



HAL
open science

Apprentissage profond pour la prévision de l'activité géomagnétique

Guillaume Bernoux

► **To cite this version:**

Guillaume Bernoux. Apprentissage profond pour la prévision de l'activité géomagnétique. Physique de l'espace [physics.space-ph]. Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-SUPAERO); Université de Toulouse, 2022. Français. NNT : 2022ESAE0037 . tel-03923415

HAL Id: tel-03923415

<https://theses.hal.science/tel-03923415v1>

Submitted on 4 Jan 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.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut Supérieur de l'Aéronautique et de l'Espace

Présentée et soutenue par :

Guillaume BERNOUX

le mercredi 6 juillet 2022

Titre :

Apprentissage profond pour la prévision de l'activité
géomagnétique

École doctorale et discipline ou spécialité :

ED SDU2E : Astrophysique, sciences de l'espace, planétologie

Unité de recherche :

Équipe d'accueil ISAE-ONERA PSI

Directeur(s) de Thèse :

Mme Angelica SICARD (directrice de thèse)

Mme Miho JANVIER (co-directrice de thèse)

Jury :

Mme Carine BRIAND Astronome LESIA - Présidente

M. Antoine BRUNET Ingénieur de recherche ONERA - Co-encadrant de thèse

Mme Dominique FONTAINE Directrice de recherche LPP Palaiseau - Rapporteur

Mme Miho JANVIER Astronome adjointe I.A.S - Co-directrice de thèse

M. Benoît LAVRAUD Directeur de recherche LAB Bordeaux - Rapporteur

Mme Angelica SICARD Ingénieure de recherche ONERA - Directrice de thèse



Daniel Bernard, *Irradiation*, 2021, acrylique et encre sur lavis. Reproduit avec l'aimable autorisation de l'artiste.

Œuvre inspirée des recherches présentées dans ce manuscrit.



Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International](#). Cette licence ne s'applique pas au matériel de tiers intégré à cette œuvre.

LAWRENCE — Oh thanks Dryden, this is going to be fun.
MR. DRYDEN — Lawrence, only two kinds of creature get
fun in the desert: Bedouins and gods, and you're neither. Take
it from me, for ordinary men, it's a burning, fiery furnace.
LAWRENCE — No, Dryden, it's going to be fun.
MR. DRYDEN — It is recognized that you have a funny
sense of fun.

(extrait de Lawrence d'Arabie, 1962)

Remerciements

Un jour, Antoine m'a dit : « Non mais tu verras, un manuscrit de thèse, ça se rédige en six semaines ». Ce jour-là j'ai compris que son optimisme pouvait atteindre des proportions véritablement cosmiques. Mais il faut bien avouer que cet enthousiasme débordant est une des qualités qui contribue à faire d'Antoine un encadrant de thèse particulièrement remarquable. Bien sûr, cet homme est également un véritable puits de science, qui possède de plus la générosité de partager ses connaissances inconditionnellement. Mais c'est aussi quelqu'un d'attentif, qui a toujours su trouver les bons mots pour me motiver lorsque l'attente de résultats nouveaux commençait à se faire pesante. Pour tout cela, et bien plus encore, je lui suis extrêmement reconnaissant.

J'adresse également de chaleureux remerciements à Angélica, Miho et Éric, sans la supervision et les conseils desquels le manuscrit que vous vous apprêtez (peut-être) à feuilleter aurait été bien différent. Merci en particulier à Angélica pour les encouragements, les conseils et l'enthousiasme, notamment pendant la dernière (longue) ligne droite, source de cet indispensable surplus d'énergie qui m'a souvent aidé à ne jamais baisser les bras. Bien que la pandémie de COVID-19 ait limité les opportunités de nous voir en personne plus souvent, je tiens à remercier vivement Miho et Éric pour leur indispensable regard éclairé et la fertilité de nos échanges.

Merci à Dominique Fontaine et Benoît Lavraud, qui ont rapporté ce manuscrit, à Carine Briand, qui a présidé le jury de soutenance, ainsi qu'à Olivier Pannekoucke pour sa participation au jury en tant que membre invité. Leurs questions, remarques et conseils ont été riches et m'ont permis de jeter un regard nouveau sur mes propres travaux. Merci aussi à Philippe Garnier et à Nicolas Balcon d'avoir accompagné cette thèse en prenant part à son comité de suivi. Un merci spécial à Daniel Bernard, qui s'est intéressé à mes travaux de recherche en y apportant son regard et sa sensibilité artistique pour en tirer la toile reproduite au début de ce manuscrit (!).

Bien qu'une thèse soit un effort individuel, je suis heureux que dans mon cas cela n'ait pas été un travail solitaire, et ce malgré le COVID. Cela, je le dois bien sûr à mes encadrants, mais aussi au reste de la formidable équipe du DPHY/ERS. Bien que ça m'en coûte, impossible de ne pas remercier sincèrement Vincent, mentor officieux à la bonne humeur bien officielle. Sans lui et la confiance qu'il m'a accordée dès ce premier projet de recherche à SUPAERO, impossible de savoir où je serai aujourd'hui. Seules certitudes : certainement pas derrière le bureau que j'occupe actuellement et sans doute très loin des ceintures de radiations.

Mes remerciements les plus vifs vont aussi à Sébastien, toujours incroyablement disponible et attentif à la bonne intégration des doctorants dans l'équipe et à la cohésion de celle-ci, une des raisons pour lesquelles je n'ai pas hésité à postuler lorsque l'opportunité de continuer mon aventure au DPHY/ERS s'est présentée. Je le remercie donc, ainsi qu'Henry et Olivier, pour leur confiance renouvelée. Merci également à Didier L., à qui j'adresse mes pensées amicales les plus réconfortantes en cette période difficile. Enfin, merci à Antoine F., Nour et Pablo, qui complètent cette joyeuse équipe, pour les conseils, le soutien et les moments partagés autour de boissons caféinées (mais pas merci pour les parties de baby-foot).

Étendons ces remerciements à l'ensemble de l'antenne toulousaine du DPHY et à tous ceux qui la composent. Un merci tout particulier à Caroline, Nathalie, Jean-Pierre et aux néo-retraitées Gisèle et Christine, grâce à qui j'ai eu à fort peu me préoccuper de la majorité des aspects administratifs, matériels et financiers relatifs à cette thèse pour pouvoir me concentrer pleinement sur le scientifique – luxe non-négligeable !

En restant dans les remerciements collectifs, mention spéciale à mes camarades de galère Lucas, Gaëtan et Julie, sans oublier mon inévitable et incorrigible co-bureau, Quentin. C'est toujours plus agréable de ramer à plusieurs, surtout lorsqu'on fait partie d'un équipage aussi sympathique que celui-ci. Cela vaut aussi pour tous les autres stagiaires, (post-)doctorants et alternants du département : Maxime, Martin, Gwendoline, Rabia et Maria (a.k.a. les autres chanceux ERSiens) mais aussi Lucas et Quentin bis, Carla, Paul, Nathan, Agnès, Lauriane, Rémi et tous les autres. Une pensée particulière pour Maria, qui, avec les encadrants de thèse qui lui ont été attribués (ceux qui savent savent), va avoir bien du pain sur la planche ! J'en profite également pour saluer amicalement les (plus ou moins) anciens : Loanne, Laura, Marine, Kévin, Eudes, Abdessamad, Maxime L., Pauline, Guillaume, Neil et Adrien, ainsi que Rafael, Hector et Mathias, que j'ai eu la chance de côtoyer (plus ou moins longtemps) au cours de cette thèse.

Ensuite, ils n'ont rien à voir avec cette thèse, en tout cas pas directement, mais impossible de ne pas profiter de ces quelques lignes pour les saluer. Je parle bien sûr de tous mes amis, qu'ils soient d'hier, d'aujourd'hui ou de demain. Comme, me connaissant, je vais sans doute en oublier, je ne me risquerai pas à tenter une liste de noms. Allez, si, faisons une petite exception* et rendons hommage aux éternels Antonin, Félix et Julien : c'est assez fou une amitié qui dure depuis aussi longtemps, non ?

Si j'ai l'opportunité aujourd'hui d'écrire maladroitement ces quelques lignes (déjà trop longues), c'est parce qu'ils ont toujours cru en moi et m'ont toujours soutenu inconditionnellement. Il n'y aurait pas d'hyperbole assez forte pour exprimer toute la gratitude que j'éprouve envers eux. Alors ne cherchons pas à faire compliqué : merci pour tout, maman, papa et Isabelle. Ce manuscrit vous est dédié. Isa, bravo pour ton courage et ta détermination à te donner les moyens d'accomplir tes rêves, ce sont des qualités qui m'ont inspiré tout au long de cette thèse et qui continuent de m'inspirer au quotidien.

*. Et une seule, désolé Crooks.

Enfin, rendre ce manuscrit dans les temps fait partie des nombreuses choses que je n'aurais pas pu faire sans toi, Clémence. Impossible de résumer ici, en à peine quelques lignes, ce que je tiens à te dire. Trichons donc un peu, et faisons appel au pouvoir évocateur du cinéma, et en particulier de cette scène magnifique concluant *Only Lovers Left Alive*, qui résume, de manière bien plus élégante, ce que je voulais exprimer :

EVE — Tell me now about entanglement. Einstein's spooky action at a distance. Is it related to quantum theory?

ADAM — Hm. No I mean, it's not a theory, it's proven.

EVE — How does it go again?

ADAM — When you separate an entwined particle and you move both parts away from the other, even at opposite ends of the universe, if you alter or affect one, the other will be identically altered or affected. Spooky.

EVE — Even at opposite ends of the universe?

ADAM — Yeah.

Merci Jim, Tom et Tilda pour ces mots et images que je vous emprunte, et surtout merci, Clémence, pour tout le reste. Que notre aventure continue, même depuis l'autre bout de l'Univers !

Avant-propos

Ce travail a été réalisé au cours d'une thèse de doctorat s'étant déroulée à l'Office national d'études et de recherches aérospatiales (ONERA), au sein du Département physique, instrumentation, environnement, espace (DPHY), dans l'unité Environnement Radiatif Spatial. Cette thèse a été encadrée par Antoine BRUNET (Ingénieur de Recherche, ONERA/DPHY), dirigée par Angélica SICARD (Ingénieure de Recherche, ONERA/DPHY) et par Miho JANVIER (Astronome Adjointe, Institut d'Astrophysique Spatiale) et co-supervisée par Éric BUCHLIN (Chargé de Recherche, Institut d'Astrophysique Spatiale).

Cette thèse a bénéficié d'un co-financement de l'ONERA et de la Région Occitanie.

Le Jury de soutenance était composé de six membres et deux invités :

Mme Carine BRIAND, Laboratoire d'études spatiales et d'instrumentation en astrophysique, Présidente du Jury ;

M. Benoît LAVRAUD, Laboratoire d'Astrophysique de Bordeaux, Rapporteur ;

Mme Dominique FONTAINE, Laboratoire de Physique des Plasmas, Rapporteur ;

Mme Angélica SICARD, ONERA/DPHY, Directrice de thèse ;

Mme Miho JANVIER, Institut d'Astrophysique Spatiale, Co-directrice de thèse ;

M. Antoine BRUNET, ONERA/DPHY, Co-encadrant de thèse ;

M. Éric BUCHLIN, Institut d'Astrophysique Spatiale, Invité ;

M. Olivier PANNEKOUCKE, INPT ENM/CNRM/CERFACS, Invité.

Résumé

L'étude des interactions Soleil–Terre, en particulier par le biais du couplage entre le vent solaire et la magnétosphère, est au cœur des enjeux liés à la météorologie de l'espace. Nous nous intéressons à la question de la prédiction à quelques jours d'indices géomagnétiques, qui peuvent servir à piloter les modèles de ceintures de radiations terrestres. Au cours de la dernière décennie, de nombreuses études ont montré que les réseaux de neurones artificiels permettaient de prédire ces indices de manière particulièrement performante, à partir des mesures du vent solaire proche de la Terre.

Au cours de nos travaux nous proposons d'abord un nouveau modèle de prédiction de l'indice géomagnétique *Dst*, composé d'un réseau de neurones possédant des couches récurrentes. Ce nouveau modèle produit des prédictions probabilistes plus performantes que l'état de l'art actuel pour des horizons de prédiction inférieurs à 6 heures. Afin de rendre notre modèle plus utile opérationnellement, nous l'adaptions pour la prédiction du nouvel indice géomagnétique *Ca*, conçu pour mieux rendre compte de la géoefficacité des événements géomagnétiques du point de vue des ceintures de radiations électroniques. En menant une évaluation complète de notre modèle, nous montrons qu'il perd de son utilité dans un contexte opérationnel pour les horizons de prédiction supérieurs à quelques heures.

Partant de ce constat, et face aux limites montrées par les modèles physiques de propagation du vent solaire actuels, nous étudions l'utilisation d'imagerie solaire pour prédire directement l'indice géomagnétique *Kp* de 2 à 7 jours en avance. Pour cela, nous construisons **SERENADE**, le premier modèle de prédiction d'un indice géomagnétique alimenté uniquement par des images du Soleil. Ce modèle est un réseau de neurones à l'architecture complexe combinant des couches de différentes natures. Nous montrons que notre modèle présente des performances au moins égalant celles des modèles empiriques simples (et pourtant actuellement les plus performants) de prédiction du maximum journalier de *Kp*. Nous mettons en évidence que celui-ci, bien qu'encore immature pour une utilisation en contexte opérationnel, est capable de rendre compte de la géoefficacité de certains événements solaires directement à partir de la seule imagerie solaire. En identifiant les limites de notre modèle et leurs causes, nos résultats ouvrent la voie à une modélisation par les données des interactions Soleil–Terre complétant les modèles physiques actuels.

Mots clés : météorologie de l'espace, apprentissage automatique, vent solaire, magnétosphère, ceintures de radiations, imagerie solaire

Abstract

Forecasting the geomagnetic activity with Deep Learning techniques

The study of Sun–Earth interactions, in particular through the solar wind–magneto-sphere coupling, is at the core of Space Weather-related issues. In this work, we focus on the forecasting of geomagnetic indices within a few days, which can be used to drive the Earth’s radiation belt models. During the last decade, many studies have shown that artificial neural networks can predict these indices in a very efficient way, based on measurements of the solar wind near the Earth.

In our work, we first propose a new model for the prediction of the geomagnetic index Dst , consisting of a neural network with recurrent layers. This new model produces better probabilistic forecasts than the current state of the art for prediction horizons shorter than 6 hours. In order to make our model more operationally useful, we adapt it for the prediction of the new geomagnetic index Ca , designed to better account for the geoeffectiveness of geomagnetic events from the perspective of the electron radiation belts. By conducting a comprehensive evaluation of our model, we show that it loses its usefulness in an operational context for prediction horizons longer than a few hours.

Based on this observation, and facing the limits shown by the current physical models of solar wind propagation, we study the use of solar imaging to directly forecast the geomagnetic index Kp from 2 to 7 days ahead. To do so, we build **SERENADE**, the first geomagnetic index forecasting model driven only by images of the Sun. This model is a neural network with a complex architecture combining layers of different nature. We show that our model performs at least as well as some simple empirical models (and yet currently the most efficient) for forecasting the daily maximum of Kp . We show that our model, although still immature for an operational context, is able to account for the geoeffectiveness of some solar events directly from solar imagery alone. By identifying the limitations of our model and their causes, our results open the way to data-driven modeling of the Sun–Earth interactions as a complement to the current physical models.

Keywords: space weather, machine learning, solar wind, magnetosphere, radiation belts, solar imaging

Table des matières

Avant-propos	ix
Résumé	x
Table des matières	xii
Table des figures	xvi
Liste des tableaux	xxiv
Liste des acronymes	xxvi
Notations	xxxiii
Introduction	1
1 Introduction aux interactions Soleil–Terre et à la météorologie de l’espace	5
Préambule	6
1.1 Introduction à l’activité solaire	7
1.1.1 La structure du Soleil	7
1.1.2 Le cycle solaire	9
1.1.3 Le(s) vent(s) solaire(s)	10
1.1.4 L’activité éruptive du Soleil	15
1.1.5 Le vent solaire à 1 au	18
1.2 Les ceintures de radiations, au cœur de la magnétosphère terrestre	21
1.2.1 Morphologie de la magnétosphère terrestre	21
1.2.2 L’activité géomagnétique	29
1.3 Les interactions Soleil–Terre vues du point de vue des activités humaines	32
1.3.1 Les conséquences de l’activité géomagnétique et de l’environnement radiatif spatial sur les activités humaines	33
1.3.2 La météorologie de l’espace	35
1.4 Bilan du chapitre	35
2 Étudier le système Soleil–Terre par le prisme des données	37
Préambule	38
2.1 Introduction à la modélisation par les données	39
2.1.1 Généralités sur la science dirigée par les données	39
2.1.2 Principes fondamentaux de l’apprentissage automatique	41
2.1.3 Du neurone formel à l’apprentissage profond	43

2.1.4	Les réseaux récurrents	55
2.1.5	Les réseaux convolutifs	59
2.1.6	La gestion des données en apprentissage profond	61
2.1.7	Autres aspects relatifs à l’implémentation et difficultés potentielles	64
2.1.8	Bilan sur la modélisation par les données et l’apprentissage profond	65
2.2	Les données à notre disposition	65
2.2.1	Observer le Soleil	65
2.2.2	Étudier le vent solaire	77
2.2.3	Les indices géomagnétiques	79
2.2.4	Mesurer les particules au cœur des ceintures	82
2.2.5	Bilan sur l’observation du système Soleil–Terre	85
2.3	Modéliser les interactions Soleil–Terre	85
2.3.1	Des modèles pour la prédiction du vent solaire dans l’héliosphère	85
2.3.2	Prédiction d’indices géomagnétiques : des modèles empiriques au <i>Deep Learning</i>	87
2.3.3	Modélisation des ceintures de radiations	89
2.4	Bilan du chapitre	90
3	Prédire l’activité géomagnétique terrestre à partir du vent solaire	93
	Préambule	94
3.1	Reproduction du modèle de GRUET et al. 2018	95
3.1.1	Étude de l’indice <i>Dst</i>	95
3.1.2	Description du modèle de GRUET et al. 2018 et de notre reproduction	97
3.1.3	Résultats obtenus avec notre reproduction du modèle	100
3.2	Proposition d’un meilleur découpage de sous-ensembles de données	103
3.3	Améliorations du modèle de prédiction de <i>Dst</i>	107
3.3.1	Changements dans l’architecture et l’entraînement du réseau de neurones	107
3.3.2	Changements dans la gestion des données manquantes	111
3.4	Un nouveau banc d’évaluation pour mieux caractériser le modèle	113
3.4.1	Évaluation de la précision en période d’orage géomagnétique	115
3.4.2	Évaluation à l’aide de métriques dédiées au problème de classification binaire	116
3.4.3	Évaluation du modèle par détection des débuts d’orage	119
3.4.4	Évaluation des distorsions temporelles entre les prédictions et les observations	121
3.4.5	Application à des horizons de prédiction plus éloignés	123
3.4.6	Bilan sur l’extension du banc d’évaluation des prédictions de <i>Dst</i>	124
3.5	Obtenir des prédictions probabilistes directement avec un réseau de neurones	126
3.5.1	Description et justification de notre approche	126
3.5.2	Présentation et analyse des prédictions probabilistes	128

3.5.3	Bilan sur l'amélioration du modèle par la prise en compte de l'incertitude	134
3.6	Bilan du chapitre	136
4	Estimer la dynamique future des ceintures de radiations électroniques	137
	Préambule	138
4.1	Prendre en compte la géoefficacité du point de vue des ceintures de radiations	139
4.1.1	Les limites des approches existantes	139
4.1.2	Définition de l'indice Ca	140
4.1.3	Étude statistique de Ca	141
4.2	Prédiction de l'indice Ca	150
4.2.1	Les données utilisées	150
4.2.2	Le modèle	152
4.2.3	Évaluation du modèle	156
4.2.4	Analyse des résultats	158
4.2.5	Bilan sur la prédiction de Ca	166
4.3	Bilan du chapitre	167
5	Prédire la géoefficacité d'événements solaires à partir d'images du Soleil : une nouvelle méthode	169
	Préambule	170
5.1	Observations solaires et apprentissage automatique : une revue	171
5.1.1	Transformation des données	171
5.1.2	Détection d'objets	173
5.1.3	Prédiction de signaux physiques	174
5.2	Philosophie de notre nouvelle approche	176
5.3	Présentation de l'architecture de SERENADE	177
5.3.1	L'architecture globale	178
5.3.2	Les données	180
5.3.3	Le réseau de neurones	184
5.3.4	L'apprentissage du modèle	190
5.4	Méthodologie de l'évaluation	199
5.4.1	Les modèles de référence	199
5.4.2	Le banc d'évaluation	202
5.5	Bilan du chapitre	203
6	Prédire la géoefficacité d'événements solaires à partir d'images du Soleil : premiers résultats	205
	Préambule	206
6.1	Analyse des résultats obtenus sur l'ensemble de la période 2010 – 2018	206
6.1.1	Résultats pour le problème de régression	208
6.1.2	Résultats pour le problème de classification	211
6.1.3	Bilan de l'évaluation de SERENADE sur l'ensemble de la période 2010 – 2018	215
6.2	Analyse des performances selon la phase du cycle solaire	215
6.3	Prédiction de l'indice $Ca_{\max,24h}$	222
6.3.1	Modifications du <i>pipeline</i>	222

6.3.2	Évaluation des performances de SERENADE- <i>Ca</i>	223
6.4	Estimation de l'influence de la période utilisée sur les performances	228
6.5	Bilan du chapitre	230
Conclusions		233
	Bilan général de nos travaux	233
	Perspectives futures	235
	À propos des modèles s'appuyant sur les paramètres du vent solaire	235
	À propos de SERENADE	236
	À propos du déploiement en contexte opérationnel	237
Bibliographie		239
Annexes		263
A	Description de la méthode de reconstruction des données manquantes par analyse spectrale singulière	263
B	Éléments supplémentaires d'évaluation du modèle SERENADE- <i>Ca</i>	265
C	Articles publiés dans des revues à comité de lecture	267
	C.1 BERNOUX ET MAGET, 2020	267
	C.2 BERNOUX ET AL. 2021	267
	C.3 BERNOUX ET AL. 2022	268

Table des figures

1.1	Schéma de la structure en couches du Soleil. Source : Wikimedia Commons, sous licence CC BY-SA 3.0.	8
1.2	Images de différentes couches du Soleil prises au même instant le 15 février 2022. Crédits : SDO (NASA).	9
1.3	Images de la photosphère solaire le 8 février 2022. Crédits : SDO (NASA). Images générées via HELIOVIEWER.	10
1.4	Évolution du nombre de taches solaires moyenné sur 13 mois entre 1749 et 2019.	11
1.5	Images du Soleil prises chaque année du cycle solaire 23 à la longueur d'onde 171 Å. Crédits : SOHO (ESA & NASA).	11
1.6	Représentation schématique de la propagation du vent solaire dans l'héliosphère suivant la spirale de Parker. Plusieurs trajectoires sont tracées en fonction de la vitesse du vent solaire. L'orbite bleue correspond à celle de la Terre, la rouge à celle de Mars et la verte à celle de Jupiter. Reproduit de LHOTKA et NARITA 2019, sous licence CC BY 4.0.	12
1.7	Superposition de plusieurs images du Soleil à différentes longueurs d'onde (visible et EUV) capturées le 2 juillet 2019 depuis le sol et l'espace, montrant la structure des <i>helmet streamers</i> en minimum solaire. Des lignes de champ ouvertes partant des trous coronaux polaires sont également visibles. Crédits : ESO/P. Horálek ; SOHO (ESA & NASA) ; SDO (NASA), sous licence CC BY 4.0.	13
1.8	Images du Soleil montrant des exemples de trous coronaux. Crédits : SDO (NASA). Images générées via HELIOVIEWER.	14
1.9	Représentation schématique d'une SIR dans l'héliosphère. Figure reproduite de LANTOS 1997, adaptée d'une figure de CROOKER et SISCOE 1986.	14
1.10	Distribution annuelle de l'observation de vents solaires rapides à 1 au cours des cycles solaires 23 et 24. Reproduite (avec autorisation) de GRANDIN et al. 2019.	15
1.11	Image du Soleil à la longueur d'onde 131 Å prise le 9 août 2011 dans laquelle un embrillancement de classe X est visible. Crédits : SDO (NASA). Image générée via HELIOVIEWER.	16
1.12	Image du Soleil à la longueur d'onde 304 Å prise le 6 décembre 2010 dans laquelle une grande protubérance est visible dans le quart inférieur gauche de l'image. Crédits : SDO (NASA). Image générée via HELIOVIEWER.	17

1.13	Images superposées du Soleil (à la longueur d'onde 304 Å) et de sa couronne (en lumière blanche) prises le 27 février 2000 montrant une CME. Crédits : SOHO (ESA & NASA). Image générée via HELIOVIEWER.	18
1.14	Distribution annuelle de l'observation d'ICMEs dirigées vers la Terre à 1 au au cours des cycles solaires 23 et 24. Reproduite (avec autorisation) de GRANDIN et al. 2019.	19
1.15	Représentation schématique du champ magnétique terrestre, vu comme un dipôle incliné et décentré. Reproduit de DAHMEN 2020.	22
1.16	Carte de l'intensité du champ magnétique terrestre à sa surface selon le modèle IGRF. Reproduit de KOSKINEN et KILPUA 2022, sous licence CC BY 4.0.	22
1.17	Représentation schématique de la magnétosphère terrestre faisant apparaître ses différentes régions. Crédits : ESA/C. T. Russel.	23
1.18	Schéma de la magnétosphère et des grands systèmes de courants la composant. Figure de T. Mäkinen, reproduite (avec autorisation) de KOSKINEN 2011b.	25
1.19	Illustration des ceintures de radiations terrestres électroniques. Crédits : NASA illustrations, sous licence CC BY-NC-ND 2.0.	27
1.20	Représentations schématiques des trois mouvements quasi-périodiques associés au piégeage d'une particule dans les ceintures de radiations.	29
1.21	Schéma de la circulation de plasma dans la magnétosphère terrestre selon le modèle du cycle de Dungey. Les croix désignent les zones où les reconnexion magnétiques du côté jour et du côté nuit ont lieu. Reproduit de EASTWOOD et al. 2017, sous licence CC BY 4.0.	30
1.22	Photographie montrant des cellules solaires de la mission EURECA ayant subi une décharge électrostatique au cours de la mission. Crédits : ESA.	34
2.1	Schéma des différentes branches de l'apprentissage automatique, avec des exemples d'application.	44
2.2	Schéma d'un neurone biologique. Adapté de Wikimedia Commons, sous licence CC BY-SA 3.0.	45
2.3	Schéma d'un neurone formel.	45
2.4	Exemple de MLP possédant quatre couches, dont deux couches cachées.	46
2.5	Schéma simplifié du fonctionnement d'un réseau de neurones récurrent ou <i>Recurrent Neural Network</i> (RNN).	56
2.6	Schéma du fonctionnement interne d'une cellule LSTM.	57
2.7	Schéma d'un neurone convolutif à deux dimensions. Crédits : FIDLE, sous licence CC BY-NC-ND 4.0.	59
2.8	Schéma du fonctionnement d'une couche de convolution à un noyau de deux dimensions. Crédits : FIDLE, sous licence CC BY-NC-ND 4.0.	60
2.9	Schéma de l'architecture classique d'une réseau de type CNN. Crédits : FIDLE, sous licence CC BY-NC-ND 4.0.	61
2.10	Exemple de courbes de fonctions de perte montrant du surapprentissage.	63

2.11	Exemple d'image capturée par Yohkoh/SXT le 28 octobre 2001. Crédits : LMSAL. Image générée via HELIOVIEWER.	66
2.12	Exemples d'images capturées par SOHO/EIT aux quatre longueurs d'ondes possibles le 28 octobre 2001. Crédits : SOHO (ESA & NASA). Images générées via HELIOVIEWER.	68
2.13	Superposition d'images capturées par SOHO/EIT à 195 Å (vert), LASCO/C2 (rouge) et LASCO/C3 (bleu) le 28 octobre 2001. Image générée via HELIOVIEWER.	69
2.14	Exemples d'images (magnétogramme et continuum) capturées par SOHO/MDI le 28 octobre 2001. Crédits : SOHO (ESA & NASA). Images générées via HELIOVIEWER.	70
2.15	Exemples d'images capturées par SDO/AIA aux dix longueurs d'ondes possibles le 23 février 2014. Crédits : SDO (NASA). Images générées via HELIOVIEWER.	73
2.16	Exemples d'images (magnétogramme et continuum) capturées par SDO/HMI le 23 février 2014. Crédits : SDO (NASA). Images générées via HELIOVIEWER.	74
2.17	Observatoires dont les mesures sont utilisées pour produire l'indice <i>Kp</i> . Crédits : ISGI.	79
2.18	Observatoires dont les mesures sont utilisées pour produire les indices <i>am</i> et <i>aa</i> . Crédits : ISGI.	80
2.19	Observatoires dont les mesures sont utilisées pour produire l'indice <i>Dst</i> . Crédits : ISGI.	81
2.20	Couverture temporelle des couples satellite/instrument disponibles dans IPODE entre 1968 et 2008. Le cycle solaire est représenté par le flux radio du Soleil à 10,7 cm (tracé jaune). De haut en bas, les missions correspondant aux orbites suivantes sont représentées : 1) missions interplanétaires, 2) HEO-GTO, 3) GEO, 4) MEO, 5) LEO-PEO. © 2013 IEEE. Reproduit, avec autorisation, de LAZARO et al. 2013.	83
2.21	Exemple de mesures de flux de protons non-contaminées (a) comparées à des mesures contaminées (b). Les flux d'électrons contaminant les mesures de protons sont données en (c).	84
3.1	Évolution de l'indice <i>Dst</i> lors de l'orage géomagnétique du 15-16 juillet 2000.	96
3.2	Auto-corrélation temporelle de <i>Dst</i> pour des décalages temporels allant jusqu'à 30 jours.	97
3.3	Architecture de notre reproduction du réseau de neurones proposé par GRUET et al. 2018 pour la prédiction de <i>Dst</i> jusqu'à six heures en avance.	99
3.4	Exemples de prédictions jusqu'à six heures en avance obtenues avec notre reproduction du modèle de GRUET et al. 2018 lors d'un orage géomagnétique survenu fin avril 2012.	102
3.5	Statistiques annuelles de l'indice <i>Dst</i> entre 1995 et 2018 sous la forme de digrammes en boîtes. Le nombre de taches solaires lissé sur 13 mois est également donné comme indicateur du cycle solaire.	104

3.6	Courbe des valeurs prises par l'indice <i>Dst</i> entre 1995 et 2018 inclus (ligne noire fine), les ensembles d'entraînement (zone bleue), de validation (zone jaune) et d'évaluation (zone rouge) sont mis en évidence. Le nombre de taches solaires lissé sur 13 mois est également représenté comme indicateur du cycle solaire (ligne grise épaisse).	105
3.7	Exemples de prédictions jusqu'à six heures en avance obtenues avec notre reproduction du modèle de GRUET et al. 2018 lors de l'orage géomagnétique extrême de juillet 2000, suivi d'un orage modéré.	106
3.8	Architecture du réseau de neurones amélioré que nous proposons pour prédiction de <i>Dst</i> jusqu'à six heures en avance.	109
3.9	Exemples de prédictions jusqu'à six heures en avance obtenues avec notre modèle amélioré lors de l'orage géomagnétique extrême de juillet 2000, suivi d'un orage modéré.	114
3.10	Comparaison des valeurs de (a) NRMSE et de (b) corrélation linéaire obtenues avec le modèle de base et le modèle amélioré jusqu'à six heures en avance lorsqu'on sépare l'ensemble de test en trois sous-ensembles (périodes calmes, orages modérés et orages intenses).	116
3.11	Exemples de prédictions une heure et six heures en avance obtenues avec le modèle de base et notre modèle amélioré lors de l'orage géomagnétique extrême de juillet 2000.	120
3.12	Pourcentage de périodes d'orage correctement prédites avant leur début.	121
3.13	Exemple de mesure de distorsion temporelle entre deux séries temporelles par DTW. Reproduit de LI et al. 2019, sous licence CC BY-NC-SA 4.0.	122
3.14	Valeurs de TDM obtenues avec le modèle de base et le modèle amélioré de prédiction de <i>Dst</i> lorsqu'on considère (a) la totalité de l'ensemble de test et (b) uniquement les périodes d'orages.	123
3.15	Exemples de prédictions 12 heures et 24 heures en avance obtenues avec le modèle amélioré lors de l'orage géomagnétique extrême de juillet 2000.	125
3.16	Schéma de l'architecture de notre modèle de prédiction probabiliste de <i>Dst</i> .	127
3.17	Exemples de prédictions obtenues une, trois et six heures en avance à l'aide du modèle probabiliste lors des orages ayant eu lieu le 26 août 1998 et mi-juillet 2000.	129
3.18	Courbes des distributions de probabilité empiriques des valeurs d'écart-type prédites par le modèle probabiliste sur l'ensemble d'évaluation complet pour des horizons de prédiction allant d'une à six heures.	130
3.19	Illustration du calcul de <i>Continuous Ranked Probability Score</i> (CRPS) pour un exemple de prédiction probabiliste.	131
3.20	Diagramme de fiabilité du modèle probabiliste pour les horizons de prédictions allant d'une à six heures.	133

3.21	Courbes PRC du modèle probabiliste pour les horizons de prédictions compris entre une et six heures. Les valeurs de l'AUPRC, des seuils optimaux et des F_{score} associés sont indiqués pour chaque horizon.	135
4.1	Évolution des flux d'électrons dans les ceintures de radiation pour trois gammes d'énergies ($0,309 < E < 0,379 \text{ MeV}$, $1,539 < E < 1,995 \text{ MeV}$, $3,299 < E < 3,969 \text{ MeV}$) mesurés par l'instrument MagEIS à bord du satellite RBSP A entre le 15 avril et le 15 mai 2017, pendant un orage induit par une CIR. L'évolution des indices géomagnétiques Ca , aa et Kp avec deux paramètres du vent solaire (pression dynamique et vitesse) est également donnée. Reproduit de BERNOUX et MAGET 2020.	142
4.2	Évolution des flux d'électrons dans les ceintures de radiation pour trois gammes d'énergies ($0,309 < E < 0,379 \text{ MeV}$, $1,539 < E < 1,995 \text{ MeV}$, $3,299 < E < 3,969 \text{ MeV}$) mesurés par l'instrument MagEIS à bord du satellite RBSP A entre le 28 septembre et le 28 octobre 2013, pendant un orage induit par une triple ICME. L'évolution des indices géomagnétiques Ca , aa et Kp avec deux paramètres du vent solaire (pression dynamique et vitesse) est également donnée. Reproduit de BERNOUX et MAGET 2020.	143
4.3	Coefficients de corrélation entre les flux d'électrons $E > 100 \text{ keV}$ (f) mesurés par l'instrument NOAA POES-15/MEPED et aa d'un côté et Ca de l'autre, en fonction du paramètre L . Différentes combinaisons ont été testées. La période utilisée pour calculer ces corrélations s'étend de juillet 1998 à décembre 2015. Adapté de BERNOUX et MAGET 2020.	144
4.4	Coefficients de corrélation entre les flux d'électrons (passés en échelle logarithmique) mesurés par l'instrument NOAA POES-15/MEPED pour différentes énergies et l'indice Ca , en fonction du paramètre L . La période utilisée pour calculer ces corrélations s'étend de juillet 1998 à décembre 2015.	145
4.5	Distribution empirique des valeurs de Ca . Quelques repères statistiques sont indiqués.	146
4.6	Évolution de la moyenne annuelle de Ca entre 1869 et 2018. La moyenne globale de Ca sur l'ensemble de la période est donnée pour repère. Le nombre mensuel de taches solaires lissé sur 13 mois est également tracé comme témoin du cycle solaire.	147
4.7	Évolution de la dynamique de Ca pour a) 2003, une année très perturbée, et b) 2009, une année très calme.	148
4.8	Auto-corrélation temporelle de Ca pour des décalages temporels allant jusqu'à 30 jours.	149
4.9	Courbe des valeurs prises par l'indice Ca entre 1995 et 2018 inclus (ligne noire fine), les ensembles d'entraînement (zone verte), de validation (zone jaune) et de test (zone rouge) étant mis en évidence. Le nombre de taches solaires lissé sur 13 mois est également représenté comme indicateur du cycle solaire (ligne rouge épaisse).	151

4.10	Représentation simple de notre modèle basé sur une architecture LSTM pour prévoir les valeurs de l'indice Ca jusqu'à n heures à l'avance. Le fonctionnement interne de la cellule LSTM n'a volontairement pas été détaillé.	154
4.11	Nombre de jours par mois pour lesquels l'instrument LANL 1991-80/MPA a mesuré une fluence électronique quotidienne pour la gamme d'énergie $10 \leq E \leq 50$ keV supérieure à $8 \times 10^{12} \text{ cm}^{-2} \text{ s}^{-1}$ ainsi que le nombre de jours par mois pour lesquels le maximum journalier de Ca quotidien était supérieur à 38 nT. Reproduit de BERNOUX et al. 2021.	156
4.12	Évaluation du modèle LSTM-NN avec trois mesures (RMSE, R et TDM) pour des valeurs d'horizon temporel allant de 1 heure à 24 heures. Trois ensembles d'évaluation (ensemble de test complet, ensemble induit par les SIRs et ensemble induit par les ICMEs) ont été utilisés. Reproduit de BERNOUX et al. 2021.	160
4.13	Exemple de prévisions obtenues avec le modèle LSTM-NN et le modèle linéaire pendant deux orages géomagnétiques, le premier (à gauche) étant un orage piloté par une ICME et le second étant piloté par une SIR (à droite). Quatre horizons de prévision différents ont été utilisés (3, 6, 12 et 24 heures). La valeur seuil de Ca utilisée pour la classification binaire est indiquée en ligne pointillée bleue. Pour chaque prédiction, la valeur de TDM correspondante est donnée. Reproduit de BERNOUX et al. 2021.	162
4.14	NRMSE et corrélation linéaire des valeurs prédites avec 3 heures d'avance par rapport aux valeurs observées, en fonction des quartiles de Ca sur l'ensemble de test. Reproduit de BERNOUX et al. 2021.	163
4.15	Pourcentage de périodes de perturbations correctement prédites jusqu'à 24 h en avance, alors qu'on se situait à l'instant T dans une période de calme, en fonction du temps restant avant le dépassement (entre 1 heure et 24 heures). Reproduit de BERNOUX et al. 2021.	164
5.1	Schéma de l'architecture globale de SERENADE.	179
5.2	Image du Soleil à la longueur d'onde 193 Å capturée le 08 septembre 2017 12h00 par SDO/AIA. (a) Image fournie dans GALVEZ et al. 2019 sans pré-traitement supplémentaire. (b) La même image après écrêtage. (c) La même image après écrêtage et passage au logarithme base 10.	183
5.3	Architecture du module Inception. Extrait de SZEGEDY et al. 2014.	186
5.4	Schéma synthétique du Feature Extractor.	187
5.5	Schéma synthétique du Temporal Encoder.	188
5.6	Schéma synthétique du Decoder. μ et σ représentent respectivement la moyenne et l'écart-type d'une distribution gaussienne.	189
5.7	Schéma synthétique de SERENADE. μ et σ représentent respectivement la moyenne et l'écart-type d'une distribution gaussienne.	191

5.8	Distribution normalisée des valeurs prises par l'indice Kp au cours des cycles solaires 23 et 24, a) pour toutes les valeurs possibles et b) uniquement pour les périodes d'orage. Le rapport entre les distributions obtenues pour le cycle solaire 23 par rapport au cycle solaire 23 est donnée en c).	192
5.9	Schéma de la méthode d'évaluation croisée imbriquée.	194
5.10	Auto-corrélation linéaire de $Kp_{\max,24h}$ pour le cycle solaire 24. . .	201
6.1	Exemples de prédictions obtenues avec SERENADE 2 jours et 5 jours en avance pour deux périodes : de septembre à fin décembre 2012 et de septembre à fin décembre 2015.	207
6.2	Évaluation de SERENADE pour six horizons de prédiction différents à l'aide de métriques dédiées au problème de régression.	209
6.3	Comparaison de SERENADE avec trois modèles de référence à l'aide de <i>Skill Scores</i> relatifs à trois métriques dédiées au problème de régression.	210
6.4	Courbes ROC de SERENADE et des modèles de référence pour chaque horizon de prédiction. Les valeurs d'AUROC ainsi que les probabilités p_{seuil} optimales pour chacun des modèles sont données.	212
6.5	Métriques d'évaluation du système d'alerte dérivé de SERENADE et des trois modèles de référence pour tous les horizons de prédiction considérés. À l'exception du <i>Brier Score</i> , plus la valeur de la métrique est élevée, plus le modèle est performant.	214
6.6	Répartition de nos données en fonction des phases du cycle solaire.	216
6.7	Évaluation de SERENADE et des modèles de référence en fonction des phases du cycle solaire pour un horizon de prédiction de 2 jours à l'aide de métriques dédiées à l'évaluation du problème de régression.	217
6.8	Évaluation de SERENADE et des modèles de référence en fonction des phases du cycle solaire pour un horizon de prédiction de 2 jours à l'aide de métriques dédiées à l'évaluation du problème de classification binaire.	219
6.9	Exemples de prédictions obtenues avec SERENADE 2 jours en avance pour deux périodes : de septembre à fin décembre 2012 et de septembre à fin décembre 2015. Les périodes d'ICMEs (resp. de SIRs) sont mises en valeur par une zone rouge (resp. verte).	220
6.10	Évaluation de SERENADE et des modèles de référence en fonction de l'horizon de prédiction lors de la phase descendante du cycle solaire, à l'aide de métriques dédiées à l'évaluation du problème de classification binaire.	221
6.11	Exemples de prédictions obtenues avec SERENADE- <i>Ca</i> 2 jours et 5 jours en avance pour deux périodes : de septembre à fin décembre 2012 et de septembre à fin décembre 2015.	223
6.12	Évaluation de SERENADE- <i>Ca</i> pour quatre horizons de prédiction différents à l'aide de métriques dédiées au problème de régression.	224
6.13	Comparaison de SERENADE- <i>Ca</i> avec trois modèles de référence à l'aide de <i>Skill Scores</i> relatifs à trois métriques dédiées au problème de régression.	225

6.14	Courbes ROC de SERENADE- <i>Ca</i> et des modèles de référence pour les horizons de prédiction 2, 5, 7 et 9 jours. Les valeurs d'AUROC ainsi que les probabilités p_{seuil} optimales pour chacun des modèles sont données.	226
6.15	Évaluation de SERENADE- <i>Ca</i> et des modèles de référence en fonction des phases du cycle solaire pour un horizon de prédiction de 2 jours à l'aide de métriques dédiées à l'évaluation du problème de régression.	227
6.16	Évaluation de SERENADE- <i>Ca</i> et des modèles de référence en fonction des phases du cycle solaire pour un horizon de prédiction de 2 jours à l'aide de métriques dédiées à l'évaluation du problème de classification binaire.	229
B.1	Métriques d'évaluation du système d'alerte dérivé de SERENADE- <i>Ca</i> et des trois modèles de référence pour tous les horizons de prédiction considérés. À l'exception du <i>Brier Score</i> , plus la valeur de la métrique est élevée, plus le modèle est performant.	265
B.2	Évaluation de SERENADE- <i>Ca</i> et des modèles de référence en fonction de l'horizon de prédiction lors de la phase descendante du cycle solaire, à l'aide de métriques dédiées à l'évaluation du problème de classification binaire.	266

Liste des tableaux

1.1	Description des principaux paramètres du vent solaire, dont on donne la valeur moyenne et l'écart-type observés à 1 au lors de vents solaires lents, rapides et pour des plasmas d'origine éruptive (valeurs issues de BOROVSKY 2020b et EBERT et al. 2009).	20
1.2	Quelques paramètres moyens du plasma rencontré dans certaines régions de la magnétosphère externe (KOSKINEN et KILPUA 2022).	26
1.3	Période de giration, de rebond et de dérive pour des électrons et des protons de 1 MeV à environ $2 R_T$ du centre de la Terre. Reproduit de HERRERA 2017.	28
2.1	Quelques fonctions d'activation usuelles.	51
2.2	Quelques fonctions de perte usuelles.	53
2.3	Bandes-passantes de l'instrument EIT, provenant de DELABOUDINIÈRE et al. 1995.	67
2.4	Caractéristiques des coronographes de l'instrument LASCO, reproduit de BRUECKNER et al. 1995.	69
2.5	Différentes longueurs d'onde d'acquisition de l'instrument AIA et caractéristiques associées (LEMEN et al. 2012).	72
3.1	Éléments statistiques de la distribution empirique des valeurs de <i>Dst</i> observées entre 1963 et 2018 (inclus).	96
3.2	Comparaison des performances de prédiction de <i>Dst</i> pour plusieurs horizons de prédiction obtenues avec le modèle originel de GRUET et al. 2018 et avec notre reproduction du modèle.	101
3.3	Comparaison des performances de prédiction de <i>Dst</i> obtenues avec le modèle de base entraîné et évalué d'abord avec l'ancien découpage des données puis avec le nouveau, pour des horizons de prédiction allant de 1 – 6 h.	107
3.4	Comparaison des performances de prédiction de <i>Dst</i> pour plusieurs horizons de prédiction obtenues avec le modèle de base et avec notre nouveau modèle, tous deux entraînés et évalués avec le nouveau découpage des données.	110
3.5	Comparaison des performances de prédiction de <i>Dst</i> obtenues avec le modèle amélioré pour trois méthodes de gestion des données manquantes, pour des horizons de prédiction de 1 – 6 h.	113
3.6	Évaluation du modèle de prédiction de <i>Dst</i> amélioré dans le contexte du problème de classification (et comparaison avec le modèle de base).	118
3.7	Évaluation du modèle amélioré entraîné pour prédire <i>Dst</i> 12 et 24 h en avance. Les métriques pour le modèle à 6 h sont données pour comparaison.	124

3.8	Évaluation du modèle probabiliste à l'aide de la MAE et le CRPS moyen.	132
4.1	Valeurs estimées par méthode méthode des pics au dessus du seuil ou <i>Peaks Over Threshold</i> (POT) des événements géomagnétiques décennaux, cinquantennaux et centennaux selon l'indice <i>Ca</i>	149
4.2	Nombre d'échantillons de données dans chaque ensemble de données, y compris le nombre d'échantillons appartenant à une période de perturbation. Adapté de BERNOUX et al. 2021.	158
4.3	Évaluation du modèle LSTM-NN et du modèle de référence linéaire dans le contexte du problème de régression. Les modèles ont été évalués avec le jeu de test complet ainsi qu'avec le jeu de test induit par les SIRs et le jeu de test induit par les ICMEs. Adapté de BERNOUX et al. 2021.	159
4.4	Évaluation du modèle LSTM-NN et du modèle de référence linéaire dans le contexte du problème de classification. Adapté de BERNOUX et al. 2021.	165
6.1	Comparaison des performances obtenues par le modèle de prédiction à 1 h de <i>Dst</i> lorsqu'on entraîne le modèle uniquement avec les données observées entre 2010 et 2018, par rapport aux performances obtenues lorsqu'on entraîne avec les données observées entre 2003 et 2018. On distingue les périodes calmes des périodes d'orages (modérés et intenses). Un pourcentage négatif indique une dégradation de la valeur de la métrique (augmentation pour NRMSE et baisse pour <i>R</i>). 230	
A.1	Valeurs optimales de M^* et K^* trouvées pour reconstruire les données manquantes dans les séries temporelles de cinq paramètres du vent solaire.	264

Liste des acronymes

ADAPT

Air Force Data Assimilative Photospheric Flux Transport. 76

AGU

American Geophysical Union. 138

AIA

Atmospheric Imaging Assembly. xviii, xxi, xxiv, 71–74, 76, 86, 172, 173, 180–183

ANN

réseau de neurones artificiel ou *Artificial Neural Network.* 43, 62, 87, 88

AUPRC

aire sous la courbe PRC ou *Area Under the PRC.* xx, 134, 135

AUROC

aire sous la courbe ROC. xxii, xxiii, 197, 198, 203, 212, 213, 215, 222, 226

BS

Brier Score. xxii, xxiii, 203, 213–215, 218, 265

CIR

région d'interaction en co-rotation ou *Corotating Interaction Region.* xx, 15, 95, 97, 139, 141, 142, 157

CME

éjection de masse coronale ou *Coronal Mass Ejection.* xvii, 17, 18, 70, 71, 86, 87, 95, 173–176, 181, 218, 221, 231, 235

CNN

réseau de neurones convolutifs ou *Convolutional Neural Network.* xvii, 41, 59–61, 86, 172, 174, 184, 234, 235

CPU

Central Processing Unit. 64

CRAND

Cosmic Ray Albedo Neutron Decay. 32

CRPS

Continuous Ranked Probability Score. xix, xxv, 130–132, 134, 197, 198, 202, 208, 211, 222, 225, 233

DPG

distribution de Pareto généralisée. 148

DPHY

Département physique, instrumentation, environnement, espace. ix, 2

DTW

Dynamic Time Warping. xix, 121, 122

EIT

Extreme ultraviolet Imaging Telescope. xviii, xxiv, 67–71

EnKF

filtre de Kalman d'ensemble. 90

EnLS

Ensemble Least Squares. 76

ESA

Agence Spatiale Européenne ou *European Space Agency*. xvi–xviii, 11, 13, 18, 23, 34, 66, 68, 70, 237

EUVI

Extreme UltraViolet Instrument. 70, 71

EVA

analyse des valeurs extrêmes ou *Extreme Value Analysis*. 147

EVE

Extreme Ultraviolet Variability Experiment. 73, 180

FAR

taux de fausses alertes ou *False Alarm Rate*. 117, 118, 133, 157, 165, 203, 211

FDP

Full-Disk Patrol. 75

FN

faux négatifs ou *False Negatives*. 117, 118

FP

faux positifs ou *False Positives*. 117, 118

GEO

Orbite géostationnaire. 145, 155, 168

GIC

courant induit géomagnétiquement ou *Geomagnetically Induced Current*. 33

GNLL

Gaussian Negative Log-Likelihood. 54, 127, 197, 222

GOES

Geostationary Operational Environmental Satellite. 76, 88, 167, 175, 235

GONG

Global Oscillations Network Group. 75, 76, 172

GP

processus gaussien ou *Gaussian Process*. 98, 126, 135, 233

GPS

Global Positioning System. 33, 88, 98, 101, 103, 235

GPU

Graphics Processing Unit. 64, 100, 108

HEK

Heliophysics Event Knowledgebase. 76, 77

HMI

Helioseismic and Magnetic Imager. xviii, xxxi, 73, 74, 76, 172, 173, 180

HSS

Heidke Skill Score. 117, 118, 157, 164, 165, 203, 213

IA

intelligence artificielle. xxxii, 41, 236, 237

IAS

Institut d'Astrophysique Spatiale. ix, 77

ICME

éjection de masse coronale interplanétaire ou *Interplanetary Coronal Mass Ejection.* xvii, xx–xxii, xxv, 18–21, 24, 29, 31, 32, 82, 87, 95, 141, 143, 157–161, 167, 168, 219, 220, 234

IGRF

International Geomagnetic Reference Field. xvii, 22

ILSVRC

ImageNet *Large Scale Visual Recognition Challenge.* 41, 61

IMF

champ magnétique interplanétaire ou *Interplanetary Magnetic Field.* xxxiii, 12, 14, 20, 21, 30–32, 78, 87, 97, 150

IPODE

Ionizing Particle ONERA Database. xviii, 82, 83

IRAP

Institut de Recherche en Astrophysique et Planétologie. 237

ISGI

International Service of Geomagnetic Indices. xviii, 79–82, 140, 153

***k*-NN**

k-Nearest Neighbors. 41, 174

LANL

Los Alamos National Laboratory. xxi, 89, 155, 156

LASCO

Large Angle and Spectrometric Coronagraph. xviii, xxiv, 67–69, 173, 175

LEO

orbite basse terrestre ou *Low Earth Orbit.* 34

LMSAL

Lockheed Martin *Solar and Astrophysics Laboratory*. xviii, 66, 76

LSTM

Long Short-Term Memory. xvii, xxi, 57, 59, 86, 88, 89, 91, 97, 108, 128, 153, 154, 174, 175, 187, 222, 233–235

MAE

erreur absolue moyenne ou *Mean Absolute Error*. xxv, 53, 131, 132, 134, 208

MDI

Michelson Doppler Imager. xviii, 67, 69–71, 73, 76, 173, 174

MEDOC

Multi Experiment Data and Operation Center. 76, 77

MEO

Medium Earth Orbit. 90

MEPED

Medium Energy Proton and Electron Detector. 84

mGNLL

modified Gaussian Negative Log-Likelihood. 127, 197

MLP

perceptron multicouche ou *Multi-Layer Perceptron*. xvii, 46, 47, 50, 51, 54–56, 61, 174, 189, 190

MLT

temps magnétique local ou *Magnetic Local Time*. 23

MPA

Magnetospheric Plasma Analyzer. xxi, 155, 156

MSE

erreur quadratique moyenne ou *Mean Squared Error*. 53, 98, 127, 153, 197, 198, 203, 222

NASA

National Aeronautics and Space Administration. xvi–xviii, 9–11, 13, 14, 16–18, 27, 67, 68, 70, 71, 73, 74, 78

NISP

NSO Integrated Synoptic Program. 75, 76

NOAA

National Oceanographic and Atmospheric Administration. xx, 76, 84, 88, 144, 145, 167, 175, 200, 202

NRMSE

racine de l'erreur quadratique moyenne normalisée ou *Normalized RMSE*. xix, xxi, xxv, 105, 107, 115, 116, 156, 161, 163, 228, 230

NSO

National Solar Observatory – l'observatoire solaire américain. xxix, 75, 76

ONERA

Office national d'études et de recherches aérospatiales. ix, xxviii, 2, 82, 95, 99, 139

PCA

analyse en composantes principales ou *Principal Component Analysis*. 42

POD

probabilité de détection ou *Probability of Detection*. 117, 212

POES

Polar Operational Environmental Satellite. xx, 84, 144, 145

POFD

probabilité de fausse détection ou *Probability of False Detection*. 117, 212

POT

méthode des pics au dessus du seuil ou *Peaks Over Threshold*. xxv, 148, 149

PRC

courbe précision-rappel ou *Precision-Recall Curve*. xx, xxvi, 133–135, 211

PROBA-2

Project for On-Board Autonomy. 71

PSP

Parker Solar Probe. 74, 75

RBSP

Radiation Belt Storm Probe. xx, 141–143

RF

Random Forest. 41, 174

RMSE

racine de l'erreur quadratique moyenne ou *Root Mean Squared Error*. xxi, 98, 100, 101, 104, 105, 107, 110, 112, 113, 115, 123, 124, 156, 158–161, 163, 168, 202, 208, 211, 217, 225, 233, 264

RNN

réseau de neurones récurrent ou *Recurrent Neural Network*. xvii, 55–57, 59, 87

ROC

fonction d'efficacité du récepteur ou *Receiver Operating Characteristic*. xxii, xxiii, xxvi, 133, 198, 203, 211, 212, 226

SAA

anomalie de l'Atlantique-Sud ou *South Atlantic Anomaly*. 22, 34

SDO

Solar Dynamics Observatory. xvi, xviii, xxi, 9, 10, 13, 14, 16, 17, 70, 71, 73, 74, 76, 86, 171–173, 175, 180–183, 190, 236

SECCHI

Sun-Earth Connection Coronal and Heliospheric Investigation. 70

SEM

Space Environment Monitor. 84

SHARP

Space Weather HMI Archive Patch. 173, 174

SIR

région à interactions de courants ou *Stream Interaction Region.* xvi, xxi, xxii, xxv, 14, 15, 20, 21, 31, 82, 97, 157–161, 167, 168, 218–220, 234, 235

SMART MF

Solar Monitor Active Region Tracker Magnetic Features. 173

SOHO

Solar and Heliospheric Observatory. xvi–xviii, 11, 13, 18, 66, 68–71, 73, 74, 76, 171, 173–175, 180, 236

SOLIS

Synoptic Optical Long-term Investigations of the Sun. 75, 76

SPE

événement solaire à particules ou *Solar Particle Event.* 16, 32

SSA

analyse spectrale singulière ou *Singular Spectrum Analysis.* 112, 113, 152, 263

STEREO

Solar Terrestrial Relations Observatory. 70, 175

SVM

Machine à Vecteurs de Support. 41, 173–175

SWAP

Sun Watcher using Active Pixel System detector and Image Processing. 71

SWIFT

Solar Wind Flux Tubes. 86, 237, 238

SWPC

Space Weather Prediction Center. 76, 200, 202, 203

TDI

indice de distorsion temporelle ou *Temporal Distortion Index.* 121, 122

TDM

Temporal Distortion Mix. xix, xxi, 121–124, 156, 159–161, 166, 168, 202, 208, 228, 233

TN

vrais négatifs ou *True Negatives.* 117, 118

TP

vrais positifs ou *True Positives.* 117, 118

TRACE

Transition Region And Coronal Explorer. 74

TS

Threat score. 117, 118, 124, 157, 164, 165, 203, 213, 218

VSM

Vector Spectromagnetograph. 75

VSO

Virtual Solar Observatory. 76

XAI

IA explicable ou *explainable AI*. 236

Notations

aa

Indice géomagnétique *aa*. xviii, xx, 80–82, 140–144, 150, 153, 158, 234

au

Unité astronomique. xvi, xvii, xxiv, 11, 15, 18–21, 75, 86, 176

|B|

Intensité de l'IMF. 20, 98, 109, 264

B_x

Composante radiale de l'IMF. 88

B_y

Composante aube-crépuscule de l'IMF. 88

B_z

Composante verticale de l'IMF. 20, 31, 86–88, 97, 98, 109, 150, 153, 177, 264

Ca

Indice géomagnétique *Ca*. x, xi, xx, xxi, xxv, xxxiii, xxxv, 3, 136, 138, 140–158, 161–163, 167, 178, 190, 206, 222, 223, 226, 231, 234

$Ca_{\max,24h}$

Valeur maximale de l'indice *Ca* sur une fenêtre de 24 h qui se termine à l'instant présent. 222, 224, 228, 234

Dst

Indice géomagnétique représentatif du courant annulaire. x, xi, xviii, xix, xxiv, xxv, 3, 80–82, 87, 88, 91, 94–101, 103–105, 107–110, 113, 115, 116, 118, 119, 122–124, 127, 136, 139, 147, 149, 151–153, 167, 168, 175, 228, 230, 233, 234, 263

ELU

Fonction Unité Linéaire Exponentielle. 52, 189

EUV

Rayonnement ultraviolet extrême. xvi, 3, 8, 9, 13, 16, 65, 70, 71, 73, 75, 87, 171, 172, 175, 176, 180, 181, 204, 235, 236

F_{score}

Moyenne harmonique de la précision et du rappel. xx, 117, 118, 134, 135, 157, 164, 165, 203, 218

Helioviewer

Plateforme de visualisation et téléchargement de données d'imagerie solaire. [xvi–xviii](#), [10](#), [14](#), [16–18](#), [66](#), [68–70](#), [73](#), [74](#)

IR

Rayonnement infrarouge. [65](#)

Kp

Indice géomagnétique K planétaire. [x](#), [xi](#), [xviii](#), [xx](#), [xxii](#), [xxxiv](#), [3](#), [79–82](#), [87–90](#), [128](#), [136](#), [139](#), [141–143](#), [147](#), [150](#), [151](#), [153](#), [167](#), [168](#), [170](#), [176–178](#), [180](#), [183](#), [184](#), [190](#), [192](#), [197](#), [198](#), [200](#), [202–204](#), [211](#), [215](#), [230](#), [234](#), [263](#)

$Kp_{\max,24\text{ h}}$

Valeur maximale de l'indice Kp sur une fenêtre de 24 h qui se termine à l'instant présent. [xxii](#), [183](#), [186](#), [190](#), [198–203](#), [206–208](#), [211](#), [215](#), [216](#), [220](#), [222](#), [234](#), [235](#)

L

Paramètre de McIlwain. [xx](#), [28](#), [144](#), [145](#), [155](#)

L^*

Paramètre de Roederer. [28](#), [31](#), [89](#), [141](#)

MHD

Magnétohydrodynamique. [86](#), [87](#), [237](#)

précision

Mesure d'évaluation d'un système de classification binaire qui répond à la question « combien de candidats sélectionnés sont pertinents ? ». [xxx](#), [xxxiii](#), [117](#), [118](#), [124](#), [133](#), [134](#), [157](#), [163–167](#), [203](#)

P_{sw}

Pression dynamique du vent solaire. [88](#)

QQ_{slope}

Pente du diagramme Quantile-Quantile. [197](#), [198](#), [222](#)

R

Coefficient de corrélation linéaire. [xxi](#), [xxv](#), [98](#), [100](#), [101](#), [107](#), [110](#), [113](#), [124](#), [144](#), [156](#), [158–161](#), [163](#), [202](#), [225](#), [230](#)

rappel

Mesure d'évaluation d'un système de classification binaire qui répond à la question « combien de candidats pertinents sont sélectionnés ? ». [xxx](#), [xxxiii](#), [117](#), [118](#), [120](#), [124](#), [133](#), [134](#), [157](#), [164–166](#), [203](#), [211](#), [213](#)

ReLU

Fonction Unité Linéaire Rectifiée. [52](#), [57](#), [189](#)

ρ_{sw}

Densité de protons du vent solaire. [20](#), [86–88](#), [97](#), [98](#), [109](#), [150](#), [153](#), [264](#)

- R_{\odot}
Rayon solaire moyen valant environ $6,957 \times 10^8$ m. 7, 69, 71
- R_{T}
Rayon terrestre, de valeur 6371 km. xxiv, 24–28, 31
- SDO/AIA – ML-Ready**
Jeu de données SDO/AIA pré-traité pour un usage en apprentissage profond (GALVEZ et al. 2019). 180, 182, 190, 192, 195, 196, 208, 214, 265
- Serenade**
Sun-EaRth intEractioN dAta-DrivEn model. x, xi, xxi, xxii, xxxv, 3, 177–181, 183–185, 187, 189–191, 193, 195, 197, 199, 201, 204, 206–222, 224–226, 228, 230, 231, 234–238
- Serenade-Ca**
Adaptation de SERENADE pour la prédiction de l'indice *Ca*. xxii, xxiii, 222–230, 265, 266
- Sn**
Statistique Sn. 197, 198, 222
- T_{sw}
Température de protons du vent solaire. 20, 86, 98, 109, 150, 151, 153, 264
- UV**
Rayonnement ultraviolet. 65, 71
- V_{sw}
Vitesse d'ensemble du vent solaire. 20, 86–88, 97, 98, 109, 150, 153, 180, 264

Introduction

Du 28 août au 2 septembre 1859, de nombreuses perturbations des réseaux télégraphiques furent observées dans le monde entier. Dans le même temps, phénomène rarissime, des aurores intenses pouvaient être observées dans le ciel à des latitudes subtropicales. Le 13 mars 1989 une panne de courant géante affecta le Québec, privant d'électricité 6 millions d'habitants de longues heures durant. Le 16 août de la même année, des défaillances du système informatique de la Bourse de Toronto la forcèrent à s'interrompre pendant trois heures. Le 30 octobre 2003, des pannes affectèrent les réseaux électriques en Suède et en Afrique du Sud, causant de longues interruptions de service. Le 3 février 2022, la compagnie américaine SpaceX lança simultanément 49 nouveaux satellites destinés à compléter sa constellation Starlink. Le 8 février, la compagnie annonça qu'une augmentation imprévue de la traînée atmosphérique avait empêché 40 de ces satellites de rejoindre leur orbite finale, causant leur perte définitive. Le point commun de tous ces événements ? Leur origine : le Soleil, et plus précisément son activité et ses interactions avec l'environnement spatial proche de la Terre.

Le système Soleil–Terre est en effet un système complexe, composé d'interactions multiples entraînant, entre autres, des transferts d'énergie de régions du Soleil jusqu'à l'environnement spatial proche de la Terre, et en particulier sa magnétosphère. Cette énergie est principalement transportée par le vent solaire, un plasma peu dense, composé de particules énergétiques chargées émises depuis la couronne solaire. Le Soleil est également à l'origine d'éruptions solaires qui sont des libérations soudaines et violentes d'une grande quantité d'énergie sous forme de rayonnement électromagnétique, parfois suivies d'une émission rapide de particules relativistes de très haute énergie, depuis des régions qualifiées « d'actives ». Ces régions actives sont également à l'origine d'éjections de masse coronales, c'est-à-dire une émission spontanée d'une grande masse de plasma solaire dans une direction donnée, qui modifie les propriétés du vent solaire. Tous ces phénomènes, lorsqu'ils sont dirigés vers la Terre, vont interagir avec la magnétosphère terrestre, notamment en la déformant fortement (causant des orages géomagnétiques) et en apportant des particules, qui peuvent parfois se retrouver alors piégées au sein de la magnétosphère, contribuant à la formation des ceintures de radiations.

Les conséquences de l'environnement radiatif spatial sont multiples et impactent essentiellement les infrastructures spatiales, pouvant causer des dégâts irréversibles (événements singuliers, effets de dose, effets de charge interne et de surface, etc.). Mais les interactions Soleil–Terre peuvent également perturber les télécommunications terrestres, les systèmes de positionnement par satellite ainsi que les systèmes aéronautiques (mettant ainsi en danger les avions et leurs

passagers). Enfin, comme nous l'avons mentionné dans le premier paragraphe, on peut même observer des conséquences sur Terre, par le biais de courants induits géomagnétiquement, qui peuvent, par exemple, endommager les transformateurs et lignes à haute tension. C'est pourquoi il devient de plus en plus important de parvenir à prévoir la dynamique de l'activité géomagnétique, représentée par des indices géomagnétiques, dans le cadre d'un domaine – relativement jeune et donc dynamique – appelé la météorologie de l'espace.

Parallèlement à cela, on assiste, depuis maintenant plus d'une dizaine d'années, à une ré-émergence massive de l'utilisation de méthodes d'apprentissage automatique (*machine learning* en anglais) dans de nombreux domaines scientifiques. L'utilisation de ces algorithmes, rendue possible notamment par l'augmentation des puissances de calcul et de stockage des calculateurs contemporains, permet souvent de modéliser fidèlement des processus (biologiques, chimiques, physiques, etc.) pour lesquels des méthodes plus « classiques » ont montré leurs limites. C'est le cas en météorologie de l'espace, où des approches utilisant (notamment) des réseaux de neurones artificiels émergent et se révèlent prometteuses, en complément des approches fondées sur une modélisation physique des différentes composantes des interactions Soleil–Terre.

Depuis plusieurs années, l'unité Environnement Radiatif Spatial du [DPHY](#) à l'[ONERA](#) développe des modèles prédictifs des indices géomagnétiques terrestres basés sur les paramètres du vent solaire mesurés au point de Lagrange L1. Les récents développements dans le domaine de l'apprentissage automatique ont permis de grands progrès dans la prévision des indices géomagnétiques en étendant leur fenêtre de prédiction jusqu'à environ six heures. Or, à des fins opérationnelles, de tels horizons de prédiction ne sont souvent pas suffisants. C'est dans ce contexte scientifique et technique que se situent nos travaux de thèse.

La finalité principale de nos travaux est d'améliorer nos capacités de prévision des perturbations géomagnétiques d'origine solaire. Pour l'atteindre, nous avons identifié trois grands axes d'étude. Le premier est d'étudier les possibilités d'amélioration des modèles actuels de prévision d'indices géomagnétiques à court terme (moins de 24 h), qui utilisent des réseaux de neurones artificiels. Dans ce contexte, améliorer les modèles signifie principalement améliorer les performances proposées, mais aussi les rendre plus utiles pour l'étude de la dynamique des ceintures de radiations terrestres. Deuxièmement, il s'agit de mieux caractériser ces modèles, afin d'avoir une meilleure idée des performances qu'on peut en tirer, notamment dans un contexte opérationnel. Enfin, le troisième axe consiste à explorer des pistes permettant d'étendre l'horizon de prévision de ces modèles à une fenêtre supérieure à 24 h. Pour traiter toutes ces problématiques nous avons construit le présent manuscrit selon la structure suivante.

Tout d'abord, dans le [Chapitre 1](#) nous commençons par justifier l'intérêt scientifique et opérationnel de nos travaux en présentant succinctement et qualitativement le système Soleil–Terre et les interactions qui le composent. Nous nous concentrerons notamment sur le couplage vent solaire–magnétosphère, à l'origine des ceintures de radiations, et leurs conséquences sur les activités humaines.

Dans le Chapitre 2 nous présenterons le concept de modélisation dirigée par les données, en nous focalisant principalement sur les méthodes d'apprentissage profond (*Deep Learning* en anglais). Nous passerons ensuite en revue les différentes sources de données permettant une observation et une description de différentes composantes du système Soleil–Terre. Tout cela nous amènera à faire une étude bibliographique sur l'état de l'art actuel de la prévision de l'activité géomagnétique terrestre, ce qui nous permettra de mieux situer nos travaux.

