droite de régression, le coefficient de détermination (R^2) et la RMSE par rapport à la droite $y = x$ (en gris tireté) sont précisés dans la légende. La deuxième ligne correspond aux mêmes graphiques mais pour le Q90 des RR1h des objets exprimé en mm. La dernière ligne compare les distributions du nombre d'objets détectés sur un champ de RR1h pour les données en entrée (objets PE, vert) et en sortie (objets AE, bleu) de l'autoencodeur. 203

9.2 Scores orientés objets pour étudier les non détections de l'autoencodeur avec un espace latent à 40 dimensions. Chaque colonne correspond à un type d'objet (pluie totale, modérée ou forte). La première ligne représente l'ensemble des objets en entrée de l'autoencodeur selon leur aire en nombre de points de grille et selon leur intensité par l'intermédiaire du Q90 en mm. Les objets non détectés en sortie de l'autoencodeur sont ensuite représentés en bleu. La deuxième ligne représente la distribution des aires pour l'ensemble des objets et celle des non détections en bleu. La troisième ligne correspond aux mêmes histogrammes mais pour le Q90 des objets. 205

9.3 Scores orientés objets pour étudier les fausses alarmes de l'autoencodeur avec un espace latent à 40 dimensions. Les graphiques sont les mêmes que ceux de la figure 9.2 mais en superposant l'ensemble des objets reconstruits par l'autoencodeurs et ceux qui sont des fausses alarmes en bleu. 207

9.4 Modification de la taille de la structure, impact sur l'espace latent à 10 dimensions. Le maximum est fixé à 10 mm, σ varie de 2.5 km à 100 km avec un pas de 2.5 km. Les images en haut représentent le premier champ (σ=2.5 km) et le dernier (σ=100 km). Les évolutions dimension par dimension entre ces deux champs de l'espace latent sont représentées en dessous ainsi que l'évolution des normes 1,2,3 et infini du vecteur dans l'espace latent. 209

9.5 Modification de l'intensité de la structure, impact sur l'espace latent à 20 dimensions. Le σ est fixé à 20 points de grille, M varie de 1 mm à 50 mm avec un pas de 1 mm. Les images en haut représentent le premier champ (M= 1 mm) et le dernier (M= 50 mm). Les évolutions dimension par dimension entre ces deux champs de l'espace latent sont représentées en dessous ainsi que l'évolution des normes 1,2,3 et infini du vecteur dans l'espace latent. 210

9.6 Modification de la position de la structure, impact sur l'espace latent à 40 dimensions (partie 1, dimension 1 à 20). Le σ est fixé à 10 points de grille et M à 10 mm. Les positions varient schématiquement comme sur la flèche verte de l'image de droite avec un pas de 10 points de grille que ce soit en X ou en Y. Les images en haut représentent le premier champ (bas à gauche) et le dernier (haut à droite). Les évolutions dimension par dimension entre ces deux champs de l'espace latent sont représentées en dessous ainsi que l'évolution des normes 1,2,3 et infini du vecteur dans l'espace latent. 211

9.7 Modification de la position de la structure, impact sur l'espace latent à 40 dimensions (partie 2, dimension 21 à 40), similaire à la figure 9.6. 212

9.8 Construction de l'étude dimension par dimension en quatre étapes. Exemple pour la dimension n°1. 214

9.9 Caractéristiques de la dimension 12 de l'espace latent à 40 dimensions à partir de la distribution calculée comme à l'étape 3 de la figure 9.8. La signification de chaque graphique de cette figure est détaillée dans la section 9.3. La première ligne correspond de gauche à droite à la répartition des valeurs prises par cette dimension selon les échéances (9.3.1), selon les zones régionales (9.3.2) et selon les mois (9.3.3). La deuxième ligne s'intéresse de nouveau à la distribution complète en se focalisant sur le champ moyen des images comprises dans le Q1 à gauche et le Q99 à droite (9.3.4). La dernière ligne s'intéresse aux images correspondants au minimum de la distribution à gauche et au maximum à droite (9.3.5). 216

9.10 Distribution mensuelle sur l'année 2021 des pluies pour les 5 zones régionales dans les prévisions PE-AROME. Ces cumuls de pluies sont moyennés sur les membres et les points de grille. 217

10.1 Représentation schématique de l'algorithme des K-Means. Un jeu de données est représenté par des croix à l'étape 1. L'objectif est de créer deux classes (bleu et orange), deux points qui représentent les futurs centroïdes des deux classes sont créés à l'étape 2. L'assignation des croix à l'une des deux classes est représentée par leur couleur. La position des centroïdes changent à l'étape 3 avec la position ancienne qui apparaît sous forme d'un rond hachuré. Les étapes 2 et 3 sont répétées jusqu'à la classification finale étape 9. . . 223

10.2 Représentation schématique d'autres méthodes de classification que la famille des K-Means. Les méthodes hiérarchiques sont représentées par un dendrogramme sur la figure de gauche, chaque classe est représentée par une couleur (Source : Minitab). Les méthodes à densité sont représentées par plusieurs points avec leurs voisinages représentés sous la forme de cercles. Les points au cœur de la classe sont en rouge, les points à la frontière de la classe sont en jaune et le point N est un point isolé qui n'appartient pas à la classe (Source : `https://commons.wikimedia.org/wiki/File:DBSCAN-Illustration.svg`). Les méthodes basées sur un mélange gaussien, où chaque classe est représentée par une gaussienne dont les paramètres sont à déterminer, sont représentées sur la figure de droite. On suppose dans cet exemple l'existence de deux classes jaune et rouge, la dernière itération est ici représentée avec les densités de chaque gaussienne en arrière plan (Source : `https://commons.wikimedia.org/wiki/File:EM_Clustering_of_Old_Faithful_data.gif`). 225

10.3 Représentation synthétique de la classe n°20 obtenue à l'aide d'un K-Means avec l'espace latent à 40 dimensions pour un nombre total de classes égal à 20. Chaque graphique ou carte est expliqué dans la section 10.3. 229

10.4 Évolution de l'AMI (ordonnée) selon le nombre de classes (abscisse) et selon les tailles d'espace latent (couleurs différentes). Les panaches représentent l'AMI minimale et maximale en comparant les différentes classifications disponibles (3 tentatives à chaque fois pour un nombre de dimensions et un nombre de classes fixés). . . . 232

10.5 Table de contingence pour un nombre total de 3 classes. Une classification K-Means a été faite sur l'espace à 10 dimensions et une autre sur l'espace à 20 dimensions. On s'intéresse à la répartition des individus parmi ces 3 classes. On peut lire par exemple dans ce tableau que seulement 11 individus sont dans la classe 3 pour l'espace à 20 dimensions et dans la classe 2 pour l'espace à 10 dimensions. 234

10.6 Cartes d'écart-type pour la classe majoritaire selon la taille de l'espace latent (colonne de gauche à droite avec 10,20 ou 40 dimensions) et selon le nombre de classes (ligne de haut en bas avec 10,20 ou 30 classes). 235

10.7 Scores pour établir le nombre de classes avec l'espace latent à 40 dimensions. Les scores de silhouette (haut,gauche), Caliński and Harabasz (haut,droite) et Davies and Bouldin (bas,centre) sont calculés pour les 3 tentatives évoquées section 10.4.1 permettant de calculer le score minimal, moyen ou maximal pour chaque nombre de classes tenté, de 2 à 50. 236

10.8 Nombre de scénarios selon les échéances. Deux réseaux PE-AROME sont utilisés avec 6h d'intervalle (PE-AROME 3h + PE-AROME 9h par exemple), l'échéance indiquée correspond à celle du réseau le plus ancien. Chaque courbe représente un nombre de classes différent allant de 10 à 30 avec un pas de 5 classes. Le graphique de gauche correspond à la moyenne pour chaque échéance, le graphique de droite au Q90 de la distribution pour chaque échéance. 238

11.1 Représentation de la synthèse par scénarios pour la zone Sud-Est sous forme condensée. Les réseaux PE-AROME représentés sont ceux du 12 septembre 2019 à 15h et 21h UTC. Les heures de prévision sont représentées en abscisse, du 13 septembre 02h UTC jusqu'au 14 septembre 12h UTC avec un pas horaire. L'ordonnée représente 16 des 20 classes ou structures météorologiques définies dans le chapitre précédent. Les points bleus représentent la répartition des 32 membres PE-AROME parmi les 16 classes. Plus le point est gros, plus le nombre de membres est important. Cette synthèse représente aussi les prévisions faites par deux réseaux AROME, ceux du 12 septembre à 18h UTC et du 13 septembre à 0h UTC. Ces prévisions sont représentées sous forme d'un cercle rose pour le réseau le plus ancien (18h UTC) et vert pour le plus récent (0h UTC). . . 243

11.2 Exemple de carte ANTILOPE sur la zone Sud-Est. Cumul de pluie entre le 3 octobre 2021 à 02h et 03h UTC. Les points de grille dépassant les 350 mm (couleur bleu canard) sur cette carte sont hors domaine pour le produit ANTILOPE. Cette couleur bleu canard correspond donc au masque ANTILOPE (Italie et extrême sud du domaine en Méditerranée). 244
11.3 Représentation de la synthèse par scénarios sous forme condensée en ajoutant la vérification ANTILOPE. La représentation est similaire à celle de la figure 11.1 en ajoutant la projection d'ANTILOPE sur les structures météorologiques à l'aide d'un carré noir. 245
11.4 Représentation des scénarios « Rien de significatif » (encadré vert), « Pyrénées » (encadré orange) et « Massif-Central Centre » (encadré violet) de la figure 11.3 pour le 13 septembre 2021 à 18h UTC. . . . 247
11.5 Situation météorologique du 14 septembre 2021. La carte du haut représente l'ANASYG produit par le CNP pour le 14 septembre 2021 à 06h UTC. La carte au centre représente la carte de vigilance émise par Météo-France le 14 septembre 2021 en fin de matinée et la carte du bas le cumul de pluie observé sur la journée du 14 septembre 2021 (Source : Okapi, Météo-France). 248
11.6 Évolution de la synthèse par scénarios en partant des réseaux PE-AROME du 12 septembre 2021 à 15h et 21h UTC (figure 11.1), jusqu'aux réseaux du 13 septembre 2021 21h et 14 septembre 2021 03h avec un pas de 6h. L'encadré en rouge représente la période d'intérêt correspondant aux heures où ANTILOPE a été classé dans la zone Cevennes/Gard/Herault-Nord (cf figure 11.3). 250
11.7 Scénario Cevennes/Gard/Herault-Nord pour les réseaux du 12 septembre 2021 à 21h UTC et du 13 septembre 2021 à 03h UTC, prévision pour le 14 septembre 2021 à 09h UTC. 251
11.8 Scénario Cevennes/Gard/Herault-Nord pour les réseaux du 13 septembre 2021 à 21h UTC et du 14 septembre 2021 à 03h UTC, prévision pour le 14 septembre 2021 à 09h UTC. 252
11.9 Situation météorologique du 2 octobre 2020. La carte du haut représente l'ANASYG produit par le CNP pour le 2 octobre 2020 à 12h UTC. La carte au centre représente la carte de vigilance émise par Météo-France le 2 octobre 2020 à 6h du matin et la carte du bas le cumul de pluie observé sur la journée du 2 octobre 2020 (Source : Okapi, Météo-France). 253
11.10 Évolution de la synthèse par scénarios en partant des réseaux PE-AROME du 30 septembre 2020 à 15h et 21h UTC jusqu'aux réseaux du 1er octobre 2020 à 15h et 21h UTC avec un pas de 6h. Le dernier graphique en bas à droite représente de nouveau la synthèse par scénarios pour les réseaux du 1er octobre 2020 à 15h et 21h UTC, mais avec ANTILOPE. 254
11.11 Scénario Alpes-Maritimes/Golfe de Gênes pour les réseaux du 30 septembre 2020 à 21h UTC et du 01 octobre 2020 à 03h UTC, prévision pour le 2 octobre 2020 à 15h UTC. 255
11.12 Scénario Italie, mêmes réseaux et même heure de prévision que la figure 11.11. 256

11.13Situation météorologique du 4 octobre 2021. La carte du haut représente l'ANASYG produit par le CNP pour le 4 octobre 2021 à 06h UTC. La carte au centre représente la carte de vigilance émise par Météo-France le 4 octobre 2021 à 6h du matin et la carte du bas le cumul de pluie observé sur les journées du 3 et 4 octobre 2021 (Source : Okapi, Météo-France). 257

11.14Évolution de la synthèse par scénarios en partant des réseaux PE-AROME du 2 octobre 2021 à 15h et 21h UTC jusqu'aux réseaux du 3 octobre 2021 à 15h et 21h UTC avec un pas de 6h. Le dernier graphique en bas à droite représente de nouveau la synthèse par scénarios pour les réseaux du 3 octobre 2021 à 15h et 21h UTC, mais avec ANTILOPE. 258

11.15Scénario Espagne pour les réseaux du 3 octobre 2021 à 15h UTC et à 21h UTC, prévision pour le 4 octobre 2021 à 03h UTC. 259

11.16Scénario Var/PACA Ouest/Drôme-Sud, mêmes réseaux et même heure de prévision que la figure 11.15. 260

11.17Répartition d'ANTILOPE à travers 19 des 20 structures météorologiques. Les noms correspondent à ceux donnés en annexe, figures E.3 à E.22. La classe « Rien de significatif » n'est pas représentée car elle représente un trop grand nombre de cas et rendrait la distinction des autres classes complexe. Les histogrammes de gauche correspondent en vert au nombre de fois qu'ANTILOPE a été assigné dans la classe et en rouge le nombre de fois où ce fut le cas seul, sans être accompagné par un membre PE-AROME. La carte de droite trace le pourcentage de fois où ce cas de figure se présente (ratio entre les valeurs en rouge et vert). 261

11.18Répartition de la taille des scénarios. Le graphique en haut à gauche représente la distribution complète de la taille des scénarios sur le jeu de données. Le graphique en haut à droite divise cette distribution en sous-catégories : si les deux réseaux AROME appartiennent au scénario (vert), si seulement le réseau AROME le plus récent appartient au scénario (bleu), si seulement le réseau AROME appartient au scénario (violet) et enfin si aucun AROME appartient au scénario (rouge). Les graphiques en bas correspondent aux mêmes types de distribution mais où l'ensemble des cas avec ANTILOPE assigné à la classe « Rien de significatif » a été retirés. 262

11.19Courbe de fiabilité avec une approche par scénarios (graphique de gauche). Toutes les échéances sont prises en compte. Cette courbe est décomposée en quatre sur le graphique de droite comme ce fut le cas pour la taille des scénarios pour la figure 11.18. 264

11.20Pourcentage de scénarios corrects selon les scénarios ou modèles choisis. Les échéances en abscisse sont celles de la PE-AROME la plus ancienne (cf section 11.1.1, correspondant au réseau de n h). Pour les scénarios, les courbes bleues désignent le scénario majoritaire, les courbes vertes le scénario majoritaire ou secondaire. Les courbes rouges et roses désignent les deux réseaux AROME et la courbe marron si on considère les deux réseaux AROME. Ces courbes sont tracées dans le cas général (avec la classe « Rien de significatif ») à gauche, seulement sur les cas pluvieux à droite. . . . 265
11.21Schéma pour prendre une décision entre trois scénarios (points bleus), dont deux accompagnés par un réseau AROME (cercles vert et rose). 266
11.22Répartition des scénarios corrects selon la présence ou non d'un réseau AROME dans le scénario. La courbe rouge représente les scénarios avec les deux réseaux AROME, la courbe jaune les scénarios avec strictement un réseau AROME, la courbe marron les scénarios avec seulement des membres PE-AROME, la courbe bleue avec exclusivement les deux réseaux AROME, la courbe rose exclusivement un réseau AROME et la courbe grise représente les scénarios corrects anticipés par aucun membre et aucun réseau AROME. La carte de gauche représente le cas général, la carte de droite les cas pluvieux. Ces courbes sont tracées en fonction des échéances comme pour la figure 11.20. 267
11.23Taille moyenne des scénarios corrects. La courbe verte représente la taille moyenne si on considère uniquement les membres de la PE-AROME. Sur la courbe bleue, les réseaux AROME sont également comptabilisés. La carte de gauche représente le cas général, la carte de droite les cas pluvieux. Ces courbes sont tracées en fonction des échéances comme pour la figure 11.20. 268
11.24Répartition des scénarios corrects selon l'ordre des scénarios du plus peuplé au moins peuplé. La courbe rouge représente les scénarios majoritaires, la courbe orange les scénarios secondaires, la courbe jaune les scénarios tertiaires, la courbe rose les scénarios quaternaires et la courbe grise les cas autres (au-delà du quatrième scénario ou non couvert par les membres). La carte de gauche représente le cas général, la carte de droite les cas pluvieux. Ces courbes sont tracées en fonction des échéances comme pour la figure 11.20. . . . 269
11.25Comparaison des scores des sections 11.3.4 à 11.3.6 entre les anciennes versions des modèles PE-AROME et AROME (courbes tiretés) et les nouvelles versions en opérationnel depuis l'été 2022 (courbes pleines). A l'exception des courbes de fiabilité, les autres graphiques se concentrent uniquement sur les cas pluvieux (le cas général étant jugé moins intéressant). 271

Liste des tableaux

4.1 Hyperparamètres à régler selon 3 catégories : phase d'entrainement/apprentissage du réseau, architecture du réseau et fonction coût. Les hyperparamètres pour l'architecture sont détaillés couche par couche. 96

5.1 Valeurs de N_{stride} et N_{patch} selon la taille de l'imagette en entrée . . 111

5.2 Tableau de contingence. Les bonnes détections (a) correspondent au nombre d'images avec au moins un écho arqué labellisé et un écho arqué détecté par le U-Net. Les fausses alarmes (b) sont le nombre d'images avec un ou des échos arqués détectés dans le U-Net mais aucun labellisé. Les non détections (c) correspondent au nombre d'images avec un ou des échos arqués labellisés mais aucun détecté. Les bonnes non détections (d) sont les images avec aucun écho arqué détecté et labellisé. 114

6.1 Pourcentage du jeu de données de départ conservé après importance sampling selon 2 valeurs de m (1.5 et 3) et 4 valeurs de s_{rr} (1mm, 3mm, 5mm et 7mm). P_{min} est fixé à 0.1. 121

6.2 Liste des hyperparamètres liés à l'entraînement de l'autoencodeur qui seront testés (colonne de gauche) et leurs différentes valeurs (colonne de droite). 124

6.3 Valeurs de σ en km ainsi que s_{obj} en mm pour les trois objets : pluie totale, modérée et forte. 125

6.4 Performances des autoencodeurs avec des tailles de batchs (B_{size}) différentes. n représente les 3 entraînements différents pour les mêmes hyperparamètres. Les scores sont divisés en trois catégories pour chaque type d'objet. Les TD et TFA sont arrondis au centième près et $\overline{dist}$ au dixième de kilomètre. La taille de batch retenue apparaît en gras. 129

6.5 Tableau similaire au tableau 6.4 pour la sélection des paramètres d'importance sampling m et s_{rr}. La configuration retenue apparaît en gras. 130

6.6 Tableau similaire au tableau 6.4 pour la sélection des hyperparamètres concernant l'architecture de l'autoencodeur pour un espace latent à 10 dimensions. Le nombre de filtres $N_{filtres}$ sera désigné sous la forme (16...128) ou (32...256) pour les suites respectivement (16,32,64,128) et (32,64,128,256). La taille du premier filtre S_{filtre} sera désignée avec le nombre 16 (pour 16×16) ou 8 (pour 8×8). L'autoencodeur sélectionné apparaît en gras. 131
6.7 Tableau similaire au tableau 6.6 pour la sélection des hyperparamètres concernant l'architecture de l'autoencodeur pour un espace latent à 20 dimensions. L'autoencodeur sélectionné apparaît en gras. 132
6.8 Tableau similaire au tableau 6.6 pour la sélection des hyperparamètres concernant l'architecture de l'autoencodeur pour un espace latent à 40 dimensions. L'autoencodeur sélectionné apparaît en gras. 133
6.9 Détail des scores région par région (figure 3.13) pour les 3 autoencodeurs sélectionnés précédemment. 134
6.10 Comparaison des scores objets après une compression de l'information par autoencodeur, ondelettes ou ACP. Cette comparaison est faite pour des espaces à 10,20 et 40 dimensions 137

Liste des acronymes

ACP	Analyse en Composantes Principales.
AEARO	Assimilation d'Ensemble AROME.
AEARP	Assimilation d'Ensemble ARPEGE.
AIES	Artificial Intelligence for the Earth Systems.
AMI	Adjusted Mutual Information.
ANASYG	ANAlyses SYnoptiques Graphiques.
ANTILOPE	ANalyse par spaTIaLisation hOraire des PrEcipitations.
AROME	Application of Research to Operations at MEsoscale.
ARPEGE	Action de Recherche Petite Echelle Grande Echelle.
AS	Adaptations Statistiques.
BGD	Batch Gradient Descent.
CEPMMT	Centre Européen pour les Prévisions météorologiques à moyen terme.
CMC	Centre Météorologique Canadien.
CNP	Centre National de Prévision.
CNRM	Centre National de Recherches Météorologiques.
CRPS	Continuous Rank Probability Score.
DIROP	Direction des Opérations pour la Prévision.
DL	Deep Learning.
DSI	Direction des Systèmes d'Information.
DSM	Direction des Services Météorologiques.
DSO	Direction des Systèmes d'Observation.
DWD	Deutscher Wetterdienst/Service météorologique allemand.
EDA	Ensemble of Data Assimilations.
EFI	Extreme Forecast Index.
EPS	Ensemble Prediction System.

GEFS	Global Ensemble Forecast System.
GMAP	Groupe de Modélisation et d'Assimilation pour la Prévision.
HarmonEPS	HARMONIE Ensemble Prediction System.
HARMONIE	HIRLAM–ALADIN Research on Mesoscale Operational NWP in Euromed.
HPE	Heavy Precipitation Event.
IA	Intelligence Artificielle.
LABO	LABOratoire de Prévisions.
MBGD	Mini Batch Gradient Descent.
MEPS	MetCoOp Ensemble Prediction System.
MJO	Madden-Julian Oscillation.
ML	Machine Learning.
MSE	Mean Squared Error.
NCEP	National Centers for Environmental Prediction.
NMI	Normalized Mutual Information.
NOAA	National Oceanic and Atmospheric Administration.
PACA	Provence-Alpes-Côte d'Azur.
PE	Prévision d'Ensemble.
PE-AROME	Prévision d'Ensemble AROME.
PEARP	Prevision d'Ensemble ARPEGE.
PNT	Prévision Numérique du Temps.
PO	Pyrénées-Orientales.
PP	Perturbed Parameter.
ReLU	Rectified Linear Unit.
RP	Random Parameter.
SERVAL	Système d'Elaboration des produits Radar et de Visualisation centrALisé.
SGD	Stochastic Gradient Descent.
SMHI	Sveriges Meteorologiska och Hydrologiska Institut/Institut suédois de météorologie et d'hydrologie.

SOT	Shift Of Tails.
SPP	Stochastic Parameter Perturbation.
SPPT	Stochastically Perturbed Parameterization Tendencies.
SURFEX	Surface Externalisée.
SVM	Support Vector Machine.
TD	Taux de Détection.
TFA	Taux de Fausses Alarmes.

Annexes

Annexe A

Synthétiser les ensembles pour la prise de décision

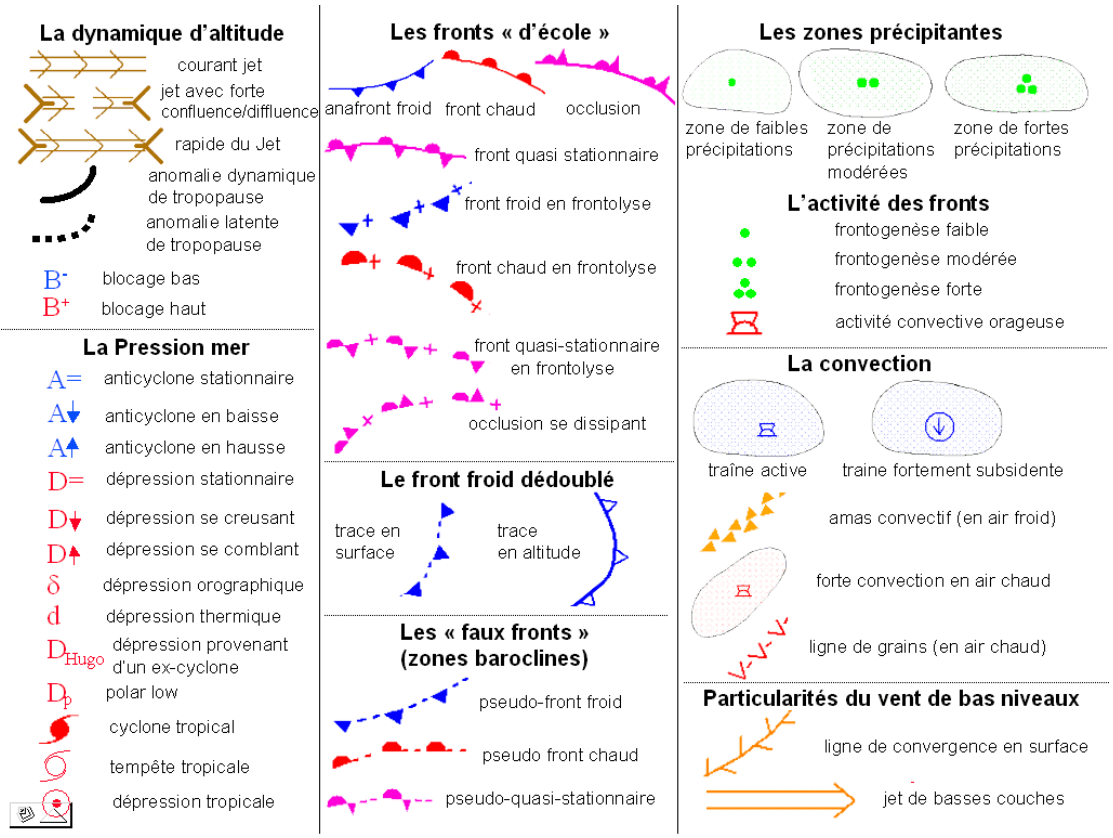


FIGURE A.1 – *Légende des objets météorologiques représentés sur un ANASYG.*

Annexe B

Réduction de dimension avec un autoencodeur

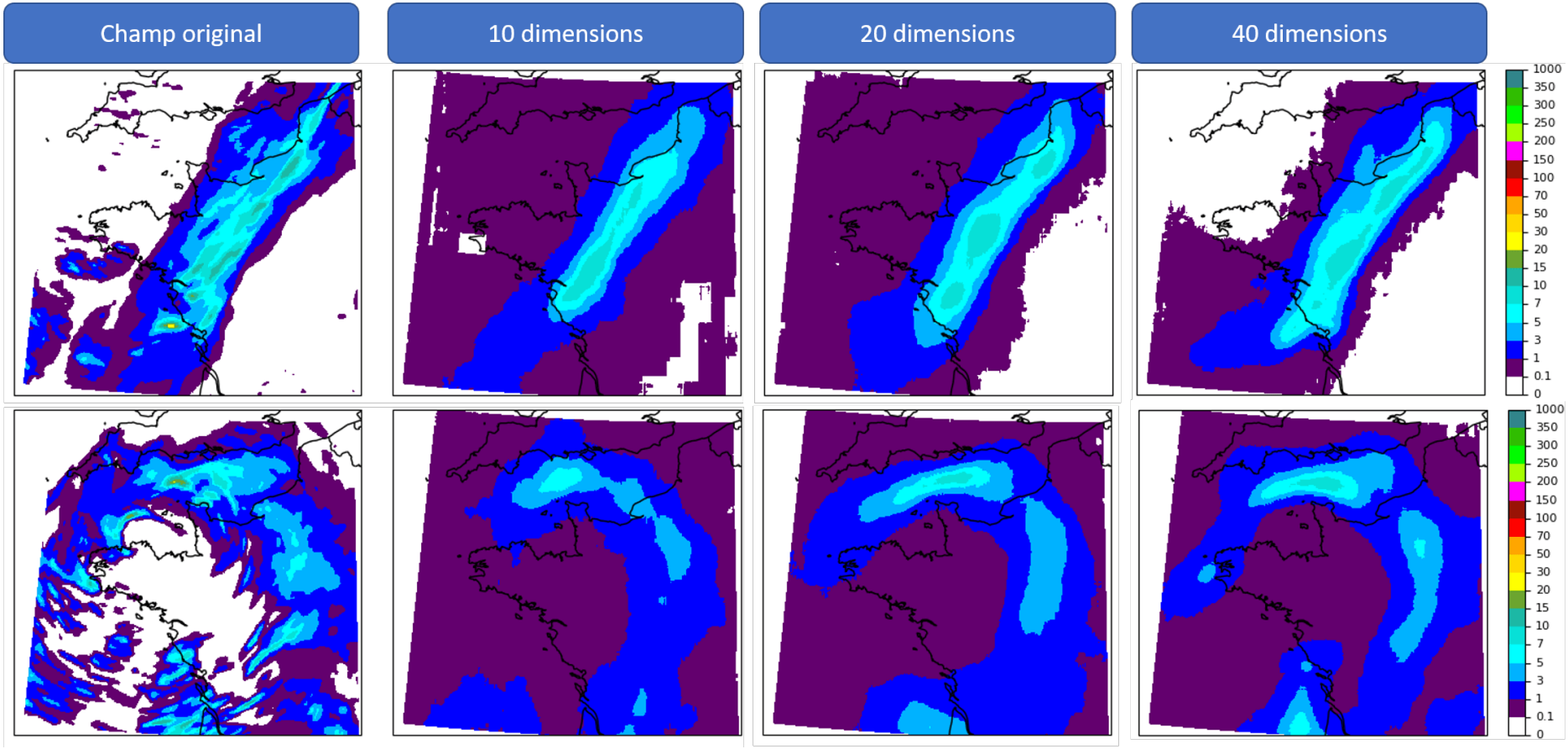


FIGURE B.1 – *Exemple d'images reconstruites pour la zone Nord-Ouest à partir des trois autoencodeurs retenus (avec un espace à 10,20 et 40 dimensions).*

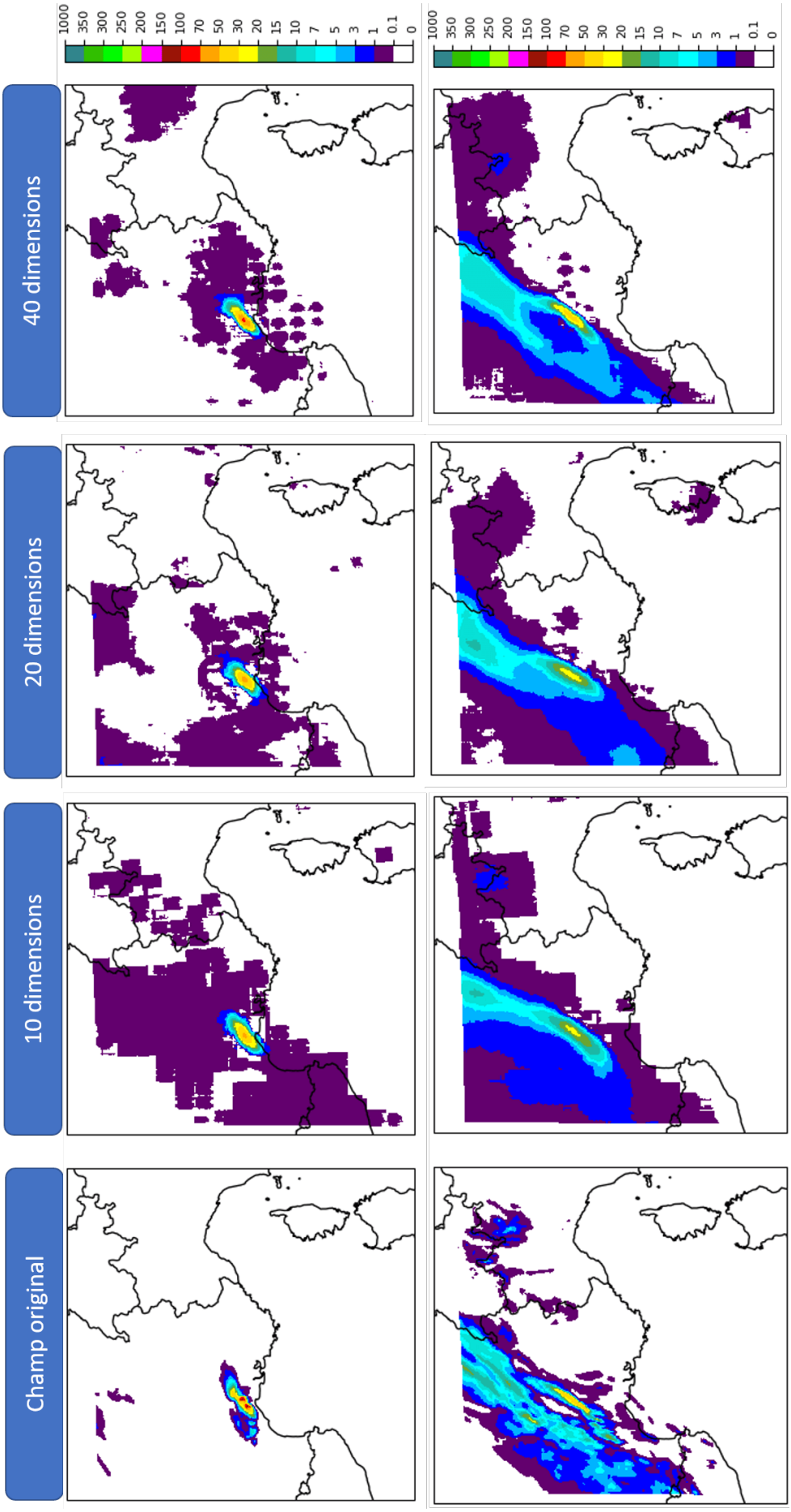


FIGURE B.2 – *Exemple d'images reconstruites pour la zone Sud-Est à partir des trois autoencodeurs retenus (avec un espace à 10,20 et 40 dimensions).*

Annexe C

Performance des modèles

Suivi diagnostic écho arqué (BE), disponible sur previlabo.meteo.fr

Date_obs	Zone	Date_run/H_run (si pertinent)	Auteur commentaire	Fausse alarme ?	Non détection ?	Bonne détection ?	Apport PEARO	Commentaires libres
10/05/2021	SO		Arnaud			1	Oui, stabilité dans les runs	Bonne anticipation par quelques membres sur plusieurs runs consecutifs. Bon zonage bien ciblé.
10/05/2021	Allemagne		Arnaud	1			Non, au contraire	Nombreuses détections. Bizarres car aucun signal sur le Qmax en rafales
11/05/2021	Allemagne		Arnaud	1			Non	Signal sur le Qmax en rafales sur ce cas. Erreur des modèles. Convection observée mais pas d'écho arqué
24/05/2021	Corse		Arnaud			1	Oui, signal moins évident dans ARO	Bonne anticipation. Signal trop étendu spatialement (démarre trop tôt sur les Baléares et se termine trop tard dans le Golfe de Gênes surtout dans les derniers runs)
03/06/2021	Aquitaine		Arnaud	1				Signal faible (~10%) pour la soirée. RAS dans les obs.
04/06/2021	IDF-Hauts-de-France		Arnaud	~1				Cas compliqué ... Diagnostic instable entre les runs. Pas de BE observé mais plutôt des multicellulaires en forme de ligne ou d'arc (particulièrement vrai pour celui qui circule sur la région parisienne en début de matinée). Confusion logique si on se met à la place de l'IA (enfin je crois, mon avis est peut-être biaisé)
16/06/2021	Ouest / Nord	cf ci-dessous	Martin	cf ci-dessous		cf ci-dessous		A J-2 21 TU et J-1 03 TU signal faible (5 à loc 10%) 4 puis 6 occurrences sur territoire (à suivre et peut-être jour d'après)
16/06/2021	Manche	15/06 15h au 16/06 3h	Arnaud	~0.5		~0.5		Signaux persistants en mer + Normandie (~20 %). Peut-être petit BE vers 22h UTC sur ce cas. Pas grand chose globalement.
16/06/2021	SO vers CO		Arnaud	1				Signal faible (~10-15%) pour la soirée. MCS mais pas BE observé
17/06/2021	SO	17/06 03h	Benoît	1		2	Oui pour le SO, non centre-ouest	Bon apport de PEARO dans le SO avec le gros bow echo de 14-16 UTC, renforçant le signal d'ARO. Signal beaucoup trop fort sur le centre-ouest vers 20UTC, le retour était déjà en train de se former (--> flux incompatible avec la formation d'un bow echo).
17/06/2021	SO	17/06 03h	Gaëtan					Ajout commentaire Benoît : globalement la PEARO est assez en retard, mais le zonage est pas mal (tout comme les ARO du jour excepté le 12TU qui est mieux dans la chronologie). Mention spéciale aux membres 16 et 12 de la PE qui ont été pas trop mal pour la chronologie. Grosse mention à ARO00 qui a une phénoménologie quasi-parfaite (signature radar + rafales), mais 1h30 de retard par rapport au obs.
17/06/2021	Nord	17/06 15h	Gaëtan	1			Bon apport pour le contexte supercellulaire, pas pour BE	Contexte plutôt supercellulaire sur le Nord-NE (deux obs dans le secteur de Reims vers 21 et 23 TU et une plus tôt (photographiée en fin de vie cf Infoclimat) à l'ouest d'Auxerre vers 17TU). La ligne orageuse à l'arrière des deux SC sur le NE n'est pas un BE. La détection dans ARO12 et le membre 14 semble lié aux supercellules dans les Rflc proposées. Idem pour les autres membres qui détectent plus au nord. Le signal diminue à juste titre sur PEARO21TU. Signal présent aussi sur le run de PEARO03TU.
17/06/2021	Nord/Belgique	17/06 15h	Arnaud					Ajout commentaire Gaëtan : Ok pour la partie française mais transition vers un BE côté belge un peu plus tard (~ 2-3hTU)
19/06/2021	Gironde (littoral)	18/06 15h	Ludo			1		Très bon apport et réaction du U-net lors de l'apparition de l'écho arqué. La trajectoire suggéré correspond bien à ce que l'on observe au moment de la rédaction de ce premier bilan (19/06 00h37 loc).
20/06/2021	Ardennes/Belgique		Arnaud			1	Oui, stabilité dans les runs. Bon ciblage de la zone à risque max	Comportement globalement OK. Même si l'écho arqué a vraiment pris forme côté Belgique encore une fois. Supercellules plutôt côté français.
21/06/2021	Languedoc		Arnaud		1 ?			Structure en forme de ligne/arc balayant la zone le matin. A voir.
21/06/2021	Nord-Est		Arnaud			1		Signal assez fort au départ de l'Auvergne vers NE mais en diminution en se rapprochant de l'évènement. Structures ressemblant à des BE à la frontière suisse et allemande. Report obs fort vent ESWD.
26/06/2021	Sud-ouest	25/06 15h	Ludo		0.5	0.5		Plusieurs échos arqués détectés par le U-net en fin d'après-midi/début de soirée sur l'occitanie et la frontière Aquitaine. Echo arqué proposé également sur la Normandie. RETEX : Bonne détection du risque de BE, trajectoire trop à l'Est, BE observé remontant sur 47 et non sur Occitanie comme proposé par la PEARO et donc le U-Net.
27/06/21	"large" Massif central		Léo					Signal de 10 à 20 % pour la soirée du 27 sur le réseau 15 TU du 26, avec plusieurs membres concordants sur la tranche horaire. A contrôler avec les obs.
28/06/2021	Grand-Est	28/06 03h	Ludo	1	0	0	Signal stable dans la durée.	Signal de 10 à 15% sur la zone en fin d'après-midi/début de soirée par 3 membres de la PEARO + par les ARO 0 et 3Z. Aucun écho arqué sur Grand-Est.
29/06/2021	Grand-Est + Doubs	28/06 15h	Ludo	0	0	1		Signaux de BE entre 12 et 16h UTC. Risque de BE également sur le Doubs. RETEX : Faibles échos arqués vers 18h UTC sur Grand-est.

FIGURE C.1 – *Suivi de la production expérimentale en mai et juin 2021.*

Suivi diagnostic écho arqué (BE), disponible sur previlabo.meteo.fr

Date_obs	Zone	Date_run/H_run (si pertinent)	Auteur commentaire	Fausse alarme ?	Non détection ?	Bonne détection ?	Apport PEARO	Commentaires libres
03/07/2021	Axe 17/33 - 03/18	02/07 03h	Ludo	1	0	0	Cadrage synoptique ok mais incertitude sur l'organisation orageuse. Meilleur zonage confirmation du scénario AROCEP sur localisation	Signal de 2,5 à 10% sur la zone le 03/07 au matin entre 7 et 11h UTC. L'incertitude sur l'organisation orageuse se retrouve dans les membres de la PEARO (cartes confettis RflcMax>40dBz) . La dispersion est telle que la carte la plus "utile" pour la prévision est celle de la proba. MAJ run 15h UTC : revision à la baisse du scénario avec un seul BE vu vers 12h UTC le 03/07. RETEX : Pas d'écho arqué.
04/07/2021	Pays-Bas/Ouest Allemagne	03/07 21h	Jean	0	0	1	Bonne info sur localisation et "risque"	De petits bow echos assez mal formés et passagers se sont produits entre l'ouest Allemagne et les Pays Bas. La PEARO de 21 h met en évidence qu'il existe un risque de bow echo sur la zone (environ 20 % des membres le proposent). Cette proposition de la PE n'est pas choquante et même pertinente vu la localisation d'une part et l'observation assez ephémère des bow echos.
06/07/2021	Axe Limousin - Lorraine	05/07 03h	Ludo	0.5	0	0.5	Information pour confirmer le positionnement des orages les plus actifs.	Plusieurs détection de BE signalées dans la PEARO entre 23h (05/07) et 08h UTC sur cet axe. RETEX : Bonne détction sur les BE entre 23h et 3h UTC. Signal trop alarmiste sur NE et Centre-est, pas de BE observés sur ces zones là.
		05/07 15h	Arnaud	0	0	1		Ajout commentaire de Ludo : Oui, net renforcement du signal sur le run de 15h. Cas atypique de ce point de vue. Rien d'évident à première vue dans les OBS. RETEX : Formation d'un BE sur Dordogne vers 22h30 UTC, écho qui poursuit sa trajectoire sur Limousin jusqu'à 23h30 UTC. autre BE sur le SO de l'Yonne vers 1h UTC.
08/07/2021	Bavière	08/07 03h	Arnaud			1	Renforcement du signal	BE (au pluriel) observés en Bavière. Fort signal sur le diagnsotic.
23/07/2021	Bretagne	22/07 15 h	Léo	1	0	0	En mode prévi : zonage pertinent avec la synoptique, incitation à zoner large le risque	MAJ run 23/07 3Z (Ludo) : forte réaction du diagnostic entre le Finistère et la Manche pour la tranche horaire 15-22h UTC. Question (Arnaud) : Le commentaire de Léo a été supprimé par erreur ? -> quand je me suis connecté il n'y avait pas/plus de commentaires. RETEX : Cellules orageuses liées à la goutte froide bien présentes et actives, pas de BE en revanche.
23/07/2021	Large partie centrale de la France	22/07 15 h	Léo	1	0	0	Confirme le zonage des déterministes pour la vigilance.	Cas intéressant : potentiel marqué pour des orages très forts (vigilance orange), mais peu de réaction du diagnostic. Les réflectivités d'AROME évoquent davantage un agrégat de cellules, sans structure linéaire vraiment bien dessinée. MAJ run 23/07 3Z (Ludo) : renforcement du signal entre Puy-de-Dome et Nièvre/Saône entre 20 et 0h UTC. RETEX : plusieurs SC mais pas de BE.
31/07/2021 pour 22/23h TU	Alpes-de-Haute-Provence	30/07 21h	Ludo	1	0	0		Quelques échos détectés mais aucun d'observés.
31/07/2021	Médit	30/07/2021	Ludo	0.25	0	0.75	Très bon zonage des zones à fort développement de la convection sévère.	Fort signal sur la méditerranée avec bcp de BE modélisés. Les obs des radars espagnols sur les Baléares montrent plusieurs BE entre 22 et 23h15 UTC. En pleine mer, l'estimation des précips convectives par satellites montrent une forte probabilité de BE en pleine mer entre 23h15 et 00h15 UTC. Le signal de la PEARO est tout de même un peu trop alarmiste par rapport aux obs d'où le 0.25 en fausse alarme.
15/08/2021	70/68	14/08 15h	Ludo	1	0	0		Convection bien observée dans la zone mais pas de BE.
21/08/2021	54/57/67	21/08 3h	Ludo	0	0	1	Bonne info sur la localisation et "risque"	Bonne détection d'une structure légèrement arquée à l'avant du système convectif qui se met en place ce soir là. 70km/h à Rodalbe (57) à 20h30 UTC à l'avant de l'orage. Le réseau de 15h fait disparaitre ce risque malheureusement. Sur le 67, un BE détecté par le run de 3h vers 22/23h UTC sur le Nord du département, bien observé. Le run de 15h est trop péssimiste sur le 67.
	21			0	1	0		Une structure légèrement arquée observable sur le radar vers 21h UTC au SE de Dijon, RAS sur les runs 3 et 15h.
01/09/2021	Espagne		Arnaud	0.5	0	0.5		Forte activité convective en Espagne pour cette journée. Organisation en MCS avec forme en ligne a minima. Le diagnostic réagissait faiblement (~10%). Sauf le run du 01/09/2021 15h qui réagissait plus fort sur Baléares.
26/09/2021	Ouest-Medit' --> Var --> Gênes	25/09 15h	Arnaud	1			1 seul fort signal sur un run	Comportement intrigant de la PE sur ce cas. Un très fort signal est observé pour la nuit/matinée du 26 mais sur un seul run. Rien dans les runs précédents ni dans le run suivant.