Dans le Chapitre 3 nous proposerons un nouveau modèle de prévision de l'indice géomagnétique Dst à partir de paramètres du vent solaire proches de la Terre et utilisant des réseaux de neurones récurrents. Pour construire ce modèle nous commencerons par reproduire le modèle proposé préalablement par GRUET et al. 2018. Nous l'améliorerons ensuite à différents niveaux (gestion des données, architecture du modèle), en détaillant les étapes. Nous attacherons une importance particulière à évaluer notre modèle à l'aide d'un banc d'évaluation complet, que nous détaillerons et justifierons.

Afin de rendre notre modèle encore plus utile opérationnellement du point de vue de l'étude des ceintures de radiations, nous débuterons le Chapitre 4 en présentant un nouvel indice géomagnétique, Ca , développé pour être mieux corrélé avec la dynamique des ceintures d'électrons. Après l'avoir analysé statistiquement, nous présentons une variante du modèle précédent que nous adapterons pour la prédiction de Ca . Nous évaluerons également ce modèle le plus exhaustivement possible.

Constatant l'incapacité des modèles étudiés dans les chapitres précédents à proposer des performances prédictives correctes pour des horizons de prédiction plus éloignés, nous présenterons dans le Chapitre 5 un nouveau modèle de prévision de l'indice géomagnétique planétaire Kp (qui est notamment utilisé pour piloter certains modèles de ceintures de radiations). Notre modèle, appelé **SERENADE**, uniquement constitué de différents types de réseaux de neurones, présente la particularité de n'être dirigé que par des séquences d'images du Soleil dans l'**EUUV**. À notre connaissance, une telle approche n'a jamais fait l'objet d'une publication scientifique à ce jour. Notre but sera précisément de déterminer si l'emploi d'une telle approche est pertinent (ou non) et mérite d'être considéré. Après avoir présenté une étude bibliographique sur l'application de méthodes d'apprentissage automatique à l'imagerie solaire, nous justifierons notre approche et détaillerons les différents sous-modules constituant notre modèle.

Enfin, dans le Chapitre 6 nous évaluerons **SERENADE** et le comparerons à des modèles empiriques simples, qui sont pourtant actuellement les meilleurs modèles de prédiction de Kp pour des horizons de prédiction supérieurs à un jour. Nous présenterons également une variante du modèle destinée à la prédiction de Ca , que nous évaluerons également. Ce manuscrit se conclura par un court récapitulatif des principaux résultats obtenus au cours de nos travaux, suivi par la mention de plusieurs pistes à explorer au cours de futurs développements.

Avant d'entrer dans le vif du sujet, précisons un petit point de vocabulaire.

Doit-on dire « prévision » ou bien « prédiction » d'indices géomagnétiques¹ ? Selon NADEAU 1999, la « prédiction » est issue d'une théorie scientifique hautement confirmée, tandis que la « prévision » découlerait d'hypothèses moins assurées. Cela signifierait qu'en fonction du type de modèle utilisé (physique, empirique, etc.) on serait amené à changer de terme. Néanmoins, cela semble aller à l'encontre de l'intuition, étant donné qu'on parle plutôt de « prédictions » astrologiques (sans fondement scientifique) et de « prévisions » météorologiques (s'appuyant sur des modèles), ce qui suggérerait que la « prédiction » serait plutôt liée à l'intuition et la « prévision » à la raison. Bien que certains auteurs utilisent exclusivement l'un ou l'autre terme, la question ne semble pas tranchée dans la littérature portant sur la météorologie de l'espace (en langues française et anglaise). C'est pourquoi dans notre manuscrit nous utiliserons les deux termes (et leurs dérivés) de manière interchangeable.

1. Notons que la même question se pose en anglais, où la distinction entre « *prediction* » et « *forecast* » n'est pas triviale.

Chapitre 1

Introduction aux interactions Soleil–Terre et à la météorologie de l’espace

Il faut reconnaître que, parmi les intellectuels, on rencontre parfois, à titre exceptionnel, des gens intelligents. On ne peut le nier.

(Mikhaïl Boulgakov, Le Maître et Marguerite)

Sommaire

Préambule	6
1.1 Introduction à l’activité solaire	7
1.1.1 La structure du Soleil	7
1.1.2 Le cycle solaire	9
1.1.3 Le(s) vent(s) solaire(s)	10
1.1.3.1 Les sources du vent solaire lent : les <i>helmet streamers</i>	12
1.1.3.2 Les sources du vent solaire rapide : les trous coronaux	13
1.1.3.3 Les régions à interactions de courants	14
1.1.4 L’activité éruptive du Soleil	15
1.1.4.1 Les embrillancements	16
1.1.4.2 Les éjections de masse coronales	17
1.1.5 Le vent solaire à 1 au	18
1.2 Les ceintures de radiations, au cœur de la magnétosphère terrestre .	21
1.2.1 Morphologie de la magnétosphère terrestre	21
1.2.1.1 Description générale du champ magnétique terrestre	21
1.2.1.2 La magnétosphère externe	24
1.2.1.3 La magnétosphère interne	25
1.2.1.4 Les ceintures de radiations	26
1.2.2 L’activité géomagnétique	29
1.2.2.1 La convection magnétosphérique	30
1.2.2.2 Les orages et sous-orages géomagnétiques	30
1.2.2.3 Le pilotage des ceintures de radiations par l’activité géomagnétique	31
1.3 Les interactions Soleil–Terre vues du point de vue des activités humaines	32
1.3.1 Les conséquences de l’activité géomagnétique et de l’environ- nement radiatif spatial sur les activités humaines	33
1.3.2 La météorologie de l’espace	35
1.4 Bilan du chapitre	35

Préambule

Ce premier chapitre a pour vocation de décrire qualitativement certaines bases de la physique des interactions Soleil–Terre, nécessaires ou utiles à la compréhension de nos travaux et de ses enjeux.

Notons bien que nous n’entrerons volontairement pas dans de trop grands détails, pour deux raisons. La première est que, de part notre sujet de thèse, nous nous intéressons aux interactions Soleil–Terre avec une vision d’ensemble, à l’échelle du système. Ainsi, il n’est pas utile d’entrer dans le détail de chacun des processus physiques mentionnés – cela nécessiterait un manuscrit à part entière. D’autre part, nos travaux portant essentiellement sur de la modélisation dirigée par les données, il n’est pas forcément nécessaire, pour le lecteur non-initié, de connaître analytiquement et exhaustivement les équations physiques régissant ces phénomènes, mais plutôt de comprendre qualitativement certains phénomènes

et les enjeux que pose leur observation. Le lecteur qui souhaiterait néanmoins approfondir ces questions, tout en conservant une vision d'ensemble des interactions Soleil–Terre et de leurs effets sur les activités humaines, pourrait se satisfaire de livres de référence tels ceux écrits par KIVELSON et RUSSEL 1995, LILENSTEN et BLELLY 1999 ou encore KOSKINEN et KILPUA 2022.

Ainsi, nous commençons par présenter les différentes composantes des interactions Soleil–Terre, de ses origines solaires à leur impact sur la magnétosphère terrestre et les ceintures de radiations, en passant par le vent solaire, vecteur de transport d'énergie du Soleil dans l'héliosphère. Nous décrivons également les conséquences de ces interactions sur les activités humaines et comment une discipline comme la météorologie de l'espace cherche à les mitiger. Nous aborderons la question des outils d'observation du système Soleil–Terre et de la modélisation des interactions dans le Chapitre 2.

1.1 Introduction à l'activité solaire

Le Soleil, l'étoile au cœur de notre système planétaire (il représente environ 99,8 % de la masse totale du système solaire), est la principale source des variations et perturbations de l'environnement radiatif spatial terrestre. C'est pourquoi il est primordial de décrire certaines des caractéristiques liées à sa structure et à son activité afin d'en comprendre les conséquences sur les activités humaines.

1.1.1 La structure du Soleil

Le Soleil est une étoile âgée d'environ 4,6 milliards d'années et composée principalement d'hydrogène ($\approx 74\%$ de sa masse) et d'hélium ($\approx 25\%$ de sa masse). Le Soleil peut être décrit selon une structure en couches successives, comme l'illustre la figure 1.1. L'intérieur du Soleil est composé d'un cœur (d'un rayon d'environ $0,25 R_{\odot}$), d'une zone de transport radiatif ($0,45 R_{\odot}$) et d'une zone de transport convectif ($0,3 R_{\odot}$). Sa surface est appelée la photosphère. Tout aussi importante, son atmosphère est composée de deux principales couches : la chromosphère et la couronne (PRIEST 2014).

Le cœur (ou noyau) est la zone dans laquelle on observe des réactions thermonucléaires exothermiques qui transforment de l'hydrogène en hélium. Ces réactions de fusion nucléaire sont la principale source d'énergie du Soleil, le cœur en est donc logiquement la région la plus chaude (avec une température de l'ordre de $1,5 \times 10^7$ K). Cette région produit également un grand nombre de neutrinos.

La zone de transport radiatif est une zone où, comme son nom l'indique, la quasi-totalité du transfert de chaleur s'effectue par rayonnement thermique. La température y est de l'ordre de 10^6 K.

Entre la zone de transport radiatif et celle de transport convectif on retrouve la tachocline, qui est la zone responsable de la dynamo solaire, c'est-à-dire l'ensemble

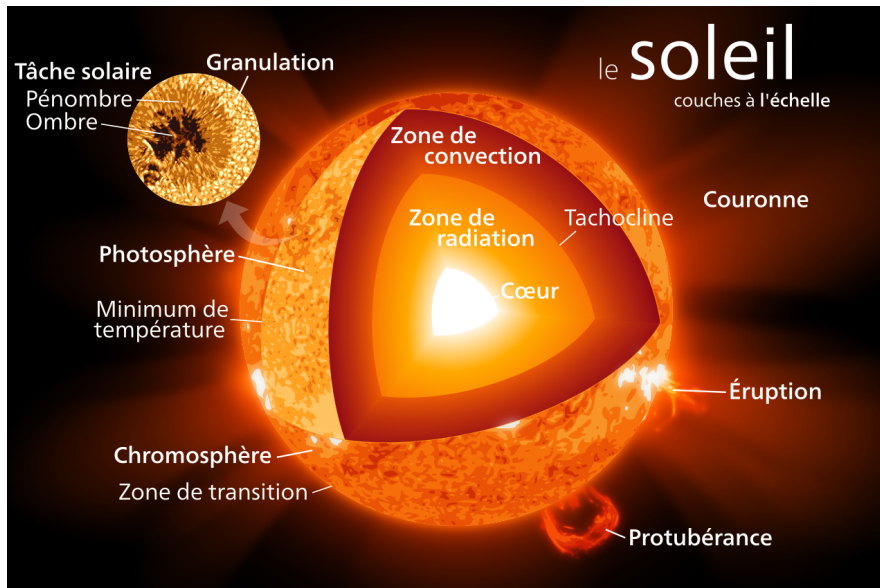


FIGURE 1.1 – Schéma de la structure en couches du Soleil. Source : Wikimedia Commons, sous licence CC BY-SA 3.0.

des processus physiques à l'origine du champ magnétique du Soleil. Celui-ci est notamment dû aux différences de rotation entre les zones radiatives et convectives (rotation rigide dans la première, différentielle dans la deuxième). La zone de transport convectif est une zone où les échanges thermiques se font essentiellement par convection verticale, la température passant de l'ordre de 10^6 K à environ 5800 K en surface.

La photosphère est la zone qui rayonne la très grande majorité de la lumière visible qui parvient à la Terre. En effet, les zones internes du Soleil ne laissent échapper que peu de photons à cause de leur densité – la surface du Soleil est d'ailleurs définie comme le rayon à partir duquel le plasma solaire laisse s'échapper les photons librement dans le système solaire. L'observation de la photosphère permet d'identifier une grande partie des caractéristiques principales de l'étoile, comme sa température apparente, sa composition, sa masse ou encore son champ magnétique. Dans le cas du Soleil, la température effective (température du corps noir associé) est de 5777 K. L'épaisseur de la photosphère du Soleil est d'environ 400 km (à mettre en perspective avec le rayon total du Soleil, d'environ 7×10^5 km).

La chromosphère est la couche basse de l'atmosphère du Soleil, composée d'un plasma peu dense. Son épaisseur est de l'ordre de 10^3 km. Dans la chromosphère la température augmente lorsqu'on s'éloigne du centre du Soleil car la pression y diminue.

La couronne est la partie haute de l'atmosphère du Soleil, en interface avec le milieu interplanétaire, s'étendant sur plusieurs millions de kilomètres. Il s'agit d'un plasma dont la structure est fortement induite par les champs magnétiques se dégageant de la photosphère. La température y est de nouveau très élevée, de l'ordre de 10^6 K. Son observation dans le visible, est donc limitée. Pour l'observer on a recours à l'utilisation de coronographes et/ou d'imageurs spatiaux dans l'EUV

(voir le Chapitre 2). La Figure 1.2 montre des images de la photosphère (1.2a), de la chromosphère (1.2b) et de la couronne (1.2c) du Soleil prises au même instant le 15 février 2022.

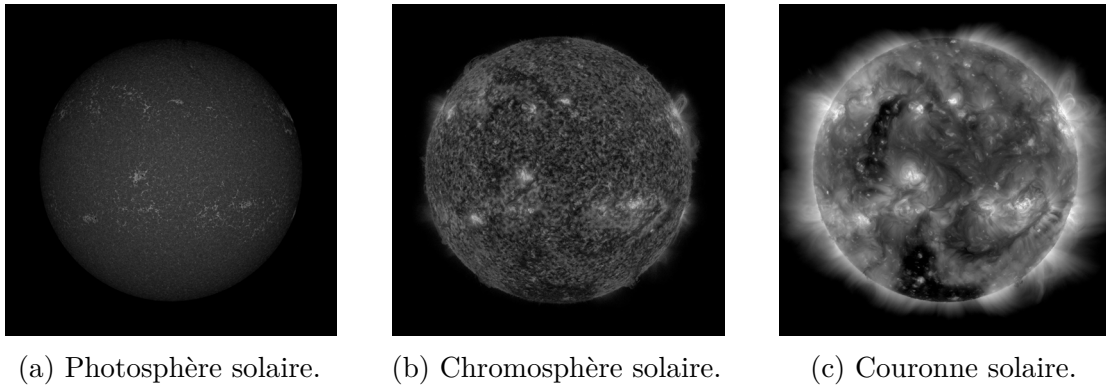


FIGURE 1.2 – Images de différentes couches du Soleil prises au même instant le 15 février 2022. Crédits : SDO (NASA).

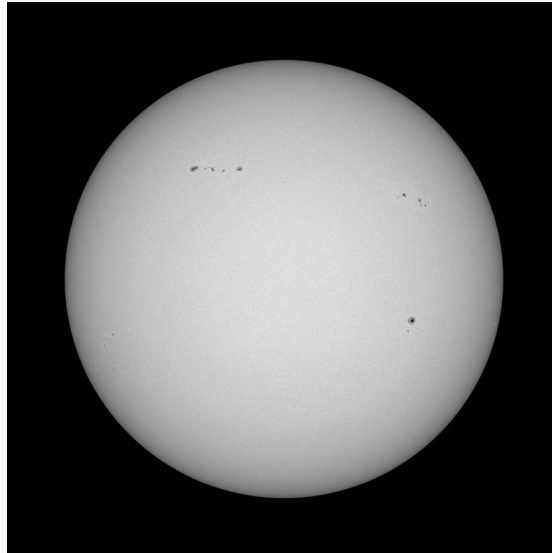
1.1.2 Le cycle solaire

Le Soleil est un astre actif, dont la dynamique possède une composante cyclique d'une période approximative de 11 ans, appelée cycle solaire (HATHAWAY 2010). Le cycle solaire se mesure notamment en observant le nombre et l'aire occupée par les taches solaires. Les taches solaires sont des régions de la photosphère plus froides (et émettant donc moins dans le visible, d'où un aspect plus sombre), causées par la présence d'un champ magnétique local très intense, qui bloque le mouvement de convection sous la photosphère et en empêche le chauffage. Les tâches solaires se forment par paire de polarité opposée, reliées par des boucles magnétiques (KIVELSON et RUSSEL 1995).

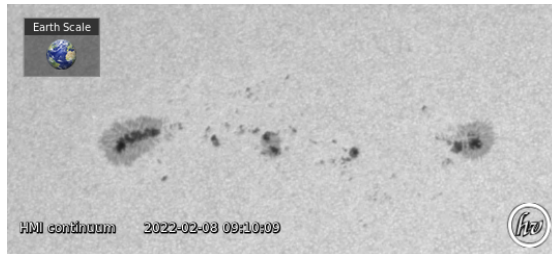
La Figure 1.3 montre une image de la photosphère solaire capturée le 8 février 2022 sur laquelle des taches solaires sont observables. Une vue agrandie d'un complexe de taches solaires est proposée, avec la Terre mise à l'échelle pour référence (1.3b). Le nombre de taches solaires est souvent utilisé comme un indicateur du cycle solaire, notamment car il est facilement observable (y compris depuis la Terre). C'est un indicateur historique : les premières valeurs mesurées et consignées rigoureusement remontent à 1749. La Figure 1.4 montre l'évolution du nombre de taches solaires (moyenné sur 13 mois) entre 1749 et 2019.

Cette figure montre bien que les cycles solaires présentent des intensités variables. En particulier, le dernier cycle solaire (cycle solaire 24, ayant débuté en janvier 2008 et s'étant terminé en décembre 2019) a été particulièrement faible. Notons également que la véritable périodicité du Soleil serait plutôt de 22 ans, la polarité du champ magnétique solaire s'inversant tous les 11 ans (MURSULA et al. 2002).

L'évolution du cycle solaire est également visible si on observe le Soleil dans l'EUV (donc depuis l'espace, l'atmosphère terrestre bloquant ces rayonnements).



(a) Photosphère solaire présentant des taches.



(b) Vue agrandie du complexe de taches solaires observable en haut à gauche de l’image 1.3a.

FIGURE 1.3 – Images de la photosphère solaire le 8 février 2022. Crédits : [SDO \(NASA\)](#). Images générées via [HELIOVIEWER](#).

Par exemple, à la longueur d’onde 171 Å on constate l’apparition puis la disparition progressive au cours d’un cycle solaire de zones plus brillantes dans la couronne, comme l’illustre la Figure 1.5. Ces zones sont appelées régions actives et sont situées au-dessus des taches solaires. Le champ magnétique local y est intense et complexe. Les taches solaires et les régions actives peuvent servir à caractériser différentes phases du cycle solaire (minimum, phase ascendante, maximum, phase descendante), au cours desquelles l’activité solaire présente des caractéristiques différentes.

1.1.3 Le(s) vent(s) solaire(s)

Comme nous l’avons dit, le Soleil est à l’origine d’une émission permanente de matière coronale thermiquement excitée, dont l’énergie permet une libération de la gravité solaire. Ce flux de particules est appelé vent solaire. Il s’agit d’un plasma peu dense, constitué de particules chargées (notamment des électrons, des ions et des protons) et au sein duquel les collisions sont rares ([MARSCH 2006](#)).

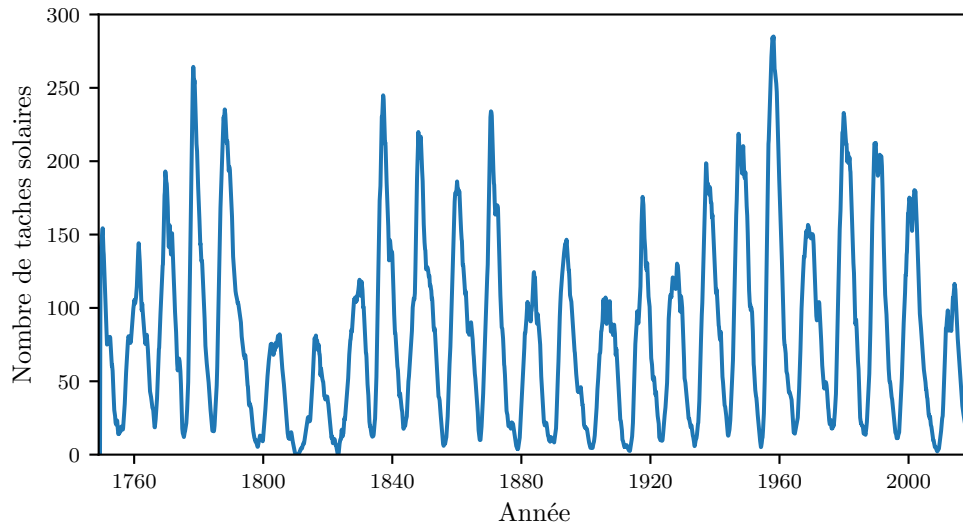


FIGURE 1.4 – Évolution du nombre de taches solaires moyenné sur 13 mois entre 1749 et 2019.

Ses propriétés (vitesse, densité, température, etc.) sont variables en fonction de l'activité solaire.

Le vent solaire est constitué de deux principales composantes. La première est un vent solaire lent dont la vitesse moyenne mesurée à 1 au est d'environ $400 \text{ km} \cdot \text{s}^{-1}$. La seconde est un vent solaire rapide, dont la vitesse mesurée à 1 au est typiquement de l'ordre de $600 - 800 \text{ km} \cdot \text{s}^{-1}$. Ci-dessous nous décrivons rapidement leurs origines solaires.

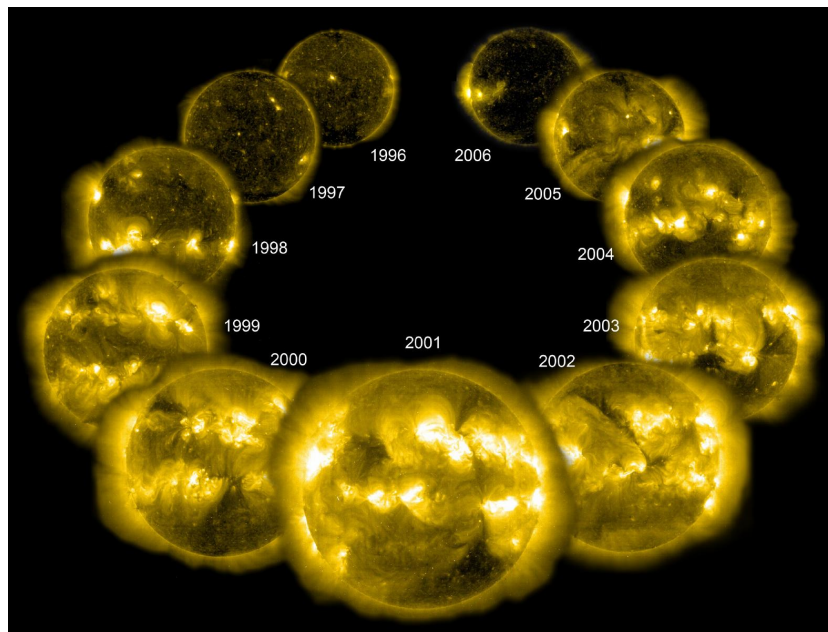


FIGURE 1.5 – Images du Soleil prises chaque année du cycle solaire 23 à la longueur d'onde 171 \AA . Crédits : SOHO (ESA & NASA).

1.1.3.1 Les sources du vent solaire lent : les *helmet streamers*

Le vent solaire est un plasma possédant une très grande conductivité électrique. Ainsi ¹, il va geler les lignes du champ magnétique du Soleil ² et les emporter dans son mouvement. Or, ces lignes de champ prennent leur source dans le Soleil, et sont donc entraînées par sa rotation sur lui-même. Le vent solaire étant émis radialement, cela lui confère un mouvement dans l’héliosphère en forme de spirale (illustré en Figure 1.6), appelée spirale de Parker, du nom de celui qui fut le premier à prédire ce phénomène avec un modèle (PARKER 1958). Le champ magnétique solaire entraîné ainsi dans l’héliosphère par le vent solaire est alors appelé *champ magnétique interplanétaire* ou *Interplanetary Magnetic Field (IMF)*.

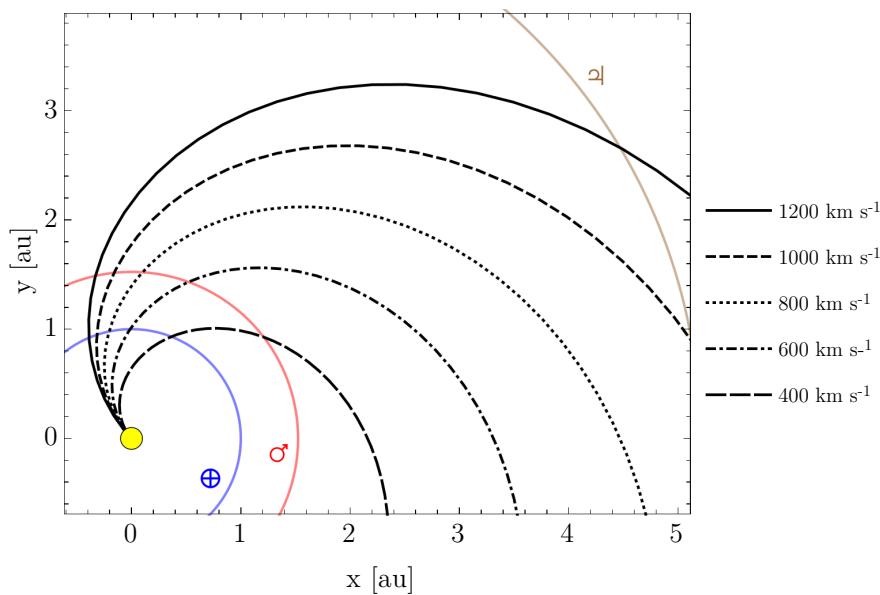


FIGURE 1.6 – Représentation schématique de la propagation du vent solaire dans l’héliosphère suivant la spirale de Parker. Plusieurs trajectoires sont tracées en fonction de la vitesse du vent solaire. L’orbite bleue correspond à celle de la Terre, la rouge à celle de Mars et la verte à celle de Jupiter. Reproduit de LHOTKA et NARITA 2019, sous licence CC BY 4.0.

Aujourd’hui nous savons, notamment grâce à la mission américaine Ulysses (MCCOMAS et al. 2000), que le vent solaire lent provient des régions de la couronne solaire appelées *helmet streamers* (parfois simplement *streamers*). Ce sont des structures en forme de boucle fermée allongée, reliant deux régions de polarité opposée sur le Soleil. Ces boucles contiennent du plasma coronal qui s’échappe progressivement, formant le vent solaire lent (OFMAN 2004). Les *streamers* sont au nombre de deux et proches de l’équateur pendant les phases de minimum solaire, dû à la nature globalement dipolaire du champ magnétique solaire à ce moment. La Figure 1.7 illustre ce phénomène.

1. C’est une conséquence du [Théorème d’Alfvén](#).

2. Cela signifie que le mouvement de la matière et celui des lignes de champ magnétique sont intriqués.

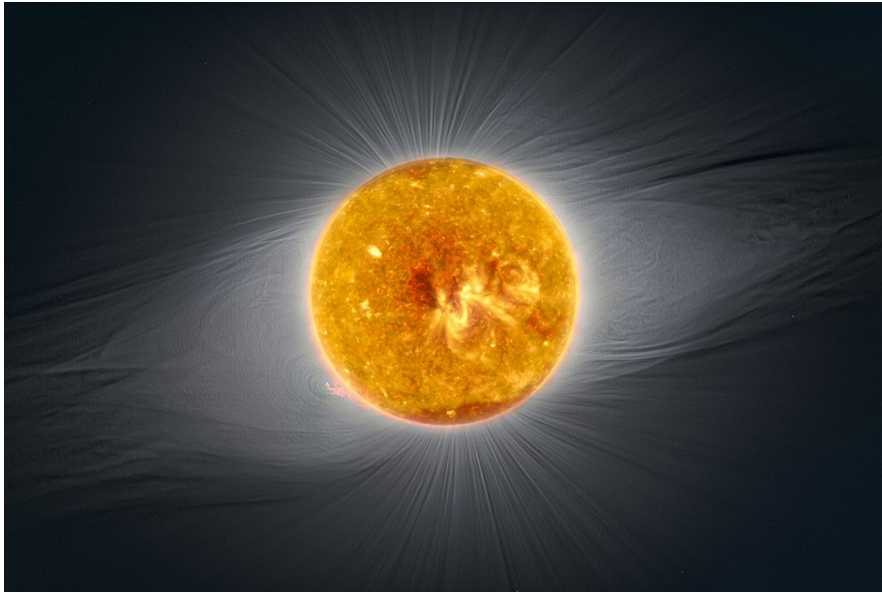


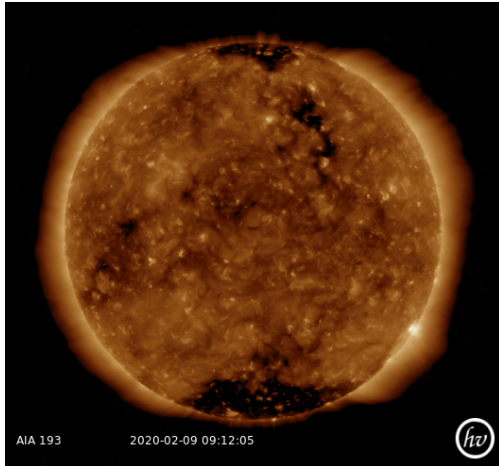
FIGURE 1.7 – Superposition de plusieurs images du Soleil à différentes longueurs d’onde (visible et EUV) capturées le 2 juillet 2019 depuis le sol et l’espace, montrant la structure des *helmet streamers* en minimum solaire. Des lignes de champ ouvertes partant des trous coronaux polaires sont également visibles. Crédits : ESO/P. Horálek; SOHO (ESA & NASA); SDO (NASA), sous licence CC BY 4.0.

Lorsque l’activité solaire s’intensifie, son champ magnétique se complexifie et le nombre et la structure des *streamers* également (RÉVILLE et BRUN 2017).

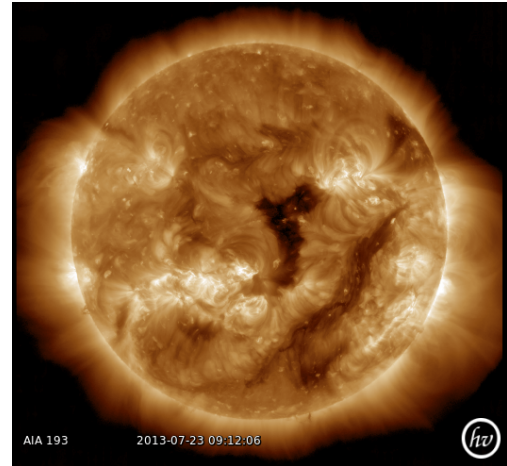
1.1.3.2 Les sources du vent solaire rapide : les trous coronaux

Le vent solaire rapide a pour source les trous coronaux, qui sont des régions de la couronne solaire de plus faible densité et température, desquelles partent des lignes de champ ouvertes vers le milieu interplanétaire (NOLTE et al. 1976). Les particules de la couronne peuvent donc être accélérées le long de ces lignes de champ et constituent le vent solaire rapide. Les trous coronaux sont observables selon certaines longueurs d’onde dans l’EUV et dans les rayons X. Ils y apparaissent alors comme de grandes surfaces sombres. Il existe différents types de trous coronaux. En phase de minimum solaire on observe le plus souvent deux grands trous coronaux situés aux pôles du Soleil. Ces trous coronaux polaires disparaissent petit à petit en phase ascendante et sont généralement absents en phase de maximum solaire. On les observe à nouveau en phase descendante. Des trous coronaux de plus petite surface sont observables à moyennes latitudes au cours des phases ascendante et maximum. Ces trous coronaux peuvent ensuite migrer vers l’équateur au cours de la phase descendante (CRANMER 2009). La Figure 1.8 montre deux exemples de trous coronaux polaires (1.8a) et de trous coronaux équatoriaux (1.8b).

Les trous coronaux sont des structures persistantes. Les trous coronaux polaires peuvent en effet être observés pendant plusieurs mois successifs et les trous coronaux équatoriaux suivent la rotation différentielle du Soleil et peuvent donc réapparaître



(a) Deux trous coronaux polaires observés le 2 février 2020.



(b) Deux trous coronaux de faible latitude observés le 23 juillet 2013.

FIGURE 1.8 – Images du Soleil montrant des exemples de trous coronaux. Crédits : SDO (NASA). Images générées via HELIOVIEWER.

au bout d’une rotation de Carrington (*i. e.* une rotation complète du Soleil, d’une période d’environ 27 jours).

1.1.3.3 Les régions à interactions de courants

Lorsque le vent solaire rapide rattrape un vent solaire lent émis précédemment, se crée alors une interface dans laquelle la densité et la pression du vent solaire du vent solaire augmente fortement. Cette région de vents solaires lents et rapides perturbés est appelée *région à interactions de courants* ou *Stream Interaction Region (SIR)*. Ces régions de compression créent des ondes de choc qui modifient la structure de l’IMF (PIZZO 1978). La Figure 1.9 schématise la formation d’une SIR dans l’héliosphère.

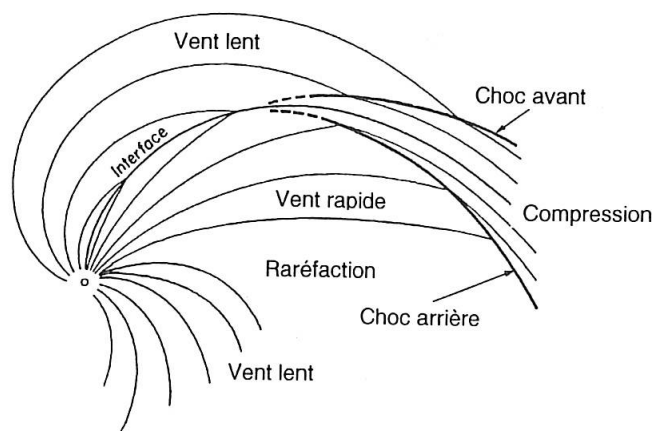


FIGURE 1.9 – Représentation schématique d’une SIR dans l’héliosphère. Figure reproduite de LANTOS 1997, adaptée d’une figure de CROOKER et SISCOE 1986.

Les trous coronaux étant persistants, certaines *SIRs* peuvent apparaître de façon récurrente après une ou plusieurs rotations de Carrington. On les appelle alors des *régions d'interaction en co-rotation* ou *Corotating Interaction Regions (CIRs)*. Du fait de l'influence du cycle solaire sur les trous coronaux, on observe plus de *SIRs* et de vents solaires rapides en phase descendante du cycle solaire. Pour illustrer cela, la Figure 1.10 (extraite de GRANDIN et al. 2019) montre la distribution annuelle des événements dûs aux vents solaires rapides (*High-Speed Streams* en anglais, incluant les *SIRs*) au cours des cycles solaires 23 et 24. On voit bien que ceux-ci sont bien plus nombreux en phase descendante du cycle, voire en minimum solaire mais ils ne sont pas inexistant non plus en phases ascendante et maximum.

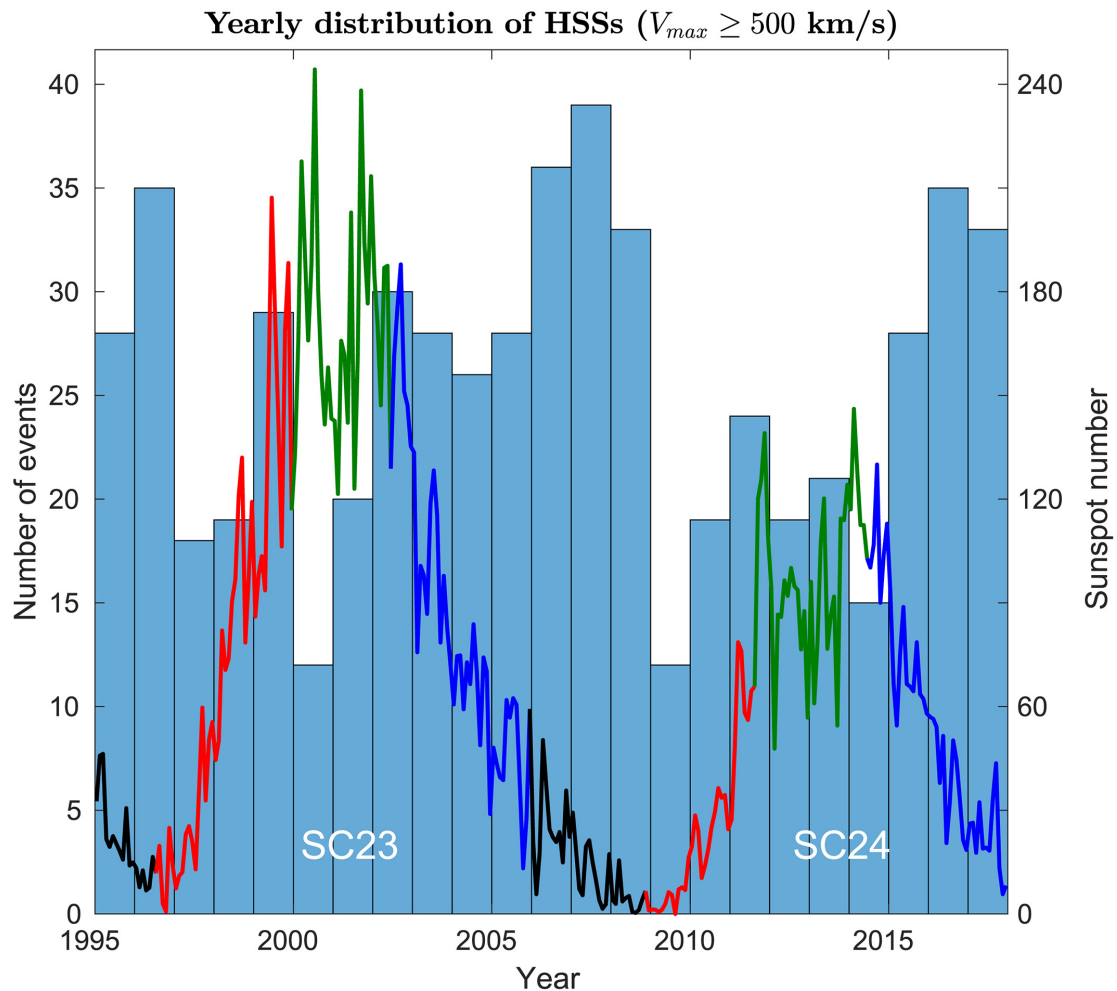


FIGURE 1.10 – Distribution annuelle de l'observation de vents solaires rapides à 1 au au cours des cycles solaires 23 et 24. Reproduite (avec autorisation) de GRANDIN et al. 2019.

1.1.4 L'activité éruptive du Soleil

Outre les émissions permanentes de vent solaire, le Soleil présente également une activité turbulente accompagnée de libérations soudaines et violentes d'énergie

sous différentes formes. Cette activité, dite éruptive, est en réalité constituée de différentes composantes plus ou moins bien corrélées entre elles, décrites ci-dessous.

1.1.4.1 Les embrillancements

Les embrillancements solaires (*solar flares* en anglais) sont des émissions soudaines de rayonnement électromagnétique à travers une vaste gamme de longueurs d’onde, dont surtout les rayons X et l’EUV. Les embrillancements surviennent le plus souvent dans les régions actives comportant des taches et leur occurrence est donc directement corrélée à l’activité cyclique du Soleil. Toutefois, on observe également des embrillancements dans des régions sans taches, ce qui rend leur prévision difficile (DODSON et HEDEMAN 1970). Leur durée d’observation va de plusieurs minutes à quelques heures. (HUDSON 1991 ; FLETCHER et al. 2011).

Les embrillancements sont très certainement la conséquence d’une libération d’énergie magnétique due à un phénomène appelé la reconnexion magnétique, au cours duquel les lignes de champ magnétique se réarrangent pour former de nouvelles structures (JANVIER 2017). Les embrillancements sont les événements éruptifs solaires les plus intenses, l’énergie libérée étant comprise entre 10^{17} et 10^{25} J (SCHRIJVER et al. 2012).

Les embrillancements peuvent être classés selon leur intensité mesurée dans différentes longueurs d’onde. Les échelles le plus utilisées reposent sur l’observation selon la longueur d’onde de la raie d’émission H_{α} ou – comme c’est plus souvent le cas aujourd’hui – selon les rayons X mous (HUDSON 2011). Dans ce dernier cas, les *flares* peuvent être classés en différentes catégories en fonction de leur intensité (du moins intense au plus intense : A, B, C, M et X). La Figure 1.11 montre un exemple d’embrillancement de classe X observé dans l’EUV le 9 août 2011.

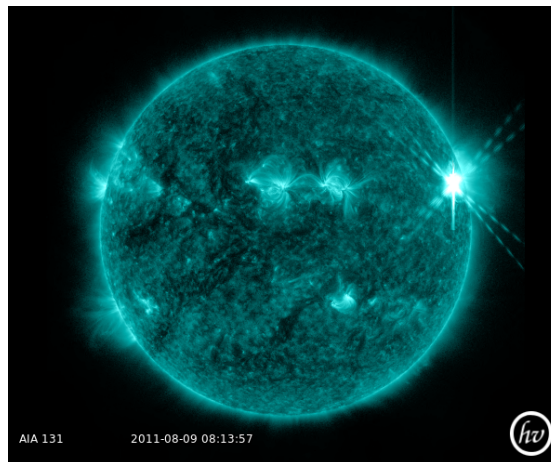


FIGURE 1.11 – Image du Soleil à la longueur d’onde 131 Å prise le 9 août 2011 dans laquelle un embrillancement de classe X est visible. Crédits : SDO (NASA). Image générée via HELIOVIEWER.

Les embrillancements, notamment les plus intenses d’entre eux, peuvent accélérer des particules en direction de l’espace interplanétaire au cours d’événements

solaires à particules ou *Solar Particle Events* (SPEs). Les particules, notamment des protons sont alors propulsés à des vitesses relativistes et, si elles sont dirigées vers la Terre, l'atteindre en moins de 10 minutes pour les plus énergétiques d'entre elles.

1.1.4.2 Les éjections de masse coronales

On observe parfois dans la couronne solaire des protubérances, comme le montre la Figure 1.12. Ces protubérances sont des arches de plasma chromosphérique en suspension dans la couronne³ à cause de structures magnétiques complexes encore mal comprises car difficiles à observer, même si plusieurs théories tentent de l'expliquer (GALSGAARD et LONGBOTTOM 1999 ; FAN 2021). Ce plasma étant issu de la chromosphère il est bien plus froid et bien plus dense que le plasma coronal qui l'entoure (la différence étant de deux ordres de grandeur).

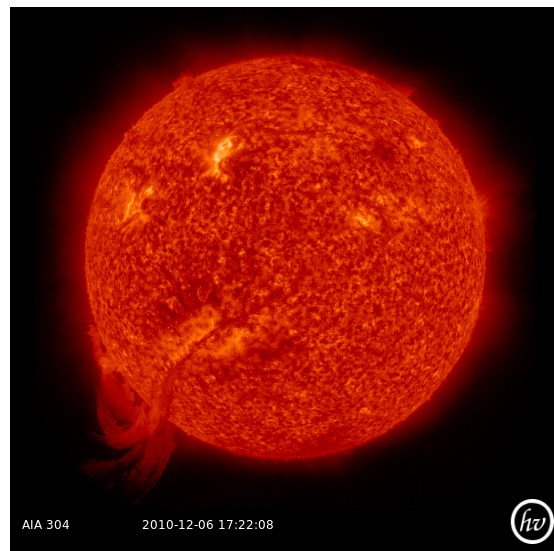


FIGURE 1.12 – Image du Soleil à la longueur d'onde 304 Å prise le 6 décembre 2010 dans laquelle une grande protubérance est visible dans le quart inférieur gauche de l'image. Crédits : SDO (NASA). Image générée via HELIOVIEWER.

Les structures magnétiques modelant les protubérances sont en équilibre permanent entre la pression exercée par le plasma piégé et la tension magnétique des lignes de champ. Cet équilibre peut être rompu dans un sens, lorsque la tension magnétique l'emporte, causant l'effondrement de la structure ; ou bien dans l'autre sens, lorsque la pression exercée par les particules leur permet de s'échapper, causant ainsi une *éjection de masse coronale* ou *Coronal Mass Ejection* (CME) (AULANIER 2013). Un exemple de CME est donné en Figure 1.13. Une CME émise dans la direction de la ligne de visée de la Terre est souvent appelée halo-CME car alors, vue depuis un coronographe situé sur cet axe, la CME apparaît comme un halo autour du Soleil. Une CME se propageant dans l'héliosphère et détectée

3. Notons que lorsqu'on observe une protubérance sur le disque solaire, elle apparaît comme une tache sombre longiligne dans la couronne et est alors appelée « filament ».

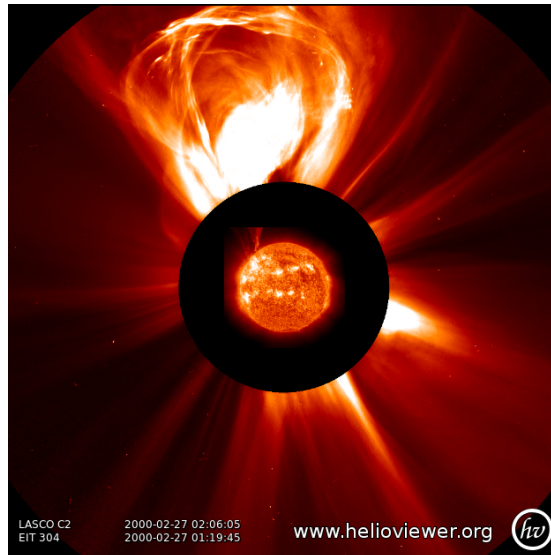


FIGURE 1.13 – Images superposées du Soleil (à la longueur d’onde 304 Å) et de sa couronne (en lumière blanche) prises le 27 février 2000 montrant une CME. Crédits : SOHO (ESA & NASA). Image générée via HELIOVIEWER.

par des instruments *in situ* est appelée éjection de masse coronale interplanétaire ou *Interplanetary Coronal Mass Ejection* (ICME).

Les phénomènes physiques menant aux embrillancements et aux CMEs sont liés, si bien qu’on observe une corrélation entre l’observation de *flares* et de CMEs. Toutefois on observe aussi bien des embrillancements non-suivis des CMEs et des CMEs non-accompagnées de *flares* (FLETCHER et al. 2011). Cela explique également le fait que l’occurrence d’apparition de CMEs soit bien corrélée au cycle solaire, avec toutefois un léger décalage. La Figure 1.14, extraite de GRANDIN et al. 2019, montre la distribution annuelle d’observation d’ICMEs dirigées vers la Terre à 1 au au cours des cycles solaires 23 et 24. On y voit bien que les ICMEs sont plus nombreuses en période de maximum solaire et jusqu’à environ 2 ans après.

1.1.5 Le vent solaire à 1 au

Il est difficile d’observer le vent solaire – et les autres plasmas d’origine solaire – partout dans l’héliosphère. Là où on le connaît le mieux, parce qu’on dispose d’instruments capables d’en mesurer les principales caractéristiques *in situ* (voir Section 2.2.2), c’est aux abords de notre planète, à environ 1 au du Soleil. Il est crucial de pouvoir caractériser le vent solaire en amont de la Terre, car c’est ce vent solaire qui interagira avec l’environnement spatial de notre planète (et plus particulièrement sa magnétosphère, dont en parlera en Section 1.2).

Pour caractériser le vent solaire on utilise souvent certains de ses paramètres physiques liés à ses propriétés dynamiques et électromagnétiques. Dans le Tableau 1.1 nous listons certains de ces paramètres et en donnons la valeur moyenne et l’écart-type observés pour différents types de plasmas solaires à 1 au. Les valeurs

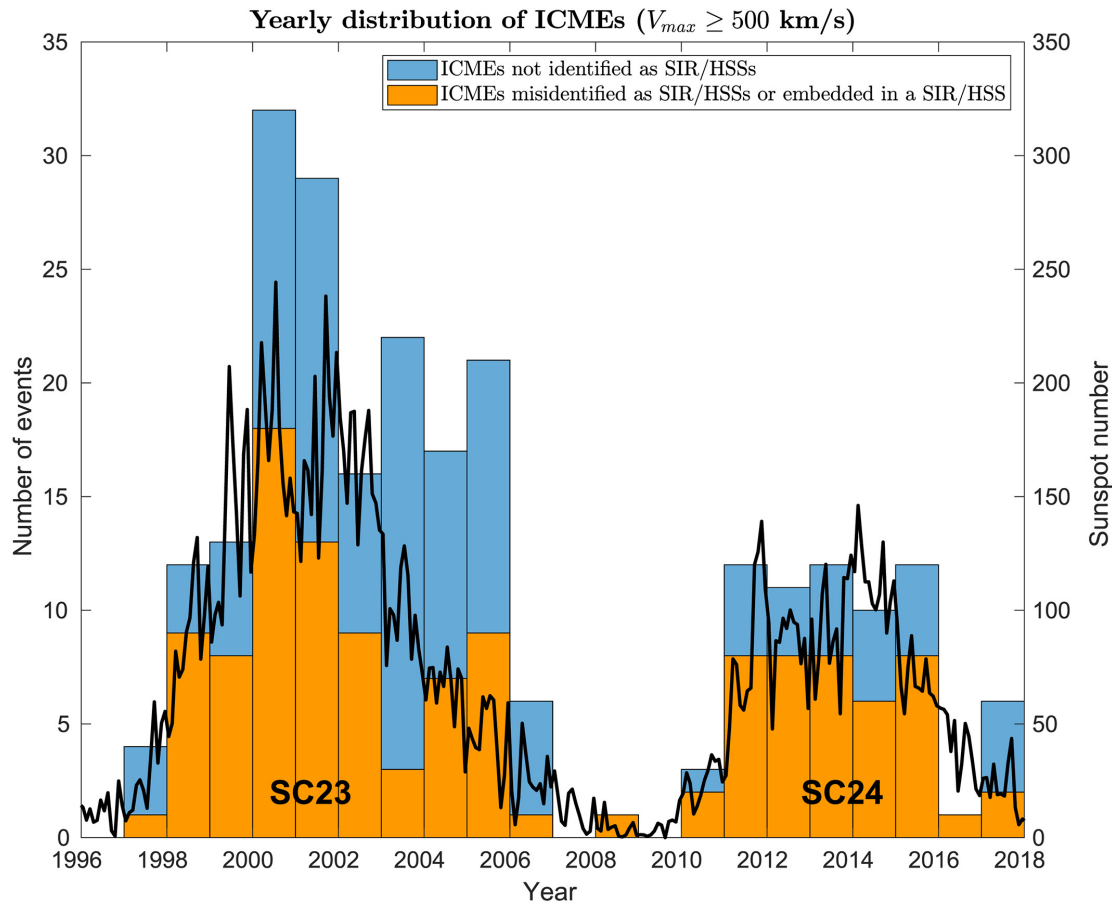


FIGURE 1.14 – Distribution annuelle de l'observation d'ICMEs dirigées vers la Terre à 1 au au cours des cycles solaires 23 et 24. Reproduite (avec autorisation) de GRANDIN et al. 2019.

proviennent de BOROVSKY 2020b (à l'exception du paramètre β , pour lequel les valeurs proviennent de EBERT et al. 2009).

Ces paramètres servent non seulement à caractériser les propriétés du vent solaire en amont de la Terre (et donc la façon et l'intensité avec laquelle ils vont interagir avec la magnétosphère terrestre), mais aussi à distinguer les différents régimes de vent solaire. En les analysant statistiquement et par l'étude de leur dynamique temporelle on peut obtenir des profils typiques pour les différents types de vent solaire, ce qui permet de comprendre l'origine solaire d'une éventuelle perturbation (voir par exemple BOROVSKY et DENTON 2010 ; BENACQUISTA et al. 2017 ; REGNAULT et al. 2020).

Ainsi, le passage d'une ICME sera souvent caractérisé par une modification importante des paramètres de champ magnétique et de plasma. Une ICME est généralement repérée par une discontinuité, définie par une augmentation importante du champ magnétique ainsi que de la vitesse, densité et température des protons. Les fluctuations importantes de ces paramètres permettent d'indiquer la présence d'une zone de compression, appelée aussi gaine. Le cœur de l'ICME est repéré par un éjecta magnétique, où les fluctuations du champ magnétique (dont

TABLEAU 1.1 – Description des principaux paramètres du vent solaire, dont on donne la valeur moyenne et l’écart-type observés à 1 au lors de vents solaires lents, rapides et pour des plasmas d’origine éruptive (valeurs issues de BOROVSKY 2020b et EBERT et al. 2009).

Paramètre	Unité	Valeur moyenne et écart-type		
		Vent solaire lent	Vent solaire rapide	Plasma d’origine éruptive
Vitesse (V_{sw})	$\text{km} \cdot \text{s}^{-1}$	410 ± 50	562 ± 80	429 ± 98
Densité de protons (ρ_{sw})	cm^{-3}	$3,2 \pm 1,7$	$5,6 \pm 3,2$	$6,4 \pm 5,7$
Intensité de l’IMF ($ B $)	nT	$5,3 \pm 2,0$	$5,8 \pm 2,1$	$10,6 \pm 4,6$
Composante verticale de l’IMF (B_z)	nT	—	—	—
Température des protons (T_{sw})	eV (parfois donnée en K)	$7,1 \pm 3,9$	$16,2 \pm 10,2$	$6,9 \pm 9,6$
Nombre d’Alfvén	1	$9,4 \pm 2,8$	$9,1 \pm 2,6$	$4,8 \pm 1,4$
Paramètre β du plasma	1	1,01	1,13	0,24
Temps d’arrivée à 1 au	heures	103 ± 12	76 ± 11	101 ± 19

l’intensité reste toujours aussi forte) diminuent, et où le β^4 du plasma est faible ($< 0,1$). Parfois, on observe le retournement d’une des composantes du champ magnétique, indiquant alors une structure plus cohérente en corde de flux torsadée. Enfin, le profil de la vitesse diminue typiquement selon une tendance linéaire au fur et à mesure de la traversée de l’éjecta (une signature qui peut être interprétée comme le résultat de l’expansion de celui-ci).

Au passage d’une SIR, on observera souvent une augmentation de la température et des autres paramètres du vent solaire (densité, vitesse, champ magnétique, pression), mais souvent de façon plus progressive que dans le cas d’une ICME. Dans les deux cas la vitesse du vent solaire augmente, mais avec des profils très différents, puisque pour la SIR, la vitesse reste importante pendant des durées bien plus longues, sans profil de décroissance linéaire. Enfin, le champ magnétique

4. Il s’agit du rapport de sa pression cinétique et de sa pression magnétique.

évolue peu. Ces différences sont partiellement illustrées dans les Figures 4.1 et 4.2 présentées dans le Chapitre 4.

Toutefois ces « signatures » caractéristiques sont très variables d'un événement à l'autre, car vents solaires lents, rapides et ICMEs peuvent interagir entre eux de différentes manières (par exemple, au passage d'une ICME ou d'une SIR on peut souvent – mais pas toujours – observer un choc), ce qui modifie les propriétés du plasma et de l'IMF à 1 au (KILPUA et al. 2017).

1.2 Les ceintures de radiations, au cœur de la magnétosphère terrestre

Nous venons de voir que le Soleil, mu par divers mécanismes pouvant se superposer les uns aux autres, émet en permanence du plasma dans l'héliosphère sous différentes formes. Ce plasma va donc interagir avec le milieu interplanétaire et les structures qui s'y trouvent, ce qui inclut notamment le champ magnétique terrestre. De ces interactions vont, entre autres, naître les ceintures de radiations.

Les ceintures de radiations terrestres sont des régions de l'environnement spatial proche de la Terre où des électrons et de protons de haute énergie se trouvent piégés et accélérés par le champ magnétique terrestre. Ces ceintures représentent une contrainte, voire une menace, pour toute activité spatiale menée dans les orbites qu'elles occupent.

1.2.1 Morphologie de la magnétosphère terrestre

Les ceintures de radiations sont au cœur de la magnétosphère terrestre, c'est-à-dire la région de l'espace dans laquelle l'influence du champ magnétique terrestre domine les processus physiques, formant une cavité magnétique plus ou moins stable. Afin d'étudier les ceintures, il est donc d'abord nécessaire de comprendre à la fois leur environnement géomagnétique et leur couplage avec le vent solaire. Pour construire les sections suivantes nous nous sommes appuyés sur les manuscrits et livres de BOURDARIE 1996 ; MAGET 2007 ; HERRERA 2017 ; DAHMEN 2020 ; KOSKINEN et KILPUA 2022.

1.2.1.1 Description générale du champ magnétique terrestre

La Terre possède un champ magnétique, dont la principale composante interne est créé par effet dynamo grâce à la rotation du noyau terrestre. À l'ordre 0 et lorsqu'on est suffisamment proche de la Terre, ce champ magnétique peut être modélisé par un dipôle magnétique, décentré d'environ 500 km par rapport au centre de la Terre et incliné de 11° par rapport à l'axe Nord-Sud géographique. La Figure 1.15 représente schématiquement cette approximation dipolaire du champ magnétique terrestre.

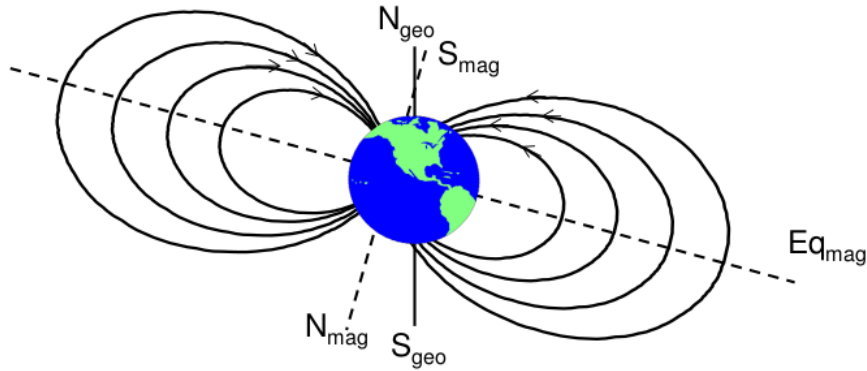


FIGURE 1.15 – Représentation schématique du champ magnétique terrestre, vu comme un dipôle incliné et décentré. Reproduit de DAHMEN 2020.

Cela a pour conséquence l'apparition d'une région, l'anomalie de l'Atlantique-Sud ou *South Atlantic Anomaly (SAA)*, dans laquelle, à altitude égale, le champ magnétique terrestre est moins intense. La Figure 1.16 montre une carte de l'intensité du champ magnétique terrestre à sa surface. En moyenne, l'intensité du champ magnétique terrestre va d'environ 65 000 nT aux pôles à 30 000 nT à l'équateur. Dans la *SAA* son intensité chute à environ 25 000 nT.

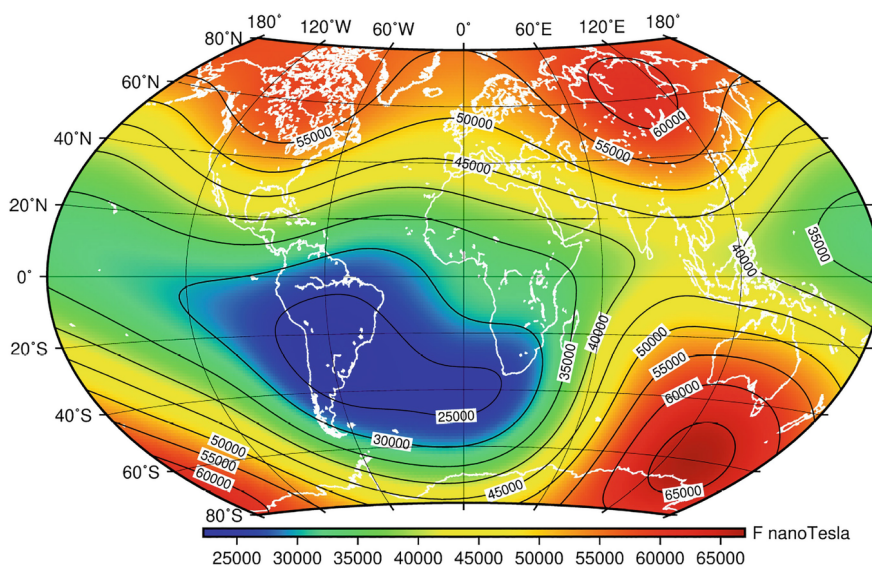


FIGURE 1.16 – Carte de l'intensité du champ magnétique terrestre à sa surface selon le modèle IGRF. Reproduit de KOSKINEN et KILPUA 2022, sous licence CC BY 4.0.

En réalité le champ magnétique terrestre est bien plus complexe qu'un simple dipôle et une modélisation réaliste nécessite la prise en compte de termes correspondant aux harmoniques sphériques d'ordre plus élevé du potentiel magnétique. C'est par exemple le cas du modèle *International Geomagnetic Reference Field (IGRF)*, qui est un des modèles de champ magnétique terrestre interne de référence aujourd'hui (ALKEN et al. 2021).

Pour bien comprendre la magnétosphère dans son ensemble, il faut, en plus

de la structure de champ magnétique interne, prendre en compte les interactions entre le vent solaire et le champ magnétique terrestre qui pilotent la physique des régions extérieures de la magnétosphère, notamment en déformant les lignes du champ géomagnétique et en injectant des particules chargées dans la cavité magnétosphérique. Ces particules verront alors leur mouvement dirigé par le champ magnétique terrestre, ce qui va contribuer à la création de courants induisant de nouvelles composantes au champ magnétosphérique.

Le vent solaire exerce en permanence une pression sur la magnétosphère terrestre, ce qui la comprime côté jour et l'allonge comme une « queue de comète » côté nuit. À cause de cette asymétrie, il est courant de se repérer longitudinalement dans la magnétosphère à l'aide d'une coordonnée appelée *temps magnétique local* ou *Magnetic Local Time (MLT)*, correspondant à l'angle formé entre le point qu'on cherche à repérer, le centre de la Terre et le centre du Soleil, le tout projeté à l'équateur magnétique terrestre. Ainsi, midi (MLT = 12 h) correspond au point face au Soleil, minuit (MLT = 0 h) au point opposé côté nuit.

La magnétosphère est composée de plusieurs régions, caractérisées par la présence de plasmas aux propriétés différentes. La Figure 1.17 représente schématiquement ces régions, vues selon une coupe longitudinale. On peut séparer grossièrement ces régions en deux grandes catégories : la magnétosphère externe et la magnétosphère interne.

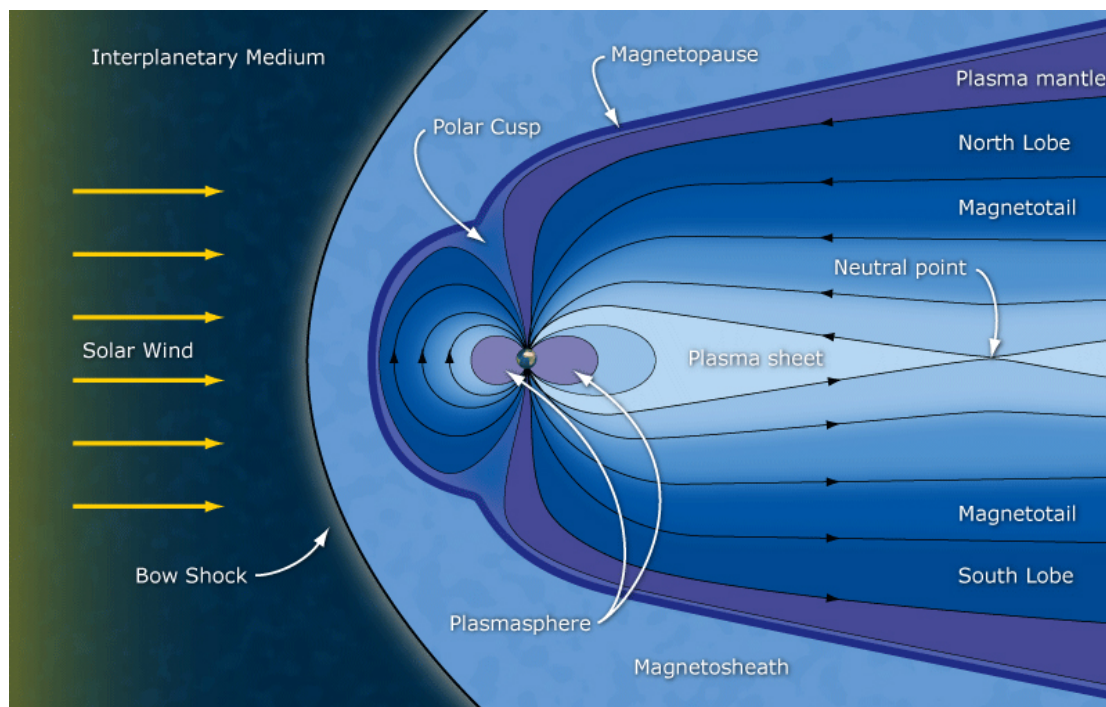


FIGURE 1.17 – Représentation schématique de la magnétosphère terrestre faisant apparaître ses différentes régions. Crédits : ESA/C. T. Russel.

1.2.1.2 La magnétosphère externe

La magnétosphère externe est constituée des régions de la magnétosphère pour lesquelles le forçage du vent solaire rend l’approximation quasi-dipolaire du champ géomagnétique inadaptée.

Le vent solaire s’approchant de la Terre se propage radialement à des vitesses supersoniques. En amont de la magnétosphère va donc se former un arc de choc (*bow choc* en anglais), situé à environ $15 R_T$ de la Terre sur l’axe Soleil–Terre. La position de l’arc de choc dépend des propriétés du vent solaire amont. Le vent solaire en aval du choc d’étrave y possède un écoulement turbulent, formant une région appelée magnétogaine (*magnetosheath*). Ce vent solaire va continuer de se propager en direction de la magnétosphère. Notons que c’est le plasma de la magnétogaine issu du choc qui va interagir directement avec la magnétosphère – et non pas le plasma du vent solaire. Cependant, comme nous le verrons dans le Chapitre 2, il est bien plus simple d’observer et mesurer *in situ* le plasma du vent solaire amont et non celui de la magnétogaine (TURC 2014).

L’interface séparant la magnétosphère – où le champ magnétique terrestre domine – et le milieu interplanétaire est appelée magnétopause. En première approximation il s’agit d’une couche de courant se formant lorsque les particules du vent solaire (des électrons et des ions) s’approchent de la Terre côté jour et deviennent alors soumises à la force de Lorentz due au champ magnétique terrestre, qui va les dévier selon des trajectoires dépendant du signe de leur charge électrique. La limite de la magnétopause sur le plan équatorial se situe à environ $10 R_T$ de la terre côté jour. Des vents solaires rapides et les ICMEs peuvent néanmoins repousser cette limite bien plus bas, sous l’orbite géostationnaire dans des cas extrêmes (LILENSTEN et BORNAREL 2001). Côtés aube et crépuscule cette distance est d’environ $16 R_T$ et côté nuit elle peut s’allonger bien au-delà de $100 R_T$. La Figure 1.18 schématise la magnétosphère terrestre et y fait notamment apparaître la magnétopause et le courant de magnétopause.

Bien que la magnétosphère dévie une grande majorité du plasma d’origine solaire, environ 2 % de l’énergie cinétique transportée par le vent solaire parvient à pénétrer la magnétosphère (SAUVAUD 2002). Une des « portes d’entrée » de la magnétosphère est constituée des cornets polaires (*polar cusps* en anglais). Ce sont les régions quasi-polaires (au sens magnétique du terme) desquelles partent toutes les lignes de champ magnétique terrestre, y compris des lignes ouvertes. La magnétopause plongeant également vers la Terre au niveau des cornets, le plasma d’origine solaire va pouvoir (partiellement) s’y engouffrer et, soit précipiter directement vers la Terre, soit se retrouver piégé dans la magnétosphère (GRISON 2006).

Une deuxième porte d’entrée de la magnétosphère est la magnétopause elle-même, qui est en réalité poreuse et permet à des particules situées sur une épaisseur appelée le manteau de plasma (*plasma mantle* en anglais) de pénétrer dans la magnétosphère. Cette entrée de particules s’effectue principalement côté nuit, où la vitesse du vent solaire est globalement parallèle aux lignes du champ géomagnétique. Les particules pénétrant ainsi dans la magnétosphère se retrouvent

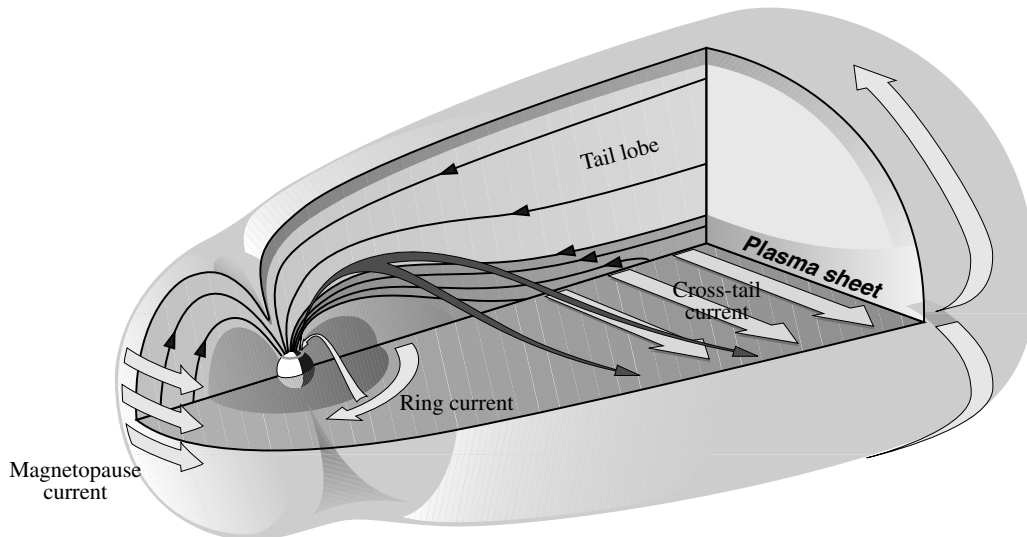


FIGURE 1.18 – Schéma de la magnétosphère et des grands systèmes de courants la composant. Figure de T. Mäkinen, reproduite (avec autorisation) de KOSKINEN 2011b.

donc d'abord dans la région de la magnétosphère s'étirant côté nuit, appelée la queue magnétosphérique (*magnetotail*). En réalité, ces deux processus d'entrée de particules dans la magnétosphère s'expliquent également en termes de reconnexion magnétique (voir Section 1.2.2.1).

La queue magnétosphérique constitue un réservoir de plasma ayant pénétré la magnétosphère. Elle est notamment composée des deux lobes magnétosphériques de polarité opposée (*tail lobe* sur la Figure 1.18), où le plasma est froid et très peu dense. Les lignes de champ y sont ouvertes, dirigées vers la Terre dans l'hémisphère nord et vers l'espace interplanétaire dans l'hémisphère sud.

Les lobes sont séparés par le feuillet de plasma (*plasma sheet*), qui est une couche de plasma chaud globalement neutre, d'une épaisseur de $3 - 7 R_T$, située dans le plan équatorial magnétique. Les lignes de champ y sont fermées et toutes les particules (chargées positivement et négativement) ont un mouvement général vers la Terre. Cela crée un courant, parfois appelé courant de Chapman-Ferraro ou courant de feuillet neutre (il est indiqué sur la Figure 1.18 sous le nom *cross-tail current*) (LILENSTEN et BLELLY 1999). Le Tableau 1.2 donne la densité moyenne du plasma et l'énergie des électrons et protons peuplant les différentes régions de la magnétosphère extérieure et la magnétogaine.

1.2.1.3 La magnétosphère interne

La magnétosphère interne est la grande région de la magnétosphère où le champ magnétique est quasi-dipolaire. Elle est composée de trois régions de haute densité peuplées différemment : la plasmasphère, le courant annulaire et les ceintures de radiations. Notons bien que ces trois régions ne sont pas forcément

TABLEAU 1.2 – Quelques paramètres moyens du plasma rencontré dans certaines régions de la magnétosphère externe (KOSKINEN et KILPUA 2022).

Région	Densité	Énergie des électrons (eV)	Énergie des protons (eV)
Magnétogaine	8	25	150
Lobes	0,01	50	300
Feuillet de plasma	0,3	600	4200

géographiquement distinctes. Bien au contraire, elles interagissent entre elles, ce qui a un impact considérable sur la dynamique des ceintures de radiations.

La première région de la magnétosphère interne est donc la plasmasphère. Il s’agit d’une région de forme toroïdale entourant la Terre et en co-rotation avec celle-ci (contrairement au courant annulaire et aux ceintures de radiations). Sa position (généralement entre $2 - 8 R_T$) et sa composition sont dynamiques et sont très dépendantes du couplage vent solaire–magnétosphère. Les particules y sont généralement de faible énergie, mais leur densité est très élevée (voir Tableau 1.2). La limite supérieure de la plasmasphère est la plasmopause.

Le courant annulaire (*ring current* en anglais) provient de la dérive azimutale des particules chargées à une distance de la Terre d’environ $2 - 6 R_T$. C’est un courant qui est proportionnel à la densité d’énergie des particules en dérive. Les principales particules contribuant au courant annulaire sont les électrons et les protons de $10 - 200$ keV, mais certains ions lourds (O^+ , He^+ , etc.) provenant de l’ionosphère⁵ et de l’atmosphère y contribuent également.

Enfin, les ceintures de radiations au cœur des motivations de nos travaux, sont décrites un peu plus en détails dans la prochaine section.

1.2.1.4 Les ceintures de radiations

L’existence des ceintures de radiations terrestres⁶ a été observée pour la première fois à la fin des années 50 de façon concomitante par les Russes (Sputnik 2, 1957) et les Américains (Explorer 1, 1958) mais ce sont ces derniers qui, par la voix de J. Van Allen, ont communiqué en premier à ce propos (ce qui a pour conséquence qu’aujourd’hui encore les ceintures de radiations sont parfois appelées ceintures de Van Allen). Notre connaissance des ceintures, de leur forme et de leur

5. Il s’agit de la partie supérieure de l’atmosphère terrestre, située à environ $80 - 600$ km d’altitude, où les rayonnements ionisants solaires et cosmiques créent une couche d’électrons et d’ions. La magnétosphère et l’ionosphère sont des systèmes couplés et on observe des transports d’énergie entre les deux systèmes, notamment via des courants électriques (MARCHAUDON 2018). C’est une autre région très importante du point de vue de la météorologie de l’espace mais elle n’est pas directement concernée par nos travaux.

6. D’autres planètes possédant un champ magnétique assez intense comme Jupiter et Saturne possèdent des ceintures de radiations, mais il n’en sera pas question dans nos études.

composition s'est accrue progressivement grâce à différentes missions successives (BAKER et PANASYUK 2017), la dernière en date de grande envergure étant sans nul doute les Van Allen Probes (lancées en 2012 et mises hors service fin 2019 FRETZ et al. 2016).

Comme nous l'avons évoqué précédemment, les ceintures de radiations sont une région de la magnétosphère interne. C'est une région où le champ magnétique est à la fois quasi-dipolaire et surtout intense, ce qui permet de piéger des particules de haute-énergie selon des mouvements quasi-périodiques.

Le mécanisme de piégeage dans les ceintures est une conséquence directe de la force de Lorentz à laquelle sont soumises les particules chargées dans un champ dipolaire stationnaire. En effet, si on a une particule de charge q et de vitesse \mathbf{v} est située dans un champ électromagnétique (\mathbf{E}, \mathbf{B}) , alors cette particule est soumise à la force de Lorentz \mathbf{F} qui s'exprime ainsi :

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \quad (1.1)$$

En supposant que l'on peut négliger la composante due au champ électrique \mathbf{E} alors on parvient à démontrer que la particule possède un mouvement périodique décomposable en trois sous-mouvements : la giration, le rebond et la dérive. En réalité l'hypothèse de stationnarité n'est pas parfaitement valable, tout comme l'hypothèse de négligence du champ \mathbf{E} . C'est pourquoi les particules ne sont pas parfaitement piégées et leurs mouvements sont qualifiés de quasi-périodiques. Le piégeage d'une particule peut toutefois durer plusieurs mois voire années.

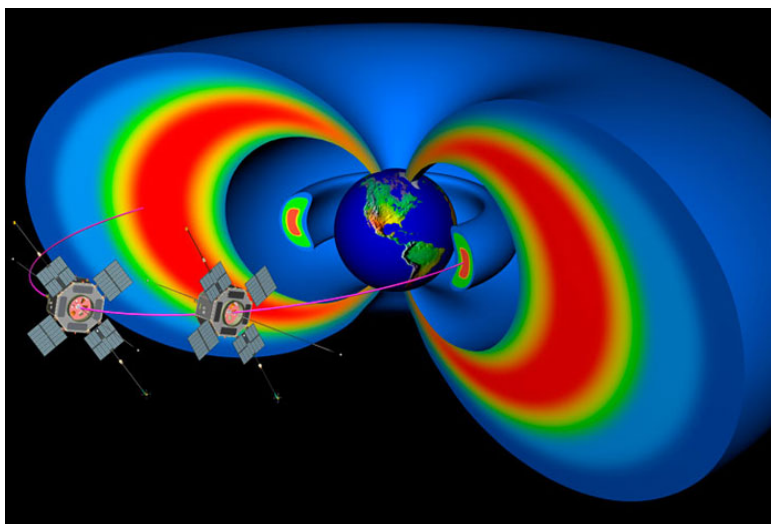


FIGURE 1.19 – Illustration des ceintures de radiations terrestres électroniques. Crédits : NASA illustrations, sous licence CC BY-NC-ND 2.0.

Les ceintures sont composées essentiellement d'électrons et de protons de haute à très haute énergie : 0,01 – 10 MeV pour les électrons et 0,01 – 1000 MeV pour les protons. Leur étendue globale est de 1 – 10 R_T pour les électrons et de 1 – 7 R_T pour les protons. Dans les faits, on observe plutôt habituellement deux ceintures d'électrons (1 – 2,5 R_T et 2,5 – 10 R_T) et une ceinture de protons (se

superposant partiellement avec la ceinture d’électrons intérieure). La Figure 1.19 propose une vue d’artiste des ceintures d’électrons.

Comme nous l’avons déjà dit, le mécanisme de piégeage des particules dans les ceintures de radiations peut être décomposé en trois mouvements : la giration, le rebond et la dérive, schématisés en Figure 1.20. La giration (1.20a) est un mouvement de rotation périodique décrit par une particule chargée autour d’une ligne de champ magnétique (au sein d’un champ électromagnétique pour lequel on peut négliger la composante électrique – au moins en première approximation). Le rebond (1.20b) est un mouvement le long de la ligne de champ, dans lequel la particule va osciller entre deux points miroirs dans chaque hémisphère. Le mouvement de dérive (1.20c) est un mouvement de rotation azimutal dans un plan perpendiculaire à l’axe du dipôle magnétique (BELMONT et al. 2013). L’ensemble de ces mouvements forme alors une coquille de dérive (1.20d), qui donne la forme toroïdale des ceintures de radiations.

Pour repérer une particule dans les ceintures de radiations on utilise souvent le paramètre de McIlwain L (MCILWAIN 1961) ou le paramètre de Roederer L^* (ROEDERER 1970). Ce sont tous deux des paramètres sans dimension. Dans le premier cas, L indique sur quelle ligne de champ se situe une particule lors de son déplacement. Toutefois, le champ magnétique n’étant pas un dipôle parfait, une particule va parcourir plusieurs valeurs de L au cours de son mouvement. L^* au contraire est défini pour être constant sur la coquille de dérive. L^* et L sont normalisés de telle sorte à ce que, pour une particule piégée dans une région où le champ magnétique peut-être considéré comme dipolaire et peu perturbé, les valeurs de L et L^* de cette particule soient quasiment identiques et équivalentes à la distance à la Terre (en rayons terrestres) de la particule sur le plan équatorial (ROEDERER et LEJOSNE 2018).

Les périodes associées à ces mouvements quasi-périodiques pour des électrons et des protons d’énergie $E = 1$ MeV à environ $2 R_T$ du centre de la Terre sont données dans le Tableau 1.3. Les valeurs précises dépendent beaucoup de l’orbite et de l’énergie de la particule, mais les différences d’ordre de grandeur persistent.

TABLEAU 1.3 – Période de giration, de rebond et de dérive pour des électrons et des protons de 1 MeV à environ $2 R_T$ du centre de la Terre. Reproduit de HERRERA 2017.

Mouvement	Période associée (s)	
	Électrons	Protons
Giration	$\approx 3,0 \times 10^{-5}$	$\approx 2,0 \times 10^{-2}$
Rebond	$\approx 1,4 \times 10^{-1}$	$\approx 2,7$
Dérive	$\approx 1,8 \times 10^3$	$\approx 1,2 \times 10^3$

La dynamique des ceintures de radiations est dictée par les arrivées et les pertes de particules, qui sont dues à différents processus physiques ayant quasiment tous pour point commun d’être liés à l’activité géomagnétique.

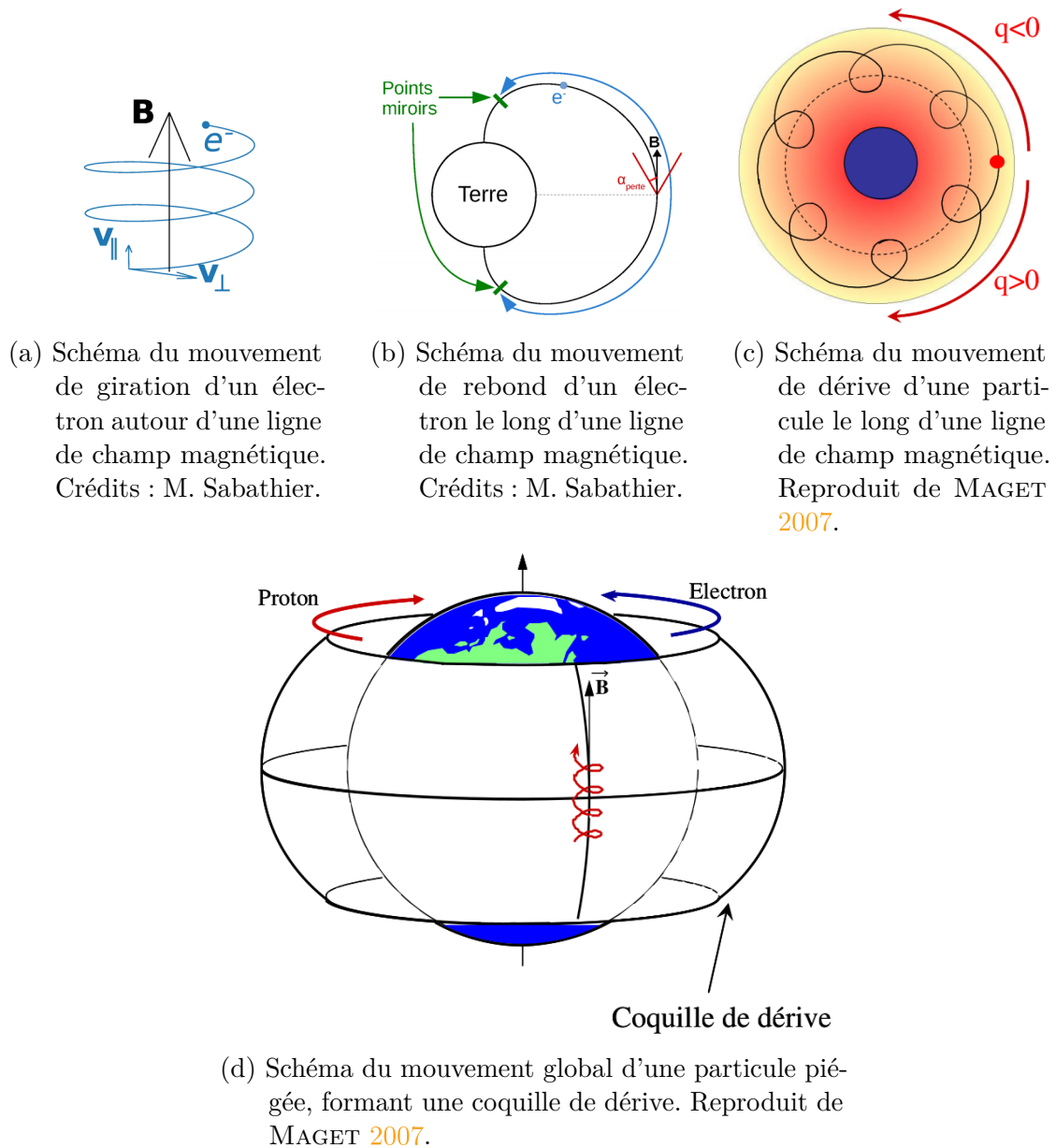


FIGURE 1.20 – Représentations schématiques des trois mouvements quasi-périodiques associés au piégeage d'une particule dans les ceintures de radiations.

1.2.2 L'activité géomagnétique

L'arrivée de vents solaires violents et autres ICMEs dans l'environnement spatial proche de la Terre peut causer des orages et sous-orages géomagnétiques. Ce sont des éléments importants du pilotage de la dynamique spatio-temporelle des ceintures de radiations.

1.2.2.1 La convection magnétosphérique

On observe que le plasma contenu dans la magnétosphère effectue des mouvements d'advection à grande-échelle, parfois appelé convection magnétosphérique. DUNGEY 1961 a expliqué ce phénomène en termes de reconnexion magnétique. Cette reconnexion magnétique survient au cours d'un cycle (appelé cycle de Dungey) qu'on pourrait décomposer en trois étapes. La première étape consiste en la scission de lignes de l'IMF et leur connexion à des lignes de champ géomagnétique du côté jour, ce qui crée une brèche dans le bouclier magnétique terrestre. Cette brèche va faire descendre la position de la magnetopause, ce qui peut libérer des particules jusqu'alors piégées dans la cavité magnétosphérique. Les lignes de champ géomagnétique dans le secteur jour de la magnétosphère étant orientées vers le nord, la connexion entre les lignes de l'IMF et de la magnétosphère est d'autant plus forte que la composante verticale de l'IMF est orientée vers le sud.

Au cours de la deuxième étape, le vent solaire repousse les lignes de champ nouvellement reconnectées en direction du côté nuit de la magnétosphère, vers la queue magnétosphérique. L'empilement ainsi créé des lignes de champ dans la queue va les pousser en direction du feuillet de plasma. Les lignes de champ du côté nord et sud de la queue magnétosphérique peuvent alors à leur tour se reconnecter côté nuit. La nouvelle boucle, étirée, va se replier en direction de la Terre, perturbant au passage les différents courants constituant la magnétosphère. Les particules ainsi amenées dans la magnétosphère interne vont pouvoir dériver longitudinalement jusqu'au côté jour, bouclant ainsi le cycle. La Figure 1.21 schématise ce phénomène de convection magnétosphérique.

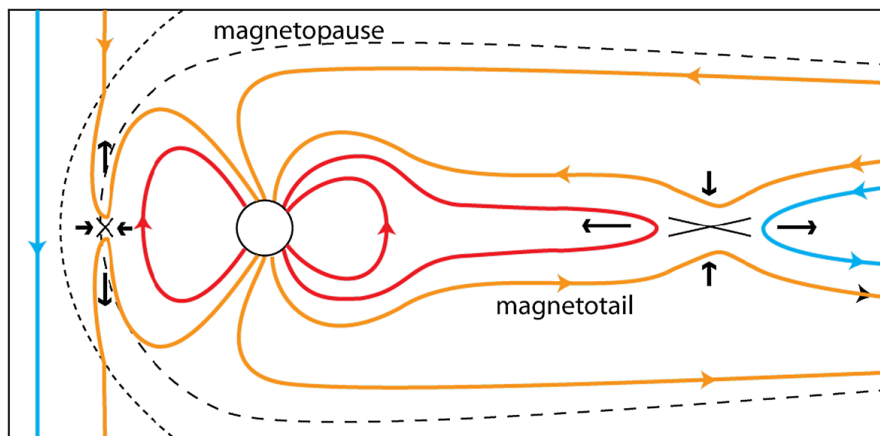


FIGURE 1.21 – Schéma de la circulation de plasma dans la magnétosphère terrestre selon le modèle du cycle de Dungey. Les croix désignent les zones où les reconnexions magnétiques du côté jour et du côté nuit ont lieu. Reproduit de EASTWOOD et al. 2017, sous licence CC BY 4.0.

1.2.2.2 Les orages et sous-orages géomagnétiques

La convection magnétique est un des processus par lequel l'énergie du vent solaire pénètre la magnétosphère. Comme les taux de reconnexion des côtés diurne

et nocturne ne sont pas nécessairement équivalents, un phénomène d'accumulation d'énergie dans la queue magnétosphérique et dans le feuillet de plasma peut avoir lieu. Cette accumulation d'énergie a très certainement pour conséquence un étirement du feuillet de plasma qui peut mener à une reconnexion magnétique bien plus proche de la Terre que celle prévue par le cycle de Dungey (vers $12 - 30 R_T$). Cette reconnexion violente déclenche alors un sous-orage géomagnétique (BAKER et al. 1996).

L'énergie magnétique libérée lors de la reconnexion va en partie chauffer et accélérer le plasma contenu dans la boucle nouvellement formée. Ces particules vont alors être injectées dans la magnétosphère intérieure, dans les ceintures de radiations et l'ionosphère. Elles sont une des principales causes des aurores polaires. Le temps caractéristique associé à un sous-orage est de l'ordre de l'heure (EASTWOOD et al. 2017).

Un orage géomagnétique survient lorsque les conditions de reconnexion magnétique pouvant mener aux sous-orages perdurent. Les orages se caractérisent alors par une intensification des courants magnétosphériques (incluant le courant annulaire) et une injection plus forte de particules dans les ceintures de radiations que lors des sous-orages. L'intensité des orages géomagnétiques est souvent caractérisée par un indice géomagnétique (voir Section 2.2.3). Leur temps caractéristique est de l'ordre du jour.

Les orages et sous-orages sont souvent déclenchés par des vents solaires rapides, des SIRs et des ICMEs, voire par des événements hybrides combinant les deux. Les orages géomagnétiques sont donc directement liés à l'activité solaire. En phase descendante du cycle solaire, ce sont les SIRs qui sont responsables de la majeure partie des orages (BOROVSKY et DENTON 2006; GRANDIN et al. 2015). Toutefois les ICMEs semblent être la cause des orages géomagnétiques les plus extrêmes (GONZALEZ et al. 2011).

À noter que tous les SIRs et ICMEs ne mènent pas forcément à un orage géomagnétique. De plus, tous les orages ne vont pas forcément impacter les ceintures de radiations de la même façon. Afin de déterminer la géoefficacité potentielle d'un orage, il est important de regarder la composante verticale de l'IMF, B_z , qui, comme nous l'avons vu, joue un rôle prépondérant dans le processus de déclenchement des orages. Notons cependant que des sous-orages peuvent survenir, certes plus rarement, même quand B_z est positif.

1.2.2.3 Le pilotage des ceintures de radiations par l'activité géomagnétique

La dynamique des ceintures de radiations est en très grande partie dirigée par l'activité géomagnétique. Ce pilotage a plusieurs composantes. Tout d'abord, l'activité géomagnétique est corrélée à des processus de source dans les ceintures de radiations (c'est-à-dire des processus qui vont contribuer au remplissage des ceintures). Parmi ceux-là, la diffusion radiale joue un rôle très important. Il s'agit du processus du transport de particules entre différentes coquilles de dérives (et donc différentes valeurs de L^*) à cause des fluctuations du champ magnétique. Ce

processus s’intensifie lors des orages géomagnétiques et peut injecter et accélérer de nouvelles particules de haute énergie dans les ceintures de radiations (LEJOSNE et KOLLMANN 2020).

Mais l’activité géomagnétique est également corrélée à des processus de perte. En effet, les perturbations géomagnétiques peuvent intensifier les ondes électromagnétiques se propageant dans la plasmasphère. Les différentes interactions onde-particule qui en résultent sont notamment source de pertes de particules par précipitation dans l’atmosphère (SICARD-PIET et al. 2014). Un autre processus de perte corrélé avec l’activité géomagnétique réside dans les *dropouts*, qui consistent en la perte très rapide (temps caractéristique inférieur à au temps de dérive des particules) de particules par la magnétopause, du côté jour. Cette perte est due à la descente de la magnétopause côté jour qui peut être causée simultanément par reconnexion magnétique avec le vent solaire et par la pression que celui-ci exerce sur le champ magnétique terrestre (SHUE et al. 1998 ; HERRERA et al. 2016).

Mentionnons également qu’il existe des processus de sources de particules indirectement corrélées avec l’activité géomagnétique (mais corrélées avec l’activité solaire), comme le remplissage des ceintures par des particules très énergétiques issues par exemple des SPE ou des interactions entre ICMEs et vent solaire. Un autre processus intéressant à mentionner est aussi le *Cosmic Ray Albedo Neutron Decay* (CRAND), processus par lequel des particules cosmiques viennent détacher des neutrons des noyaux de particules neutres de la haute atmosphère ou des ceintures. Ces neutrons, instables par nature, vont se décomposer et donner lieu notamment à des protons très énergétiques qui vont contribuer au remplissage des ceintures (NÉNON 2018). Comme les rayons cosmiques peuvent interagir avec le vent solaire et l’IMF avant d’atteindre la Terre, le CRAND est inversement corrélé avec l’activité solaire (POTGIETER 2008).

La combinaison de tous ces processus physiques résulte en des changements d’équilibre fréquents des ceintures de radiations, pilotés en très grande partie par l’activité géomagnétique. Cette dynamique n’est de plus pas la même en fonction des espèces considérées et de leur énergie, ce qui rend l’ensemble des interactions très complexes.

1.3 Les interactions Soleil–Terre vues du point de vue des activités humaines

En introduction de nos travaux nous avons mentionné certains événements causés par les activités solaires et géomagnétiques ayant eu des conséquences marquantes. Les conséquences potentielles des interactions Soleil–Terre sont aujourd’hui bien connues et leur impact sur les sociétés modernes fait l’objet d’estimations économiques qui en justifient la nécessité d’une prise en compte accrue (HAPGOOD et al. 2021). Ci-dessous nous décrivons succinctement certains des effets de l’activité géomagnétique sur les activités humaines.

1.3.1 Les conséquences de l’activité géomagnétique et de l’environnement radiatif spatial sur les activités humaines

Les effets des interactions Soleil–Terre peuvent être classés en quatre grandes catégories : les effets sur les longs conducteurs, sur les télécommunications et sur les infrastructures aérospatiales (LANZEROTTI et BAKER 2017), auxquels s’ajoutent les effets biologiques. Nous en mentionnons quelques exemples (sans être exhaustifs) ci-dessous.

Les effets sur les longs conducteurs sont principalement connus sous le nom de courants induits géomagnétiquement ou *Geomagnetically Induced Currents* (GICs). Ces courants se produisent lorsque les courants magnétosphériques et ionosphériques varient fortement, ce qui induit la création d’un courant dans les conducteurs au sol. Ces courants affectent principalement les longs conducteurs tels les pipelines et les lignes à haute tension, ce qui peut les endommager. Ce sont des GICs qui ont causé les perturbations du réseau de télégraphie observées aux XIX^e et XX^e siècles, ainsi que la coupure d’électricité de mars 1989 au Québec (BOLDUC 2002). Le fait que cet événement ait eu lieu au Québec n’est pas surprenant car l’intensité des GICs est plus forte près des pôles. Néanmoins il est attendu qu’un orage extrême comme celui de Carrington puisse causer des GICs même à des latitudes plus faibles (SISCOE et al. 2006). Les réseaux de distribution d’électricité étant un des piliers de nos sociétés modernes, on considère que les GICs sont un des plus grands dangers sociétaux liés aux interactions Soleil–Terre (EASTWOOD et al. 2017).

Les effets des interactions Soleil–Terre sur les télécommunications sont multiples. Parmi ceux-ci, les coupures radio sont les plus connues. Elles arrivent principalement lorsque le rayonnement X intense émis par le Soleil lors d’un embrillancement vient perturber la magnétosphère et surtout l’ionosphère. Cela peut parasiter ou empêcher la propagation de signaux radio, utilisés notamment par les systèmes de télécommunications, mais aussi par les systèmes de positionnement par satellite comme le *Global Positioning System* (GPS). Dans ce dernier cas la précision des signaux GPS peut être fortement dégradée (on parle alors de scintillement GNSS), ce qui peut avoir des conséquences importantes pour la sécurité de l’avion civile ou encore pour les activités militaires.

La dernière catégorie d’effets concerne les systèmes et infrastructures spatiales et est principalement imputable à l’environnement radiatif spatial, dû par l’activité solaire et l’activité géomagnétique. Au sein de cette catégorie d’effets on peut en distinguer deux sortes : les effets cumulés et les effets sporadiques. Les effets cumulés comprennent tout d’abord des effets de vieillissement, dus aux rayonnements, qui affectent tous les matériaux, y compris au sol. Ces effets sont encore plus intenses dans l’espace, où l’atmosphère terrestre ne protège plus les matériaux d’une grande partie du spectre électromagnétique. Parmi les conséquences du vieillissement on trouve des altérations des propriétés structurelles et optiques des matériaux utilisés, ce qui peut nuire au bon fonctionnement des composants affectés. En plus de ces effets non-ionisants on trouve également des effets de dose ionisante cumulée, qui peuvent être à l’origine de dérives des propriétés électriques de composants

électroniques (INGUIMBERT 2018).

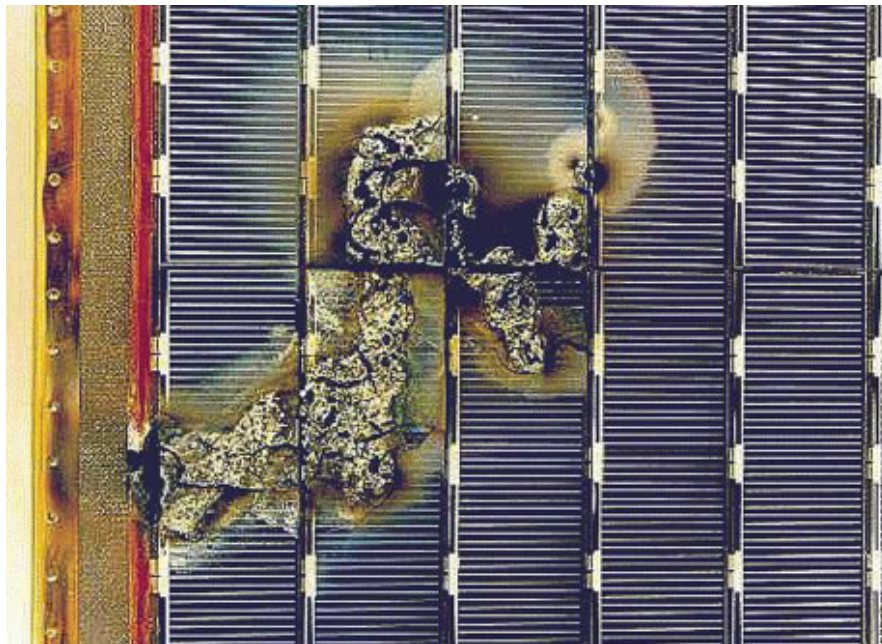


FIGURE 1.22 – Photographie montrant des cellules solaires de la mission EURECA ayant subi une décharge électrostatique au cours de la mission. Crédits : ESA.

Parmi les effets sporadiques causés par l'environnement spatial on trouve en premier lieu les effets de charge. Pouvant affecter les matériaux de surface, les panneaux solaires ou encore les matériaux isolants au cœur d'un satellite et de ses instruments, ils consistent en une accumulation locale et temporaire de charge à cause des radiations environnantes. Lorsque la charge accumulée dépasse certains seuils, la différence de potentiel électromagnétique ainsi créé peut entraîner des décharges électrostatiques violentes, pouvant définitivement endommager un satellite ou un de ses composants. La Figure 1.22 est une photographie d'une partie des panneaux solaires du satellite européen EURECA ayant subi une décharge qui les a irrémédiablement endommagés, dégradant ainsi l'espérance de vie de la mission. Mentionnons également les effets singuliers, qui consistent en la dégradation transitoire ou permanente de l'information transmise par un composant électronique (par exemple l'inversion d'un bit) suite au passage d'une particule chargée de haute énergie (HORNE et al. 2013). Notons qu'on observe un plus grand nombre d'anomalies au-dessus de la SAA, car les satellites en orbite basse terrestre ou *Low Earth Orbit (LEO)* y traversent alors fréquemment les ceintures de radiations internes. Cette corrélation entre la position de la SAA et l'observation d'anomalies montre bien que les ceintures de radiations représentent une menace permanente mais dynamique pour toutes les missions spatiales.

Enfin, à l'heure où les grandes puissances spatiales s'apprêtent de nouveau à envoyer l'humain sur la Lune, les effets biologiques de l'environnement radiatif spatial sont à nouveau source de beaucoup d'intérêt. Ces effets sont principalement dus aux rayonnements ionisants qui, en interagissant avec les tissus biologiques et les cellules, altèrent les propriétés physico-chimiques des molécules, incluant l'ADN.

Cela peut provoquer des effets déterministes (lorsqu'on dépasse un seuil de dose reçue), ou encore stochastiques, par définition plus imprévisibles et qui peuvent se manifester plusieurs années après l'exposition aux radiations. En réalité, nul besoin d'être dans l'espace pour commencer à observer ces effets, puisqu'aux altitudes de l'aviation civile on mesure déjà une plus grande intensité de radiations, corrélée avec l'activité géomagnétique. Il faut donc prendre en compte ces effets lorsqu'on effectue des missions habitées, que ce soit dans l'orbite terrestre ou même au-delà, notamment au cours de leur passage dans les ceintures de radiations (UNGER 2019 ; ZENCHENKO et BREUS 2021 ; DIMITROVA et BABAYEV 2021).

1.3.2 La météorologie de l'espace

Les effets de l'environnement radiatif spatial sur les activités humaines sont fortement corrélés à l'activité solaire et à l'activité géomagnétique, qui perturbent en permanence l'équilibre radiatif des régions traversées par les orbites des satellites. La météorologie de l'espace (*Space Weather* en anglais) est une discipline à la croisée de l'héliophysique, de la géophysique spatiale et de l'aéronomie, au sein de laquelle on s'intéresse aux interactions Soleil-Terre du point de vue de leurs conséquences sociétales, dans le but de les mitiger du mieux possible (BAKER 1998).

Ainsi, la météorologie de l'espace consiste notamment à modéliser (physiquement, empiriquement) les différentes composantes des interactions Soleil-Terre liées aux activités solaires et géomagnétiques. Les modèles qui en résultent sont souvent utilisés à des fins de prévision, dans le but d'anticiper avec la plus grande avance possible un potentiel effet néfaste, ou bien, si ce sont des modèles de climatologie, à des fins d'ingénierie (mieux comprendre l'environnement radiatif spatial permet de mieux le prendre en compte dans la conception d'une mission spatiale BOTHMER et DAGLIS 2007).

Aujourd'hui toutes les plus grandes agences et puissances spatiales s'intéressent à la météorologie de l'espace, certaines d'entre elles proposant des services opérationnels dédiés à l'observation et la prévision de l'activité solaire et de l'activité géomagnétique, avec émission d'alertes lors de périodes de dangers (CALDWELL et al. 2017 ; HAPGOOD et al. 2021). Toutefois, la météorologie de l'espace est une discipline encore jeune, qui a progressivement émergé depuis la deuxième moitié du XX^e siècle et qui, malgré son dynamisme actuel, n'est pas encore aussi mature que ses contreparties terrestres – la météorologie et la climatologie « classiques » (CADE III et CHAN-PARK 2015).

1.4 Bilan du chapitre

Nous venons de voir que le Soleil est à l'origine d'une émission permanente de particules via le vent solaire et des phénomènes éruptifs. Ceux-ci se propagent dans l'héliosphère et interagissent avec la magnétosphère terrestre. La réponse de la magnétosphère à la pression exercée par le vent solaire est complexe, et

résulte notamment en la création de ceintures de radiations autour de la Terre, qui couvrent des régions se superposant aux orbites occupées par des satellites artificiels. Ces ceintures de radiations sont en équilibre permanent entre l’arrivée et la perte de particules de haute énergie. Cette dynamique est principalement dirigée par l’activité géomagnétique, elle-même résultant du couplage vent solaire–magnétosphère, et donc à fortiori de l’activité solaire.

Les ceintures de radiations posent un danger pour les activités humaines. Afin de mitiger les divers effets de l’environnement spatial sur les missions spatiales, on s’intéresse à la prévision de leur dynamique, au sein d’une discipline appelée météorologie de l’espace. C’est précisément dans ce contexte que se situent nos travaux. Notre objectif, déjà évoqué en introduction de ce manuscrit, sera d’améliorer nos capacités de prédiction de l’activité géomagnétique terrestre, en gardant en ligne de mire la géoeffectivité des événements du point de vue des ceintures de radiations.

Pour cela, il nous faut d’abord étudier comment les interactions Soleil–Terre peuvent être observées, mesurées et surtout modélisées à l’échelle du sous-système vent solaire–magnétosphère, voire du système Soleil–Terre entier. Dans le chapitre suivant nous allons passer en revue l’état de l’art relatif à ce sujet. Comme nous le verrons, parmi les méthodes utilisées aujourd’hui, celles relevant de l’apprentissage automatique sont parmi les plus prometteuses. C’est pourquoi nous commencerons le Chapitre 2 par une courte, mais nécessaire, introduction à la modélisation dirigée par les données, avec une emphase mise sur les méthodes d’apprentissage profond.

Chapitre 2

Étudier le système Soleil–Terre par le prisme des données

All we ever see of stars are their old photographs.

(Alan Moore, Watchmen)

Sommaire

Préambule	38
2.1 Introduction à la modélisation par les données	39
2.1.1 Généralités sur la science dirigée par les données	39
2.1.2 Principes fondamentaux de l'apprentissage automatique	41
2.1.3 Du neurone formel à l'apprentissage profond	43
2.1.3.1 Le concept de neurone formel	43
2.1.3.2 Le perceptron multicouche	45
2.1.3.3 L'apprentissage par rétropropagation du gradient	47
2.1.3.4 Les fonctions d'activation	51
2.1.3.5 Les fonctions de perte	53
2.1.3.6 Bilan sur le perceptron multicouche	54
2.1.4 Les réseaux récurrents	55
2.1.5 Les réseaux convolutifs	59
2.1.6 La gestion des données en apprentissage profond	61
2.1.7 Autres aspects relatifs à l'implémentation et difficultés potentielles	64
2.1.8 Bilan sur la modélisation par les données et l'apprentissage profond	65
2.2 Les données à notre disposition	65
2.2.1 Observer le Soleil	65
2.2.1.1 Les missions d'observation du Soleil	66
2.2.1.2 Les autres sources de données	75
2.2.1.3 Les données dérivées de l'imagerie	76
2.2.1.4 Autres données utiles	76
2.2.1.5 Accéder aux données	76
2.2.1.6 Bilan sur l'imagerie solaire	77
2.2.2 Étudier le vent solaire	77
2.2.2.1 Les missions d'étude du vent solaire	77
2.2.2.2 L'accès aux données	78
2.2.3 Les indices géomagnétiques	79
2.2.4 Mesurer les particules au cœur des ceintures	82
2.2.5 Bilan sur l'observation du système Soleil–Terre	85
2.3 Modéliser les interactions Soleil–Terre	85
2.3.1 Des modèles pour la prédiction du vent solaire dans l'héliosphère	85
2.3.2 Prédiction d'indices géomagnétiques : des modèles empiriques au <i>Deep Learning</i>	87
2.3.3 Modélisation des ceintures de radiations	89
2.4 Bilan du chapitre	90

Préambule

Dans ce chapitre nous commençons par introduire les concepts essentiels de l'apprentissage automatique, et plus particulièrement de l'apprentissage profond, que nous avons utilisé au cours des nos travaux et qui seront employés dans les

prochains chapitres. Ensuite, comme notre étude se situe dans un contexte de modélisation dirigée par les données (*data-driven modeling*), nous nous intéressons aux données issues de l’observation du système Soleil–Terre. À cet effet nous porterons notre attention plus particulièrement sur les données issues de l’observation du Soleil et aux mesures *in situ* des propriétés locales du vent solaire ou encore de l’état de la magnétosphère terrestre. C’est pourquoi nous détaillons les différentes missions spatiales, instruments et bases de données dédiés à la collecte et à l’analyse de ces observations et nous nous penchons sur certaines de leurs caractéristiques. Finalement, afin de mieux situer nos travaux, nous passerons en revue diverses études portant sur la modélisation de différentes composantes des interactions Soleil–Terre. Nous y verrons notamment que parmi ces études, certaines des plus récentes d’entre elles montrent la pertinence d’une approche de modélisation dirigée par les données à des fins de météorologie de l’espace.

2.1 Introduction à la modélisation par les données

Dans cette section nous commençons par présenter le concept de science dirigée par les données, avant de nous pencher plus en détail sur l’apprentissage automatique en général, puis sur les réseaux de neurones en particulier, en présentant différentes architectures de réseaux de neurones que nous utilisons dans nos travaux.

Cette section n’a pas pour vocation d’être un cours complet sur la théorie et l’implémentation de méthodes relevant de l’apprentissage automatique, mais plutôt à aider le lecteur novice à comprendre, au moins qualitativement, les principes de fonctionnement des réseaux de neurones, les processus mis en jeu lors de l’apprentissage de ceux-ci, ainsi que les diverses considérations et difficultés qu’on peut rencontrer lorsqu’on cherche à mettre en place une étude s’appuyant sur ces méthodes – dont les enjeux diffèrent parfois des méthodes de modélisation empirique ou physique plus « classiques ». Le lecteur qui souhaiterait se plonger plus profondément dans les détails mathématiques et/ou relatifs à l’implémentation de ces méthodes est encouragé à se référer aux livres de référence qui seront mentionnés dans les paragraphes qui suivent.

L’utilisation de ces méthodes dans nos travaux sera justifiée en Section 2.3. Nous débutons ce chapitre par la présentation de ces méthodes car il nous semble important que le lecteur soit déjà conscient des techniques et enjeux de l’apprentissage profond lorsque nous décrirons les travaux ayant précédé les nôtres et qui nous ont menés à orienter nos recherches dans cette direction.

2.1.1 Généralités sur la science dirigée par les données

Historiquement, le premier paradigme scientifique fut la science expérimentale, où la connaissance est accumulée essentiellement par le biais d’expériences, par nature empiriques. À partir du XVII^e siècle, un nouveau paradigme a émergé, dans lequel la modélisation théorique des phénomènes, par exemple par des équa-

tions, devient prépondérante. Au XX^e siècle, avec l'émergence de moyens de calcul informatiques de plus en plus puissants, est apparu le troisième paradigme scientifique, à savoir la science computationnelle, qui s'appuie sur la résolution (ou l'approximation) numérique des modèles théoriques (KUHN et MEYER 1983).

Selon HEY et al. 2009, nous vivons actuellement l'apparition d'un quatrième paradigme scientifique, appelé la science dirigée par les données. Cette évolution repose sur un ensemble de faits concomitants :

- la démultiplication massive du nombre de données à notre disposition, qu'elles soient issues d'observations, d'expériences ou bien encore de modèles numériques ;
- l'augmentation des capacités de stockage locales, sur des outils de travail accessibles, permettant de travailler avec des bases de données de très grande taille ;
- l'augmentation de la puissance de calcul disponible, permettant d'effectuer des calculs utilisant ces bases de données en un temps de plus en plus restreint ;
- le développement d'une nouvelle discipline, l'apprentissage automatique, reposant sur un ensemble d'outils mathématiques et statistiques permettant à une machine d'apprendre à résoudre automatiquement des tâches – incluant la reproduction, voire la découverte, de phénomènes physiques.

Il ne faut pas confondre la science dirigée par les données, qui est un paradigme scientifique, avec la science des données. Cette dernière désigne un domaine interdisciplinaire, utilisant des outils mathématiques et informatiques pour fouiller, traiter, analyser, visualiser et extraire de l'information d'un ensemble de données (DHAR 2013). Ce domaine a connu un gain de popularité avec l'accroissement massif des bases de données (on parle alors souvent de *Big Data*). Ces techniques sont souvent utilisées conjointement aux méthodes d'apprentissage automatique.

Une des nouvelles approches ayant émergé au cours des dernières années, s'inscrivant au sein de ce nouveau paradigme est la modélisation dirigée par les données (MONTÁNS et al. 2019). Par opposition à la modélisation physique, où on cherche à modéliser un phénomène en le contraignant par le biais d'équations et de conditions limites, la modélisation par les données vise à résoudre un problème physique à l'aide d'algorithmes capables d'approcher – voire de découvrir –, souvent par le biais d'un apprentissage, des fonctions très complexes (WILLARD et al. 2021).

Puisque les quatre points mentionnés plus hauts sont particulièrement vrais en ce qui concerne l'étude des interactions Soleil–Terre, il apparaît qu'il s'agit d'un domaine d'application pertinent des méthodes de modélisation par les données. Pour nous en convaincre, dans ce chapitre (Section 2.2) nous décrivons une partie des données d'observation accumulées depuis des décennies, permettant une étude de l'héliosphère dirigée par les données (Section 2.3). Toutefois, avant cela, il nous faut exposer et expliquer les fondements de l'apprentissage automatique et des réseaux de neurones, qui sont les méthodes de modélisation par les données que nous utiliserons dans les chapitres suivants.

2.1.2 Principes fondamentaux de l'apprentissage automatique

L'apprentissage automatique, généralement plus connu sous son nom en langue anglaise *Machine Learning*, est un des sous-domaines de l'intelligence artificielle (IA). L'intelligence artificielle (IA) peut se définir comme « [l']ensemble des théories et des techniques développant des programmes informatiques complexes capables de simuler certains traits de l'intelligence humaine »¹. Au sein de l'IA, l'apprentissage automatique est constitué des techniques et méthodes visant à créer des programmes capables d'apprendre comment résoudre une tâche à partir de données (CHOLLET 2021).

L'apprentissage automatique est une discipline qui a émergé au milieu du XX^e siècle, peu de temps après les travaux de TURING 1950 et la conception de son fameux test. SAMUEL 1959 est le premier à utiliser le terme de *Machine Learning* pour qualifier son programme capable d'apprendre à jouer au jeu de dames. Peu de temps auparavant ROSENBLATT 1958 proposait le perceptron, la première architecture reposant sur le neurone formel artificiel – qui avait lui été théorisé dès 1943 par MCCULLOCH et PITTS –, ancêtre de tous les réseaux de neurones. Depuis, l'apprentissage automatique a connu plusieurs vagues d'engouement successives, entrecoupées de périodes de moindre popularité. On peut estimer le début du troisième et actuel âge d'or de l'apprentissage automatique, centré principalement autour des réseaux de neurones, par la publication par KRIZHEVSKY et al. 2012 d'un modèle construit autour d'un réseau de neurones convolutifs ou *Convolutional Neural Network* (CNN) qui a remporté le concours de classification d'images *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) (RUSSAKOVSKY et al. 2015), dépassant de loin les performances des modèles concurrents.

Au sein de l'apprentissage automatique, on distingue plusieurs grandes catégories d'algorithmes d'apprentissage, schématisées en Figure 2.1. Tout d'abord on a la classe des algorithmes d'apprentissage supervisé, c'est-à-dire les algorithmes étant capables d'apprendre comment résoudre une tâche à partir d'exemples précis préalablement sélectionnés et labélisés par l'utilisateur. Parmi les problèmes d'apprentissage supervisé on distingue deux classes de problèmes : les problèmes de régression et les problèmes de classification. Les premiers consistent à estimer une variable quantitative, alors que les seconds reviennent à estimer une variable qualitative. Parmi les algorithmes d'apprentissage les plus utilisés on peut citer :

- les différentes variantes de la régression linéaire (moindres carrés, régression LASSO (KIM et al. 2007), régression RIDGE (RIFKIN et LIPPERT 2007), etc.) ;
- la régression logistique (BERKSON 1944) ;
- l'algorithme de *Machine à Vecteurs de Support* (SVM) (SMOLA et SCHÖLKOPF 2004) ;
- l'algorithme de *Random Forest* (RF) (LOUPPE 2014) ;
- l'algorithme de *k-Nearest Neighbors* (*k*-NN) (GOLDBERGER et al. 2005) ;
- les Processus Gaussiens (WILLIAMS et RASMUSSEN 2006) ;

1. Définition du dictionnaire Le Robert.

- les réseaux de neurones ;
- etc.

Il existe également la classe des algorithmes d'apprentissage non-supervisé. Dans ce cas l'algorithme est chargé de détecter des structures sous-jacentes à un ensemble de données non-étiquetées. Parmi les applications de l'apprentissage supervisé on peut citer le regroupement (*clustering*), c'est-à-dire une forme de classification pour laquelle les classes n'ont pas été formellement identifiées au préalable ; ou encore la réduction de données, qui consiste à obtenir une représentation de nos données sous une autre forme, le plus souvent de dimension inférieure. Parmi les algorithmes d'apprentissage non-supervisé couramment utilisés on trouve :

- les modèles de mélange (MCLACHLAN et al. 2019) ;
- l'algorithme de *k*-moyennes (HARTIGAN et WONG 1979) ;
- l'algorithme de analyse en composantes principales ou *Principal Component Analysis* (PCA) (PEARSON 1901) ;
- etc.

La troisième grande classe des algorithmes d'apprentissage automatique est l'apprentissage par renforcement. L'apprentissage par renforcement consiste à apprendre les décisions à prendre dans un contexte donné à partir d'expériences passées, à l'aide d'un système de récompenses et de pénalités (SUTTON 1988). L'apprentissage par renforcement a donné lieu à certaines des avancées les plus médiatiques dans le domaine de l'apprentissage automatique, comme par exemple le désormais célèbre programme AlphaGo, capable de battre les plus grands champions du jeu de Go (SILVER et al. 2016). Depuis, ce programme s'est encore amélioré en s'entraînant uniquement contre lui-même, si bien que sa nouvelle version, AlphaGo Zéro, a largement battu AlphaGo premier du nom (SILVER et al. 2017).

Enfin, certains considèrent également l'apprentissage par transfert comme une branche à part entière de l'apprentissage automatique (PAN et YANG 2010). L'apprentissage par transfert consiste à transférer des connaissances acquises lors d'un apprentissage préalable (qu'il soit supervisé, non-supervisé ou par renforcement) vers une nouvelle tâche cible, parfois pour des applications différentes.

Dans nos travaux nous avons principalement utilisé des techniques relevant de l'apprentissage supervisé, pour résoudre des problèmes de régression mais aussi de classification. Dans les derniers chapitres nous avons également eu recours à des procédés d'apprentissage par transfert, qui y seront alors expliqués et détaillés.

Les algorithmes mentionnés tout au long de cette section ont déjà montré leur efficacité pour traiter de nombreux problèmes, y compris dans le domaine de la météorologie de l'espace (CAMPOREALE et al. 2018). Parmi ces méthodes, les réseaux de neurones ont démontré une très bonne capacité à traiter les problèmes de haute dimension pour lesquels nous possédons beaucoup d'échantillons de données, ce qui est souvent le cas en météorologie de l'espace, et reste vrai dans le contexte de nos études. Certaines études comme LAZZÚS et al. 2017 ; WINTOFT et al. 2017 ;

GRUET 2018 ; CHAKRABORTY et MORLEY 2020 ayant démontré l'excellence des réseaux de neurones à traiter les problèmes de prédiction d'indices géomagnétiques, c'est presque exclusivement à cette classe d'algorithmes et de méthodes que nous ferons appel tout au long de ce manuscrit. C'est pourquoi dans les sections qui suivent nous expliquons plus en détail le fonctionnement général d'un réseau de neurones et comment programmer son apprentissage, avant de passer en revue différentes architectures que nous utiliserons dans les chapitres qui suivent.

Nous renvoyons le lecteur souhaitant approfondir ses connaissances sur les algorithmes d'apprentissage automatique ne relevant pas des réseaux de neurones vers le livre de référence de GÉRON 2019, dans lequel il décrit de nombreuses méthodes d'apprentissage supervisé et non-supervisé et explique comment les implémenter en langage Python à l'aide de la librairie de référence `scikit-learn` (PEDREGOSA et al. 2011). En ce qui concerne l'apprentissage par renforcement, nous suggérons au lecteur de se référer à SUTTON et BARTO 2018.

2.1.3 Du neurone formel à l'apprentissage profond

Pour construire cette section nous nous sommes en partie appuyés sur des livres de référence incluant GOODFELLOW et al. 2016 ; BUDUMA et LOCASCIO 2017 ; CHOLLET 2021. Pour plus de détails concernant ce qui suit, le lecteur est invité à se référer à ces livres.

2.1.3.1 Le concept de neurone formel

En s'inspirant du fonctionnement du cerveau humain, constitué de neurones connectés, MCCULLOCH et PITTS 1943 ont été les premiers à proposer un modèle mathématique visant à imiter le fonctionnement d'un neurone biologique : le neurone formel. Un neurone biologique peut être schématisé assez simplement, comment le montre la Figure 2.2. Le neurone reçoit un signal par ses dendrites. Son noyau traite ensuite cette information et la transforme. L'axone transporte le nouveau signal, qui sera transmis à un autre neurone au niveau des synapses, qui sont les points de connexion entre l'arborisation terminale du premier neurone et les dendrites du second.

L'idée du neurone formel est de modéliser les principales parties du neurone biologique sous la forme de fonctions mathématiques, comme le montre la Figure 2.3. Ici, les dendrites correspondraient aux entrées $\mathbf{x} = (x_1, \dots, x_m)$. Ces entrées sont pondérées par des poids $\mathbf{w} = (w_1, \dots, w_m)$, qui confèrent une importance relative à chacune des entrées. Ces entrées pondérées sont sommées, avec un éventuel biais w_0 supplémentaire. Cette somme est ensuite transformée par une fonction dite d'activation σ . Ces transformations correspondraient au rôle du noyau. Enfin, le neurone possède comme sortie \hat{y} , qui représente un nouveau signal (qui serait transmis par l'arborisation terminale dans le cas du neurone biologique). Ce nouveau signal pourra lui-même servir d'entrée à un autre neurone, formant ainsi un *réseau de neurones artificiel* ou *Artificial Neural Network (ANN)*.

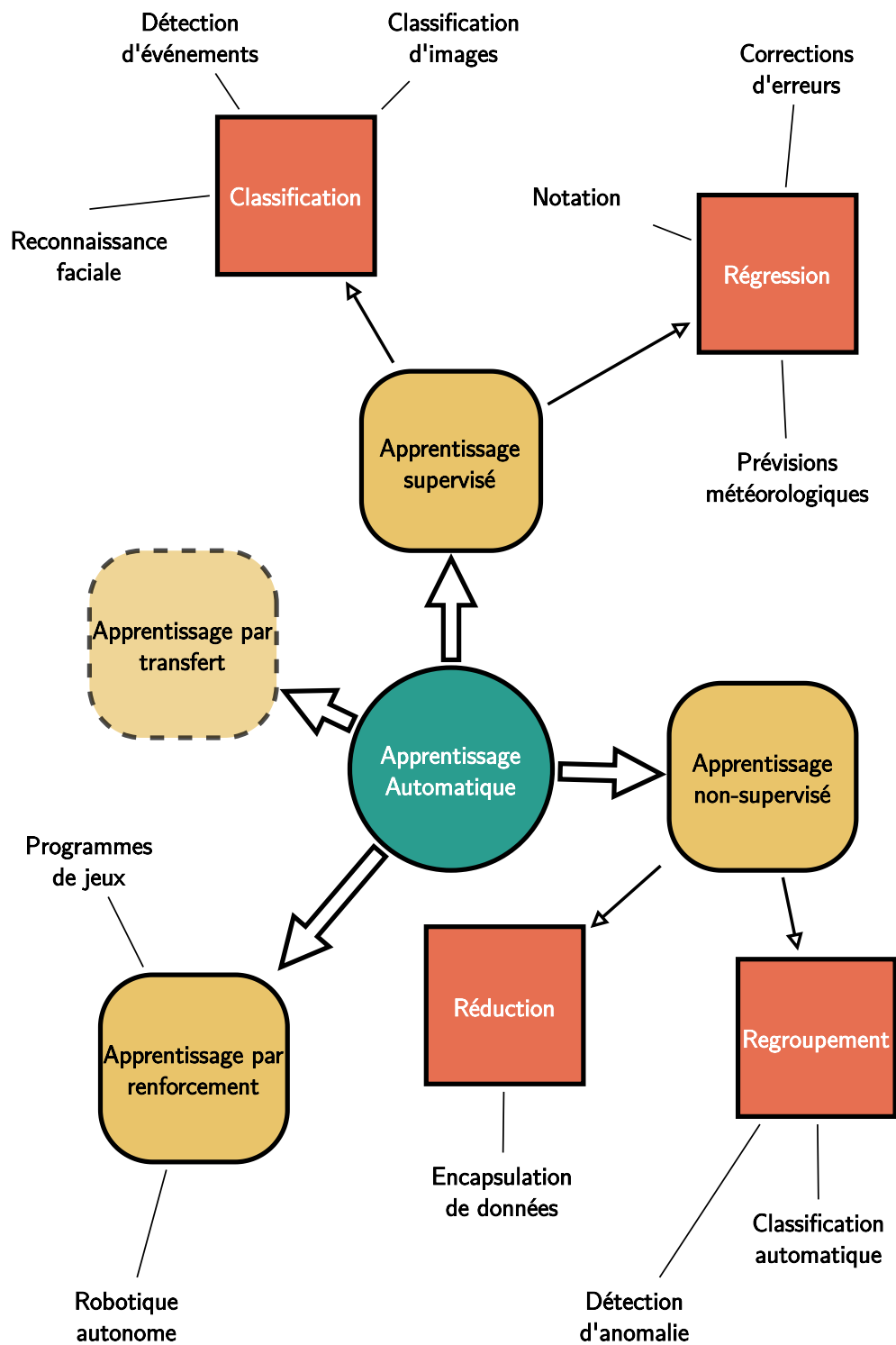


FIGURE 2.1 – Schéma des différentes branches de l'apprentissage automatique, avec des exemples d'application.

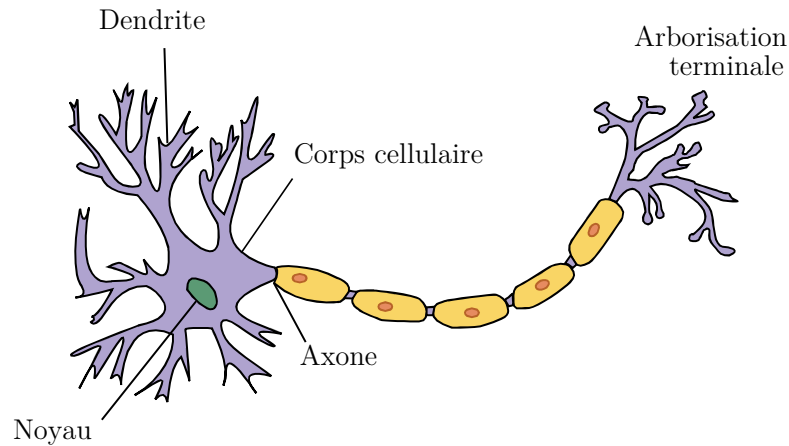


FIGURE 2.2 – Schéma d'un neurone biologique. Adapté de Wikimedia Commons, sous licence CC BY-SA 3.0.

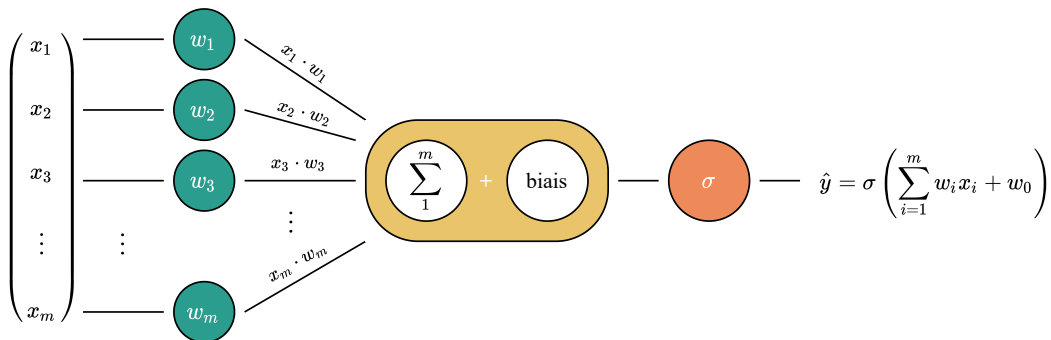


FIGURE 2.3 – Schéma d'un neurone formel.

2.1.3.2 Le perceptron multicouche

À partir du neurone formel, ROSENBLATT 1958 propose le perceptron. Le perceptron est un classifieur linéaire simple, construit à partir d'un neurone formel dont la fonction d'activation σ est la fonction de Heaviside², ou bien la fonction \tanh . Le perceptron permet d'associer à une entrée, représentée par un vecteur \mathbf{x} , un label de classe binaire (typiquement, 0 ou 1). Afin que le classifieur soit performant (*i. e.* qu'il classe correctement les entrées qui lui sont données), il faut que les poids \mathbf{w} du réseau soient adaptés au problème. Cela peut se faire à l'aide d'une base de données d'apprentissage, c'est-à-dire un ensemble de données préalablement labélisées, servant d'exemples pour le perceptron. Le perceptron peut alors « apprendre » (mettre à jour) la valeur de ses poids \mathbf{w} en suivant un algorithme. L'un des premiers algorithmes d'apprentissage utilisés était le suivant :

$$\mathbf{w}' = \mathbf{w} + \alpha (y - \hat{y}) \mathbf{x}, \quad (2.1)$$

où \mathbf{w}' désigne le vecteur des poids du perceptron actualisés, y la véritable classe binaire correspondant aux entrées \mathbf{x} et α est un coefficient appelé taux d'apprentissage (fixé par l'utilisateur). Notons que dans ce modèle le biais (seuil) w_0 est

2. Aussi appelée fonction échelon unitaire.

également fixé par l'utilisateur.

En connectant plusieurs perceptrons (afin d'imiter le fonctionnement d'un cerveau biologique) on peut alors obtenir un **perceptron multicouche** ou *Multi-Layer Perceptron (MLP)*, schématisé en Figure 2.4. Dans cet exemple, nos entrées \mathbf{x} sont de dimension $l_0 = 3$, et nos sorties sont de dimension $l_3 = 2$. Chaque cercle représente un neurone. On définit une couche l_k comme l'ensemble des p_k neurones ayant les mêmes entrées, mais possédant chacun un ensemble de poids \mathbf{w}_i^k , $i \in [1, p_k]$, distincts. Dans notre exemple, la couche l_1 possède $p_1 = 8$ neurones et la couche l_2 possède $p_2 = 6$ neurones. Une des caractéristiques principales d'un réseau de neurones est l'organisation de son graphe, représentant les connexions entre les différents neurones.

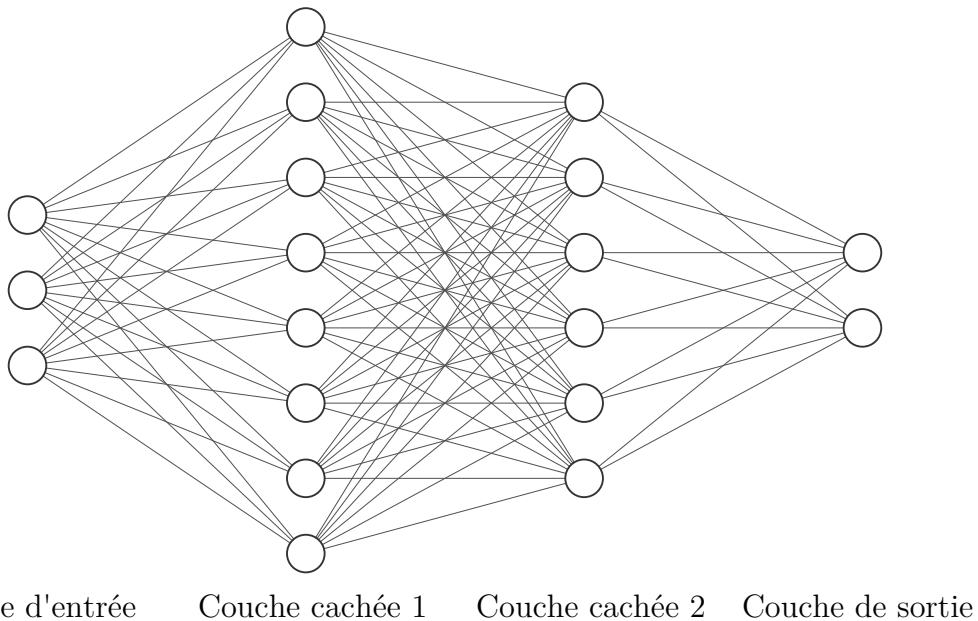


FIGURE 2.4 – Exemple de **MLP** possédant quatre couches, dont deux couches cachées.

Dans notre exemple, les sorties des neurones d'une couche ne sont pas connectées à l'entrée des neurones de couches précédentes : on parle alors de réseau de type *feedforward* (réseau à propagation avant). De plus, tous les neurones appartenant à une même couche sont connectés à l'ensemble des neurones de la couche suivante, on parle donc de réseau *fully-connected* (entièrement connecté). Ce type d'architecture de réseau est très classique.

Les couches situées entre la couche d'entrée et la couche de sortie sont appelées couches cachées. Plus un réseau possède de couches cachées, plus il est profond : c'est de là que provient le terme *Deep Learning* (apprentissage profond), désignant l'ensemble des méthodes d'apprentissage automatique reposant sur l'utilisation de réseaux de neurones profonds. Notons qu'il n'existe pas de définition officielle du nombre de couches nécessaires pour qu'un réseau de neurones puisse être qualifié de profond, bien qu'un usage fréquent soit de considérer un réseau profond dès

que son nombre de couches cachées est supérieur ou égal à trois (HINTON et al. 2006; SCHMIDHUBER 2015).

2.1.3.3 L'apprentissage par rétropropagation du gradient

Bien que le MLP ait été théorisé assez tôt, les approches reposant sur les réseaux de neurones n'ont pas été immédiatement et massivement adoptées, faute d'une méthode d'apprentissage efficace, permettant au réseau de mettre à jour l'ensemble de ses paramètres de manière efficace.

Cela a changé lorsque RUMELHART et al. 1986 ont formalisé l'algorithme de rétropropagation du gradient – qui avait été redécouvert au cours des années 70 et 80 par différents chercheurs en parallèle. L'algorithme de rétropropagation du gradient (*gradient backpropagation*) est aujourd'hui une des pierres angulaires de l'apprentissage profond et est certainement à l'origine d'une grande partie de l'engouement porté aux réseaux de neurones. C'est grâce à cet algorithme que LECUN et al. 1989 proposa un modèle alors révolutionnaire de reconnaissance automatique des chiffres manuscrits, ce qui fut l'un des premiers grands succès de l'apprentissage profond (CHOLLET 2021).

Le concept de l'algorithme de rétropropagation du gradient peut s'expliquer assez simplement. Prenons un réseau de neurones dédié à une certaine tâche (régression ou classification). L'idée de l'algorithme de rétropropagation est d'estimer l'erreur commise par le réseau, puis d'adapter les paramètres de l'ensemble des neurones afin de minimiser cette erreur, à l'aide d'un algorithme d'optimisation de type « descente du gradient » adapté à la structure en couches successives d'un réseau de neurones. Une fois les paramètres du réseau mis à jour, on peut réitérer le processus jusqu'à convergence vers un modèle optimal.

Nous pouvons formaliser tout cela. Considérons un MLP à n couches (qui possède donc $n - 2$ couches cachées). Pour $k = 1, \dots, n$, on note :

- l_k la $k^{\text{ième}}$ couche du réseau ;
- p_k le nombre de neurones dans la couche l_k ;
- $\mathbf{w}_i^k = (w_{1,i}^k, \dots, w_{p_{k-1},i}^k)$ le vecteur des poids du $i^{\text{ième}}$ neurone de la couche l_k ;
- $w_{0,i}^k$ le biais du $i^{\text{ième}}$ neurone de la couche l_k ;
- Θ l'ensemble de tous les poids et biais du réseau ;
- o_i^k la sortie du $i^{\text{ième}}$ neurone de la couche l_k ;
- $\mathbf{x} = \mathbf{o}^0 = (o_1^0, \dots, o_{p_0}^0)$ les entrées du modèle ;
- $\hat{\mathbf{y}} = \mathbf{o}^n = (o_1^n, \dots, o_{p_n}^n)$ les sorties du modèle ;
- σ la fonction d'activation utilisée, supposée dérivable³ ;

3. Ce qui exclut la fonction de Heaviside et justifie l'utilisation d'une fonction comme tanh. Notons également qu'on peut utiliser conjointement plusieurs fonctions d'activation différentes dans un même réseau sans que cela ne change la démonstration qui suit.

- $a_i^k = \mathbf{w}_i^k \cdot \boldsymbol{o}^{k-1} + w_{0,i}^k$ la sortie du neurone avant sa transformation par la fonction d'activation – on a donc $o_i^k = \sigma_i^k(a_i^k)$.