FIGURE C.2 – *Suivi de la production expérimentale en juillet, août et septembre 2021.*

Annexe D

Étude approfondie de l'espace latent

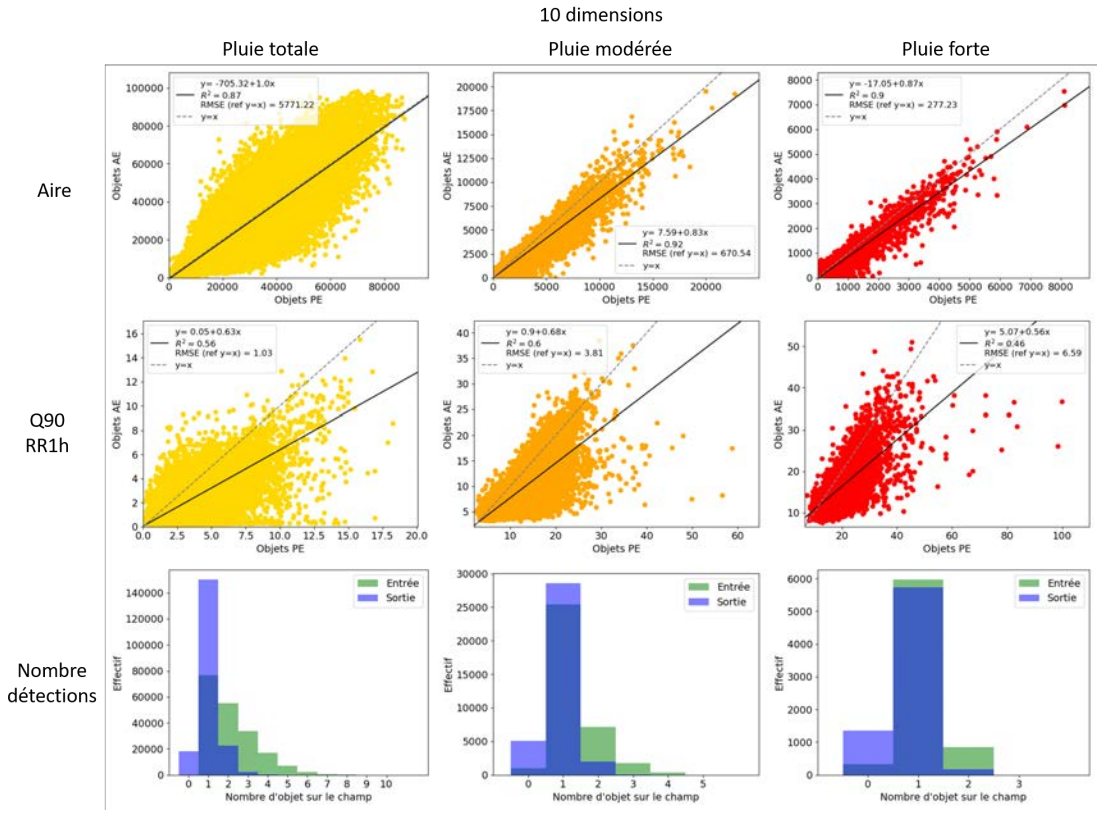


FIGURE D.1 – *Autoencodeur avec 10 dimensions, graphiques similaires à ceux de la figure 9.1(objets entrée/sortie).*

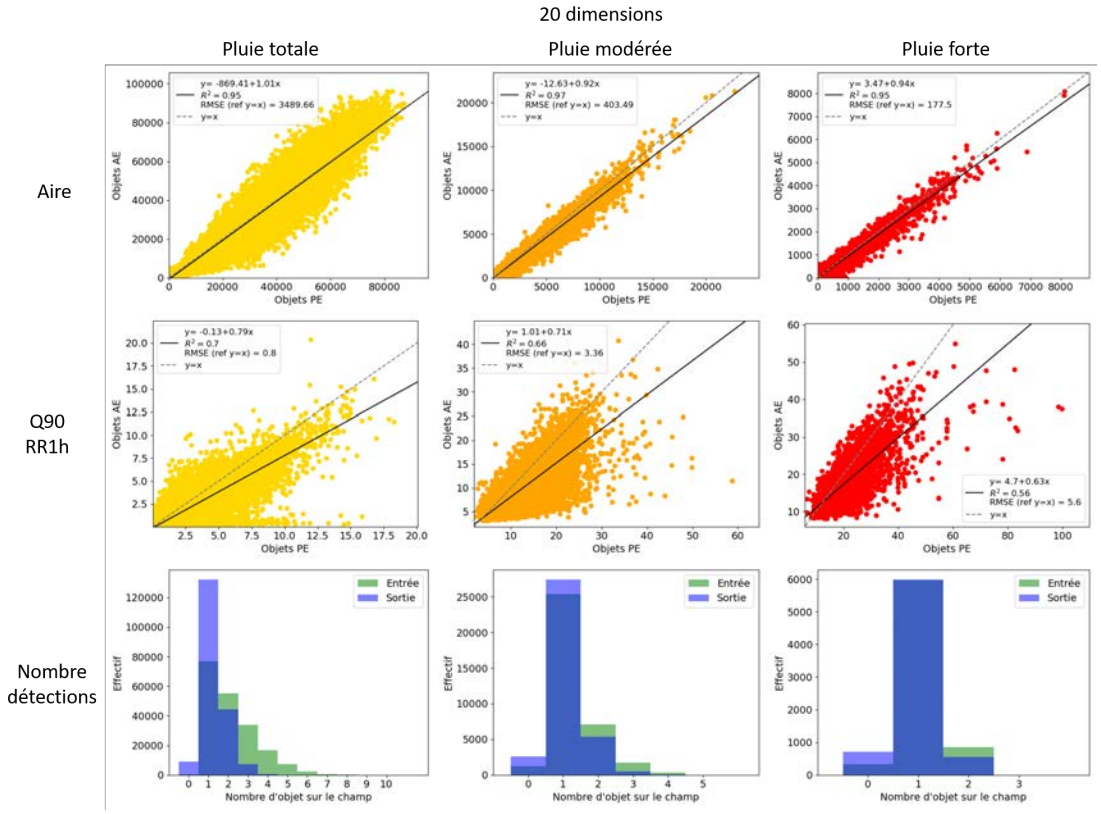


FIGURE D.2 – *Autoencodeur avec 20 dimensions, graphiques similaires à ceux de la figure 9.1 (objets entrée/sortie).*

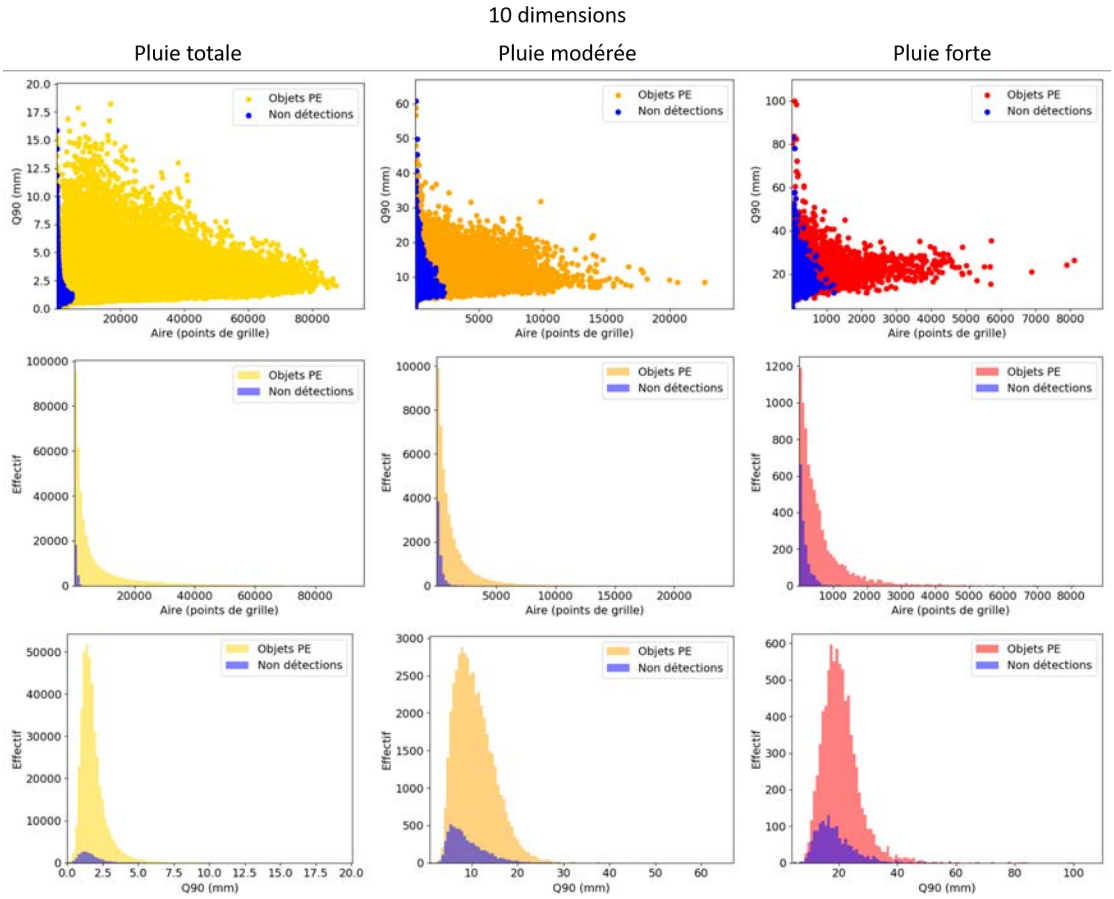


FIGURE D.3 – *Autoencodeur avec 10 dimensions, graphiques similaires à ceux de la figure 9.2 (non détections).*

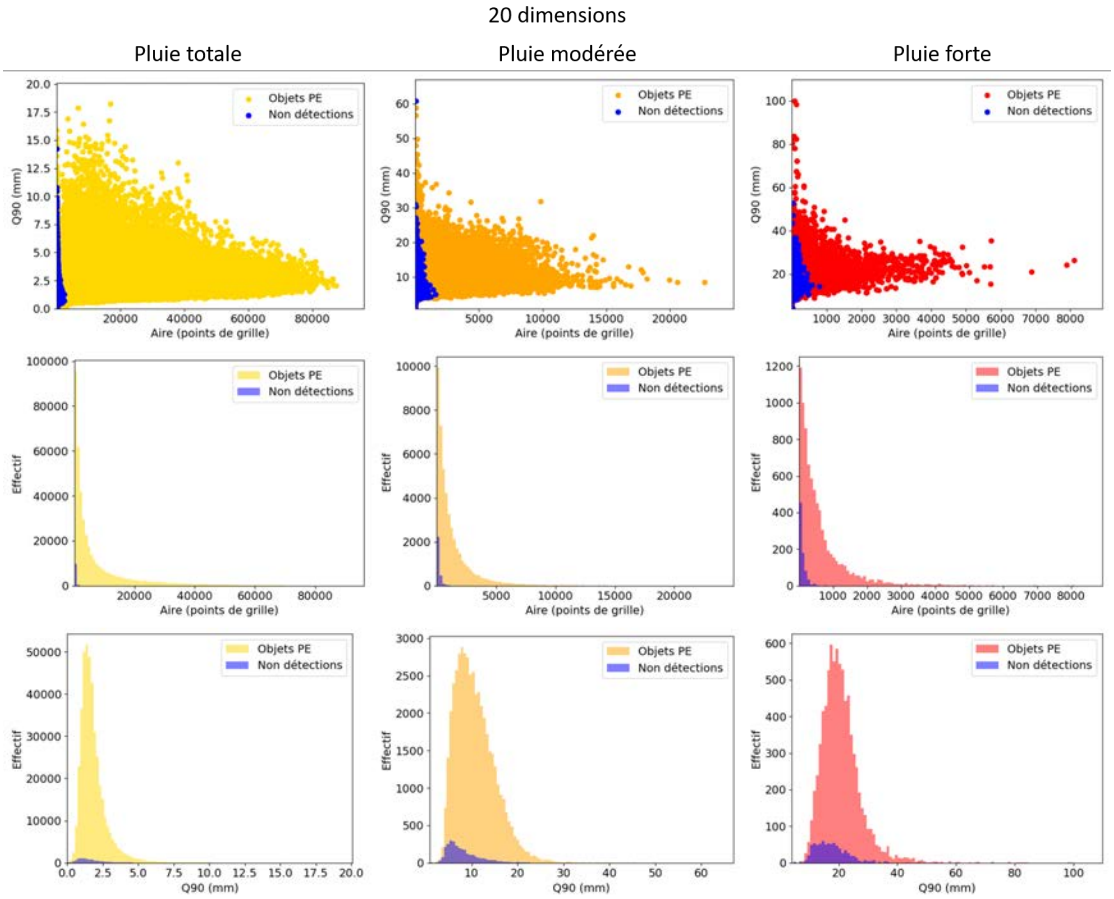


FIGURE D.4 – *Autoencodeur avec 20 dimensions, graphiques similaires à ceux de la figure 9.2 (non détections).*

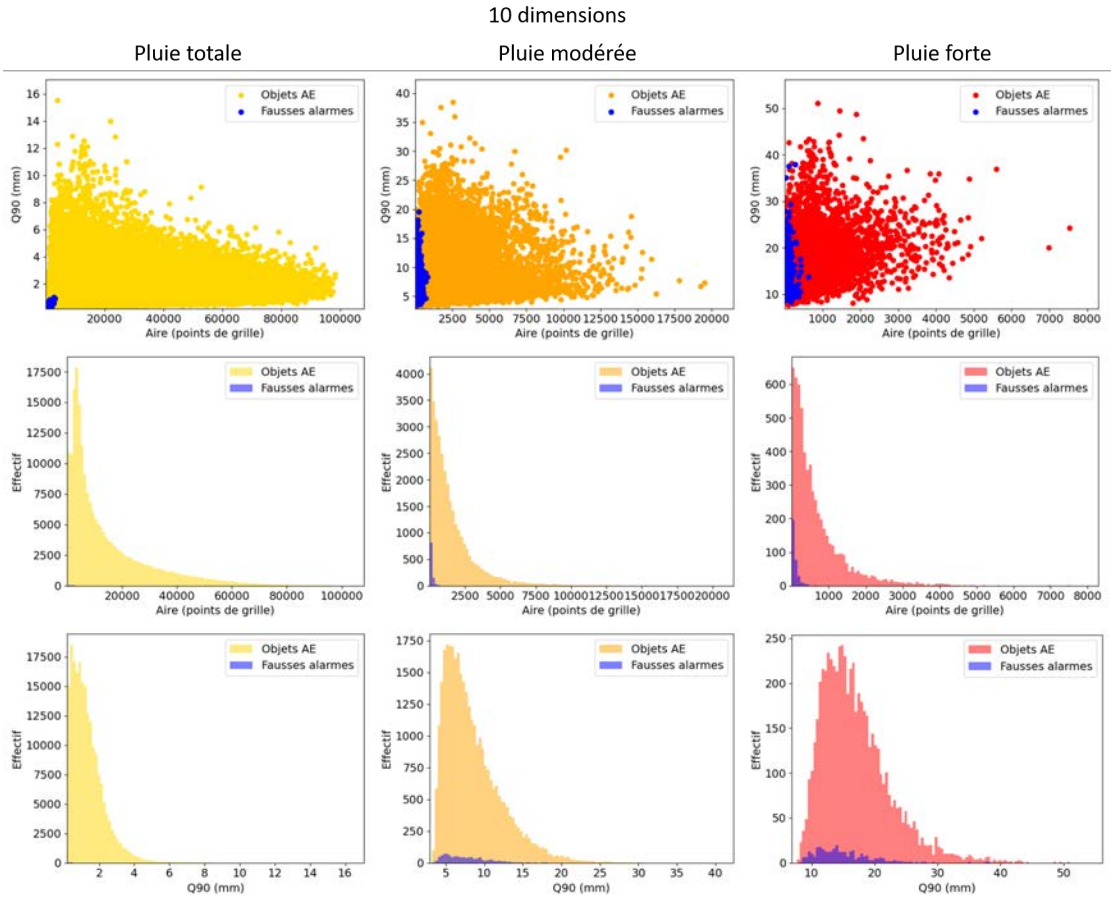


FIGURE D.5 – *Autoencodeur avec 10 dimensions, graphiques similaires à ceux de la figure 9.3 (fausses alarmes).*

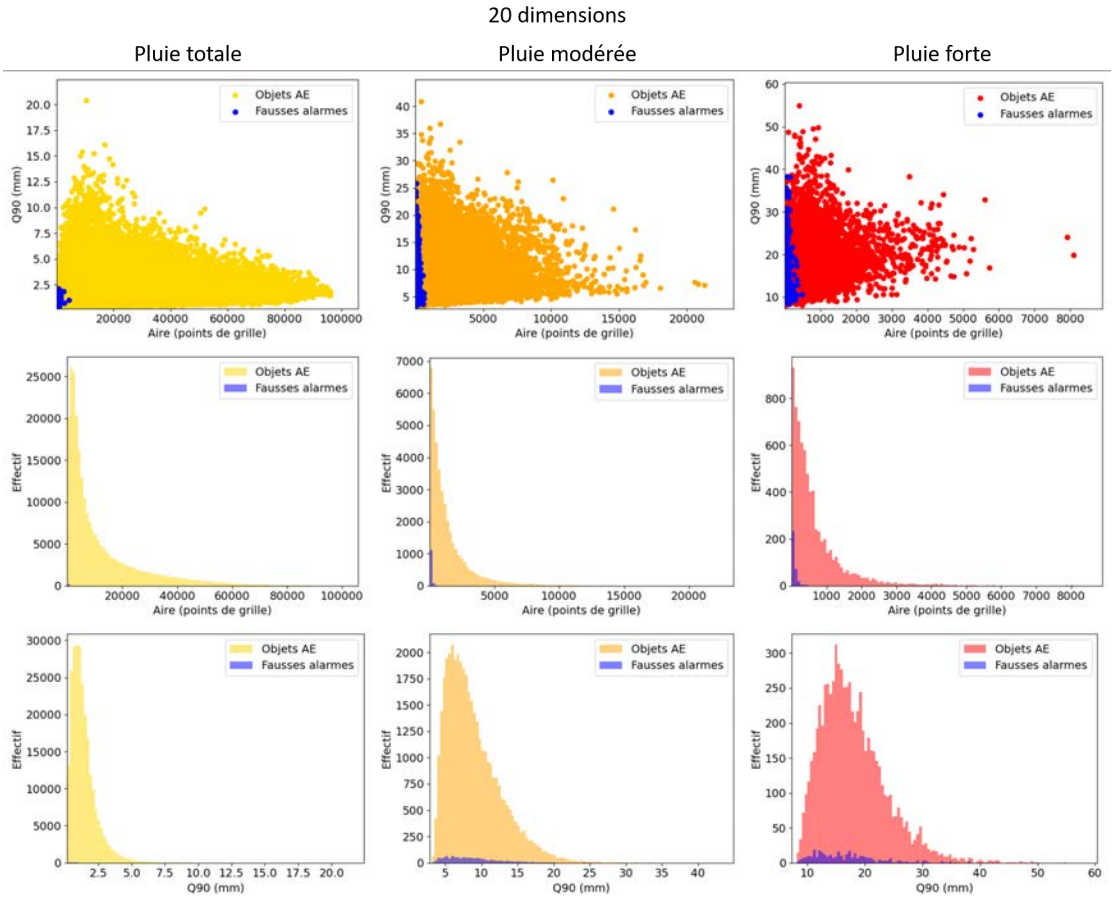
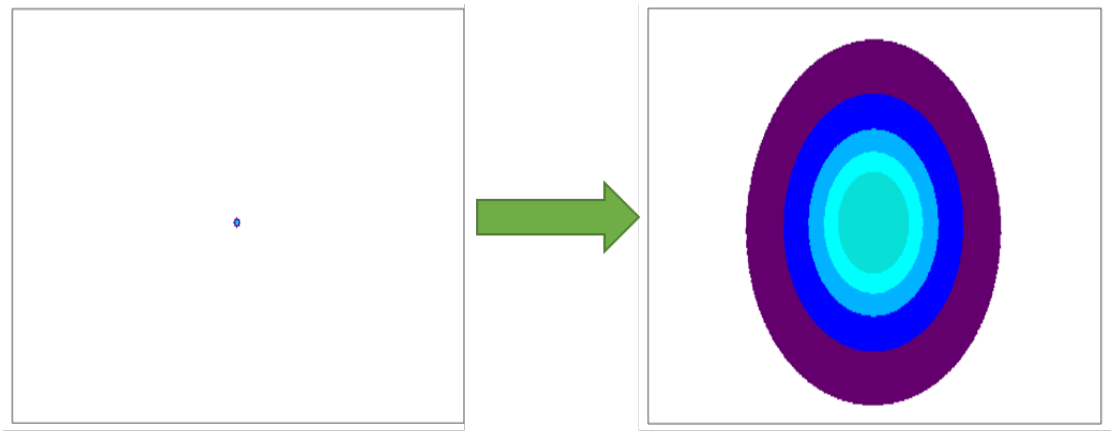


FIGURE D.6 – *Autoencodeur avec 20 dimensions, graphiques similaires à ceux de la figure 9.3 (fausses alarmes).*



20 dimensions

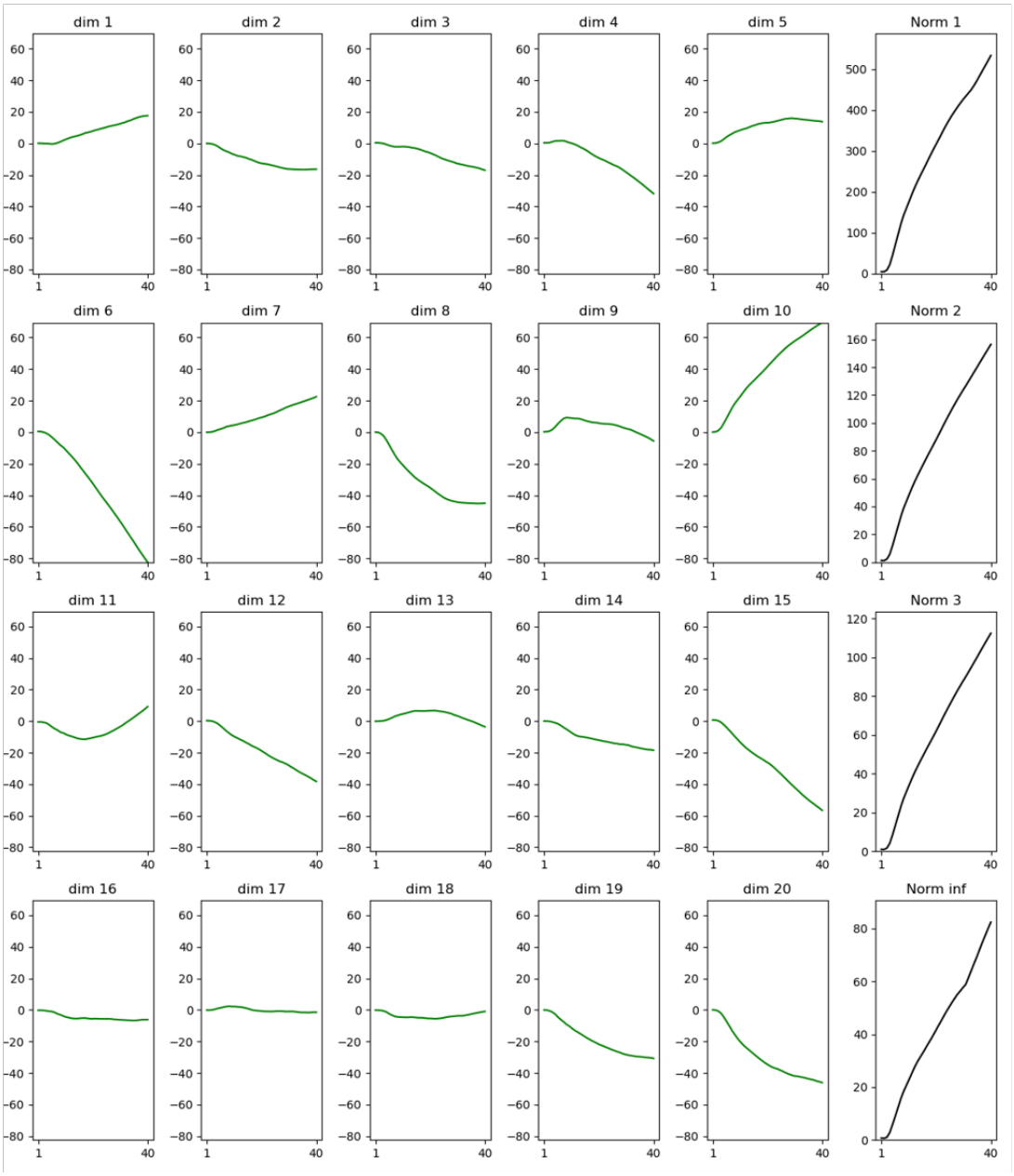


FIGURE D.7 – *Autoencodeur avec 20 dimensions, graphiques similaires à ceux de la figure 9.4 (variation de σ).*

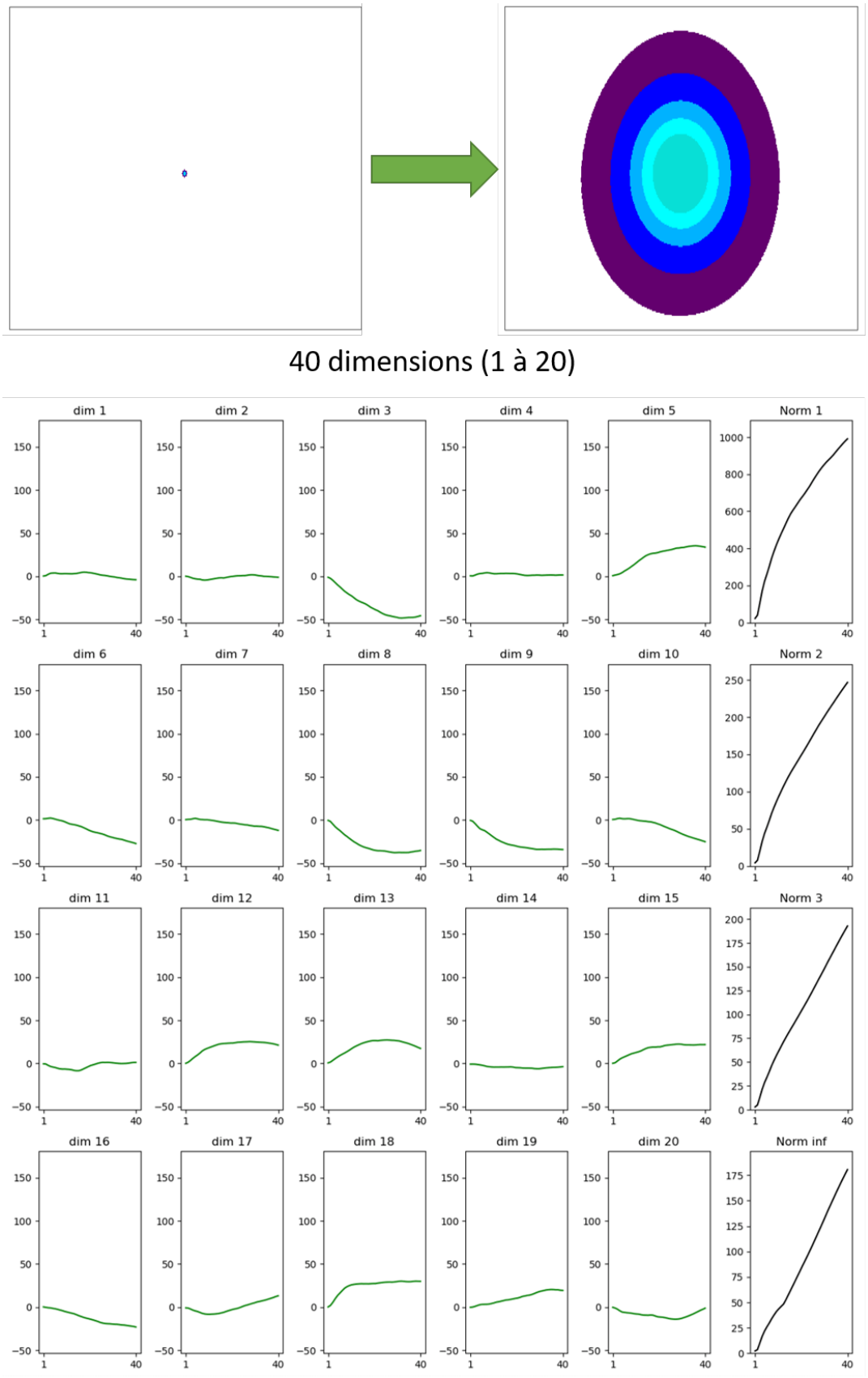
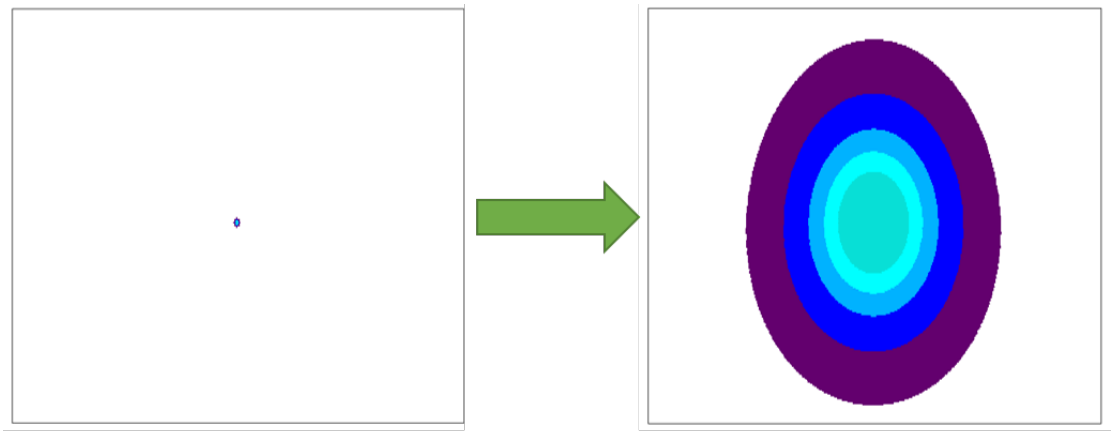


FIGURE D.8 – *Autoencodeur avec 40 dimensions (Partie 1), graphiques similaires à ceux de la figure 9.4 (variation de σ). Les dimensions 1 à 20 sont représentées.*



40 dimensions (21 à 40)

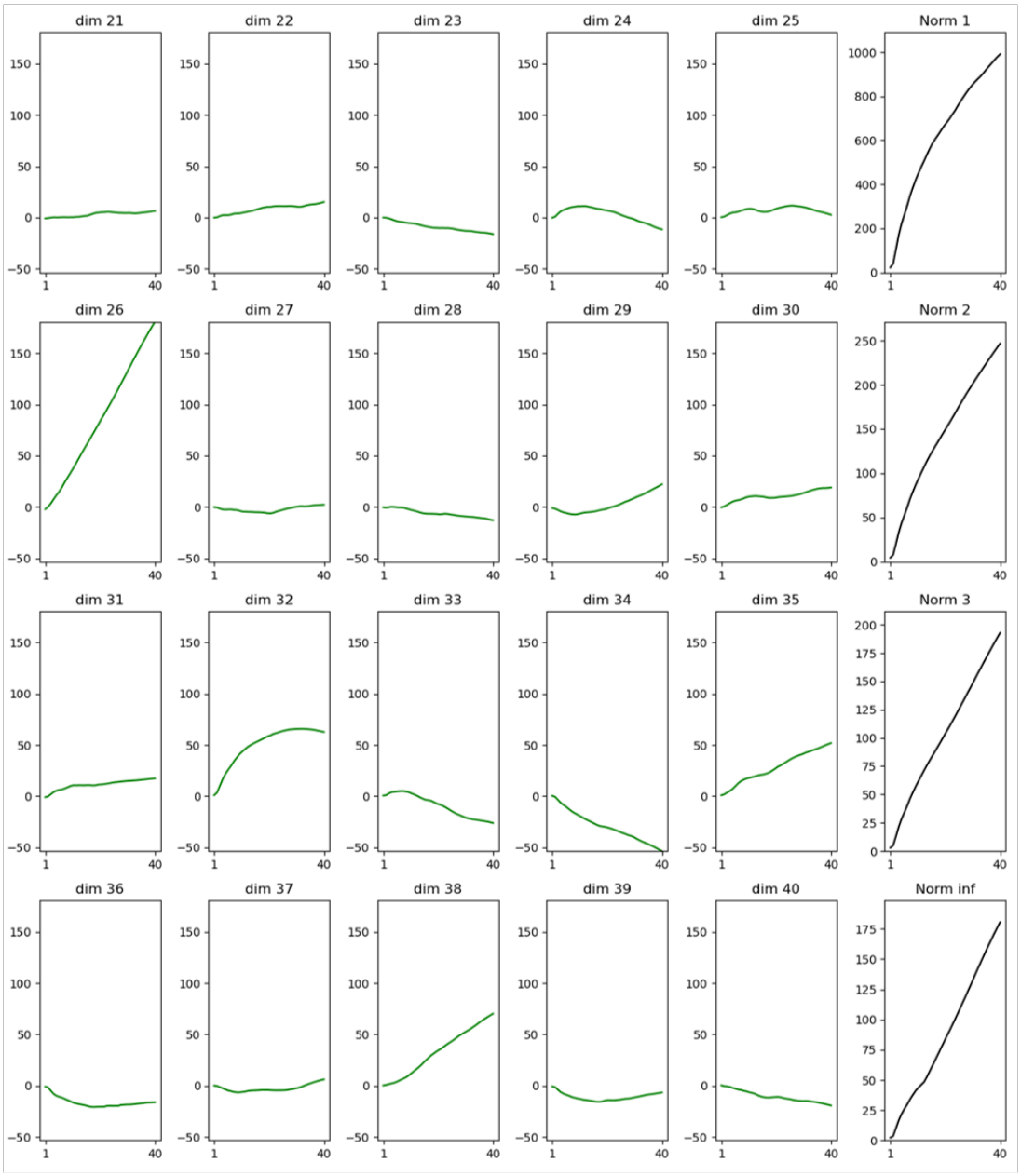
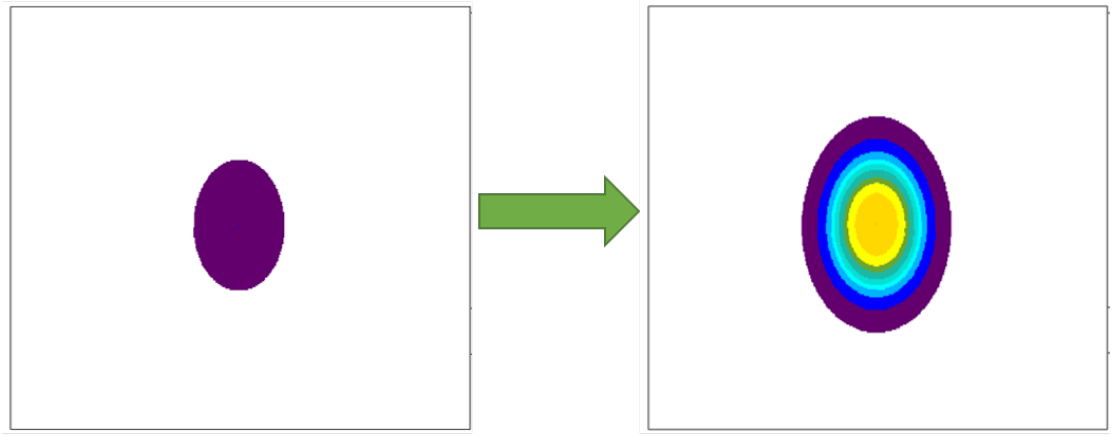


FIGURE D.9 – *Autoencodeur avec 40 dimensions (Partie 2), graphiques similaires à ceux de la figure 9.4 (variation de σ). Les dimensions 21 à 40 sont représentées.*



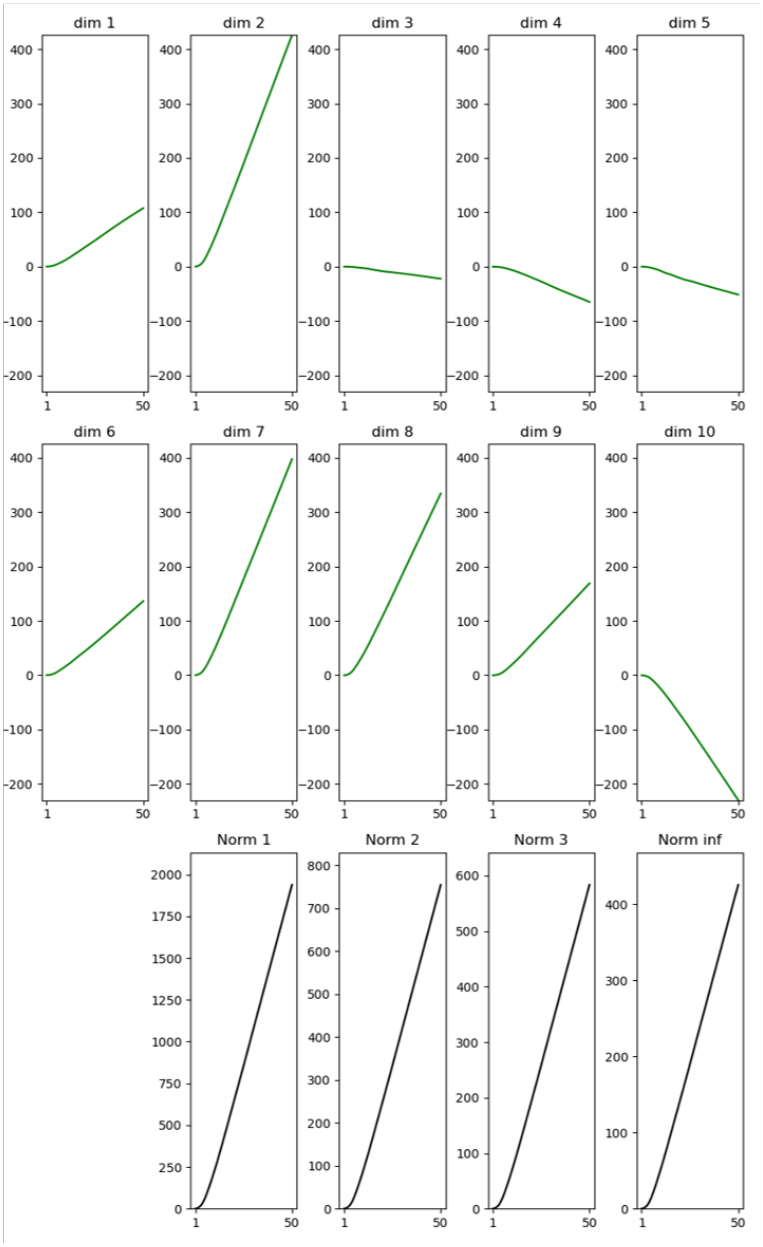
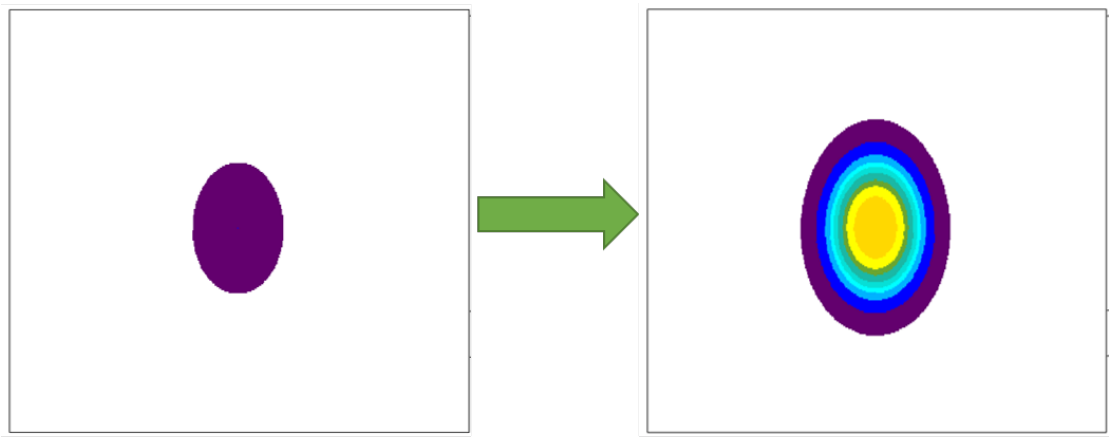


FIGURE D.10 – *Autoencodeur avec 10 dimensions, graphiques similaires à ceux de la figure 9.5 (variation de M). Les dimensions 21 à 40 sont représentées.*