Notons également \mathbf{x}_j une entrée (vectorielle) de notre modèle, associée à une sortie « vraie » \mathbf{y}_j (qui peut être un label, une classe, un ensemble de valeurs continues, etc.). Notons N le nombre d'échantillons dans notre base de données d'apprentissage. Notons cette dernière $D = \{(\mathbf{x}_j, \mathbf{y}_j)_{j=1,\dots,N}\}$. Notons enfin $\mathbf{X} = \{\mathbf{x}_j\}_{j=1,\dots,N}$, $\mathbf{Y} = \{\mathbf{y}_j\}_{j=1,\dots,N}$ et $\hat{\mathbf{Y}} = \{\hat{\mathbf{y}}_j\}_{j=1,\dots,N}$.

Définissons $E_{\Theta}(\mathbf{Y}, \hat{\mathbf{Y}}) = \sum_{j=1}^N E_{\Theta}(\mathbf{y}_j, \hat{\mathbf{y}}_j)$ l'erreur commise par notre modèle sur la base de données d'apprentissage. $E_{\Theta}(\mathbf{Y}, \hat{\mathbf{Y}})$ est d'une certaine manière une mesure de l'écart entre les $\hat{\mathbf{y}}_j$ et les \mathbf{y}_j (nous verrons plus bas des exemples de fonctions pouvant servir à quantifier l'erreur E), dépendant des entrées \mathbf{x}_j (via les $\hat{\mathbf{y}}_j$), des « vraies » sorties \mathbf{y}_j et des paramètres Θ du modèle. De plus, nous supposons que E_{Θ} est dérivable par rapport aux \mathbf{y}_j et que nous connaissons analytiquement sa dérivée.

Comme nous l'avons dit plus haut, entraîner le modèle revient à essayer de minimiser $E_{\Theta}(\mathbf{Y}, \hat{\mathbf{Y}})$. Pour cela nous allons utiliser la méthode classique de descente du gradient⁴ (CAUCHY 1847) et chercher à mettre à jour Θ tel que :

$$\Theta^{t+1} = \Theta^t - \alpha^t \cdot \nabla_{\Theta^t} E_{\Theta^t}(\mathbf{Y}, \hat{\mathbf{Y}}), \quad (2.2)$$

où :

- Θ^t désigne l'ensemble des paramètres du modèle lors de la $t^{\text{ième}}$ itération de l'algorithme d'apprentissage ;
- α^t désigne le taux d'apprentissage ;
- $\nabla_{\Theta^t} E_{\Theta^t}(\mathbf{Y}, \hat{\mathbf{Y}})$ est le gradient de $E_{\Theta^t}(\mathbf{Y}, \hat{\mathbf{Y}})$ par rapport à Θ^t . Dans la suite nous l'abrégeons $\nabla_{\Theta^t} E$.

La difficulté réside donc dans le calcul de $\nabla_{\Theta^t} E$, *i. e.* dans le calcul de tous les $\frac{\partial E}{\partial w_{q,i}^k}$. Commençons par noter que, par composition, on a :

$$\frac{\partial E}{\partial w_{q,i}^k} = \frac{\partial E}{\partial a_i^k} \frac{\partial a_i^k}{\partial w_{q,i}^k}. \quad (2.3)$$

4. Algorithme d'optimisation différentiable, où l'on cherche à minimiser une fonction en se déplaçant dans l'espace de ses paramètres par itérations successives. Le gradient de la fonction détermine la direction du déplacement, tandis que le taux d'apprentissage détermine la longueur de celui-ci. Si α est trop petit, alors la convergence sera lente, et il y a un risque de convergence vers un minimum local. Si α est trop grand, alors il y a un risque d'osciller autour du minimum et de ne jamais converger, voire diverger.

Or comme $a_i^k = \mathbf{w}_i^k \cdot \mathbf{o}^{k-1} + w_{0,i}^k$, on a donc :

$$\begin{aligned} \frac{\partial a_i^k}{\partial w_{q,i}^k} &= \frac{\partial}{\partial w_{q,i}^k} (\mathbf{w}_i^k \cdot \mathbf{o}^{k-1} + w_{0,i}^k) \\ &= \frac{\partial}{\partial w_{q,i}^k} (\mathbf{w}_i^k \cdot \mathbf{o}^{k-1}) \\ &= \frac{\partial}{\partial w_{q,i}^k} \left(\sum_{j=1}^{p_{k-1}} w_{j,i}^k o_j^{k-1} \right) = o_i^{k-1}. \end{aligned} \quad (2.4)$$

Ainsi :

$$\frac{\partial E}{\partial w_{q,i}^k} = \frac{\partial E}{\partial a_i^k} o_i^{k-1} \stackrel{\text{def}}{=} \delta_i^k o_i^{k-1}. \quad (2.5)$$

Comme o_i^{k-1} est connu, il nous reste à calculer δ_i^k . L'astuce de l'algorithme de rétropropagation est de montrer qu'il est possible de calculer un δ_i^k d'une couche k à partir de l'ensemble des $\delta_{j=1, \dots, p_{k+1}}^{k+1}$ de la couche $k+1$.

En effet, partons de la dernière couche. On a :

$$\begin{aligned} \delta_i^n &= \frac{\partial E}{\partial a_i^n} = \frac{\partial E}{\partial o_i^n} \frac{\partial o_i^n}{\partial a_i^n} \\ &= \frac{\partial E}{\partial o_i^n} \frac{\partial \sigma(a_i^n)}{\partial a_i^n} = \frac{\partial E}{\partial o_i^n} \sigma'(a_i^n). \end{aligned} \quad (2.6)$$

Or $\frac{\partial E}{\partial o_i^n}$ est connu par hypothèse, qui, dans les cas concrets, est facilement vérifiée en choisissant bien E . De plus $\sigma'(a_i^n)$ est également connu. Ainsi, $\frac{\partial E}{\partial w_{q,i}^n}$ est connu pour tous les poids de tous les neurones de la dernière couche.

Prenons maintenant une couche l_k quelconque, qui ne soit pas la dernière couche ($k < n$). On a, à nouveau par dérivée de fonctions composées :

$$\begin{aligned} \delta_i^k &= \frac{\partial E}{\partial a_i^k} = \sum_{j=1}^{p_{k+1}} \frac{\partial E}{\partial a_j^{k+1}} \frac{\partial a_j^{k+1}}{\partial a_i^k} \\ &= \sum_{j=1}^{p_{k+1}} \delta_j^{k+1} \frac{\partial a_j^{k+1}}{\partial a_i^k}. \end{aligned} \quad (2.7)$$

Par définition, $a_j^{k+1} = \sum_{l=1}^{p_k} w_{l,j}^{k+1} o_l^k = \sum_{l=1}^{p_k} w_{l,j}^{k+1} \sigma(a_l^k)$. Donc :

$$\delta_i^k = \sum_{j=1}^{p_{k+1}} \delta_j^{k+1} \frac{\partial a_j^{k+1}}{\partial a_i^k} = \sum_{j=1}^{p_{k+1}} \delta_j^{k+1} w_{i,j}^{k+1} \sigma'(a_i^k). \quad (2.8)$$

Enfin :

$$\frac{\partial E}{\partial w_{q,i}^k} = \delta_i^k o_i^{k-1} = \sigma'(a_i^k) o_i^{k-1} \sum_{j=1}^{p_{k+1}} \delta_j^{k+1} w_{i,j}^{k+1} \quad (2.9)$$

On remarque alors que, par récurrence en partant de la dernière couche, tous les termes de cette somme sont connus. Ainsi, on peut calculer $\nabla_{\Theta^t} E$ en calculant tous les $\frac{\partial E}{\partial w_{q,i}^k}$ en partant de la dernière couche : d'où le nom de l'algorithme de rétropropagation du gradient. Par cette méthode, le **MLP** va minimiser automatiquement son erreur par itérations successives de la formule 2.2, et ce faisant, apprendre le lien entre les entrées \mathbf{X} et les sorties \mathbf{Y} . Dans la littérature de l'apprentissage profond, chaque itération d'apprentissage est appelée une *epoch*. Généralement, l'erreur commise par le modèle à chaque *epoch* est surveillée et reportée dans un graphe, ce qui nous indique – parfois en temps réel – si le modèle est en train d'apprendre (si l'erreur diminue), ou non.

Grâce à cette méthode nous pouvons utiliser des réseaux très profonds, et donc apprendre des tâches très complexes. De nos jours cette méthode est implémentée directement dans la plupart des libraires permettant de créer des réseaux de neurones, à l'aide de mécanismes de calcul automatique des gradients et de rétention du graphe reliant ceux-ci (voir par exemple PASZKE et al. 2017). Notons également que le plus souvent les poids du réseau de neurones ne sont aujourd'hui plus actualisés en suivant l'algorithme de descente du gradient classique, mais plutôt l'algorithme de descente du gradient stochastique (KIEFER et WOLFOWITZ 1952) ou une de ses variantes. De plus, afin de pouvoir utiliser des jeux de données de plus en plus grands, on procède à un apprentissage par *mini-batches*. Cela signifie que nos données d'entraînement sont découpées en sous-ensemble de petite taille (appelés *mini-batches*, souvent de quelques dizaines à quelques centaines d'éléments) qu'on donne successivement au réseau, les poids de celui-ci étant mis-à-jour après le passage de chaque *mini-batch*. Cette méthode permet non seulement d'utiliser de grands jeux de données (qui ne tiennent entièrement pas dans la mémoire d'une machine), mais aussi parfois d'accélérer l'apprentissage. Toutefois, la taille idéale d'un *mini-batch* dépend de chaque problème et est un paramètre supplémentaire à optimiser (BENGIO 2012; MASTERS et LUSCHI 2018).

Maintenant que nous avons montré les fondements de l'analyse profonde, constitués d'un côté par l'architecture du **MLP** et de l'autre par l'algorithme de rétropropagation du gradient, nous allons nous intéresser à certaines façons différentes d'implémenter ces algorithmes, en passant en revue divers types de fonctions d'activation et d'erreur couramment utilisées.

2.1.3.4 Les fonctions d'activation

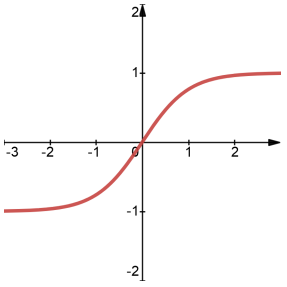
La (ou les) fonction(s) d'activation d'un réseau de neurones est très importante, pour deux principales raisons. La première est qu'utiliser des fonctions d'activation introduit des non-linéarités dans le réseau. Celles-ci sont essentielles. En effet, sans fonction d'activation un réseau de neurones n'est qu'une combinaison de transformations linéaires. Or une combinaison de transformations linéaires est également une transformation linéaire. Ainsi, sans fonction d'activation on perd tout l'intérêt d'avoir une structure en couches multiples.

De surcroît, l'introduction de non-linéarités via les fonctions d'activations est une des hypothèses du théorème d'approximation universelle, selon lequel un MLP avec au moins une couche cachée et un nombre fini de neurones peut approximer n'importe quelle fonction continue sur des sous-ensembles compacts de \mathbb{R}^n (HORNİK et al. 1989 ; HORNİK 1991). Ce théorème, très puissant, et qui a connu beaucoup de raffinements (LU et al. 2017 ; KIDGER et LYONS 2020), est l'un des principaux arguments théoriques justifiant l'engouement autour des réseaux de neurones, venant s'ajouter aux réussites empiriques publiées dans la littérature.

La deuxième raison justifiant l'importance des fonctions d'activation est d'ordre numérique. En effet, selon la fonction d'activation choisie, on peut contraindre les sorties de différents neurones dans certains intervalles, soit pour des raisons d'interprétabilité (par exemple contraindre la sortie du modèle à prendre des valeurs entre 0 et 1 pour que cela corresponde à une probabilité), ou bien pour mitiger des problèmes d'instabilité numérique, comme le problème de l'évanescence du gradient (voir section 2.1.4).

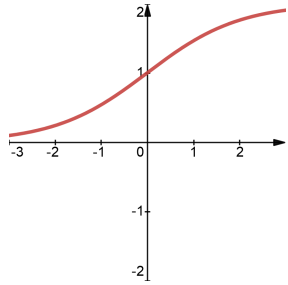
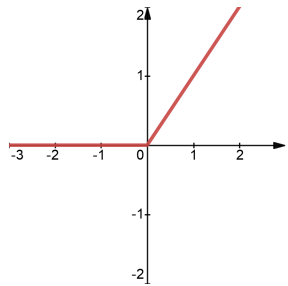
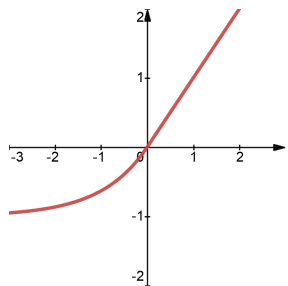
Dans le Tableau 2.1 nous listons certaines des fonctions d'activations usuelles, en donnant leur formule et justifiant de leur rôle et propriétés particulières éventuelles.

TABLEAU 2.1 – Quelques fonctions d'activation usuelles.

Nom	Formule	Graphe	Caractéristiques
tanh	$f(x) = \frac{2}{1 + e^{-2x}} - 1$		Utilisée dans les problèmes de régression, où les données ont été normalisées entre -1 et 1.

Continue sur la prochaine page

TABLEAU 2.1 – Quelques fonctions d’activation usuelles. (Cont.)

Nom	Formule	Graphes	Caractéristiques
sigmoïde	$f(x) = \frac{2}{1 + e^{-x}}$		Utilisée dans les problèmes de classification binaire, pour obtenir une probabilité d’appartenir à une classe.
ReLU	$f(x) = \max(0, x)$		Fonction la plus simple permettant d’introduire une non-linéarité. Permet d’accélérer l’apprentissage. A montré empiriquement son efficacité dans de nombreuses études.
ELU	$f(x) = \begin{cases} \alpha(e^x - 1) & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$		Fonction se rapprochant de ReLU, mais pouvant avoir des valeurs négatives. La fonction ne sature donc pas à 0, ce qui permet de conserver plus d’information, améliorant ainsi le caractère généralisable du modèle (CLEVERT et al. 2016).

Le choix des fonctions d’activation à utiliser dans un modèle est souvent empirique. Lorsque plusieurs fonctions d’activation sont possibles pour résoudre un même problème, il n’y a le plus souvent pas de règle théorique permettant d’affirmer qu’une fonction d’activation fonctionnera mieux qu’une autre. Empiriquement, de nombreuses études ont tendance à montrer que certaines fonctions d’activation fonctionnent mieux pour certaines applications, mais l’utilisateur est le plus souvent amené à en essayer plusieurs. Le choix des fonctions d’activation à utiliser fait

partie des nombreux choix que le concepteur d'un réseau de neurones doit faire. Nous en verrons d'autres dans les sections suivantes, et plus particulièrement en Section 2.1.7.

2.1.3.5 Les fonctions de perte

Le nom de « fonction de perte » (*loss function* en anglais) est souvent donné à la fonction représentant l'erreur E à minimiser lors de l'apprentissage d'un modèle, comme présenté précédemment. Cette fonction, mesurant l'écart entre les sorties du modèle et les vraies observations, doit être dérivable et sa dérivée doit être connue analytiquement.

Dans les faits, c'est un nombre plutôt restreint de fonctions de perte qui est le plus souvent utilisé dans la littérature. Dans le Tableau 2.2 nous en décrivons certaines parmi les plus utilisées dans différents contextes d'apprentissage supervisé.

TABLEAU 2.2 – Quelques fonctions de perte usuelles.

Nom	Formule	Problème	Caractéristiques
MAE	$f(y, \hat{y}) = y - \hat{y} $	Régression	Fonction de perte la plus simple pour le problème de régression. Mesure simplement l'écart absolu entre une prédiction et une observation.
MSE	$f(y, \hat{y}) = (y - \hat{y})^2$	Régression	Similaire à la MAE, mais pénalise plus les grosses erreurs (et donc, à proportions égales, les erreurs commises sur les valeurs cibles les plus élevées).

Continue sur la prochaine page

TABLEAU 2.2 – Quelques fonctions de perte usuelles. (*Cont.*)

Nom	Formule	Problème	Caractéristiques
GNLL	$f(y, \hat{\mu}, \hat{\sigma}) = \frac{1}{2} * (\log(\hat{\sigma}^2) + \frac{(\hat{\mu}-y)^2}{\hat{\sigma}^2})$	Régression probabiliste	Permet d'entraîner un réseau effectuant des prédictions probabilistes sous la forme d'une loi gaussienne (<i>i. e.</i> au lieu d'estimer une valeur \hat{y} , on estime une distribution de probabilités suivant une loi normale).
Cross-Entropy	$f(\mathbf{y}, \hat{\mathbf{p}}) = -\sum_{i=1}^{N_c} y_i \log(\hat{p}_i)$	Classification	Utile pour entraîner un classifieur à $N_c \geq 2$ classes.

Notes : Dans le cas d'une régression probabiliste sous la forme d'une loi normale, $\hat{\mu}$ désigne la moyenne de la loi normale estimée, et $\hat{\sigma}$ son écart-type. Dans le cas d'un problème de classification multi-classes, $N_c \geq 2$ désigne le nombre de classes et \hat{p}_i désigne la probabilité estimée par le modèle d'appartenir à la classe $i \leq N_c$.

Bien que le choix de la fonction de perte soit primordial – car de lui va dépendre directement la façon d'apprendre du modèle –, il est dans la plupart des cas conditionné par le problème que doit résoudre le réseau de neurones et, dans la majorité des cas, les possibilités sont assez limitées.

2.1.3.6 Bilan sur le perceptron multicouche

Dans cette section nous avons donné un aperçu du fonctionnement et de l'intérêt du MLP, qui est l'architecture de réseaux de neurones la plus classique. Nous avons identifié ses constituants fondamentaux, que sont le neurone formel, les fonctions d'activation, la fonction de perte, et l'algorithme de rétropropagation du gradient – clef de voûte de l'apprentissage d'un réseau de neurones.

Dans les sections suivantes nous allons explorer deux autres types d'architectures de réseaux de neurones, développés afin de mieux tenir compte de la spécificité de certains types de données et qui seront utilisés dans les prochains chapitres de ce manuscrit.

2.1.4 Les réseaux récurrents

Le **MLP** est une architecture éprouvée, mais elle n'est pas sans défauts. L'un d'entre eux est le fait qu'il n'est pas possible pour un **MLP** de prendre en compte explicitement la structure séquentielle de données comme des séries temporelles ou du langage naturel. Or, dans le cas de séries temporelles, une partie de l'information est souvent contenue dans la dynamique (actuelle et passée) du signal considéré. Il semble donc primordial de pouvoir prendre en compte de façon explicite le caractère séquentiel des données si on veut, par exemple, prédire les valeurs futures d'une séquence. De plus, utiliser des séquences de données en entrée d'un **MLP** nécessite l'utilisation de réseaux encore plus grands, ce qui peut poser un problème si on n'a pas assez de données pour l'entraîner.

Afin de répondre à ce problème on peut utiliser des modèles appartenant à la classe des **RNNs**. L'idée principale du **RNN** est de posséder des cellules qui prennent en compte de façon successive les données d'une séquence. Pour accomplir cela, chaque couche d'un **RNN** possède une couche de rétroaction à sa sortie, afin qu'une sortie du réseau dépende de ses entrées, mais aussi de ce qu'il a déjà vu. On introduit ainsi dans le réseau une mémoire, sous la forme d'un état caché \mathbf{h} .

Ce fonctionnement peut être schématisé de deux façons : avec une vue « repliée » (Figure 2.5a), ou bien « dépliée » (Figure 2.5b), où les connexions entre les différents pas de temps sont explicitées. Ici l'entrée $\mathbf{x} = (\mathbf{x}_t)_{0 \leq t \leq T}$, peut représenter les valeurs prises par un signal physique (multi-dimensionnel) à $t + 1$ instants successifs. Le modèle est composé d'une couche cachée, d'un état $\mathbf{h} = (\mathbf{h}_t)_{0 \leq t \leq T}$ et possède une sortie $\hat{\mathbf{y}} = (\hat{\mathbf{y}}_t)_{0 \leq t \leq T}$. Les paramètres (apprenables) du modèle sont \mathbf{W}_{xh} , \mathbf{W}_{hh} et \mathbf{W}_{hy} . Le **RNN** possède également une fonction d'activation σ (souvent \tanh) – non représentée sur le schéma – et peut également comporter des biais \mathbf{w}_0 .

Plus formellement, le **RNN** est, dans le cas général, régi par des équations de la forme :

$$\mathbf{h}_t = F_{\mathbf{W}_{wh}, \mathbf{W}_{hh}}(\mathbf{h}_{t-1}, \mathbf{x}_t) \quad (2.10)$$

$$\mathbf{y}_t = \mathbf{W}_{hy}\mathbf{h}_t + \mathbf{w}_{0,y}. \quad (2.11)$$

Le principal atout d'une telle architecture est que les valeurs de l'état \mathbf{h} sont mises à jour itérativement, au fur et à mesure que les entrées \mathbf{x}_t (et \mathbf{h}_{t-1}) sont propagées dans la couche cachée, créant de se fait une sorte de « mémoire » interne au modèle, prenant en compte la dynamique passée de la série temporelle qui lui a été donnée en entrée. Plusieurs variantes existent parmi les **RNNs**. Dans les plus connues on trouve les réseaux de Jordan (JORDAN 1990) et ceux d'Elman (ELMAN 1990). Dans ce dernier cas, les équations du **RNN** deviennent plus spécifiquement :

$$\mathbf{h}_t = \sigma(\mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{w}_{0,h} + \mathbf{W}_{xh}\mathbf{x}_t + \mathbf{w}_{0,x}) \quad (2.12)$$

$$\mathbf{y}_t = \mathbf{W}_{hy}\mathbf{h}_t + \mathbf{w}_{0,y}. \quad (2.13)$$

Notons qu'il est possible d'empiler plusieurs couches de type **RNNs**, voire des

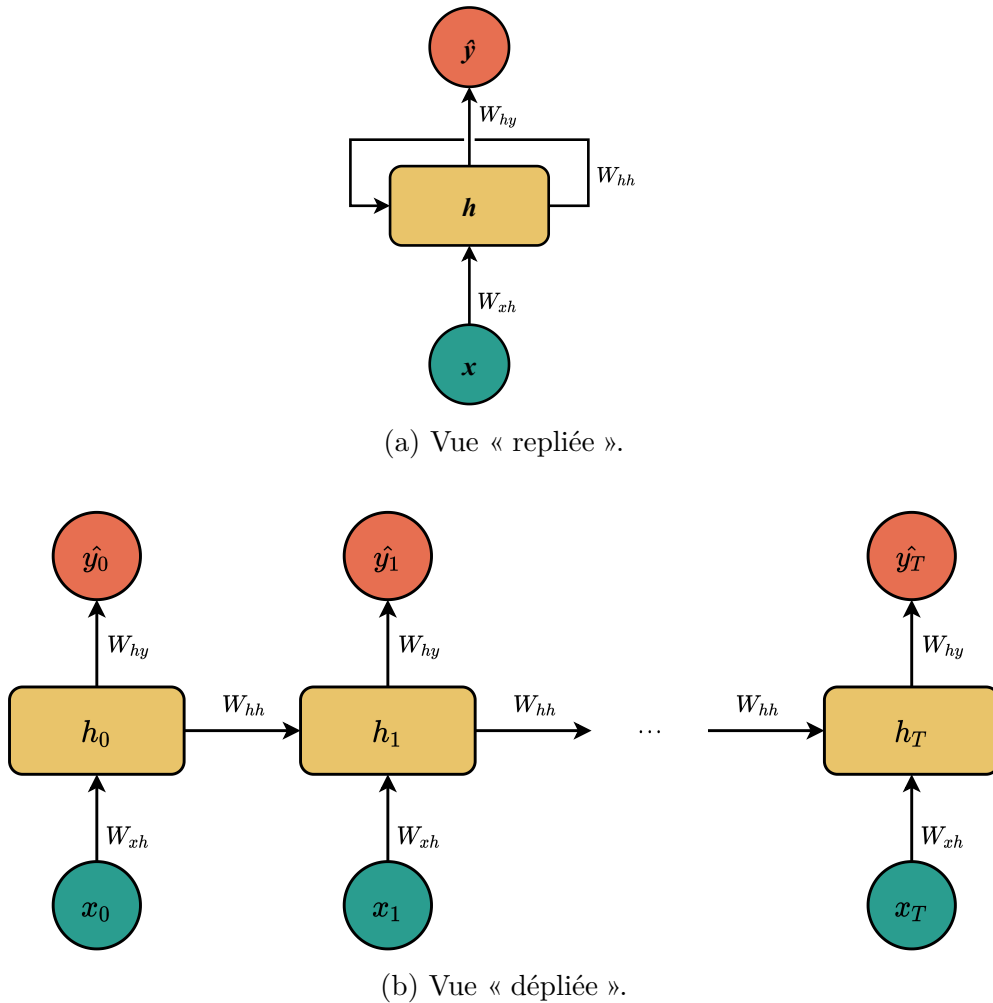


FIGURE 2.5 – Schéma simplifié du fonctionnement d'un RNN.

couches RNN avec des couches linéaires, pour faire un réseau plus profond.

Les RNNs peuvent être entraînés à l'aide d'une méthode dérivée de la rétropropagation du gradient, appelée rétropropagation dans le temps (*backpropagation through time*). L'idée de cette méthode est de rétropropager le gradient dans le réseau non-seulement à travers les couches, mais également en remontant les différentes itérations « temporelles ».

Le principal problème que pose l'utilisation d'un RNN est celui de l'évanescence du gradient. En effet, comme nous l'avons vu, lorsqu'on entraîne un réseau par la méthode de rétropropagation, le calcul des gradients d'une couche dépend des dérivées de la fonction d'activation de cette couche, ainsi que des couches supérieures. Or, des fonctions comme la sigmoïde ont des valeurs de dérivée faibles (dans le cas de la sigmoïde, le gradient est toujours inférieur à 0,3). En rétropropageant le gradient à travers les différentes couches, *i. e.* en multipliant successivement les valeurs (faibles) des dérivées, le gradient peut numériquement « disparaître » au fur et à mesure, ce qui empêche un apprentissage correct des couches les plus profondes (HOCHREITER 1991). Ce problème, qui existe déjà avec les MLPs est encore plus présent avec les RNNs à cause de leur utilisation de rétropropagation

dans le temps, qui multiplie par la longueur des séquences utilisées le nombre de valeurs multipliées entre elles. À noter que le problème inverse « d'explosion » du gradient existe aussi, dans le cas où les multiplications successives entraînent une hausse exponentielle du gradient, ce qui mène à des problèmes similaires. Une façon de mitiger ce problème est l'utilisation de fonctions d'activations spécifiques comme **ReLU**, dont le gradient vaut toujours 0 ou 1, ce qui permet au moins d'éviter le problème d'explosion du gradient, mais cela n'est pas entièrement satisfaisant.

HOCHREITER et SCHMIDHUBER 1997 proposèrent une solution à ce problème sous la forme d'une nouvelle architecture de **RNN** baptisée *Long Short-Term Memory* (**LSTM**). L'idée principale du **LSTM** est de proposer un réseau de neurones possédant deux états, afin de pouvoir maintenir une mémoire de long terme en plus de celle de plus court terme. La Figure 2.6 schématise le fonctionnement interne d'une cellule **LSTM**. Dans cette figure, σ symbolise la fonction sigmoïde (généralisée au cas d'un vecteur). Les poids et biais ne sont pas représentés afin d'alléger la figure.

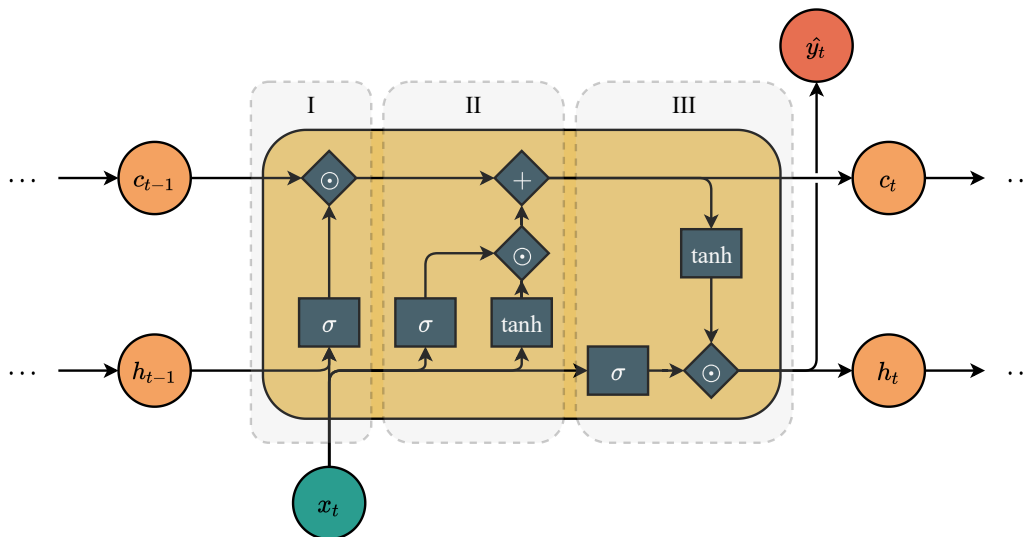


FIGURE 2.6 – Schéma du fonctionnement interne d'une cellule **LSTM**.

Les deux états d'un **LSTM** sont généralement appelés « état de la cellule » (\mathbf{c}) et « état caché » (\mathbf{h}). L'état de la cellule contient la mémoire à long terme que le **LSTM** va conserver au fur et à mesure des itérations temporelles. L'état caché correspond à la mémoire à plus court terme du modèle. La valeur de l'état caché sert également de sortie à la couche⁵. La dimension des états d'un **LSTM** est un paramètre choisi par l'utilisateur, que nous notons D ici. Une cellule **LSTM** est composée de multiples couches qui contrôlent, régulent et transforment le flux d'information contenue dans l'entrée et dans les mémoires à court et à long terme. Nous détaillons ci-dessous la propagation d'un élément \mathbf{x}_t (possiblement un vecteur de dimension d_{in}) d'une entrée (séquentielle) \mathbf{x} , en suivant trois principales étapes.

5. Ces appellations sont donc assez peu explicites, voire contre-intuitives, car l'état caché ne l'est justement pas, vu qu'il s'agit aussi de la sortie de la couche, à laquelle nous avons un accès direct. Mais comme ces dénominations sont unanimement utilisées dans la littérature nous les conservons ici.

- I. L'entrée \mathbf{x}_t et l'état caché de l'itération temporelle précédente \mathbf{h}_{t-1} sont concaténés, ramenés entre 0 et 1 par une fonction sigmoïde et multipliés par l'état de la cellule obtenu lors de l'itération temporelle précédente \mathbf{c}_{t-1} . Cette étape sert à décider, à partir de \mathbf{x}_t et \mathbf{h}_{t-1} , quelle part de la mémoire de long terme \mathbf{c}_t la cellule doit oublier. C'est pourquoi l'opération de concaténation et le passage par la sigmoïde est souvent appelé *forget gate* (« porte de l'oubli »), et la sortie de cette porte est notée \mathbf{f}_t . On a donc :

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{w}_{0,f}), \quad (2.14)$$

où \mathbf{W}_f est une matrice des poids de dimensions $D(D + d_{in})$ et $\mathbf{w}_{0,f}$ un vecteur de biais de dimension D . La première étape répond donc à la question : « est-ce que cela vaut la peine de continuer à garder les informations contenues dans \mathbf{c}_{t-1} en mémoire ? »

- II. \mathbf{x}_t et \mathbf{h}_{t-1} sont concaténés et ramenés entre 0 et 1 par une deuxième fonction sigmoïde (*input gate*, ou « porte d'entrée »). Cela nous donne donc \mathbf{i}_t tel que :

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{w}_{0,i}), \quad (2.15)$$

où \mathbf{W}_i est une matrice des poids de dimensions $D(D + d_{in})$ et $\mathbf{w}_{0,i}$ un vecteur de biais de dimension D . Parallèlement, \mathbf{x}_t et \mathbf{h}_{t-1} sont concaténés et transformés en une information « éligible » (notée $\tilde{\mathbf{c}}_t$) par passage dans une fonction \tanh :

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{w}_{0,c}), \quad (2.16)$$

où \mathbf{W}_c est une matrice des poids de dimensions $D(D + d_{in})$ et $\mathbf{w}_{0,c}$ un vecteur de biais de dimension D . En multipliant \mathbf{i}_t par $\tilde{\mathbf{c}}_t$ on détermine quel part de l'information nouvelle contenue dans \mathbf{x}_t on souhaite garder en mémoire et rajouter dans \mathbf{c}_{t-1} , pour obtenir \mathbf{c}_t . La seconde étape répond donc à la question : « est-ce que cela vaut la peine de sauvegarder l'information contenue dans \mathbf{x}_t en mémoire ? »

En combinant les étapes I. et II. on peut écrire :

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t, \quad (2.17)$$

où \odot désigne un produit membre à membre.

- III. \mathbf{x}_t et \mathbf{h}_{t-1} sont concaténés et ramenés entre 0 et 1 par une troisième fonction sigmoïde (*output gate*, ou « porte de sortie »), ce qui nous donne \mathbf{o}_t :

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{w}_{0,o}), \quad (2.18)$$

où \mathbf{W}_o est une matrice des poids de dimensions $D(D + d_{in})$ et $\mathbf{w}_{0,o}$ un vecteur de biais de dimension D . En parallèle, \mathbf{c}_t passe par une deuxième fonction \tanh , qui détermine quels éléments de la mémoire sont pertinents pour le calcul du prochain état caché \mathbf{h}_t . D'où :

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t). \quad (2.19)$$

Notons que dans le cas du LSTM on a simplement $\hat{\mathbf{y}}_t = \mathbf{h}_t$.

Les LSTMs sont devenus très populaires après avoir démontré une excellente capacité à traiter des problèmes comme la reconnaissance automatique de la parole (GRAVES et al. 2013; AMODEI et al. 2015) ou la prédiction de séries temporelles (SAGHEER et KOTB 2019; CHIMMULA et ZHANG 2020). Les LSTMs sont utilisés aujourd'hui au sein de modèles représentant l'état de l'art de leurs domaines d'application respectifs, y compris en météorologie de l'espace (voir Section 2.3).

2.1.5 Les réseaux convolutifs

Nous avons vu que les RNNs, et en particuliers les LSTMs sont des réseaux adaptés à l'étude de données séquentielles. Il existe également une classe de réseaux dédiés à l'étude des images (et, plus généralement, des données sous forme de tenseurs de dimension ≥ 2) : les réseaux de neurones convolutifs ou *Convolutional Neural Networks* (CNNs). Les CNNs sont des réseaux composés essentiellement de deux types de couches : les couches de convolution, et les couches de *pooling* (« mise en commun »).

Les couches de convolution sont des couches utilisant des neurones convolutifs. Un neurone convolutif est constitué d'un noyau, c'est-à-dire un tenseur de dimensions $kx \times ky$ (dans le cas à deux dimensions), ainsi que d'un biais et d'une fonction d'activation. Le neurone convolutif est schématisé en Figure 2.7. \otimes désigne le produit de Hadamard.

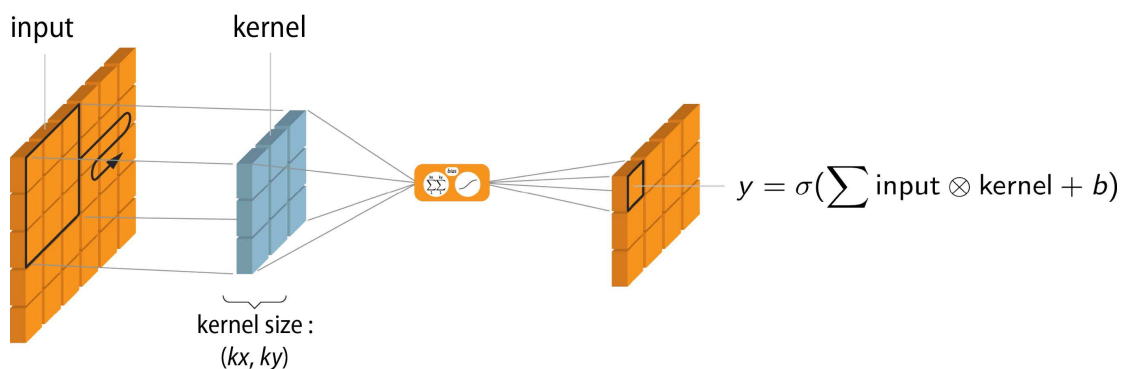


FIGURE 2.7 – Schéma d'un neurone convolutif à deux dimensions. Crédits : FIDDLE, sous licence CC BY-NC-ND 4.0.

Les couches de convolutions sont alimentées par des données s'apparentant à des images. Le noyau de convolution est appliqué à l'ensemble de l'image par décalages successifs. La sortie d'une couche de convolution est donc une nouvelle image, appelée carte d'activation. Ce processus est schématisé en Figure 2.8.

À cause des effets de bord, l'image qui en résulte est souvent de dimensions plus petite que l'image initiale, à moins d'utiliser une méthode comme le *padding* (« rembourrage »), qui consiste à rajouter des pixels nuls autour de l'image d'entrée afin que la sortie de la couche de convolution soit de la dimension souhaitée. Un

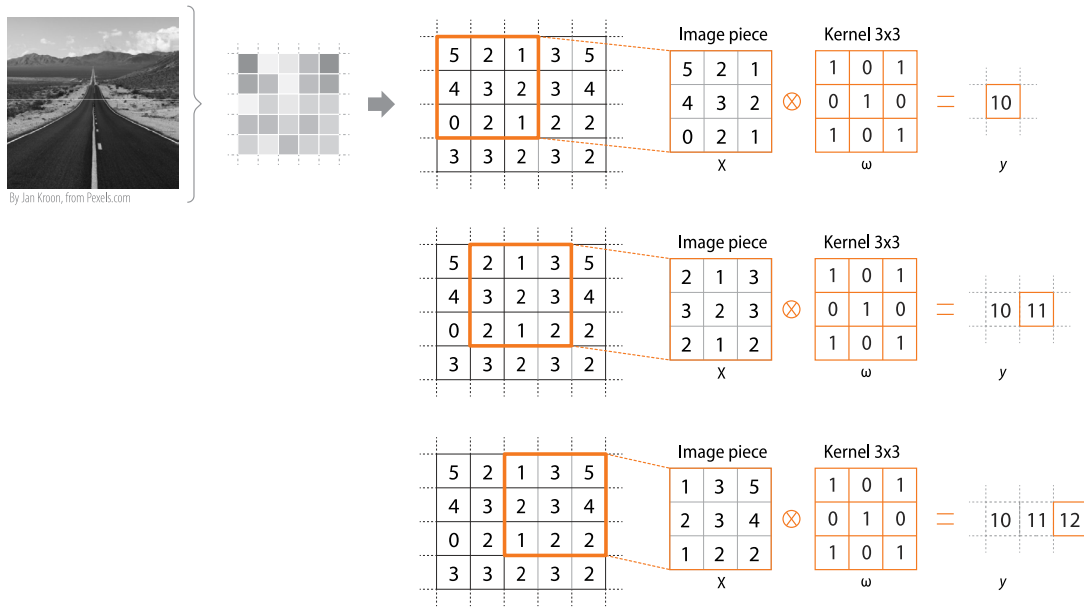


FIGURE 2.8 – Schéma du fonctionnement d’une couche de convolution à un noyau de deux dimensions. Crédits : FIDLE, sous licence CC BY-NC-ND 4.0.

autre paramètre modifiable d’une couche de convolution est le *stride* (« foulée »), qui désigne le nombre de pixels que l’on saute lorsqu’on décale le noyau de convolution. Par ailleurs, notons que les poids du neurone convolutif qui est appliqué à l’image d’entrée sont les mêmes sur toute l’image. Afin de pouvoir extraire de l’image des propriétés différentes (par exemple à différentes échelles spatiales), une couche de convolution sera la plus souvent composée d’un grand nombre de neurones convolutifs. Cela nous donne en sortie plusieurs cartes d’activation (qu’on peut considérer comme une image multispectrale). Remarquons également qu’il est facile d’étendre le concept du cas à deux dimensions à des cas de dimension plus grande. Il suffit alors d’utiliser un noyau également de plus grande dimension, et de le faire parcourir l’image selon tous les axes.

Comme nous l’avons évoqué plus haut, les CNNs sont également constitués de couches de *pooling*. Concrètement, une couche de pooling va combiner plusieurs pixels d’un même voisinage dans une carte d’activation en un seul pixel, afin de réduire la dimension de cette carte d’activation. Les stratégies de *pooling* les plus utilisées sont le *Max Pooling* et le *Average Pooling*. Dans le premier cas, on va garder uniquement la valeur maximale des pixels dans un voisinage, dans le second on prendra la valeur moyenne de ces pixels. Il a été montré empiriquement que l’utilisation de couches de *pooling* entre deux couches de convolution permet d’augmenter significativement les performances d’un modèle (SCHERER et al. 2010). Ainsi, un CNN complet est composé d’une succession de couches de convolution et de couches de *pooling*, et se termine le plus souvent par une ou plusieurs couches « classiques » entièrement connectées à propagation avant, qui interprètent l’information contenue dans les cartes d’activation en une information utile pour l’utilisateur. La structure classique d’un réseau de type CNN est donnée en Figure 2.9.

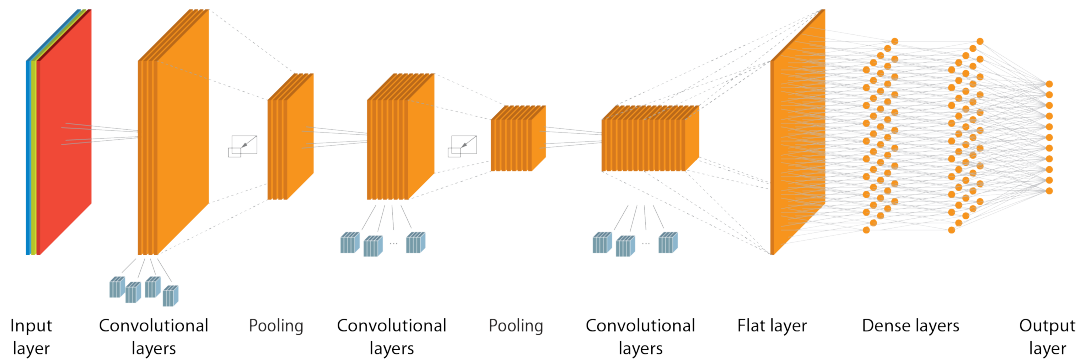


FIGURE 2.9 – Schéma de l'architecture classique d'un réseau de type CNN. Crédits : FIDLE, sous licence CC BY-NC-ND 4.0.

Dans l'exemple de la Figure 2.9, on alimente le réseau initialement avec une image à trois canaux RGB. Cette image passe à travers une première couche de convolution possédant quatre neurones convolutifs, ce qui nous donne en sortie quatre cartes d'activation. La dimension de ces cartes d'activation est réduite par passage dans une couche de pooling. Ces cartes d'activation réduites passent à nouveau deux fois par un ensemble de couches de convolution et de pooling, ce qui nous donne *in fine* un grand nombre de cartes d'activation de dimension très réduite. Ces cartes d'activation sont ensuite « aplaties » (par concaténation verticale de ses colonnes) et passent à travers deux couches cachées entièrement connectées à propagation avant, qui nous donnent la sortie définitive du CNN. Cette architecture permet aussi de comprendre que l'un des principaux avantages des couches de convolution est de permettre d'analyser des données de très haute dimension comme des images, tout en ayant un modèle possédant relativement peu de paramètres par rapport à un MLP ayant le même nombre de couches.

Les CNNs ont été la clef de certains des succès les plus retentissants de l'apprentissage automatique. Nous avons déjà évoqué plus tôt dans ce chapitre le modèle de LECUN et al. 1989, pionnier dans l'utilisation de CNNs pour la reconnaissance et l'interprétation d'écriture manuscrite. Le concours ILSVRC, qui a lieu chaque année, voit régulièrement émerger de nouvelles architectures de plus en plus performantes pour la classification d'images. Parmi les plus connues, on peut citer AlexNet (KRIZHEVSKY et al. 2012), GoogLeNet (SZEGEDY et al. 2014) ou encore ResNet (HE et al. 2015). Les CNNs sont également utilisés pour détecter des pathologies à partir d'imagerie biomédicale (KAYALIBAY et al. 2017) ou pour reconnaître des actions sur des vidéos (BACCOUCHE et al. 2011).

2.1.6 La gestion des données en apprentissage profond

Nous venons de décrire succinctement les principes fondamentaux de l'apprentissage profond et nous avons détaillé trois des principales architectures de réseaux de neurones les plus utilisées. Nous utiliserons ces méthodes dans les chapitres suivants pour procéder à une modélisation dirigée par les données des interactions

Soleil–Terre. Nous décrivons donc ici certaines des particularités de la gestion des données dans un contexte d'apprentissage profond, qui peuvent s'avérer tout aussi – voire plus – importantes que le choix de l'architecture d'ANN utilisée.

Développer un modèle s'appuyant sur l'apprentissage profond implique le passage par deux étapes essentielles : une étape d'entraînement d'un réseau de neurones, et une étape d'évaluation de ses performances et de sa capacité à généraliser. Au cours de l'étape d'entraînement nous alimentons le réseau de neurones avec certaines données et faisons en sorte que le modèle adapte ses paramètres pour minimiser l'erreur qu'il commet. Au cours de l'étape d'évaluation, nous allons tester le modèle avec des nouvelles données qu'il n'a pas vu au cours de son apprentissage, afin d'évaluer ses performances en introduisant le moins de biais possible. En effet, un des écueils possibles lorsqu'on entraîne un réseau de neurones est le surapprentissage, également appelé *overfitting*. L'*overfitting* survient quand un réseau de neurones (ou un autre modèle d'apprentissage automatique) ajuste ses paramètres pour qu'ils correspondent exactement au jeu de données d'entraînement qui lui est donné lors de l'entraînement, dans ce qui ressemble alors à un processus d'apprentissage « par cœur ». Cela se fait au détriment de sa capacité à généraliser à des données qu'il n'aura pas vu pendant l'entraînement. C'est pourquoi il est primordial d'évaluer un modèle d'apprentissage automatique à l'aide d'un jeu de données « nouveau », le plus indépendant possible du jeu de données utilisé pour l'entraînement (CAWLEY et TALBOT 2010).

C'est pourquoi avant même le début de la conception d'un modèle d'apprentissage profond il faut analyser les données dont on dispose et les séparer en trois sous-ensembles : l'ensemble d'entraînement, l'ensemble de validation et l'ensemble d'évaluation (ou de test). L'ensemble d'entraînement est l'ensemble qui va effectivement servir à entraîner le modèle : ce sont ces données que le modèle va utiliser pour minimiser sa fonction de perte. L'ensemble de validation est un premier ensemble servant à évaluer le modèle. Plus précisément, il a pour fonction d'aider à ajuster les hyperparamètres du modèle (c'est-à-dire les paramètres ajustables par l'utilisateur), y compris le nombre d'*epochs* d'entraînement utilisées. C'est l'ensemble de validation qui va nous indiquer si notre modèle est en train de surapprendre ou non, comme l'illustre la Figure 2.10. Sur cette figure on observe un apprentissage correct jusqu'à l'*epoch* 30 (les courbes représentant la valeur de la fonction de perte calculée sur les ensembles de validation et d'entraînement décroît). À partir de l'*epoch* 30 en revanche la valeur de la fonction de perte sur l'ensemble de validation remonte, alors que sur l'ensemble d'entraînement elle continue de décroître, indiquant donc une situation de surapprentissage sur l'ensemble d'entraînement.

Enfin l'ensemble d'évaluation servira à fournir une évaluation juste des performances du modèle sur des données que celui-ci n'aura jamais vues, y compris lors de la validation. En effet, certaines études utilisent parfois le même ensemble pour valider et évaluer le modèle. Toutefois, en faisant cela, on risque d'introduire un biais dû au fait que les hyperparamètres du modèle sont ajustés à partir de ses performances sur l'ensemble de validation. C'est pourquoi, lorsqu'on dispose d'assez de données, il est préférable d'utiliser des ensembles de validation et d'éva-

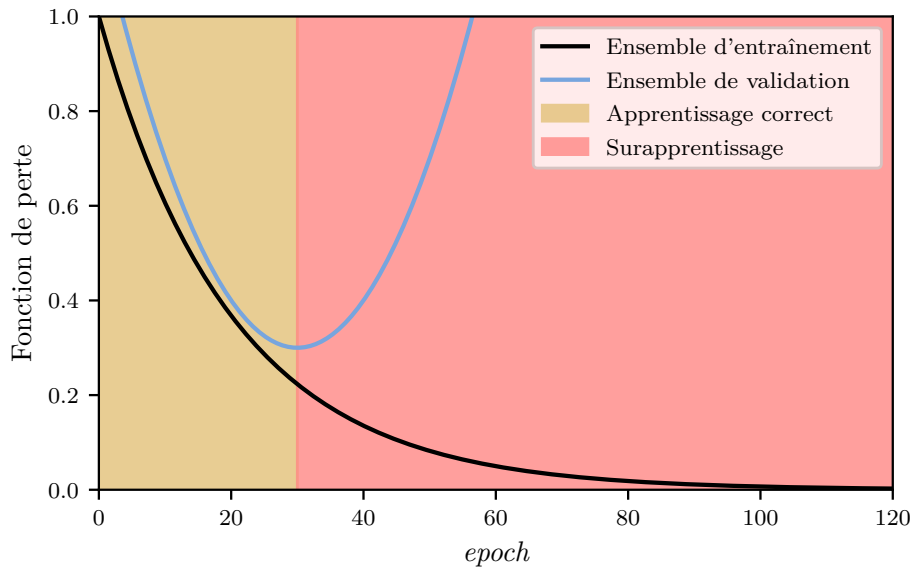


FIGURE 2.10 – Exemple de courbes de fonctions de perte montrant du surapprentissage.

luation distincts. Pour les mêmes raisons il convient de s'assurer que ces différents sous-ensembles sont statistiquement indépendants les uns des autres.

La séparation de l'ensemble de données complet en sous-ensembles doit également prendre en compte la distribution statistique des données considérées. En effet, il convient de s'assurer que chacun des sous-ensembles d'entraînement, de validation et de test soit le plus représentatif possible de la distribution générale des valeurs, afin de ne pas biaiser l'entraînement et/ou l'évaluation.

Un autre aspect très important relatif à la gestion des données réside dans la qualité des données utilisées. En effet, si nous entraînons un réseau de neurones avec des données de mauvaise qualité (rapport signal/bruit trop faible, présence d'une dérive dans le temps, contamination, trous dans les données, etc.) alors le modèle apprendra à reproduire les défauts des données, ce qui amoindrira ses performances et son caractère généralisable. C'est pourquoi avant d'entraîner un réseau de neurones il convient de s'assurer que ses données soient de la meilleure qualité possible, et si nécessaire, utiliser des méthodes adaptées pour mitiger les éventuels problèmes (interpolation pour combler les trous, réduction du bruit, etc.).

Enfin, une dernière particularité de la gestion des données pour une utilisation dans un modèle d'apprentissage profond est la nécessité de pré-traiter les données. Le pré-traitement peut servir à faire ressortir une caractéristique particulière des données (par exemple augmenter le contraste d'une image), ou bien à assurer de bonnes propriétés numériques lors de l'apprentissage. En effet, si on alimente un réseau de neurones avec des données dont les valeurs numériques diffèrent de plusieurs ordres de grandeurs alors on crée un déséquilibre qui va influencer l'entraînement. C'est pourquoi dans la très grande majorité des *pipelines*⁶ impliquant

6. Ensemble d'instructions pour le traitement de données organisées en série.

l'utilisation d'un réseau de neurones une des premières étapes consiste à normaliser les données. Cet objectif peut être atteint à l'aide de différentes méthodes, comme une simple remise à l'échelle (souvent entre 0 et 1 ou entre -1 et 1), ou bien par une modification de la distribution des valeurs via une transformation non-linéaire. Le choix de la méthode de normalisation dépend du modèle utilisé et des données considérées. Ce choix (souvent empirique) peut avoir une influence notable sur les performances finales du modèle. Notons que dans le cas où la méthode de normalisation choisie est une méthode paramétrique, alors ces paramètres sont ajustés en utilisant uniquement le sous-ensemble d'entraînement, encore une fois afin de ne pas inclure de biais lors de l'évaluation.

2.1.7 Autres aspects relatifs à l'implémentation et difficultés potentielles

Nous avons passé en revue différents types d'architectures de réseaux de neurones et décrit certaines des particularités relatives à la gestion de données dans un contexte d'apprentissage profond. Il apparaît donc déjà que la conception et l'implémentation d'un modèle d'apprentissage profond nécessitent plusieurs arbitrages, qui reposent le plus souvent sur des arguments de nature empirique. Ainsi, outre le type de réseau de neurones, il faut déterminer le nombre de couches utilisées, ainsi que le nombre de neurones par couches. Il faut également choisir les fonctions d'activation et la fonction de perte utilisée. Nous avons également évoqué les choix à faire en termes de normalisation des données. Toutefois nous n'avons pas évoqué certaines contraintes qui peuvent également avoir un impact sur la conception d'un modèle.

Tout d'abord, entraîner un réseau de neurones nécessite des moyens de calcul importants, qui doivent être à la fois puissants, véloce, et posséder assez de mémoire pour qu'on puisse y conserver – au moins temporairement – nos données. Il a de plus été montré que l'ensemble des calculs nécessaires à l'entraînement d'un réseau de neurones s'effectue beaucoup plus rapidement sur un processeur de type GPU plutôt que CPU (RAINA et al. 2009). Ainsi l'entraînement d'un réseau de neurones peut nécessiter l'utilisation de moyens de calcul haute performance, ce qui peut s'avérer contraignant.

De plus, l'entraînement d'un réseau de neurones peut prendre un temps très variable en fonction de l'objectif visé et du modèle et des données utilisés. Cela peut aller de quelques minutes à plus d'une semaine (JUMPER et al. 2021). Lorsqu'on rajoute à cela le fait que beaucoup d'hyperparamètres d'un modèle doivent être sélectionnés empiriquement par itérations successives, il apparaît que le temps de développement d'un modèle satisfaisant peut être très long.

Enfin nous avons volontairement omis d'évoquer de nombreuses techniques utiles lorsqu'on développe un réseau de neurones : les différentes méthodes de régularisation (MORADI et al. 2020), le *dropout* (BALDI et SADOWSKI 2013), la « *batch normalization* » (IOFFE et SZEGEDY 2015), la validation croisée, etc. Ces techniques, trop nombreuses pour être évoquées ici en détail ici, peuvent pourtant

parfois améliorer les performances d'un modèle de façon notable. Lorsque nous en utiliserons certaines dans les prochains chapitres, nous les y décrirons alors plus en détail.

Ici, nous voulions simplement attirer l'attention du lecteur sur le fait que la construction d'un modèle d'apprentissage automatique peut vite devenir très complexe et coûteuse – tant du point de vue des ressources informatiques que du temps. Lorsqu'on essaie les différentes techniques mentionnées plus haut, ayant certes fait leurs preuves (bien que souvent seulement empiriquement) mais qui nécessitent un ajustement de leurs hyperparamètres, la complexité de la conception du modèle s'accroît encore, ce qui peut représenter un obstacle non-négligeable pour l'utilisation de ces méthodes par des physiciens ou des chercheurs dont l'apprentissage automatique n'est pas le cœur de compétence.

2.1.8 Bilan sur la modélisation par les données et l'apprentissage profond

Dans cette section nous avons décrit les principes fondamentaux de l'apprentissage profond, ainsi que diverses architectures de réseaux de neurones répondant à des problèmes différents. Nous avons donné autant de détails que nous avons estimés nécessaires afin qu'un lecteur qui ne soit pas familier avec ces méthodes puisse comprendre les techniques utilisées dans la suite de ce manuscrit, et les enjeux qu'elles impliquent. Toutefois ces quelques pages ne sauraient se substituer à une véritable monographie de référence, et nous encourageons à nouveau le lecteur qui souhaiterait approfondir ces sujets à s'appuyer sur les références données.

Nous avons abordé cette thématique par le point de vue d'une modélisation dirigée par les données dans un contexte d'apprentissage supervisé. C'est pourquoi dans la suite de ce chapitre nous nous intéressons désormais aux différentes données issues de l'observation de l'héliosphère – qui peuvent servir à une étude dirigée par les données des interactions Soleil-Terre –, avant de passer en revue diverses études ayant déjà appliqué des méthodes d'apprentissage automatique ou profond pour résoudre des problèmes relevant de la météorologie de l'espace.

2.2 Les données à notre disposition

2.2.1 Observer le Soleil

L'observation du Soleil depuis la Terre est possible, mais présente des limites. En effet, l'atmosphère terrestre absorbe une grande partie des longueurs d'onde émises par le Soleil, laissant principalement passer les longueurs d'onde du visible et masquant partiellement voire totalement d'autres gammes de fréquence. Or certains phénomènes ayant lieu dans la chromosphère ou la couronne solaire ne sont observables que dans ces longueurs d'ondes occultées (IR, UV, EUV, rayons X et gamma). Afin de pouvoir observer le rayonnement du Soleil dans ces longueurs d'onde il est nécessaire d'utiliser des instruments embarqués dans des satellites

dédiés à des missions d’observation du Soleil.

2.2.1.1 Les missions d’observation du Soleil

Depuis les années 1990, plusieurs missions ayant pour but d’observer le Soleil pour mieux le comprendre et l’étudier se sont succédées. Parmi celles-ci, certaines sont encore en activité et continuent de fournir des données exploitables. Ci-dessous nous listons certaines parmi les plus notables de ces missions et en détaillons les instruments embarqués et les données générées.

Yohkoh Yohkoh (aussi connue sous le nom Solar-A) était une mission nippo-americano-britannique ayant pour but l’étude de la physique solaire, lancée en 1991 et ayant duré jusqu’en 2001 (OGAWARA et al. 1991). Le satellite était placé dans une orbite basse légèrement elliptique. Yohkoh embarquait quatre instruments : un télescope à rayons X durs (HXT), un télescope à rayons X mous (SXT) et deux spectromètres. Les images fournies par HXT et SXT avaient une résolution de 1024×1024 pixels, et une résolution spatiale de 5 arcsec (pour un champ de vision total de 35×35 arcsec). Un exemple d’image fournie par Yohkoh/SXT est donné en Figure 2.11. La durée de vie de Yohkoh d’environ 10 ans en a fait la première mission spatiale dédiée à l’observation du Soleil à produire des données au cours d’une période correspondant à la durée d’un cycle solaire.

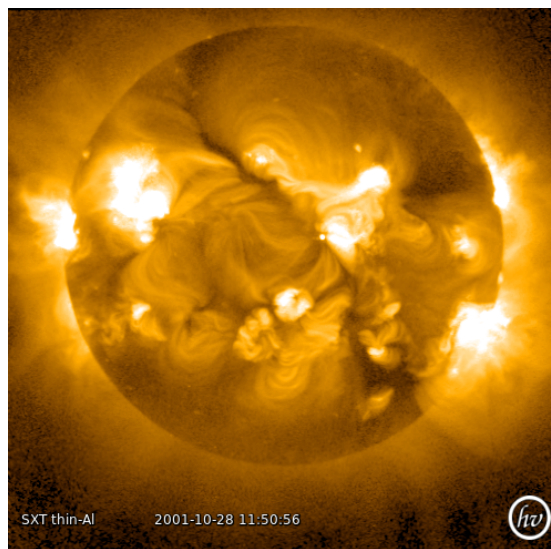


FIGURE 2.11 – Exemple d’image capturée par Yohkoh/SXT le 28 octobre 2001. Crédits : LMSAL. Image générée via HELIOVIEWER.

Yohkoh a eu un successeur en 2006 (et toujours actif) nommé Hinode (Solar-B), qui possède un imageur dans le visible, un imageur à rayons X et un spectromètre (KOSUGI et al. 2007).

Solar and Heliospheric Observatory (SOHO) La mission *Solar and Heliospheric Observatory (SOHO)* est le fruit d’un effort commun entre l’Agence Spatiale

Européenne ou *European Space Agency* (ESA) et la *National Aeronautics and Space Administration* (NASA). C'est une mission qui possède trois objectifs scientifiques principaux : l'étude de la composition du Soleil par héliosismologie, l'étude de la couronne solaire et de ses mécanismes de réchauffement et l'étude du vent solaire (DOMINGO et al. 1995). La mission a été lancée fin 1995 pour une durée de vie initialement prévue d'environ 6 années mais est à ce jour toujours en activité. Le satellite est placé au niveau du point de Lagrange L1 du système Soleil–Terre. Pour atteindre ses objectifs la sonde possède 12 instruments, parmi lesquels on retrouve notamment EIT, LASCO et MDI, que l'on détaille ci-dessous.

Extreme ultraviolet Imaging Telescope *Extreme ultraviolet Imaging Telescope* (EIT) est un imageur dans l'extrême ultra-violet (DELABOUDINIÈRE et al. 1995). Celui-ci permet d'obtenir des images de la couronne solaire et d'une partie de la chromosphère. Cet imageur permet d'observer le Soleil selon quatre longueurs d'onde différentes, correspondant à différentes raies d'émissions, récapitulées dans le Tableau 2.3 ci-dessous, reproduite de DELABOUDINIÈRE et al. 1995 :

TABLEAU 2.3 – Bandes-passantes de l'instrument EIT, provenant de DELABOUDINIÈRE et al. 1995.

Longueur d'onde	Ion	Température	Objectif d'observation
304 Å	He II	$8,0 \times 10^4$ K	chromosphère ; trous coronaux
171 Å	Fe IX et Fe X	$1,3 \times 10^6$ K	région de transition de la couronne ; structures internes des trous coronaux
195 Å	Fe XII	$1,6 \times 10^6$ K	couronne calme hors des trous coronaux
284 Å	Fe XV	$2,0 \times 10^6$ K	régions actives

L'imageur capture l'intégralité du disque solaire avec un champ de vision de 45×45 arcmin et une résolution spatiale d'environ 5 arcsec. La résolution d'une image fournie par EIT est de 1024×1024 pixels. La cadence est d'une image toutes les 12 minutes environ. Cet instrument fournit des images jusqu'à ce jour depuis le 01/01/1996, constituant maintenant une base de données de plus de 25 ans, ce qui est un atout très précieux lorsqu'on développe des modèles pilotés par les données. Des exemples d'images fournies à chaque longueur d'onde par EIT sont donnés en Figure 2.12.

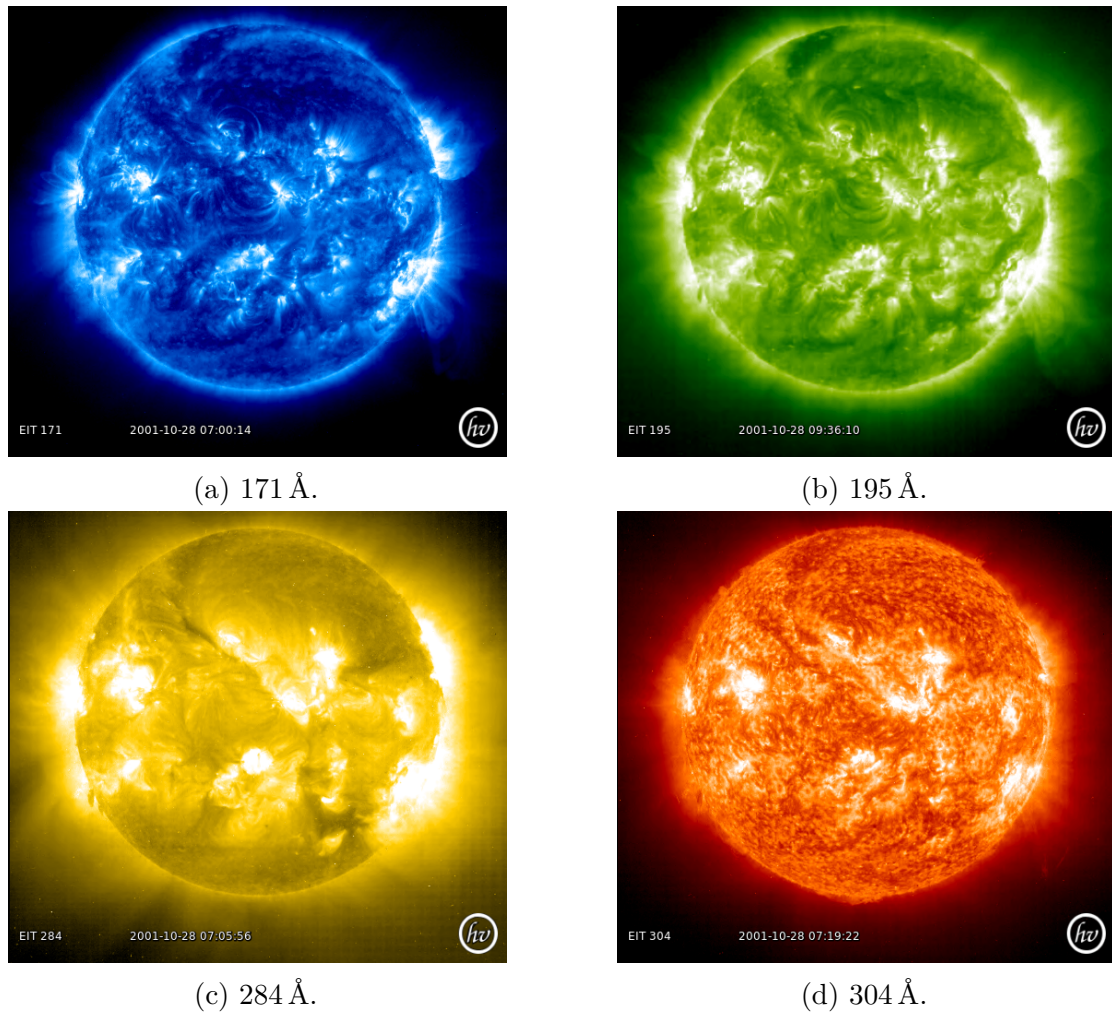


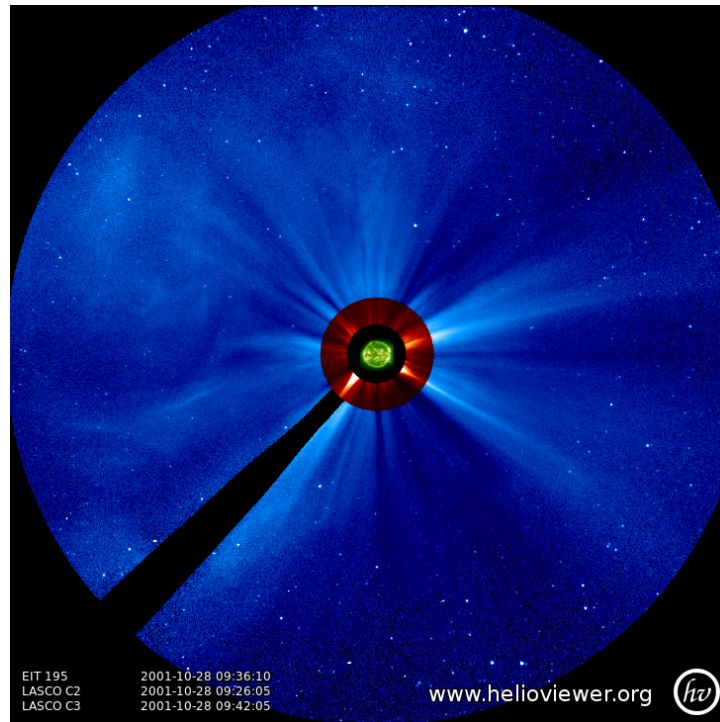
FIGURE 2.12 – Exemples d’images capturées par **SOHO/EIT** aux quatre longueurs d’ondes possibles le 28 octobre 2001. Crédits : **SOHO** (ESA & NASA). Images générées via **HELIOVIEWER**.

Large Angle and Spectrometric Coronagraph *Large Angle and Spectrometric Coronagraph* (**LASCO**) est un instrument embarqué à bord de **SOHO** (BRUECKNER et al. 1995). Il est composé de trois coronographes (C1, C2 et C3) qui permettent d’observer la couronne solaire dans le domaine du visible et par spectroscopie. Un coronographe est un instrument d’observation de la couronne solaire qui reconstitue une éclipse solaire artificielle à l’aide d’un système optique dédié. Les caractéristiques principales des trois coronographes sont détaillées dans le Tableau 2.4 ci-dessous :

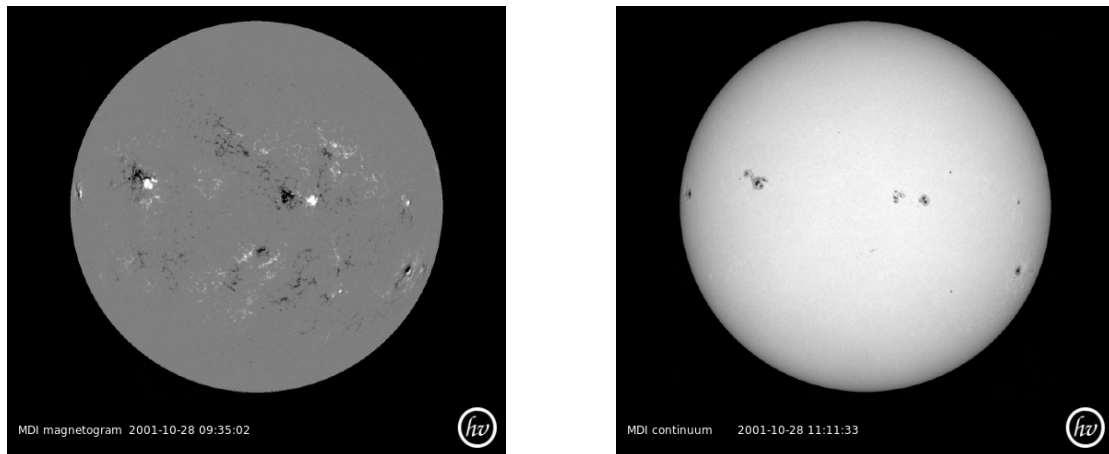
Malheureusement le coronographe C1 a cessé de fonctionner en juin 1998, lorsque **SOHO** a subi un échec de son système de contrôle d’attitude. En revanche C2 et C3 continuent jusqu’à aujourd’hui à capturer des images, avec une cadence d’environ une dizaine d’images par heure. Un exemple de coronographes obtenus avec C2 et C3 est donné en Figure 2.13.

TABLEAU 2.4 – Caractéristiques des coronographes de l'instrument **LASCO**, reproduit de BRUECKNER et al. 1995.

Nom	Champ de vision (R_{\odot})	Type d'occul-teur	Bande-passante spectrale	Objectif	Résolution spatiale (par pixel)	Gamme de luminosité (B_{\odot})
C1	1,1 – 3,0	Interne	Fabry-Pérot	Miroir	5,6 arcsec	2×10^{-5} – 2×10^{-8}
C2	1,5 – 6,0	Externe	Large-bande	Lentille	11,4 arcsec	2×10^{-7} – 5×10^{-10}
C3	3,7 – 30	Externe	Large-bande	Lentille	56 arcsec	3×10^{-9} – 10^{-11}

FIGURE 2.13 – Superposition d'images capturées par **SOHO/EIT** à 195 Å (vert), **LASCO/C2** (rouge) et **LASCO/C3** (bleu) le 28 octobre 2001. Image générée via **HELIOVIEWER**.

Michelson Doppler Imager *Michelson Doppler Imager (MDI)* est un magnéto-mètre et dopplermètre embarqué à bord de **SOHO** (SCHERRER et al. 1995). C'est un instrument d'héliosismologie ayant pour but d'analyser les modes de vibration du Soleil. Il permet également d'obtenir des magnétogrammes de la photosphère solaire. Concrètement, **MDI** propose trois types de produits : des dopplergrammes, des magnétogrammes et des images du Soleil dans le continuum du visible. Ces images sont également de résolution 1024×1024 pixels. Cet instrument est toujours en activité aujourd'hui. Nous donnons ici un exemple de magnétogramme et



(a) Magnétogramme.

(b) Continuum.

FIGURE 2.14 – Exemples d’images (magnétogramme et continuum) capturées par **SOHO/MDI** le 28 octobre 2001. Crédits : **SOHO (ESA & NASA)**. Images générées via **HELIOVIEWER**.

d’image dans le continuum obtenus avec **MDI** en Figure 2.14.

Bien que le satellite soit toujours en état de fonctionnement, dépassant ainsi toutes les attentes, et ce malgré la défaillance de juin 1998, il a depuis 2011 un successeur : *Solar Dynamics Observatory (SDO)*, présenté plus bas dans cette section.

Solar Terrestrial Relations Observatory (STEREO) *Solar Terrestrial Relations Observatory (STEREO)* était une mission de la NASA, composée d’un ensemble de deux sondes jumelles en orbite héliosynchrone dédiées à l’observation du Soleil. Lancées en 2006, nous avons perdu le contact avec la sonde STEREO-B en 2014, tandis que STEREO-A fournit toujours des données à ce jour. Le principal objectif de la mission **STEREO** était d’observer les **CMEs** et en étudier l’origine et leur propagation dans l’héliosphère. La présence de deux sondes permettait d’obtenir des représentations stéréoscopiques du Soleil et de **CMEs** pendant un temps restreint. Les sondes comportaient chacune 16 instruments réparties dans 4 suites. En particulier, la suite *Sun–Earth Connection Coronal and Heliospheric Investigation (SECCHI)*, composée des quatre instruments listés ci-dessous, était dédiée à l’imagerie solaire (HOWARD et al. 2008).

Extreme UltraViolet Instrument *Extreme UltraViolet Instrument (EUVI)* est un imageur dans l’**EUV** (WUELSER et al. 2004). Il produit des images dans les mêmes longueurs d’onde que l’imageur de **SOHO**. Néanmoins le capteur CCD fournit cette fois des images de résolution 2048×2048 pixels, avec une meilleure résolution spatiale et cadence temporelle que celles de **SOHO/EIT**.

Les coronographes COR1 et COR2 Chaque sonde STEREO embarquait deux coronographes, COR1 et COR2. Ces deux instruments observent dans le visible et fournissent des coronogrammes de la couronne interne ($1,3 - 4 R_{\odot}$ pour COR1) et externe ($2 - 15 R_{\odot}$ pour COR1) de résolution 2048×2048 pixels.

Heliospheric Imagers (HI-1 et HI-2) HI-1 et HI-2 sont deux imageurs dédiés à l'observation de l'héliosphère, afin notamment d'y étudier la propagation des CMEs une fois émises par le Soleil. HI-1 a un champ de vision allant de 15 à $84 R_{\odot}$ et HI-2 de 66 à $318 R_{\odot}$. Ce sont des caméras observant dans le domaine du visible. Les images fournies par cet instrument aussi sont de résolution 2048×2048 pixels.

Project for On-Board Autonomy (PROBA-2) Le satellite *Project for On-Board Autonomy (PROBA-2)* est une petite sonde (130 kg) conçue par l'ESA et servant de démonstrateur pour de nouvelles technologies spatiales. Mis en orbite héliosynchrone en 2009, ce satellite emporte notamment un imageur solaire, *Sun Watcher using Active Pixel System detector and Image Processing (SWAP)*, évolution des imageurs EIT ou EUVI. Celui-ci est notamment plus petit, plus léger et comporte des procédures de sauvegarde plus sophistiquées (BERGHMANS et al. 2006). Toutefois il ne fournit des observations que selon la longueur d'onde 174 \AA . Bien qu'ayant une durée de vie initialement prévue de seulement 2 ans, la sonde et l'instrument SWAP sont toujours fonctionnels.

Solar Dynamics Observatory (SDO) SDO est une mission de la NASA conçue dans le cadre du programme *Living with a star* (PESNELL et al. 2012). Prenant le relais de SOHO, SDO propose d'étudier l'activité solaire et son impact sur la météorologie de l'espace à l'aide de trois instruments. Le satellite a été lancé en février 2010 et est placé en orbite géosynchrone autour de la Terre. Sa durée de vie initialement prévue était de 5 ans mais à ce jour le satellite est toujours fonctionnel. Nous détaillons ci-dessous les instruments embarqués dans SDO.

Atmospheric Imaging Assembly *Atmospheric Imaging Assembly (AIA)* est un imageur embarqué dans SDO, composé de 4 télescopes produisant des images dans 10 longueurs d'onde différentes (7 dans l'EUV, 2 dans l'UV et 1 dans le visible) avec une résolution spatiale de $1,5$ arcsec et une période d'acquisition par canal de 12 secondes (LEMEN et al. 2012). Les images comprennent le disque solaire complet jusqu'à une distance radiale de $0,5 R_{\odot}$ au-delà du limbe solaire. La résolution des images est de 4096×4096 pixels, ce qui est une augmentation forte par rapport à SOHO/EIT. Le Tableau 2.5 ci-dessous résume les différentes longueurs d'onde d'observation d'AIA et leurs caractéristiques. La Figure 2.15 montre des images prises par AIA dans les 10 longueurs d'onde possibles.

On constate que plusieurs longueurs d'onde assurent une parfaite continuité avec l'imageur EIT de SOHO (notamment 171 , 195 et 304 \AA) voire avec les images dans le continuum fournies par SOHO/MDI.

TABLEAU 2.5 – Différentes longueurs d’onde d’acquisition de l’instrument *AIA* et caractéristiques associées (LEMEN et al. 2012).

Longueur d’onde	N° Télescope	Ion	$\log(T)$ caractéristique	Objectif d’observation
4500 Å	3	continuum	3,7	photosphère
1700 Å	3	continuum	3,7	température minimum ; photosphère
304 Å	4	He II	4,7	chromosphère ; région de transition
1600 Å	3	C IV + continuum	5,0	région de transition ; photosphère supérieure
171 Å	3	Fe IX	5,8	couronne calme ; région de transition supérieure
193 Å	2	Fe XII + Fe XXIV	6,2 et 7,3	couronne et flare de plasma chaud
211 Å	2	Fe XIV	6,3	régions actives de la couronne
335 Å	1	Fe XVI	6,4	régions actives de la couronne
94 Å	4	Fe XVIII	6,8	flares coronaux
131 Å	1	Fe VIII + Fe XXI	5,6 et 7,0	région de transition et flares coronaux

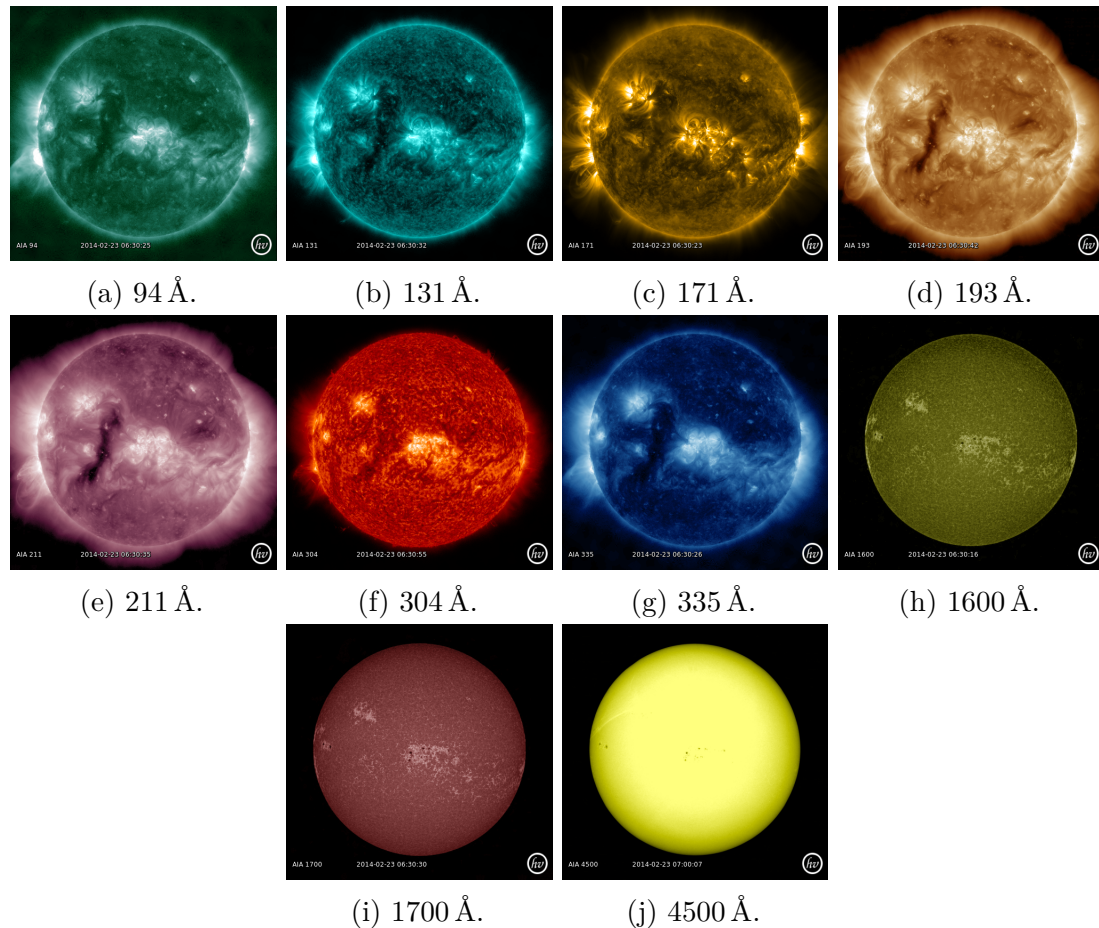


FIGURE 2.15 – Exemples d’images capturées par [SDO/AIA](#) aux dix longueurs d’ondes possibles le 23 février 2014. Crédits : [SDO \(NASA\)](#). Images générées via [HELIOVIEWER](#).

Helioseismic and Magnetic Imager *Helioseismic and Magnetic Imager (HMI)* est un dopplermètre et magnétomètre embarqué à bord de [SDO](#). Ses objectifs principaux sont de comprendre et caractériser les dynamiques solaires et les composantes intervenant dans son activité magnétique. Cet instrument reprend et améliore les capacités de [SOHO/MDI](#). La résolution spatiale des magnétogrammes de HMI est de 0,5 arcsec/pixel, chaque image ayant une résolution de 4096×4096 pixels. Les dopplégrammes sont fournis toutes les 45 secondes, tandis que les magnétogrammes le sont toutes les 90 à 135 secondes. Un exemple de magnétogramme et un exemple d’image dans le continuum obtenus à l’aide de [HMI](#) sont donnés en 2.16.

Extreme Ultraviolet Variability Experiment *Extreme Ultraviolet Variability Experiment (EVE)* est un instrument dédié à la mesure de l’irradiation dans l’EUV du Soleil (WOODS et al. 2012). À l’aide de deux spectromètres, [EVE](#) est capable de fournir des données d’irradiance pour des longueurs d’onde allant de 0,1 à 135 nm, avec une résolution spectrale de 0,1 nm et une cadence temporelle de 10 s.

Notons que la mission [SDO](#) fournit à la fois des données de niveau 1 (données

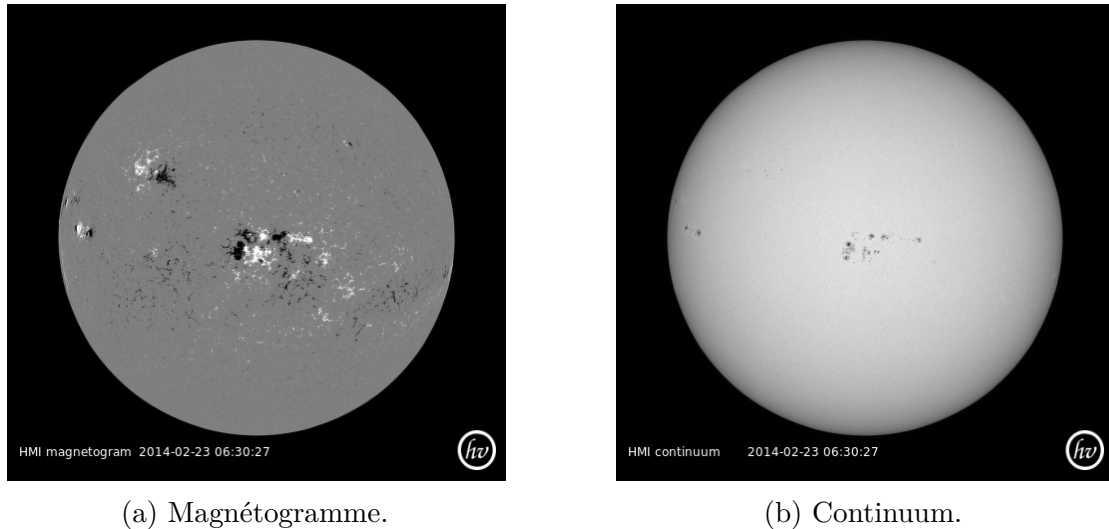


FIGURE 2.16 – Exemples d’images (magnétogramme et continuum) capturées par SDO/HMI le 23 février 2014. Crédits : SDO (NASA). Images générées via HELIOVIEWER.

calibrées), mais aussi des données de niveau 1.5. Les données de niveau 1.5 sont intéressantes car elles correspondent par exemple à des légères modifications des données de niveau 1, afin que les images fournies par AIA et HMI soient correctement alignées, occupent la même surface sur l’image et aient la même résolution spatiale. Cela permet la superposition de ces produits et donc rend leur utilisation conjointe plus aisée.

Dans nos travaux nous utiliserons essentiellement les données fournies par SDO, plus particulièrement par son instrument AIA. La qualité des données fournies, leur fréquence d’acquisition, la multiplicité des longueurs d’onde d’observation proposées et la pré-existence d’études utilisant ces données dans un contexte d’apprentissage automatique en font des atouts considérables pour les objectifs de notre étude (voir à ce sujet le Chapitre 5).

Les autres missions D’autres missions d’observations du Soleil autres que celles mentionnées précédemment ont eu lieu. Nous pouvons par exemple mentionner le micro-satellite français Picard (130 kg), qui, placé en orbite héliosynchrone, observait le Soleil à l’aide de trois instruments, incluant un télescope, mais dont la principale tâche est de mesurer de façon précise le diamètre du Soleil, ainsi que d’étudier l’intérieur du Soleil par héliosismologie (MEFTA et al. 2014). Nous pouvons également citer la mission américaine *Transition Region And Coronal Explorer* (TRACE), lancée en 1998 et ayant duré 12 ans, qui possède un télescope d’observation du Soleil, complémentaire aux observations faites par SOHO (HANDY et al. 1999).

Enfin plus récemment deux missions d’observation et d’étude du Soleil très ambitieuses ont été lancées : la sonde américaine *Parker Solar Probe* (PSP) et la sonde européenne Solar Orbiter. Celles-ci ont pour particularité d’être placées en

orbite autour du Soleil, de sorte à pouvoir observer le Soleil de plus près (périapside de 0,04 au pour PSP et 0,28 au pour Solar Orbiter). Ces sondes ont été lancées respectivement en 2018 et 2020, chacune pour une durée de vie prévue d'environ 7 ans (FOX et al. 2016 ; MÜLLER et al. 2013).

2.2.1.2 Les autres sources de données

L'imagerie solaire disponible ne se résume pas aux produits de missions spatiales, et des données d'imagerie solaire prises depuis le sol sont également disponibles. Comme nous l'avons indiqué plus haut, un des avantages de l'imagerie solaire spatiale réside dans la possibilité d'observer le Soleil dans des longueurs d'onde filtrées par l'atmosphère terrestre, par l'exemple l'EUUV. En revanche cela a également des désavantages liés au fait que ce sont des missions spatiales : les instruments sont soumis à des conditions environnementales extrêmes, notamment thermiques et radiatives, il peut y avoir des pertes dues à des problèmes de télémetrie, la disponibilité des données n'est pas sans faille, le satellite peut avoir des défaillances mécaniques et surtout toute réparation physique (sauf cas exceptionnel) est rendue impossible. C'est pourquoi il est également intéressant d'avoir des données obtenues à partir d'instruments au sol, surtout dans un contexte opérationnel où la continuité et la disponibilité des données sont des enjeux importants.

Parmi les observatoires du Soleil basés au sol et proposant en particulier des données orientées pour un usage en météorologie de l'espace on trouve notamment *Global Oscillations Network Group* (GONG). Géré par le *National Solar Observatory* – l'observatoire solaire américain (NSO), GONG est un des projets phares de son programme *NSO Integrated Synoptic Program* (NSO). Il s'agit d'un ensemble de six télescopes identiques disposés à travers le monde, permettant une observation du Soleil quasi-permanente (la couverture temporelle actuelle est d'environ 90%). Chaque télescope de GONG est construit selon le principe de l'interféromètre de Michelson et l'image est capturée par un capteur CCD ayant une résolution spatiale de 5 arcsec. Depuis 1995, GONG est capable de fournir divers types de données, incluant des magnétogrammes du disque solaire (toutes les minutes), des images du soleil (selon la longueur d'onde H_α) et des données d'héliosismologie (HARVEY et al. 1996).

Le programme NISP inclut également le projet *Synoptic Optical Long-term Investigations of the Sun* (SOLIS), qui est un observatoire abritant trois instruments : un spectromagnétomètre, un imageur et un spectrographe. Le spectromagnétomètre *Vector Spectromagnetograph* (VSM) fournit des cartes vectorielles du champ magnétique de la chromosphère et de la photosphère solaires avec une cadence journalière. L'imageur *Full-Disk Patrol* (FDP) fournit des images du disque solaire selon plusieurs longueurs d'onde (dont H_α et dans le visible KELLER et al. 2003).

2.2.1.3 Les données dérivées de l'imagerie

Que ce soit à partir des données d'imagerie solaire spatiale ou bien obtenue à l'aide des observatoires terrestres, il existe un certain nombre de données dérivées qui peuvent servir de base à la construction d'un modèle météorologie de l'espace. Parmi ces données on trouve notamment les cartes synoptiques, qui sont des reconstructions de magnétogrammes de toute la surface du Soleil à partir de plusieurs magnétogrammes du disque solaire tel qu'il est vu par l'observatoire (terrestre ou spatial) en question. Ainsi NISP propose des cartes synoptiques construites à partir de magnétogrammes GONG et SOLIS (en général une par heure pour GONG par exemple). Il existe également des cartes synoptiques construites à partir de magnétogrammes obtenus par SOHO/MDI et SDO/HMI.

À partir des cartes synoptiques elles-même et à l'aide de modèles de physique solaire il est possible de dériver de nouvelles bases de données, comme la base de données de cartes *Air Force Data Assimilative Photospheric Flux Transport (ADAPT)*, composée de cartes du champ magnétique solaire obtenues par assimilation de données (ARGE et al. 2010). Les données utilisées pour l'assimilation de données sont des cartes synoptiques GONG et SOLIS, voire HMI. Le modèle utilisé est un modèle de transport de flux (WORDEN et HARVEY 2000) et la méthode d'assimilation est l'*Ensemble Least Squares (ENLS)*. ADAPT fournit, à partir des données GONG, une carte synoptique toutes les 6 heures.

2.2.1.4 Autres données utiles

Pour compléter toutes ces bases de données relevant de l'imagerie solaire il existe d'autres sources de données de nature différente. Nous pouvons citer par exemple la *Heliophysics Event Knowledgebase (HEK)*, mise en place par le Lockheed Martin *Solar and Astrophysics Laboratory (LMSAL)* qui comprend une base de données d'événements solaires, ainsi que des cartes de régions d'intérêt (régions actives, trous coronaux, etc) se superposant aux images fournies par SDO/AIA et HMI (HURLBURT et al. 2012). On peut également citer le *Space Weather Prediction Center (SWPC)* de la *National Oceanographic and Atmospheric Administration (NOAA)* qui propose des données diverses, comme des relevés manuels des caractéristiques quotidiennes de l'activité solaire, ou encore des mesures de flux de rayons X mesurés par des satellites de la série *Geostationary Operational Environmental Satellite (GOES)*, servant à observer et caractériser les embrillancements solaires.

2.2.1.5 Accéder aux données

Généralement, les données provenant de chaque instrument sont disponibles auprès des organismes responsables de l'aspect scientifique de la mission en question. Très souvent, ces données sont mises à disposition publiquement via un site internet. Par ailleurs, il existe divers agrégateurs de données issues de l'observation du Soleil. On peut par exemple citer le *Virtual Solar Observatory (VSO)*, géré par le NSO, mais aussi la base de données *Multi Experiment Data and Operation Center*

(MEDOC). MEDOC, tenu par l'Institut d'Astrophysique Spatiale (IAS), propose également des archives de données solaires, ainsi qu'un ensemble de données dérivées issues de plusieurs sources (dont la HEK ou encore des catalogues d'événements), mais aussi des outils, des modèles et des résultats de simulations.

On peut aussi citer la plateforme HELIOVIEWER, qui permet de facilement visualiser et accéder à de l'imagerie solaire provenant des missions citées plus haut. Notons qu'il existe une librairie Python dédiée à la recherche, acquisition et traitement de données d'imagerie solaire s'appelant SunPy et qui propose une interface unifiée pour acquérir les données provenant de certaines des bases de données citées ci-dessus (COMMUNITY et al. 2020).

2.2.1.6 Bilan sur l'imagerie solaire

Il apparaît que depuis les premières missions d'observation du Soleil il y ait eu une volonté de la part des agences spatiales et des laboratoires de maintenir une capacité continue d'observation du Soleil. La quantité et la qualité des produits fournis ont augmenté au cours des trois dernières décennies, tout en gardant une volonté d'assurer une certaine continuité, par exemple via le choix des longueurs d'onde d'observation des imageurs.

Néanmoins les nombreuses différences entre les différentes missions (instruments, résolution spatiale, fréquence d'acquisition, orbite du satellite, etc.) rendent l'exploitation conjointe de ces données complexe. En particulier, dans un contexte de modélisation par les données, il est important que la base de données utilisée soit le plus homogène possible. Or, parvenir à rendre homogène une base de données utilisant des données provenant de missions différentes semble particulièrement difficile. C'est pourquoi la plupart des études utilisant ces données au sein de modèles d'apprentissage automatique ou profond n'utilisent que des données provenant d'une unique source à la fois (voir Section 5.1).

2.2.2 Étudier le vent solaire

2.2.2.1 Les missions d'étude du vent solaire

L'étude des interactions Soleil-Terre passe également par l'observation et l'analyse du vent solaire. Le principal moyen de mesurer l'état du vent solaire à différents endroits de l'héliosphère consiste en l'utilisation d'instruments *in situ*.

Depuis le début de l'ère spatiale, de nombreuses missions dédiées à l'étude du vent solaire en amont de la Terre se sont succédé, nous fournissant des données et des mesures de qualité variable. Parmi les premières missions on peut citer notamment les satellites IMP-1 (1963 – 1964), 3 (1965 – 1967), 4 (1967 – 1968), 5 (1969 – 1972), 6 (1971 – 1974), 7 (1972 – 1978) et 8 (1974 – 2001), tous ayant des orbites géocentriques très elliptiques, approche présentant le défaut notable de ne pas garantir une couverture permanente (car proches du périhélie les satellites se trouvent dans la magnétosphère).

Par la suite, les agences spatiales ont préféré placer les satellites d'étude du vent solaire autour du point de Lagrange L1, situé sur l'axe reliant le Soleil à la Terre, légèrement en amont de celle-ci, il s'agit d'un point idéal où placer un satellite dont l'objectif est la mesure des paramètres du vent solaire se dirigeant vers la Terre.

Parmi les satellites d'étude du vent solaire placés autour de L1 on peut citer principalement WIND (1994 – aujourd'hui) et ACE (1995 – aujourd'hui). Ces satellites, encore actifs aujourd'hui, sont la principale source de données d'observation du vent solaire dans le contexte de l'étude du système Soleil–Terre. Comme leur positionnement spatial leur assure une excellente continuité de service en amont de notre planète, ces satellites ont également un rôle clef au sein de la communauté de météorologie de l'espace. Les paramètres mesurés par ces satellites sont notamment les paramètres du vent solaire déjà listés en section 1.1.5, dont notamment les différentes composantes de sa vitesse et de son champ magnétique (IMF-B) mais aussi sa densité, sa température, et ses paramètres plasma.

2.2.2.2 L'accès aux données

Il est souvent possible d'accéder aux mesures faites par les différents instruments directement auprès de leur *Principal Investigator*. Une autre possibilité, que nous avons privilégiée ici, est l'utilisation de la base de données OMNIWeb⁷. Cette base de données, tenue par la NASA, propose un historique complet des paramètres du vent solaire, en regroupant les mesures effectuées par les différentes missions qui se sont succédé. De plus les données sont intercalibrées, et toutes sont projetées au nez du choc d'étrave, afin d'être spatialement homogènes (KING et PAPITASHVILI 2005).

De ce fait, les paramètres du vent solaire mis à disposition sur OMNIWeb sont tout à fait adaptés à une utilisation dans des modèles pilotés par les données. La seule difficulté réside dans la présence dans certaines séries de données dues à des trous temporels. Cette absence de données peut avoir de multiples causes : latence entre deux missions successives, panne momentanée d'un instrument, télémétrie corrompue, etc. Ces trous temporels, lorsqu'ils ne sont pas trop grands, peuvent être remplis à l'aide de différentes techniques d'interpolation. Ce sujet, et l'influence de la méthode d'interpolation, sera notamment abordé dans le Chapitre 3.

Deux catégories de données OMNI sont disponibles : les données à basse fréquence (pas de temps d'une heure), et les données à haute fréquence (pas de temps de l'ordre de la minute). Dans nos différentes études nous avons toujours utilisé les données à basse fréquence, suffisantes pour nos usages.

7. <https://omniweb.gsfc.nasa.gov/index.html>

2.2.3 Les indices géomagnétiques

D'une façon générale, un indice peut être vu comme un « nombre permettant de résumer et de caractériser la variation relative d'une grandeur simple ou complexe entre deux situations, dont l'une sert de base (temps ou lieu de référence)⁸. » Un indice géomagnétique est, plus spécifiquement, une mesure de la perturbation de la magnétosphère terrestre causée par les différents processus inhérents au couplage vent solaire–magnétosphère (MENVIELLE et al. 2011).

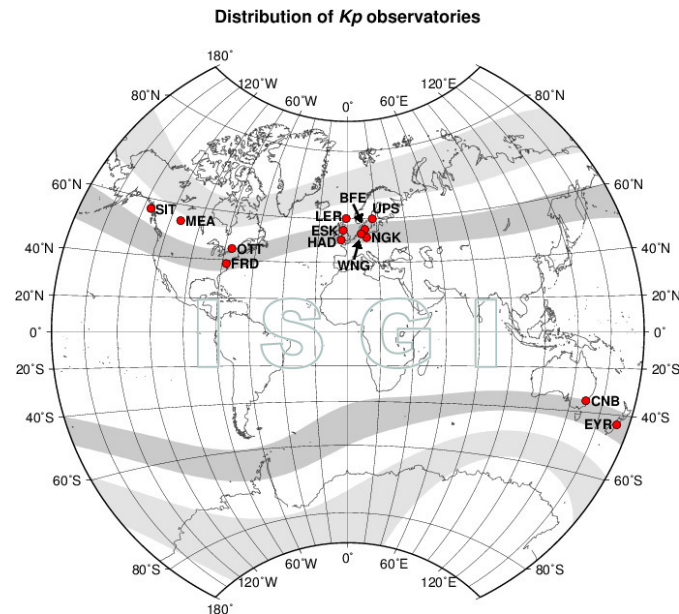


FIGURE 2.17 – Observatoires dont les mesures sont utilisées pour produire l'indice K_p . Crédits : ISGI.

Les indices géomagnétiques sont le plus souvent calculés à partir de mesures faites au sol, par des magnétomètres situés dans des observatoires répartis à la surface du globe. En fonction de l'emplacement des observatoires choisis, on peut créer des indices géomagnétiques représentatifs de perturbations de nature physique différente.

Historiquement, la communauté scientifique a ressenti le besoin de définir et calculer des indices géomagnétiques assez tôt. Le premier indice proposé était l'indice C , un indice assez simple, ne pouvant prendre que trois valeurs discrètes (0, 1 ou 2) et calculé à partir des écarts maximaux de l'intensité du champ magnétique mesurés quotidiennement (MAYAUD 1980). BARTELS et al. 1939 a ensuite proposé les indices K , qui sont rapidement devenus essentiels. Un indice K est un indice spécifique à l'observatoire qui le calcule, à partir des perturbations des deux composantes horizontales du champ magnétique terrestre local, avec un pas de temps de 3 heures entre deux valeurs produites. À partir de la mesure de perturbation obtenue, qui est un signal continu s'exprimant en nanotesla, on calcule un indice adimensionnel, prenant des valeurs discrètes comprises entre 0 (pour une absence de perturbation) à 9 (pour un événement extrême, correspondant à moins de 0,068 % des valeurs selon BARTELS et al. 1939). Cette échelle est souvent

8. Définition du dictionnaire Larousse.

qualifiée de « logarithmique » de façon abusive, car elle est en réalité obtenue à partir d’une table de conversion.

À partir des indices K de multiples observatoires il est possible de définir d’autres indices. Le plus utilisé est sans doute l’indice Kp , proposé dans un premier temps par BARTELS 1949, construit à partir des données de 13 observatoires de latitude moyenne – autour de 50° Nord pour 11 stations et Sud pour deux stations –, représentés sur la Figure 2.17. Comme son nom l’indique, Kp (le p signifie « planétaire ») est un indice qui se veut représentatif de l’état de perturbation globale de la magnétosphère (MATZKA et al. 2021). L’historique des valeurs de Kp remonte à 1932, ce qui nous donne aujourd’hui presque 90 ans de données tri-horaires.

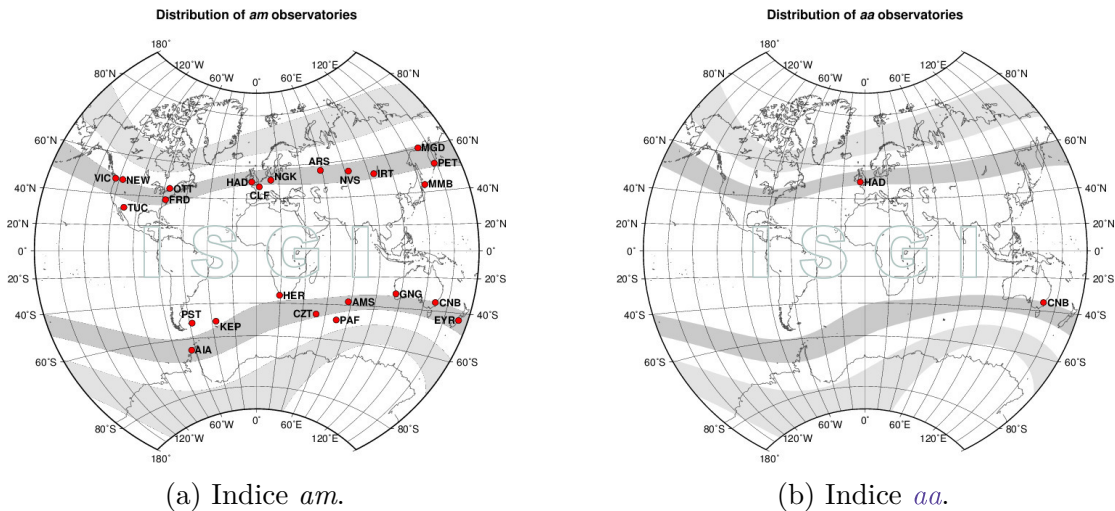


FIGURE 2.18 – Observatoires dont les mesures sont utilisées pour produire les indices am et aa . Crédits : ISGI.

Un autre indice planétaire, calculé à partir des mêmes stations, est l’indice ap . À la différence de Kp , il représente la mesure de perturbation du champ magnétique avant sa conversion en une échelle discrète. C’est donc un indice quasi-continu, s’exprimant en nanotesla. Les indices aa et am sont construits selon la même logique, mais à partir de stations différentes : 23 stations pour am contre seulement 2 stations pour aa (voir les Figures 2.18a et 2.18b). Ainsi, am se veut plus précis, tandis que aa , grâce à sa simplicité, propose un historique de données remontant à 1868 (soit plus de 150 ans de données), alors que am n’est disponible que depuis 1959 (MAYAUD 1968 ; MAYAUD 1971).

En outre, il existe des indices géomagnétiques qui ne sont pas dérivés de l’indice K , et représentatifs d’autres perturbations géomagnétiques. On peut notamment citer l’indice Dst , qui est également très utilisé dans la communauté. Il s’agit d’un indice horaire, calculé à partir de 4 observatoires répartis autour de l’équateur terrestre à faible latitude (voir Figure 2.19. C’est un indice prenant des valeurs continues, exprimées en nanotesla. C’est un indice principalement représentatif de l’intensité du courant annulaire (SUGIURA 1964). Lors des orages géomagnétiques, Dst prend des valeurs négatives, et il est souvent utilisé pour définir l’intensité d’un orage géomagnétique. Ses premières valeurs remontent à 1957.

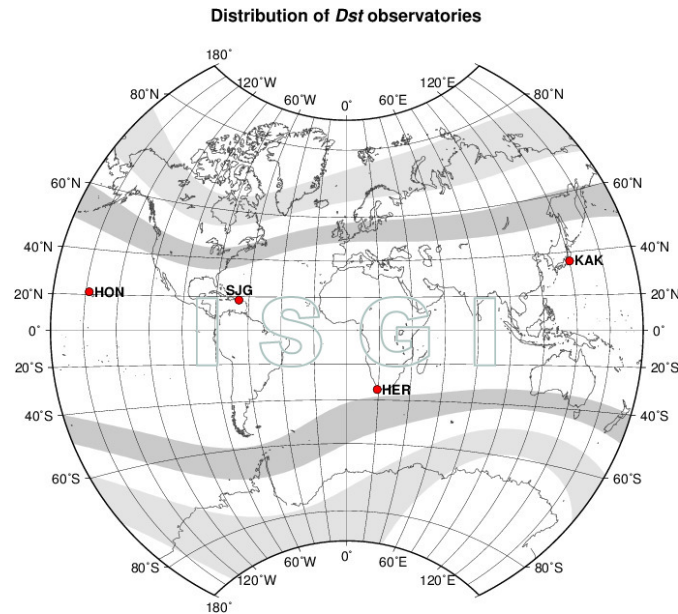


FIGURE 2.19 – Observatoires dont les mesures sont utilisées pour produire l’indice *Dst*. Crédits : ISGI.

Parmi les autres indices largement utilisés on peut citer les indices *AU*, *AL* et *AE*, qui sont des indices représentatifs des électrojets auroraux dans l’hémisphère nord (DAVIS et SUGIURA 1966 ; NOSE et al. 2015). On peut également citer les indices *PC*, signalant les perturbations magnétosphériques au niveau des calottes polaires (TROSHICHEV et al. 1979).

Les indices géomagnétiques sont des données présentant de nombreux avantages pour une utilisation dans un contexte de modélisation pilotée par les données. En effet, ce sont des séries temporelles mono-dimensionnelles, possédant le plus souvent une grande couverture temporelle, avec peu, voire pas de trous de données (grâce au fait qu’elles soient issues d’instruments au sol), et dont les données sont calibrées. Ce processus de vérification et de calibration peut d’ailleurs parfois être long, et certains indices géomagnétiques, comme *aa*, sont parfois disponibles dans leur version définitive seulement plusieurs mois, voire années après leur mesure. Toutefois, des versions *quick-look*, puis des versions temporaires, sont souvent fournies par les institutions produisant ces indices (pour des usages opérationnels).

De nombreuses études ont montré que les indices géomagnétiques sont directement corrélés d’un côté avec les paramètres du vent solaire mesurés en L1 (voir par exemple BOROVSKY 2020a), et de l’autre côté avec les populations de particules piégées dans les ceintures de radiations (par exemple ZHAO et LI 2013 ; XU et al. 2019).

Dans nos travaux nous avons principalement utilisé les indices *Dst*, *Kp* et *aa*. Nous avons utilisé *Dst* (voir le Chapitre 3) car il s’agit d’un indice très étudié dans la littérature, et qui a déjà fait l’objet de modélisations dirigées par les données. De plus, cet indice est souvent utilisé pour définir – parfois abusivement (voir BOROVSKY et SHPRITS 2017) – les périodes d’orage géomagnétique. Nous avons

également étudié Kp (voir les Chapitres 5 et 6) car il s’agit de l’indice planétaire le plus largement utilisé, à la fois pour définir des seuils d’alerte (MATZKA et al. 2021), mais aussi pour piloter des modèles physiques des ceintures de radiations (BOURDARIE et MAGET 2012). Enfin, un indice dérivé de aa a été utilisé (voir les Chapitres 4 et 6) car il s’agit de l’indice présentant le plus long historique, ce qui permet des analyses statistiques poussées, notamment en ce qui concerne les événements extrêmes.

Notons également que les indices Dst et Kp sont non seulement disponibles auprès des différents organismes les produisant et les distribuant, mais aussi dans la base de données OMNIWeb. L’indice aa est quant à lui disponible par exemple auprès de l’*International Service of Geomagnetic Indices* (ISGI).

En dernier mot, évoquons le fait que ces indices, notamment Dst et parfois Kp , utilisés conjointement aux paramètres du vent solaire, servent de référence pour définir des catalogues de périodes pendant lesquelles la magnétosphère terrestre est perturbée suite à des ICMEs (JIAN et al. 2006a; KILPUA et al. 2015; CHI et al. 2016) ou à des SIRs (JIAN et al. 2006b; KILPUA et al. 2015; CHI et al. 2018).

2.2.4 Mesurer les particules au cœur des ceintures

Depuis le début de l’ère spatiale, de très nombreuses missions spatiales embarquant des moniteurs de radiations se sont succédé, afin d’observer, comprendre, mesurer et, *in fine*, modéliser les ceintures de radiations terrestres. L’étude des ceintures de radiations est en effet particulièrement complexe. En effet, elle repose principalement sur l’analyse de données *in situ*. En plus des problèmes classiques que pose le recours aux missions spatiales, les satellites mesurant les flux de particules au cœur des ceintures sont justement très exposés à ces particules, et donc potentiellement victimes des effets de l’environnement spatial.

De plus, à cause de leur structure complexe, avec plusieurs régions dans lesquelles les phénomènes physiques dominants diffèrent, les ceintures de radiation ne peuvent pas être cartographiées par une unique mission. Il faut, pour avoir un aperçu le plus exhaustif possible, démultiplier les orbites et les gammes d’énergie mesurées par les moniteurs de radiations. Ainsi, à la difficulté déjà établie d’avoir une couverture temporelle exhaustive, se rajoute la difficulté de couvrir l’intégralité des orbites importantes (en prenant en plus en compte les asymétries jour/nuit par exemple) pour toutes les gammes d’énergie pertinentes.

Tout cela justifie l’existence (passée et présente) d’un nombre très important de missions embarquant un moniteur de radiations dédié à l’étude des ceintures de Van Allen. Les données produites par un très grand nombre de ces instruments sont disponibles à l’ONERA, dans la base de données *Ionizing Particle ONERA Database* (IPODE) (LAZARO et al. 2014). La Figure 2.20, qui répertorie certains des instruments dont les données sont disponibles dans IPODE, avec leur couverture spatio-temporelle, illustre bien la difficulté d’étudier ces données dans un contexte d’analyse statistique ou de modélisation pilotée par les données. En effet, les problèmes liés à l’intercalibration des données, qui ont souvent des attributs

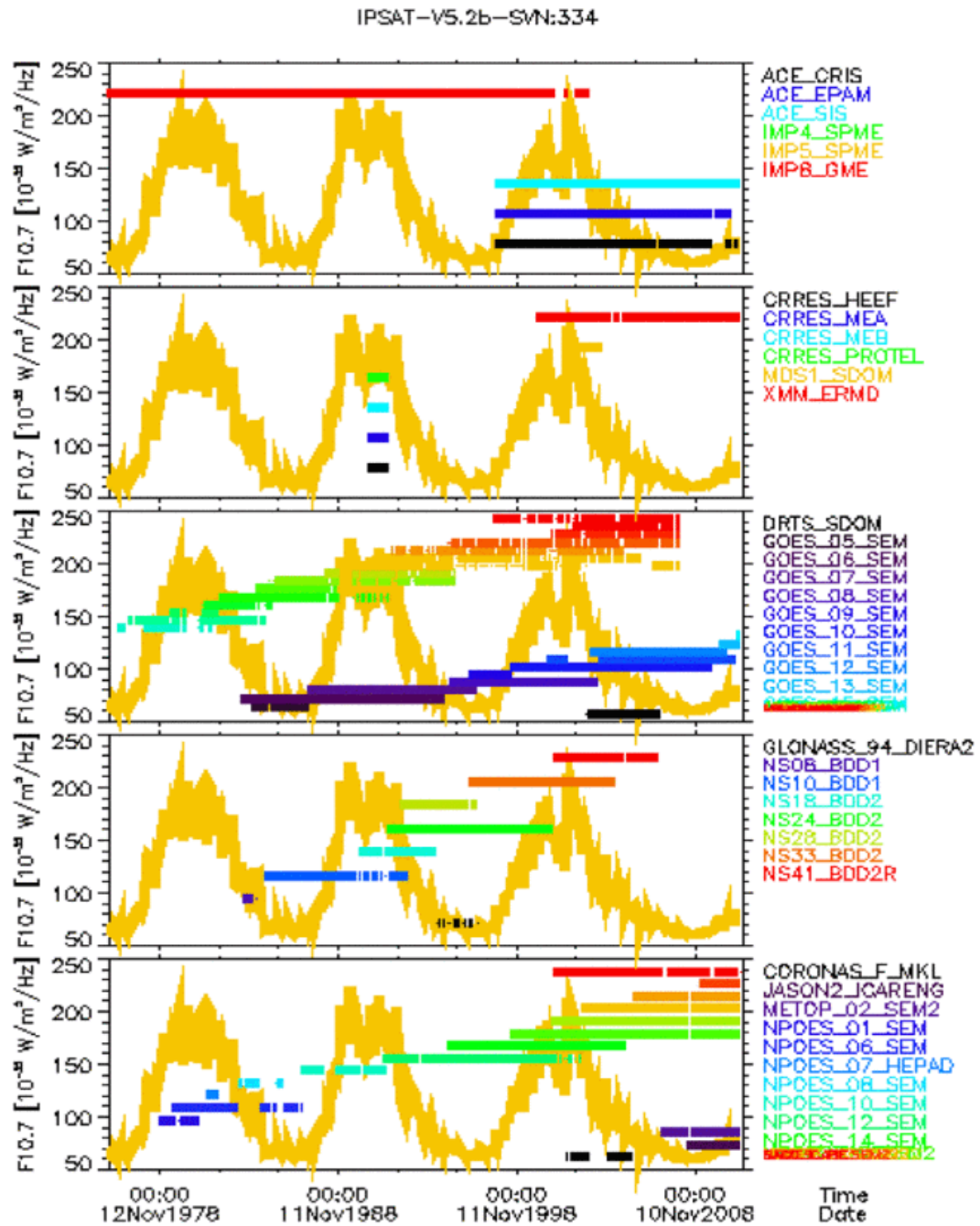


FIGURE 2.20 – Couverture temporelle des couples satellite/instrument disponibles dans IPODE entre 1968 et 2008. Le cycle solaire est représenté par le flux radio du Soleil à 10,7 cm (tracé jaune). De haut en bas, les missions correspondant aux orbites suivantes sont représentées : 1) missions interplanétaires, 2) HEO-GTO, 3) GEO, 4) MEO, 5) LEO-PEO. © 2013 IEEE. Reproduit, avec autorisation, de LAZARO et al. 2013.

très différents (rapport signal/bruit, fréquence d’acquisition, gammes d’énergie, directivité, etc.) deviennent très complexes. À cela se rajoutent, entre autres, les problèmes de contamination, quand un moniteur mesure, à tort, des flux correspondant en réalité à une autre espèce et/ou à une autre gamme d’énergie que celle prévue, comme c’est le cas par exemple avec le moniteur *Medium Energy Proton and Electron Detector* (MEPED) de l’instrument *Space Environment Monitor* (SEM), à bord du satellite NOAA POES-14, dont les mesures de flux de protons d’énergie $E > 2,5$ MeV sont contaminés par les électrons, comme on peut le voir sur la Figure 2.21.

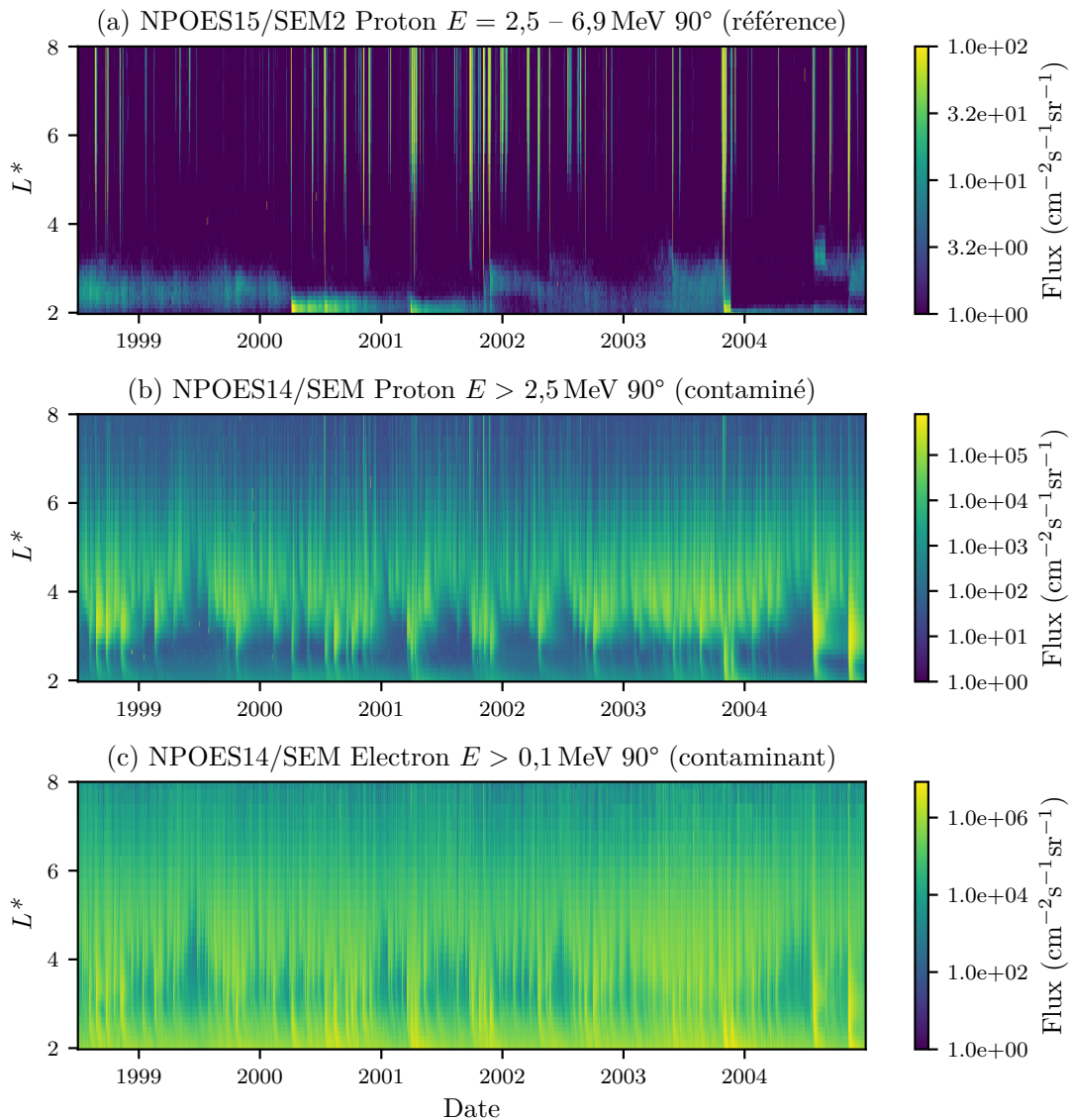


FIGURE 2.21 – Exemple de mesures de flux de protons non-contaminées (a) comparées à des mesures contaminées (b). Les flux d’électrons contaminant les mesures de protons sont données en (c).

Ces données de flux de particules au sein des ceintures de radiations sont très précieuses et elles nous permettent de mieux comprendre les processus physiques qui y sont à l’œuvre. Certaines difficultés mentionnées ci-dessus rendent leur utilisation

au sein de modèles pilotés par les données plus difficile que l'utilisation d'autres données (comme les indices géomagnétiques), bien que cela ne reste malgré tout pas impossible et même potentiellement indispensable en fonction des objectifs visés.

2.2.5 Bilan sur l'observation du système Soleil–Terre

Nous avons passé en revue les différentes sources de données permettant d'observer, d'étudier et d'analyser les interactions Soleil–Terre dans leur ensemble, depuis la couronne solaire jusqu'aux ceintures de radiations, en passant par le vent solaire. Ces données peuvent être des images mais également des séries temporelles scalaires.

Nous avons vu qu'une grande partie de ces données est obtenue à partir d'instruments embarqués à bord de missions spatiales, ce qui rend complexe la constitution de bases de données de longue durée et de qualité homogène. Les séries temporelles d'indices géomagnétiques et de paramètres du vent solaire disponibles sur OMNIWeb semblent être les données possédant les meilleures qualités pour une utilisation au sein d'un modèle d'apprentissage profond. Toutefois cela ne signifie pas qu'une utilisation des autres types de données pour des applications de *Deep Learning* est impossible, mais que des précautions supplémentaires sont alors à prendre.

2.3 Modéliser les interactions Soleil–Terre

Pour rappel, nos travaux ont pour objectif l'étude de la prédiction des perturbations géoeffectives de la magnétosphère terrestre. Dans cette section nous allons passer en revue différentes approches utilisées pour modéliser une partie des interactions Soleil–Terre d'intérêt pour la météorologie de l'espace, à savoir la propagation du vent solaire depuis la couronne solaire dans l'héliosphère, le couplage vent solaire–magnétosphère à une échelle macroscopique et la dynamique des ceintures de radiations terrestres.

Bien évidemment les interactions Soleil–Terre possèdent bien d'autres composantes. Toutefois nous ne les étudierons pas ici et nous nous concentrerons sur celles qui ont un intérêt direct pour notre thèse.

2.3.1 Des modèles pour la prédiction du vent solaire dans l'héliosphère

La question de la modélisation de la propagation du vent solaire, depuis la couronne solaire jusqu'à un point donné de l'héliosphère est abondamment traitée dans la littérature et est un des sujets recevant le plus d'intérêt au sein de la communauté de météorologie de l'espace. La modélisation de la propagation du

vent solaire (excluant les éjections de type CME) dans l'héliosphère peut se faire à l'aide de méthodes très différentes.

Tout d'abord, certains modèles utilisent des méthodes empiriques ou semi-empiriques. Parmi ces modèles, on peut distinguer deux grandes catégories : les modèles s'appuyant sur des magnétogrammes de la photosphère solaire, et ceux utilisant uniquement des méthodes statistiques appliquées à l'historique des paramètres du vent solaire mesurés en L1. Dans la première catégorie on peut citer le modèle WSA (ARGE et PIZZO 2000) ou encore le modèle HAF (FRY et al. 2001). Dans la deuxième catégorie on peut citer le modèle PDF (BUSSY-VIRAT et RIDLEY 2014), qui propose des prédictions de la vitesse du vent solaire en L1 jusqu'à 5 jours en avance sous la forme de distributions de probabilités, en exploitant notamment la permanence de structures du vent solaire entre deux rotations de Carrington. D'autres modèles empiriques utilisant des statistiques en L1 sont ceux de OWENS et al. 2017 et RILEY et al. 2017, qui utilisent des approches de correspondance avec des « jours similaires ».

Une autre catégorie de modèles sont les modèles physiques de MHD. Ces modèles sont plus coûteux mais permettent la prise en compte plus fine de la physique du vent solaire. Parmi les modèles de référence citons par exemple ENLIL (ODSTRCIL 2003) ou encore EUHFORIA (POMOELL et POEDTS 2018). Un autre modèle est le modèle *Solar Wind Flux Tubes* (SWIFT), qui combine bout à bout le modèle MULTI-VP qui propage le vent solaire de la couronne solaire jusqu'à 0,1 au (PINTO et ROUILLARD 2017), puis jusqu'à n'importe quel point de l'héliosphère interne à l'aide d'un second modèle comme par exemple le modèle HELIO1D⁹.

MACNEICE et al. 2018 compare les performances opérationnelles de ces modèles et parvient à la conclusion que les modèles les plus performants sont aujourd'hui les modèles empiriques, suivis par les modèles semi-empiriques qui devancent de peu les modèles physiques. Toutefois les modèles empiriques ne peuvent fournir que des prédictions en L1, alors que les modèles physiques permettent des prédictions dans d'autres régions de l'héliosphère. De plus, l'utilisation de modèles physiques permet de mieux comprendre la physique inhérente aux processus modélisés. Dans tous les cas MACNEICE et al. 2018 conclut que ces modèles sont assez efficaces pour prédire V_{sw} et ρ_{sw} , mais échouent globalement à prédire B_z ou la température du vent solaire T_{sw} .

Très récemment, en parallèle de nos travaux, un premier modèle d'apprentissage profond, combinant un CNN et un LSTM a été proposé par UPENDRAN et al. 2020 pour prédire la vitesse du vent solaire en L1 jusqu'à 5 jours en avance à partir d'images SDO/AIA à 193 et 211 Å. Selon les auteurs, ce modèle produirait des performances équivalentes aux modèles empiriques. S'agissant d'un premier modèle de ce type, dont les auteurs reconnaissent que beaucoup d'aspects (relatifs aux données et à l'architecture du modèle) peuvent être optimisés, cela laisse présager que l'utilisation de réseaux de neurones est une piste d'amélioration des prédictions du vent solaire en L1 très prometteuse.

9. <http://storms-service.irap.omp.eu/forecast>

Enfin il existe également des modèles de propagation des CMEs dans l’héliosphère. Beaucoup d’entre eux sont physiques, et reposent sur une modélisation MHD en 3D : SUSANOO-CME (SHIOTA et KATAOKA 2016), AWSoM-SWMF (JIN et al. 2017), MAS (TÖRÖK et al. 2018) et EUHFORIA (POEDTS et al. 2020). Toutefois ces modèles présentent de fortes limitations, notamment en ce qui concerne la prédiction de B_z (REISS et al. 2021). Ici encore les meilleures performances opérationnelles sont obtenues à l’aide de modèles empiriques (entre autres SALMAN et al. 2018 ; MÖSTL et al. 2018).

Récemment, de nouveaux modèles s’appuyant sur des méthodes de *Deep Learning* et utilisant des images EUV ou des magnétogrammes solaires proposent la prédiction non pas de paramètres du vent solaire au sein d’une ICME, mais des caractéristiques des ICMEs, comme leur temps d’arrivée sur Terre. Ces méthodes, qui sont aujourd’hui très prometteuses, seront discutées plus en détail en Section 5.1.

2.3.2 Prédiction d’indices géomagnétiques : des modèles empiriques au *Deep Learning*

Les premiers modèles proposant de modéliser la dynamique des indices géomagnétiques s’appuyaient sur des formules empiriques, faisant intervenir ou non le couplage vent solaire–magnétosphère. En ce qui concerne l’indice *Dst*, BURTON et al. 1975 proposa un algorithme de prédiction sous la forme d’une équation différentielle ordinaire pilotée par trois paramètres du vent solaire : sa vitesse V_{sw} , sa densité ρ_{sw} et la composante nord-sud de l’IMF B_z . IYEMORI et al. 1979 proposa un modèle linéaire de prédiction de l’indice *Dst* et des indices auroraux *AE*, *AU* et *AL* piloté par B_z .

Assez rapidement, des méthodes de prévision immédiate ou à court terme utilisant des réseaux de neurones ont émergé. Ainsi LUNDSTEDT et WINTOFT 1994 proposèrent l’utilisation d’un ANN pour prédire *Dst* avec une avance d’une heure, toujours à partir de V_{sw} , ρ_{sw} et B_z . GLEISNER et al. 1996 améliorèrent ce modèle en incluant également la dernière valeur de *Dst* parmi les entrées du modèle. (WU et LUNDSTEDT 1997) montrèrent qu’une architecture de RNN permettait d’obtenir de meilleurs résultats et étendirent leur horizon de prédiction à 6 heures. Ce dernier modèle fut encore amélioré pour des conditions opérationnelles par LUNDSTEDT et al. 2002.

En parallèle, COSTELLO 1998 puis BOBERG et al. 2000 proposèrent un modèle de prédiction de *Kp* utilisant des ANNs alimentés par les mêmes trois paramètres du vent solaire. WING et al. 2005 montrèrent également que des RNNs incluant en entrée les valeurs passées de l’indice *Kp* pouvaient améliorer les performances de prédiction et proposèrent des prédictions de *Kp* jusqu’à 4 heures en avance.

Au cours de la dernière décennie, avec l’augmentation de la puissance de calcul accessible aux chercheurs et la démocratisation des méthodes d’apprentissage automatique (avec la mise en place de bibliothèques informatiques dédiées) le nombre d’études portant sur la prédiction d’indices géomagnétiques à l’aide de réseaux de

neurones a grandement augmenté, et des modèles encore plus performants ont été proposés. Ainsi, BALA et REIFF 2012 rendirent opérationnels des prédictions de Kp , Dst et AE jusqu'à 6 heures en avance en utilisant des ANNs pilotés par les paramètres du vent solaire et un indice empirique approximant le potentiel de la calotte polaire. WINTOFT et al. 2017 proposèrent une étude portant sur l'influence de certains paramètres du vent solaire sur la prédiction de Kp 3 heures en avance avec un ANN.

Mais surtout, GRUET et al. 2018 puis TAN et al. 2018 montrèrent que les réseaux de type LSTM étaient capables d'améliorer notablement les performances de prédiction des indices Dst et Kp jusqu'à 6 heures en avance. Ces modèles utilisent un historique récent des paramètres du vent solaire (V_{sw} , ρ_{sw} et B_z pour GRUET et al. 2018, auxquels se rajoutent B_x , B_y , B , P_{sw} et des indices empiriques pour TAN et al. 2018) et de l'indice étudié pour fournir des prédictions jusqu'à 6 heures en avance. La particularité de TAN et al. 2018 est de proposer une prédiction en deux temps, avec un modèle qui prédit d'abord si on sera (ou non) en période d'orage géomagnétique (classification binaire), puis qui prédit une valeur de Kp avec un modèle entraîné différemment selon le cas (régression). De nombreuses études ayant suivi semblent confirmer que l'utilisation de LSTMs pour la prédiction d'indices géomagnétiques semble être aujourd'hui la méthode la plus performante (GRUET 2018; CHAKRABORTY et MORLEY 2020; LAPERRÉ et al. 2020; COLLADO-VILLAVERDE et al. 2021).

D'autres pistes ont récemment été explorées pour améliorer les résultats de prédictions de ces indices. Parmi celles-ci, l'utilisation d'autres données que les paramètres géomagnétiques, comme des mesures de perturbation du champ magnétique par les satellites GPS (GRUET et al. 2018) ou encore les flux de rayons X solaires mesurés par les satellites NOAA GOES (CHAKRABORTY et MORLEY 2020). Par ailleurs, une autre piste est la fourniture de prédictions probabilistes (et non plus déterministes), qui possèdent indubitablement une plus grande valeur ajoutée pour un utilisateur final. Pour cela, CHANDORKAR et al. 2017; GRUET et al. 2018; CHAKRABORTY et MORLEY 2020 ont montré que différentes variantes de processus gaussiens produisent de bonnes performances, tandis que CAMPOREALE et al. 2019 suggère l'utilisation d'un deuxième réseau de neurones pour prédire un intervalle de confiance qui compléterait une prévision déterministe préalablement obtenue. D'un autre côté, d'autres études continuent d'explorer l'utilisation de méthodes empiriques pour prédire Kp à court voire moyen terme, comme par exemple OWENS et al. 2017 qui proposent une méthode dite « analogue-ensemble » ou LUO et al. 2017 qui utilise certaines fonctions de couplages empiriques plus anciennes en les réadaptant.

Ce qui ressort de ce corpus est qu'aujourd'hui la méthode de prédiction d'indices géomagnétiques tels Kp et Dst la plus efficace est l'utilisation de réseaux récurrents de type LSTM, alimentés par un historique de valeurs passées récentes des paramètres du vent solaire mesurés au point de Lagrange L1 et de l'indice géomagnétique que l'on souhaite prédire. Une telle méthode fournit d'excellentes prédictions immédiates ou à court terme.

En revanche, d’autres questions restent encore ouvertes. Ainsi, des études comme GRUET 2018 ; SHPRITS et al. 2019 concluent que ces modèles perdent de leur efficacité pour des horizons de prédictions plus grands que 6 heures, ce qui semble en accord avec la physique du problème. Néanmoins, d’autres études comme LETHY et al. 2018 ; SEXTON et al. 2019 proposent des prédictions jusqu’à 12, voire 24 heures en avance, ce qui suggère soit une capacité des réseaux de neurones à extrapoler la dynamique de ces indices au-delà du temps de réaction de la magnétosphère déformée par le vent solaire, soit une erreur d’interprétation des résultats. De plus, les études mentionnées ici utilisent souvent des méthodologies d’évaluation différentes, en faisant usage d’ensembles de données d’évaluations différents, et de multiples métriques d’évaluation des performances, ce qui rend les comparaisons difficiles. Enfin, même si l’idée générale consistant à exploiter la capacité d’un réseau de neurones de type LSTM à traiter les séries temporelles fait consensus aujourd’hui, les différentes études citées utilisent toutes un *pipeline* différent. Les divergences peuvent se situer au niveau des données utilisées, de la manière dont elles sont pré-traitées, ou bien dans certains hyperparamètres du modèle, comme l’algorithme d’optimisation utilisé ou encore la dimension des états de la cellule LSTM. Tout cela suggère que la question de la prédiction immédiate ou à court terme des indices géomagnétiques est encore loin d’être parfaitement réglée. Enfin, comme le montre SHPRITS et al. 2019 pour l’indice Kp , plus l’horizon de prédiction est grand moins les modèles actuels sont satisfaisants, voire se font battre par des modèles extrêmement simples comme le modèle de récurrence (qui dans sa version la plus basique consiste à prédire la même valeur de Kp que celle observée une rotation de Carrington plus tôt).

2.3.3 Modélisation des ceintures de radiations

Aujourd’hui, la physique des ceintures de radiations est suffisamment bien comprise pour permettre une modélisation physique. Parmi les modèles existant on peut citer les modèles BAS du British Antarctic Survey (GLAUERT et al. 2014), DREAM3D du *Los Alamos National Laboratory* (LANL) (TU et al. 2013) ou encore le modèle SALAMMBÔ développé depuis plus de 20 ans à l’ONERA (BEUTIER 1993 ; BOURDARIE 1996 ; DAHMEN et al. 2020 ; DAHMEN 2020 ; BRUNET et al. 2021).

Le modèle SALAMMBÔ propose une modélisation des ceintures d’électrons mais aussi de protons. Salammbô s’appuie sur une résolution de l’équation de Fokker-Planck sur une grille à trois dimensions (E, y, L^*). Ce code permet la prise en compte de nombreux processus physiques, comme la diffusion radiale, les phénomènes de frictions et de collisions, les interactions onde-particules (à l’aide du code WAPI, SICARD-PIET et al. 2014) ou encore les *dropouts*¹⁰ (HERRERA et al. 2016). Ce modèle est piloté en partie par l’indice Kp , qui sert notamment à modéliser la localisation de la plasmopause et va donc diriger les phénomènes d’injection et de pertes de particules chargées. Notons également que ce modèle peut être utilisé conjointement à des outils d’assimilation de données, comme

10. À ne pas confondre avec la technique homonyme parfois utilisée lors de l’apprentissage d’un réseau de neurones.

l’assimilation directe (MAGET et al. 2008) ou le filtre de Kalman d’ensemble (ENKF) (MAGET 2007).

Bien que ce modèle représente l’état de l’art de la modélisation des ceintures de radiations, son utilisation dans un contexte de prédictions présente des limites. En effet, SALAMMBÔ est piloté par des paramètres comme Kp et des paramètres du vent solaire, et la résolution des équations nécessite des conditions initiales et limites. Pour utiliser SALAMMBÔ pour prédire l’état des ceintures de radiations, il faut donc à minima avoir des prédictions de Kp et des paramètres du vent solaire de qualité. De plus, ces modèles des ceintures de radiations sont coûteux en termes de ressources informatiques. Par conséquent, obtenir un résultat peut prendre beaucoup de temps, ce qui rend inutile l’utilisation de cette méthode pour des prédictions à court terme.

Afin de résoudre partiellement ces problèmes, certaines études très récentes proposent de modéliser (partiellement) les flux de particules dans les ceintures de radiations avec des algorithmes d’apprentissage statistique. Ainsi, GUERRA et al. 2020 propose un modèle, PreMeV 2.0 pour la prédiction de flux d’électrons d’énergie $E \geq 1$ MeV dans les ceintures de radiations avec un à deux jours d’avance, à partir de mesures de flux d’électrons et la vitesse du vent solaire. Par ailleurs, SMIRNOV et al. 2020 décrit le modèle MERLIN, qui permet la prédiction immédiate des flux d’électron d’énergie $E = 120 - 600$ keV dans la région *Medium Earth Orbit* (MEO) à partir de paramètres du vent solaire et d’indices géomagnétiques. La performance de ces modèles peut sans doute être encore améliorée, mais ils montrent que l’apprentissage automatique peut représenter une voie d’avenir pour la modélisation des ceintures de radiations.