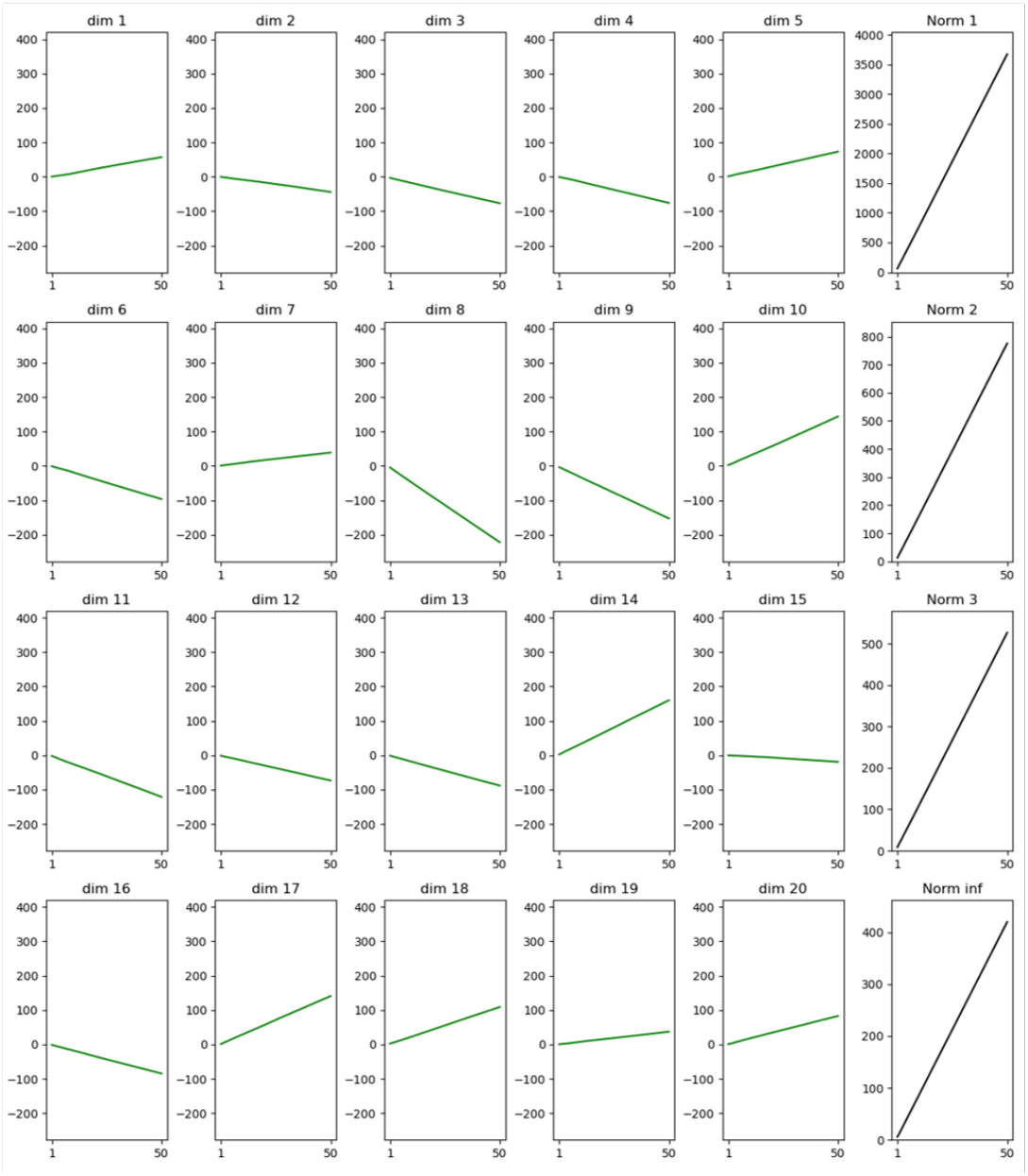
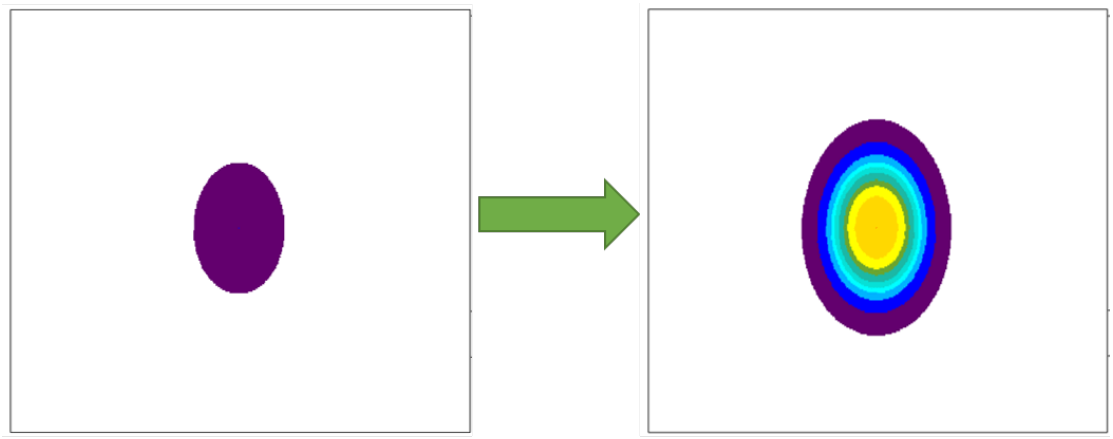


FIGURE D.11 – *Autoencodeur avec 40 dimensions (Partie 2), graphiques similaires à ceux de la figure 9.5 (variation de M). Les dimensions 21 à 40 sont représentées.*



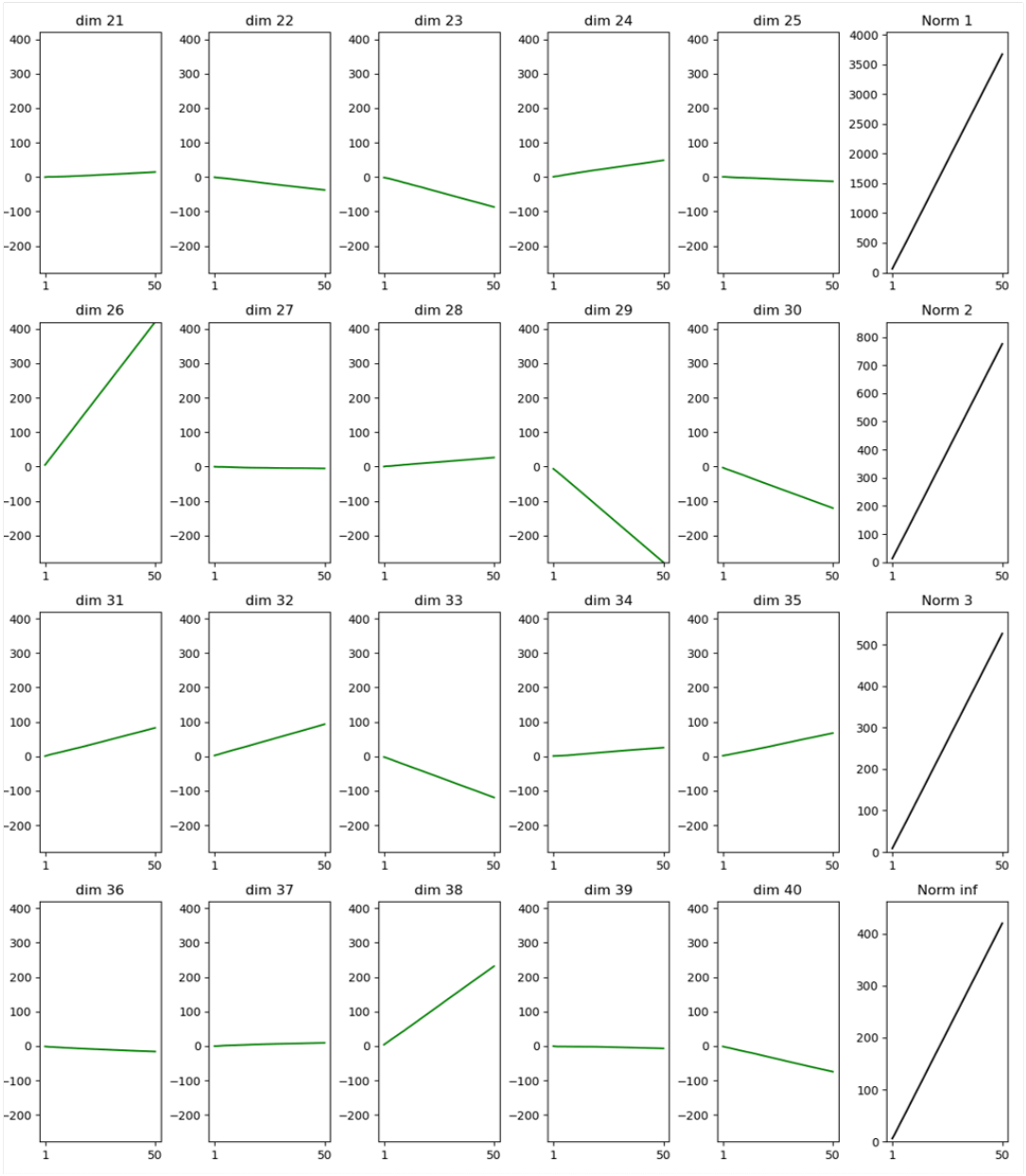


FIGURE D.12 – *Autoencodeur avec 40 dimensions (Partie 2), graphiques similaires à ceux de la figure 9.5 (variation de M). Les dimensions 21 à 40 sont représentées.*

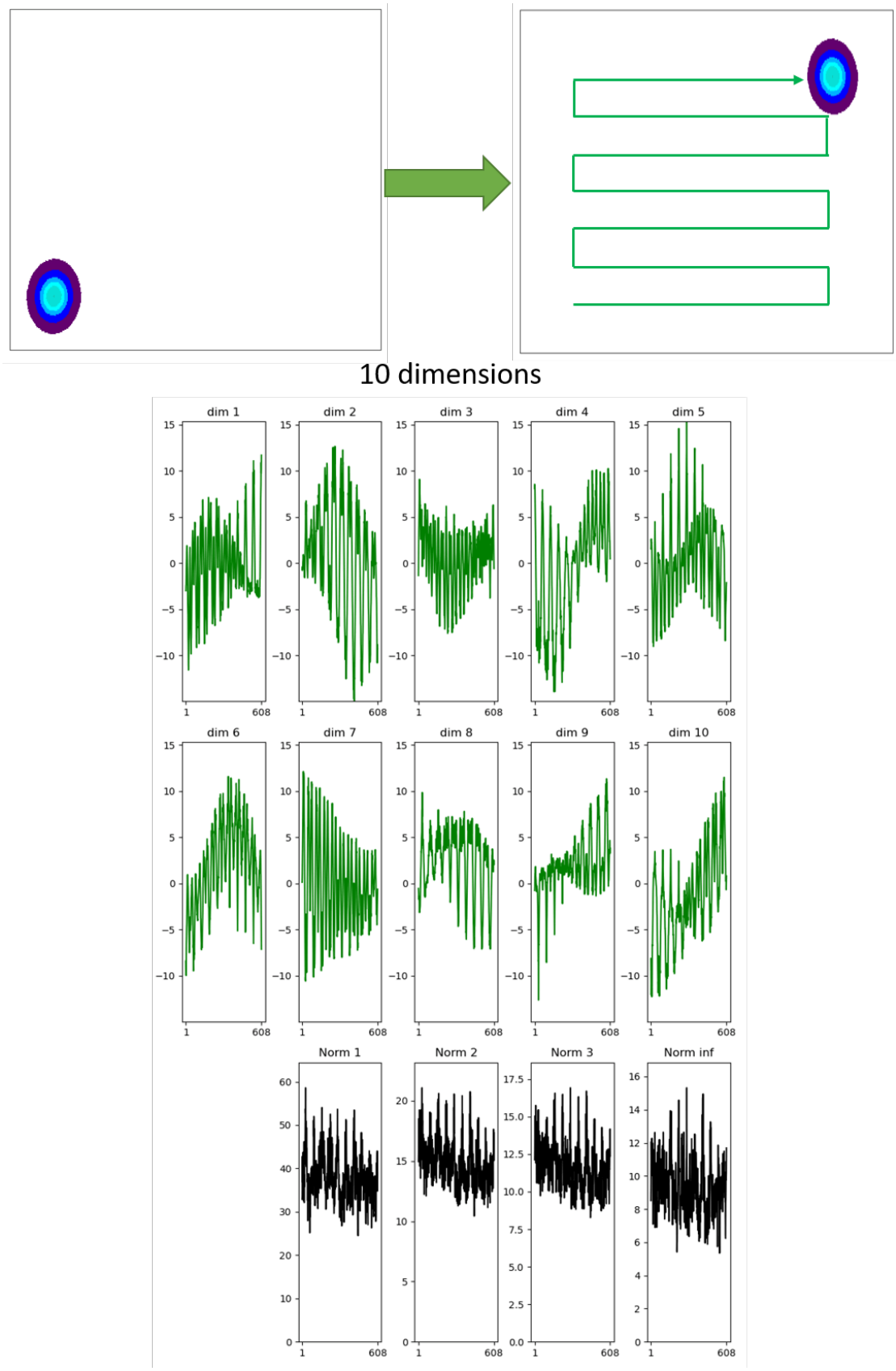


FIGURE D.13 – *Autoencodeur avec 10 dimensions, graphiques similaires à ceux de la figure 9.6 (variation de position).*

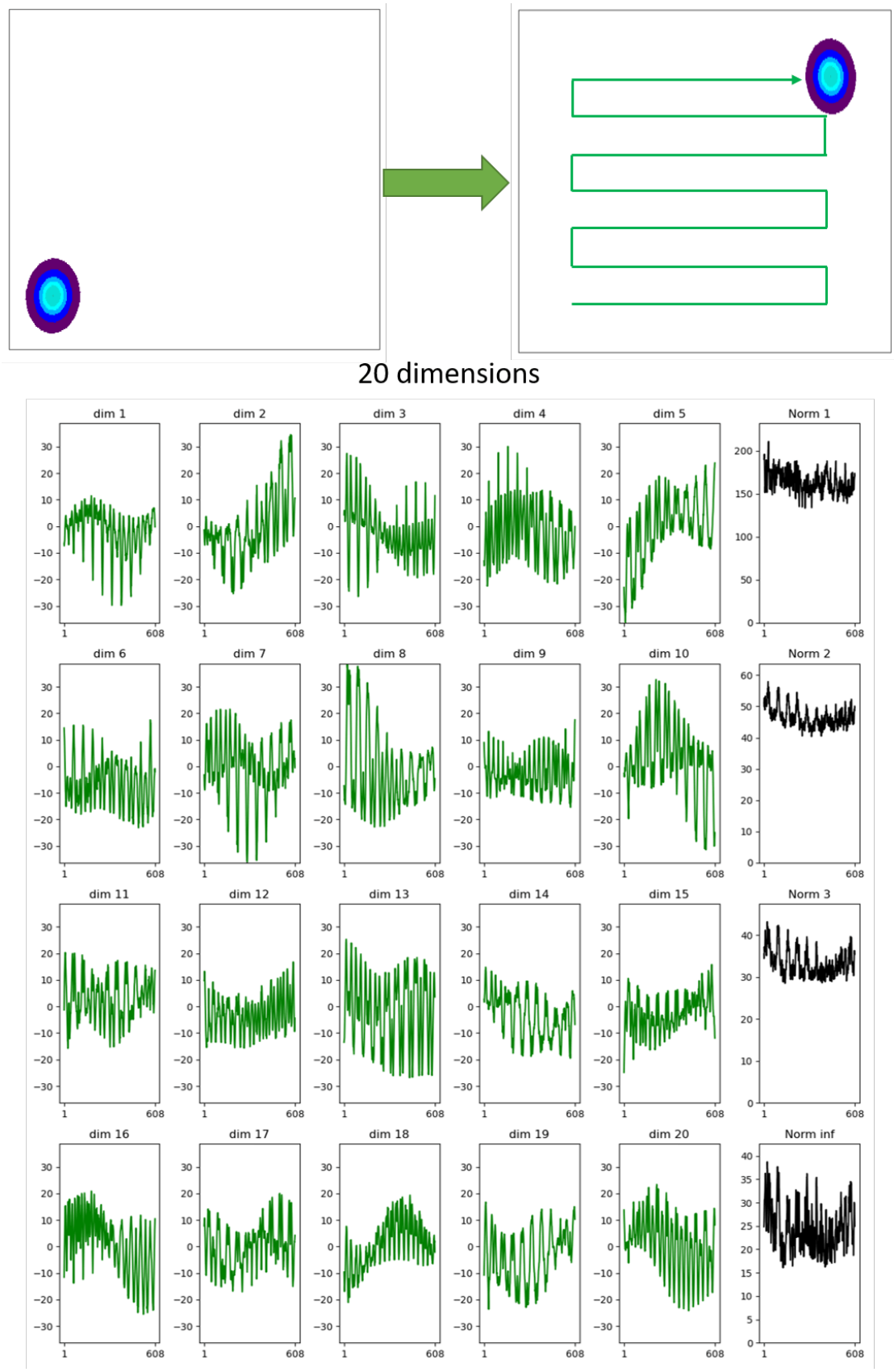


FIGURE D.14 – *Autoencodeur avec 20 dimensions, graphiques similaires à ceux de la figure 9.6 (variation de position).*