2.4 Bilan du chapitre

L’apprentissage profond est un ensemble de méthodes permettent de modéliser empiriquement des fonctions complexes en exploitant de grandes quantités de données. Depuis le début de l’ère spatiale, de nombreuses missions se sont succédé pour observer et mesurer *in situ* les multiples composantes des interactions Soleil–Terre. Les données à notre disposition sont nombreuses, leurs natures sont diverses (images, séries temporelles, etc.) et elles proviennent de sources et d’instruments multiples. Cette richesse est certes un très grand atout, mais elle impose des contraintes spécifiques (gestion de l’inter-calibration, des données manquantes, des données contaminées, etc.) qui contraignent leur utilisation dans un contexte de modélisation dirigée par les données.

Depuis quelques années, en plus des modèles physiques et des modèles empiriques statistiques plus « traditionnels », on a assisté à une émergence d’études démontrant la grande efficacité des modèles d’apprentissage automatique – et notamment des réseaux de neurones – à modéliser le couplage vent solaire–magnétosphère. Ces modèles tirent justement leur force de la quantité et qualité des données d’observation du système Soleil–Terre disponibles. La qualité des prévisions d’indices géomagnétiques a en particulier été grandement améliorée pour de faibles horizons

de prédiction (environ six heures d'avance au maximum) à l'aide de modèles exploitant des réseaux récurrents de type LSTM, alimentés par des séries temporelles comme les paramètres du vent solaire. Cependant la plupart de ces modèles récents n'ont pas encore fait leurs preuves dans un contexte opérationnel.

Dans nos travaux nous souhaitons améliorer la qualité des prévisions des événements solaires géoeffectifs et en étendre leur horizon. Suite à notre revue bibliographique, il apparaît donc qu'une approche fondée sur la prédiction d'indices géomagnétiques à partir des paramètres du vent solaire mesuré en L1 est une des voies les plus prometteuses, au moins pour des horizons plus courts. À l'heure actuelle, s'intéresser directement aux flux de particules dans les ceintures de radiations semble être un problème plus complexe, dû au fait que ce sont des données de bien plus haute dimension, et dont l'acquisition repose sur des instruments *in situ* (contrairement aux magnétomètres au sol). De plus, des modèles physiques des ceintures de radiations comme Salammbô sont dirigés notamment par des indices géomagnétiques globaux, ce qui rend une approche reposant sur la prévision d'indices géomagnétiques d'autant plus pertinente. C'est pourquoi dans le chapitre suivant nous commencerons nos travaux par une étude du modèle de prédiction de l'indice *Dst* proposé par GRUET et al. 2018, qui fait partie de l'état de l'art actuel en la matière et à partir duquel nous proposerons un nouveau modèle amélioré que nous caractériserons de façon complète.

Chapitre 3

Prédire l'activité géomagnétique terrestre à partir du vent solaire

Another One

When you're a child
you learn
there are three dimensions:
height, width, and depth.
Like a shoebox.
Then later you hear
there's a fourth dimension:
time.
Hmm.
Then some say
there can be five, six, seven . . .

I knock off work,
have a beer
at the bar.
I look down at the glass
and feel glad.

(Ron Padgett)

Sommaire

Préambule	94
3.1 Reproduction du modèle de GRUET et al. 2018	95
3.1.1 Étude de l'indice <i>Dst</i>	95
3.1.2 Description du modèle de GRUET et al. 2018 et de notre reproduction	97
3.1.3 Résultats obtenus avec notre reproduction du modèle	100
3.2 Proposition d'un meilleur découpage de sous-ensembles de données	103
3.3 Améliorations du modèle de prédiction de <i>Dst</i>	107
3.3.1 Changements dans l'architecture et l'entraînement du réseau de neurones	107
3.3.2 Changements dans la gestion des données manquantes	111
3.4 Un nouveau banc d'évaluation pour mieux caractériser le modèle	113
3.4.1 Évaluation de la précision en période d'orage géomagnétique	115
3.4.2 Évaluation à l'aide de métriques dédiées au problème de classification binaire	116
3.4.3 Évaluation du modèle par détection des débuts d'orage	119
3.4.4 Évaluation des distorsions temporelles entre les prédictions et les observations	121
3.4.5 Application à des horizons de prédiction plus éloignés	123
3.4.6 Bilan sur l'extension du banc d'évaluation des prédictions de <i>Dst</i>	124
3.5 Obtenir des prédictions probabilistes directement avec un réseau de neurones	126
3.5.1 Description et justification de notre approche	126
3.5.2 Présentation et analyse des prédictions probabilistes	128
3.5.3 Bilan sur l'amélioration du modèle par la prise en compte de l'incertitude	134
3.6 Bilan du chapitre	136

Préambule

Afin d'atteindre les objectifs de cette thèse – améliorer la qualité des prédictions de l'activité géomagnétique terrestre et augmenter l'horizon de prédiction à quelques jours – nous commençons par reproduire le modèle proposé par GRUET et al. 2018 pour la prédiction de *Dst*. Ensuite nous proposons de multiples pistes d'améliorations originales à ce modèle, à la fois au niveau des données utilisées, de l'architecture du modèle et de son entraînement, mais aussi dans la façon d'évaluer les performances du modèle. Cela nous permettra de mieux caractériser – et donc comprendre – les performances que l'on peut obtenir avec ce type d'approche, par exemple en ce qui concerne l'horizon de prédiction maximal qu'on peut espérer atteindre avec cette méthode. Les méthodes que nous décrivons, développons et/ou utilisons dans ce chapitre nous serviront également dans la suite de nos travaux.

3.1 Reproduction du modèle de Gruet et al. 2018

Dans cette première section nous essayons de reproduire le plus fidèlement possible le modèle de prédiction de Dst proposé par GRUET et al. 2018. Le but ici est de disposer d'un modèle pouvant nous servir de référence tout au long du chapitre.

Nous avons choisi ce modèle en particulier comme point de départ de notre étude pour plusieurs raisons. Tout d'abord, au début de notre thèse en 2019 il s'agissait d'un des modèles constituant l'état de l'art de la prédiction à court terme de Dst (cela est toujours vrai à la date de rédaction de ce manuscrit), qui est, comme nous l'avons vu, l'un des indices les plus utilisés pour caractériser l'activité géomagnétique terrestre. Leurs auteurs ont en particulier montré que leur modèle était plus performant qu'un modèle de référence comme le modèle de persistance, mais aussi que d'autres modèles présents dans la littérature, comme ceux de LAZZÚS et al. 2017 et de BALA et REIFF 2012. Il s'agit de plus d'un modèle n'utilisant que des données publiques, téléchargeables par exemple depuis la base de données OMNIWeb. Enfin, ce modèle a été développé au cours d'une précédente thèse (GRUET 2018) ayant eu lieu à l'ONERA, ce qui nous a permis un accès plus détaillé au code utilisé pour construire et entraîner le modèle.

Dans cette section nous commençons par une courte étude statistique de Dst , afin que le lecteur qui n'est pas familier avec cet indice ait une meilleure idée de son comportement. Ensuite nous détaillons les caractéristiques du modèle de GRUET et al. 2018 avant d'expliquer comment nous l'avons implémenté et de commenter les résultats obtenus.

3.1.1 Étude de l'indice Dst

Pour rappel l'indice Dst est un indice horaire, exprimé en nanotesla (nT). Représentatif du courant annulaire, il est souvent utilisé pour définir les orages géomagnétiques. Il est en particulier très sensible aux perturbations causées par des ICMEs, bien que les perturbations causées par les CIRs l'affectent également (BOROVSKY et DENTON 2006). Dst peut prendre des valeurs négatives et c'est la valeur minimale de Dst atteinte lors d'un événement qui sert souvent à le caractériser.

Plusieurs seuils de Dst sont souvent utilisés dans la littérature pour caractériser l'intensité d'un orage géomagnétique. Par exemple, un seuil de -50 nT (KOSKINEN 2011a) est souvent utilisé pour caractériser des orages modérés. ACERO et al. 2018 estime, suite à une analyse statistique dédiée, qu'un seuil de -250 nT peut être retenu pour caractériser les orages extrêmes du point de vue de Dst . À titre d'illustration, la Figure 3.1 montre un exemple d'orage extrême survenu les 15-16 juillet 2000 et causé par une halo-CME observée le 14 juillet (cet orage est donc connu sous le nom de *Bastille Day geomagnetic storm* dans la littérature). Les seuils de -50 nT et -250 nT sont mis en valeur.

Certains repères statistiques relatifs à la distribution empirique des valeurs

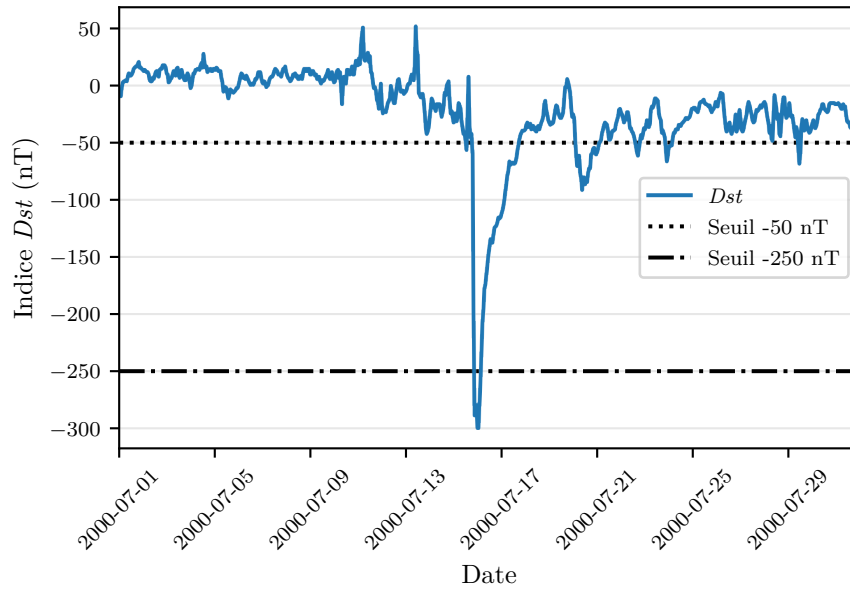


FIGURE 3.1 – Évolution de l'indice Dst lors de l'orage géomagnétique du 15-16 juillet 2000.

de Dst observées entre 1963 et 2018 (inclus) sont donnés dans le Tableau 3.1. On remarque que les valeurs de Dst sont majoritairement très concentrées autour de sa moyenne. Néanmoins les valeurs extrêmes s'en éloignent très fortement, en atteste la valeur minimale (représentant l'orage le plus intense) de Dst de -589 nT (atteinte le 13 mars 1989). Notons de plus que sur cette période, seulement 5,8 % des valeurs de Dst sont inférieures au seuil de -50 nT, et moins de 0,03 % sous le seuil de -250 nT.

TABLEAU 3.1 – Éléments statistiques de la distribution empirique des valeurs de Dst observées entre 1963 et 2018 (inclus).

Repère statistique	Valeur
Moyenne	$-14,5$ nT
Écart-type	$22,4$ nT
Minimum	-589 nT
5 ^e percentile	-52 nT
1 ^{er} quartile	-23 nT
Médiane	-10 nT
3 ^e quartile	-1 nT
95 ^e percentile	11 nT
Maximum	81 nT

Ce tableau et la précédente figure donnent une bonne idée du comportement de l'indice Dst , qui a tendance à rester très longtemps autour de valeurs faibles en temps calme et à soudainement prendre des valeurs bien plus fortes, mais pour

une courte durée, avant un retour rapide à l'équilibre. En conséquence, Dst est un indice qui présente une auto-corrélation linéaire forte à l'échelle de l'heure, mais qui diminue rapidement à l'échelle du jour, comme le montre la Figure 3.2. Cela n'est pas forcément le cas de tous les indices géomagnétiques, comme nous le verrons au Chapitre 4. Par ailleurs, on ne remarque pas de hausse de l'auto-corrélation autour de 27 jours, ce qui illustre bien que Dst est peu sensible aux structures récurrentes par rotation de Carrington de la couronne solaire à l'origine des vents solaires rapides et donc des $SIRs$ et $CIRs$.

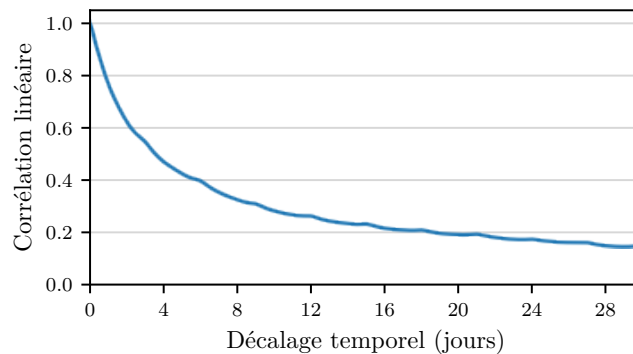


FIGURE 3.2 – Auto-corrélation temporelle de Dst pour des décalages temporels allant jusqu'à 30 jours.

Maintenant que nous avons analysé statistiquement Dst et illustré son comportement en période d'orage géomagnétique, nous allons décrire le modèle de prédiction proposé par GRUET et al. 2018.

3.1.2 Description du modèle de Gruet et al. 2018 et de notre reproduction

Pour rappel, le modèle proposé par GRUET et al. 2018 est un modèle de prédiction de l'indice Dst jusqu'à six heures en avance. Il s'agit d'un réseau de neurones utilisant une couche récurrente de type $LSTM$. Le modèle est dirigé par l'indice Dst lui-même et par quatre paramètres du vent solaire, à savoir sa vitesse V_{sw} , sa densité ρ_{sw} , la composante Nord-Sud du champ magnétique interplanétaire ou *Interplanetary Magnetic Field* B_z et sa norme $|B|$. Plus précisément, ce modèle est dirigé par un historique des valeurs de ces cinq séries temporelles. Nos données d'entrée sont donc de nature séquentielle, ce qui explique le choix d'utiliser un réseau $LSTM$. Les données utilisées sont les données horaires comprises entre le 14 janvier 2001 et le 31 décembre 2016.

Plus précisément, le réseau de neurones est composé d'une unique couche $LSTM$ dont la dimension des états est de 20, suivie d'une couche classique de type *fully-connected feedforward* (que nous appellerons couche « linéaire » dans la suite). La sortie du réseau de neurones est un scalaire représentant la prédiction de Dst à un instant $t + \tau$ futur, τ étant compris entre 1 et 6 h. Notons bien que le réseau ne propose de prédiction que pour un unique instant futur. Pour chaque

horizon de prédiction τ il faut donc entraîner un nouveau modèle (ce qui fait donc six modèles en tout).

L'algorithme d'optimisation utilisé est RMSProp¹. La fonction de perte minimisée lors de l'apprentissage est l'erreur quadratique moyenne ou *Mean Squared Error* (MSE). La précision du modèle par rapport aux vraies observations est déterminée à l'aide de deux métriques : la racine de l'erreur quadratique moyenne ou *Root Mean Squared Error* (RMSE) et la corrélation linéaire R .

En réalité le modèle proposé par GRUET et al. 2018 est un peu plus complexe, puisque les auteurs proposent d'utiliser également en entrée de leur modèle des mesures du champ magnétique terrestre faites par un satellite GPS, afin de l'améliorer légèrement. De plus, afin d'obtenir des prédictions probabilistes (*i. e.* la prédiction n'est plus un simple scalaire mais une distribution de probabilités), les auteurs proposent de combiner leur réseau de neurones à un modèle de processus gaussien ou *Gaussian Process* (GP). Par simplicité nous décidons pour l'instant de laisser de côté ces deux aspects, sur lesquels nous reviendrons plus tard.

Toutefois, de nombreux paramètres requis pour la reproduction du modèle ne sont pas indiqués explicitement dans l'article de GRUET et al. 2018. Par exemple, selon l'article, la relation entrées-sortie du modèle peut-être résumée ainsi :

$$\begin{aligned} \overline{Dst}(t + \tau) = \text{NN}(V_{\text{sw}}(t), \rho_{\text{sw}}(t), B_z(t), |B|(t), \\ Dst(t - 1 \text{ h}), Dst(t - 2 \text{ h}), \dots, Dst(t - 6 \text{ h})) \end{aligned} \quad (3.1)$$

où \overline{Dst} est la prédiction produite par le réseau de neurones NN. Le problème est que cette relation ne fait pas apparaître le fait que nous utilisons l'historique des valeurs passées des paramètres du vent solaire, mais uniquement de Dst (pour lequel la valeur à l'instant t n'est pas utilisée contrairement aux autres paramètres). En nous appuyant sur le manuscrit de thèse et le code à notre disposition nous en déduisons que le modèle est en réalité bien alimenté par des séries temporelles, de longueur 6, et que la relation entrées-sortie s'exprime mieux ainsi :

$$\begin{pmatrix} V_{\text{sw}}(t - 5) & \cdots & V_{\text{sw}}(t - 1) & V_{\text{sw}}(t) \\ \rho_{\text{sw}}(t - 5) & \cdots & \rho_{\text{sw}}(t - 1) & \rho_{\text{sw}}(t) \\ B_z(t - 5) & \cdots & B_z(t - 1) & B_z(t) \\ |B|(t - 5) & \cdots & |B|(t - 1) & T_{\text{sw}}(t) \\ Dst(t - 1 - 5) & \cdots & Dst(t - 1 - 1) & Dst(t - 1) \\ \vdots & \vdots & \vdots & \vdots \\ Dst(t - 6 - 5) & \cdots & Dst(t - 6 - 1) & Dst(t - 6) \end{pmatrix} \xrightarrow{\text{NN}} \overline{Dst}(t + \tau), \quad (3.2)$$

ce qui expliquerait pourquoi l'indice Dst est démultiplié dans l'équation 3.1. La Figure 3.3 schématise l'architecture du réseau de neurones de GRUET et al. 2018 tel que nous l'avons compris.

1. Cet algorithme a été proposé par Geoffrey HINTON dans un cours (http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf) et n'a pas fait l'objet d'une publication en soi, bien qu'il soit aujourd'hui largement utilisé et implémenté dans de nombreuses bibliothèques dédiées à l'apprentissage profond.

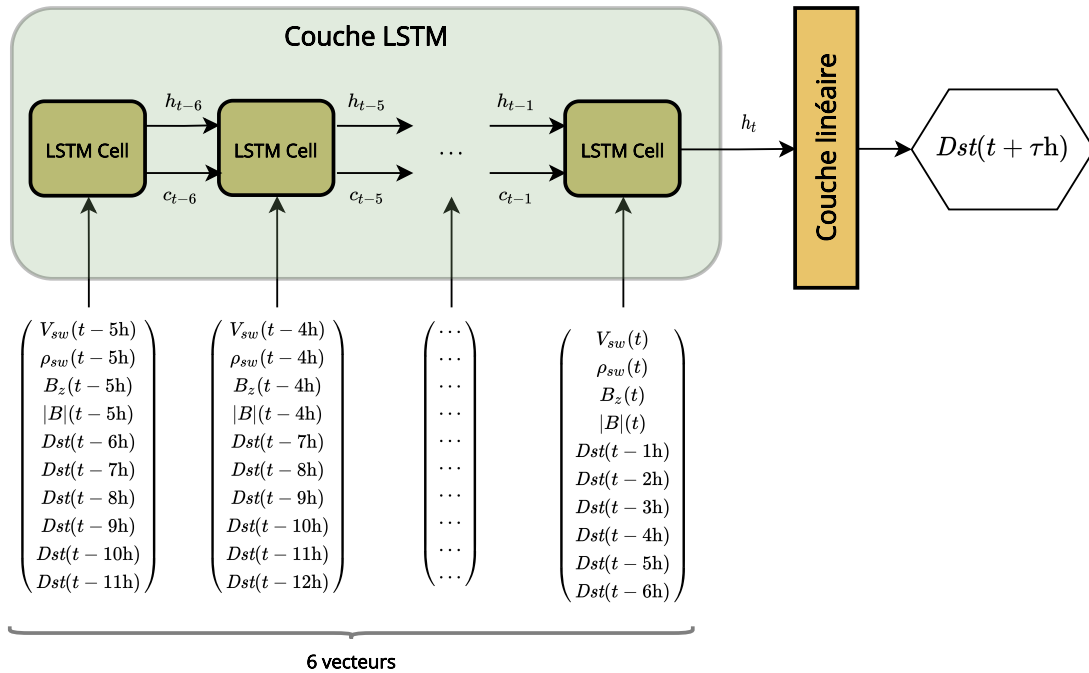


FIGURE 3.3 – Architecture de notre reproduction du réseau de neurones proposé par GRUET et al. 2018 pour la prédiction de Dst jusqu'à six heures en avance.

De nombreux autres paramètres et hyperparamètres² du modèle ne sont pas explicités dans l'article. On peut citer :

- les paramètres de l'algorithme d'optimisation RMSProp ;
- le nombre d'*epochs* d'apprentissage nécessaires à l'entraînement ;
- la taille des *mini-batches* d'apprentissage utilisés ;
- la méthode d'initialisation des poids et des états du réseau.

Notons que ces paramètres sont des paramètres qu'il faut ajuster manuellement, et que des valeurs différentes peuvent mener au même résultat final, la différence étant alors, par exemple, la vitesse de convergence de l'algorithme d'optimisation et donc la durée de l'entraînement. L'absence de ces paramètres dans l'article pour une reproduction rapide n'est donc pas bloquante. En nous appuyant sur le code à notre disposition à l'ONERA nous avons décidé d'utiliser les paramètres suivants :

- paramètres de l'algorithme d'optimisation : taux d'apprentissage (*learning rate*) $lr = 1e - 4$, alpha = 0,9, epsilon = 10^{-3} ;³
- l'apprentissage dure 200 *epochs* avec un arrêt précoce⁴ ;
- les *mini-batches* contiennent 300 séquences chacun ;
- tous les états du réseau sont initialisés avec une valeur nulle, les poids sont initialisés en suivant une loi de probabilités uniforme entre -1 et 1.

2. Le terme « hyperparamètres » désigne ici les paramètres d'ajustement des algorithmes d'apprentissage automatique.

3. Cf. <https://pytorch.org/docs/stable/generated/torch.optim.RMSprop.html>.

4. L'arrêt précoce consiste à arrêter l'apprentissage du réseau dès que des signes explicites de surapprentissage sont observés.

Enfin, des aspects plus importants relatifs à la gestion et au traitement des données sont omis dans l'article. Ainsi, il n'est pas clair comment les données sont séparées en sous-ensembles d'entraînement, de validation et d'évaluation, ni comment les données manquantes sont traitées. Il n'est pas question non plus de l'utilisation (ou non) d'une méthode pour pré-traiter les données. Or, nous avons vu en section 2.1.6, que s'agissant de méthodes de modélisation dirigée par les données, ces aspects-là sont très importants. Concernant la séparation des données en sous-ensembles, il est seulement mentionné dans l'article que 70 % des données servent à entraîner, 10 % à valider et 20 % à évaluer le modèle. À l'aide des supports à notre disposition il semblerait que le découpage des données soit tout simplement séquentiel (les sous-ensembles se suivent) de la façon suivante :

- entraînement : du 14/01/2001 au 18/03/2012 (97 960 valeurs horaires) ;
- évaluation : du 18/03/2012 au 28/05/2015 (27 990 valeurs horaires) ;
- validation : du 28/05/2015 au 31 décembre 2018 (13 996 valeurs horaires).

Cela représente un total de 139 944 valeurs horaires, c'est-à-dire environ 5000 de plus ($\approx 4\%$) que les 134 398 valeurs horaires mentionnées dans l'article. Nous n'avons pas réussi à expliquer cette différence, qui ne devrait pas avoir une grande influence. Quant aux données manquantes, il semblerait qu'elles soient remplacées par la valeur moyenne (calculée sur l'ensemble d'entraînement) de la série temporelle considérée. Enfin, en ce qui concerne le pré-traitement, les entrées et sorties sont re-échelonnées de telle sorte que leur moyenne soit égale à 0 et leur écart-type à 1. Les coefficients pour effectuer ces transformations sont calculés uniquement à partir de l'ensemble de données d'entraînement afin de ne pas inclure de biais lors des étapes de validation et de test. Cette procédure est standard lorsqu'on travaille avec des réseaux récurrents (SHERSTINSKY 2020).

Enfin nous avons implémenté ce modèle à l'aide du langage de programmation Python (v. 3.8). Nous avons en particulier utilisé les bibliothèques PyTorch (v. 1.9.1 PASZKE et al. 2019), scikit-learn (v. 0.23.1 PEDREGOSA et al. 2011), NumPy (v. 1.19.2 HARRIS et al. 2020) et Pandas (v. 1.0.5 WES MCKINNEY 2010). Nous avons entraîné ce modèle sur une machine disposant d'un processeur de type GPU NVIDIA Titan V. Chaque *epoch* d'apprentissage dure environ 2,5 s, ce qui nous donne une durée d'entraînement totale (par horizon de prédiction) inférieure à 10 minutes.

3.1.3 Résultats obtenus avec notre reproduction du modèle

Nous avons évalué notre reproduction du modèle de GRUET et al. 2018 à l'aide des mêmes métriques d'évaluation (RMSE, R) que dans l'article original. Les résultats pour les deux modèles sont donnés dans le Tableau 3.2.

Tout d'abord on remarque que nous avons effectivement réussi à reproduire un modèle de prédiction de l'indice Dst . Cependant, on ne retrouve pas tout à fait les mêmes valeurs de métriques d'évaluation. Par exemple, nous obtenons systématiquement des valeurs de RMSE plus faibles, la différence étant de l'ordre de $\approx 10 - 25\%$ avec notre reproduction du modèle, ce qui est notable. En revanche,

TABLEAU 3.2 – Comparaison des performances de prédiction de Dst pour plusieurs horizons de prédiction obtenues avec le modèle original de GRUET et al. 2018 et avec notre reproduction du modèle.

Horizon de prédiction (heures)	GRUET ET AL. 2018		Notre reproduction du modèle	
	RMSE (nT)	R	RMSE (nT)	R
1	5,94	0,966	4,35	0,970
2	6,65	0,946	5,82	0,945
3	7,86	0,923	7,08	0,917
4	8,86	0,902	8,05	0,892
5	9,59	0,882	8,76	0,871
6	10,24	0,865	9,34	0,852

à l'exception de l'horizon $\tau = 1$ h notre reproduction du modèle donne des valeurs de corrélation linéaire légèrement moins bonnes (l'écart est inférieur à $\approx 2\%$). Ces écarts s'expliquent certainement par le fait que nous avons dû estimer ou deviner de nombreux paramètres du modèle. Bien que nos mesures d'évaluation soient globalement meilleures (la RMSE est même meilleure que pour le modèle de GRUET et al. 2018 utilisant les données GPS), il ne serait pas pertinent de conclure immédiatement que notre implémentation du modèle est plus efficace. En effet, bien que nous ayons fait de notre mieux pour reproduire le plus fidèlement possible le modèle, une légère incertitude demeure dans le découpage des données utilisées (notamment pour l'évaluation). Or, sans certitudes sur cet aspect-là, nous ne pouvons conclure définitivement.

La Figure 3.4 montre des exemples de prédiction obtenus avec notre reproduction du modèle, pour l'orage géomagnétique du 23-24 avril 2012 (situé dans notre ensemble d'évaluation). Cette figure illustre bien la légère dégradation de la qualité de la prédiction lorsque l'horizon de prédiction augmente. De plus, elle semble aller dans le sens des conclusions de l'article, estimant que le modèle est efficace jusqu'à des horizons de prédiction d'environ six heures.

Toutefois, comme nous le verrons dans les sections suivantes, il est possible d'améliorer le modèle et son *pipeline* à de nombreux égards. Dans la suite de ce chapitre nous comparerons les nouveaux modèles que nous développerons à notre reproduction du modèle de GRUET et al. 2018 (et non pas au modèle original) car nous en maîtrisons parfaitement les paramètres. C'est pourquoi nous appellerons ce modèle le « modèle de base ».

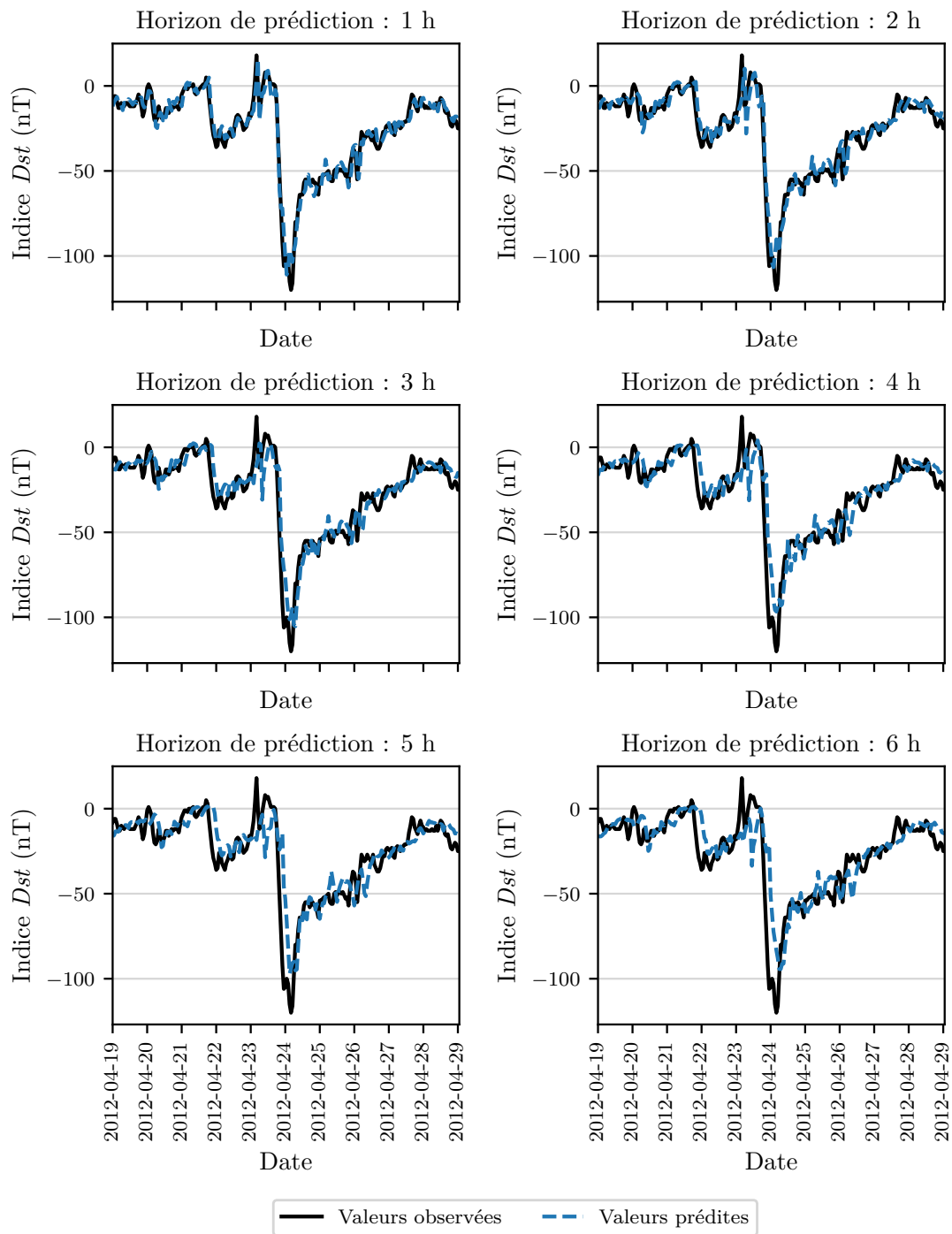


FIGURE 3.4 – Exemples de prédictions jusqu’à six heures en avance obtenues avec notre reproduction du modèle de GRUET et al. 2018 lors d’un orage géomagnétique survenu fin avril 2012.

3.2 Proposition d'un meilleur découpage de sous-ensembles de données

Tout d'abord nous proposons ici de modifier le découpage des données utilisées. En effet, dans la section précédente nous utilisons les valeurs des paramètres du vent solaire (et l'indice *Dst*) comprises entre 2001 et 2016. Comme ici nous n'utilisons pas les données GPS nous pouvons utiliser des données plus anciennes que 2001. De plus, ce travail ayant démarré en 2019, nous utilisons les données observées jusqu'à fin 2018. Une analyse rapide de la base de données OMNIWeb montre qu'avant la mise à disposition des données des satellites WIND et ACE à partir de 1995 et 1998, la base de données OMNIWeb possède un pourcentage annuel élevé de données manquantes (dépassant certaines années les 60 %). C'est pourquoi nous nous restreignons aux données comprises entre 1995 et 2018 (inclus). Durant cette période il y avait en moyenne 2,41 % de données manquantes.

La période choisie correspond approximativement aux cycles solaires 23 et 24. Or ces deux cycles ont présenté une activité très différente. La Figure 3.5 montre les diagrammes en boîtes de la distribution annuelle des valeurs de *Dst* entre 1995 et 2018. Pour chaque boîte, sont indiqués successivement le premier décile, le premier quartile, la médiane, le troisième quartile et le neuvième décile. Les valeurs inférieures au premier décile et supérieures au neuvième décile sont également indiquées par de petits traits. On remarque tout de suite qu'en fonction de l'année l'indice *Dst* peut prendre des valeurs très différentes et avoir une distribution plus ou moins resserrée autour de ses quartiles. De plus, il apparaît clairement que les années incluses dans le cycle solaire 24 présentent des valeurs de *Dst* globalement plus faibles qu'au cycle solaire 23. En particulier il apparaît que *Dst* n'a jamais atteint des valeurs extrêmes au cours du cycle solaire 24, alors qu'il l'a fait à de nombreuses reprises au cours de quatre années différentes du cycle solaire 23.

Or, afin d'entraîner un réseau de neurones correctement, il faut s'assurer que l'ensemble d'entraînement soit statistiquement et physiquement représentatif des différents types de perturbations géomagnétiques existantes. Toutefois, il faut aussi de préférence que l'ensemble d'évaluation soit représentatif des mêmes événements. En effet, il est bien plus facile pour un modèle d'être performant en périodes calmes, au cours desquelles la valeur de *Dst* fluctue légèrement au cours de sa moyenne. Or, ces périodes sont bien plus nombreuses que les périodes d'orage géomagnétique. Il faut donc faire attention, afin de pouvoir proposer une évaluation juste et équitable du modèle, à ce que le sous-ensemble de test utilisé présente assez de périodes d'orages (modérés et intenses).

Alors que la période d'entraînement utilisée par GRUET et al. 2018 semble satisfaisante, la période d'évaluation pourrait être meilleure. C'est pourquoi nous avons choisi de séparer nos données par années, avec la répartition suivante, représentée en Figure 3.6 :

- entraînement : de 2003 à 2018 compris ;
- validation : de 1995 à 1996 compris ;
- évaluation : de 1997 à 2002 compris.

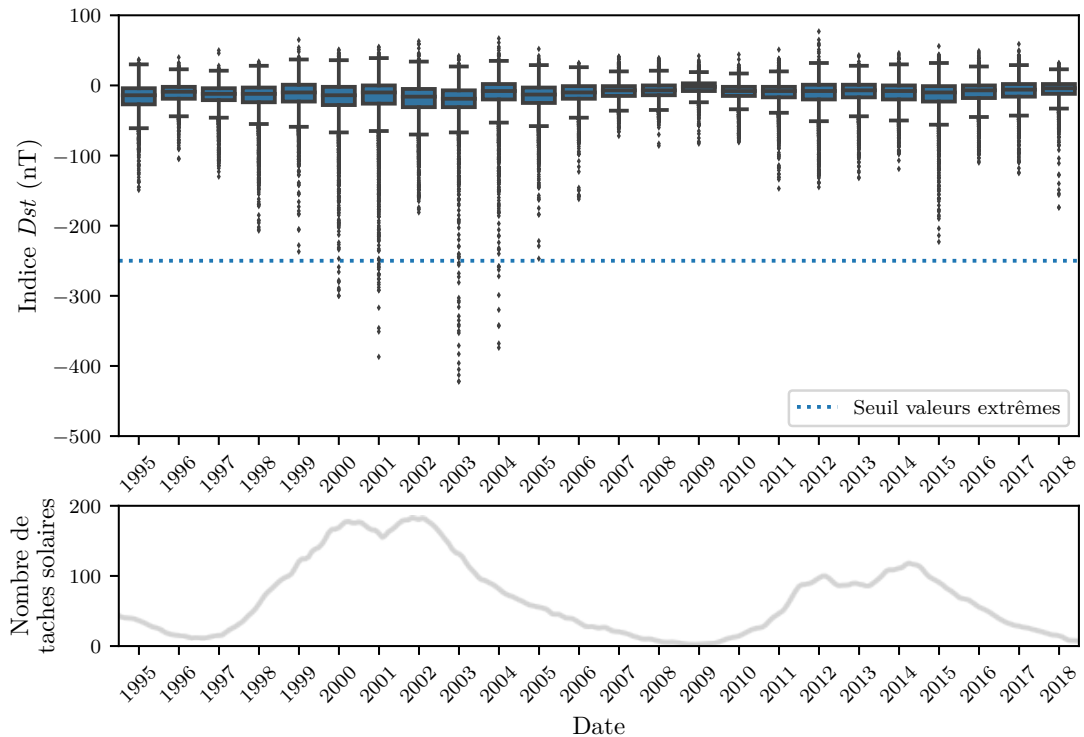


FIGURE 3.5 – Statistiques annuelles de l'indice Dst entre 1995 et 2018 sous la forme de digrammes en boîtes. Le nombre de taches solaires lissé sur 13 mois est également donné comme indicateur du cycle solaire.

Nous proposons un découpage des données par années (et non par mois) afin de ne pas biaiser l'apprentissage et l'évaluation avec des variations saisonnières ou semi-annuelles (KATSAVRIAS et al. 2021), mais aussi car les données utilisées peuvent être fortement auto-corrélées entre deux rotations de Carrington à cause des structures solaires persistantes.

Avec cette répartition, nous avons des orages géomagnétiques extrêmes dans notre ensemble d'entraînement, mais aussi dans notre ensemble d'évaluation. Avec ce nouveau découpage, nous avons 140 249 échantillons dans l'ensemble d'entraînement, 17 532 dans l'ensemble de validation et 52 578 dans l'ensemble d'évaluation. De plus, avec le nouveau découpage 7,4 % des valeurs de l'ensemble de test sont inférieures à -50 nT (seuil d'orage modéré), contre 3,2 % avec le précédent découpage. Nous avons donc ré-entraîné et évalué le modèle de base avec ce découpage des données. Les résultats sont donnés dans le Tableau 3.3.

On constate que le nouveau découpage des données produit des valeurs de $RMSE$ moins bonnes que l'ancien découpage. Cela ne signifie pas pour autant que le nouveau découpage produit un modèle moins performant. Il s'agit plutôt d'une conséquence de la différence entre les distributions des valeurs Dst , au cours de la période allant du 18/03/2012 au 28/05/2015 d'un côté, et de celle allant du 01/01/1997 au 31/12/2002 de l'autre. En particulier, la $RMSE$ est une métrique qui pénalise plus les erreurs lorsqu'elles sont commises sur des valeurs plus élevées (en valeur absolue), à cause du passage au carré de l'erreur. Pour compenser ce

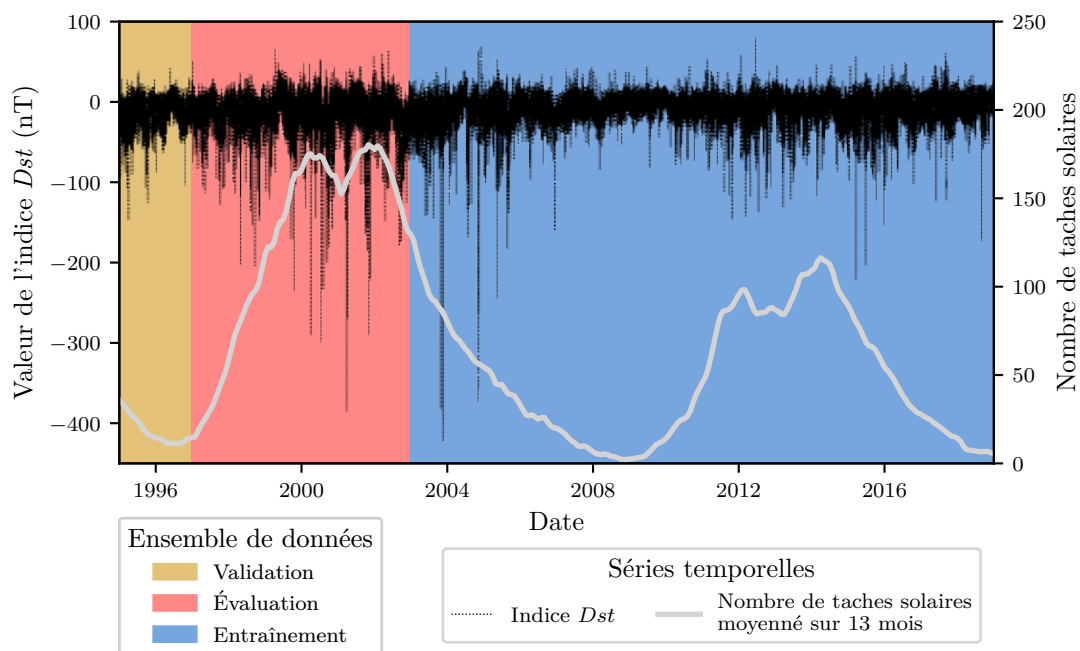


FIGURE 3.6 – Courbe des valeurs prises par l'indice Dst entre 1995 et 2018 inclus (ligne noire fine), les ensembles d'entraînement (zone bleue), de validation (zone jaune) et d'évaluation (zone rouge) sont mis en évidence. Le nombre de taches solaires lissé sur 13 mois est également représenté comme indicateur du cycle solaire (ligne grise épaisse).

biais, il est possible d'utiliser plutôt la *racine de l'erreur quadratique moyenne normalisée* ou *Normalized RMSE (NRMSE)*. Le facteur de normalisation que nous utilisons ici est l'écart-type⁵ de Dst , calculé séparément dans chacun des sous-ensembles d'évaluation (celui-ci vaut 17,8 nT avec l'ancien découpage et de 25,2 nT avec le nouveau). En utilisant la *NRMSE* on se rend compte que nous obtenons des valeurs meilleures avec le nouveau découpage.

De même, les valeurs de corrélation linéaire sont légèrement meilleures, mais la différence est faible. Le fait qu'elles ne soient pas dégradées comme la *RMSE* corrobore notre analyse selon laquelle le modèle n'est pas devenu moins performant.

En Figure 3.7 nous montrons des exemples de prédictions obtenues pour les six horizons de prédiction lors de l'orage extrême de juillet 2000 (le même qu'en Figure 3.1) et d'un orage modéré qui a suivi. Ces orages, et notamment l'orage extrême, sont maintenant dans notre ensemble de test. On observe un peu plus nettement la dégradation de la qualité des prédictions obtenues avec l'augmentation de l'horizon de prédiction sur cet exemple.

Dans la suite de ce chapitre nous conserverons donc ce découpage des données, que nous estimons plus représentatif.

5. D'autres méthodes de normalisation de *RMSE* existent, par exemple à l'aide de la moyenne. Avec Dst cela ne convient pas car c'est un indice qui peut prendre des valeurs positives et négatives.

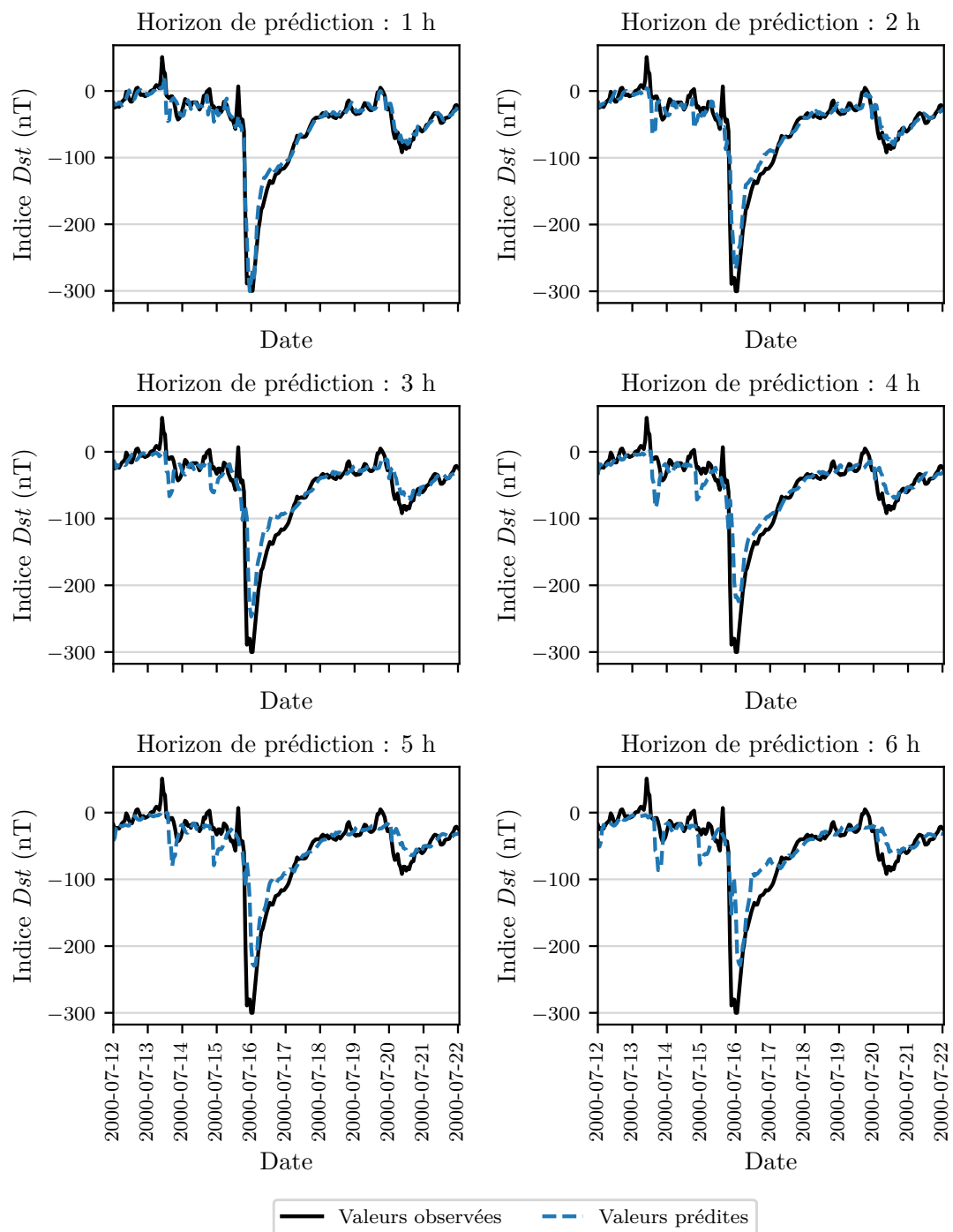


FIGURE 3.7 – Exemples de prédictions jusqu'à six heures en avance obtenues avec notre reproduction du modèle de GRUET et al. 2018 lors de l'orage géomagnétique extrême de juillet 2000, suivi d'un orage modéré.

TABLEAU 3.3 – Comparaison des performances de prédiction de *Dst* obtenues avec le modèle de base entraîné et évalué d’abord avec l’ancien découpage des données puis avec le nouveau, pour des horizons de prédiction allant de 1 – 6 h.

Horizon de prédiction (heures)	Modèle de base Ancien découpage			Modèle de base Nouveau découpage		
	RMSE (nT)	NRMSE	<i>R</i>	RMSE (nT)	NRMSE	<i>R</i>
1	4,35	0,245	0,970	5,80	0,231	0,973
2	5,82	0,327	0,945	7,67	0,305	0,953
3	7,08	0,398	0,917	9,55	0,380	0,926
4	8,05	0,453	0,892	10,77	0,428	0,905
5	8,76	0,493	0,871	11,91	0,474	0,884
6	9,34	0,525	0,852	12,90	0,513	0,860

3.3 Améliorations du modèle de prédiction de *Dst*

Dans cette section nous analysons la construction du modèle de base et nous en proposons des améliorations de différentes natures menant à de meilleures performances.

3.3.1 Changements dans l’architecture et l’entraînement du réseau de neurones

Les premières voies d’amélioration que nous explorons ici sont directement liées au réseau de neurones, que ce soit au niveau de son architecture, de sa méthode d’entraînement, ou encore de son implémentation pratique. En effet, l’apprentissage automatique est un domaine actuellement très actif et de nouvelles méthodes à l’efficacité démontrée émergent constamment. De plus, la puissance de calcul à notre disposition a également beaucoup augmenté au cours de ces dernières années. Comme l’optimisation des hyperparamètres d’un réseau de neurones se fait le plus souvent par recherche empirique et itérative, le fait d’accélérer le processus d’entraînement (grâce à la puissance de calcul plus élevée) permet de faire des recherches d’hyperparamètres plus fines.

Nous ne détaillerons pas l’ensemble de tous les modèles que nous avons testés, cela serait fastidieux et peu intéressant (bien que cette étape ait été longue). Nous allons tout simplement passer en revue l’ensemble des caractéristiques du modèle que nous avons modifiées, en justifiant les changements, avant de donner les résultats obtenus avec le meilleur modèle vers lequel nos efforts ont convergé.

Propriétés du LSTM Nous utilisons ici un réseau possédant deux couches LSTM (au lieu d'une seule auparavant), ayant chacune des états de dimension 256 (contre seulement 20 dans le modèle de base). Cela nous permet d'avoir un réseau plus complexe, que nous pouvons entraîner grâce à l'utilisation d'un *Graphics Processing Unit* (GPU) puissant comme le Nvidia Titan V à notre disposition. Entre les deux couches LSTM nous utilisons une couche de type *dropout* (BALDI et SADOWSKI 2013). Le *dropout* consiste à « désactiver » les neurones d'une couche du réseau (*i. e.* rendre leur valeur de sortie nulle) de façon aléatoire à chaque *epoch* avec une probabilité $p_{dropout}$ pour chaque neurone. Cette technique permet d'éviter un surapprentissage du modèle. Ici nous utilisons une valeur de $p_{dropout} = 0,3$.

Méthode d'optimisation Nous n'utilisons plus l'algorithme d'optimisation RMSPprop, mais l'algorithme Adam (KINGMA et BA 2017), construit pour être plus efficace que RMSProp, et qui est aujourd'hui sans doute l'algorithme le plus utilisé dans la communauté d'apprentissage profond, aux côtés de l'algorithme de descente du gradient stochastique⁶.

On utilise une valeur de *learning rate* initiale $lr = 10^{-4}$, qui est divisée par 2 toutes les 10 *epochs* jusqu'à l'*epoch* 60 inclus. Notre apprentissage se fait sur 120 *epochs*, avec arrêt précoce.

De plus on utilise une méthode de régularisation. La régularisation consiste à pénaliser la fonction de perte à l'aide d'une valeur dépendant des poids du réseau de neurones, afin de forcer le réseau de neurones à avoir des poids les plus petits possibles pour éviter un surapprentissage (MORADI et al. 2020). Ici nous utilisons une régularisation L_2 . La régularisation L_2 consiste à ajouter la somme au carré des poids du réseau (avec une constante multiplicative) à la fonction de perte. Nous utilisons une constante multiplicative ayant pour valeur $wd = 5 \times 10^{-5}$.

Données utilisées Dans le modèle de base nous utilisons six séquences distinctes de *Dst*, décalées d'une heure chacune. Ce choix avait probablement été fait pour aider le réseau à mieux prendre en compte les valeurs passées de *Dst*. Or l'architecture LSTM a justement été développée pour que nous n'ayons pas à procéder ainsi mais plutôt en fournissant une seule séquence pour chaque série temporelle, pour laquelle le réseau serait capable de retenir les informations importantes à court et à long terme. De plus nous avons rendu notre réseau plus complexe, donc théoriquement capable de mémoriser plus d'informations. C'est pourquoi nous n'utilisons plus qu'une seule séquence de *Dst*.

Par ailleurs, dans le modèle de base la séquence de *Dst* s'arrêtait à un instant $t - 1$, tandis que les séquences des paramètres du vent solaire s'arrêtaient à t . Or cette distinction n'a pas lieu d'être, car le WDC Kyoto Observatory (qui calcule les valeurs officielles de *Dst*) fournit des valeurs prévisionnelles de *Dst* de type

6. Voir par exemple le post de blog d'Andrej Karpathy « *A Peek at Trends in Machine Learning* » publié en 2017 (<https://karpathy.medium.com/a-peek-at-trends-in-machine-learning-ab8a1085a106>).

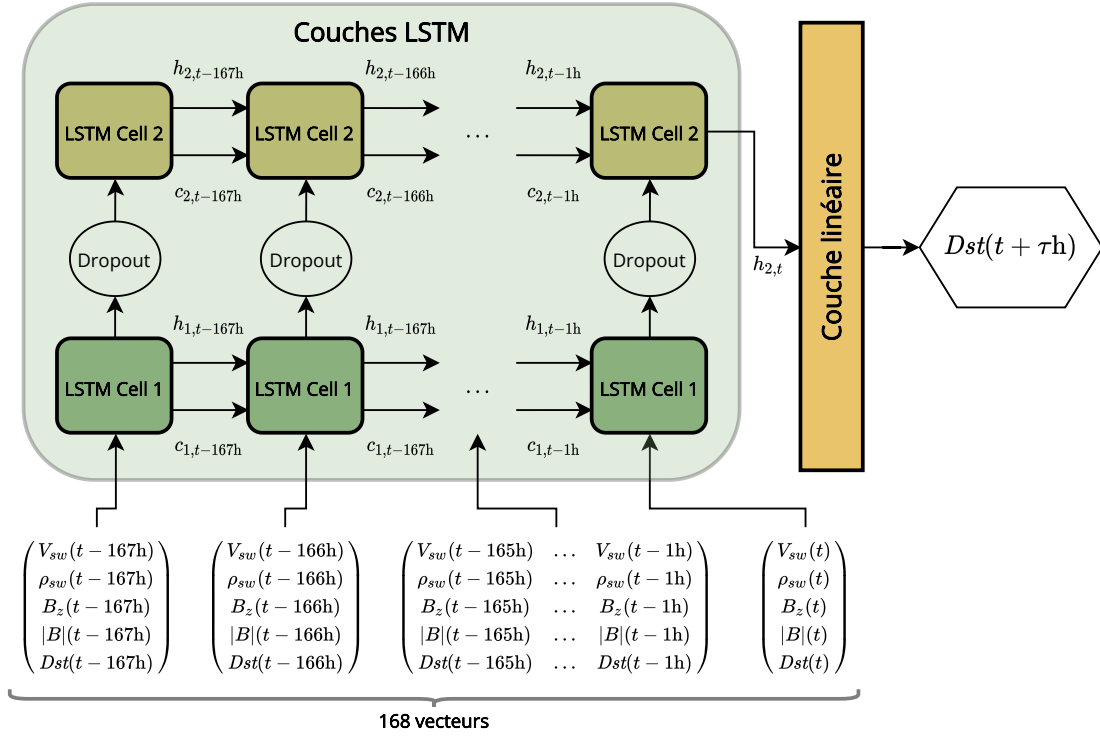


FIGURE 3.8 – Architecture du réseau de neurones amélioré que nous proposons pour prédiction de Dst jusqu'à six heures en avance.

quick-look en temps quasi-réel. C'est pourquoi nous prenons les valeurs de Dst jusqu'à l'instant t . Les relations entrées-sorties de notre nouveau modèle peuvent se résumer ainsi :

$$\begin{pmatrix} V_{sw}(t-l+1) & \dots & V_{sw}(t-1) & V_{sw}(t) \\ \rho_{sw}(t-l+1) & \dots & \rho_{sw}(t-1) & \rho_{sw}(t) \\ B_z(t-l+1) & \dots & B_z(t-1) & B_z(t) \\ |B|(t-l+1) & \dots & |B|(t-1) & T_{sw}(t) \\ Dst(t-l+1) & \dots & Dst(t-1) & Dst(t) \end{pmatrix} \xrightarrow{NN} \overline{Dst}(t + \tau), \quad (3.3)$$

où l désigne la longueur de la séquence. Dans le modèle de base, $l = 6h$. Ici nous exploitons nos capacités de calcul augmentées et utilisons une longueur $l = 168h$ (c'est-à-dire 7 jours). Étant donné la décroissance de l'auto-corrélation de Dst pour des longueurs plus grandes et les temps caractéristiques de corrélation de Dst avec les paramètres du vent solaire (BOROVSKY 2020a), il est inutile ici de prendre des séquences plus longues. La Figure 3.8 schématise l'architecture de notre modèle « amélioré ».

Avec tous ces nouveaux changements, chaque *epoch* d'entraînement dure maintenant $\approx 25s$ et l'entraînement complet dure donc 50 minutes (par horizon de prédiction). En revanche, le temps d'inférence est quasiment inchangé.

Résultats Les résultats obtenus avec notre nouveau modèle sont donnés dans le Tableau 3.4, dans lequel ils sont comparés aux résultats obtenus avec le modèle

de base. Comme les deux modèles ont été entraînés, validés et évalués en suivant le nouveau découpage des données, nous pouvons les comparer directement. Pour chaque horizon de prédiction nous avons mis un gras la valeur de **RMSE** et de corrélation linéaire obtenues avec le meilleur modèle.

TABLEAU 3.4 – Comparaison des performances de prédiction de *Dst* pour plusieurs horizons de prédiction obtenues avec le modèle de base et avec notre nouveau modèle, tous deux entraînés et évalués avec le nouveau découpage des données.

Horizon de prédiction (heures)	Modèle de base		Nouveau modèle	
	RMSE (nT)	<i>R</i>	RMSE (nT)	<i>R</i>
1	5,80	0,973	3,90	0,988
2	7,67	0,953	6,70	0,964
3	9,55	0,926	8,87	0,937
4	10,77	0,905	10,42	0,912
5	11,91	0,884	11,73	0,889
6	12,90	0,860	12,87	0,862

Note : Les résultats obtenus avec le meilleur modèle sont indiqués en gras.

On remarque que les deux métriques d'évaluation sont améliorées pour tous les horizons de prédiction. Cette amélioration est particulièrement forte pour un horizon de prédiction d'une heure (par rapport au modèle de base, la **RMSE** décroît d'environ 33% et la corrélation linéaire augmente de 1,5%). Toutefois cette amélioration devient de moins en moins marquée à mesure que l'horizon de prédiction augmente, pour devenir quasi nulle pour un horizon de prédiction de six heures.

L'amélioration des performances par rapport au modèle de base est le résultat conjoint de l'utilisation d'un modèle plus complexe, optimisé à l'aide d'algorithmes plus performants, et exploitant au mieux les données dont on dispose. Le fait que l'amélioration des performances soit plus marquée pour les horizons de prédiction courts suggère que le nouveau modèle, plus complexe, a été capable de mieux « apprendre » le couplage vent solaire–magnétosphère, dont l'information est contenue dans les paramètres du vent solaire qui s'apprête à immédiatement perturber la magnétosphère terrestre. En effet, les paramètres du vent solaire de la base OMNIWeb sont projetés au nez de la magnétopause, donc le temps de propagation restant du vent solaire jusqu'à la magnétosphère est quasi-nul. L'avance que nous pouvons avoir avec nos prédictions est donc imputable au temps de réaction de la magnétosphère au forçage du vent solaire puis éventuellement aux temps caractéristiques de retour à l'équilibre de la magnétosphère. Les processus de transfert d'énergie du vent solaire à la magnétosphère étant rapides (de l'ordre de l'heure selon par exemple AKASOFU 1981), ce sont eux que le réseau de neurones va principalement modéliser pour nos prédictions à très court terme. Comme ces processus sont très complexes, il n'est pas surprenant que l'amélioration du

modèle de base ait pour conséquence une amélioration notable de la performance prédictive.

Pour les horizons de prédiction plus éloignés, le fait que le modèle de base produise des performances similaires au modèle plus complexe suggère que la quantité d'information contenue dans les entrées du modèle (par rapport à la sortie espérée) est relativement faible, et donc que la complexité supplémentaire du modèle amélioré est moins utile que pour les horizons plus courts. Ces résultats indiquent donc que l'information d'une éventuelle perturbation du vent solaire impactant la magnétosphère n'est pas encore contenue dans les paramètres du vent solaire six heures en avance, ce qui est en accord avec la physique du problème.

3.3.2 Changements dans la gestion des données manquantes

Avoir un jeu de données bien préparé est essentiel lorsqu'on entraîne un réseau de neurones. La gestion des données manquantes (trous de données) en est un des aspects essentiels. En effet, en remplaçant les données manquantes on peut introduire deux types d'erreurs. Premièrement, on risque de modifier la distribution statistique des données, et donc biaiser l'entraînement. Deuxièmement, nous risquons d'inclure des valeurs qui n'ont pas de sens physique. Or le réseau de neurones va s'appuyer sur ces données pour comprendre les relations entre les entrées et les sorties. Une mauvaise gestion des données manquantes peut donc mener à des performances de prédiction dégradées. Notons que la question de la gestion des données manquantes est assez fréquemment oubliée par les auteurs d'études portant sur la prédiction d'indices géomagnétiques à partir des paramètres du vent solaire (par exemple chez BALA et REIFF 2012 ; CHANDORKAR et al. 2017 ; LAZZÚS et al. 2017 ; LETHY et al. 2018 ; SEXTON et al. 2019 ; SHPRITS et al. 2019, et d'autres.)

Comme nous l'avons vu en section 3.2, depuis 1995, il manque en moyenne 2,41 % de données dans les séries temporelles des paramètres du vent solaire. Cette absence de données s'explique certainement par des incidents sporadiques, soit au niveau de l'acquisition des données par les instruments embarqués, soit au niveau de la télécommunication avec les stations sol. Les causes de ces défaillances peuvent être multiples et il n'est pas à exclure qu'elles puissent avoir pour origine un événement solaire. La majeure partie de ces « trous » sont de l'ordre de l'heure ou de quelques heures. Toutefois, il existe des trous aussi longs que trois ou quatre jours. Il faut ainsi prévoir une stratégie adéquate pour les gérer.

Dans GRUET et al. 2018, les auteurs remplacent les données manquantes par la valeur moyenne de la série concernée. Nous avons vu qu'avec cette stratégie assez simple il est déjà possible d'obtenir de bons résultats de prédiction. Toutefois cette méthode n'est pas très satisfaisante statistiquement ou physiquement – d'autant moins si le trou de données est dû à un événement solaire, auquel cas on aurait certainement mesuré des valeurs de paramètres du vent solaire éloignées de la moyenne.

Afin de gérer les données manquantes, la méthode la plus simple respectant les

exigences statistiques et physiques consiste à ne pas utiliser les séquences contenant des données manquantes, comme le font LAPERRE et al. 2020. Cette méthode est possible lorsqu'on a un assez grand nombre d'échantillons (ce qui est notre cas), mais aussi si la longueur L des séquences utilisées est assez faible (une seule donnée manquante peu entraîner le rejet de L séquences). C'est par exemple le cas chez LAPERRE et al. 2020 (où les auteurs utilisent $L = 6$, comme pour le modèle de base), mais pas ici où nous utilisons désormais une longueur $L = 168$.

Enfin la méthode qui semble actuellement la plus répandue dans la communauté, et utilisée dans des études faisant partie de l'état de l'art dans leurs domaines respectifs, consiste à interpoler linéairement les données manquantes. Cette interpolation peut se faire de manière directe (voir par exemple COLLADO-VILLAVERDE et al. 2021), ou bien de façon un peu plus subtile, en exploitant l'auto-corrélation forte d'une heure à l'autre des paramètres du vent solaire (TAN et al. 2018 ; CHAKRABORTY et MORLEY 2020). Toutefois, comme le relèvent CHAKRABORTY et MORLEY 2020, cette méthode n'est physiquement pertinente que pour les trous de données de longueur réduite (moins d'une demi-journée).

Cependant, d'autres méthodes plus sophistiquées existent, comme celle proposée par KONDRASHOV et al. 2010. Cette méthode repose sur l'analyse spectrale singulière ou *Singular Spectrum Analysis* (SSA), une méthode d'estimation spectrale non-paramétrique conçue pour rendre compte de la dynamique sous-jacente de séries temporelles (GHIL et al. 2002). Il s'agit ici de reconstruire les trous temporels de séries de données en exploitant les corrélations spatio-temporelles, pour mettre en avant des modes cohérents tout en rejetant le « bruit ». La SSA a déjà été utilisée pour remplir les trous de données dans la base de données OMNI-web dans une étude portant sur la modélisation empirique du champ magnétique terrestre, permettant à leurs auteurs d'obtenir de meilleurs résultats qu'avec une interpolation linéaire (KONDRASHOV et al. 2014). Il n'est pas pour autant garanti que l'utilisation de la SSA pour remplacer les données manquantes améliore le modèle dans notre cas.

C'est pourquoi nous avons entraîné deux nouveaux modèles, l'un remplaçant les données manquantes par une simple interpolation linéaire, l'autre par la méthode SSA. Les détails concernant le fonctionnement de celle dernière et sa mise en œuvre dans notre cas sont donnés en Annexe A. Les résultats obtenus avec ces méthodes sont donnés dans le Tableau 3.5, où ils sont comparés aux résultats précédemment obtenus (où les données manquantes étaient remplacées par la valeur moyenne).

Il apparaît que la méthode de gestion des données manquantes ait un impact non-négligeable sur l'entraînement du modèle et donc sur ses performances. La méthode de reconstruction par SSA produit quasi-systématiquement les meilleurs résultats, suivie de peu par l'interpolation linéaire puis de la méthode utilisée dans le modèle de base. Toutefois, l'amélioration qui en résulte est relativement faible (inférieure à 2 % pour la RMSE). Nous pouvons en conclure que l'utilisation d'une méthode de gestion des trous de données plus sophistiquée permet d'améliorer les résultats, mais que cette amélioration reste faible dans le cas où peu de données sont manquantes (pour rappel, environ 2,4 % dans notre cas).

TABLEAU 3.5 – Comparaison des performances de prédiction de Dst obtenues avec le modèle amélioré pour trois méthodes de gestion des données manquantes, pour des horizons de prédiction de 1 – 6 h.

Horizon de prédiction (heures)	Sans interpolation		Interpolation linéaire		Reconstruction SSA	
	RMSE (nT)	R	RMSE (nT)	R	RMSE (nT)	R
1	3,90	0,988	3,85	0,988	3,83	0,988
2	6,70	0,964	6,65	0,965	6,64	0,965
3	8,87	0,937	8,87	0,937	8,87	0,938
4	10,42	0,912	10,44	0,913	10,27	0,914
5	11,73	0,889	11,60	0,889	11,62	0,889
6	12,87	0,862	12,75	0,865	12,67	0,867

Note : Les résultats obtenus avec le meilleur modèle sont indiqués en gras.

La Figure 3.9 montre des exemples de prédictions obtenues avec le modèle amélioré (entraîné avec les données manquantes reconstruites par SSA) lors de l'orage extrême de juillet 2000. Les prédictions obtenues avec le modèle de base sont également indiquées à titre comparatif. On remarque que dans cet exemple, le modèle amélioré est capable de prédire la bonne amplitude de l'orage extrême pour tous les horizons de prédiction, ce qui n'était pas le cas avec le modèle de base.

Jusqu'ici nous avons exclusivement évalué le modèle avec la RMSE et la corrélation linéaire. Cependant, ces métriques ne peuvent rendre compte exhaustivement du comportement de notre modèle en toutes circonstances. Or, il est très important d'évaluer un modèle de façon exhaustive et le caractériser dans les détails, afin de comprendre son comportement dans toutes les situations. Tout d'abord, cela permet de comparer plus finement deux modèles. Cela permet également de pouvoir éventuellement déployer un modèle en contexte opérationnel en connaissant précisément ses qualités et défauts – conditions nécessaires à une prise de décision éclairée et donc utile. Ainsi, il est inutile de posséder un modèle très performant s'il n'est pas correctement évalué et caractérisé. En outre, dans le contexte de cette étude, cela permettra d'apporter des éléments de réponse précis pour répondre à notre problématique. C'est pourquoi dans la suite nous proposons de nouvelles façons d'évaluer notre modèle.