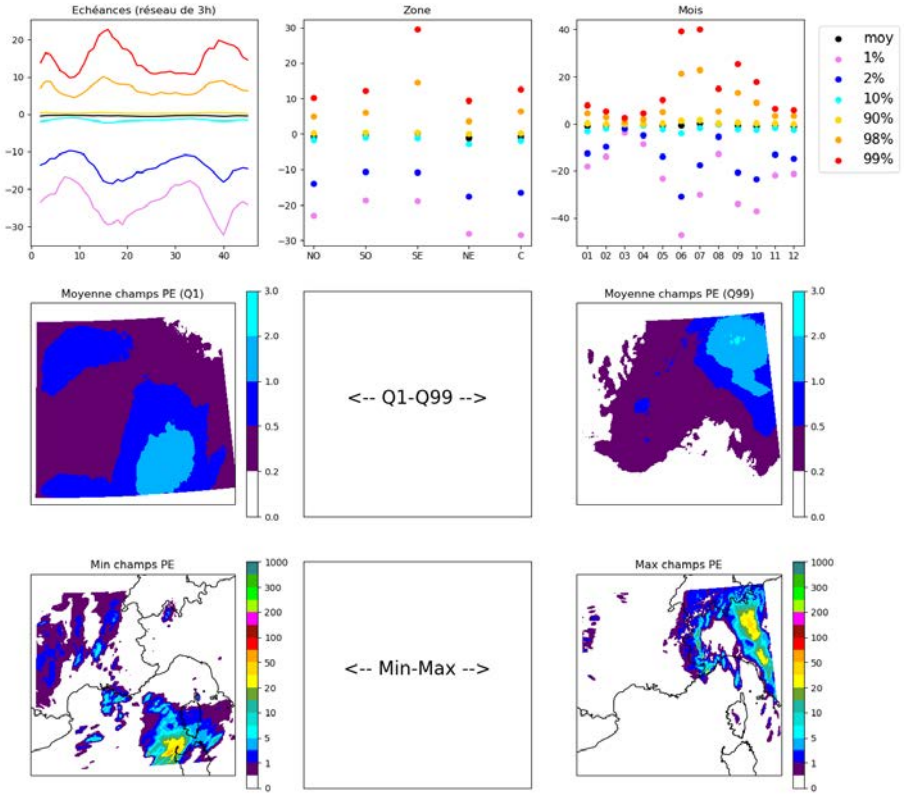


FIGURE D.15 – *Espace latent à 10 dimensions : dimension 1*

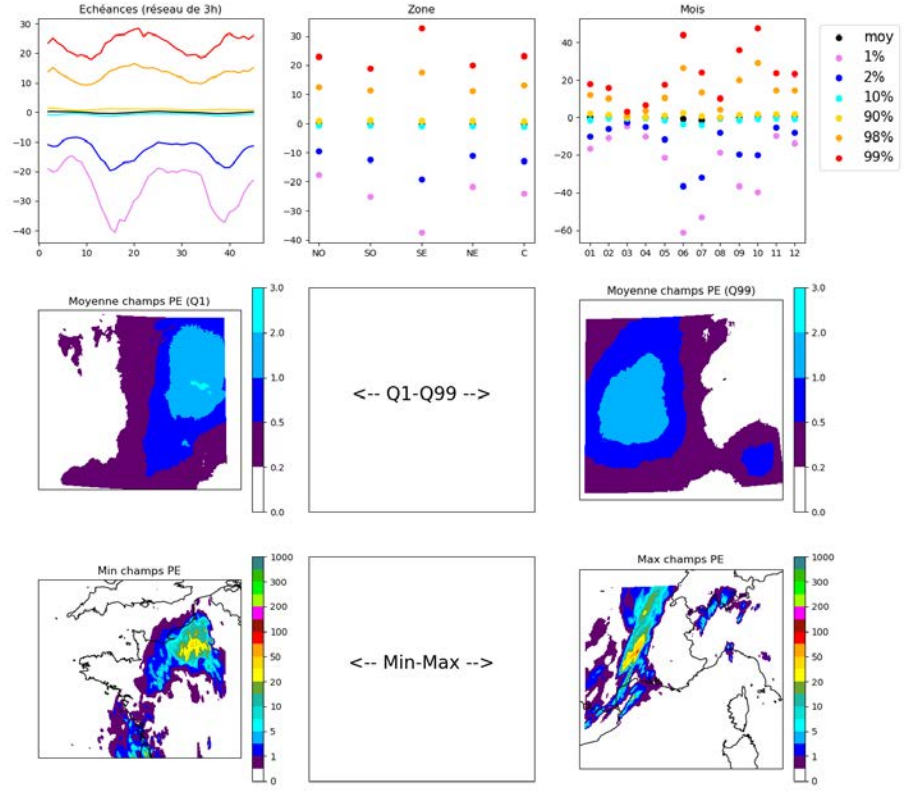


FIGURE D.16 – *Espace latent à 10 dimensions : dimension 2*

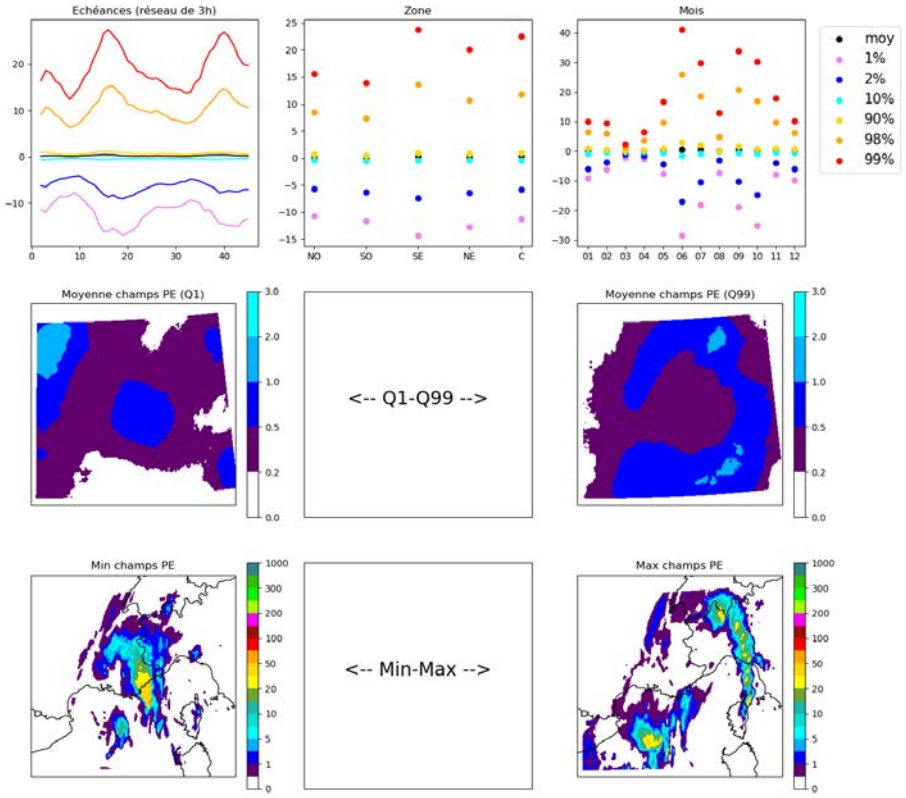


FIGURE D.17 – *Espace latent à 10 dimensions : dimension 3*

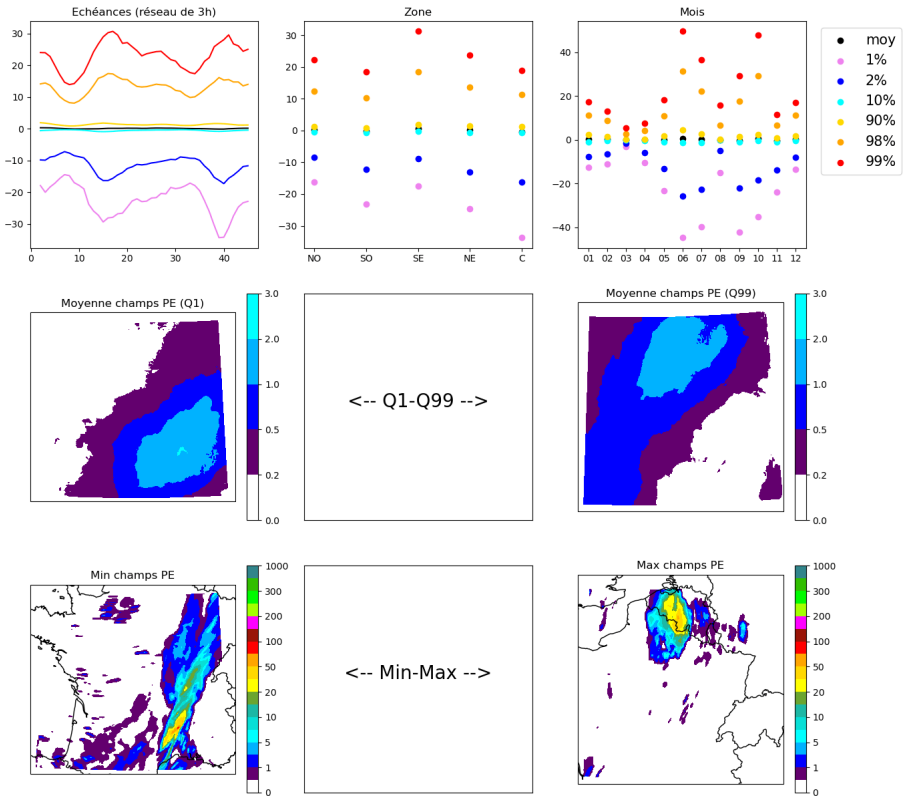


FIGURE D.18 – *Espace latent à 10 dimensions : dimension 4*

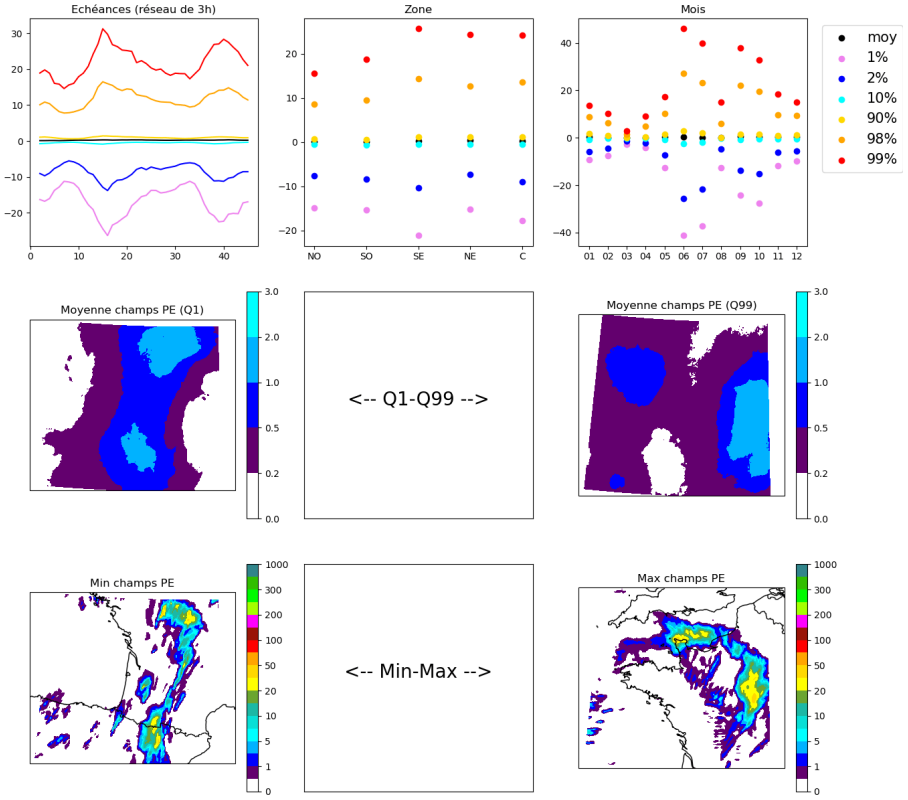


FIGURE D.19 – *Espace latent à 10 dimensions : dimension 5*

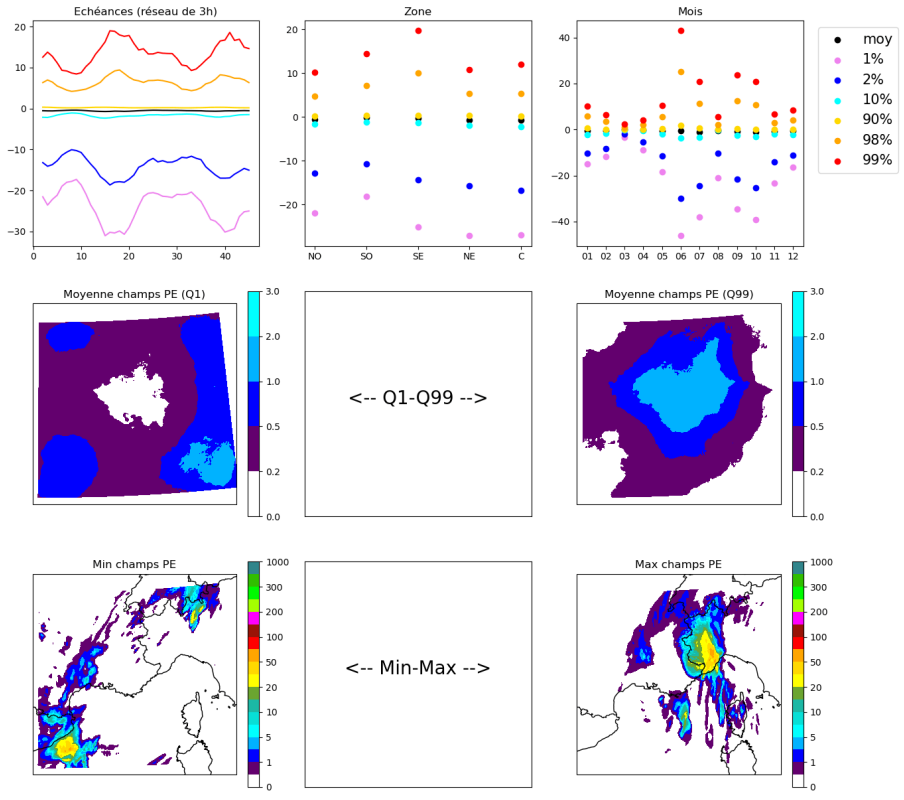


FIGURE D.20 – *Espace latent à 10 dimensions : dimension 6*

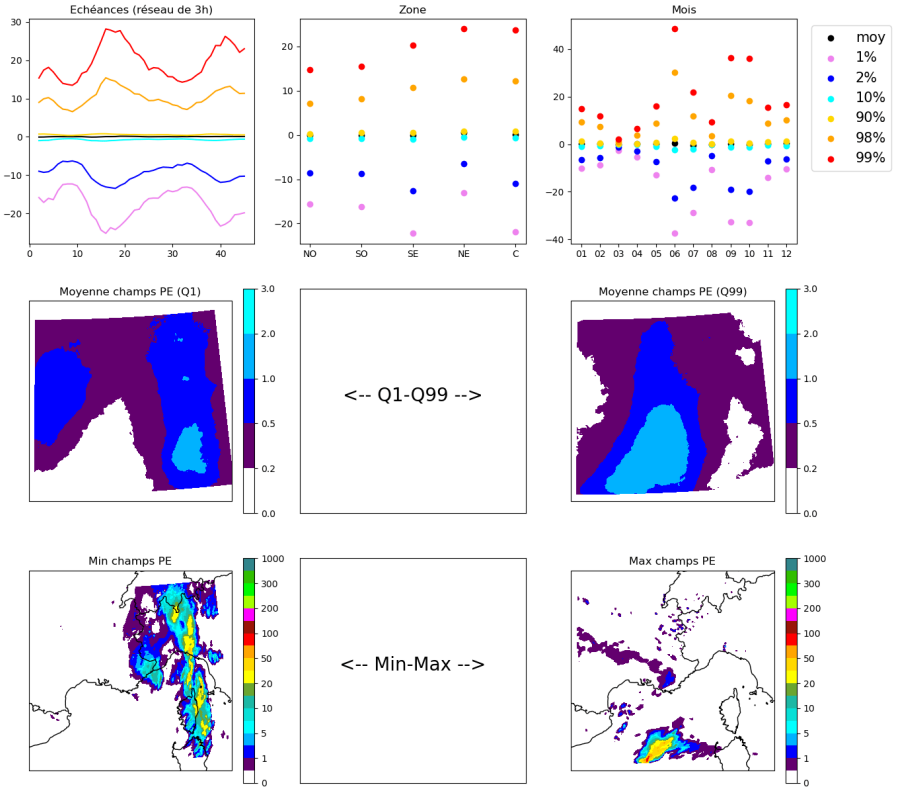


FIGURE D.21 – *Espace latent à 10 dimensions : dimension 7*

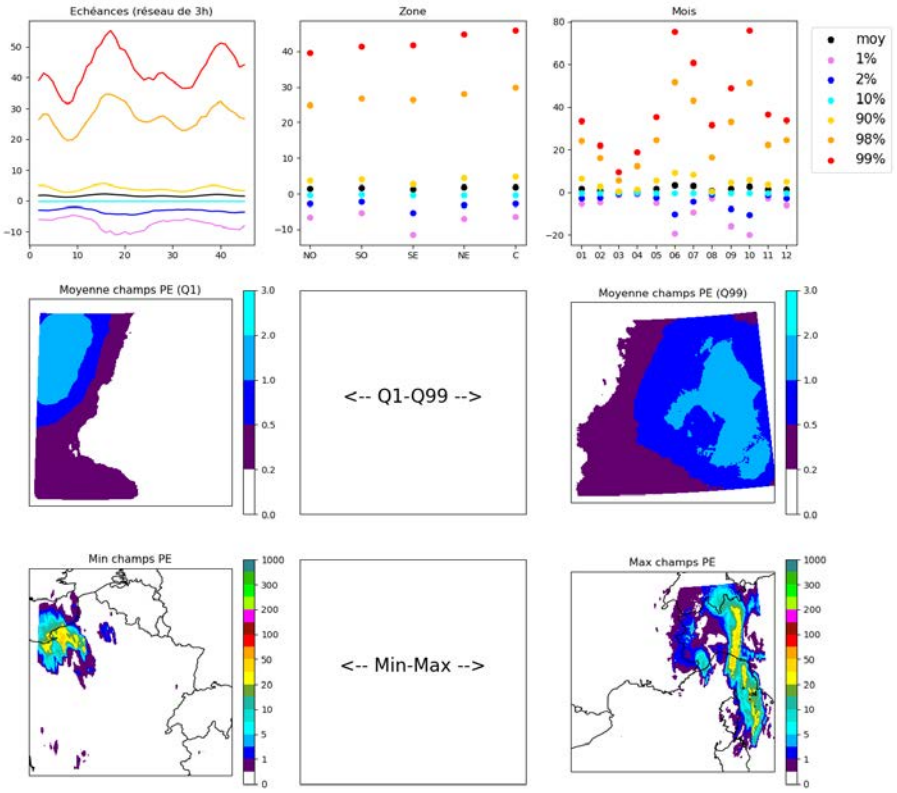


FIGURE D.22 – *Espace latent à 10 dimensions : dimension 8*

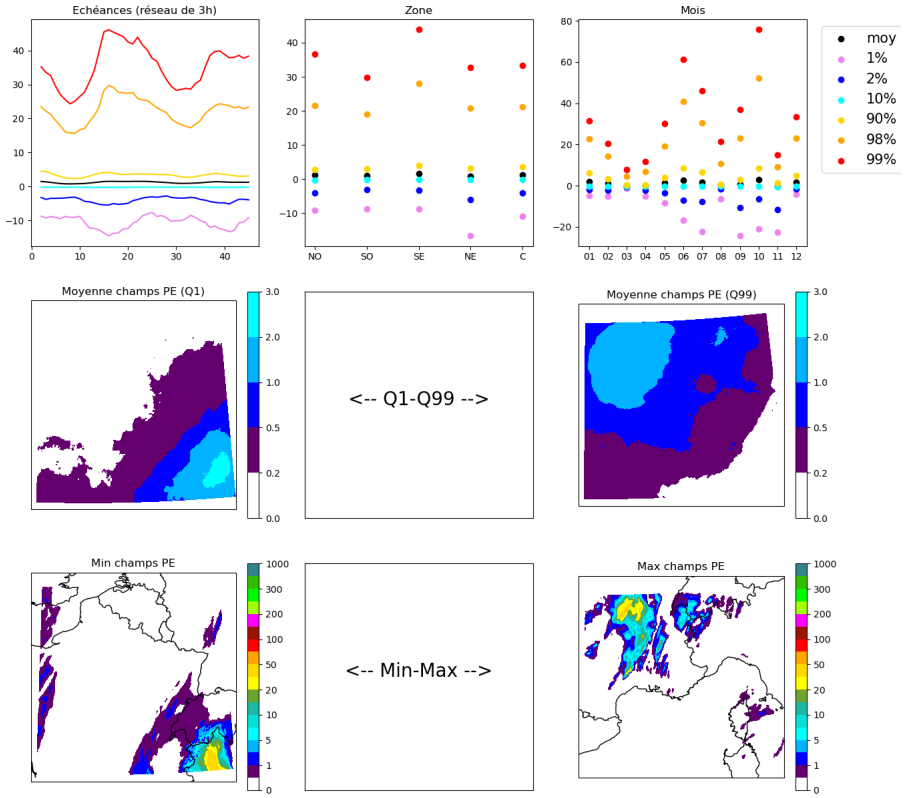


FIGURE D.23 – *Espace latent à 10 dimensions : dimension 9*

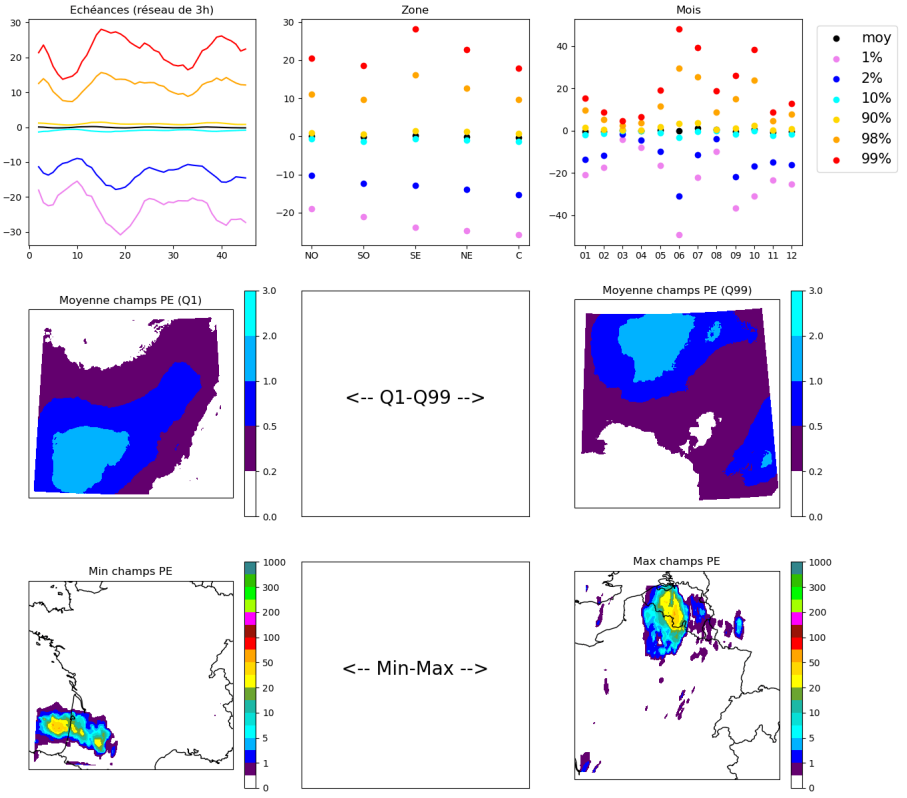


FIGURE D.24 – *Espace latent à 10 dimensions : dimension 10*

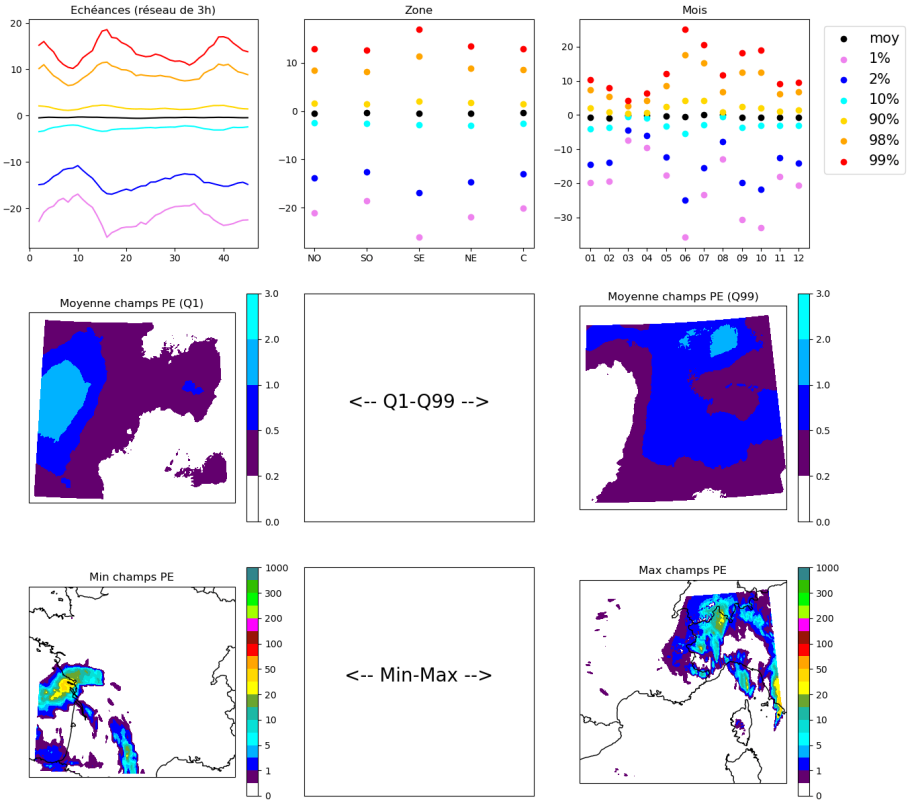


FIGURE D.25 – *Espace latent à 20 dimensions : dimension 1*

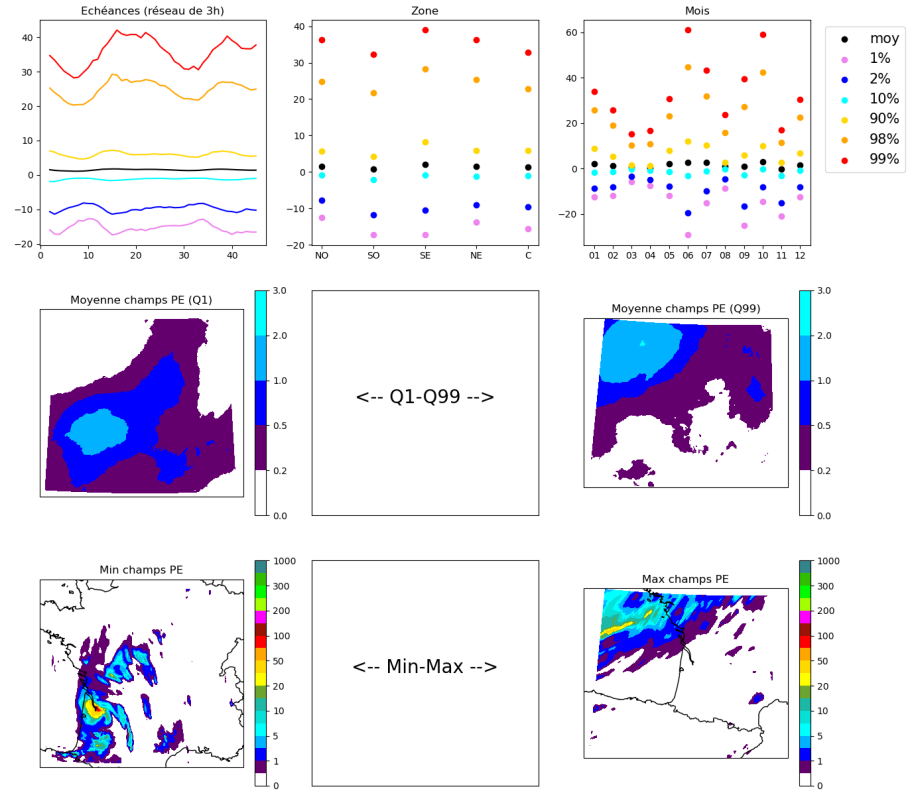


FIGURE D.26 – *Espace latent à 20 dimensions : dimension 2*

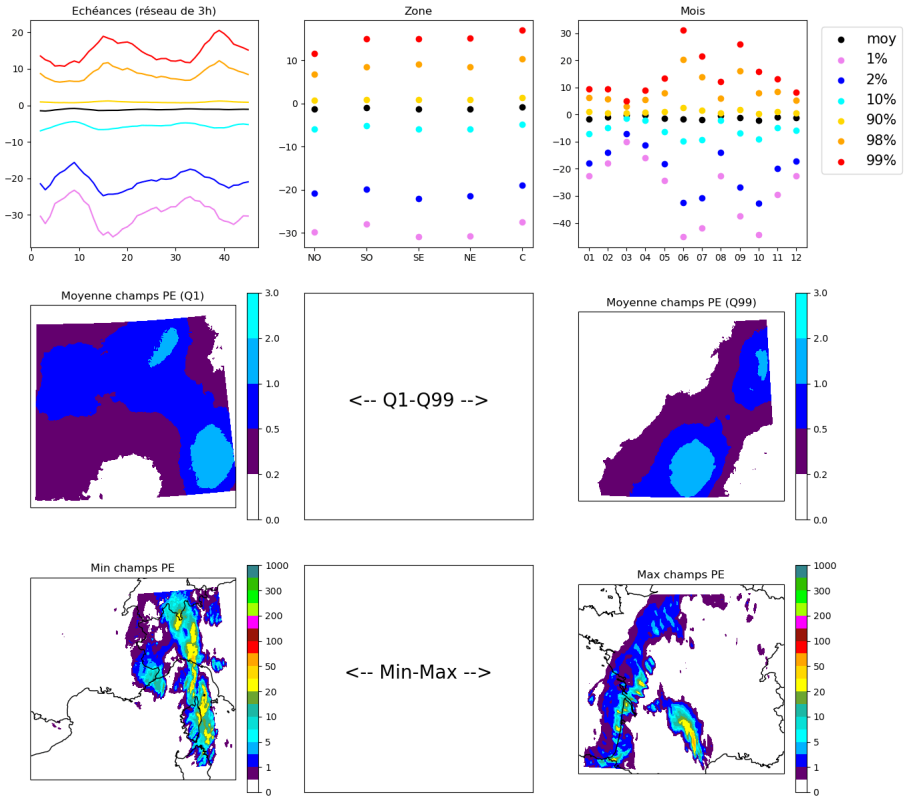


FIGURE D.27 – *Espace latent à 20 dimensions : dimension 3*

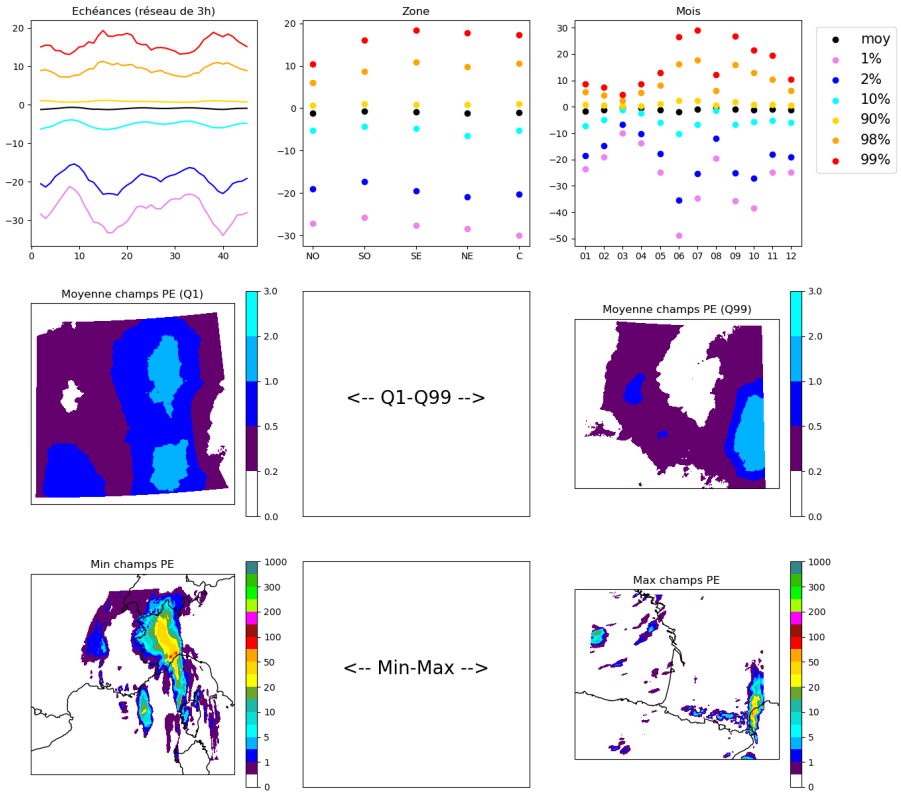


FIGURE D.28 – *Espace latent à 20 dimensions : dimension 4*

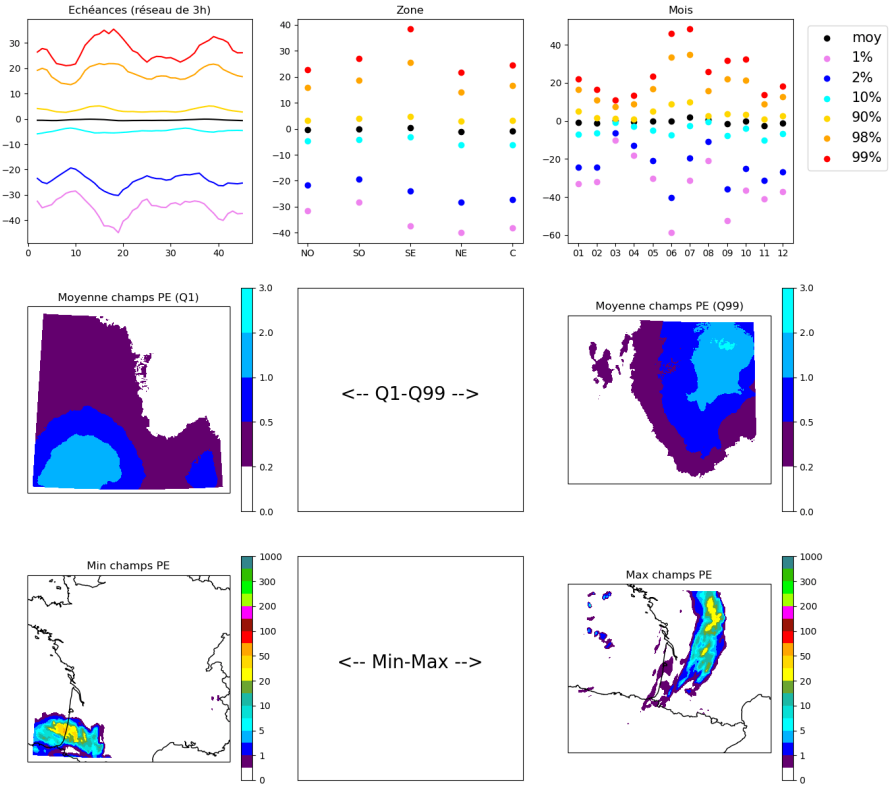


FIGURE D.29 – *Espace latent à 20 dimensions : dimension 5*

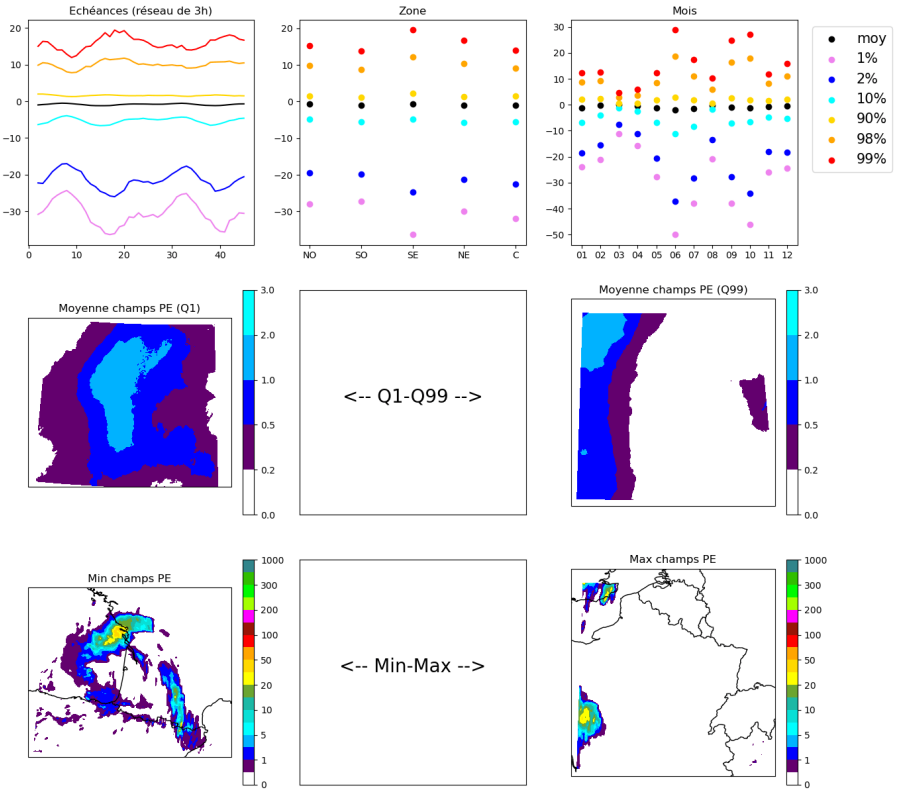


FIGURE D.30 – *Espace latent à 20 dimensions : dimension 6*

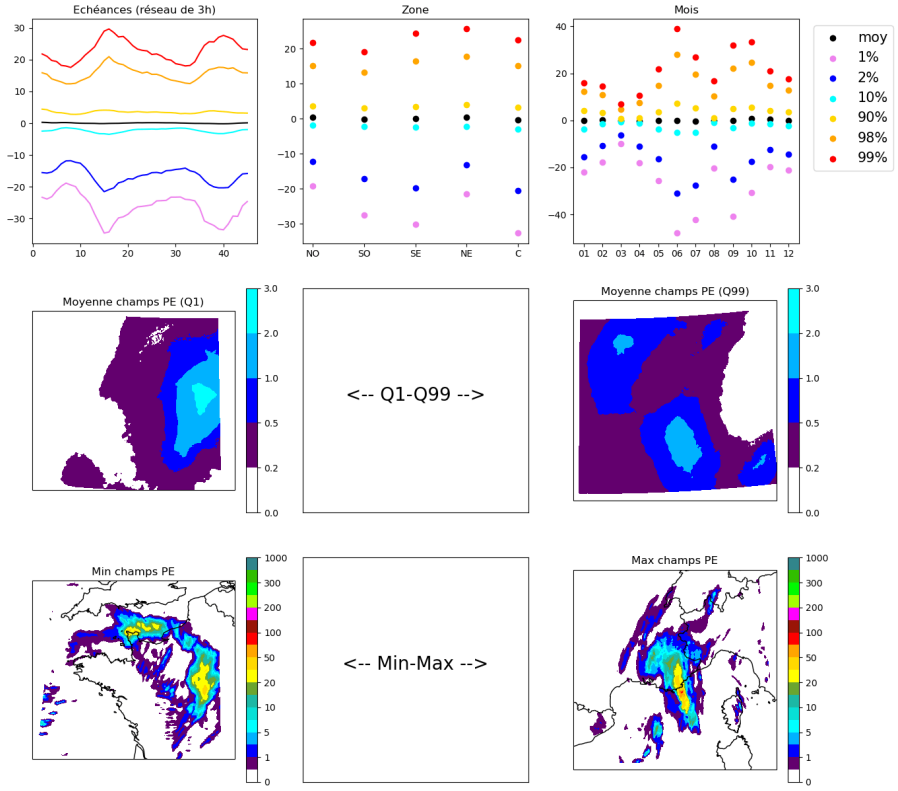


FIGURE D.31 – *Espace latent à 20 dimensions : dimension 7*

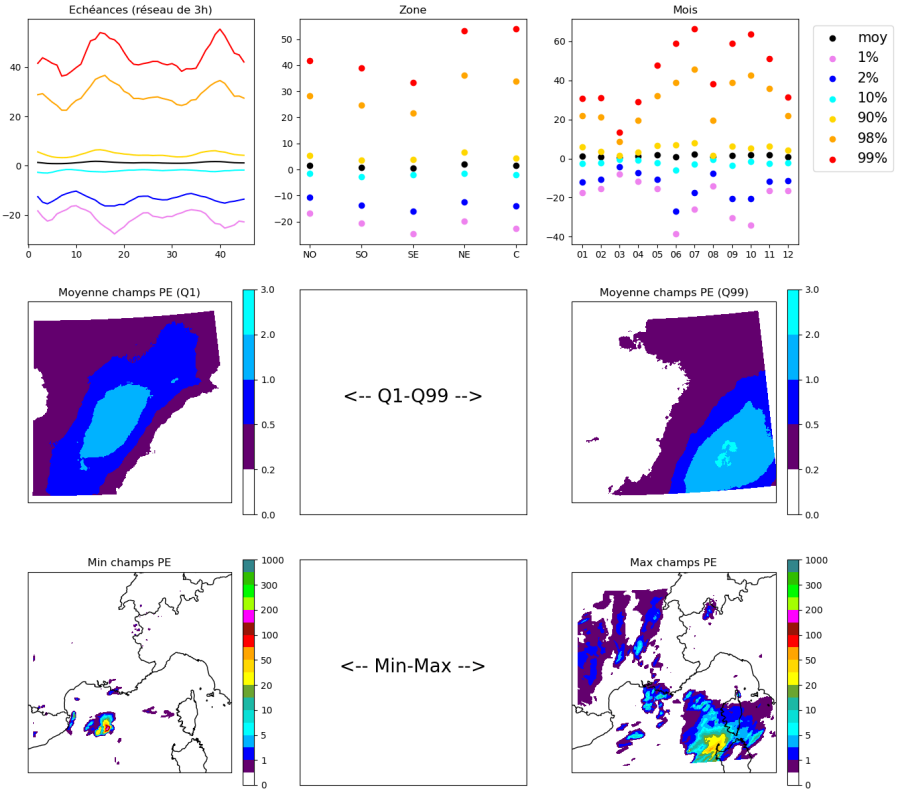


FIGURE D.32 – *Espace latent à 20 dimensions : dimension 8*

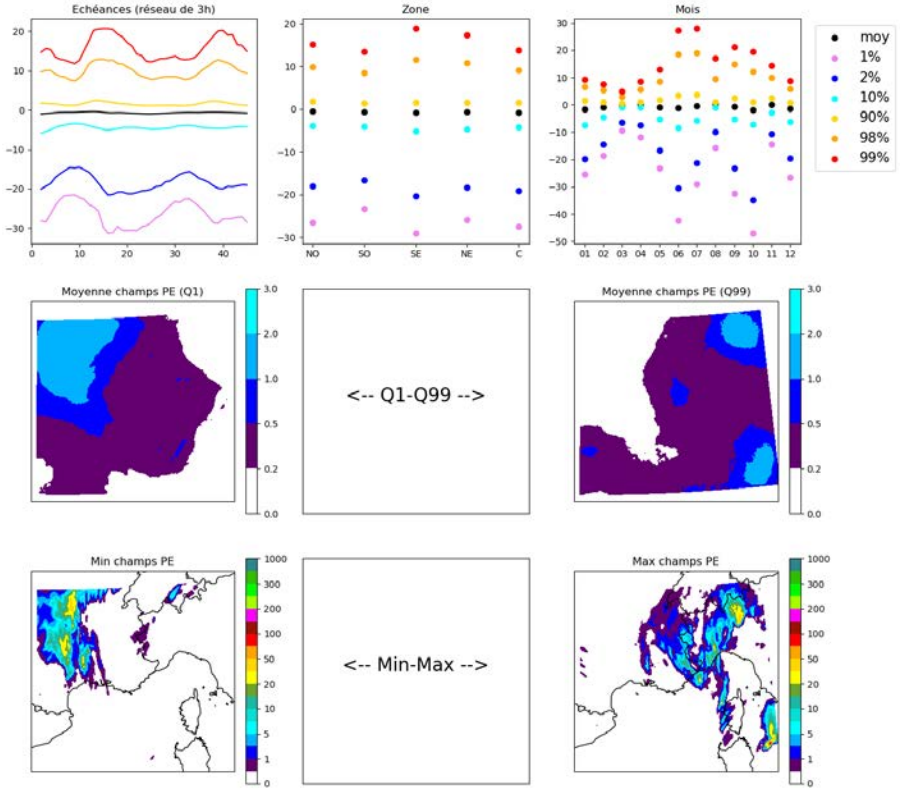