3.4 Un nouveau banc d'évaluation pour mieux caractériser le modèle

La RMSE et la corrélation linéaire indiquent des tendances générales. Or, la plupart du temps, l'activité géomagnétique est calme, donc les périodes calmes vont avoir un poids plus fort dans le calcul de ces métriques que les périodes

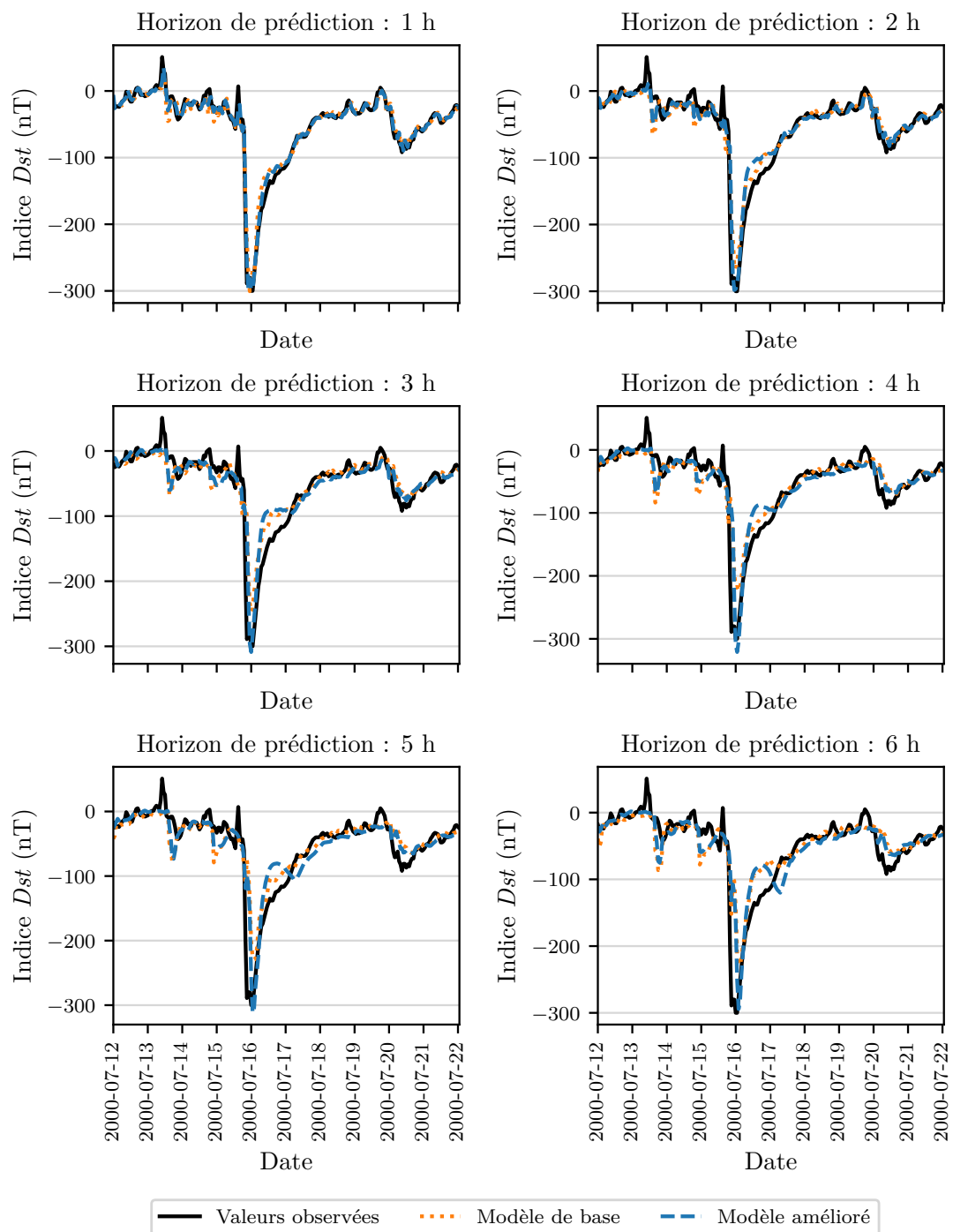


FIGURE 3.9 – Exemples de prédictions jusqu'à six heures en avance obtenues avec notre modèle amélioré lors de l'orage géomagnétique extrême de juillet 2000, suivi d'un orage modéré.

d'activité intense, ce qui peut créer un biais. Il est certes intéressant de prédire de façon précise l'indice Dst au cours des périodes calmes, mais aussi (et surtout) au cours des périodes d'orages géomagnétiques.

De plus, il faut non seulement que nos prédictions soient précises au cours des périodes d'activité intense, mais il faut aussi parvenir à détecter ces périodes d'activité intense en avance. Or, comme le soulignent LIEMOHN et al. 2018, la $RMSE$ et la corrélation linéaire ne permettent pas d'évaluer cet aspect du modèle.

Dans cette section nous proposons donc de compléter notre banc d'évaluation du modèle de prédiction de Dst à l'aide de différentes méthodes, que nous détaillons successivement.

3.4.1 Évaluation de la précision en période d'orage géomagnétique

Une méthode simple et pourtant efficace pour évaluer la précision de notre modèle en période d'orage géomagnétique consiste à séparer l'ensemble d'évaluation en périodes calmes et périodes actives et utiliser les mêmes métriques que précédemment. Afin de ne pas biaiser nos résultats à cause des valeurs numériques que prend l'indice Dst en période d'orage géomagnétique, il ne convient pas d'utiliser la $RMSE$ simple mais plutôt la $NRMSE$. Le facteur de normalisation que nous utilisons ici est toujours l'écart-type de Dst , calculé séparément dans chacun des sous-ensembles d'évaluation.

La Figure 3.10 montre les valeurs de $NRMSE$ et de corrélation linéaire pour chacun des six horizons de prédiction obtenues en séparant notre ensemble de test en trois sous-ensembles : $Dst > -50$ nT (périodes calmes), -150 nT $< Dst \leq -50$ nT (orages modérés) et $Dst \leq -150$ nT (orages intenses). Les performances du modèle de base et du modèle amélioré sont comparées.

Sans grande surprise on constate que les modèles sont plus performants en période calme qu'en période d'orage. De même, les modèles apparaissent plus précis en période d'orage modéré plutôt qu'intense. Cette figure nous apprend en revanche que le domaine dans lequel le modèle amélioré est notablement meilleur que le modèle de base est celui des orages intenses. On y constate que les valeurs de corrélation linéaire sont bien meilleures pour le modèle amélioré que pour le modèle de base, alors que dans les autres domaines (périodes calmes et orages modérés) cet écart est bien moindre. Ainsi le modèle amélioré est non seulement meilleur, mais il apporte un surplus de précision surtout en période d'orage intense, ce qui est un gain très précieux d'un point de vue opérationnel.

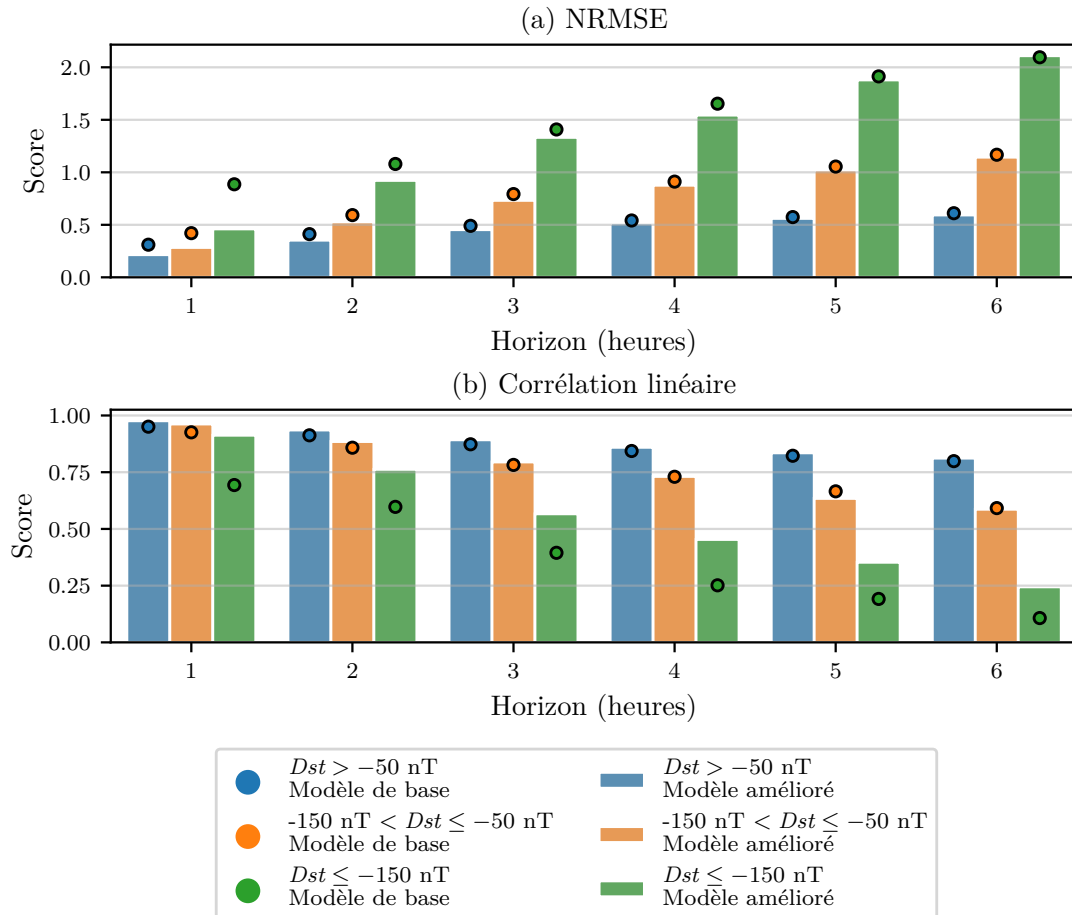


FIGURE 3.10 – Comparaison des valeurs de (a) **NRMSE** et de (b) corrélation linéaire obtenues avec le modèle de base et le modèle amélioré jusqu'à six heures en avance lorsqu'on sépare l'ensemble de test en trois sous-ensembles (périodes calmes, orages modérés et orages intenses).

Note : Sur cette figure et dans toute la suite du manuscrit, le terme « Score » désigne simplement la valeur d'une mesure d'évaluation atteinte par un modèle.

3.4.2 Évaluation à l'aide de métriques dédiées au problème de classification binaire

Maintenant que nous avons évalué la précision de notre modèle en périodes calmes et d'orage, il serait intéressant d'évaluer sa capacité à distinguer ces deux types de périodes. En effet, il est souvent utile pour un opérateur de satellite cherchant à protéger sa mission de bénéficier d'un système d'alertes plutôt que d'une valeur brute de Dst . Le seuil pour déclencher une alerte peut dépendre de nombreux facteurs (architecture du satellite, orbite, instruments embarqués, etc.). Dans cette étude et en l'absence d'exemple concret, nous utilisons le seuil de -50 nT pour définir une alerte binaire : si notre modèle prédit que Dst chute en dessous de ce seuil alors nous émettons une alerte, sinon non.

Il existe plusieurs métriques dédiées à l'évaluation du problème de classifica-

tion binaire. Celles-ci s'appuient sur les quatre composantes de la « matrice de confusion », à savoir le nombre de vrais positifs ou *True Positives* (TP), vrais négatifs ou *True Negatives* (TN), faux positifs ou *False Positives* (FP) et faux négatifs ou *False Negatives* (FN). À partir de ces quantités nous pouvons définir les mesures suivantes :

- la *précision* : répond à la question « quelle proportion d'alertes émises sont pertinentes ? » Ses valeurs sont comprises entre 0 et 1 (le plus haut le meilleur). Elle s'obtient par la formule :

$$\text{précision} = \frac{\text{TP}}{\text{TP} + \text{FP}};$$

- le *rappel* – parfois appelé *probabilité de détection* ou *Probability of Detection* (POD) : répond à la question « quelle proportion de périodes d'orages sont détectées ? » Ses valeurs sont comprises entre 0 et 1 (le plus haut le meilleur). Il s'obtient par la formule :

$$\text{rappel} = \frac{\text{TP}}{\text{TP} + \text{FN}};$$

- le F_{score} : il s'agit de la moyenne harmonique de la *précision* et du *rappel*. Ses valeurs sont comprises entre 0 et 1 (le plus haut le meilleur). Il s'obtient par la formule :

$$F_{\text{score}} = 2 \cdot \frac{\text{précision} \times \text{rappel}}{\text{précision} + \text{rappel}};$$

- le *taux de fausses alertes* ou *False Alarm Rate* (FAR) – parfois appelé *probabilité de fausse détection* ou *Probability of False Detection* (POFD) : il s'agit de la proportion de non-événements pour lesquels une alerte a été faussement émise. Ses valeurs sont comprises entre 0 et 1 (le plus bas le meilleur). Il s'obtient par la formule :

$$\text{FAR} = \frac{\text{FP}}{\text{FP} + \text{TN}};$$

- le *Threat score* (TS) : donne une indication de la quantité de vrais orages détectés, en pénalisant à la fois les fausses alertes et les faux négatifs, sans prendre en compte les vrais négatifs (qui peuvent être très nombreux avec des jeux de données déséquilibrés). Ses valeurs sont comprises entre 0 et 1 (le plus haut le meilleur). Il s'obtient par la formule

$$\text{TS} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}};$$

- le *Heidke Skill Score* (HSS) : peut être vu comme une mesure de compétence généraliste, qui compare les prédictions de notre modèle à celles obtenues de façon aléatoire. Ses valeurs sont comprises entre -1 et 1 (le plus haut le meilleur). Un score positif indique une compétence par rapport à un modèle aléatoire. Un score nul une absence de compétence. Un score négatif indique

que le modèle aléatoire est plus compétent. Il s'obtient par la formule :

$$\text{HSS} = \frac{2 \times (\text{TP} \times \text{TN} - \text{FP} \times \text{FN})}{(\text{TP} + \text{FN})(\text{FN} + \text{TN}) + (\text{TP} + \text{FP})(\text{FP} + \text{TN})}$$

Nous calculons l'ensemble de ces mesures avec le modèle de base et le modèle amélioré (pour les six horizons de prédiction). Les résultats sont donnés dans le Tableau 3.6.

TABLEAU 3.6 – Évaluation du modèle de prédiction de *Dst* amélioré dans le contexte du problème de classification (et comparaison avec le modèle de base).

Horizon temporel (heures)	précision	rappel	F_{score}	FAR	TS	HSS
1	0,941 (0,923)	0,899 (0,837)	0,919 (0,878)	0,005 (0,006)	0,851 (0,782)	0,913 (0,868)
2	0,909 (0,890)	0,821 (0,793)	0,863 (0,839)	0,007 (0,008)	0,759 (0,722)	0,853 (0,827)
3	0,879 (0,875)	0,760 (0,721)	0,815 (0,791)	0,008 (0,008)	0,688 (0,654)	0,801 (0,775)
4	0,856 (0,855)	0,710 (0,682)	0,776 (0,759)	0,010 (0,009)	0,634 (0,611)	0,760 (0,742)
5	0,833 (0,856)	0,686 (0,629)	0,752 (0,723)	0,011 (0,009)	0,603 (0,569)	0,734 (0,706)
6	0,834 (0,829)	0,624 (0,598)	0,714 (0,694)	0,010 (0,010)	0,555 (0,532)	0,694 (0,674)

Note : Les résultats obtenus avec le meilleur modèle sont indiqués en gras. Les résultats obtenus avec le modèle de base sont indiqués entre parenthèses.

Nous constatons sans surprise que les performances de classification se dégradent avec l'augmentation de l'horizon temporel. Néanmoins, par rapport au modèle de base, nous constatons une nette amélioration avec notre modèle amélioré. Pour un horizon d'une heure, notre modèle émet des alertes qui sont à 94,1 % pertinentes (contre 92,1 % pour le modèle de base) et 89,9 % des dépassements de seuil sont correctement prédits (contre 83,7 % pour le modèle de base). Cela s'accompagne d'un très bon taux de fausses alertes : 0,05 % seulement. Certes, pour des horizons de prédiction plus lointains (de quatre à six heures) notre modèle amélioré présente un FAR légèrement moins bon que le modèle de base (écart inférieur à 0,02 %), mais cela est largement compensé par une meilleure capacité à détecter les orages, comme en atteste la nette amélioration du **rappel**. Cette compensation est visible sur les valeurs de **TS**, qui sont toujours meilleures pour le modèle amélioré. Enfin notons que les valeurs de **HSS** sont également toujours meilleures pour le modèle amélioré, ce qui confirme que le modèle amélioré est plus

compétent que le modèle de base, bien que celui-ci présente déjà de très hautes valeurs. Ainsi, non seulement le modèle amélioré est plus précis lors des périodes d'orages que le modèle de base, mais il arrive à mieux les distinguer des périodes calmes au sein d'un système d'alerte.

Il est toutefois difficile d'utiliser ces métriques pour quantifier l'horizon maximal pour lequel nos prédictions restent acceptables. Il faudrait pour cela avoir un cahier des charges avec des exigences précises. Par exemple, si on souhaite que 90 % de nos alertes soient pertinentes, alors l'horizon de prédiction maximal est de deux heures. Si on se fixe comme critère que trois-quarts des périodes d'orages doivent être correctement détectées, alors l'horizon maximal est de trois heures, etc. En l'absence de telles contraintes nous ne pouvons définitivement conclure.

Mais surtout, ces métriques nous indiquent si notre modèle arrive à distinguer les périodes d'orages des périodes calmes en général. Or si nous prenons l'exemple de l'orage extrême de juillet 2000, nous constatons que la valeur de *Dst* reste sous le seuil de -50 nT pendant 48 heures – ce qui nous donne 48 échantillons pour lesquels notre modèle a l'occasion d'émettre une alerte. Or, afin que la prédiction soit réellement utile, il faut surtout que le modèle soit capable de détecter le début de l'orage géomagnétique. En Figure 3.11 nous montrons les prédictions obtenues toujours pour l'orage extrême de juillet 2000 mais en resserrant la vue autour du pic de l'orage. Pour plus de clarté nous ne montrons ici que les prédictions obtenues avec une et six heures d'avance. On constate assez clairement que pour un horizon de prédiction d'une heure, le modèle de base semble clairement « en avance » sur les observations, tandis que le modèle amélioré a l'air d'être plus synchrone avec les observations. En revanche, pour l'horizon de prédiction de six heures on constate clairement que les deux modèles semblent très « en retard » par rapport au début d'orage.

Dans la suite nous allons essayer de quantifier ce phénomène, qui devrait pouvoir nous indiquer plus précisément le degré d'utilité de nos modèles.

3.4.3 Évaluation du modèle par détection des débuts d'orage

En considérant toujours le problème de classification binaire, nous nous concentrons ici uniquement sur les débuts d'orage. Ainsi, nous évaluons uniquement la capacité de notre modèle à détecter les dépassements du seuil -50 nT lorsqu'à l'instant présent (de la prédiction) nous sommes en période calme. Cela nous permet d'obtenir un pourcentage de périodes d'orages que nous avons réussi à effectivement prédire en avance, donné pour chaque horizon de prédiction en Figure 3.12.

Nous avons identifié 395 débuts d'orage géomagnétique dans notre ensemble de test. Parmi ceux-ci, un peu plus des deux-tiers (68,4 %) sont correctement prédits une heure en avance. Toutefois ce pourcentage chute rapidement : il est de 52,2 % deux heures en avance, et seulement de 31,4 % six heures en avance. Nous remarquons que le modèle amélioré est plus performant que le modèle de base, mais cette différence est surtout marquée pour l'horizon de prédiction d'une heure (68,4 % contre 50,6 % pour le modèle de base).

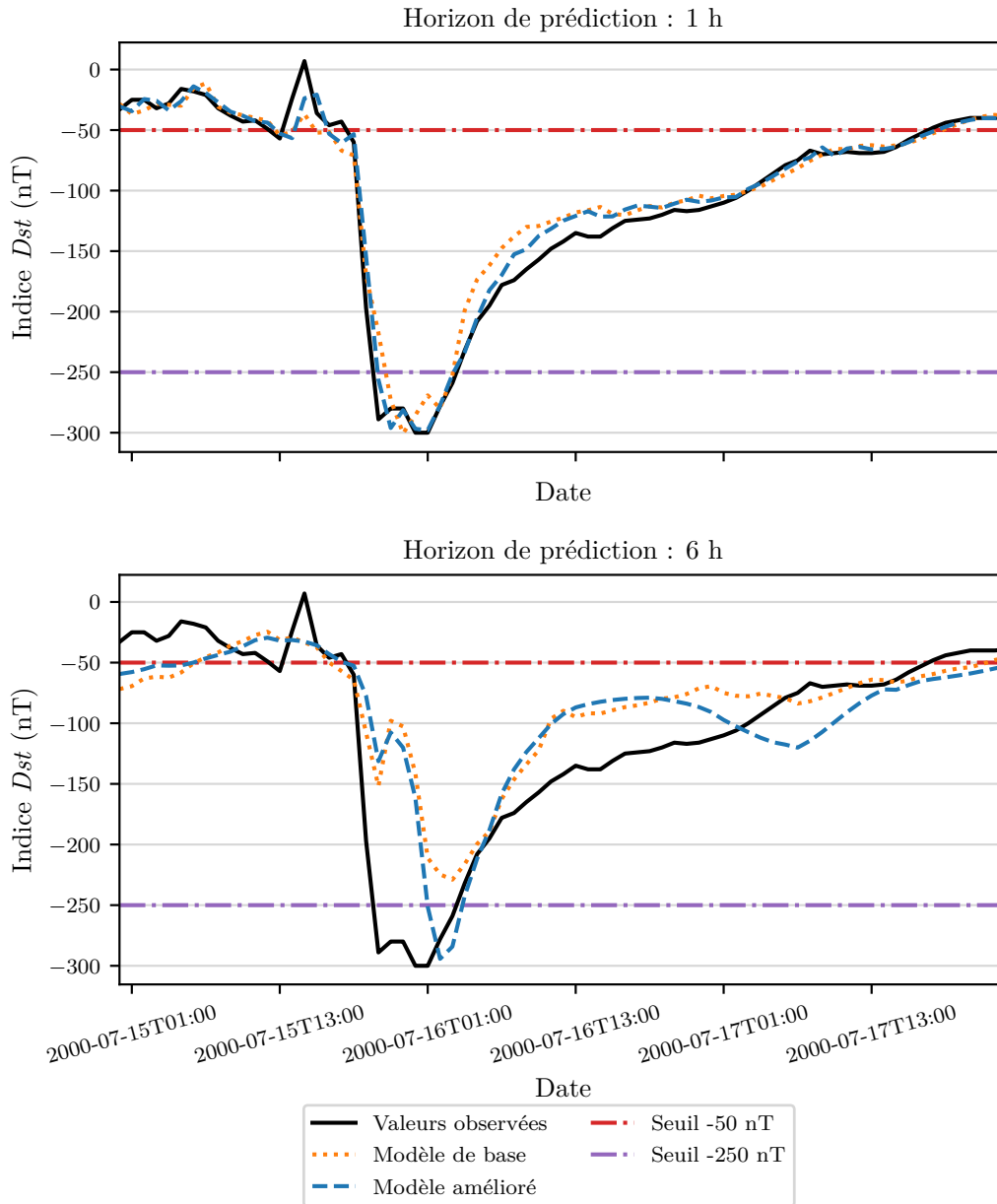


FIGURE 3.11 – Exemples de prédictions une heure et six heures en avance obtenues avec le modèle de base et notre modèle amélioré lors de l'orage géomagnétique extrême de juillet 2000.

L'écart entre les valeurs présentées ici et les valeurs de [rappel](#) présentées dans le Tableau 3.6 montre que les deux modèles sont plus compétents à prédire la continuité d'une période d'orage plutôt que son début. De plus, on peut aisément définir un critère de performance minimal à partir de ces valeurs, par exemple : si on souhaite prédire au moins deux-tiers des orages en avance, alors l'horizon de prédiction maximal est d'une heure ; si on souhaite prédire au moins la moitié des orages en avance, alors l'horizon de prédiction maximal est de deux heures ; si on souhaite prédire au moins un-tiers des orages en avance, alors l'horizon de prédiction maximal est de cinq heures.

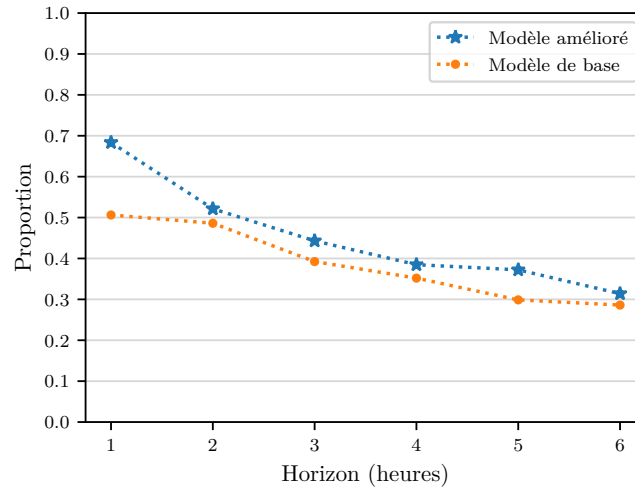


FIGURE 3.12 – Pourcentage de périodes d'orage correctement prédites avant leur début.

3.4.4 Évaluation des distorsions temporelles entre les prédictions et les observations

Afin de compléter l'approche précédente, nous proposons ici d'évaluer notre modèle à l'aide d'une métrique appelée le *Temporal Distortion Mix* (TDM). Cette métrique, proposée par VALLANCE et al. 2017 en s'appuyant sur des travaux précédents de FRÍAS-PAREDES et al. 2016, permet de caractériser la propension d'une série temporelle à exhiber (ou non) un décalage temporel systématique par rapport à une série de référence.

Cette métrique est basée sur l'algorithme de *Dynamic Time Warping* (DTW), qui permet de mesurer la déformation temporelle entre deux séries de données (SAKOE et CHIBA 1978). Pour ce faire, cet algorithme associe, sous certaines contraintes, chaque élément de la première série temporelle à un élément de la seconde série (et inversement) de telle sorte à minimiser un score de déformation temporelle. L'ensemble des paires d'indices d'éléments associés constitue alors un « chemin optimal ». Ce processus est illustré en Figure 3.13. Dans cet exemple, LI et al. 2019 appliquent l'algorithme de DTW à deux séries temporelles (bleue et rose), ce qui leur permet d'obtenir le chemin optimal (en vert dans la figure du milieu) symbolisant la déformation temporelle entre les deux séries. Si deux séries sont parfaitement alignées, alors le chemin optimal est la diagonale (en bleu clair dans l'image du milieu).

À partir de cet algorithme, FRÍAS-PAREDES et al. 2016 construit l'indice de distorsion temporelle ou *Temporal Distortion Index* (TDI) en mesurant l'aire par rapport à la diagonale du chemin optimal. Le TDI indique dans quelle mesure les deux séries temporelles sont systématiquement (ou pas) en retard (ou en avance). Le TDM est ensuite une mesure normalisée du TDI, possédant les propriétés suivantes :

- ses valeurs sont comprises entre -1 et 1 ;

- si $TDM = -1$ alors la première série est systématiquement en avance par rapport à la série de référence ;
- si $TDM = 1$ alors la première série est systématiquement en retard par rapport à la série de référence ;
- si $TDM = 0$ alors la première série est, en moyenne, parfaitement alignée (temporellement) avec la série de référence.

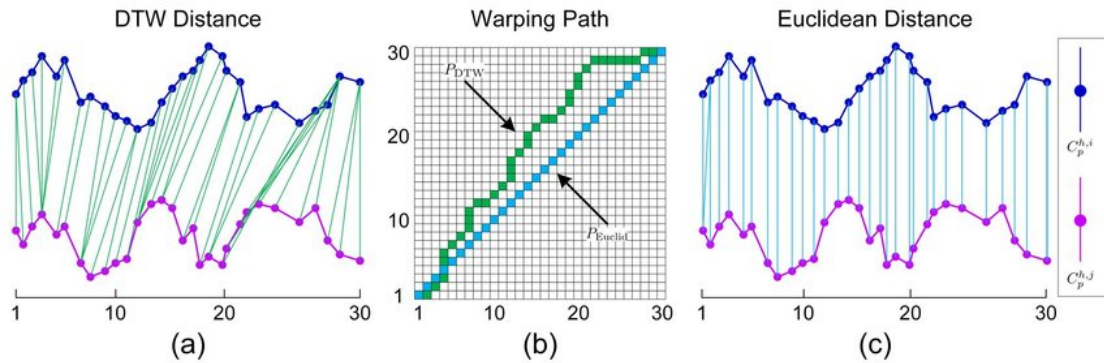


FIGURE 3.13 – Exemple de mesure de distorsion temporelle entre deux séries temporelles par DTW. Reproduit de LI et al. 2019, sous licence CC BY-NC-SA 4.0.

L'objectif est donc d'avoir une valeur de TDM la plus proche possible de 0, sachant que dans notre cas, il est préférable d'avoir une valeur légèrement négative plutôt que positive. Pour une définition mathématique plus précise de l'algorithme de DTW, du TDI et du TDM nous referons le lecteur aux articles sus-cités. Ici nous utilisons le TDM pour quantifier la tendance de nos modèles de prédiction de Dst à être temporellement alignés (ou non) avec les observations. Nous avons implémenté l'algorithme de DTW à l'aide de la librairie pour langage Python `tslearn` (v. 0.4.1, TAVENARD et al. 2020).

Notons que d'autres études ayant été publiées au cours de cette thèse ont également fait utilisation de l'algorithme de DTW pour mesurer l'écart temporel entre des prédictions d'indices géomagnétiques et les valeurs observées, sans toutefois utiliser le TDM (voir par exemple LAPERRE et al. 2020). Les mesures de TDM obtenues avec le modèle de base et le modèle amélioré sont données en Figure 3.14.

On constate que sur l'ensemble de test complet, les deux modèles présentent des valeurs de TDM très élevées, comprises entre 0,6 et 0,7. Ces valeurs suggèrent que nos modèles ont tendance à présenter un retard systématique. Cela est sans doute dû à la prédominance de périodes calmes, au cours desquelles les valeurs de Dst fluctuent peu et sont très fortement auto-corrélées d'une heure à l'autre. Dans cette situation, le réseau de neurones adopte pour stratégie de prédire une valeur proche de la dernière valeur observée de Dst (qui, rappelons-le fait partie des *inputs* du modèle). Or quand on ne considère que les périodes d'orages, les valeurs de TDM chutent fortement et sont comprises entre 0 et 0,3. En particulier, celles-ci sont particulièrement faibles avec le modèle amélioré pour des horizons de deux et trois heures (valant alors 0,085 et 0,064), montrant que pour ces horizons

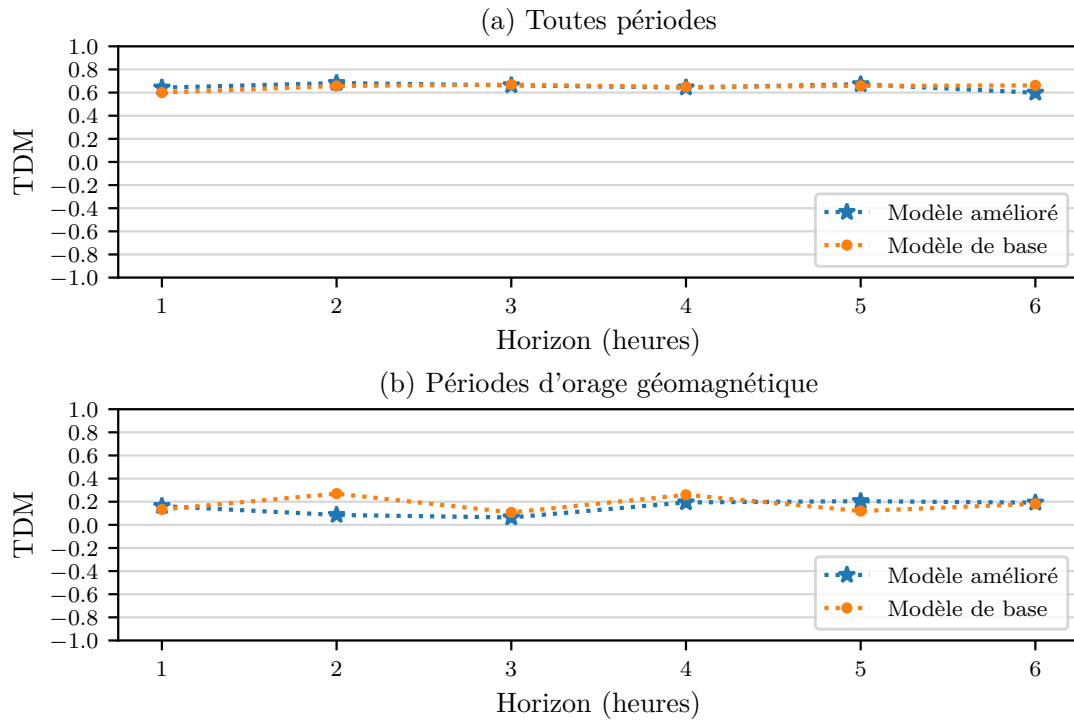


FIGURE 3.14 – Valeurs de TDM obtenues avec le modèle de base et le modèle amélioré de prédiction de Dst lorsqu'on considère (a) la totalité de l'ensemble de test et (b) uniquement les périodes d'orages.

le modèle est en moyenne très synchronisé avec les observations. Pour des horizons de prédiction supérieurs, la valeur remonte jusqu'à $\approx 0,2$, suggérant que le réseau prend un léger retard, ce qui va dans le sens des conclusions tirées de l'évaluation du modèle par détection des débuts d'orage (section 3.4.3).

3.4.5 Application à des horizons de prédiction plus éloignés

Nous avons vu plus tôt dans ce chapitre que, dès un horizon de prédiction de cinq-six heures, le modèle amélioré, malgré sa plus grande complexité, n'était pas notablement meilleur que le modèle de base, ce qui suggère que l'information contenue dans les paramètres du vent solaire et dans Dst six heures en avance sont insuffisantes pour le prédire de façon plus précise. Il serait tentant d'extrapoler cette affirmation pour tout horizon de prédiction supérieur, mais il n'est pas à exclure le fait que le réseau de neurones parvienne à détecter, si elles existent, des corrélations statistiques plus fortes pour des horizons de prédiction plus lointains, permettant d'utiliser ce modèle pour des horizons plus éloignés. Pour en avoir le cœur net, nous avons entraîné le modèle pour des horizons de 12 et 24 h. Nous récapitulons les métriques d'évaluation obtenues dans le Tableau 3.7.

Sans surprise, ces performances sont bien plus faibles que celles observées jusqu'à six heures en avance. Pour un horizon de 24 h notamment nous ne parvenons à prédire plus que 15,4 % des orages en avance, et la valeur de $RMSE$ se rapproche

TABLEAU 3.7 – Évaluation du modèle amélioré entraîné pour prédire *Dst* 12 et 24 h en avance. Les métriques pour le modèle à 6 h sont données pour comparaison.

Métrique	Horizon 6 h	Horizon 12 h	Horizon 24 h
RMSE (nT)	12,67	17,91	20,92
<i>R</i>	0,87	0,70	0,56
TDM	0,60	0,52	0,70
précision	0,834	0,610	0,594
rappel	0,624	0,399	0,161
TS	0,555	0,318	0,145
Proportion d'orages prédits en avance	0,314	0,321	0,154

de l'écart-type de la distribution des valeurs de *Dst* sur l'ensemble de test (valant 25,19 nT) ce qui suggère que le modèle n'a quasiment plus aucune compétence. Pour l'horizon de 12 h, les performances sont certes dégradées en absolu, mais en l'absence de modèle de référence il est difficile de conclure quant à l'utilité (ou l'inutilité) d'un tel modèle. Cela est illustré par la Figure 3.15, montrant les prédictions obtenues pour l'orage extrême de juillet 2000 pour des horizons de 12 et 24 h.

3.4.6 Bilan sur l'extension du banc d'évaluation des prédictions de *Dst*

La RMSE et la corrélation linéaire sont des métriques globales qui ne rendent pas compte exhaustivement de la performance d'un modèle. En complétant notre banc d'évaluation nous avons montré que le modèle amélioré est en particulier plus précis que le modèle de base en période d'orage intense et qu'il distingue mieux les périodes calmes des périodes d'orage géomagnétique, mais aussi que cette capacité de distinction chute si on se concentre uniquement sur le début d'un orage. Cela est probablement dû à une certaine tendance à l'inertie, que nous avons également pu quantifier. Il est difficile d'affirmer, sans cahier des charges précis, quel serait l'horizon de prédiction limite de notre modèle. Toutefois, en conservant des performances correctes, cet horizon se situe sans doute aux environs de 5 – 6 h, confirmant ainsi certaines des conclusions de GRUET et al. 2018 et des autres études ayant trouvé des horizons de prédiction maximaux similaires.

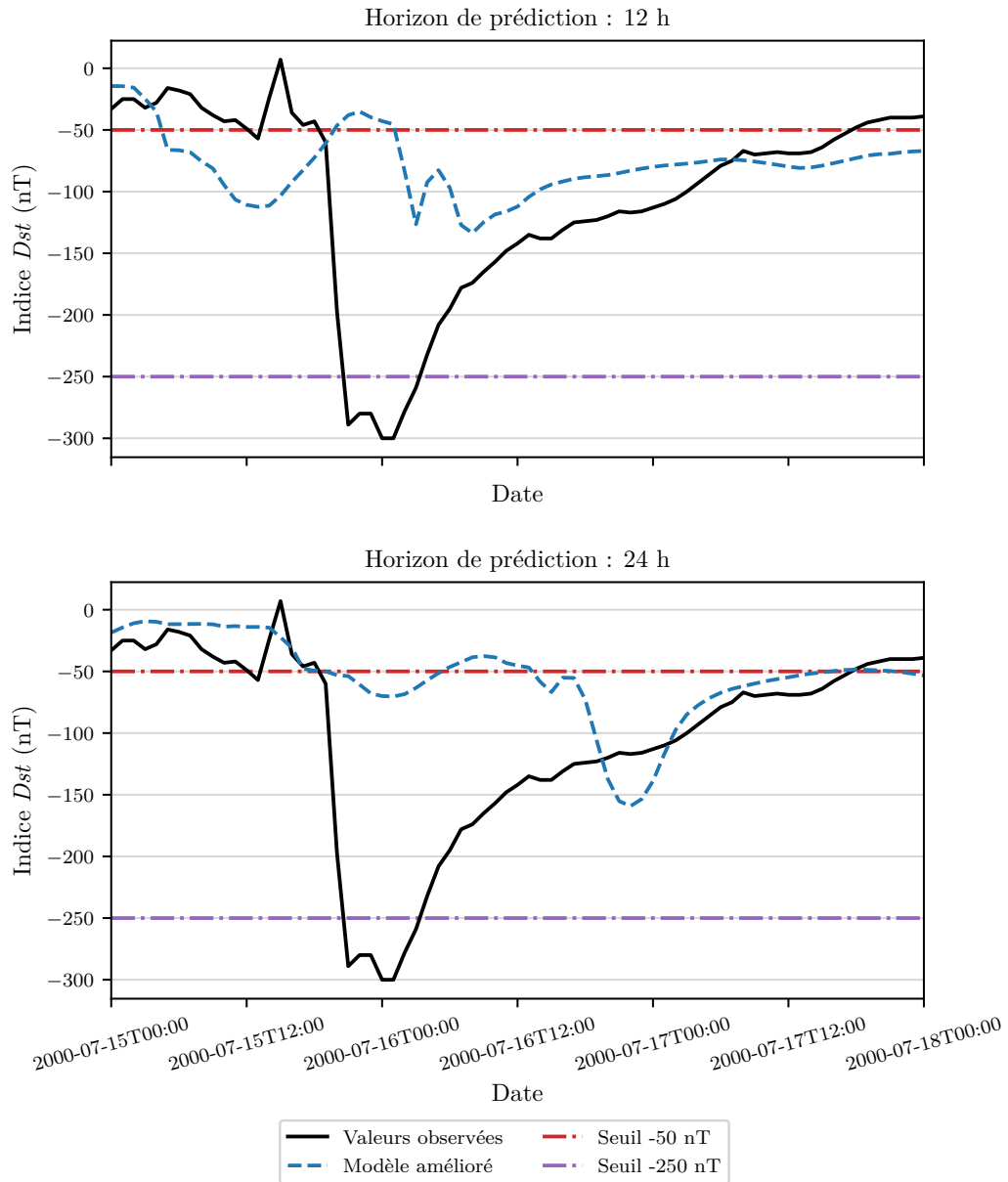


FIGURE 3.15 – Exemples de prédictions 12 heures et 24 heures en avance obtenues avec le modèle amélioré lors de l'orage géomagnétique extrême de juillet 2000.

3.5 Obtenir des prédictions probabilistes directement avec un réseau de neurones

Un des aspects de l'étude de GRUET et al. 2018 que nous avons volontairement mis de côté jusqu'ici est la capacité de leur modèle à produire des prédictions probabilistes. Nous n'avons pas encore évoqué ce problème car nous estimions qu'il était déjà possible, et intéressant, d'essayer dans un premier temps de caractériser la capacité de nos modèles à produire des prédictions à six heures de façon déterministe. Néanmoins, comme il est sans aucun doute plus intéressant dans un contexte de prise de décision de bénéficier d'un modèle probabiliste, nous ne pouvions pas ignorer cette question.

3.5.1 Description et justification de notre approche

Pour obtenir des prédictions probabilistes (c'est-à-dire des prédictions accompagnées d'une incertitude quantifiée), GRUET et al. 2018 utilisent une deuxième classe d'algorithmes d'apprentissage automatique appelés *processus gaussiens* ou *Gaussian Processes* (GPs). Les GPs sont une méthode de régression bayésienne capable de fournir une mesure explicite de l'incertitude d'une prédiction en fournissant non pas un unique modèle, mais une distribution de probabilité dans l'espace des modèles. Un GP est entièrement défini par une fonction moyenne et un noyau de covariance. Dans leur étude, GRUET et al. 2018 utilisent les prédictions obtenues à l'aide de leur réseau de neurones comme fonction moyenne du GP. Pour plus de détails sur ce type de méthodes le lecteur est invité à consulter un livre de référence comme WILLIAMS et RASMUSSEN 2006.

Ce type de méthode a l'avantage de fournir une incertitude quantifiée et explicite à des prévisions. Le principal défaut des GP est leur coût informatique. En effet, entraîner un GP classique nécessite, dans notre cas, plusieurs heures – voire jours. Cela est notamment dû au fait qu'il n'est pas possible d'entraîner un GP par *mini-batches* comme nous le faisons pour les réseaux de neurones⁷. Une solution pour remédier à ce problème serait de prendre uniquement un sous-ensemble de nos données pour entraîner un GP, mais nous avons vu qu'il est important dans notre cas d'utiliser de données couvrant des périodes diverses et longues, représentant différents processus physiques à l'échelle du cycle solaire.

Nous proposons ici une méthode d'obtention de prédictions probabilistes directement à l'aide d'un réseau de neurones, sans utilisation d'autres classes d'algorithmes. Pour ce faire, nous supposons que nous voulons des prédictions sous la forme d'une distribution gaussienne (pour chaque instant, au lieu d'obtenir une prédiction ponctuelle, nous souhaitons obtenir une distribution de probabilité suivant une loi normale). Or une distribution gaussienne est entièrement caractérisée par une moyenne et un écart-type. Ainsi, nous construisons un réseau ayant non

7. Récemment des méthodes pour entraîner des GPs par *mini-batches* ont été proposées par, entre autres, HENSMAN et al. 2014; WILSON et al. 2015; WILSON et al. 2016; LIU et al. 2019b. Au cours de notre thèse nous avons essayé ces méthodes, qui, bien que prometteuses, se sont montrées très complexes à entraîner, pour un gain incertain. Nous les avons donc mises de côté.

plus une seule sortie, mais deux sorties et nous entraînons ce réseau pour que les sorties du modèle correspondent à la moyenne et à l'écart-type d'une distribution gaussienne.

Concrètement, nous utilisons un réseau de neurones ayant quasiment la même architecture que précédemment, à l'exception que ce réseau possède une deuxième couche linéaire, en parallèle de la première, qui sert à prédire l'écart-type de nos prédictions. Le schéma de notre nouveau modèle est donné en Figure 3.16. μ_{Dst} et σ_{Dst} désignent respectivement la moyenne et l'écart-type d'une prédiction.

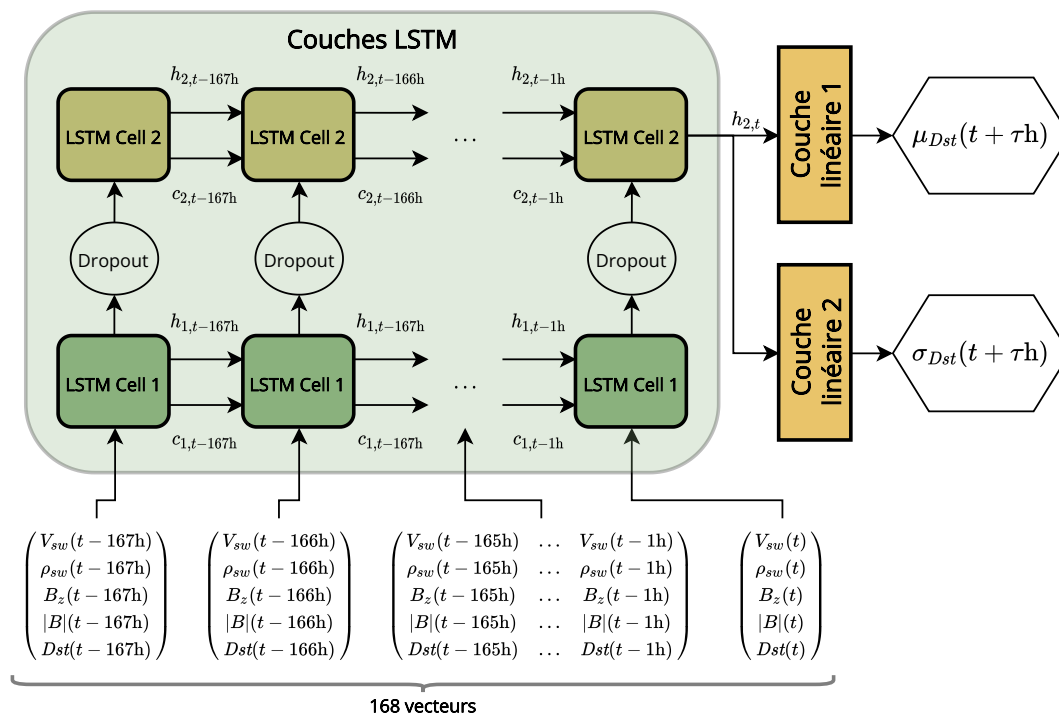


FIGURE 3.16 – Schéma de l'architecture de notre modèle de prédiction probabiliste de Dst .

Afin d'entraîner correctement le modèle il nous faut changer la fonction de perte utilisée jusqu'ici, à savoir la **MSE**, et utiliser une fonction qui permet de comparer une distribution de probabilité à un point. C'est pourquoi nous utilisons ici comme fonction de perte une fonction dérivée de la log-vraisemblance négative gaussienne ou *Gaussian Negative Log-Likelihood* (GNLL), à laquelle nous ajoutons la **MSE**. On obtient ainsi une *modified Gaussian Negative Log-Likelihood* (MGNLL) qui s'exprime :

$$\begin{aligned} \text{MGNLL}(\mu, \sigma, y_{\text{vrai}}) &= \text{GNLL}(\mu, \sigma, y_{\text{vrai}}) + \text{MSE}(\mu, y_{\text{vrai}}) \\ &= \frac{1}{2} * (\log(\sigma^2) + \frac{(\mu - y_{\text{vrai}})^2}{\sigma^2}) + (\mu - y_{\text{vrai}})^2, \end{aligned} \quad (3.4)$$

où μ et σ sont la moyenne et l'écart-type de la distribution gaussienne prédite par le réseau de neurones et y_{vrai} est la vraie valeur observée de Dst .

Les avantages de cette méthode sont multiples. Tout d'abord, comme nous

l'avons dit, cette méthode permet d'exploiter l'intégralité des données à notre disposition. Cela est notamment permis par le fait que l'entraînement de ce modèle est à peine plus long que celui des modèles présentés précédemment (avec notre ordinateur, une *epoch* d'entraînement dure en moyenne 26 s contre 25 s avec le modèle déterministe). Le réglage des hyperparamètres du modèle se fait très rapidement dès lors qu'on a déjà développé le modèle déterministe. Dans notre cas, le seul hyperparamètre ayant été modifié est le *learning rate* initial, fixé ici à $lr = 5 \times 10^{-4}$.

Enfin, nous pouvons nous appuyer sur l'apprentissage effectué avec le modèle déterministe et procéder à de l'apprentissage par transfert, en initialisant les valeurs des poids des couches LSTM et de la première couche linéaire avec les valeurs des poids de ces mêmes couches obtenues à l'issue de l'entraînement du modèle déterministe⁸. En procédant ainsi notre approche se rapproche dans celle proposée par CAMPOREALE et al. 2019, qui suggère l'utilisation d'un deuxième réseau de neurones pour prédire des incertitudes à des prédictions déterministes préalablement obtenues.

Notons qu'une telle démarche n'avait jamais été proposée pour des applications en météorologie de l'espace à notre connaissance avant le début de cette thèse. Depuis, TASISTRO-HART et al. 2021 ont également utilisé un modèle similaire pour obtenir des prédictions probabilistes de l'indice *Kp*.

Du fait de la très grande proximité entre ce modèle probabiliste et le modèle amélioré précédent, nous n'allons pas proposer ici de nouveau un banc d'évaluation complet de ce modèle probabiliste avec les mêmes métriques déjà utilisées, car les valeurs obtenues sont identiques ou extrêmement proches de celles obtenues avec le modèle déterministe. Par contre, nous proposons d'évaluer ce nouveau modèle à l'aide de métriques dédiées à l'évaluation de prédictions probabilistes, ce qui permettra de compléter le banc d'évaluation précédent avec de nouvelles informations.

3.5.2 Présentation et analyse des prédictions probabilistes

Nous analysons ici succinctement les prédictions obtenues à l'aide du modèle probabiliste. Tout d'abord, la Figure 3.17 montre des exemples de prédictions probabilistes obtenues avec notre modèle pour un orage survenu le 26 août 1998 ainsi que pour l'orage extrême de juillet 2000 (pour des horizons de prédictions d'une, trois et six heures). Les intervalles de confiance à 70 % et 95 % sont indiqués sous la forme d'une aire bleue.

D'après ces deux exemples, il semblerait que l'incertitude prédite par le modèle devienne plus grande lors de la phase principale d'un orage par rapport aux périodes calmes et aux phases de retour à l'équilibre. De plus, il semblerait que le modèle

8. Cet apprentissage par transfert n'est pas obligatoire pour faire converger le modèle, mais dans notre cas il a permis de l'accélérer car une grande partie du réseau de neurones a été préalablement optimisée.

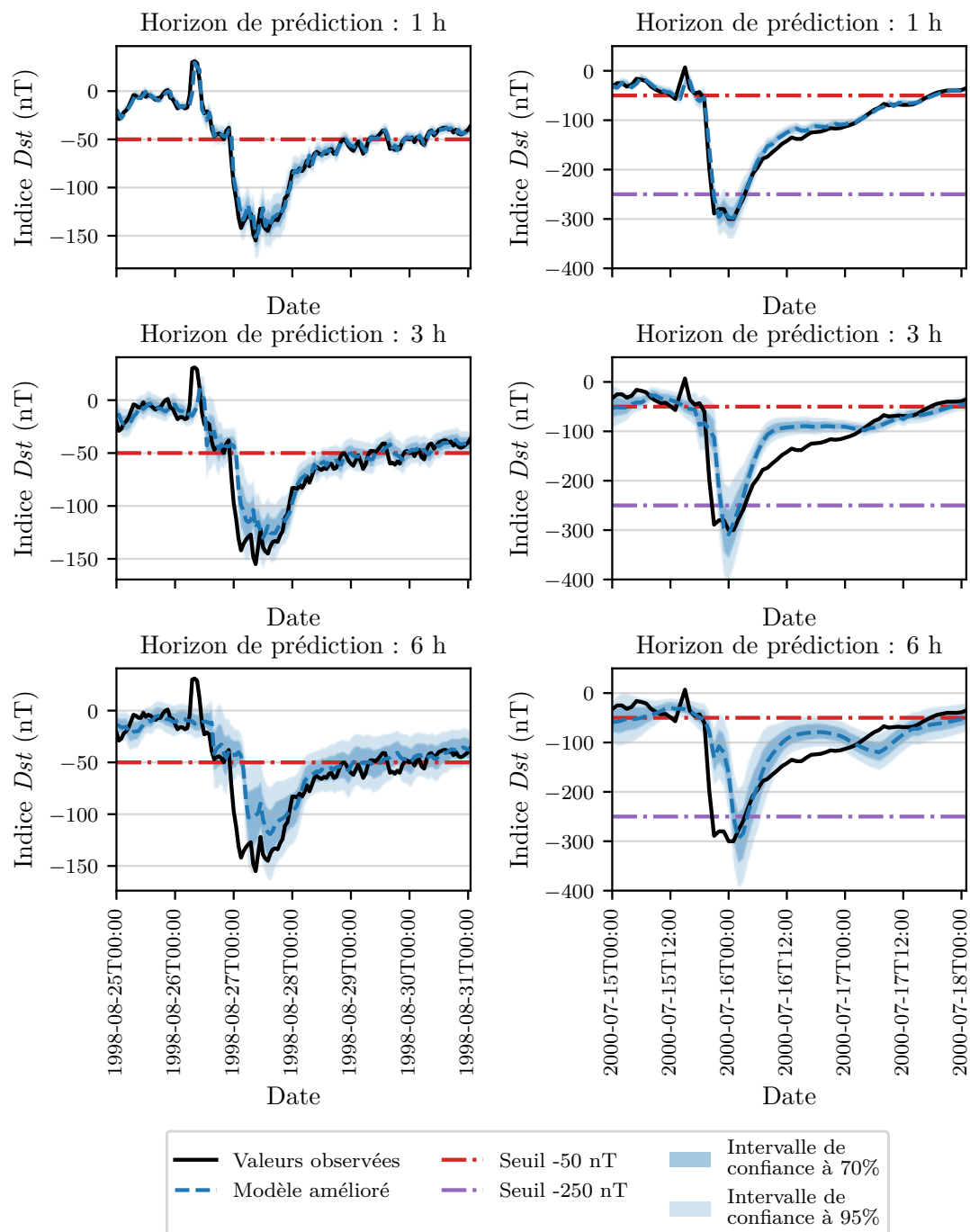


FIGURE 3.17 – Exemples de prédictions obtenues une, trois et six heures en avance à l’aide du modèle probabiliste lors des orages ayant eu lieu le 26 août 1998 et mi-juillet 2000.

produise des incertitudes globalement plus fortes lorsqu'on augmente l'horizon de prédiction. Cette analyse est confirmée par les résultats présentés en Figure 3.18. Dans cette figure nous avons tracé la distribution de probabilité empirique des valeurs d'écart-type prédites par notre modèle sur la totalité de l'ensemble de test.

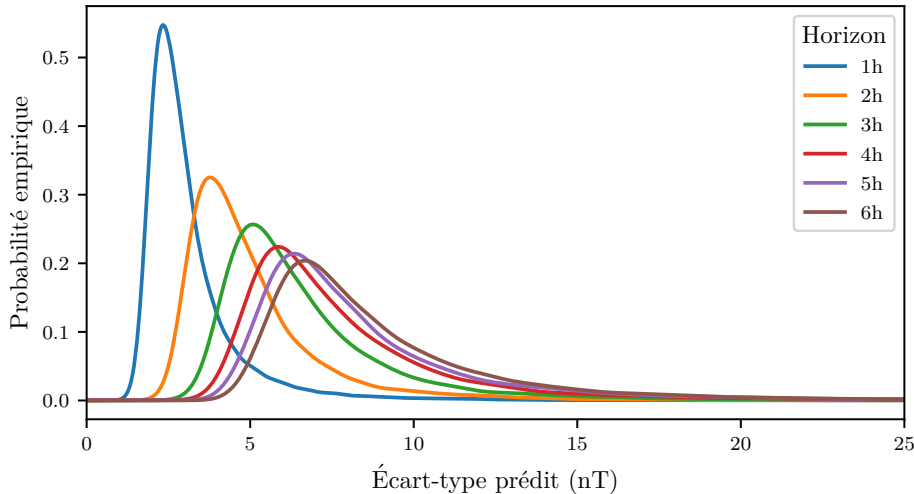


FIGURE 3.18 – Courbes des distributions de probabilité empiriques des valeurs d'écart-type prédites par le modèle probabiliste sur l'ensemble d'évaluation complet pour des horizons de prédiction allant d'une à six heures.

Note : Nous avons coupé la figure selon l'axe des abscisses à une valeur de 25 nT pour de raisons de lisibilité. En réalité, pour des horizons supérieurs à quatre heures, l'écart-type prédit peut atteindre environ 60 nT dans les cas les plus extrêmes.

Il apparaît clairement que plus l'horizon de prédiction est lointain, plus l'incertitude prédite est généralement élevée. En outre, plus l'horizon augmente, plus la courbe représentant la distribution des écart-types « s'aplatit ». Concrètement, cette figure nous montre que pour un horizon d'une heure, le modèle « estime » être très fiable, et produit donc presque toujours des incertitudes faibles. D'un autre côté, pour un horizon de six heures, le modèle estime généralement être moins fiable (la distribution est déplacée vers les valeurs plus élevées) et, de plus, il reconnaît que sa fiabilité est variable, en fonction du type d'activité géomagnétique auquel il est confronté (d'où l'étalement de la distribution des écarts-types). Ce résultat est très satisfaisant, car il montre que notre modèle a appris à rendre compte correctement de la croissance de l'incertitude avec l'augmentation de l'horizon de prédiction.

Comme notre méthode semble effectivement capable de faire apprendre au réseau de neurones à quantifier l'incertitude de ses prédictions, nous proposons quelques méthodes dédiées à l'évaluation de prédictions probabilistes. La première métrique que nous utilisons est le **CRPS** (MATHESON et WINKLER 1976). Si on note $F_{\mu,\sigma}^t$ la fonction de répartition associée à une prédiction probabiliste pour un instant t et y_{vrai}^t la valeur observée à cet instant t , le **CRPS** est défini par :

$$\text{CRPS} = \int_{-\infty}^{+\infty} \left(F_{\mu,\sigma}^t(x) - \mathbb{1}(x - y_{\text{vrai}}^t) \right)^2 dx \quad (3.5)$$

Pour nous aider à comprendre cette formule, la Figure 3.19 montre un calcul de **CRPS** pour un exemple de prédiction. Dans cet exemple, la valeur observée vaut $y_{\text{vrai}} = -35 \text{ nT}$, et la prédiction est une distribution normale de moyenne -30 nT et d'écart-type 10 nT . Le **CRPS** est la valeur de l'aire (surface rouge) entre la courbe de la fonction de répartition de la prédiction (en bleu) et la droite verticale représentant l'observation (en noir pointillé). Dans cet exemple, le **CRPS** vaut $3,31 \text{ nT}$. Sur cette figure on voit bien que le **CRPS** est une métrique qui pénalise à la fois le biais de la prédiction (la moyenne de la prédiction doit être proche de l'observation) et sa variance (un écart-type trop grand « aplatit » la fonction de répartition et donc augmente la valeur de l'aire rouge). Le **CRPS** peut être vu comme une généralisation de l'*erreur absolue moyenne* ou *Mean Absolute Error (MAE)* au cas probabiliste. En effet, dans le cas d'une prédiction avec un écart-type nul alors le **CRPS** est strictement égal à l'erreur absolue. Dans notre exemple la **MAE** vaut 5 nT , soit 51% de plus que le **CRPS**, ce qui indique que le fait d'adjoindre une distribution à la prédiction la rend meilleure. Notons que ce n'est pas toujours le cas. Si dans notre exemple l'écart-type prédit valait 20 nT au lieu de 10 nT , alors le **CRPS** aurait été égal à $5,17 \text{ nT}$, soit une valeur moins bonne que la **MAE**. Dans ce dernier cas le fait d'adjoindre une distribution de probabilité à la prédiction aurait dégradé sa précision.

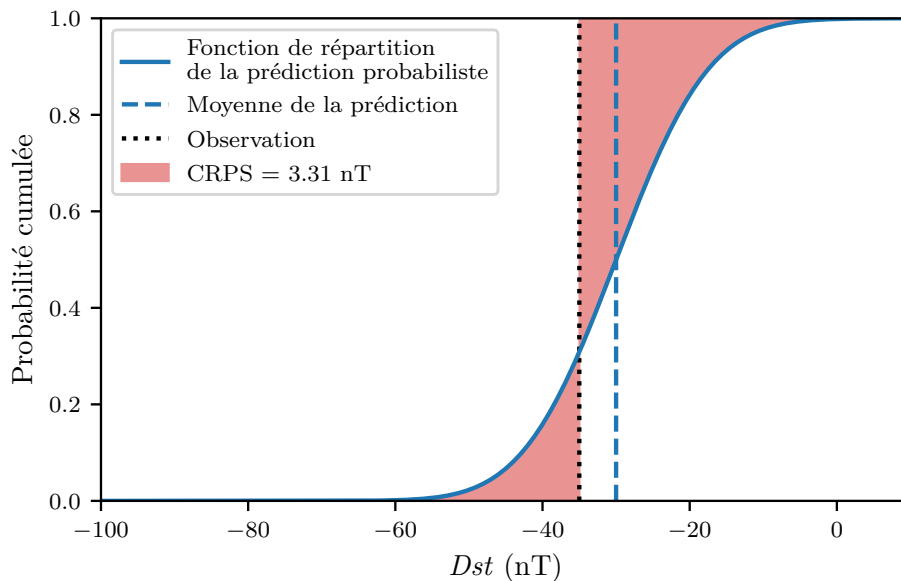


FIGURE 3.19 – Illustration du calcul de **CRPS** pour un exemple de prédiction probabiliste.

Ici nous utiliserons plus précisément le **CRPS** moyen calculé sur l'ensemble d'évaluation complet, que nous noterons $\overline{\text{CRPS}}$. Le Tableau 3.8 donne les valeurs de **CRPS** et de **MAE** (pour comparaison) obtenues à partir de nos prédictions⁹. Nous avons également indiqué le $\overline{\text{CRPS}}$ qu'on obtiendrait si, au lieu de faire prédire un écart-type au réseau de neurones, on se contentait d'utiliser un écart-type fixe (qui minimise le $\overline{\text{CRPS}}$) pour toutes les prédictions. Nous appelons ce modèle

9. Les valeurs de **MAE** du modèle probabiliste ont un écart inférieur à 1% à celles du modèle déterministe donc on ne les indique pas dans le tableau pour une meilleure lisibilité.

« modèle probabiliste de référence ». Pour chaque horizon, nous indiquons entre parenthèses la valeur de l'écart-type fixe utilisé par le modèle probabiliste de référence.

TABLEAU 3.8 – Évaluation du modèle probabiliste à l'aide de la MAE et le $\overline{\text{CRPS}}$.

Horizon (heures)	Notre modèle		Modèle probabiliste de référence
	MAE (nT)	$\overline{\text{CRPS}}$ (nT)	$\overline{\text{CRPS}}$ (nT)
1	2,49	1,81	1,87 ($\sigma = 3 \text{ nT}$)
2	4,28	3,10	3,21 ($\sigma = 5 \text{ nT}$)
3	5,62	4,08	4,22 ($\sigma = 6 \text{ nT}$)
4	6,47	4,71	4,87 ($\sigma = 7 \text{ nT}$)
5	7,15	5,24	5,42 ($\sigma = 8 \text{ nT}$)
6	7,66	5,64	5,83 ($\sigma = 8 \text{ nT}$)

On constate que pour tous les horizons de prédiction le $\overline{\text{CRPS}}$ est largement inférieur à la MAE, ce qui montre que notre modèle gagne en compétence lorsqu'on lui fait prédire sa propre incertitude. Pour tous les horizons de prédiction, l'amélioration est comprise entre 26 et 28 %, ce qui montre que même les prédictions faites avec les horizons de prédiction les plus éloignés gagnent à être probabilistes. Lorsqu'on compare notre modèle probabiliste avec le modèle probabiliste de référence on remarque également que notre modèle est toujours plus performant. Cela montre bien que le réseau de neurones fournit des mesures d'incertitude qui sont plus précises et pertinentes que si on se contentait d'ajouter une valeur d'écart-type fixe aux prédictions déterministes.

Ensuite, nous proposons d'évaluer de nouveau le système d'alerte binaire en prenant en compte l'incertitude de nos prédictions. En effet, maintenant que nous avons des prédictions probabilistes, nos alertes sont également probabilistes (elles indiquent la probabilité d'être sous le seuil de -50 nT). De plus, il est toujours possible d'émettre une alerte déterministe à l'aide d'un seuil de probabilité. Par exemple, si on veut retrouver le même système d'alerte que celui obtenu à partir du modèle déterministe, il nous suffit d'émettre une alerte lorsque la probabilité d'être sous le seuil de -50 nT excède 0,5.

Pour évaluer le système de prédiction probabiliste, nous commençons par étudier son diagramme de fiabilité, donné en Figure 3.20. Un diagramme de fiabilité (parfois appelé courbe de calibrage) permet de vérifier la qualité du calibrage des prédictions probabilistes d'un classifieur binaire. Il représente la fréquence réelle observée de la classe positive (ici, les périodes d'orage) par rapport à sa probabilité prédite, pour des prédictions regroupées. L'axe des abscisses représente la probabilité moyenne prédite dans chaque groupe. L'axe des ordonnées est la fréquence de positifs observée, c'est-à-dire la proportion d'échantillons dont la classe véritable est la classe positive (dans chaque groupe). Un modèle parfaitement calibré produit la droite identité.

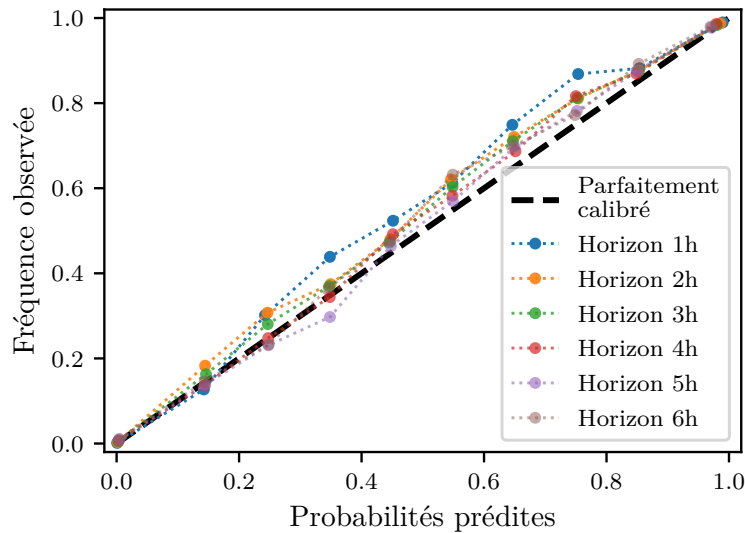


FIGURE 3.20 – Diagramme de fiabilité du modèle probabiliste pour les horizons de prédictions allant d’une à six heures.

On constate que nos modèles semblent assez bien calibrés, avec une légère tendance à la sous-estimation. Comme cette tendance est légère nous ne procédons pas à un calibrage supplémentaire.

Pour mesurer la performance d’un système de prédiction probabiliste il est habituel dans notre domaine (voir par exemple GRUET et al. 2018 ; CHAKRABORTY et MORLEY 2020, etc.) d’utiliser la courbe de la fonction d’efficacité du récepteur ou *Receiver Operating Characteristic (ROC)* (BRADLEY 1997). Cette courbe s’obtient simplement en comparant les valeurs de *rappel* et de *FAR* obtenues avec plusieurs seuils de probabilité. Or, des études ont montré que l’usage de la courbe *ROC* n’était pas idéal dans le cas où les proportions de classes positives et négatives (ici, périodes d’orage et périodes calmes) sont très déséquilibrées (DAVIS et GOADRICH 2006 ; SAITO et REHMSMEIER 2015). En effet, la courbe *ROC* est insensible aux différences de proportions entre deux classes : le *rappel* ne prenant en compte que les instants où un orage a effectivement eu lieu et le *FAR* uniquement les instants où un orage n’a pas eu lieu, modifier la proportion d’observation des classes ne modifie pas la valeur de ces métriques (et donc ne modifie pas la courbe *ROC*). Or, dans le cas d’un ensemble de données déséquilibré, cela peut mener à des erreurs d’interprétation, par exemple si le modèle est très bon à prédire la classe majoritaire, mais qu’il est plus important de parvenir à prédire la classe minoritaire. Ici, nous sommes dans cette situation : il est plus important de parvenir à prédire correctement la classe minoritaire, c’est-à-dire les périodes d’orage, représentant 7,4% de l’ensemble de test.