FIGURE D.33 – *Espace latent à 20 dimensions : dimension 9*

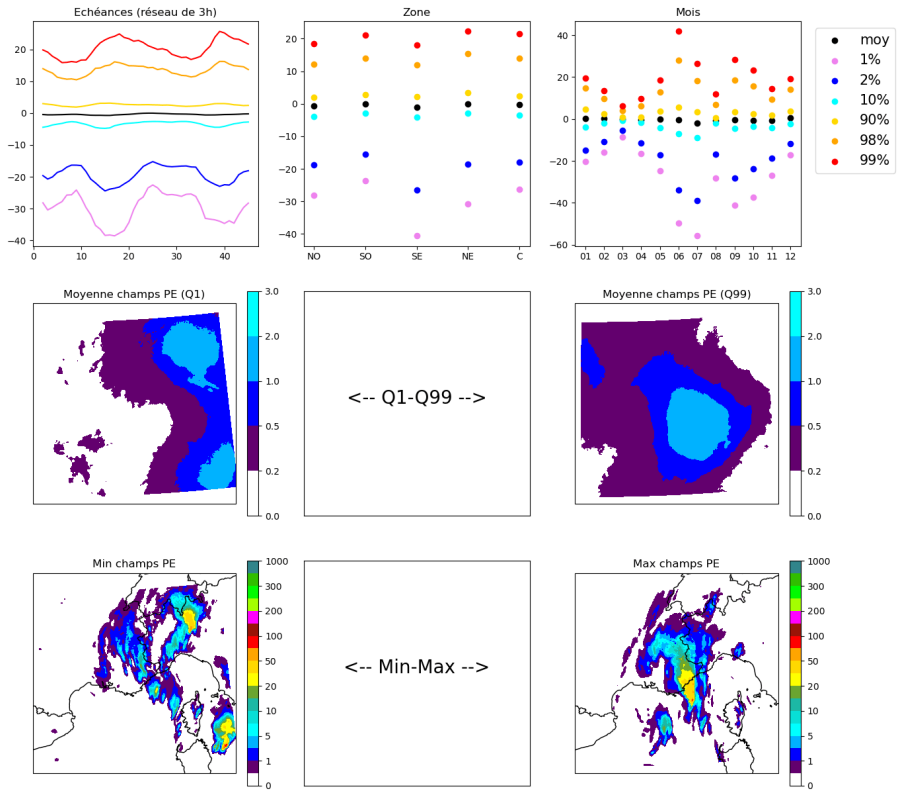


FIGURE D.34 – *Espace latent à 20 dimensions : dimension 10*

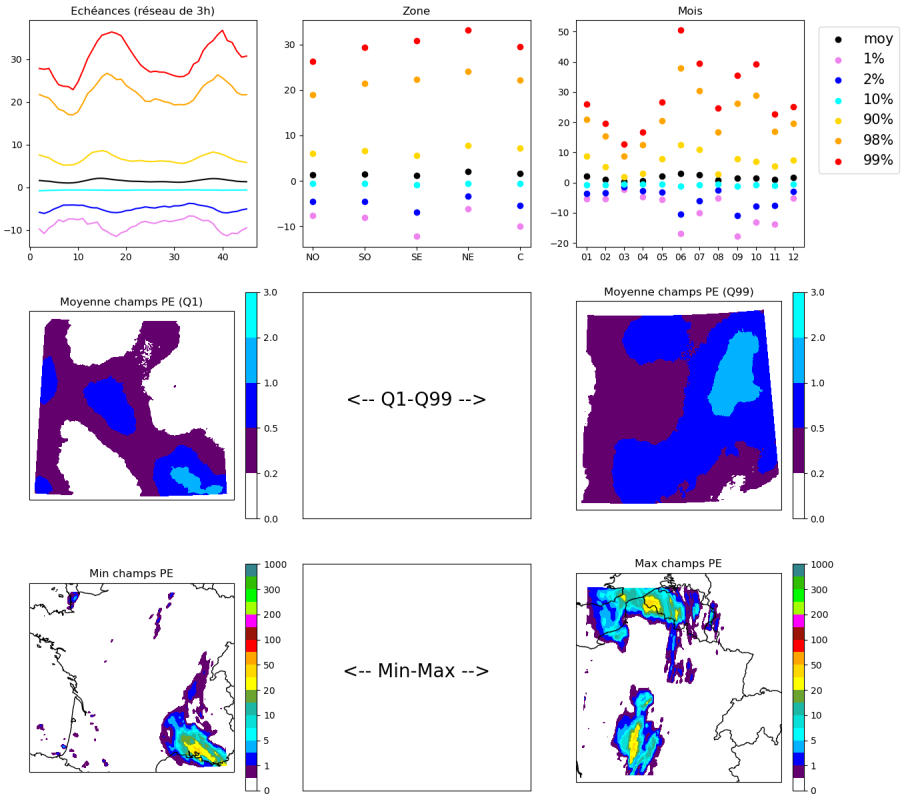


FIGURE D.35 – *Espace latent à 20 dimensions : dimension 11*

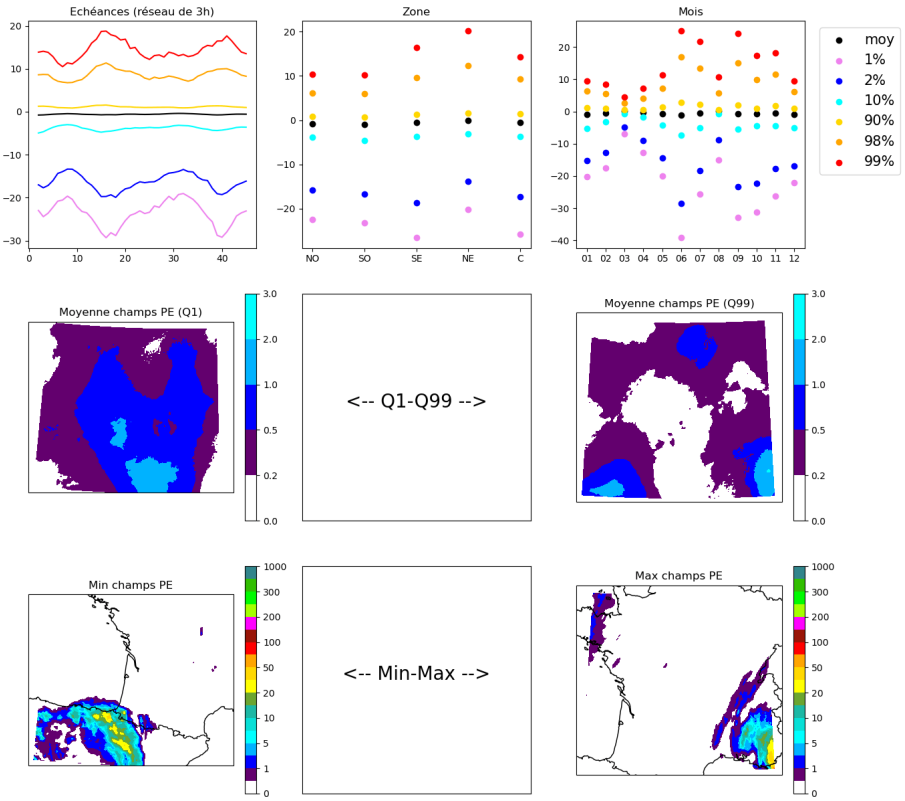


FIGURE D.36 – *Espace latent à 20 dimensions : dimension 12*

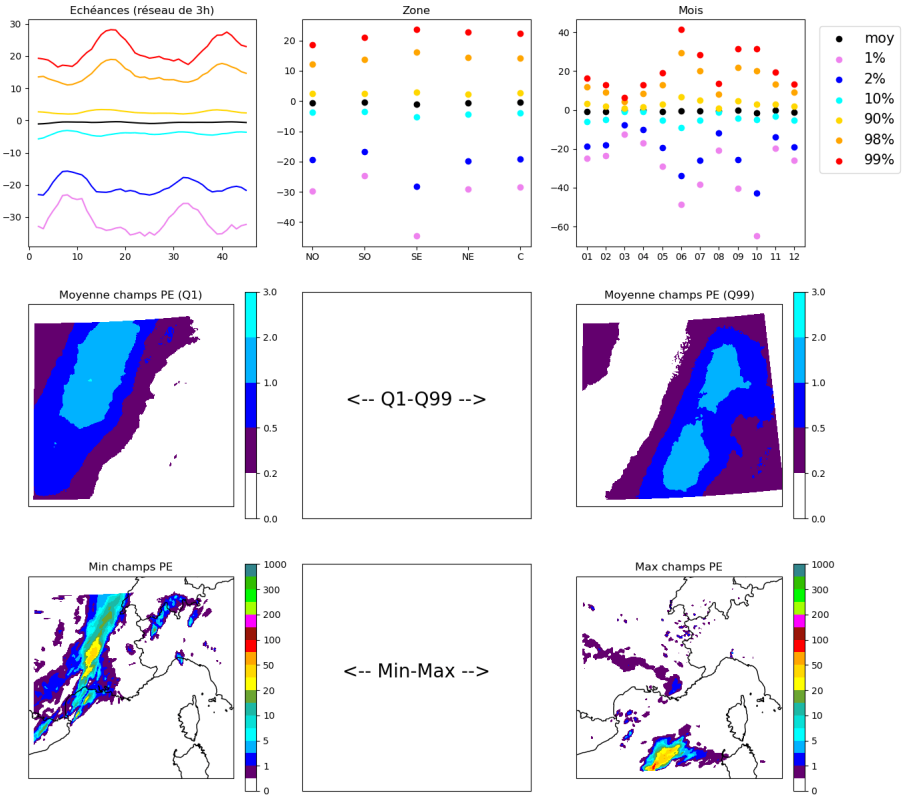


FIGURE D.37 – *Espace latent à 20 dimensions : dimension 13*

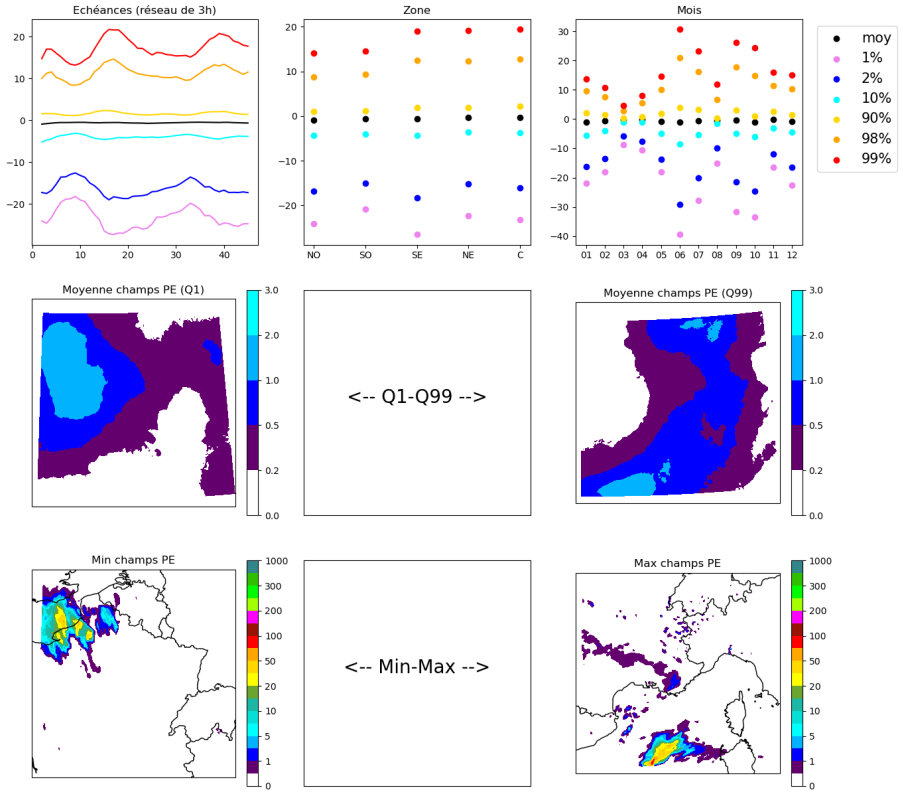


FIGURE D.38 – *Espace latent à 20 dimensions : dimension 14*

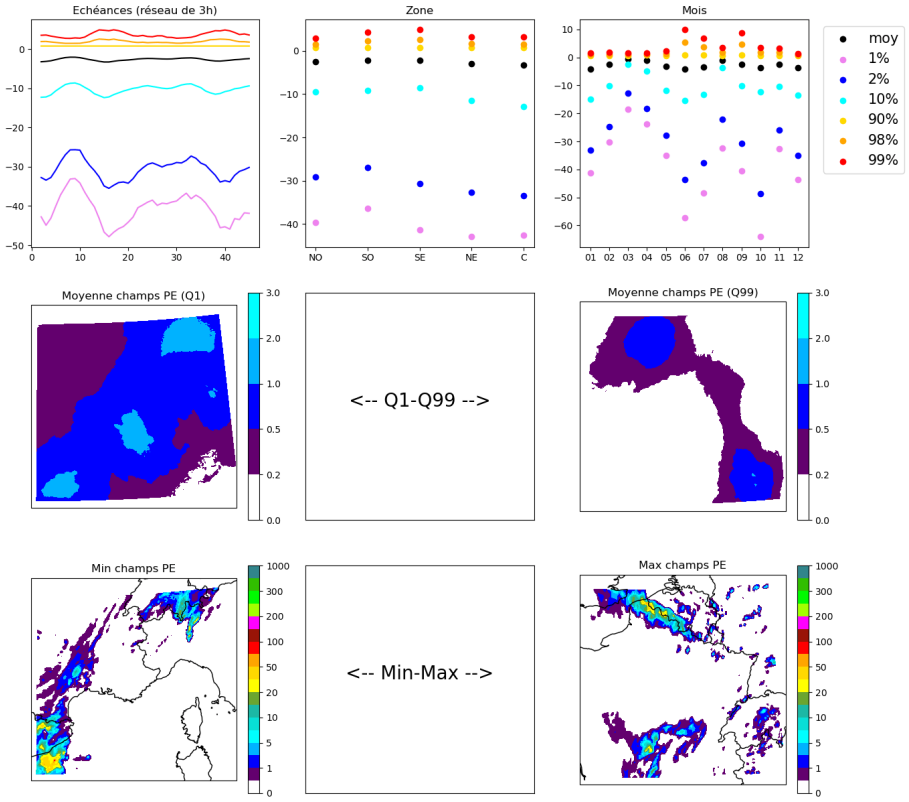


FIGURE D.39 – *Espace latent à 20 dimensions : dimension 15*

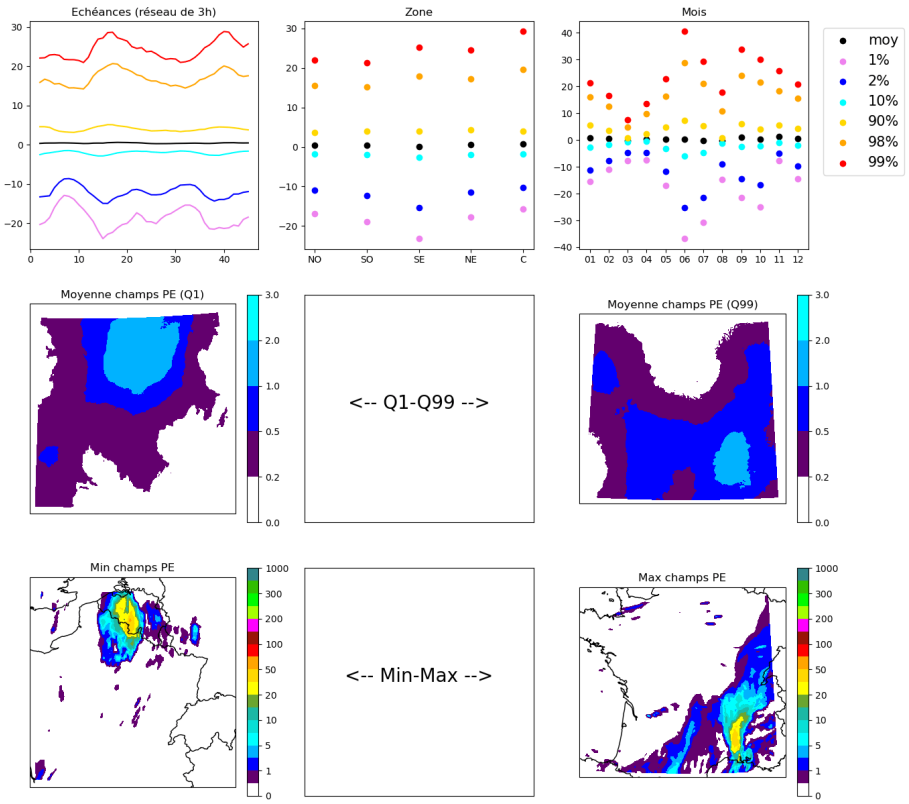


FIGURE D.40 – *Espace latent à 20 dimensions : dimension 16*

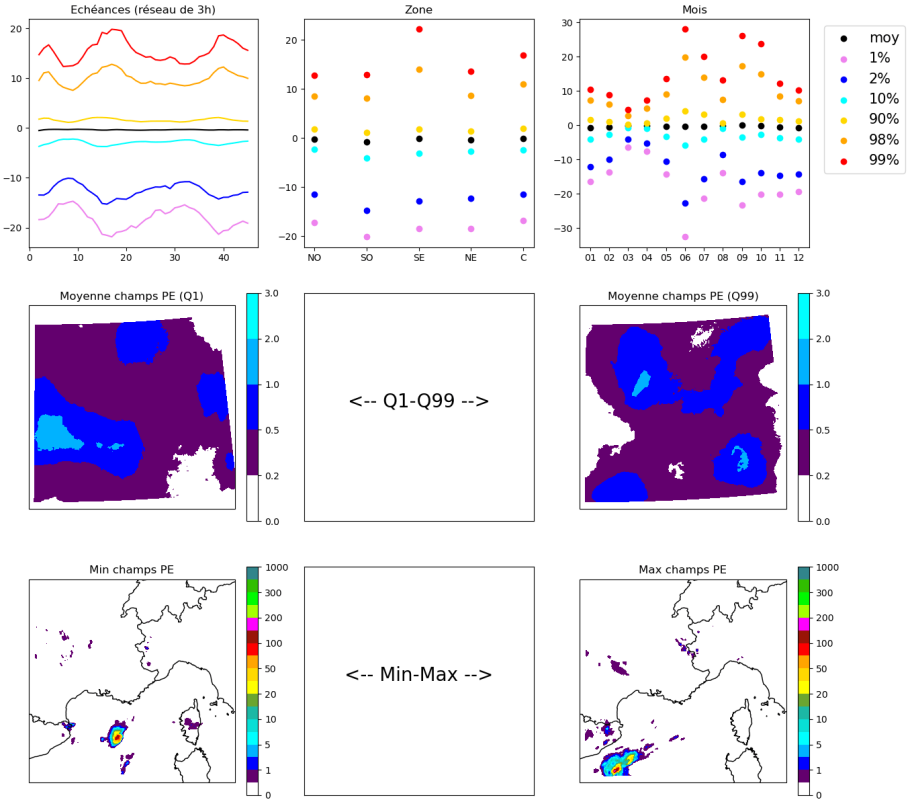


FIGURE D.41 – *Espace latent à 20 dimensions : dimension 17*

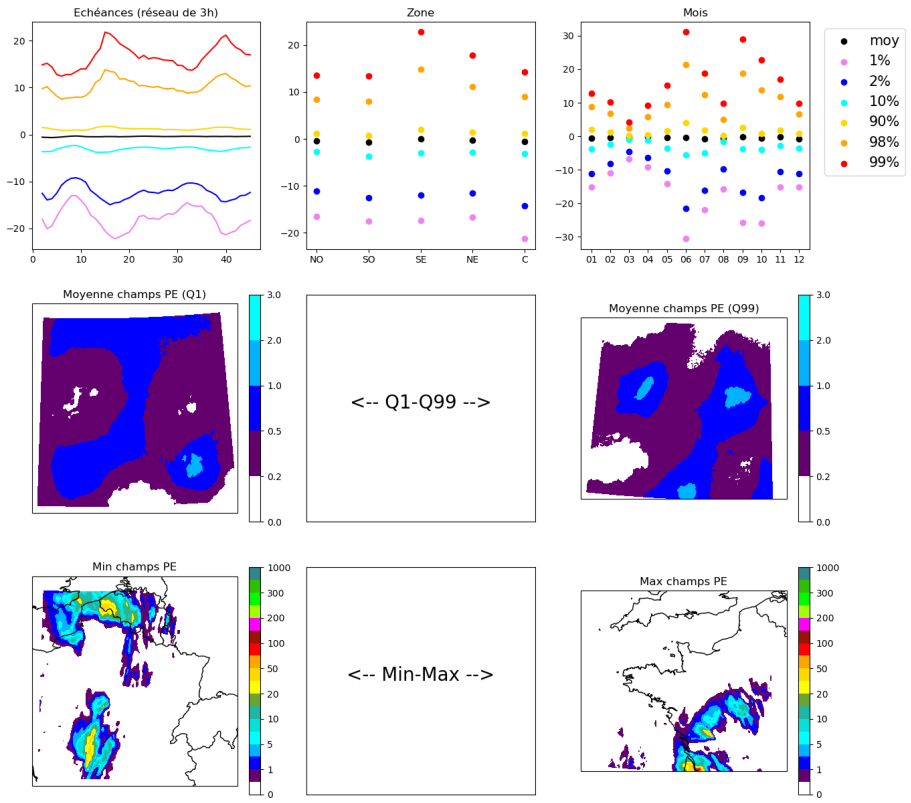


FIGURE D.42 – *Espace latent à 20 dimensions : dimension 18*

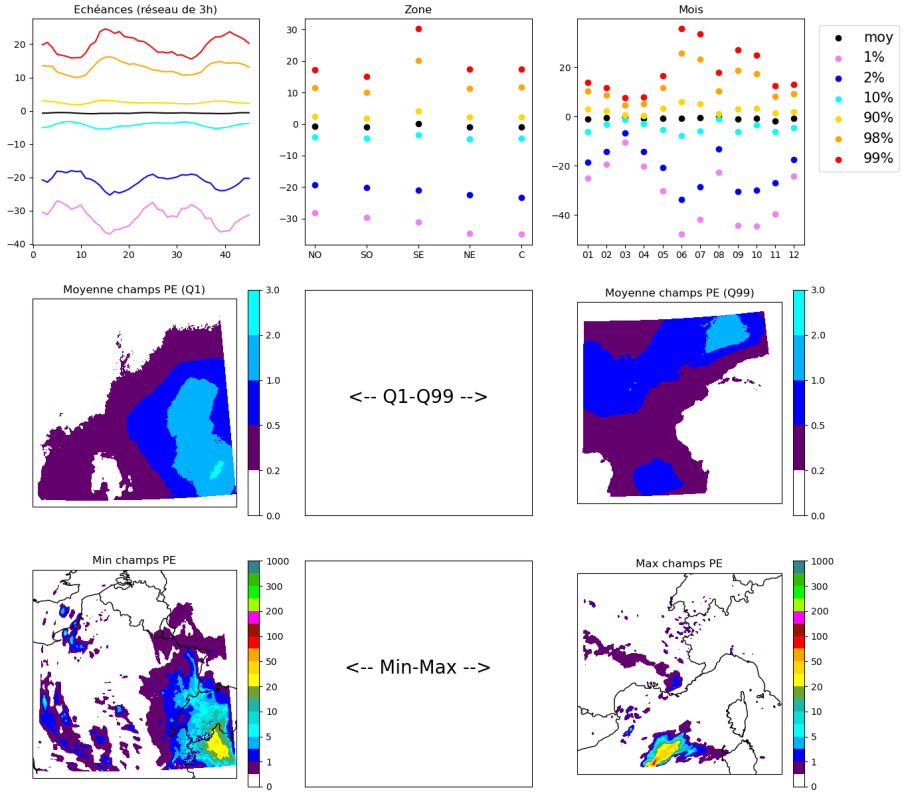


FIGURE D.43 – *Espace latent à 20 dimensions : dimension 19*

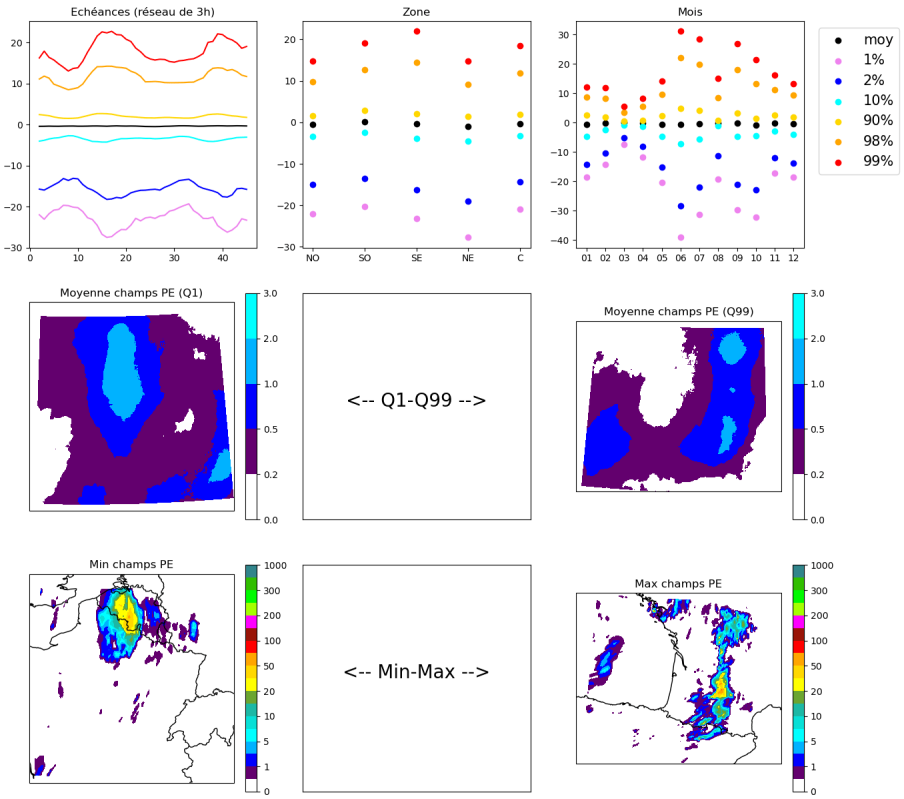


FIGURE D.44 – *Espace latent à 20 dimensions : dimension 20*

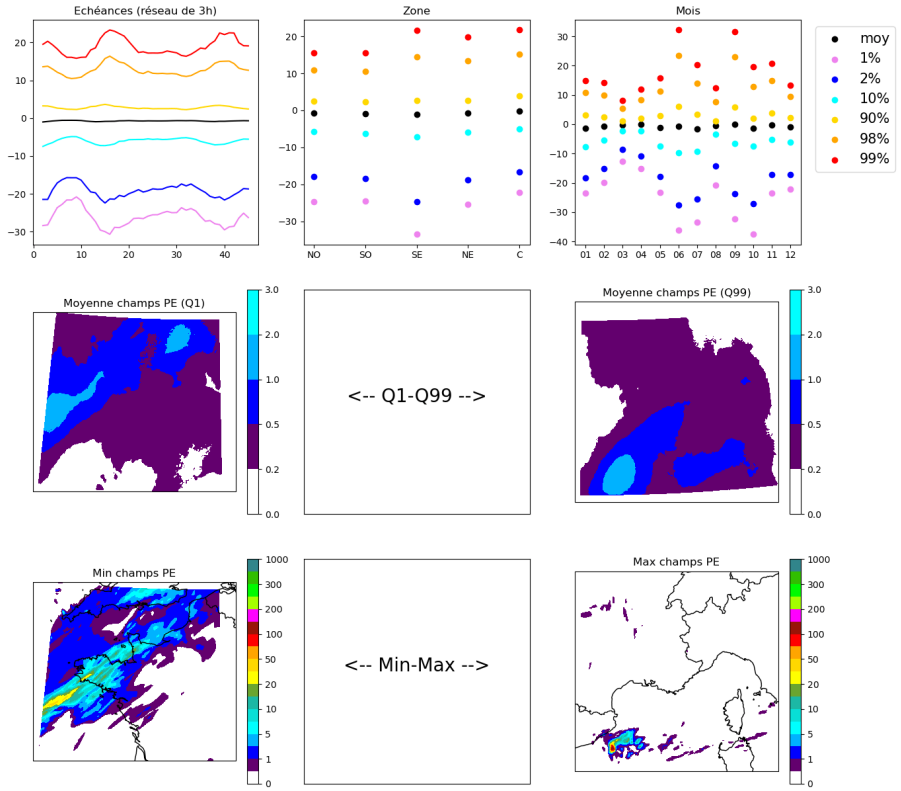


FIGURE D.45 – *Espace latent à 40 dimensions : dimension 1*

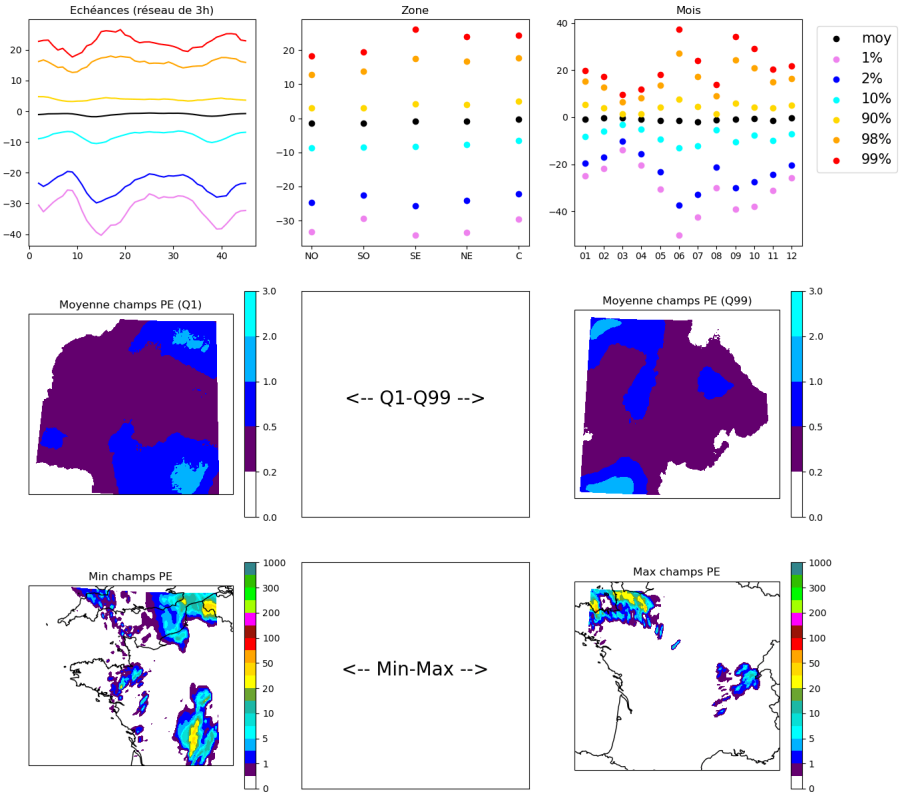


FIGURE D.46 – *Espace latent à 40 dimensions : dimension 2*

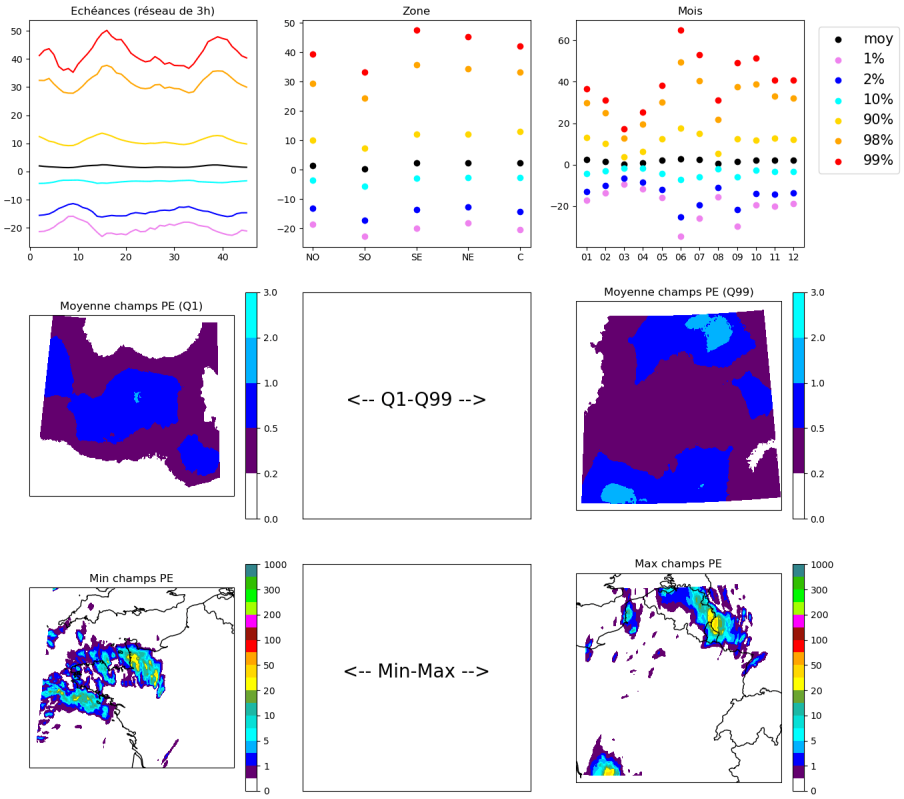


FIGURE D.47 – *Espace latent à 40 dimensions : dimension 3*

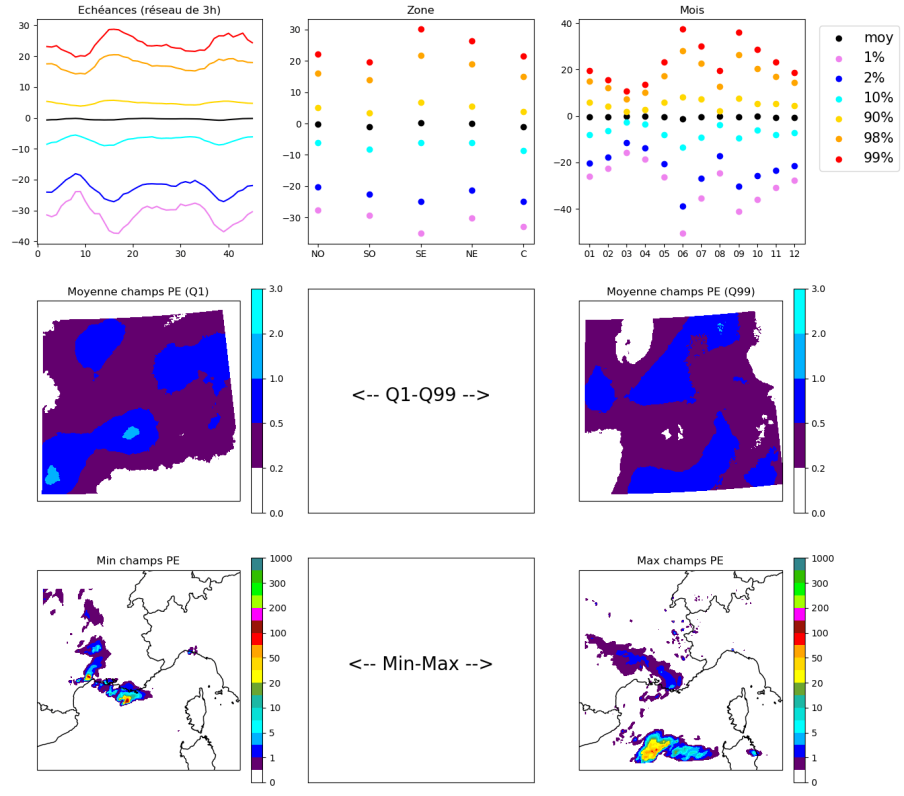


FIGURE D.48 – *Espace latent à 40 dimensions : dimension 4*

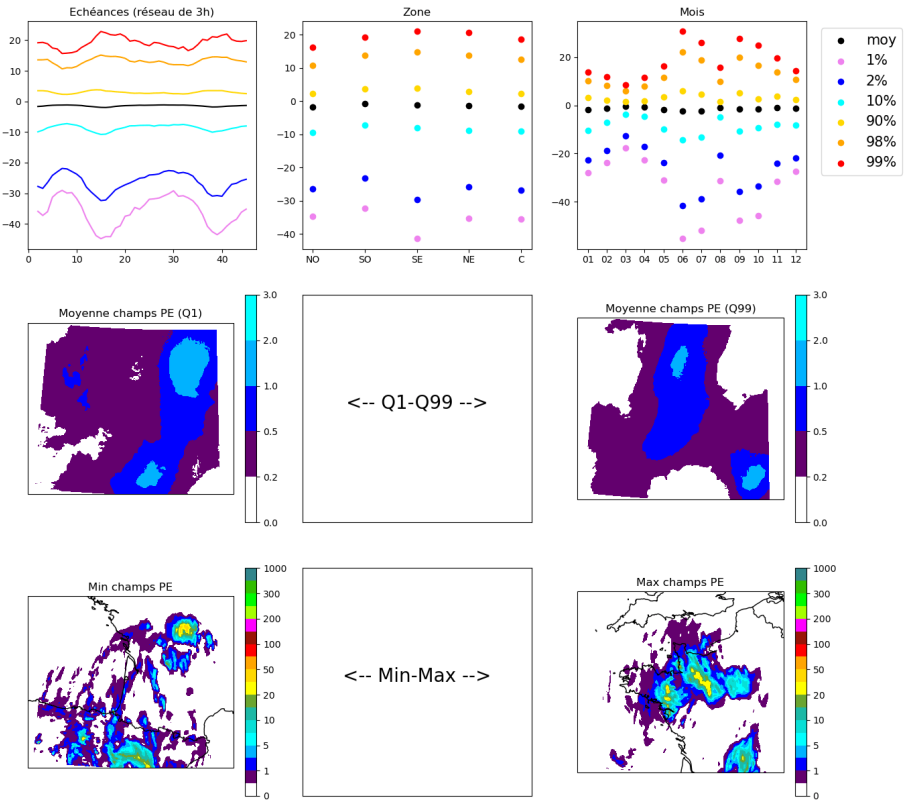


FIGURE D.49 – *Espace latent à 40 dimensions : dimension 5*

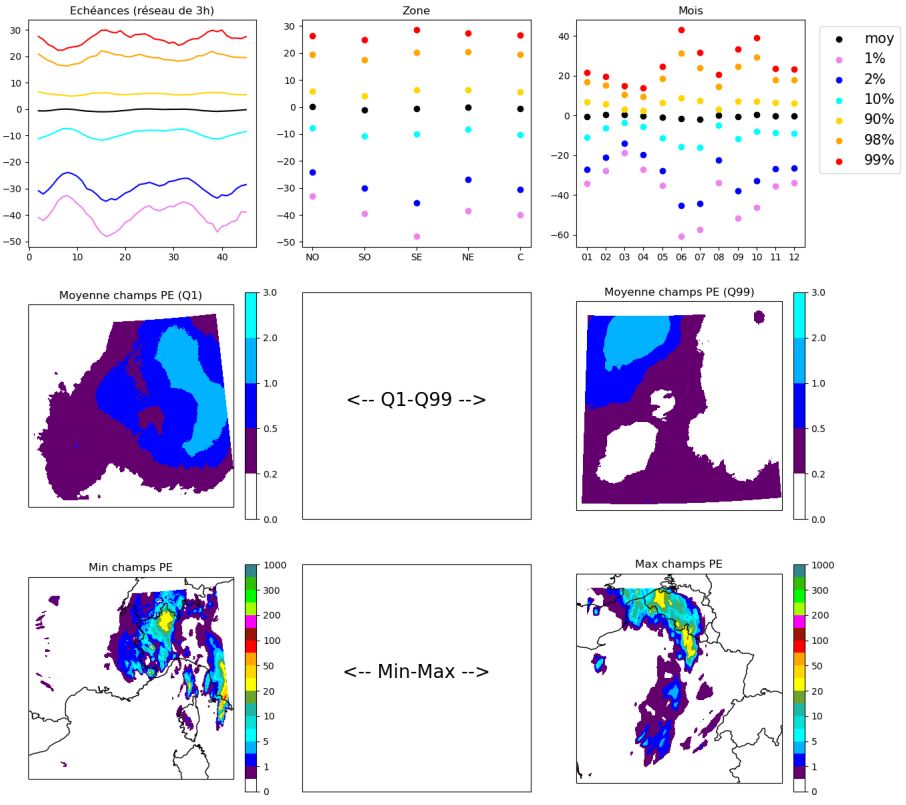


FIGURE D.50 – *Espace latent à 40 dimensions : dimension 6*

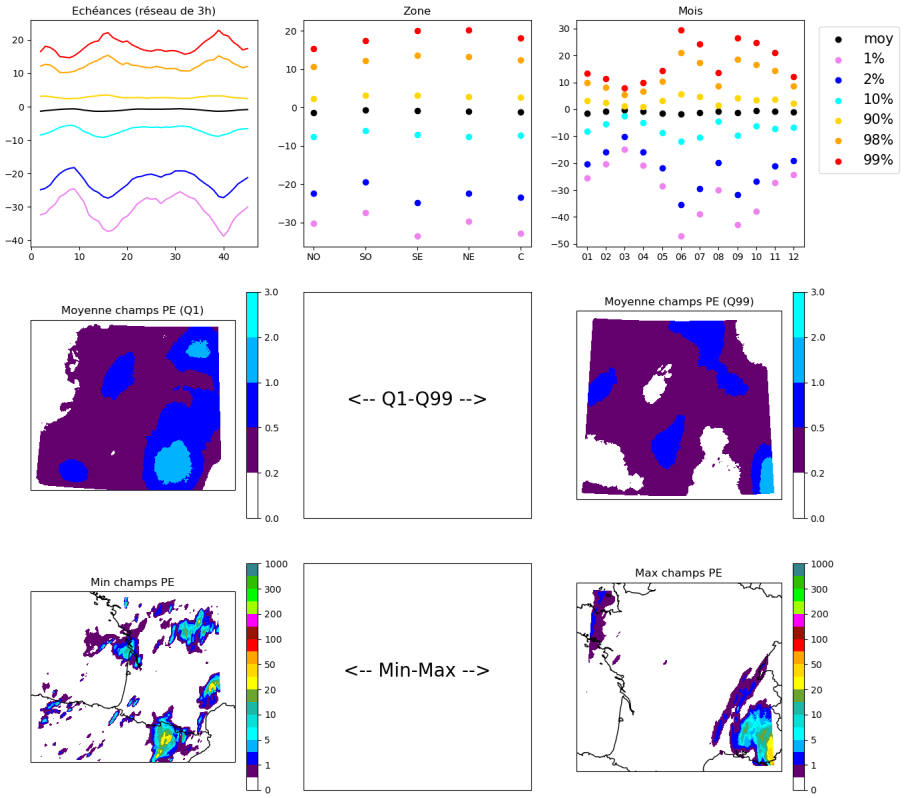


FIGURE D.51 – *Espace latent à 40 dimensions : dimension 7*

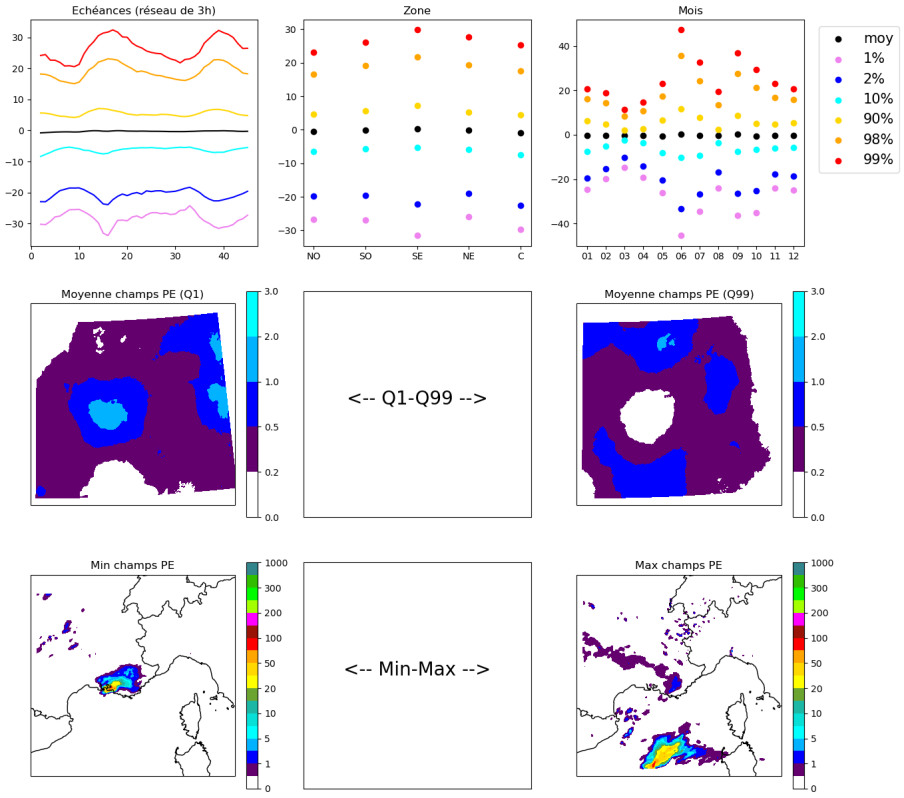


FIGURE D.52 – *Espace latent à 40 dimensions : dimension 8*

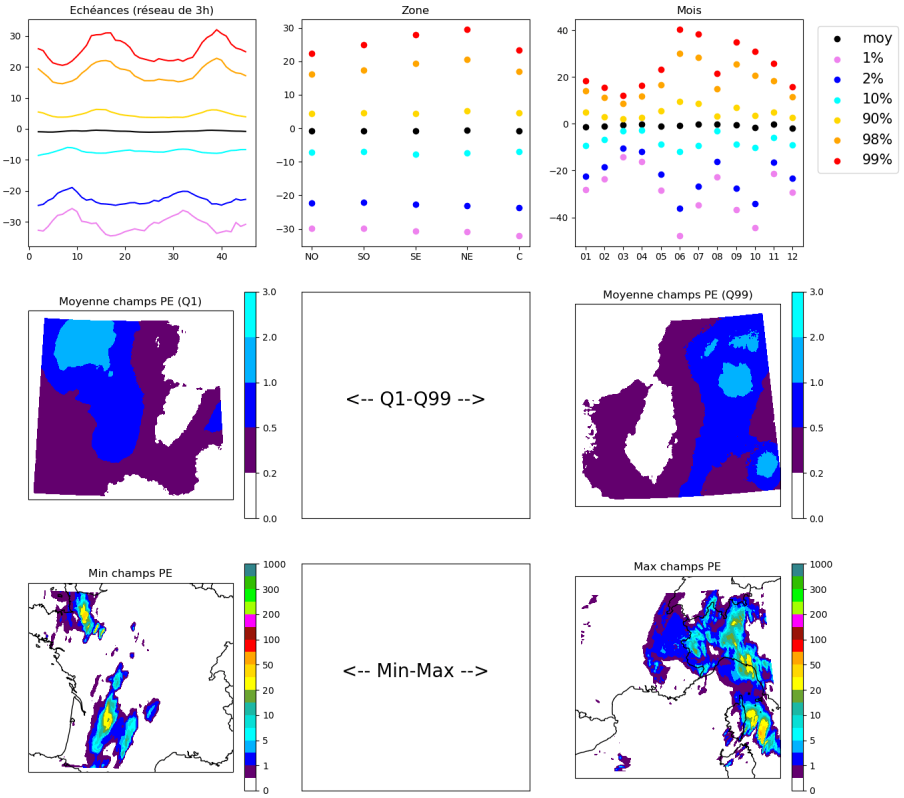


FIGURE D.53 – *Espace latent à 40 dimensions : dimension 9*

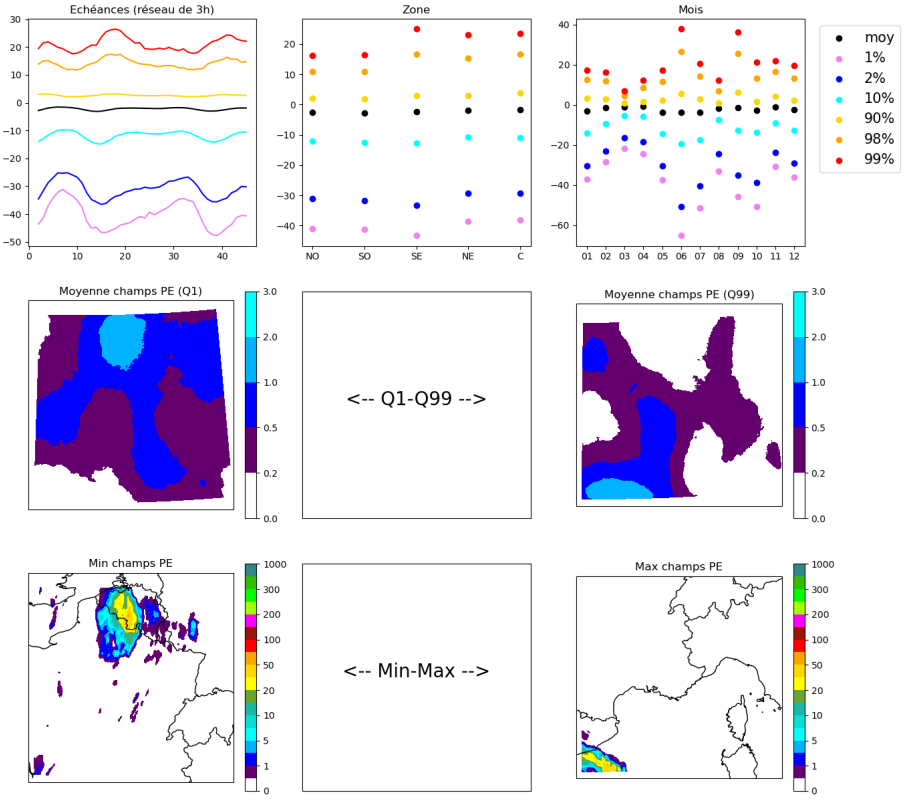


FIGURE D.54 – *Espace latent à 40 dimensions : dimension 10*

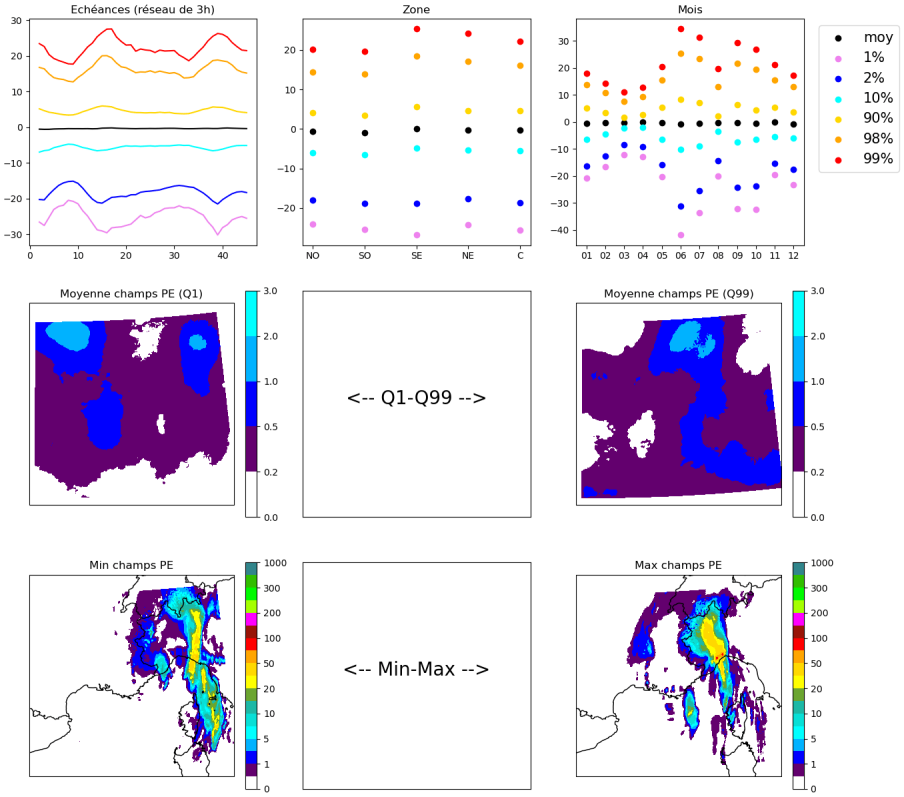


FIGURE D.55 – *Espace latent à 40 dimensions : dimension 11*

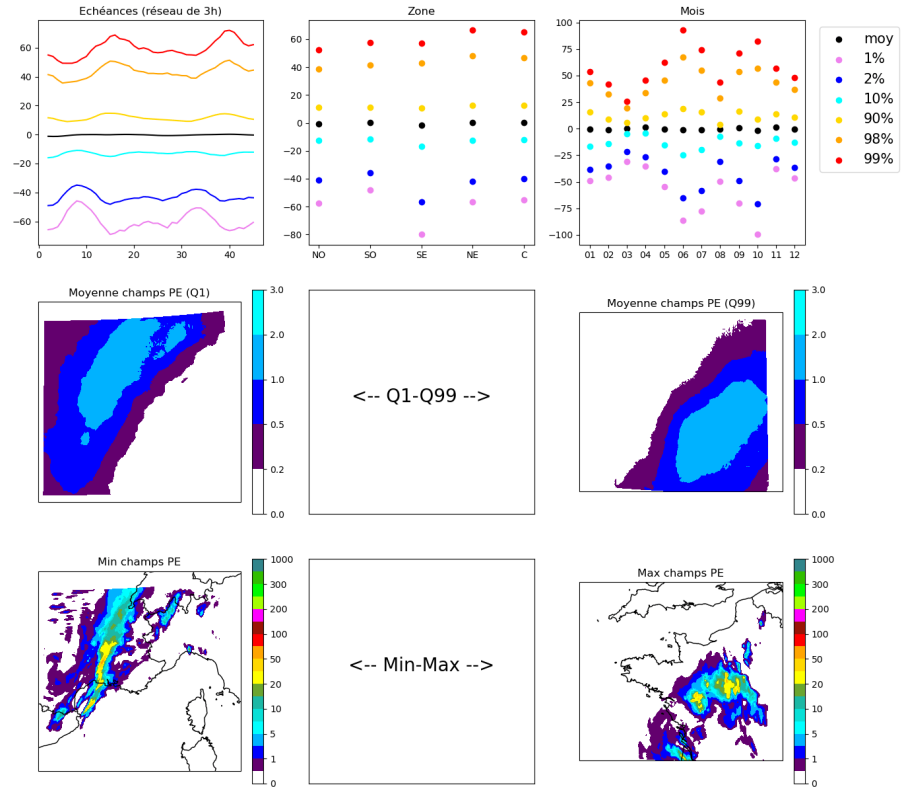


FIGURE D.56 – *Espace latent à 40 dimensions : dimension 12*

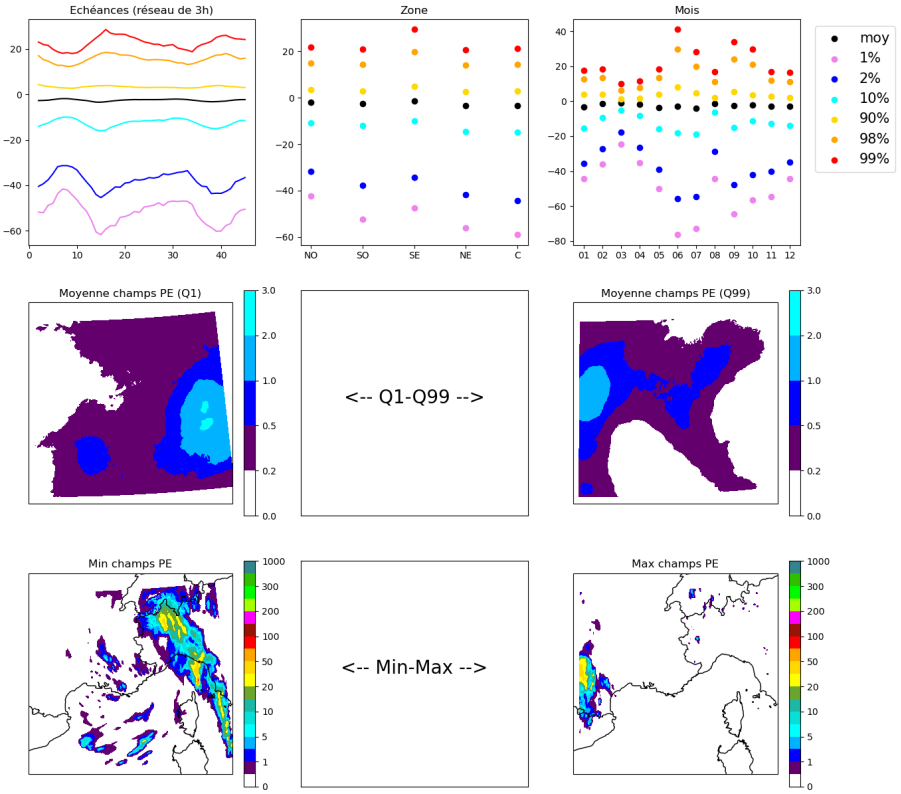


FIGURE D.57 – *Espace latent à 40 dimensions : dimension 13*

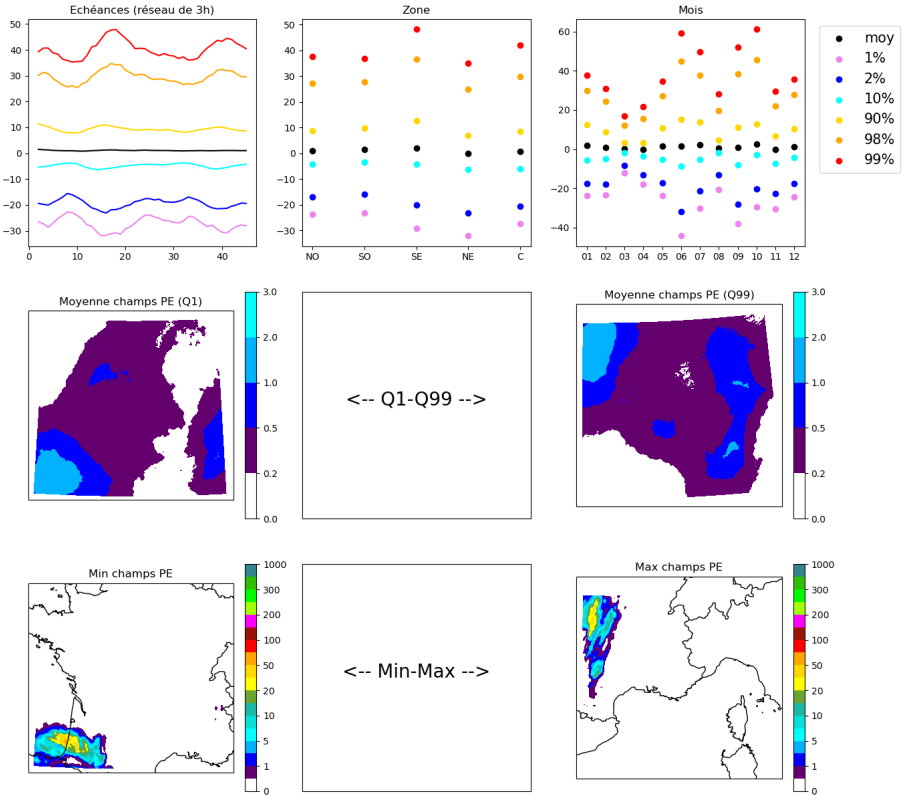


FIGURE D.58 – *Espace latent à 40 dimensions : dimension 14*

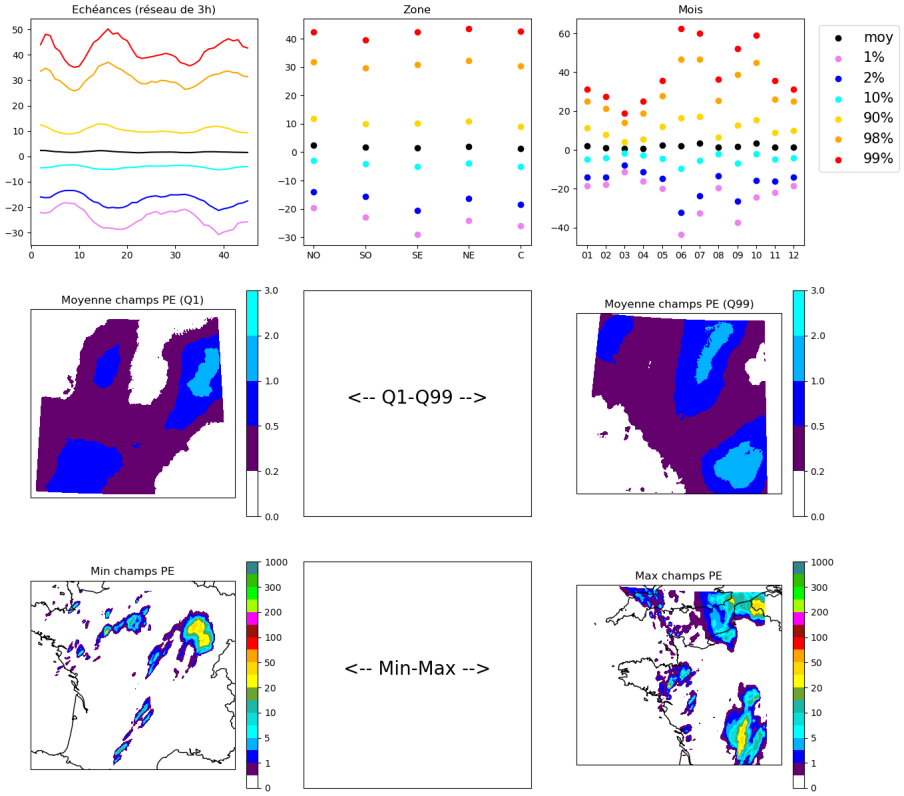


FIGURE D.59 – *Espace latent à 40 dimensions : dimension 15*

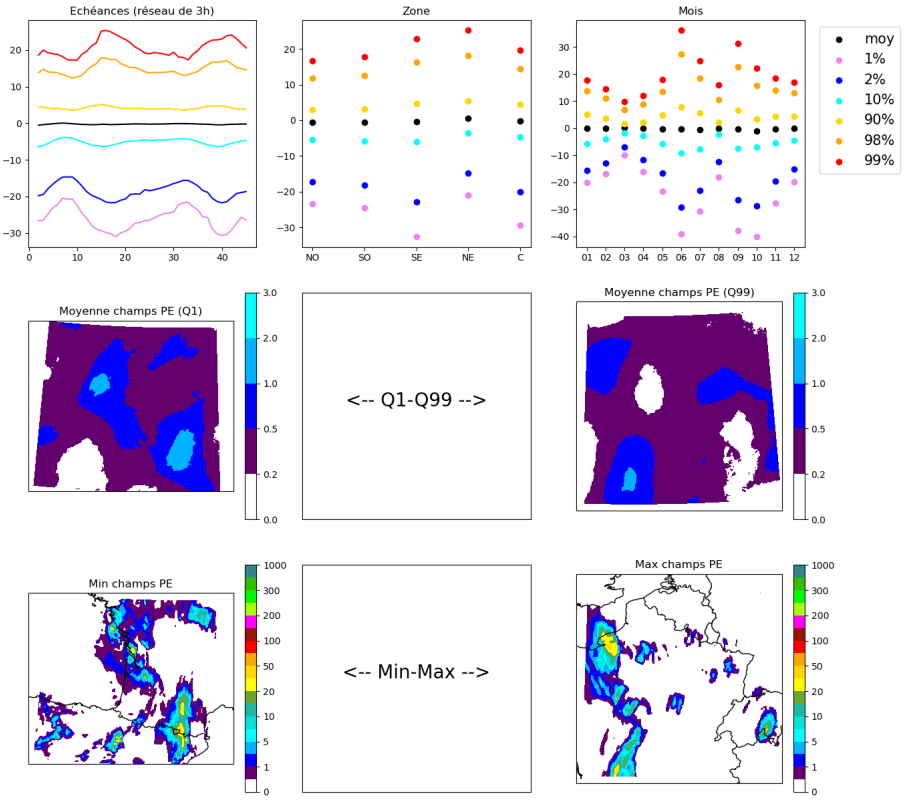


FIGURE D.60 – *Espace latent à 40 dimensions : dimension 16*

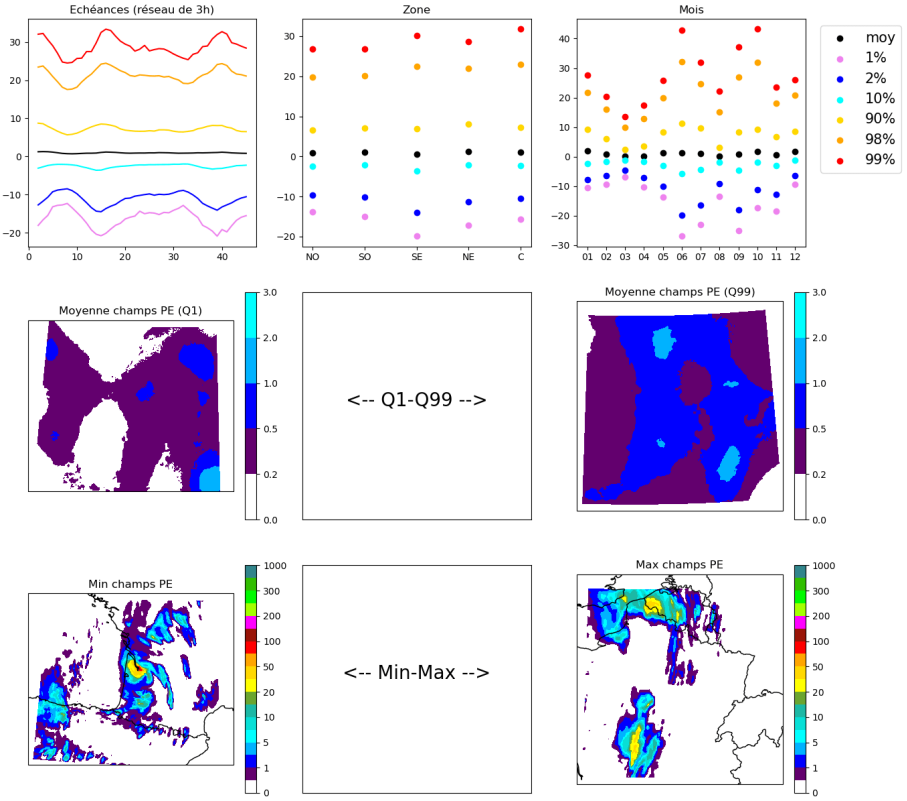


FIGURE D.61 – *Espace latent à 40 dimensions : dimension 17*

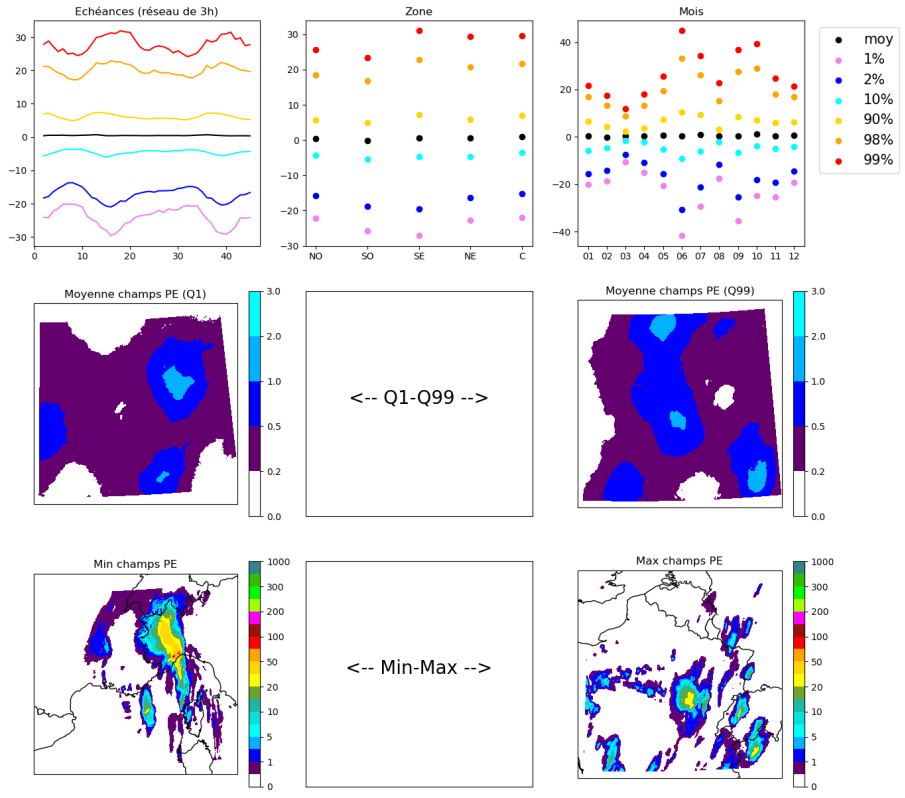


FIGURE D.62 – *Espace latent à 40 dimensions : dimension 18*

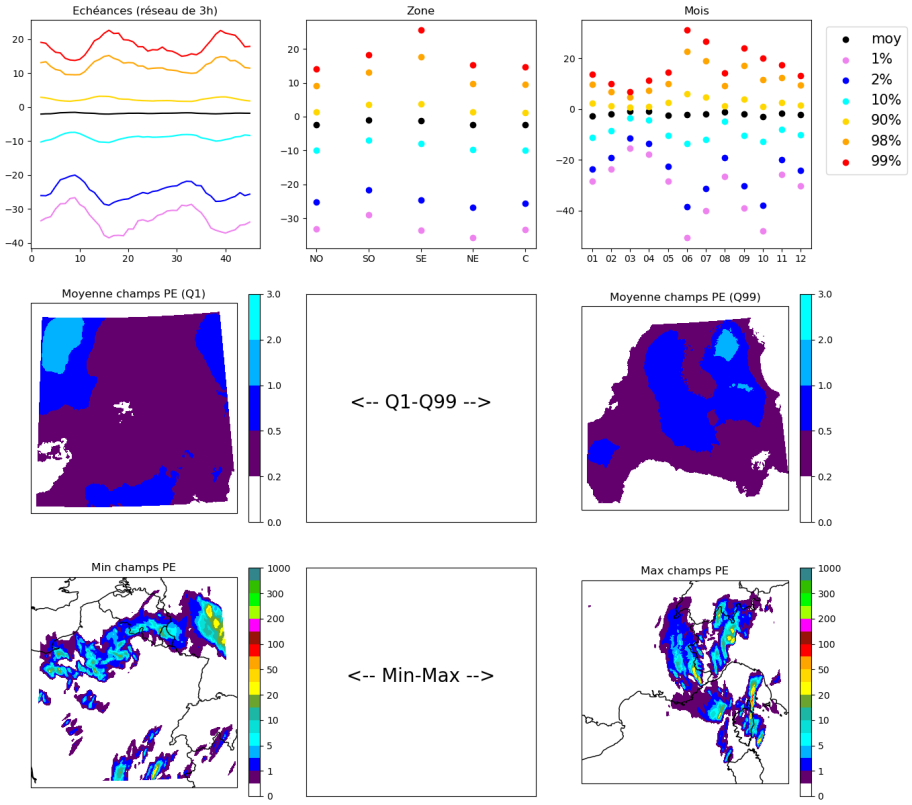


FIGURE D.63 – *Espace latent à 40 dimensions : dimension 19*

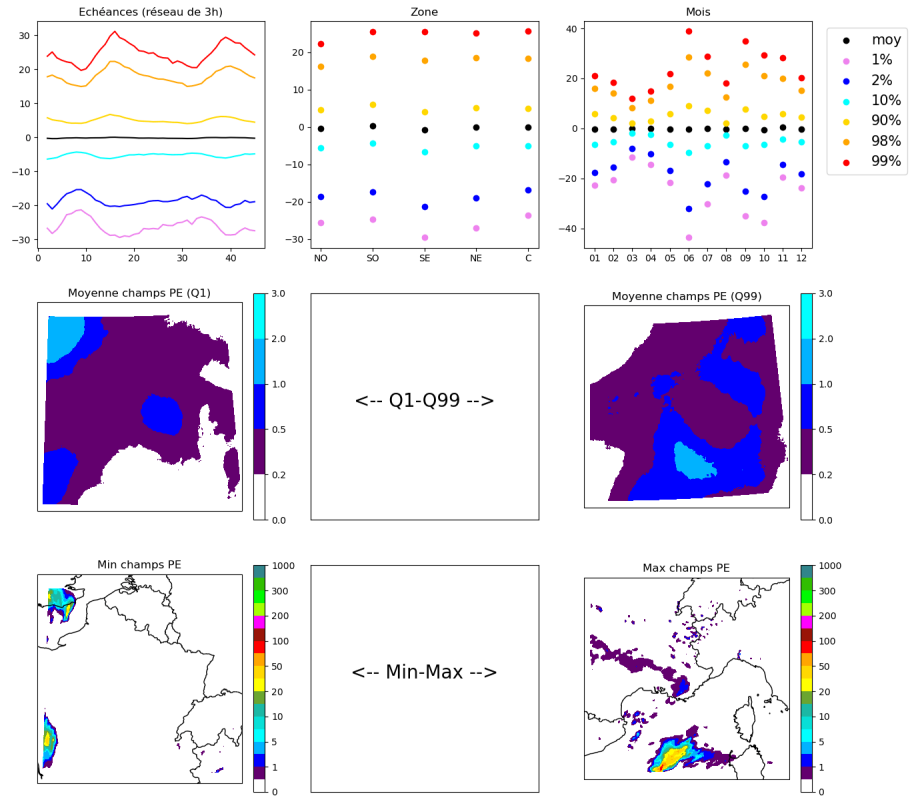


FIGURE D.64 – *Espace latent à 40 dimensions : dimension 20*

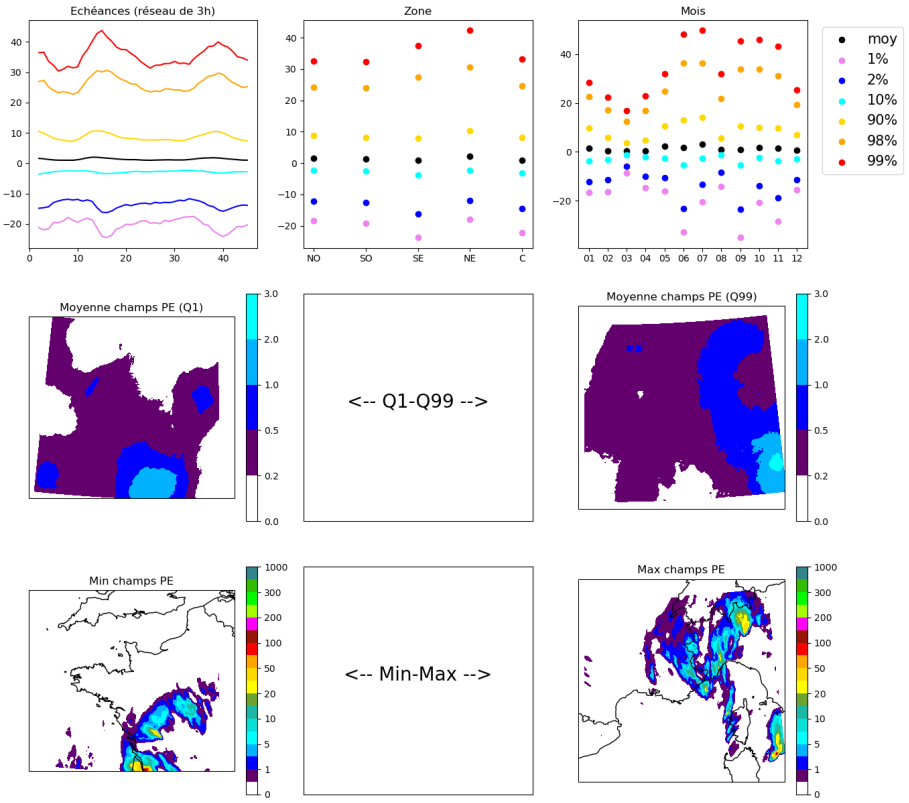


FIGURE D.65 – *Espace latent à 40 dimensions : dimension 21*

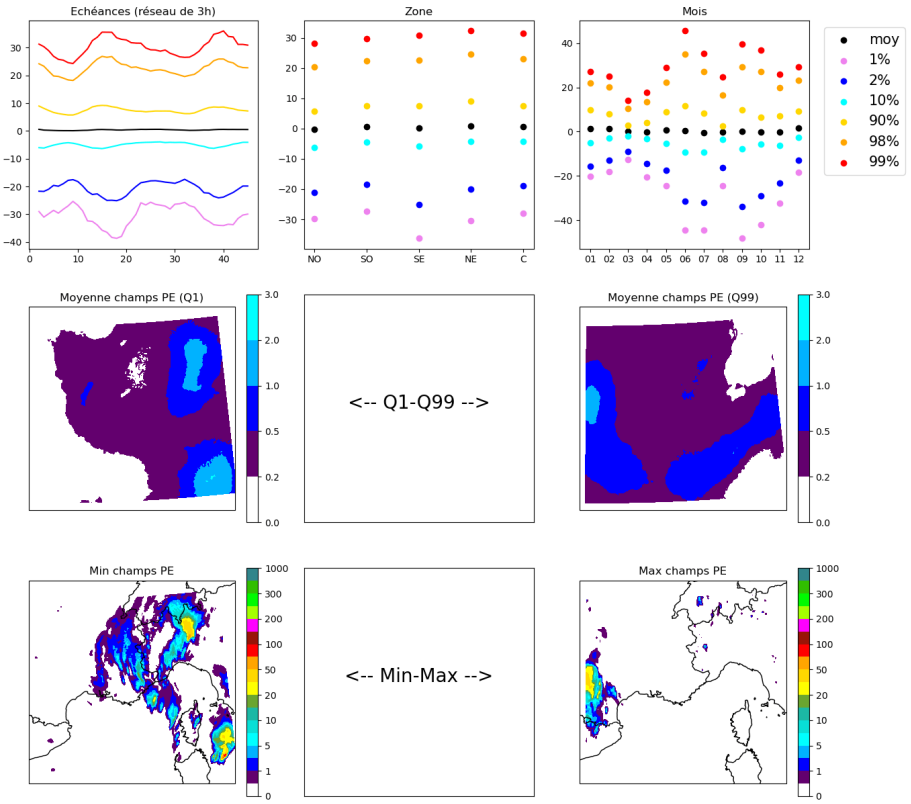


FIGURE D.66 – *Espace latent à 40 dimensions : dimension 22*

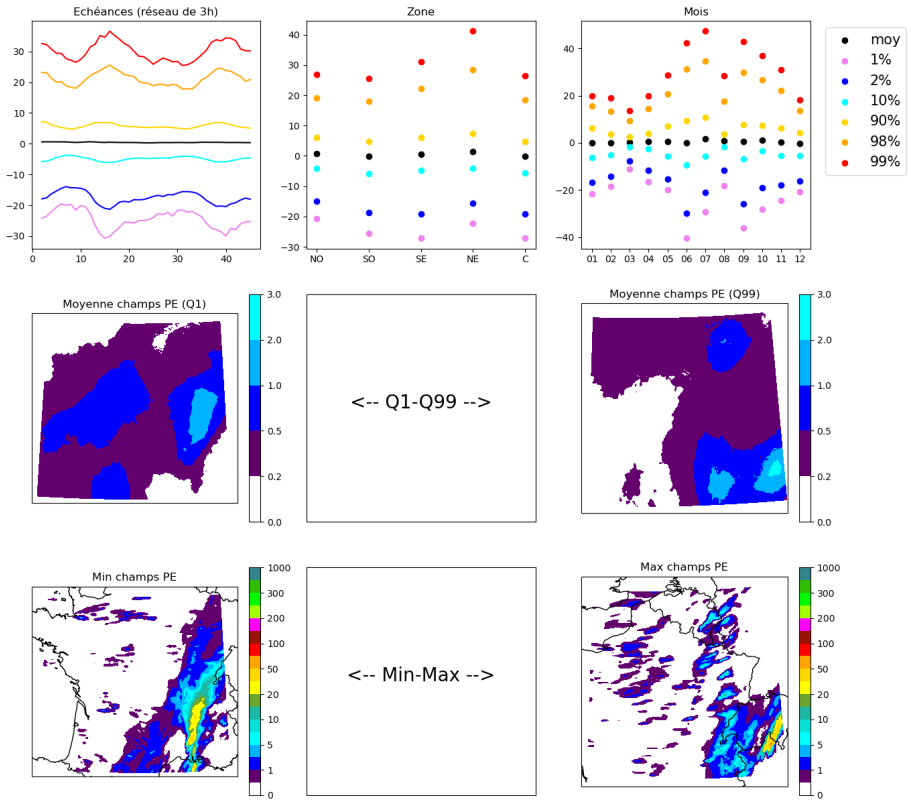


FIGURE D.67 – *Espace latent à 40 dimensions : dimension 23*

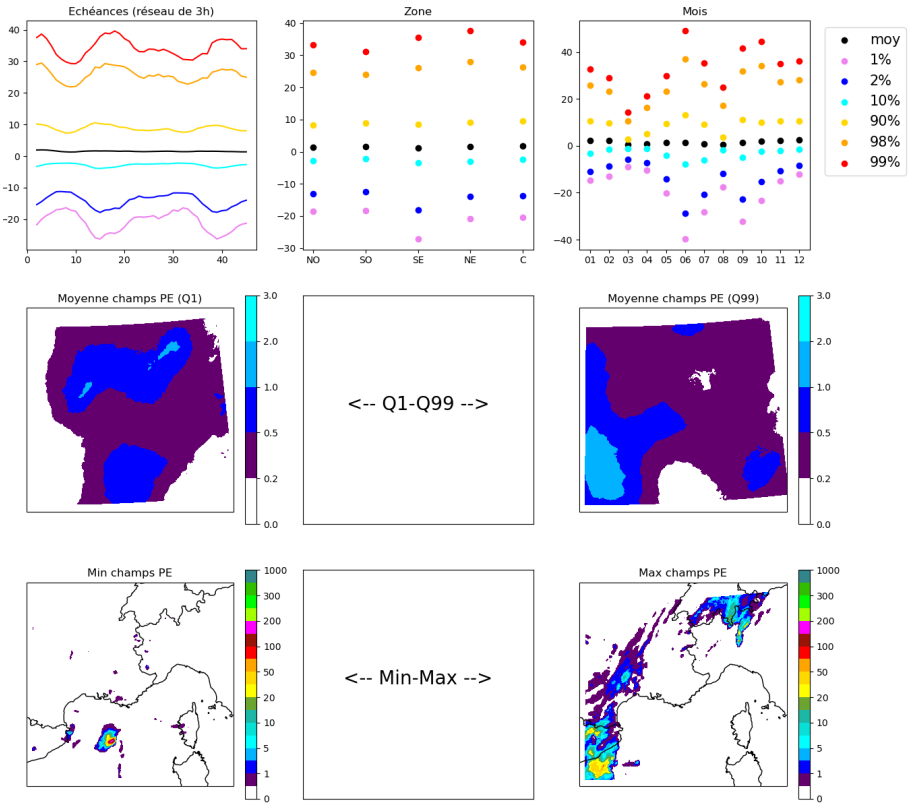


FIGURE D.68 – *Espace latent à 40 dimensions : dimension 24*

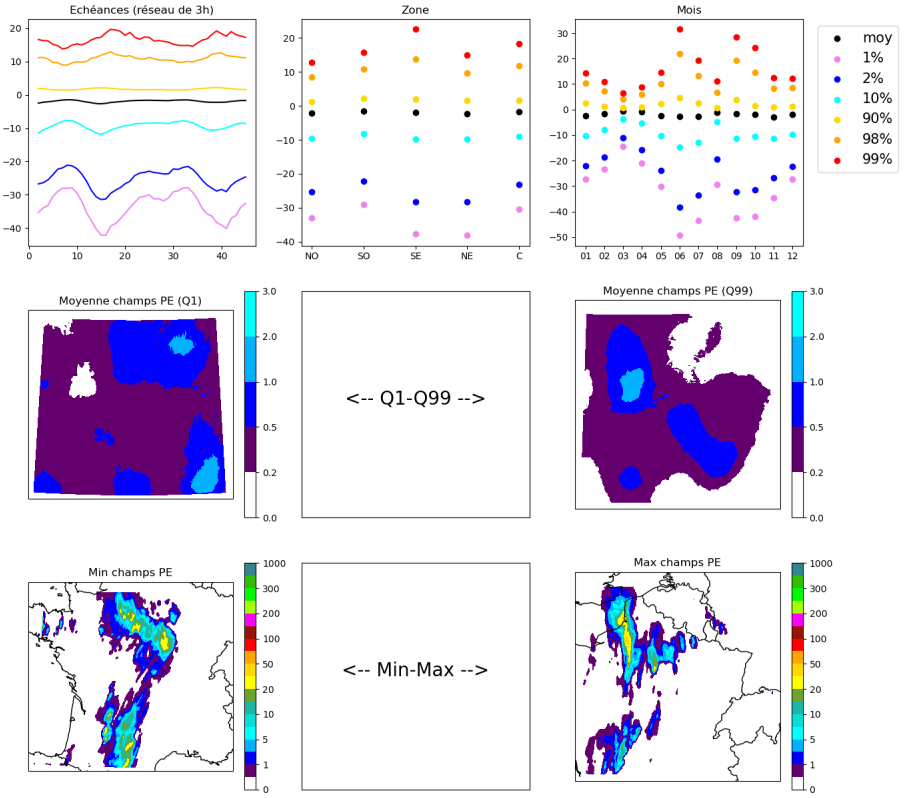


FIGURE D.69 – *Espace latent à 40 dimensions : dimension 25*

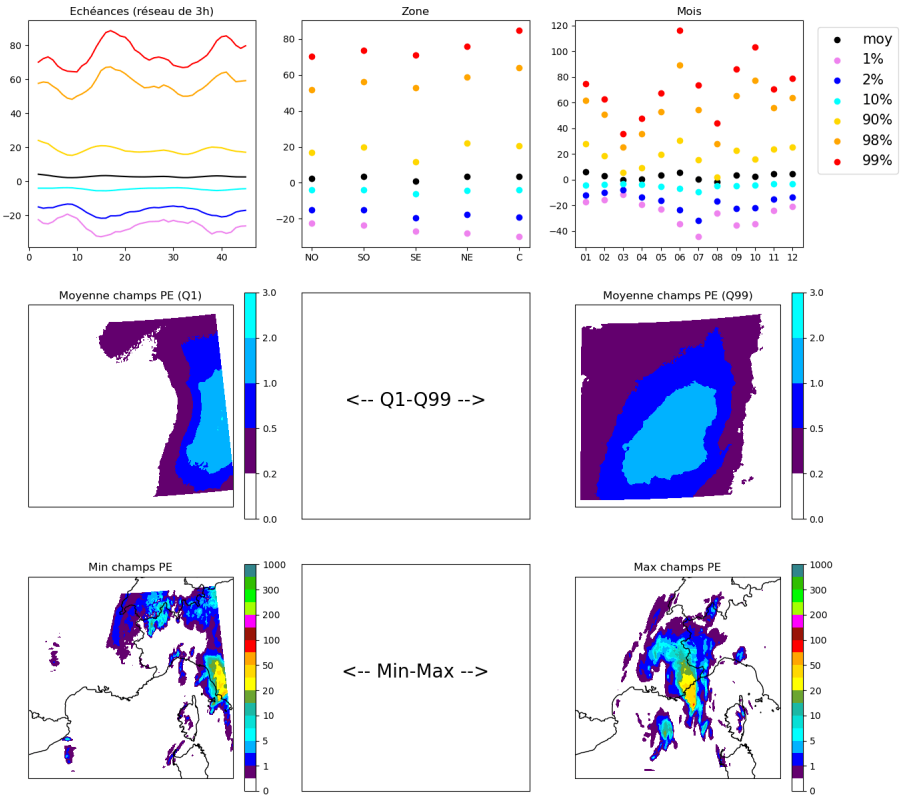


FIGURE D.70 – *Espace latent à 40 dimensions : dimension 26*

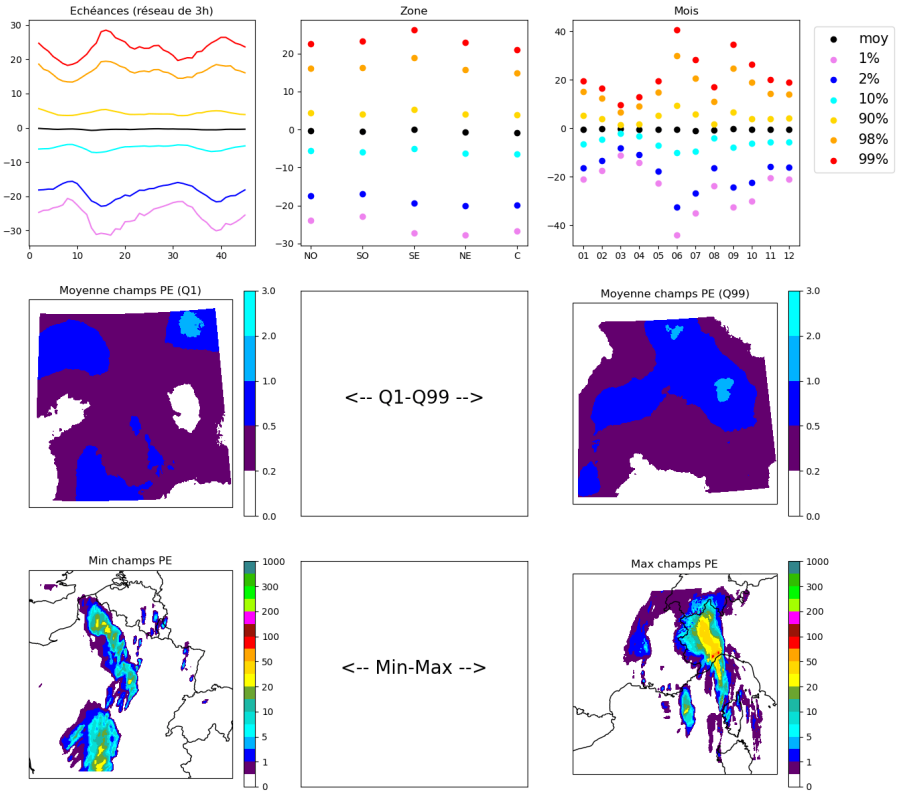


FIGURE D.71 – *Espace latent à 40 dimensions : dimension 27*

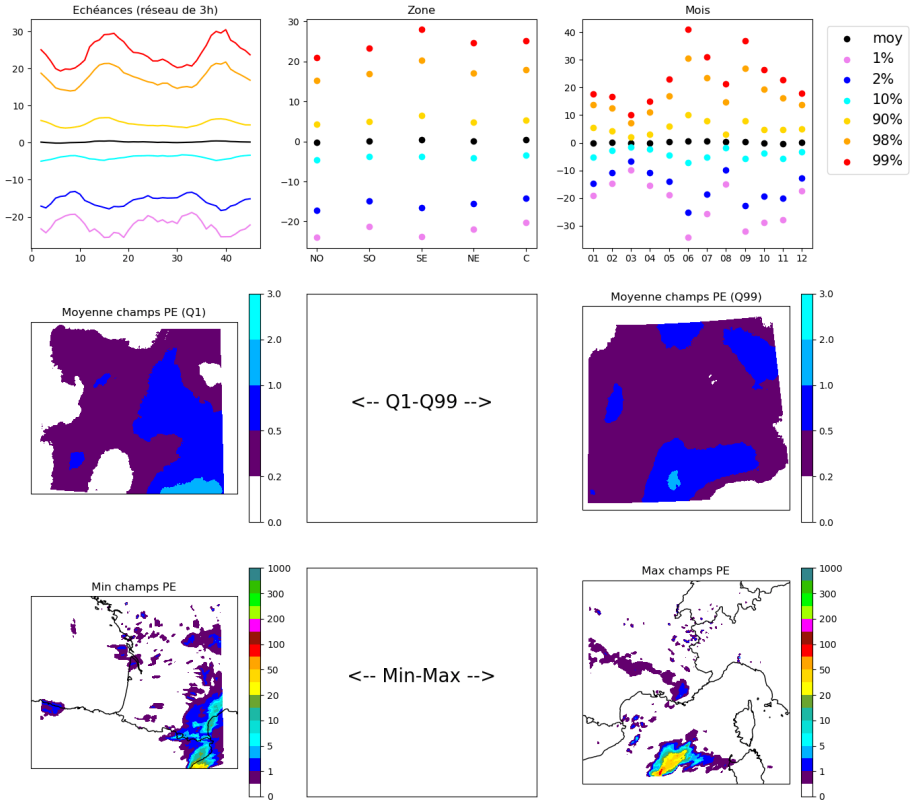


FIGURE D.72 – *Espace latent à 40 dimensions : dimension 28*

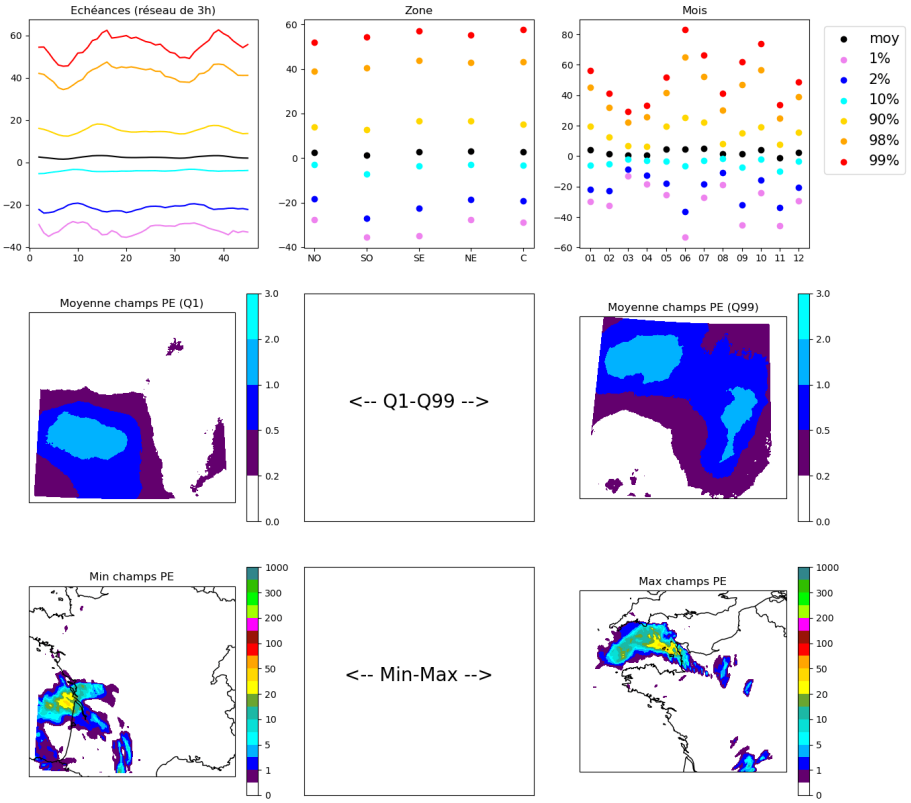


FIGURE D.73 – *Espace latent à 40 dimensions : dimension 29*

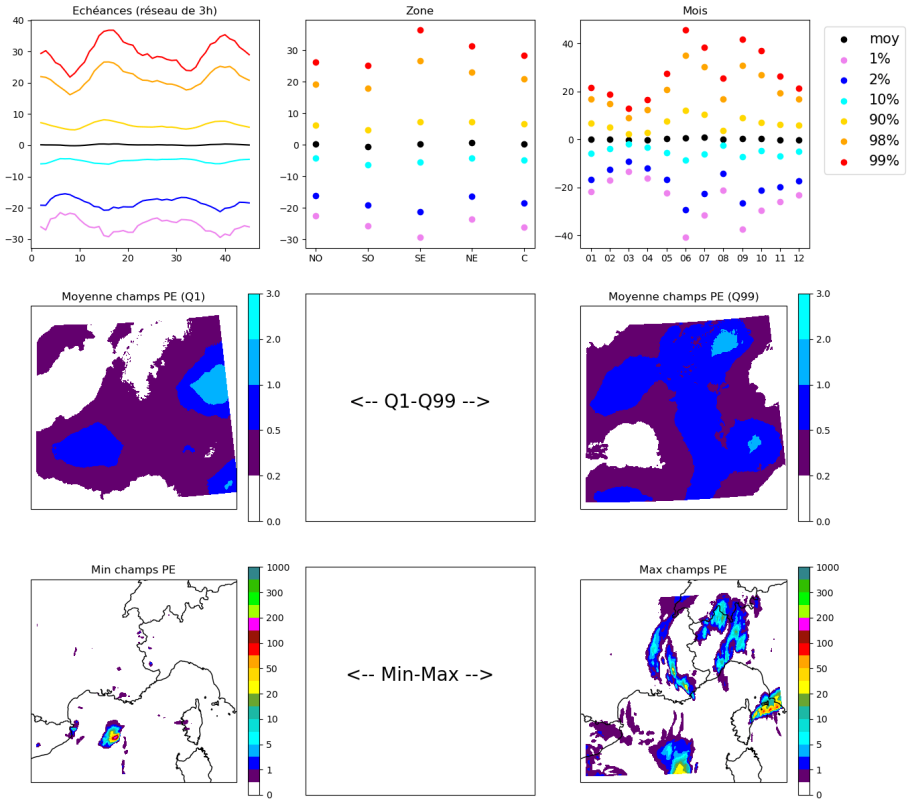


FIGURE D.74 – *Espace latent à 40 dimensions : dimension 30*

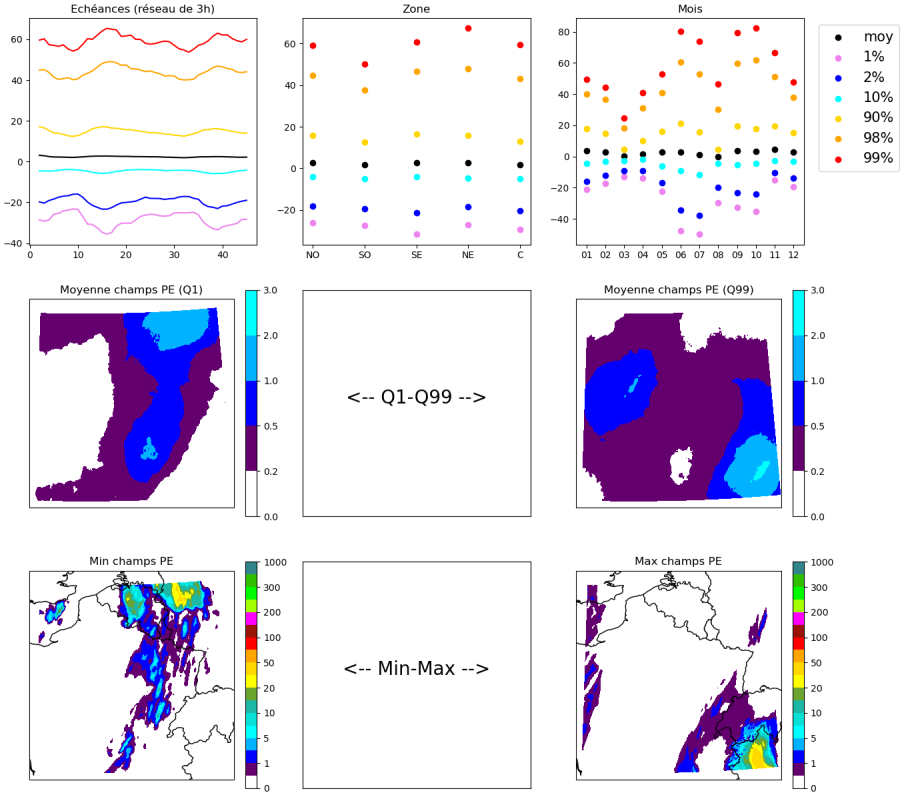


FIGURE D.75 – *Espace latent à 40 dimensions : dimension 31*

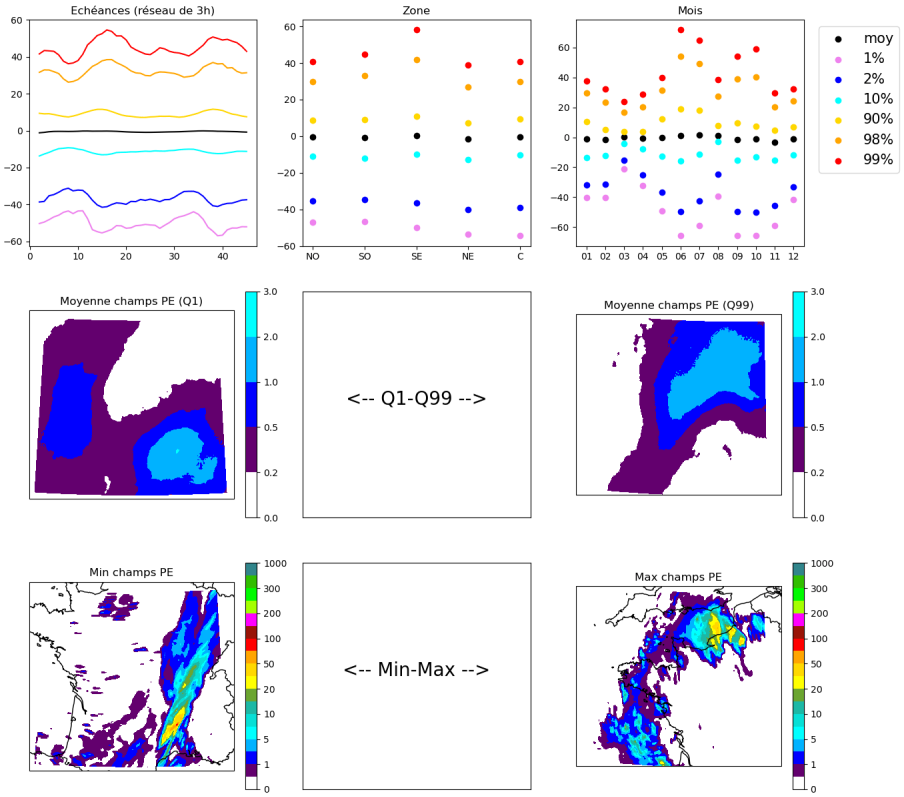


FIGURE D.76 – *Espace latent à 40 dimensions : dimension 32*

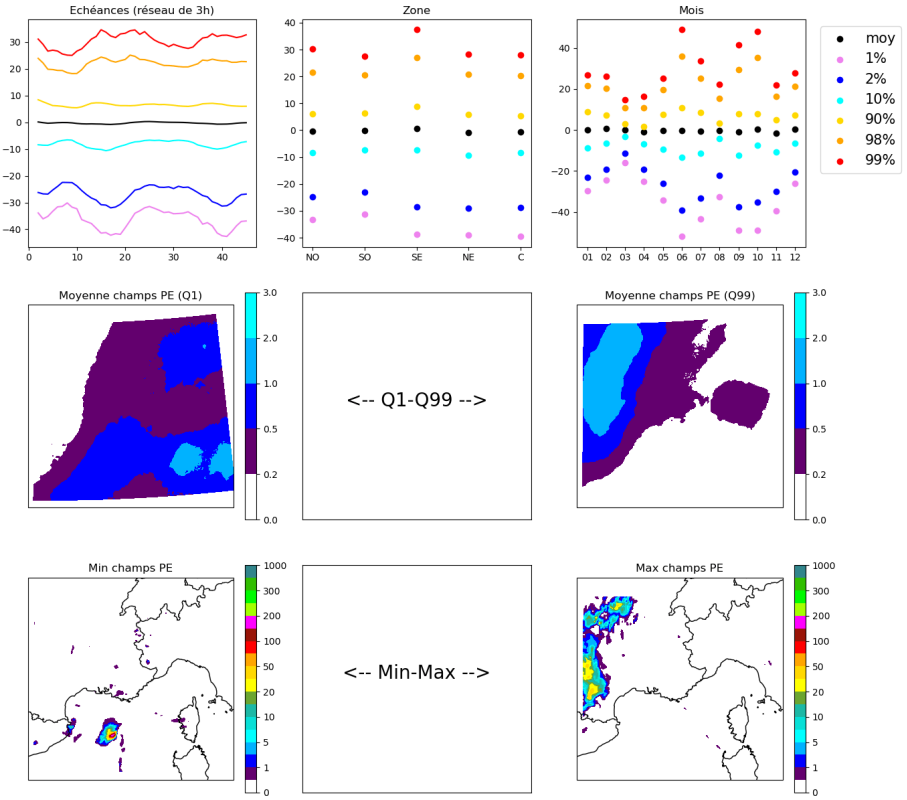


FIGURE D.77 – *Espace latent à 40 dimensions : dimension 33*

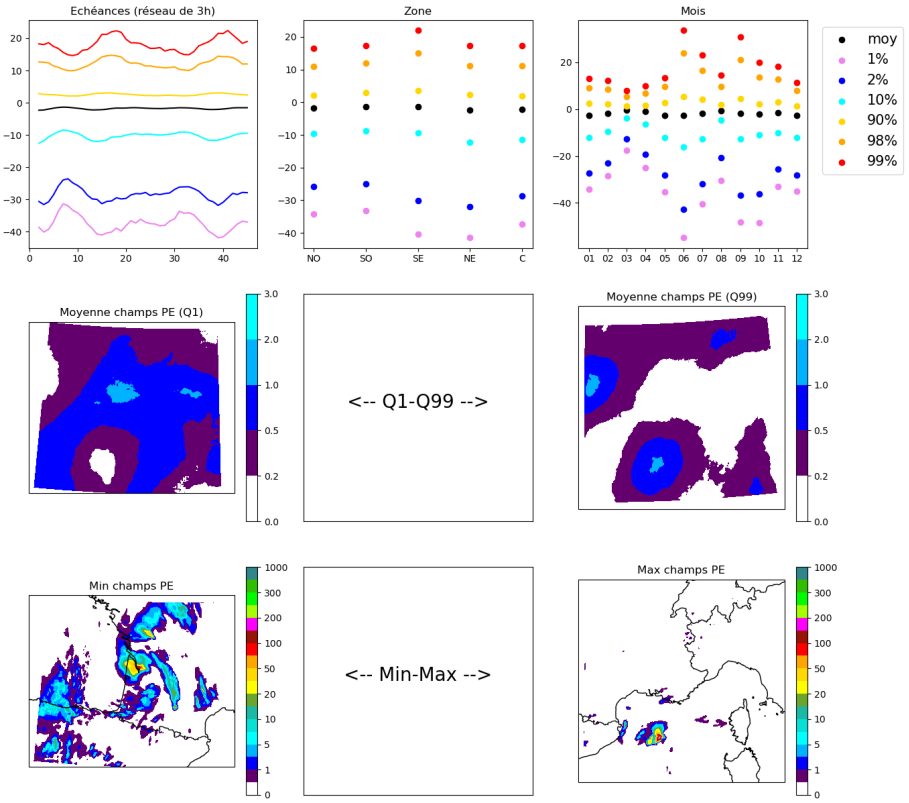


FIGURE D.78 – *Espace latent à 40 dimensions : dimension 34*

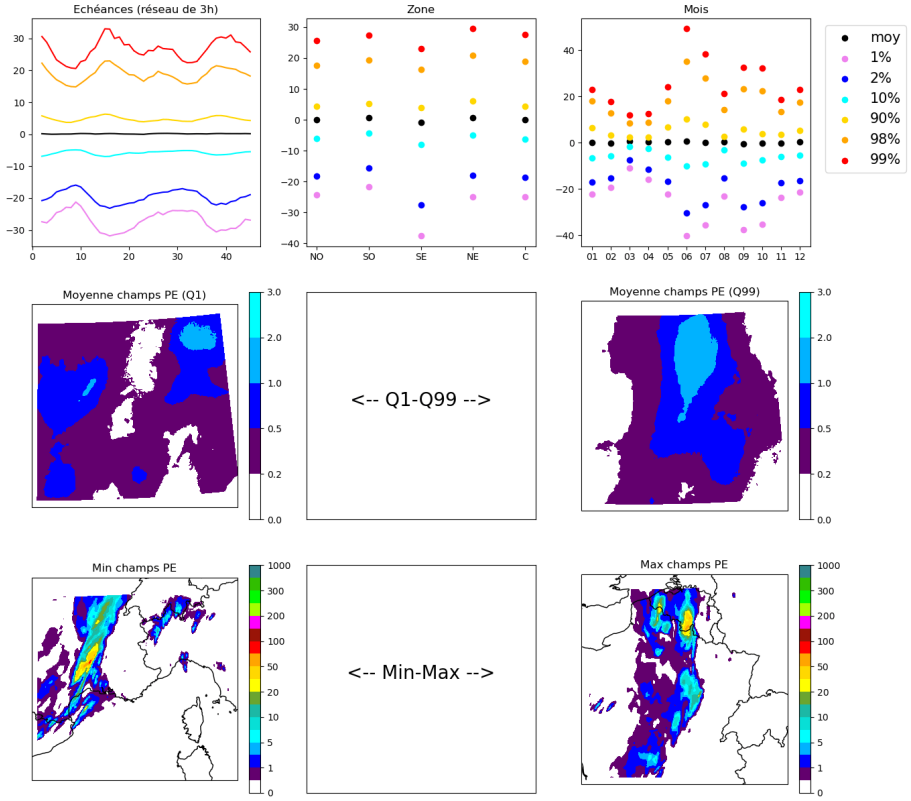


FIGURE D.79 – *Espace latent à 40 dimensions : dimension 35*

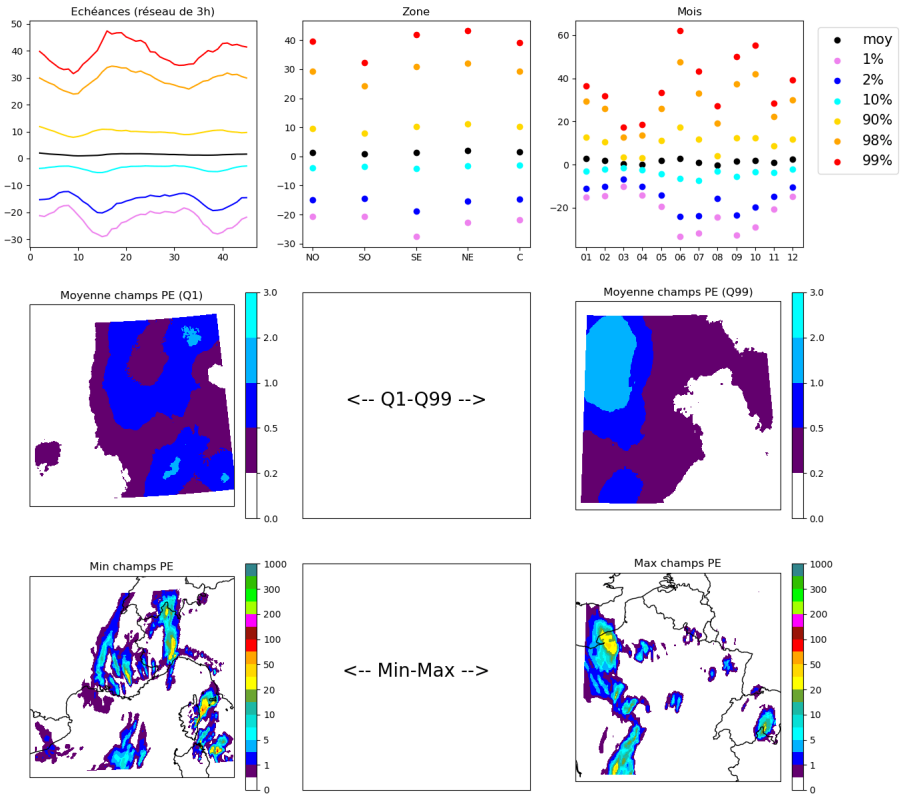


FIGURE D.80 – *Espace latent à 40 dimensions : dimension 36*

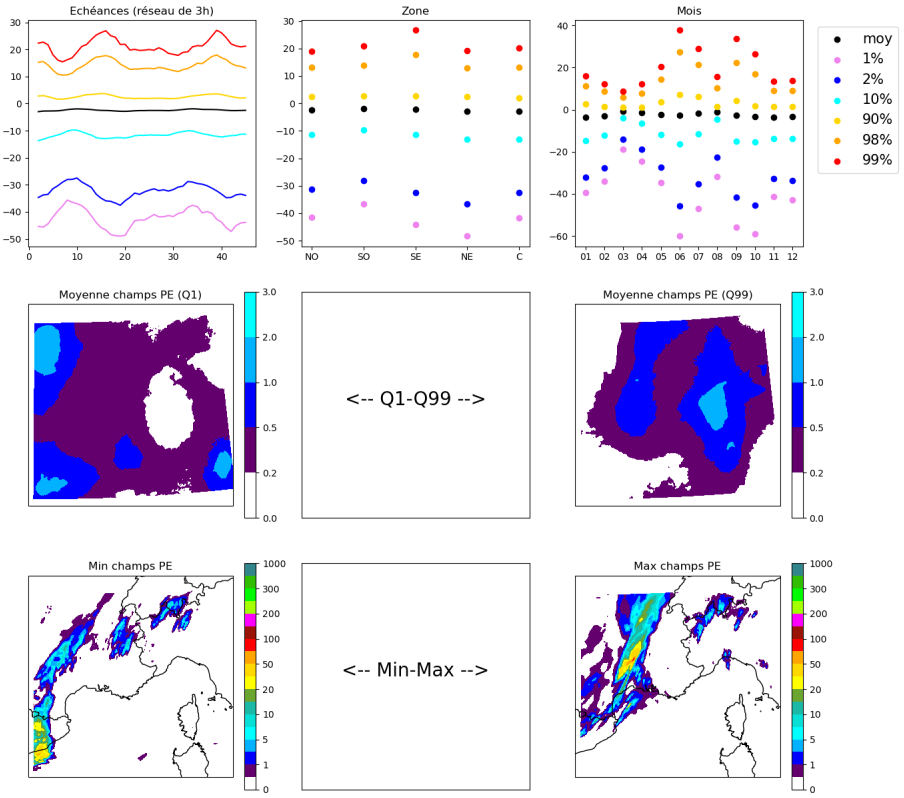


FIGURE D.81 – *Espace latent à 40 dimensions : dimension 37*

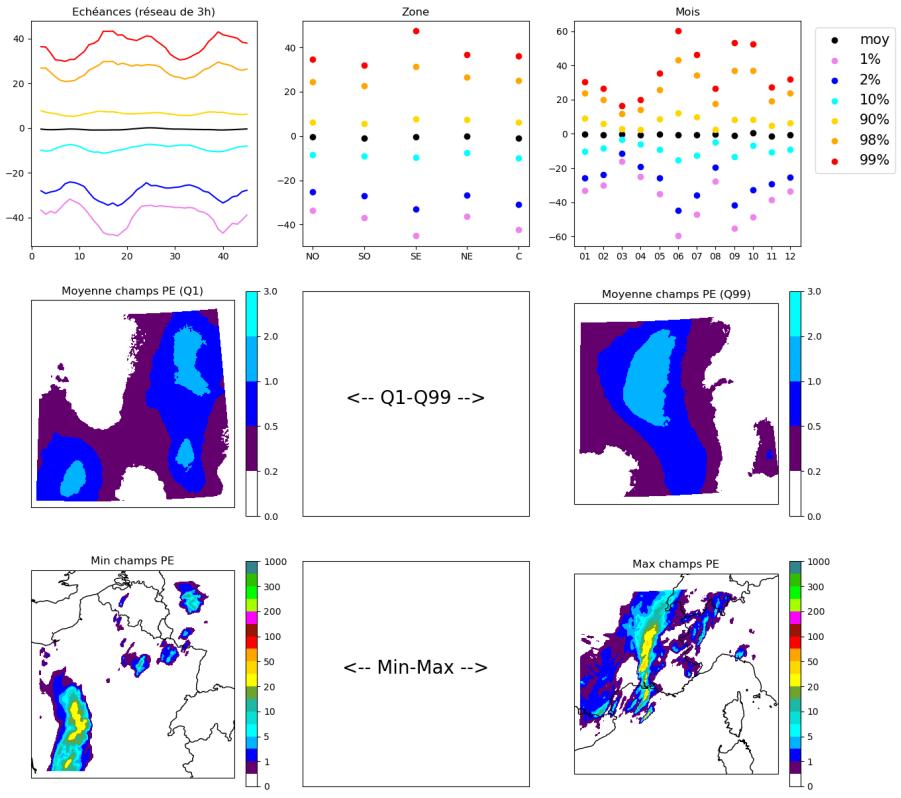


FIGURE D.82 – *Espace latent à 40 dimensions : dimension 38*

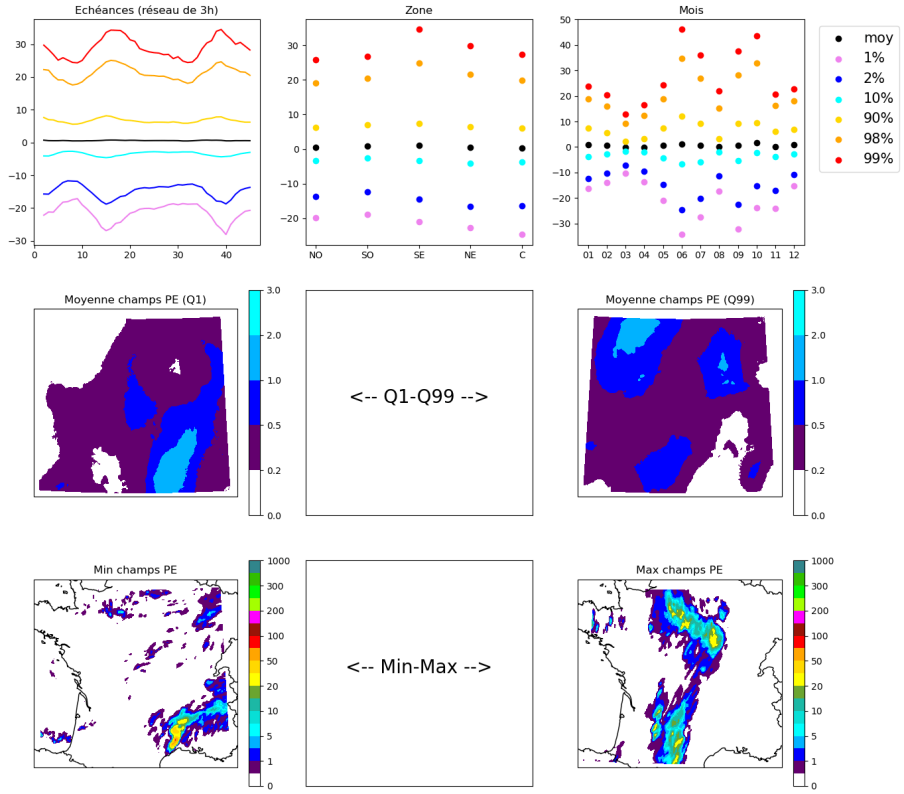


FIGURE D.83 – *Espace latent à 40 dimensions : dimension 39*

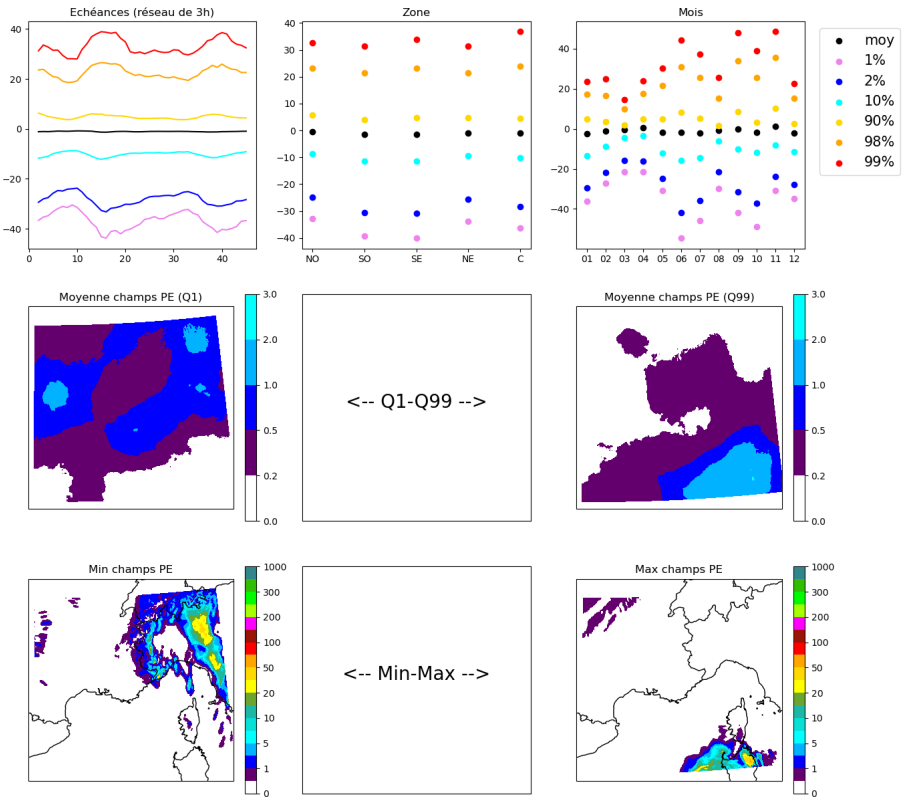


FIGURE D.84 – *Espace latent à 40 dimensions : dimension 40*

Annexe E

Structures météorologiques récurrentes

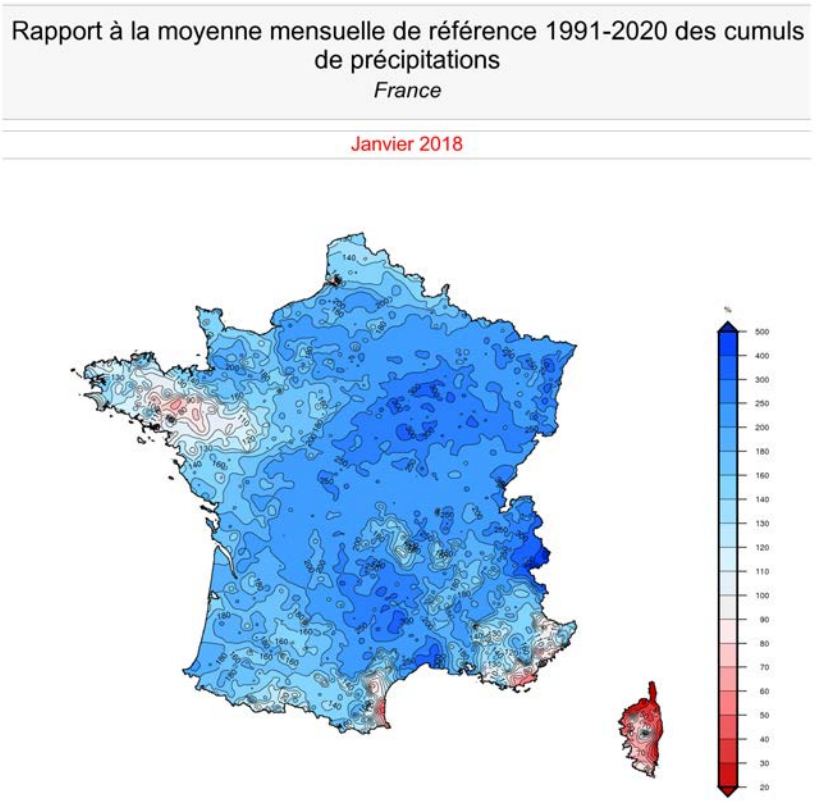


FIGURE E.1 – *Anomalie de précipitation pour le mois de janvier 2018 (Source : Climascope, site interne de Météo-France).*

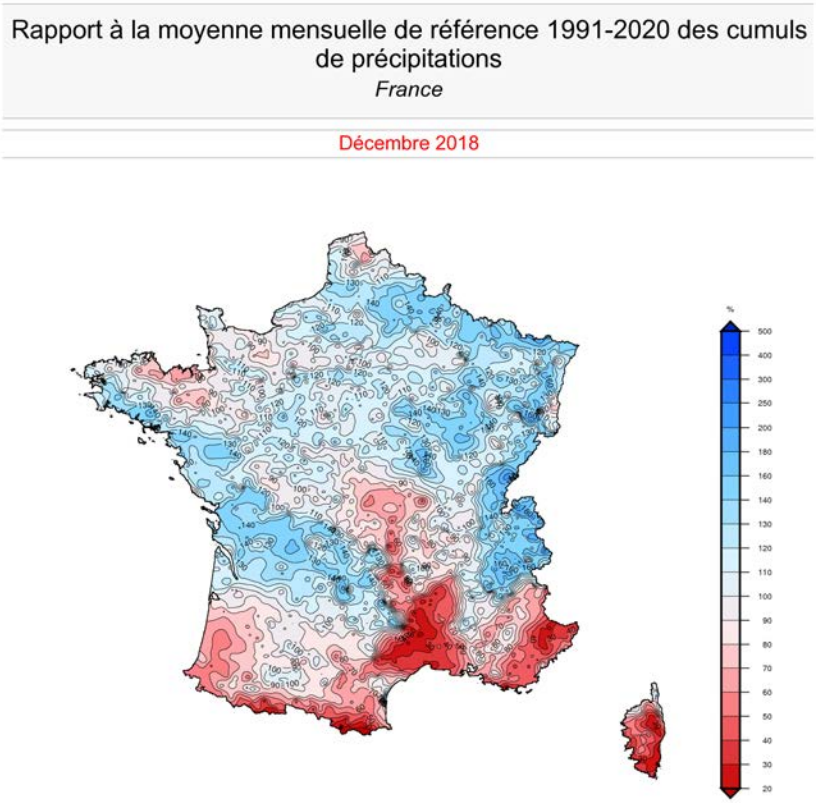


FIGURE E.2 – *Anomalie de précipitation pour le mois de décembre 2018 (Source : Climascope, site interne de Météo-France).*

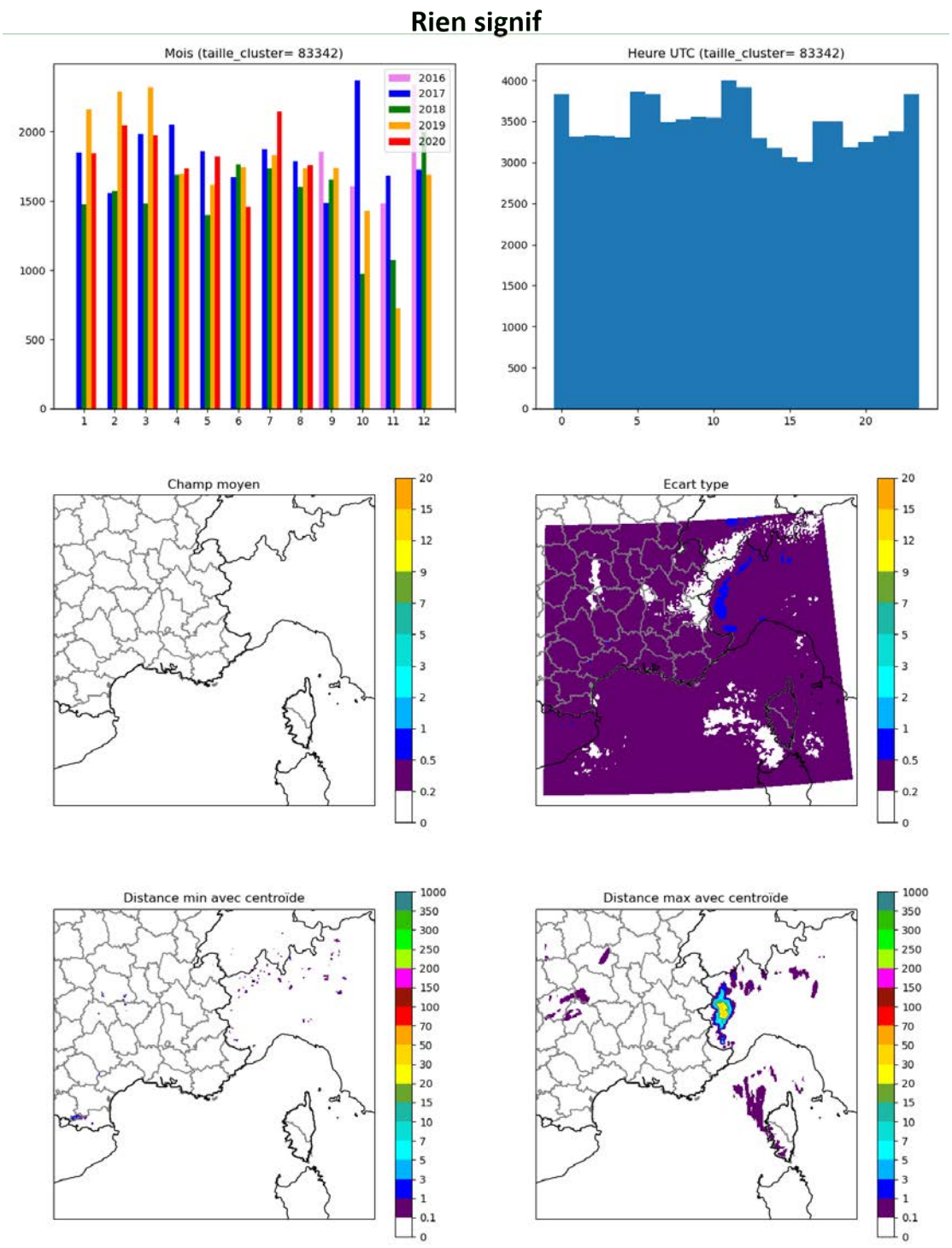


FIGURE E.3 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°1.*

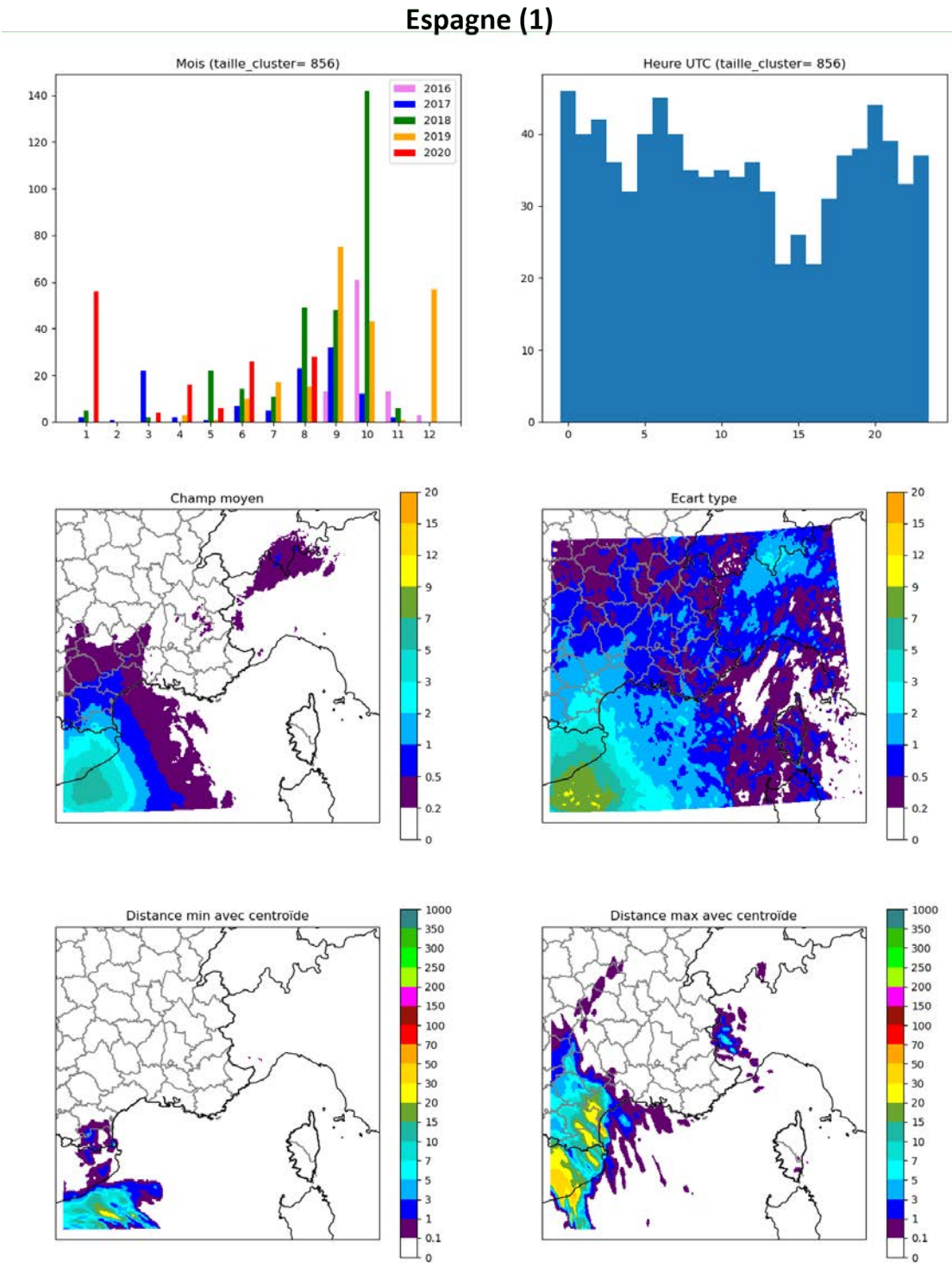


FIGURE E.4 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°2.*

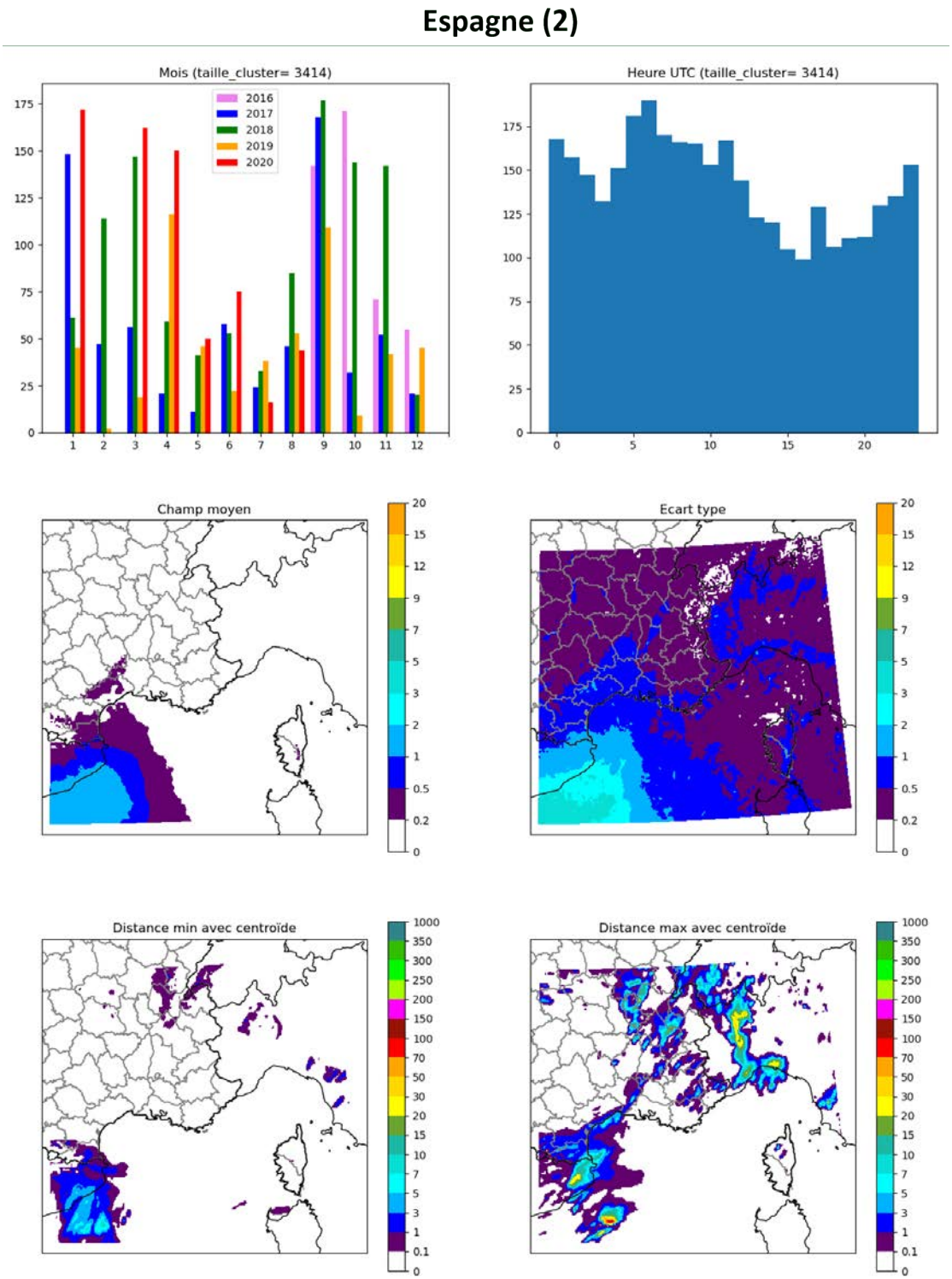


FIGURE E.5 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°3.*

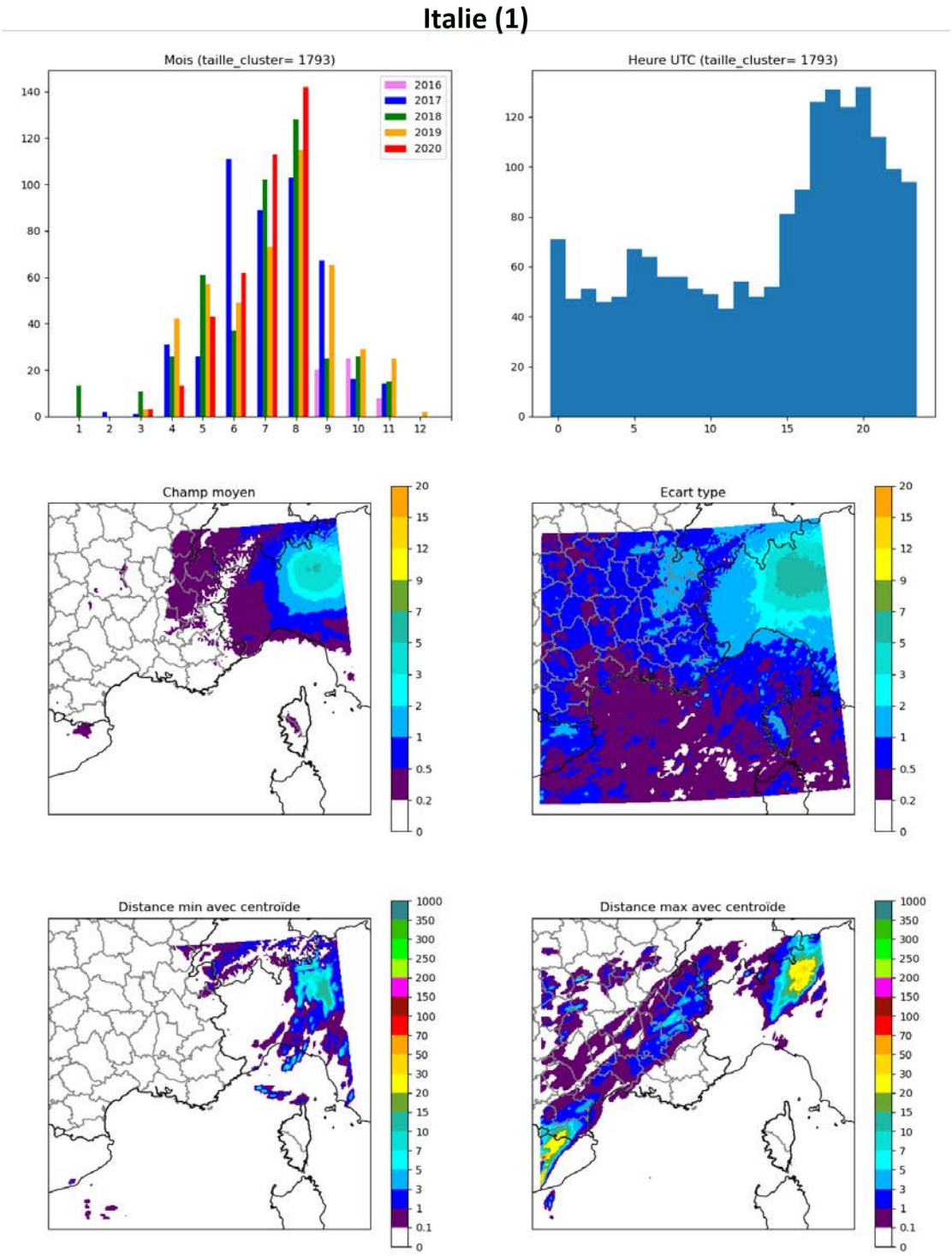


FIGURE E.6 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°4.*

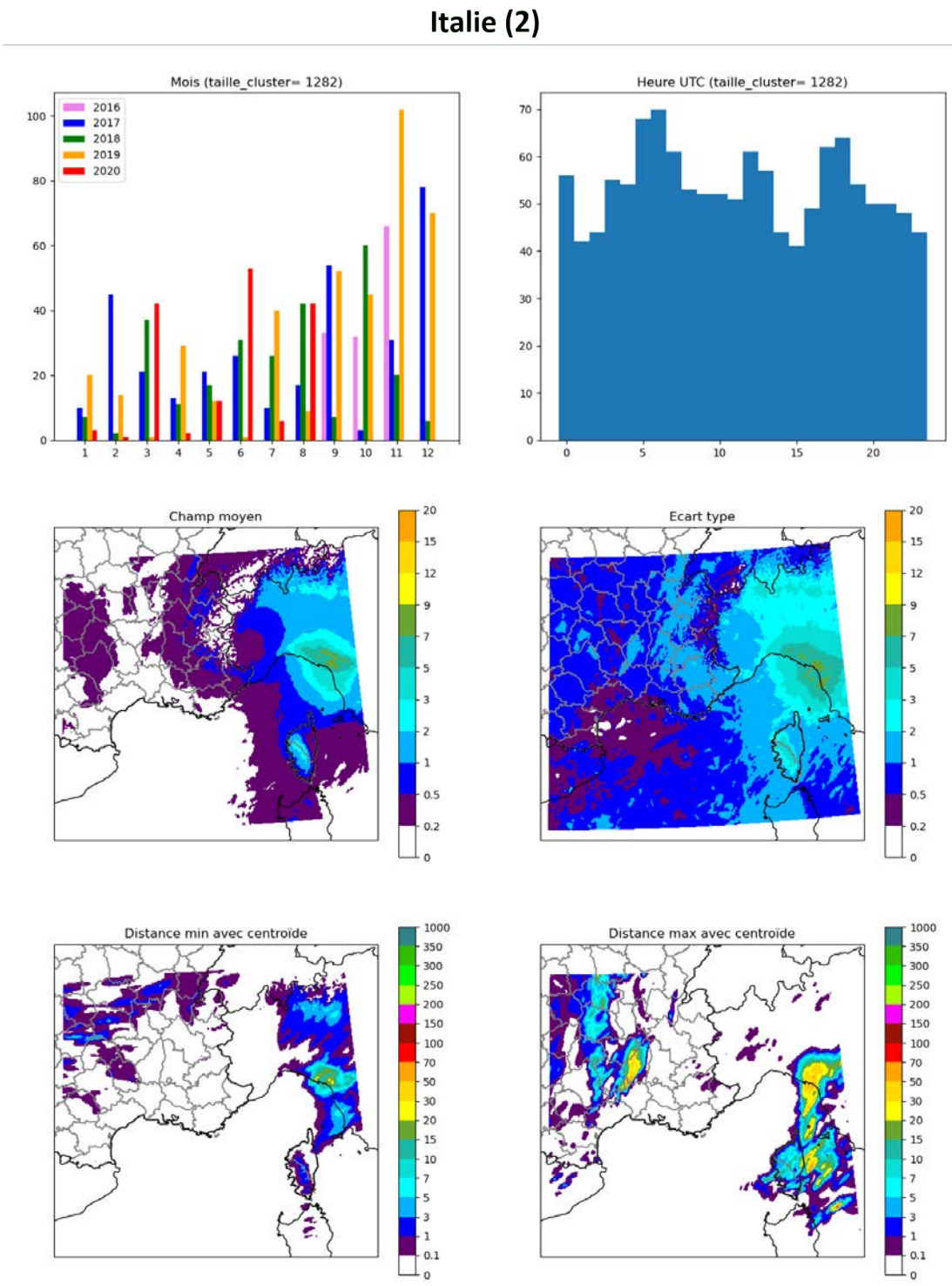


FIGURE E.7 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°5.*

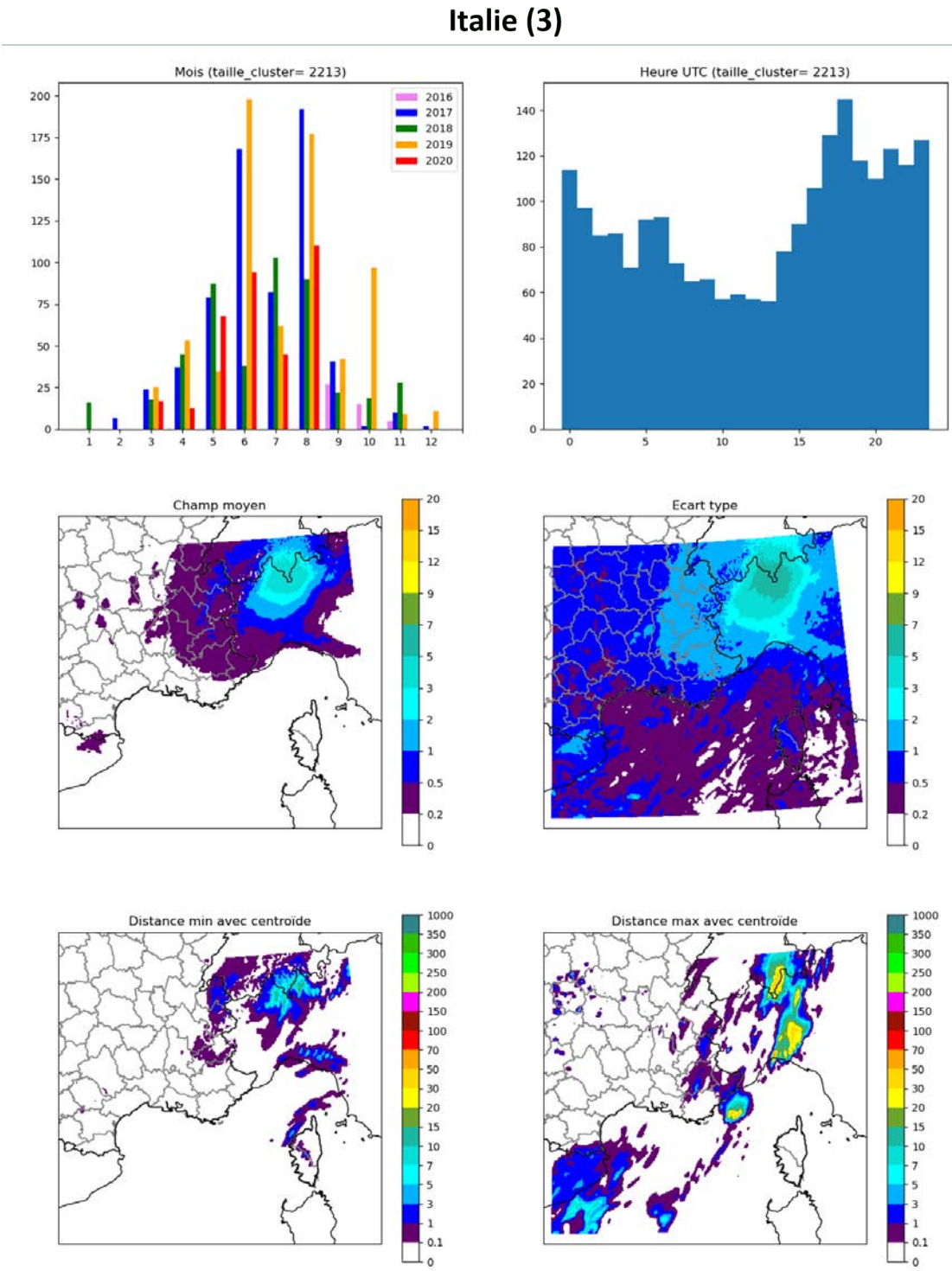


FIGURE E.8 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°6.*

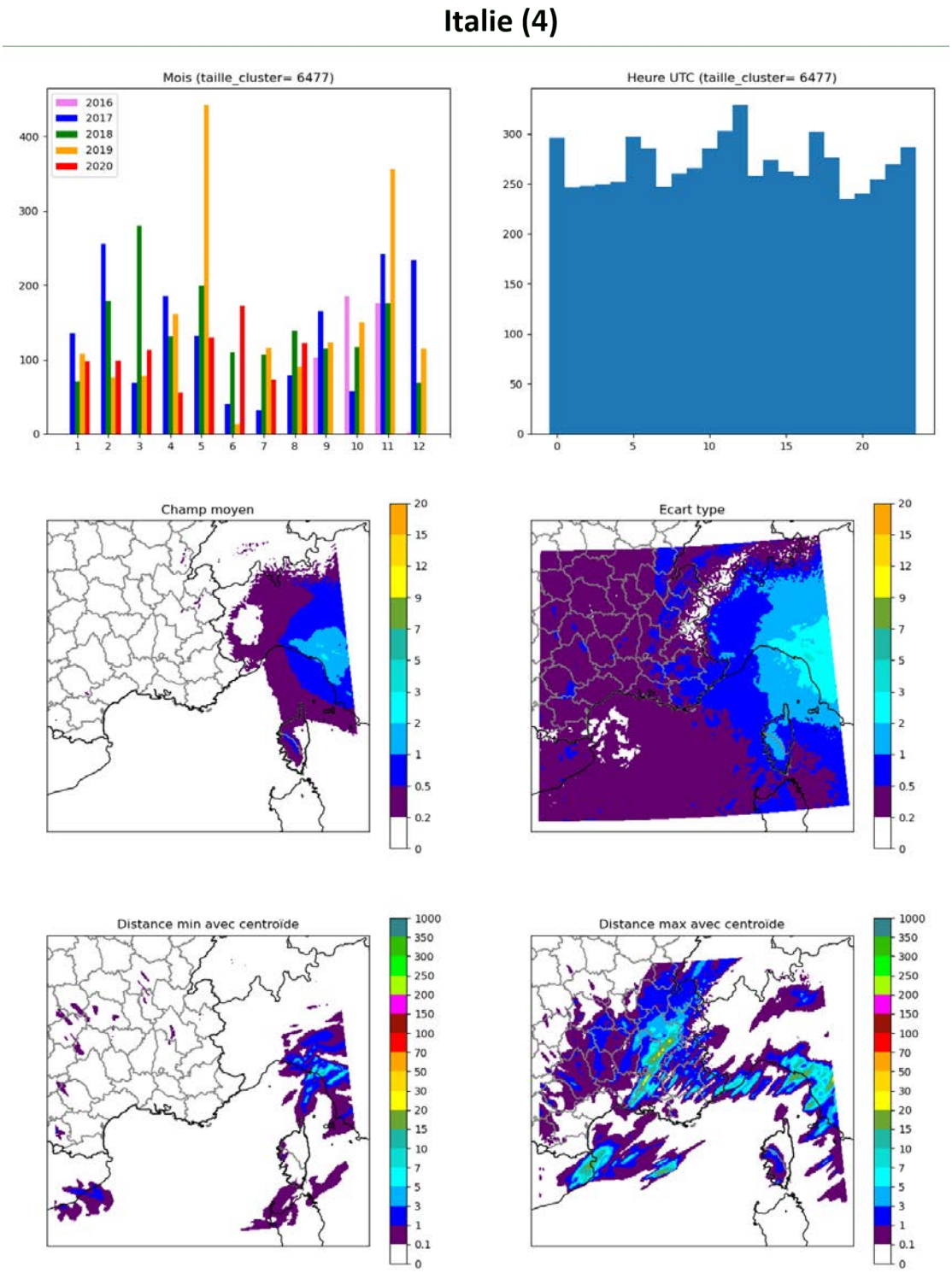


FIGURE E.9 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°7.*

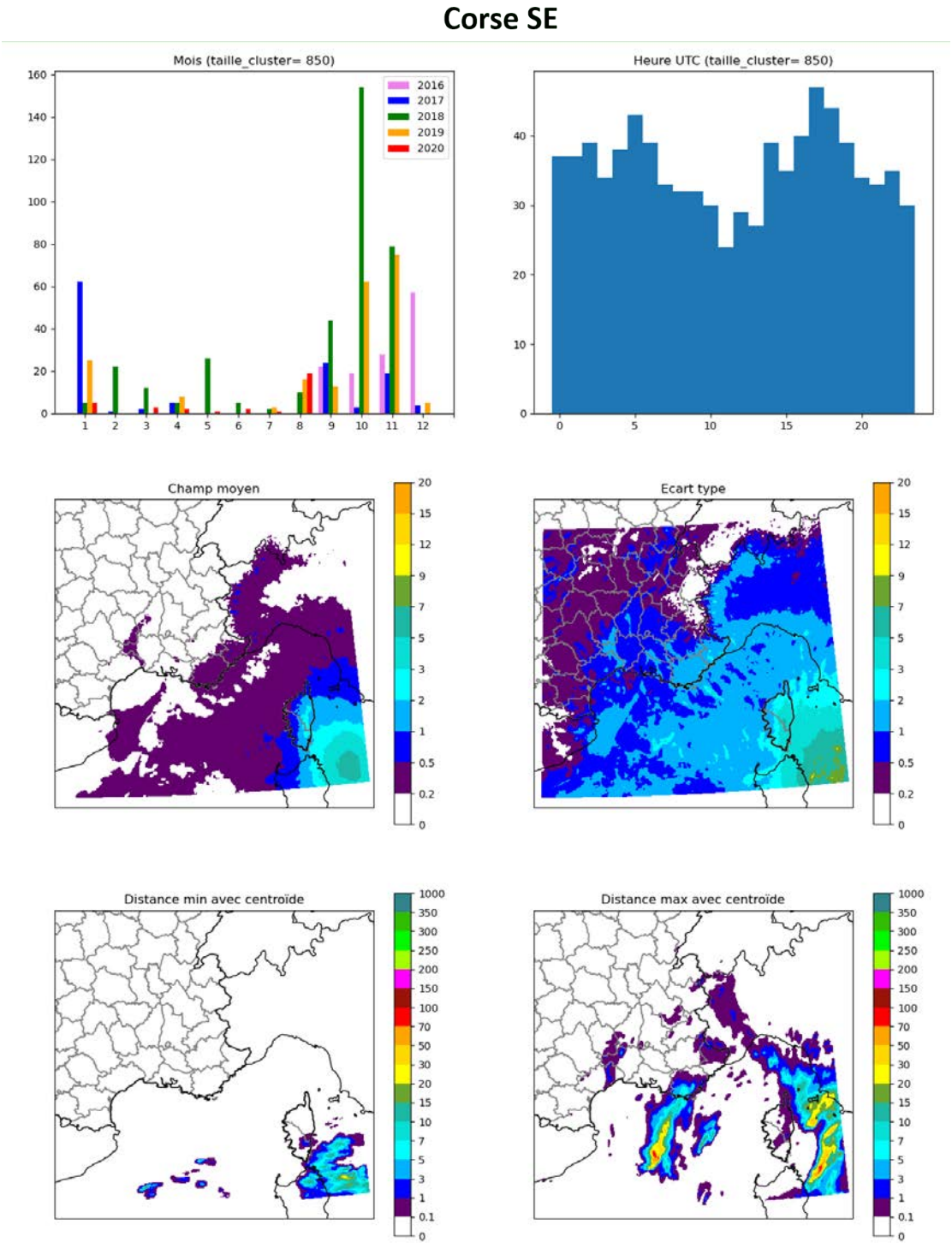


FIGURE E.10 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°8.*

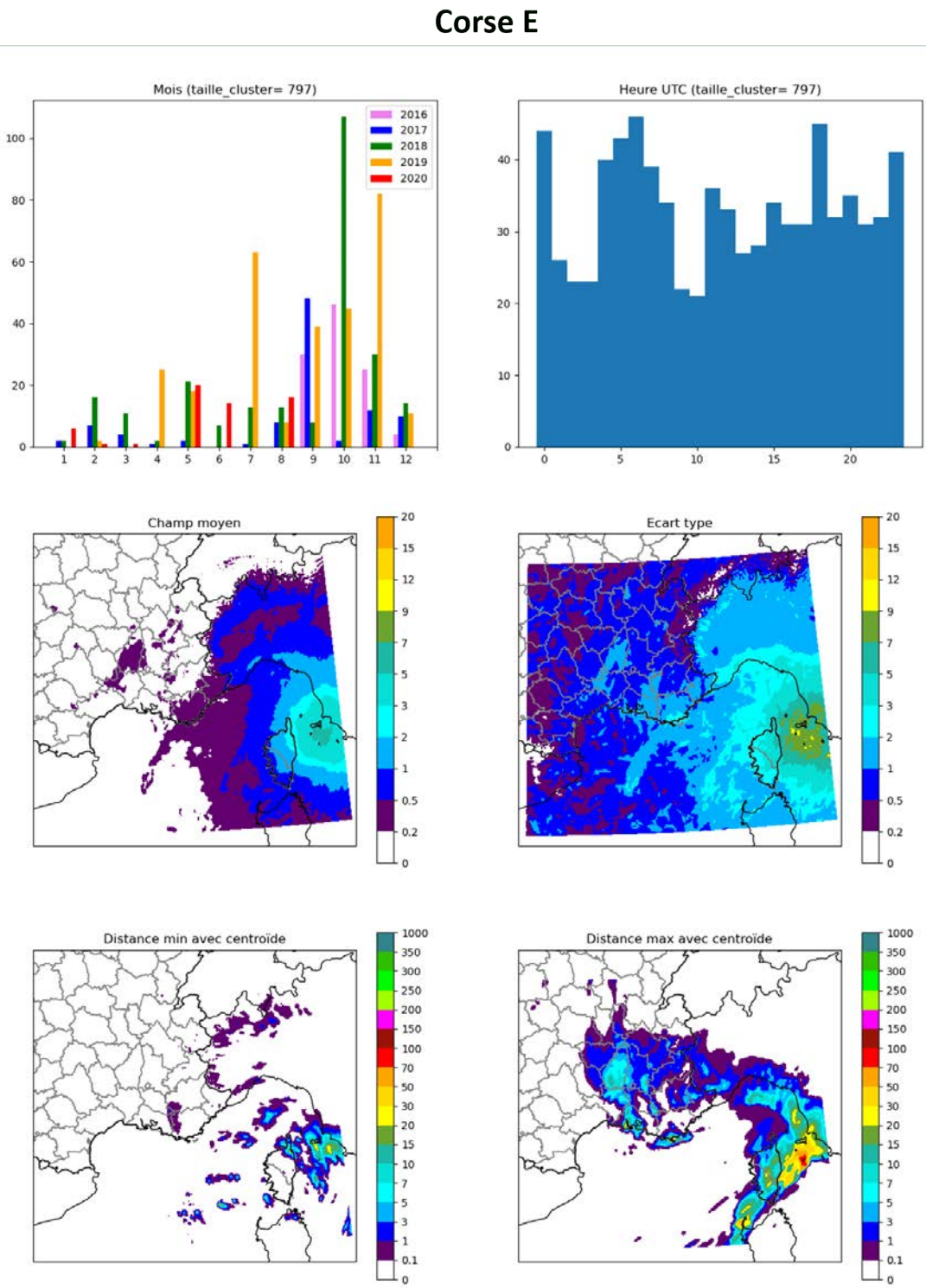


FIGURE E.11 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°9.*

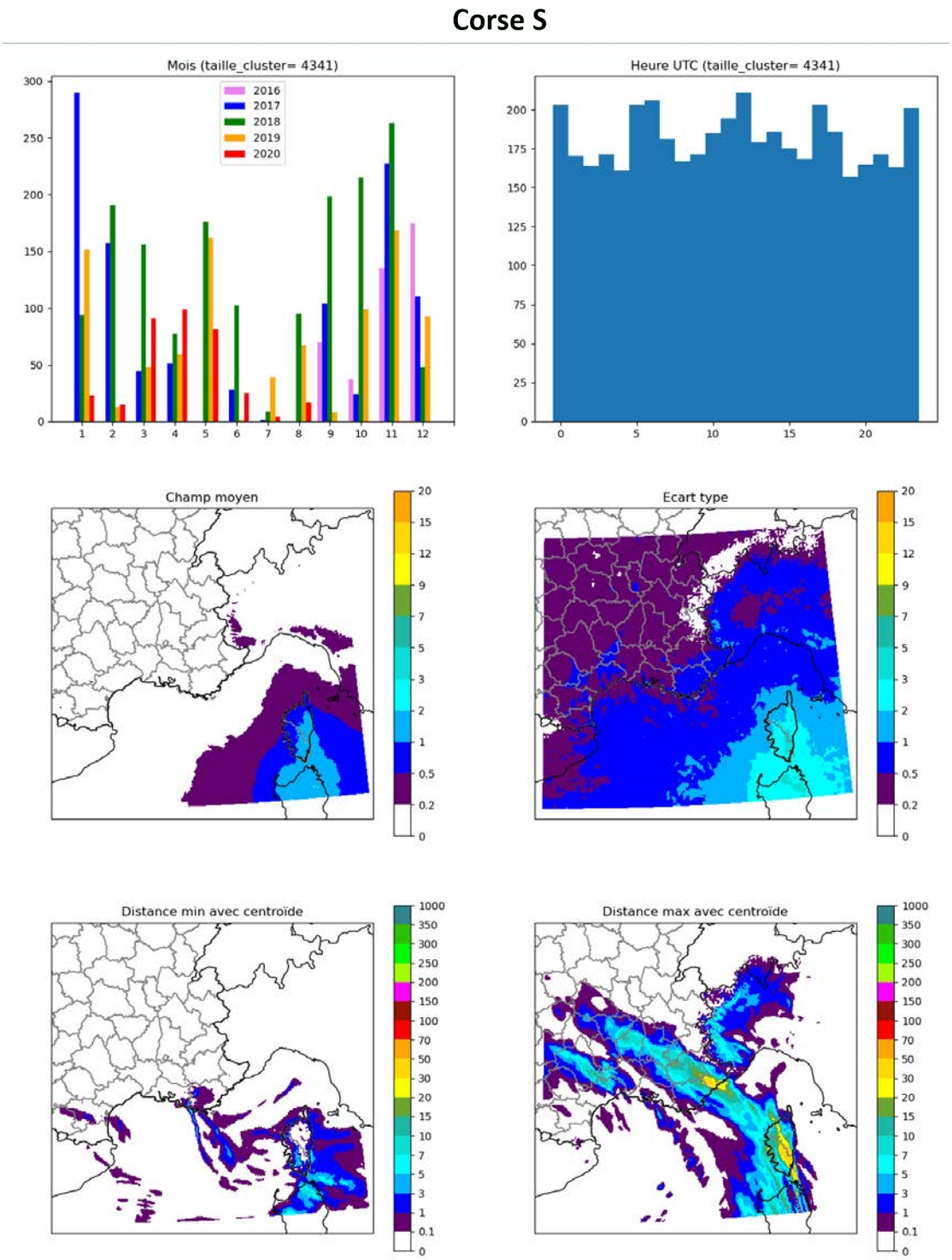


FIGURE E.12 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°10.*

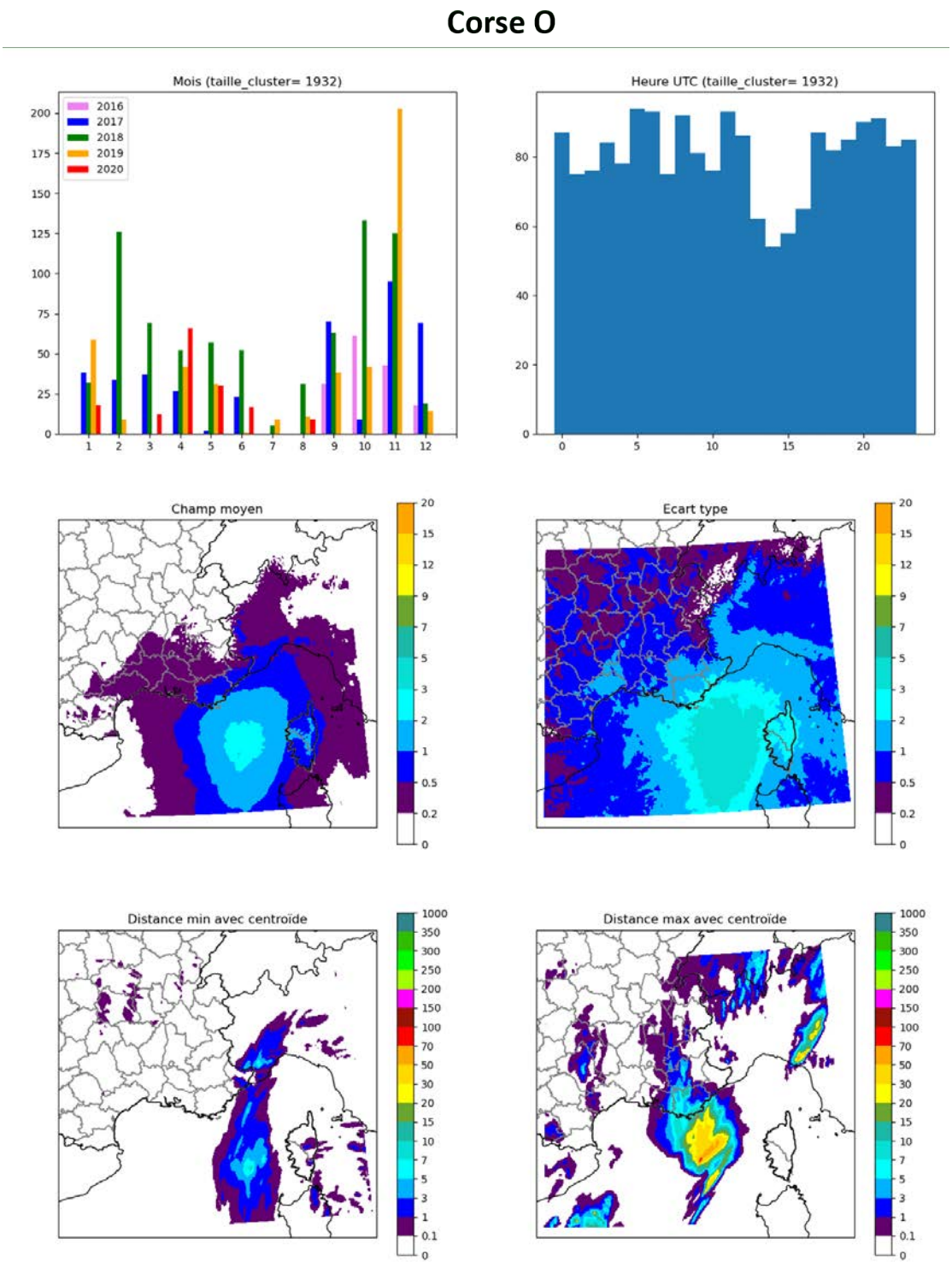


FIGURE E.13 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°11.*

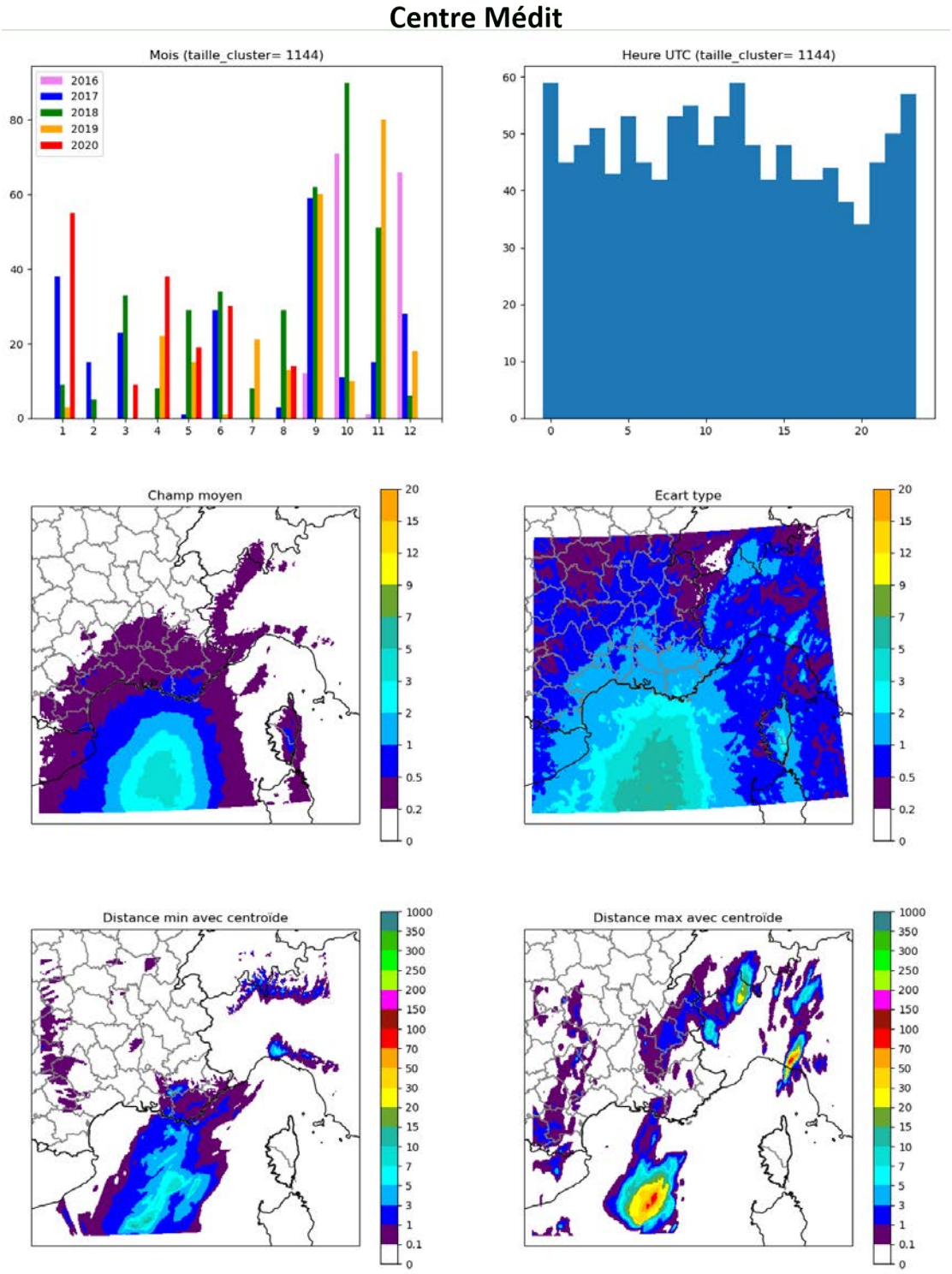


FIGURE E.14 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°12.*

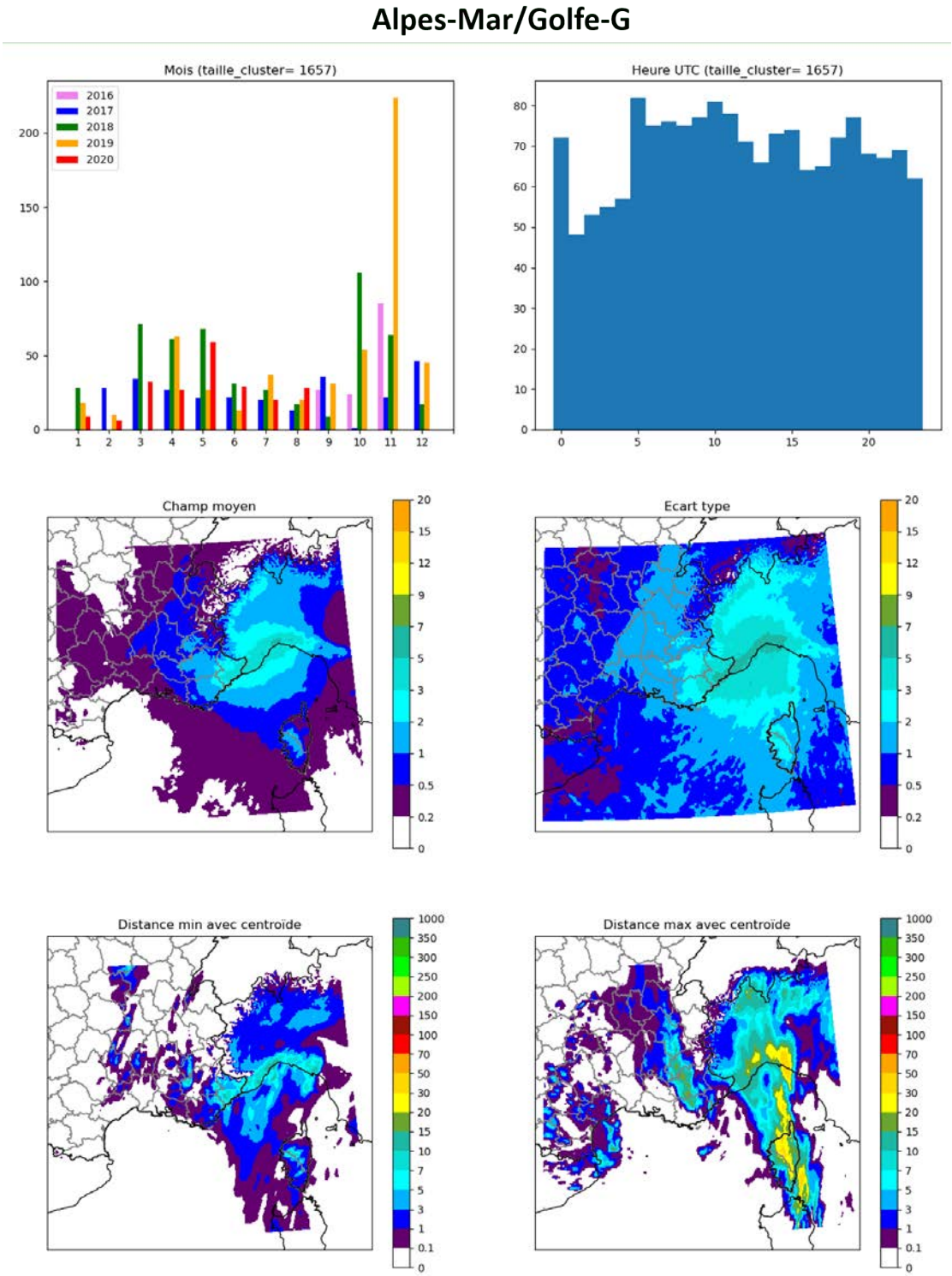


FIGURE E.15 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°13.*

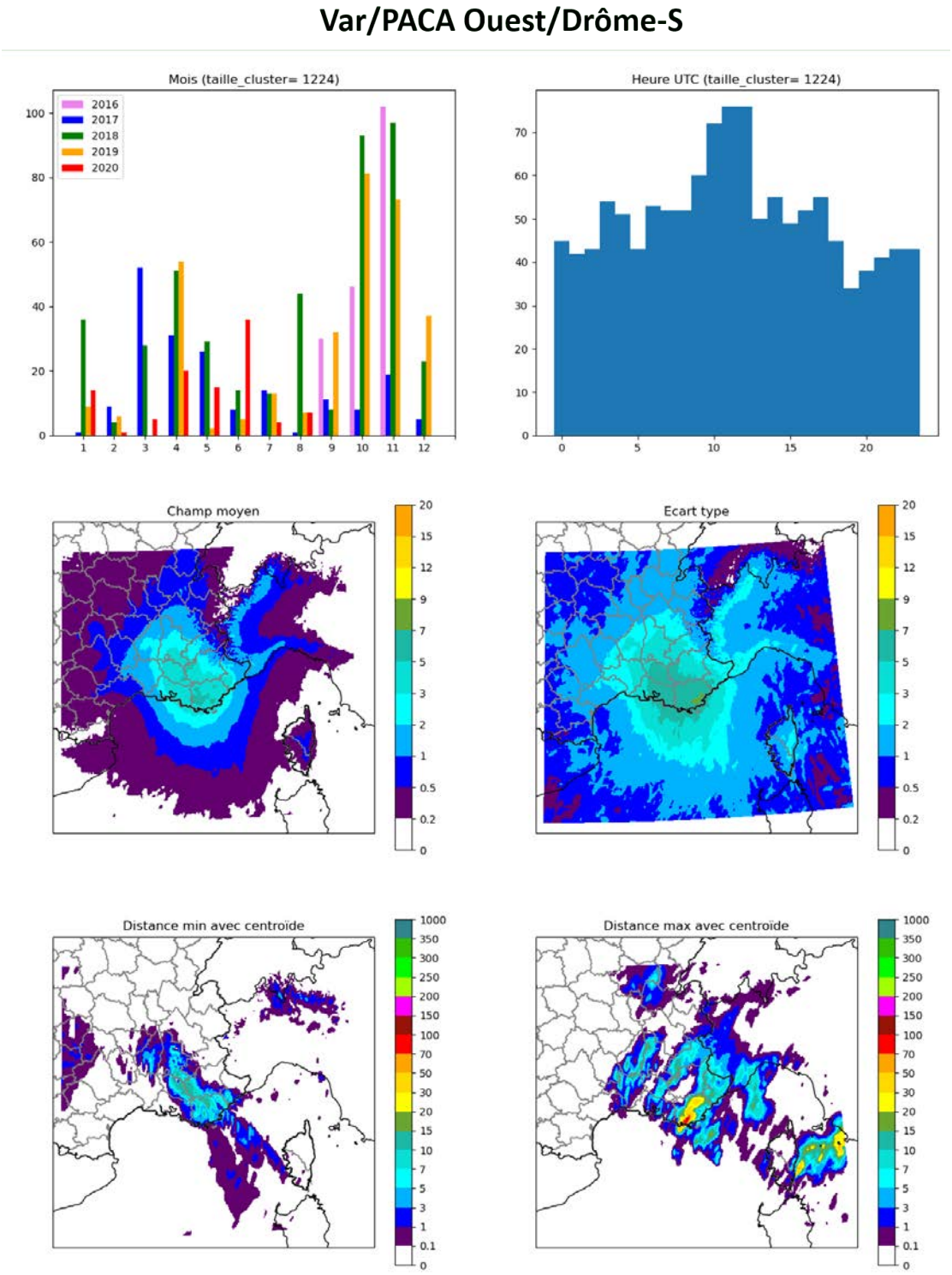


FIGURE E.16 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°14. PACA désigne la région Provence-Alpes-Côte d'Azur.*

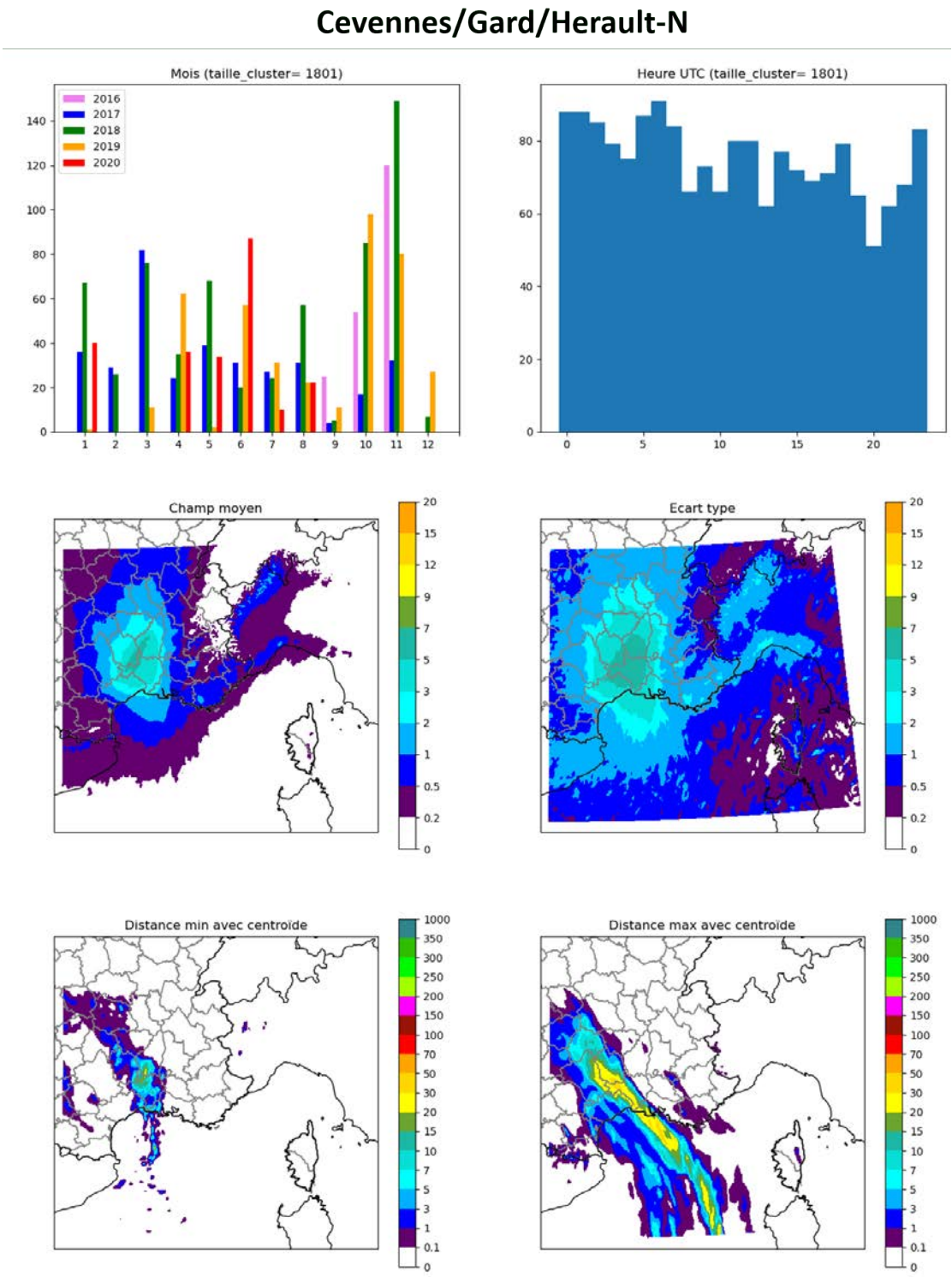


FIGURE E.17 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°15.*

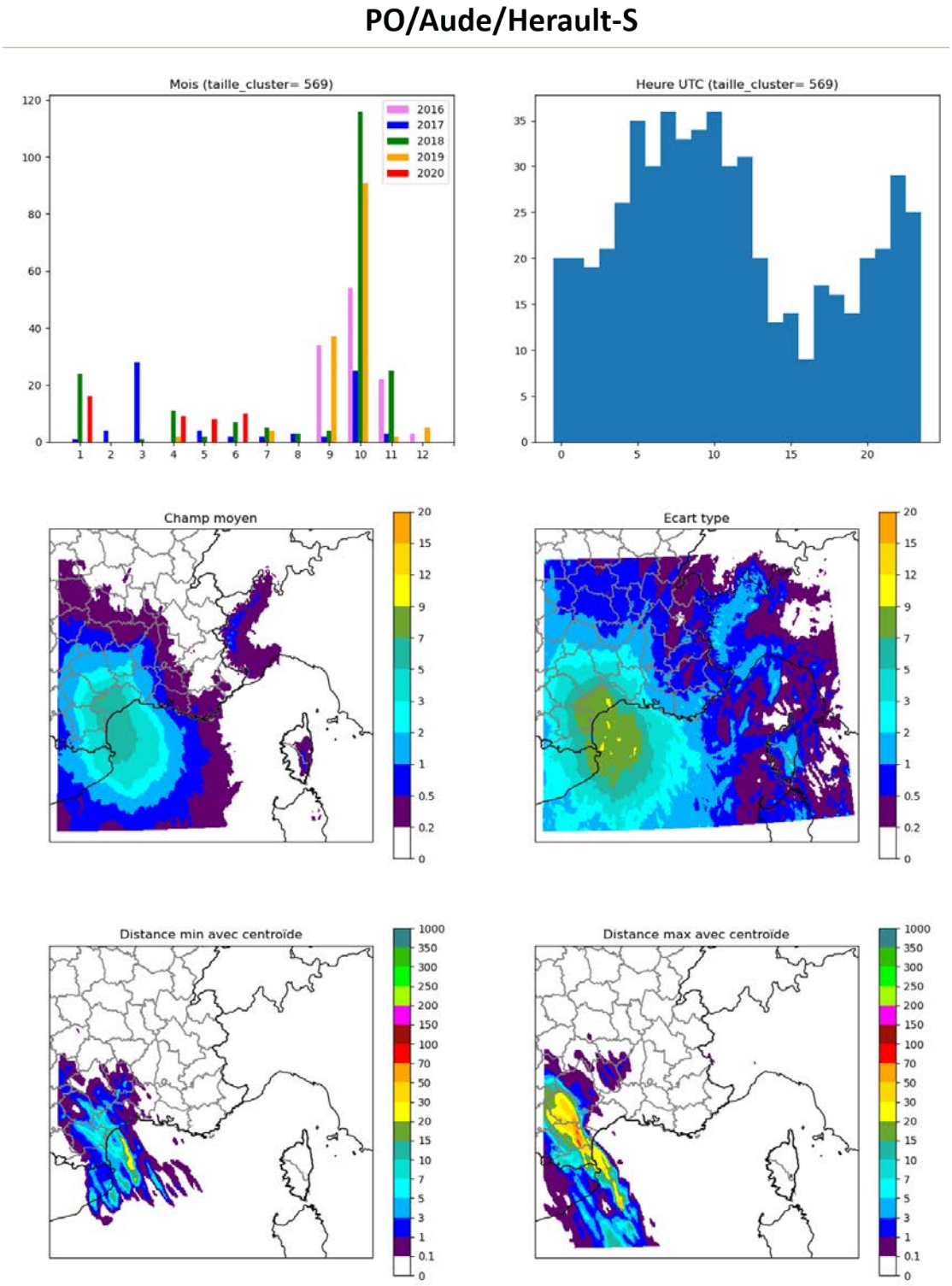


FIGURE E.18 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°16. PO désigne le département des Pyrénées-Orientales.*

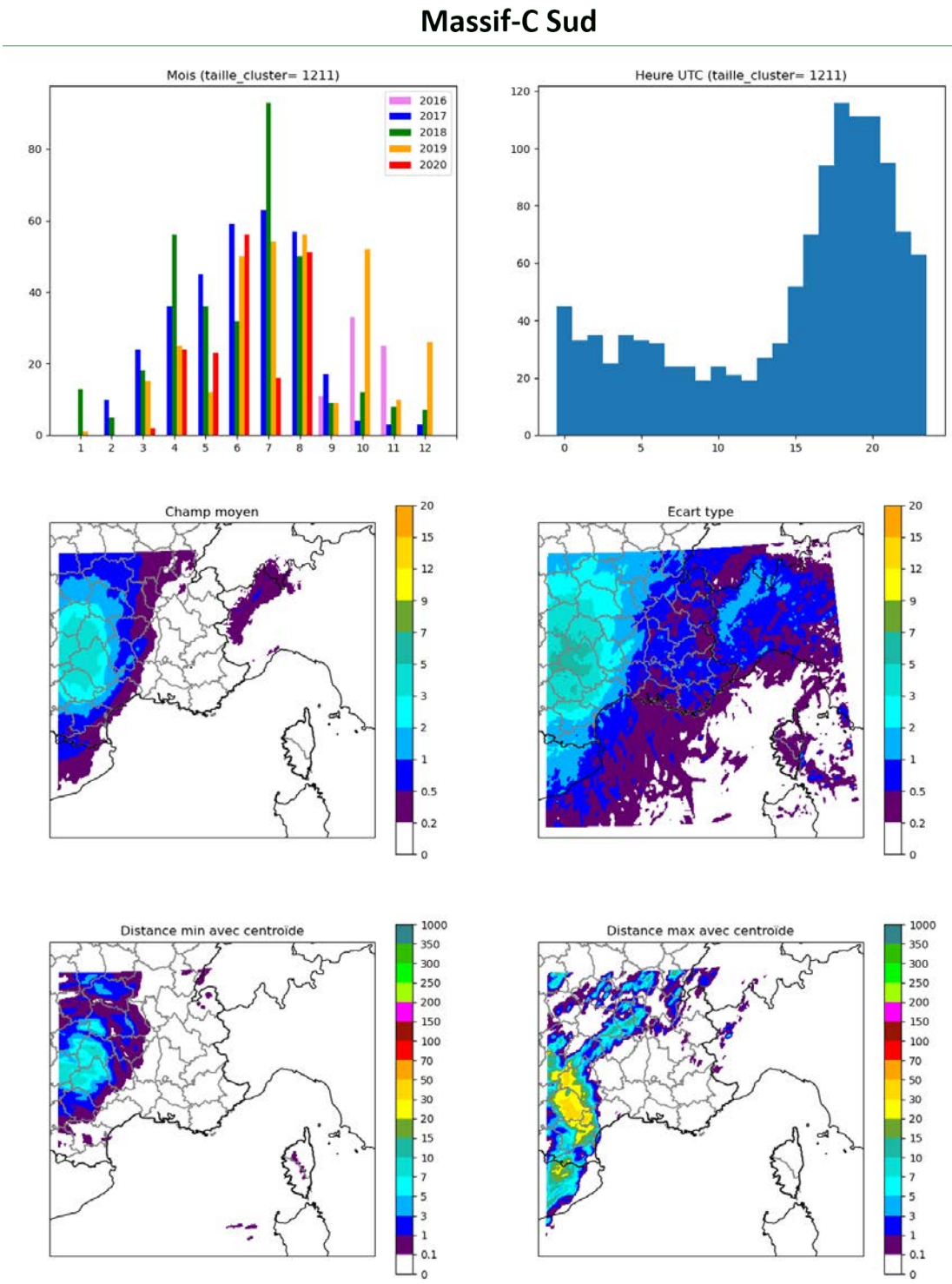


FIGURE E.19 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°17.*

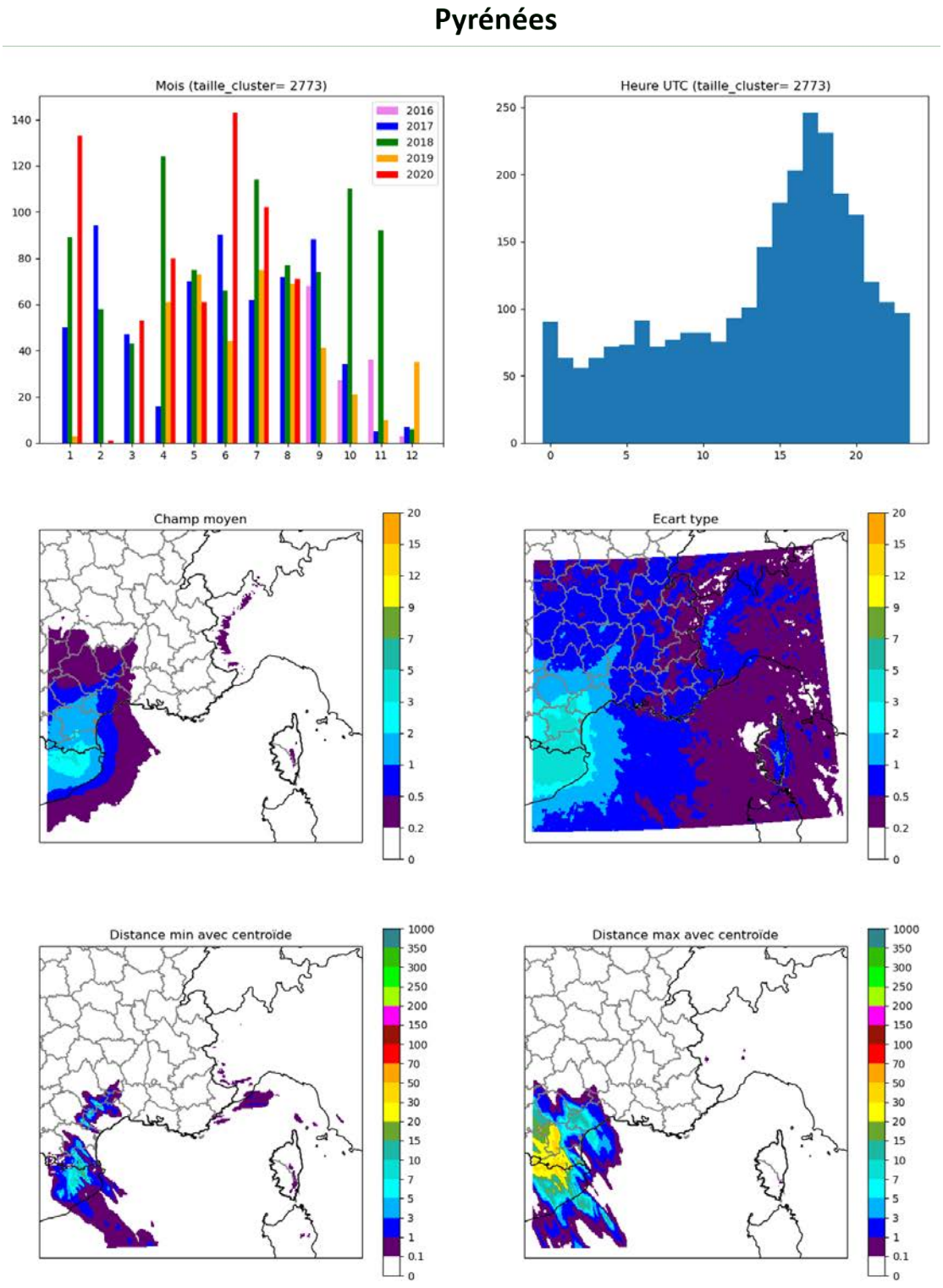


FIGURE E.20 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°18.*

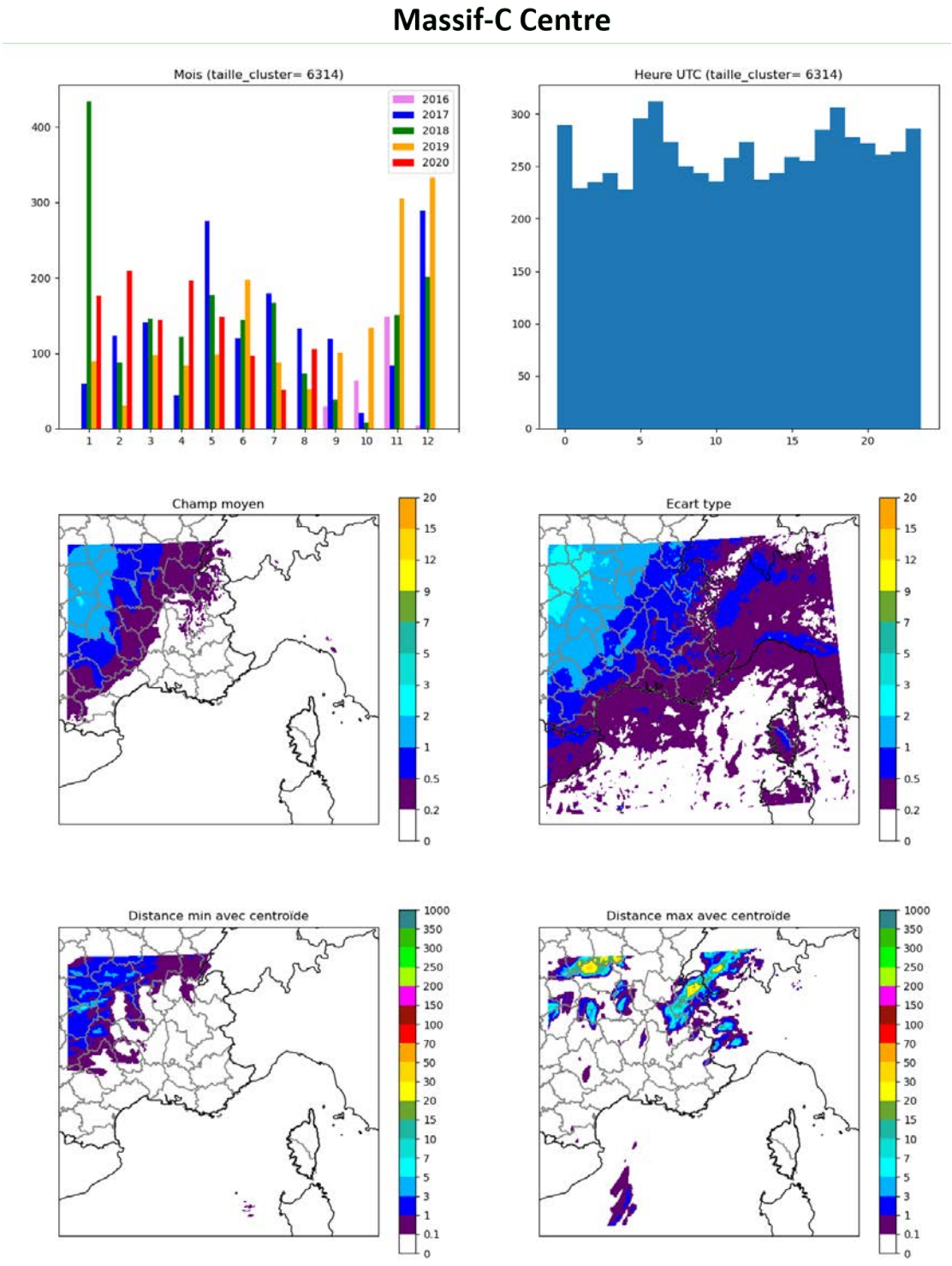


FIGURE E.21 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°19.*

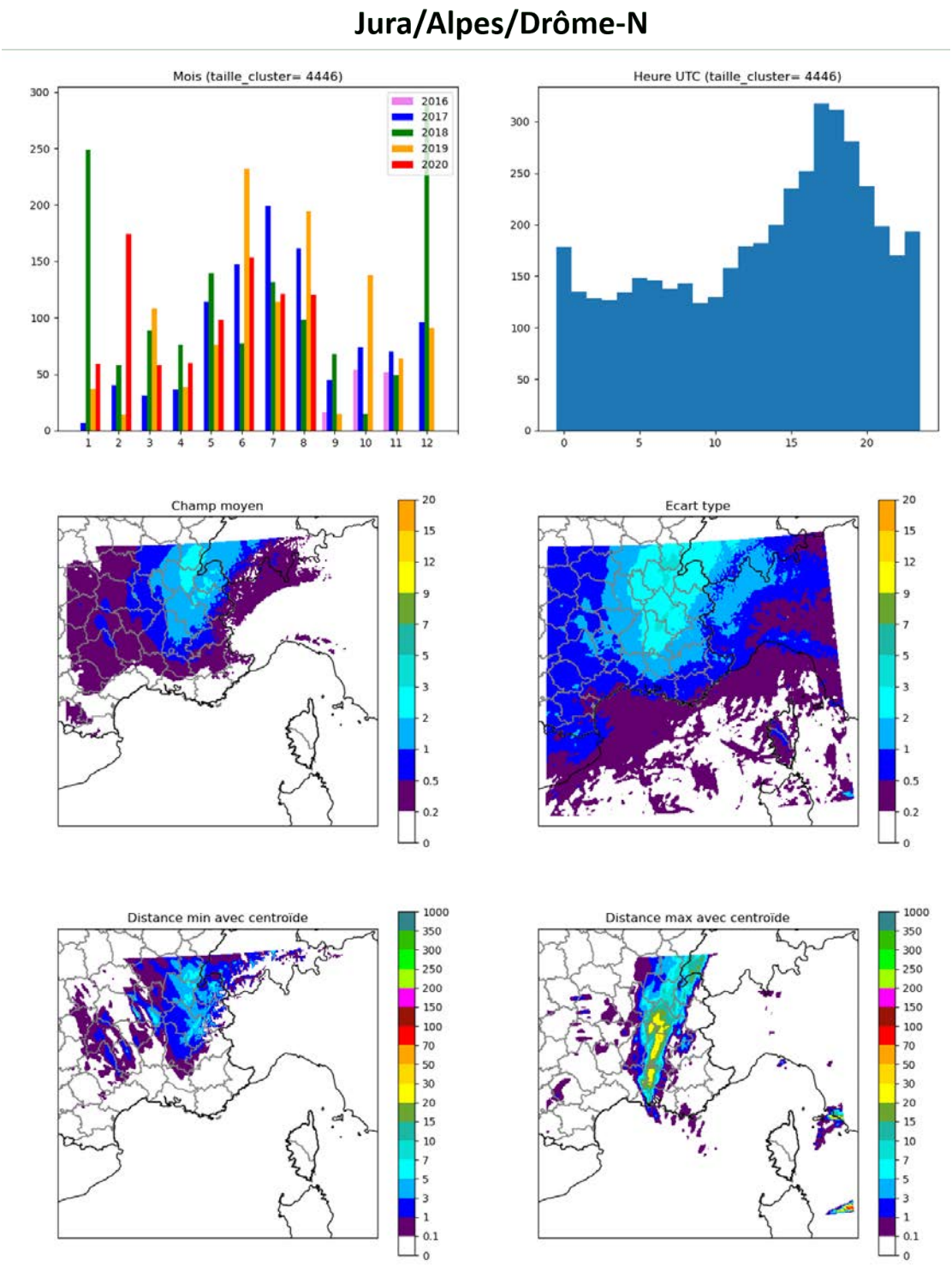


FIGURE E.22 – *Classification pour un espace latent à 40 dimensions avec un total de 20 classes. Classe n°20.*