Dans ce type de situation il est donc conseillé par SAITO et REHMSMEIER 2015 de préférer l’utilisation de la courbe *précision-rappel* ou *Precision-Recall Curve*. La *PRC* compare les valeurs de *précision* et de *rappel* d’un système d’alerte probabiliste obtenues avec plusieurs seuils de probabilité. La *précision* étant une métrique sensible aux déséquilibres de classe, il s’agit d’une métrique adaptée

à notre situation. À partir de la **PRC** il est possible d'obtenir deux principales informations :

- le seuil de probabilité « idéal »¹⁰, qui maximise la valeur du F_{score} (qui est, pour rappel, la moyenne géométrique de la **précision** et du **rappel**) ;
- l'aire sous la courbe **PRC** ou *Area Under the PRC (APRC)*, qui est une mesure générale de la performance du classifieur probabiliste.

Un modèle idéal présente une courbe horizontale dont l'ordonnée est égale à 1 (et donc une **AUPRC** valant 1), et son seuil optimal correspond au point de coordonnées (1, 1). Un modèle aléatoire (et donc sans compétence) donne une courbe horizontale dont l'ordonnée est égale à la proportion d'apparition de la classe « orage géomagnétique » dans notre ensemble de données, qui vaut ici 0,074. La Figure 3.21 fournit les **PRCs** pour chacun des six horizons de prédiction.

Tout d'abord on constate que les valeurs d'**AUPRC** sont globalement élevées, le modèle étant même assez proche de la perfection pour un horizon d'une heure (**AUPRC** = 0,980). Sans surprise, la performance se dégrade progressivement avec l'augmentation de l'horizon de prédiction. Sur chaque sous-figure nous avons indiqué la valeur du seuil de probabilité maximisant le F_{score} , ainsi que la valeur de ce F_{score} . À titre de comparaison, nous avons également indiqué l'emplacement du point correspondant à un seuil de probabilité de 0,5 (seuil correspondant au système d'alerte déterministe que nous avons présenté dans les sections précédentes). On constate que le passage à un système d'alerte probabiliste permet, en utilisant un seuil légèrement plus faible (compris entre 0,36 et 0,42 en fonction de l'horizon de prédiction) d'augmenter encore la valeur de F_{score} . Le fait que le seuil idéal soit ici légèrement plus faible que le seuil médian est sans doute dû au fait, vu plus haut avec le diagramme de fiabilité, que notre modèle a une tendance à la sous-estimation. Ainsi, émettre une alerte lorsque la probabilité d'observer un orage est légèrement inférieure à 0,5 permet de compenser cela et obtenir des alertes plus fiables.

3.5.3 Bilan sur l'amélioration du modèle par la prise en compte de l'incertitude

Nous avons proposé une approche permettant d'obtenir des prédictions probabilistes directement à l'aide d'un réseau de neurones directement dérivé de notre modèle déterministe amélioré. Cette approche – qui présente les avantages d'être facile à implémenter, de permettre l'utilisation de l'ensemble d'entraînement complet pour l'apprentissage et de présenter un surcoût informatique négligeable – nous fournit un modèle capable de rendre compte de son incertitude (bien que nous ne soyons pas capables de distinguer la part d'incertitude due aux données de la part d'incertitude imputable au modèle). Grâce à des outils d'évaluation dédiés, nous avons montré que notre modèle gagnait en compétence par rapport au modèle déterministe, avec des valeurs de **CRPS** toujours inférieures à la **MAE**

10. Ce seuil est idéal seulement en théorie. En pratique, si on souhaite un modèle respectant certaines contraintes en termes de **précision** ou de **rappel** il peut être préférable de choisir un autre seuil.

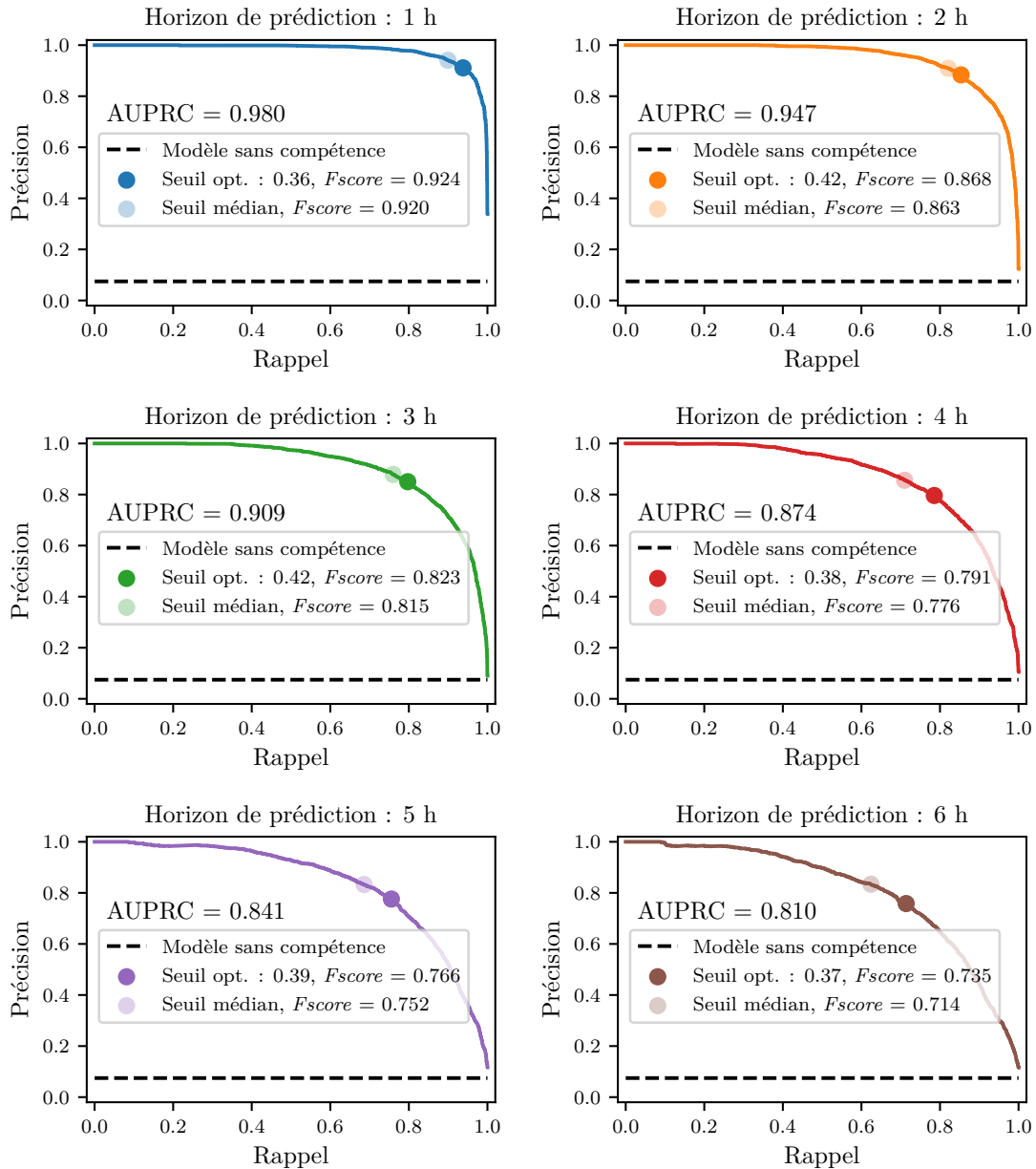


FIGURE 3.21 – Courbes PRC du modèle probabiliste pour les horizons de prédictions compris entre une et six heures. Les valeurs de l’AUPRC, des seuils optimaux et des F_{score} associés sont indiqués pour chaque horizon.

et la possibilité d’émettre des alertes probabilistes à partir desquelles on peut utiliser des seuils de décision meilleurs que le simple seuil médian (correspondant au modèle déterministe). Une telle approche représente donc une alternative intéressante aux approches proposées par CHANDORKAR et al. 2017; GRUET et al. 2018; CHAKRABORTY et MORLEY 2020 qui s’appuient sur des modèles coûteux combinant réseaux de neurones et processus gaussiens ou *Gaussian Processes*.

3.6 Bilan du chapitre

Dans ce chapitre nous avons construit un nouveau modèle de prédiction de Dst à partir des paramètres du vent solaire mesurés en L1. Ce modèle, que nous avons bâti à partir du modèle pré-existant de GRUET et al. 2018 propose des prédictions plus précises, surtout en périodes d'orages en intenses. De plus, son utilisation en tant que système d'alerte est plus performante : avec un horizon d'une heure, nous prédisons 20 % d'orages en plus que le modèle de base. Néanmoins, en complétant notre banc d'évaluation à l'aide de métriques dédiées à l'évaluation de caractéristiques précises du modèle, nous montrons que celui-ci présente une tendance à être en retard. Nous pouvons améliorer la compétence du modèle en le transformant en modèle probabiliste avec un faible coût computationnel. Toutefois cela n'empêche que nos prédictions se dégradent avec l'augmentation de l'horizon de prédiction et, comme d'autres études avant nous, nous trouvons que les performances de notre modèle restent acceptables pour un horizon maximal autour de 6 h.

Notons qu'ici nous nous sommes intéressés uniquement à la prédiction de l'indice Dst . Dans les faits, nous avons également adapté notre modèle pour la prédiction de l'indice Kp et nous avons obtenu des performances similaires à celles d'autres études ayant été publiées au cours des trois dernières années, comme celles de TAN et al. 2018 ; SHPRITS et al. 2019 ; CHAKRABORTY et MORLEY 2020. Nous ne détaillons pas dans ce manuscrit notre modèle de prédiction de Kp à court terme, car cela serait redondant à la fois avec ce qui a déjà été présenté dans ce chapitre, et ce qui est décrit dans les papiers que nous venons de citer. De plus, avec notre modèle de prédiction de Kp nous parvenons aux mêmes conclusions et retrouvons un horizon de prédiction limite de l'ordre de quelques heures. Pour des horizons de prédictions supérieurs, des modèles statistiques simples, comme le modèle de persistance ou le modèle de récurrence, restent les meilleurs. Nous nous intéressons à la question de la prédiction de Kp pour des horizons supérieurs à deux jours à l'aide d'une méthode innovante dans les Chapitres 5 et 6.

Avant cela, c'est une autre limite de nos modèles que nous souhaitons étudier. En effet, Kp et Dst sont des indices ayant une utilité opérationnelle limitée, notamment pour des opérateurs de satellites, pour lesquels connaître plus précisément la dynamique des flux de particules dans les ceintures de radiations est primordial. Afin de rendre nos prédictions plus utiles dans un tel contexte, nous étudions dans le Chapitre 4 un nouvel indice géomagnétique, Ca , que nous décrivons, dont nous justifions la pertinence, que nous étudions statistiquement et pour lequel nous proposons un modèle de prédiction dérivé de celui que nous venons de présenter.

Chapitre 4

Estimer la dynamique future des ceintures de radiations électroniques

Un homme courait après un autre alors que celui qui s'enfuyait courait à son tour après un troisième, lequel, sans pressentir qu'on lui courait après, marchait simplement d'un pas vif sur le trottoir.

(Daniil Harms, 1940)

Sommaire

Préambule	138
4.1 Prendre en compte la géoefficacité du point de vue des ceintures de radiations	139
4.1.1 Les limites des approches existantes	139
4.1.2 Définition de l'indice Ca	140
4.1.3 Étude statistique de Ca	141
4.1.3.1 Comparaison de Ca avec les flux d'électrons dans les ceintures	141
4.1.3.2 Éléments sur la distribution de Ca	145
4.1.3.3 Étude et caractérisation des événements extrêmes	146
4.2 Prédiction de l'indice Ca	150
4.2.1 Les données utilisées	150
4.2.1.1 Les jeux de données	150
4.2.1.2 Séparation des données pour entraînement et évaluation	151
4.2.1.3 Pré-traitement des données	152
4.2.2 Le modèle	152
4.2.2.1 Architecture du modèle	152
4.2.2.2 Émission d'une alerte	154
4.2.3 Évaluation du modèle	156
4.2.3.1 Les métriques d'évaluation	156
4.2.3.2 Catalogues d'événements	157
4.2.3.3 Modèle de référence	157
4.2.4 Analyse des résultats	158
4.2.4.1 Régression	158
4.2.4.2 Classification	163
4.2.5 Bilan sur la prédiction de Ca	166
4.3 Bilan du chapitre	167

Préambule

Dans ce chapitre nous souhaitons mieux prendre en compte dans nos prédictions la géoeffectivité des perturbations magnétosphériques du point de vue des ceintures de radiations terrestres. Pour ce faire, nous commençons par définir un nouvel indice géomagnétique Ca , conçu pour mieux rendre compte de la géoeffectivité des événements géomagnétiques du point de vue des ceintures de radiations électroniques. Après l'avoir présenté, nous proposons une analyse statistique de cet indice, en se concentrant notamment sur sa corrélation avec les flux d'électrons dans les ceintures de radiations. Enfin, nous appliquons des méthodes similaires à celles présentées dans le Chapitre 3 afin de prédire l'indice Ca .

Les travaux originaux présentés dans ce chapitre ont fait l'objet de deux publications dans des revues à comité de lecture. Le premier article, BERNOUX et MAGET 2020, a été publié en mai 2020 dans la revue *Space Weather* de l'*American Geophysical Union* (AGU). Le second, BERNOUX et al. 2021, a été

publié en novembre 2021 dans le Journal of Space Weather and Space Climate (EDP Sciences). Ces articles ont été publiés de façon ouverte, le premier selon une licence CC BY-NC 4.0¹, le second selon une licence CC BY 4.0². Les *abstracts* des deux articles sont donnés en Annexe C.

4.1 Prendre en compte la géoefficacité du point de vue des ceintures de radiations

4.1.1 Les limites des approches existantes

Dans le Chapitre 3 nous avons utilisé des réseaux de neurones afin de prédire les indices géomagnétiques *Dst* et *Kp* à partir de paramètres du vent solaire. Comme nous l'avons vu, ces indices sont très utilisés dans la communauté afin de définir des seuils correspondant à des orages géomagnétiques de différentes intensités. On peut donc, à partir de ces indices, définir des seuils d'alerte correspondant à des niveaux de danger généraux pour des opérateurs de satellites.

Il convient toutefois de rappeler que ces indices ont des limitations physiques. Par exemple, BOROVSKY et SHPRITS 2017 indique clairement que l'indice *Dst* est incapable de rendre compte de tous les « types » d'orages géomagnétiques et qu'il s'agit en réalité d'un indice très médiocre pour l'étude de la dynamique des flux d'électrons dans les ceintures de radiations externes, pilotée par des perturbations de longue durée provoquées par des CIRs. C'est pourquoi il est important de ne pas diriger l'effort de recherche uniquement sur le problème de la prévision des indices *Kp* et *Dst*, mais de diversifier les indices étudiés, afin d'inclure une plus grande diversité de phénomènes liés à la météorologie de l'espace.

Ici nous souhaitons mieux prendre en compte la géoeffectivité des événements du point de vue des ceintures d'électrons. Une façon d'aborder ce problème est de combiner un modèle de prédiction de l'indice *Kp* avec un modèle physique des ceintures de radiations, comme le modèle Salammbô de l'ONERA, qui est piloté par *Kp* (BOURDARIE et MAGET 2012). La principale difficulté que pose cette approche réside alors dans la gestion des incertitudes et erreurs dues à chacun des modèles et à leur couplage. De plus, on perd alors un des avantages de l'utilisation des réseaux de neurones qui réside dans leur rapidité à obtenir une prédiction (alors que les modèles physiques peuvent être plus coûteux en termes de temps de calcul).

Une autre approche serait donc de construire un réseau de neurones pour prédire directement les flux de particules dans les ceintures de radiations à partir des paramètres du vent solaire et d'indices géomagnétiques, comme le proposent par exemple CHEN et al. 2019. Bien que cette approche, si elle aboutit, peut donner les résultats les plus fins, elle présente de nombreuses difficultés, comme par exemple le fait qu'il faille potentiellement plusieurs modèles par gamme d'énergie

1. Creative Commons Attribution – Pas d'Utilisation Commerciale 4.0 International
2. Creative Commons Attribution 4.0 International

des flux de particules, mais aussi en fonction des orbites, ce qui en fait un problème de haute dimension. Pour pouvoir correctement le résoudre, il faut donc beaucoup de données, ce qui est difficile à obtenir car les moniteurs de radiations *in situ* ont une durée de vie limitée, et l'inter-calibration des séries temporelles obtenues à l'aide de différents moniteurs n'est pas forcément un problème évident.

Ainsi l'approche que nous privilégions ici est de construire puis utiliser un indice qui soit un indicateur général de l'état de remplissage des ceintures de radiations, obtenu à partir de mesures effectuées au sol. Une telle méthode est certes potentiellement moins précise que travailler directement avec les flux de particules dans les ceintures de radiations, mais elle présente de nombreux avantages, que nous détaillons dans les sections suivantes.

4.1.2 Définition de l'indice Ca

Afin de mieux prendre en compte la géoeffectivité d'un événement, ROCHEL et al. 2016 propose une méthode pour définir l'indice géomagnétique Cm , construit à partir de l'indice am (MAYAUD 1968). En suivant la méthodologie proposée par ROCHEL et al. 2016, nous avons fondé notre étude sur l'utilisation d'un indice géomagnétique intégré temporellement, que nous appelons indice Ca , et qui est construit à partir de l'indice géomagnétique aa . Ca a été conçu pour quantifier la géoefficacité des structures du vent solaire impactant la magnétosphère du point de vue des ceintures de radiations électroniques.

L'indice aa – que nous avons déjà mentionné au Chapitre 2 – est un indice tri-horaire dérivé de K , proposé pour la première fois par MAYAUD 1971. Ses premières valeurs pour la période 1868-1968 ont été publiées pour la première fois dans MAYAUD 1972. L'indice aa est calculé à partir des mesures de deux observatoires subauroraux opposés, le premier étant au Royaume-Uni et le second en Australie. L'intérêt d'utiliser aa est de disposer d'un indice pouvant remonter jusqu'à 1868, ce qui nous donne plus de 150 ans de données, soit environ 14 cycles solaires (c'est d'ailleurs la principale raison qui nous a poussé à utiliser Ca plutôt que Cm , qui ne dispose « que » d'environ 60 ans de données). Il s'agit de l'indice géomagnétique ayant le plus grand historique. Malgré sa grande étendue temporelle, cet indice est considéré comme très homogène et constitue un moyen simple et facile de surveiller l'activité géomagnétique planétaire (MAYAUD 1980). L'indice aa prend des valeurs positives non-bornées et son unité est le nanotesla (nT). Plus la valeur de aa est élevée, plus la perturbation de la magnétosphère est forte. L'indice aa est disponible sur le site de l'ISGI³.

Le temps caractéristique de relaxation dans les ceintures de radiation pour les électrons de haute énergie (ordre de grandeur supérieur à 100 keV) après une forte perturbation magnétosphérique est de l'ordre de 4 jours (MEREDITH et al. 2006; ROCHEL et al. 2016). Par conséquent nous définissons l'indice Ca comme suit :

3. http://isgi.unistra.fr/indices_aa.php

$$Ca(t) = \frac{1}{\tau} \int_0^{\infty} aa(t-t') e^{-\frac{t'}{\tau}} dt', \quad (4.1)$$

avec $\tau = 4$ jours représentant le temps caractéristique de relaxation des ceintures et aa représentant l'activité géomagnétique.

Définir un indice intégré dans le temps permet une meilleure prise en compte de multiples événements rapprochés et donc de l'accumulation d'énergie dans les ceintures qui en résulte et qui pourrait expliquer certains événements extrêmes (LUGAZ et al. 2015 ; BENACQUISTA et al. 2017). Par ailleurs, il a été montré dans ROCHEL et al. 2016 – et nous le détaillons encore plus dans la section suivante – que Ca est bien corrélé aux flux d'électrons énergétiques dans les ceintures de radiations (les corrélations les plus élevées étant trouvées autour de $L^* \approx 4$).

Afin que le lecteur ait une idée plus précise du comportement typique de Ca pendant les périodes de perturbation causées par des CIRs et des ICMEs, les Figures 4.1 et 4.2 montrent la dynamique des indices géomagnétiques Kp , aa et Ca avec deux paramètres du vent solaire (la vitesse et la pression dynamique), et les flux d'électrons dans les ceintures de radiation pour trois gammes d'énergies ($0,309 < E < 0,379$ MeV, $1,539 < E < 1,995$ MeV, $3,299 < E < 3,969$ MeV) mesurés par l'instrument MagEIS à bord du satellite *Radiation Belt Storm Probe* (RBSP) A (BLAKE et al. 2013) entre le 15 avril et le 15 mai 2017 (Figure 4.1) et entre le 28 septembre et le 28 octobre 2013 (Figure 4.2).

L'événement induit par une CIR (Figure 4.1) se caractérise par l'arrivée d'un vent solaire à grande vitesse et le remplissage des ceintures de radiation pour toutes les gammes d'énergie, en particulier avec des électrons de très haute énergie (3,299 – 3,969 MeV). L'événement (triple) induit par les ICMEs (Figure 4.2) est caractérisé par une forte augmentation spontanée de la pression dynamique ainsi que par une augmentation modérée de la vitesse du vent solaire, tandis que les ceintures de radiations sont remplies principalement d'électrons dans la gamme d'énergie $0,309 < E < 0,379$ MeV, aucune augmentation n'étant observée dans les flux d'électrons de très haute énergie (3,299 – 3,969 MeV). Alors que l'événement induit par une CIR est celui qui remplit le plus les ceintures de radiation dans toutes les gammes d'énergie et cause la plus grande accumulation d'énergie, on observe des pics plus élevés pour les indices géomagnétiques aa et Kp lors de l'événement induit par la triple ICME. On peut alors voir que Ca augmente et diminue plus lentement que aa , mais il atteint des valeurs plus élevées (au-dessus de 37,5 nT) dans le premier cas que dans le second (au-dessous de 25 nT), montrant une meilleure cohérence avec la géoefficacité du point de vue des ceintures de radiation que les indices aa et Kp .

4.1.3 Étude statistique de Ca

4.1.3.1 Comparaison de Ca avec les flux d'électrons dans les ceintures

Comme mentionné dans la section précédente, l'indice Ca a été défini pour caractériser la géoefficacité du point de vue des ceintures de radiations d'électrons.

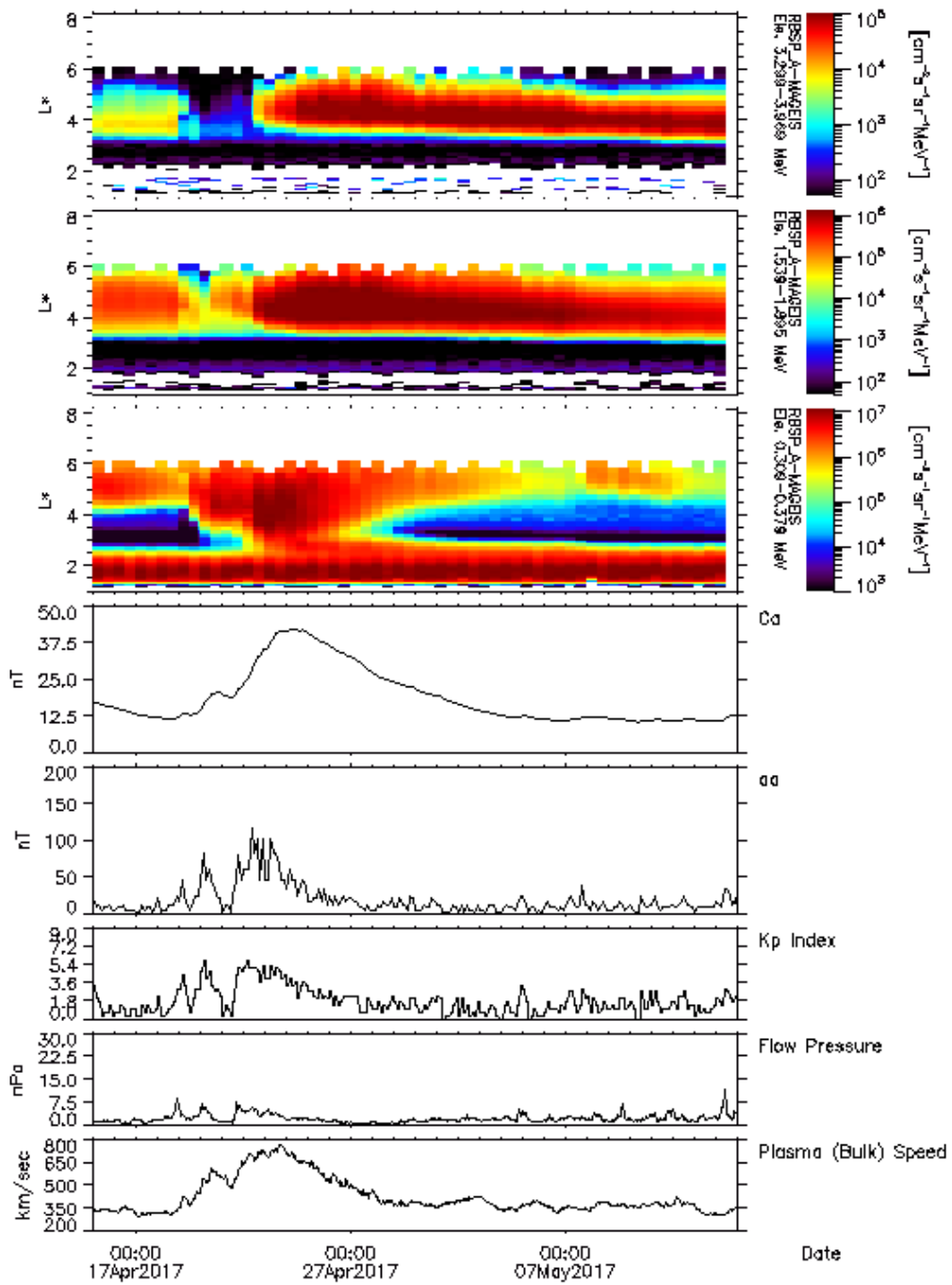


FIGURE 4.1 – Évolution des flux d'électrons dans les ceintures de radiation pour trois gammes d'énergies ($0,309 < E < 0,379$ MeV, $1,539 < E < 1,995$ MeV, $3,299 < E < 3,969$ MeV) mesurés par l'instrument MagEIS à bord du satellite RBSP A entre le 15 avril et le 15 mai 2017, pendant un orage induit par une CIR. L'évolution des indices géomagnétiques Ca , aa et Kp avec deux paramètres du vent solaire (pression dynamique et vitesse) est également donnée. Reproduit de BERNOUX et MAGET 2020.

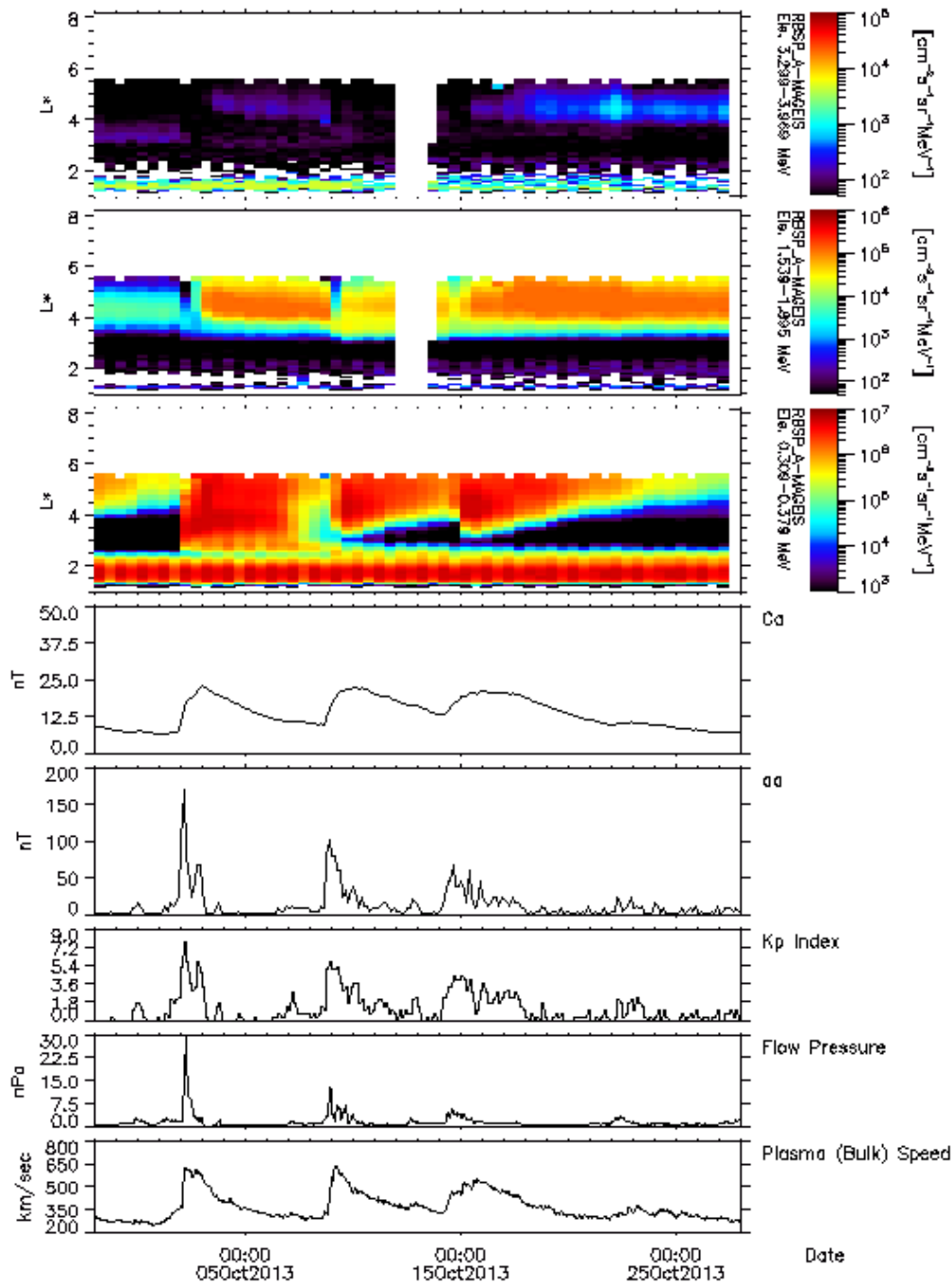


FIGURE 4.2 – Évolution des flux d'électrons dans les ceintures de radiation pour trois gammes d'énergies ($0,309 < E < 0,379$ MeV, $1,539 < E < 1,995$ MeV, $3,299 < E < 3,969$ MeV) mesurés par l'instrument MagEIS à bord du satellite RBSP A entre le 28 septembre et le 28 octobre 2013, pendant un orage induit par une triple ICME. L'évolution des indices géomagnétiques Ca , aa et Kp avec deux paramètres du vent solaire (pression dynamique et vitesse) est également donnée. Reproduit de BERNOUX et MAGET 2020.

La convolution temporelle effectuée dans le calcul de l'indice Ca permet de prendre en compte la réaction induite et non-immédiate du système des ceintures de radiation lorsqu'une perturbation intense affecte la magnétosphère. L'énergisation induite par les interactions ondes-particules est en particulier bien corrélée avec cet indice. Ainsi, le transfert global d'énergie du système couplé vent solaire-magnétosphère-ceintures de radiation est bien quantifié au premier ordre. La Figure 4.3 présente la corrélation linéaire (R) directe des indices aa et Ca avec les flux d'électrons d'énergie $E \geq 100$ keV (avec et sans passage au logarithme) mesurés par le satellite NOAA POES-15, entre juillet 1998 et décembre 2015. Elle montre clairement l'amélioration du coefficient de corrélation linéaire lorsque l'on utilise l'indice Ca . En particulier, cette amélioration est bien marquée pour toute valeur de L à l'intérieur du domaine des ceintures de radiation ($2.5 \leq L \leq 6$).

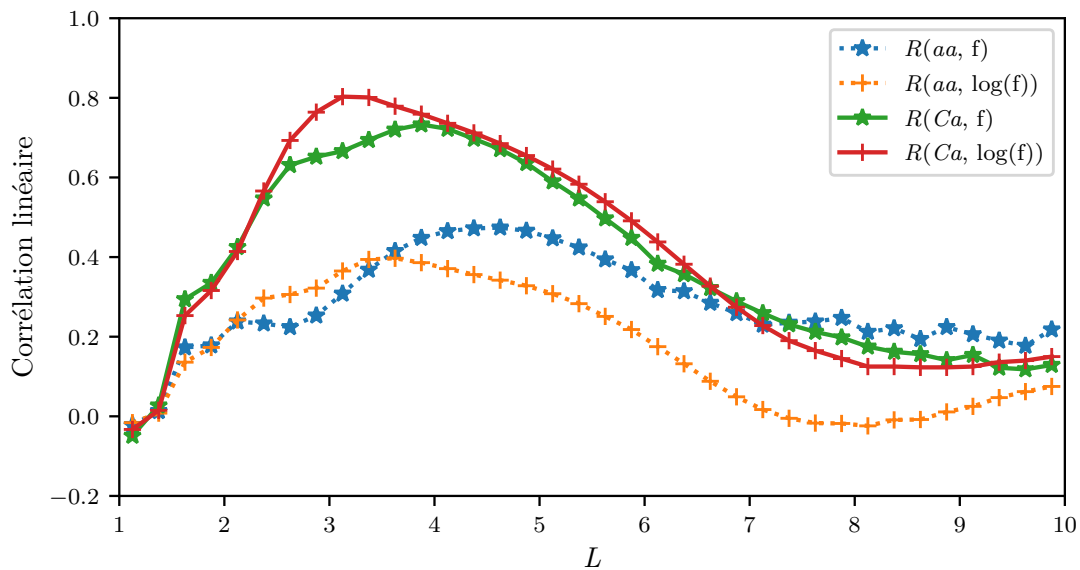


FIGURE 4.3 – Coefficients de corrélation entre les flux d'électrons $E > 100$ keV (f) mesurés par l'instrument NOAA POES-15/MEPED et aa d'un côté et Ca de l'autre, en fonction du paramètre L . Différentes combinaisons ont été testées. La période utilisée pour calculer ces corrélations s'étend de juillet 1998 à décembre 2015. Adapté de BERNOUX et MAGET 2020.

Ces simples estimations de corrélation sont très intéressantes et montrent la capacité d'un tel indice à mieux caractériser les événements du point de vue de la quantité d'énergie transférée aux ceintures de radiation. Bien sûr, comme tout indice simple, il ne peut pas quantifier totalement ce processus complexe. Certaines limites sont évidemment observables. Pour nous en rendre compte, nous avons tracé dans la Figure 4.4 les coefficients de corrélation entre Ca et le logarithme du flux pour différentes énergies. En particulier, lorsque l'énergie augmente, le coefficient de corrélation a tendance à diminuer, comme on peut l'observer pour le flux d'énergie $E > 900$ keV dans la Figure 4.4.

Ceci est principalement dû au fait que le temps caractéristique de 4 jours utilisé pour la convolution a été choisi pour s'adapter au mieux aux observations

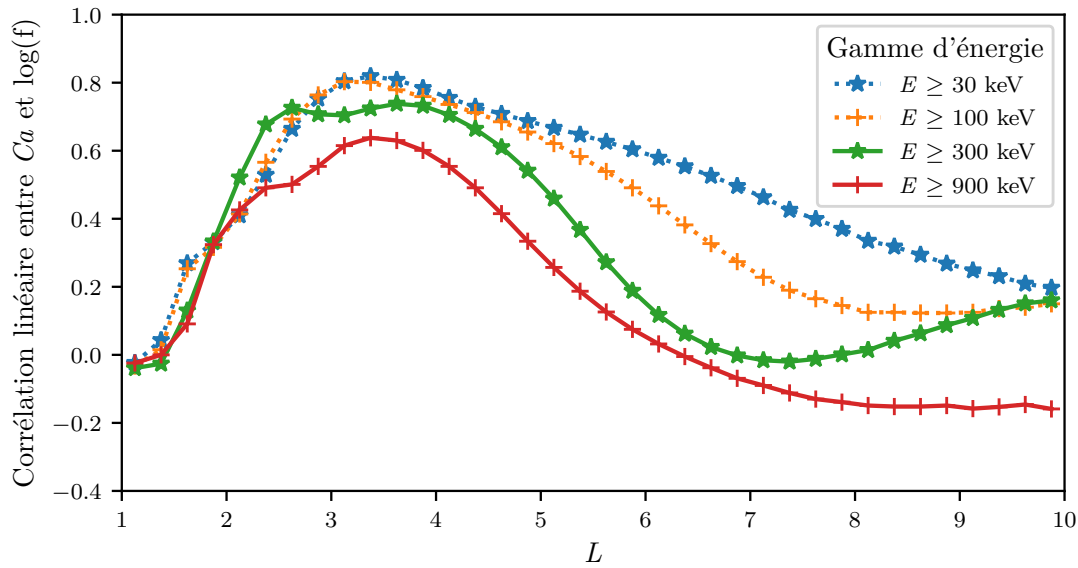


FIGURE 4.4 – Coefficients de corrélation entre les flux d'électrons (passés en échelle logarithmique) mesurés par l'instrument NOAA POES-15/MEPED pour différentes énergies et l'indice Ca , en fonction du paramètre L . La période utilisée pour calculer ces corrélations s'étend de juillet 1998 à décembre 2015.

pour toutes les valeurs de L et toutes les énergies. Par conséquent, l'indice Ca sera moins bien corrélé aux observations qui se situent dans les valeurs extrêmes de la distribution de ces deux paramètres. Cependant, ceci peut être modifié si on souhaite se concentrer sur une gamme d'énergie ou une orbite particulière. Enfin, on peut voir qu'en *Orbite géostationnaire (GEO)*, pour des valeurs de L proche de 6, le coefficient de corrélation tend à diminuer pour toutes les énergies. Cela met en évidence les différences dans la dynamique des ceintures de radiations entre cette orbite particulière, proche de la limite extérieure des ceintures de radiations, et d'autres valeurs de L plus petites, au cœur des ceintures de radiations. Les processus prédominants n'y sont pas les mêmes, et la dynamique n'est pas pilotée de la même manière.

4.1.3.2 Éléments sur la distribution de Ca

La distribution empirique des valeurs de Ca prises entre 1869 et 2018 (inclus) est donnée en Figure 4.5. On y remarque une très forte concentration de la distribution dans les faibles valeurs de Ca . Cela se traduit par des valeurs de quartiles très resserrées : le premier quartile vaut $Ca_{Q1} = 12,4$ nT, la médiane vaut $Ca_{Q2} = 17,6$ nT et le troisième quartile vaut $Ca_{Q3} = 24,4$ nT, tandis que 95 % des valeurs de Ca sont inférieures à $Ca_{0,95} = 38,6$ nT. Notons que la moyenne de Ca sur cette période est de $Ca_{moy} = 19,6$ nT, sa valeur minimale est de $Ca_{min} = 2,8$ nT, et son maximum est de $Ca_{max} = 160,9$ nT, ce qui est une valeur plus de quatre fois plus grande que $Ca_{0,95}$.

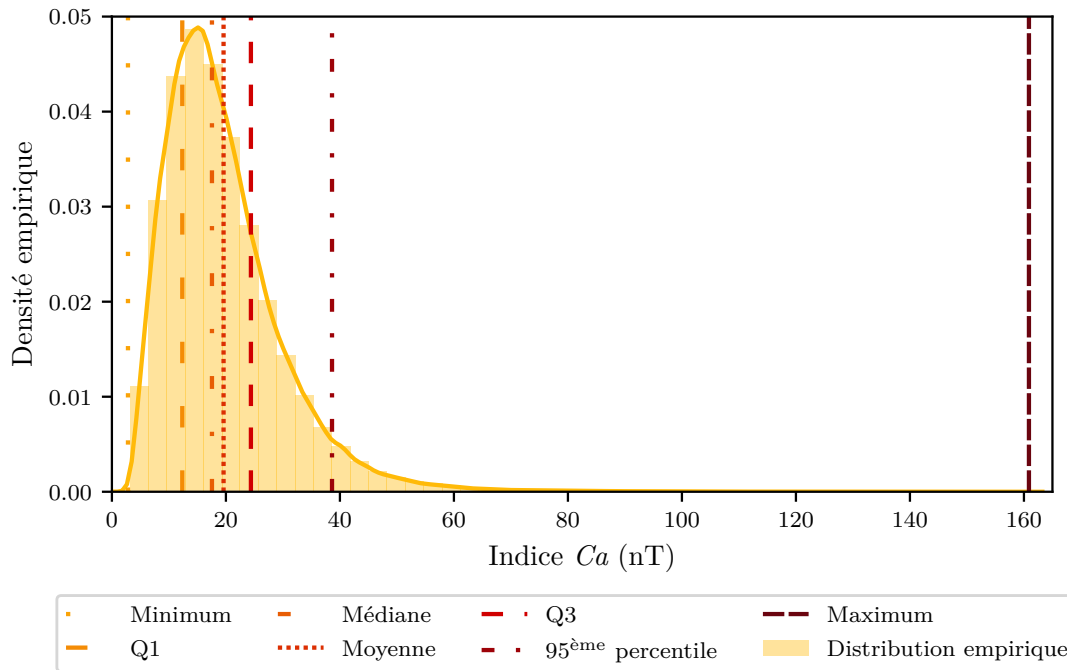


FIGURE 4.5 – Distribution empirique des valeurs de Ca . Quelques repères statistiques sont indiqués.

On observe de plus que Ca peut adopter une dynamique très différente en fonction de la période du cycle solaire dans laquelle on se situe. La Figure 4.6 montre l'évolution de la moyenne annuelle de Ca entre 1869 et 2018. On voit bien que différentes années peuvent avoir des valeurs moyennes de Ca très éloignées, traduisant des périodes à l'activité très différente.

Ainsi, une année comme 2003 va avoir une valeur moyenne annuelle de Ca égale à 36,8 nT, alors que 2009 va avoir une valeur moyenne annuelle de Ca égale à 8,8 nT. Cette différence de dynamique à l'échelle annuelle se retrouve également dans des échelles plus courtes, comme le montre la Figure 4.7. On y observe que Ca varie beaucoup en 2003 (écart-type de 14,9 nT), y atteignant même les valeurs les plus hautes mesurées depuis le début de l'ère spatiale (147,6 nT en novembre 2003), alors qu'en 2009 il n'y a quasiment pas de dynamique (écart-type de 2,5 nT), ce qui est la conséquence directe d'une activité solaire de faible intensité.

Enfin, la Figure 4.8 nous montre que Ca est un indice qui présente une très forte auto-corrélation temporelle, comme on pouvait s'y attendre étant donné qu'il s'agit d'un indice intégré dans le temps. La remontée de l'auto-corrélation pour un décalage temporel d'environ 27 jours indique que Ca est bien corrélé avec les structures persistantes entre deux rotations du Soleil.

4.1.3.3 Étude et caractérisation des événements extrêmes

Outre sa bonne corrélation avec l'état de remplissage des ceintures de radiations, Ca présente la caractéristique de posséder un historique de 150 ans de

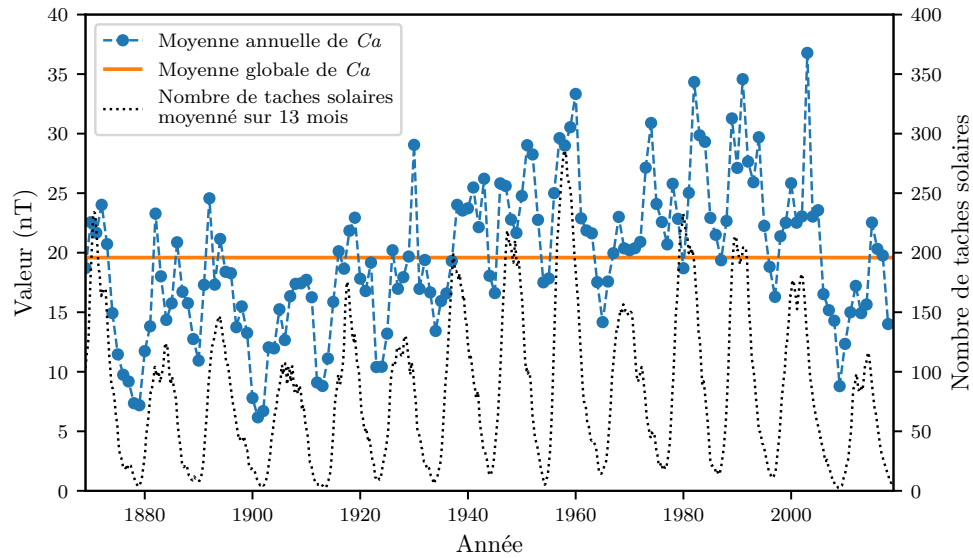


FIGURE 4.6 – Évolution de la moyenne annuelle de Ca entre 1869 et 2018. La moyenne globale de Ca sur l'ensemble de la période est donnée pour repère. Le nombre mensuel de taches solaires lissé sur 13 mois est également tracé comme témoin du cycle solaire.

données. Cela nous permet d'étudier de manière précise les événements géomagnétiques extrêmes – et donc rares – vus selon Ca . L'intérêt d'une telle étude est de nous permettre de mieux comprendre cet indice nouveau en caractérisant la queue de sa distribution et en répondant aux questions suivantes : peut-on définir un événement décennal ? Centennal ? A-t-on déjà connu des événements extrêmes depuis le début de l'ère spatiale ? Répondre à ces questions est primordial, surtout si on veut mieux caractériser d'éventuelles prédictions de Ca . En effet, une prédiction de l'indice Ca n'est véritablement utile que si l'on peut estimer qualitativement la dangerosité d'une perturbation à partir d'une valeur quantitative de Ca . Comme les indices Kp et Dst sont très utilisés dans la communauté, il y a déjà de nombreuses études portant sur les statistiques de ces indices, ce qui permet, par exemple, de définir des seuils d'alerte. Comme ce n'est pas le cas pour Ca , nous proposons ici un récapitulatif de notre étude sur l'analyse des événements extrêmes vus selon l'indice Ca , publiée dans l'article BERNOUX et MAGET 2020.

Afin d'étudier les événements géomagnétiques extrêmes, nous utilisons un ensemble de méthodes spécifiques d'un domaine appelé *analyse des valeurs extrêmes* ou *Extreme Value Analysis (EVA)*. L'EVA est largement utilisée dans de nombreux domaines de recherche différents, y compris en météorologie de l'espace. Ainsi, des études appliquent la théorie des valeurs extrêmes afin d'analyser les flux d'électrons énergétiques dans les ceintures de radiation (par exemple KOONS 2001 ; MEREDITH et al. 2015) ou pour modéliser l'activité géomagnétique, soit du point de vue des indices géomagnétiques (SISCOE et al. 2006) ou directement à partir de données magnétiques brutes au sol (THOMSON et al. 2011).

Dans son livre de référence sur l'EVA, COLES 2001 met en évidence deux principales techniques dédiées à l'étude des événements extrêmes. La première

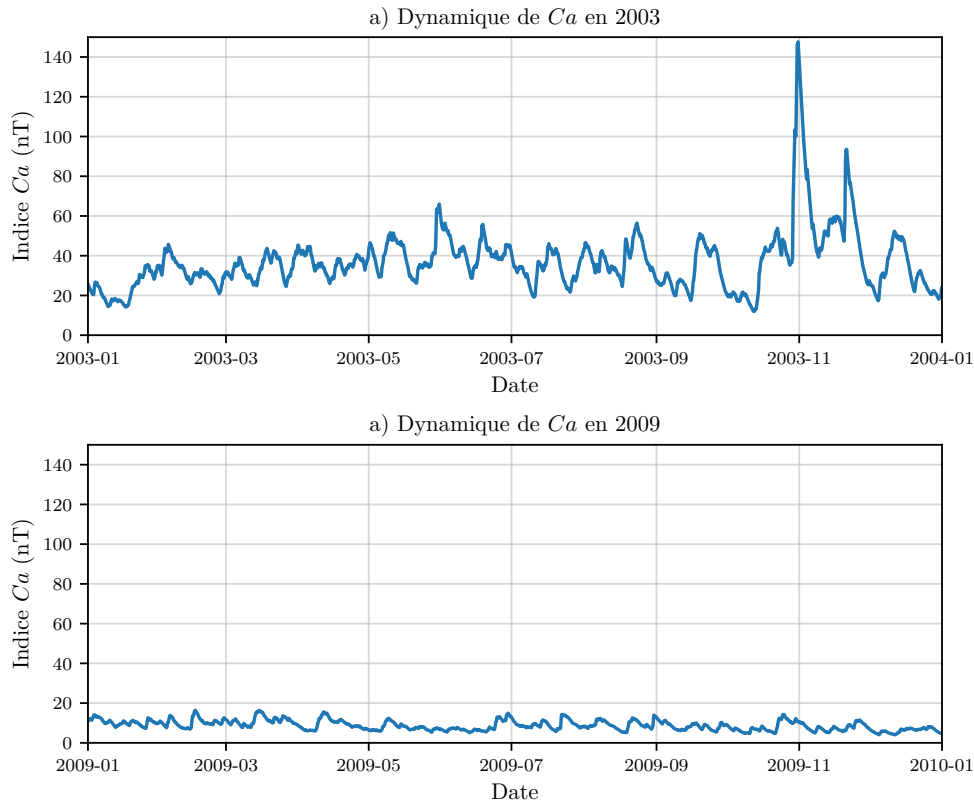


FIGURE 4.7 – Évolution de la dynamique de Ca pour a) 2003, une année très perturbée, et b) 2009, une année très calme.

méthode est basée sur une approche dite de « maximums par blocs », adaptée aux situations où des blocs de données sont clairement identifiables ou bien lorsque seules les valeurs maximales pour une période fixe sont disponibles, ce qui n'est pas notre cas. La seconde méthode, appelée **POT**, consiste à estimer les paramètres d'une **distribution de Pareto généralisée (DPG)** à partir de tous les échantillons supérieurs à une certaine valeur seuil (PICKANDS 1975). Cette deuxième approche est plus adaptée à notre étude puisqu'elle permet la prise en compte de multiples événements extrêmes survenus successivement, et, en contrepartie, le modèle ne prend pas en compte les périodes où aucun orage géomagnétique ne s'est produit.

Notons Z une variable aléatoire suivant une **DPG**, u une valeur dite seuil et $z > u$. Alors on a :

$$P(Z > z | z > u) = 1 - \left[1 + K \frac{z - u}{S} \right]^{-\frac{1}{2}}, \quad (4.2)$$

où K et S sont les paramètres de forme et d'échelle. Lorsqu'on applique la méthode **POT**, la première étape consiste à trouver (à l'aide de méthodes empiriques dédiées) la valeur optimale pour le seuil u , au-delà de laquelle nous nous trouvons dans la queue de la distribution de Ca et donc au-delà de laquelle nous pouvons dire qu'il s'agit d'un événement extrême. Il faut ensuite estimer les paramètres K et S (par exemple par la méthode du maximum de vraisemblance). Lorsque les trois paramètres K , S et u sont connus nous connaissons alors la distribution

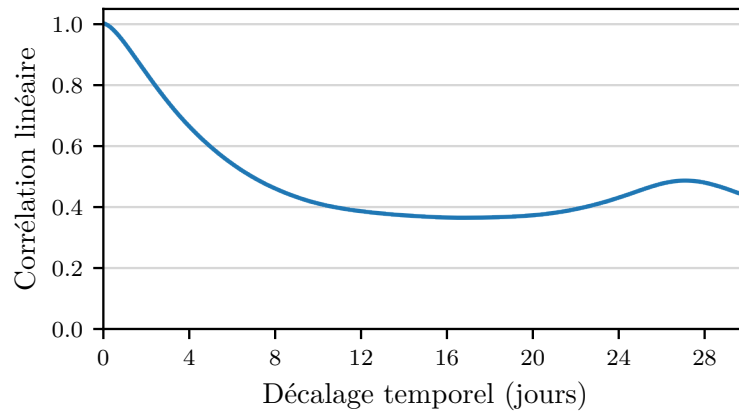


FIGURE 4.8 – Auto-corrélation temporelle de Ca pour des décalages temporels allant jusqu'à 30 jours.

des événements extrêmes de Ca et nous pouvons par exemple en déduire le niveau correspondant à un événement décennal⁴.

Toute la méthodologie et les calculs sont détaillés dans BERNOUX et MAGET 2020. Ici nous donnons simplement les principaux résultats obtenus. Tout d'abord il apparaît que la valeur de Ca au-dessus de laquelle nous pouvons considérer qu'un événement est extrême est $u = 67$ nT. Les valeurs correspondant aux événements décennaux, cinquennaux et centennaux sont donnés en Table 4.1.

TABLEAU 4.1 – Valeurs estimées par méthode POT des événements géomagnétiques décennaux, cinquennaux et centennaux selon l'indice Ca .

Période de retour de l'événement (ans)	Niveau de Ca estimé (nT)	Intervalle de confiance à 95 % (nT)
10	100,4	[93,1, 109,2]
50	131,4	[119,7, 156,1]
100	142,8	[128,9, 182,0]

Ainsi selon cette étude nous avons déjà observé de nombreux événements extrêmes depuis le début de l'ère spatiale. Par exemple, l'orage du 14 mars 1989 serait un événement quasi-cinquennal, tandis que l'orage de Halloween 2003 serait un événement plus que centennal. Notons que l'événement de mars 1989 était l'événement le plus extrême du point de vue de Dst (depuis 1963). Cela illustre bien que l'utilisation de Ca n'est pas redondante avec celle de Dst par exemple.

Maintenant que nous avons caractérisé statistiquement l'indice Ca , y compris du point de vue des événements extrêmes, nous présentons un modèle de prédiction de Ca s'appuyant sur nos acquis du Chapitre 3.

4. C'est-à-dire un événement qui revient en moyenne tous les 10 ans.

4.2 Prédiction de l'indice Ca

Dans cette section nous allons utiliser des modèles s'inspirant de ceux que nous avons présentés au Chapitre 3 pour prédire Ca jusqu'à 24 h en avance. Nous commencerons par décrire les données utilisées, avant de détailler l'architecture du modèle développé et les outils d'évaluation employés, ce qui inclut la définition d'un seuil d'alerte. Enfin nous présenterons les résultats obtenus et les analyserons.

D'un point de vue opérationnel, la prédiction de l'indice Ca pourrait servir de base à un service d'alerte pour l'accumulation d'électrons de haute énergie dans les ceintures de radiations. Dans un tel contexte, l'indice Ca agirait comme une approximation des flux d'électrons relativistes, obtenue grâce à des magnétomètres au sol. Utiliser Ca pour prédire la dynamique globale des flux d'électrons à l'aide de méthodes pilotées par les données a de nombreux avantages.

Premièrement, il s'agit d'un indice scalaire, donc de dimension faible, ce qui facilite l'apprentissage d'un modèle. On dispose également de plus de 150 années de données calibrées et sans « trous », ce qui est un atout pour toute étude statistique, incluant la prédiction de l'indice par des réseaux de neurones. Par ailleurs, la simplicité dans la construction de cet indice (rappelons que l'indice aa est obtenu à partir des mesures faites par seulement deux observatoires, contre par exemple 13 pour l'indice Kp) est une garantie pour la disponibilité future de cet indice. Enfin, dans un contexte opérationnel, il peut être plus fiable en termes de continuité de service de s'appuyer sur des instruments au sol plutôt que sur des instruments embarqués qui sont soumis aux risques liés à leur présence dans l'espace.

Ainsi, la prédiction de l'indice Ca est d'un intérêt immédiat pour les opérateurs de systèmes spatiaux. Par conséquent dans ce chapitre, de nombreux choix que nous effectuerons dans la construction de nos modèles le seront en prenant en compte le point de vue d'un tel opérateur, et nous essaierons autant que possible d'arbitrer nos décisions pour favoriser le caractère opérationnel du modèle.

4.2.1 Les données utilisées

4.2.1.1 Les jeux de données

Tout comme dans le Chapitre 3, nous utilisons comme entrées de nos modèles des paramètres du vent solaire mis à disposition dans la base de données OMNIweb. Ici, nous utiliserons les valeurs horaires des paramètres suivants :

- la vitesse du vent solaire V_{sw} ;
- la densité du vent solaire ρ_{sw} ;
- la composante B_z de l'IMF ;
- la température du vent solaire T_{sw} .

Nous avons choisi ces paramètres du vent solaire car ils correspondent, pour les trois premiers, à ceux le plus souvent utilisés pour étudier le couplage vent solaire–magnétosphère, notamment à des fins de prédiction, et déjà utilisés dans le chapitre

précédent pour prédire Dst et Kp . Nous leur avons adjoint ici la température du vent solaire T_{sw} , car il s'agit d'une des sorties fréquentes des modèles physiques de propagation du vent solaire dans l'héliosphère, comme par exemple SWIFT ou EUHFORIA. Ainsi, il sera à l'avenir possible de facilement « chaîner » les modèles.

4.2.1.2 Séparation des données pour entraînement et évaluation

Pour les mêmes raisons qu'en section 3.2, nous utilisons les données comprises entre le 1er janvier 1995 et le 31 décembre 2018, et nous séparons nos jeux de données en sous-ensembles d'entraînement, de validation et de test de la façon suivante :

- entraînement : de 2003 à 2018 compris ;
- validation : de 1995 à 1996 compris ;
- évaluation : de 1997 à 2002 compris.

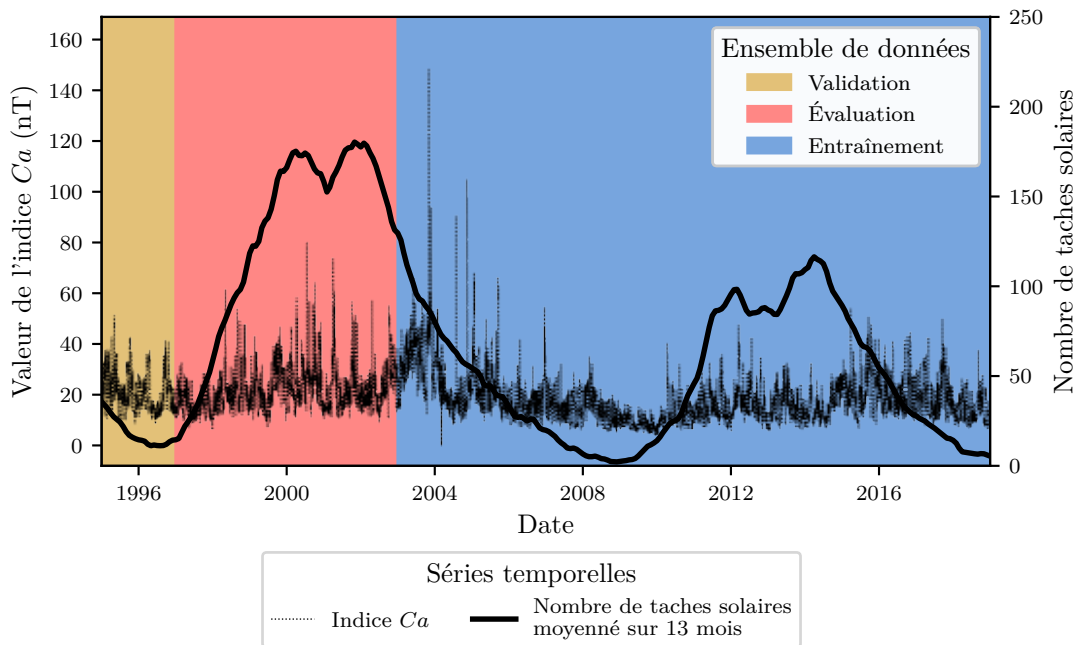


FIGURE 4.9 – Courbe des valeurs prises par l'indice Ca entre 1995 et 2018 inclus (ligne noire fine), les ensembles d'entraînement (zone verte), de validation (zone jaune) et de test (zone rouge) étant mis en évidence. Le nombre de taches solaires lissé sur 13 mois est également représenté comme indicateur du cycle solaire (ligne rouge épaisse).

Dans le cas de Ca cette séparation est toujours pertinente, et grâce à l'analyse des événements extrêmes menée dans BERNOUX et MAGET 2020 nous pouvons même affirmer que les ensembles d'entraînement et d'évaluation contiennent tous deux des événements extrêmes ($Ca \geq 67$ nT). L'ensemble d'entraînement contient même l'orage d'halloween de 2003, qui est, en termes de Ca , un événement plus rare que centennal (qui arrive, en moyenne, plus rarement qu'une fois tous les 100 ans), au cours duquel Ca a atteint une valeur de 147,6 nT. Le fait d'avoir à la

fois des valeurs usuelles de Ca , mais aussi des événements très rares dans nos jeux d'entraînement et d'évaluation nous permet d'assurer à la fois que nous entraînons notre modèle avec un large éventail de phénomènes (ce qui améliore le caractère généralisable du modèle), mais aussi que nous l'évaluons au cours de périodes de fortes perturbations, qui sont les périodes présentant le danger le plus grand pour les systèmes spatiaux. La Figure 4.9 montre la courbe de Ca (et celle du nombre de taches solaires, indiquant le cycle solaire) entre le 1er janvier 1995 et le 31 décembre 2018. Les sous-ensembles d'entraînement, de validation et d'évaluation y sont différenciés à l'aide de couleurs d'arrière-plan différentes.

4.2.1.3 Pré-traitement des données

Avant d'être injectées dans le réseau de neurones, les données sont pré-traitées comme suit :

- afin de nous ajuster à la fréquences des paramètres du vent solaire, nous interpolons les valeurs de l'indice Ca afin d'avoir des valeurs horaires au lieu d'une valeur toutes les 3 heures (cette opération a du sens puisque Ca est un indice intégré dans le temps très lisse et donc faire cette interpolation ne change ni la physique ni la statistique du problème) ;
- les valeurs manquantes dans les séries temporelles des paramètres du vent solaire sont remplies en utilisant la méthode de SSA ;
- les entrées sont re-échelonnées de manière à ce que leur moyenne soit égale à 0 et leur écart-type à 1. Les sorties sont re-échelonnées pour s'adapter à l'intervalle $[0, 1]$. Les coefficients pour effectuer ces transformations sont calculés uniquement à partir de l'ensemble de données d'entraînement afin de ne pas inclure de biais lors des étapes de validation et de test. Cette procédure est standard lorsqu'on travaille avec des réseaux récurrents.

4.2.2 Le modèle

4.2.2.1 Architecture du modèle

Mentionnons en tout premier lieu que le modèle que nous étudierons dans ce chapitre fournit des prédictions déterministes. Nous aurions pu appliquer la même méthode que celle présentée au Chapitre 3 pour prédire Dst de façon probabiliste à l'indice Ca . En réalité nous l'avons fait, mais nous présentons ici uniquement le modèle déterministe car les conclusions auxquelles nous aboutissons sont exactement les mêmes peu importe le modèle. Par conséquent, afin de nous concentrer sur l'essentiel et de ne pas diluer notre propos (qui est de savoir s'il est possible, ou non, de prédire Ca avec des réseaux de neurones et si oui, dans quelle mesure et avec quelle avance) nous présentons uniquement la version déterministe du modèle.

Le modèle développé dans cette étude reçoit en entrée les valeurs passées de quatre paramètres du vent solaire listés dans la section 4.2.1.1. Contrairement à d'autres études, et à ce que nous avons effectué au Chapitre 3, nous choisissons

cette fois-ci de ne pas inclure les valeurs passées de l'indice géomagnétique Ca en entrée du modèle car, comme nous l'avons expliqué plus haut, nous nous plaçons dans un contexte opérationnel. En effet, même si l'ISGI fournit des valeurs de l'indice aa de type *quick-look*⁵, le fait de s'appuyer sur deux sources de données différentes présente toujours un risque plus élevé d'indisponibilité des données d'une source, ce qui est préjudiciable lors de la mise en place d'un service de prévision en temps quasi réel. L'idéal pour un tel service serait d'avoir deux modèles, avec et sans les indices géomagnétiques historiques comme entrées, mais cela sort du cadre de cette étude et pour plus de clarté et de concision nous n'étudions ici qu'une seule approche. Nous utilisons les valeurs des 30 derniers jours pour chaque entrée (*i. e.* les 720 dernières valeurs horaires). Le lien entre les entrées et les sorties du modèle peut être résumé comme suit :

$$\begin{pmatrix} V_{sw}(t-719) & \dots & V_{sw}(t-1) & V_{sw}(t) \\ \rho_{sw}(t-719) & \dots & \rho_{sw}(t-1) & \rho_{sw}(t) \\ B_z(t-719) & \dots & B_z(t-1) & B_z(t) \\ T_{sw}(t-719) & \dots & T_{sw}(t-1) & T_{sw}(t) \end{pmatrix} \longrightarrow \begin{pmatrix} Ca(t+1) \\ Ca(t+2) \\ \dots \\ Ca(t+n) \end{pmatrix},$$

où n est l'horizon de prédiction maximal. En pratique nous prendrons $n = 24$ h dans la suite.

Notons qu'ici nous obtenons les prédictions Ca pour tous les horizons temporels inférieurs à n avec un seul modèle, contrairement à ce que nous faisons avec Dst et Kp dans le Chapitre 3, où pour chaque horizon temporel nous avons un modèle (et donc un apprentissage) dédié. Il est probable que notre approche ici ne soit pas optimale, mais elle présente l'avantage d'être environ n fois plus rapide et de simplifier le *pipeline* général, notamment en ce qui concerne l'émission d'une alerte (voir la section 4.2.3). Comme il s'agit ici de la première étude portant sur la prédiction de Ca , le fait d'accélérer le processus est un avantage pour développer le modèle par itérations successives, mais surtout cela nous permettra d'obtenir une première idée de l'horizon de prédiction *efficace*⁶ maximal, ce qui permettrait ensuite éventuellement d'entraîner un modèle différent par horizon de prédiction inférieur à l'horizon de prédiction efficace.

Tout comme au Chapitre 3, notre modèle principal est un réseau de neurones, se composant d'un réseau récurrent de type LSTM à une seule couche, combiné à une couche de type *feedforward* linéaire entièrement connectée. L'architecture du modèle est résumée en Figure 4.10.

Notre modèle a été entraîné en utilisant la méthode classique de rétropropagation du gradient (RUMELHART et al. 1986). La méthode d'optimisation utilisée est l'algorithme Adam (KINGMA et BA 2017). Nous avons utilisé un taux d'apprentissage $lr = 10^{-4}$ qui est divisé par deux une première fois après l'*epoch* 15 et une seconde fois après l'*epoch* 50. La fonction de perte est la MSE. Les paramètres

5. Ce sont des valeurs mises à disposition très rapidement après leur mesure, mais qui ne sont pas encore passées par un processus complet de vérification et de calibration.

6. C'est-à-dire l'horizon de prédiction maximal pour lequel nos prédictions peuvent être considérées comme utiles, ou encore opérationnelles.

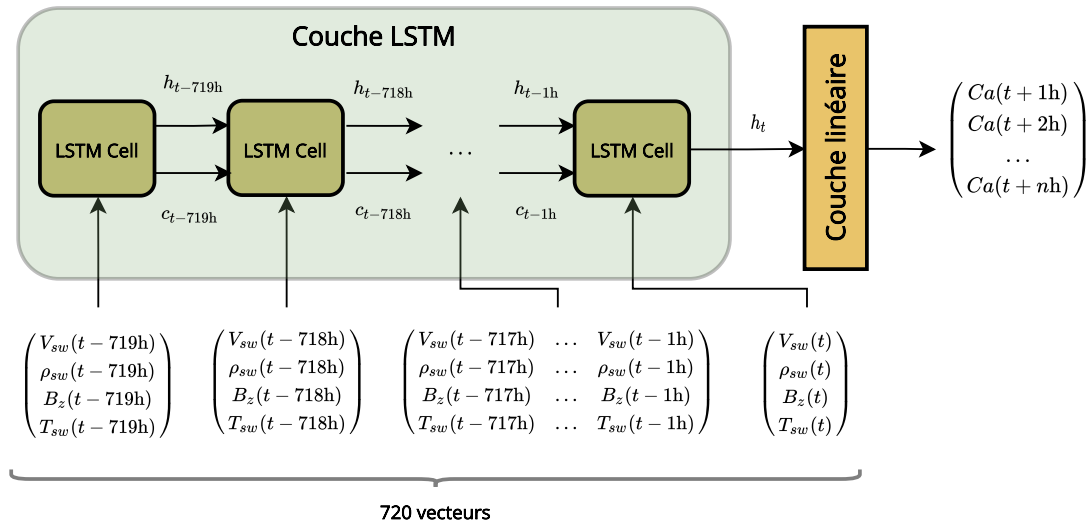


FIGURE 4.10 – Représentation simple de notre modèle basé sur une architecture LSTM pour prévoir les valeurs de l'indice Ca jusqu'à n heures à l'avance. Le fonctionnement interne de la cellule LSTM n'a volontairement pas été détaillé.

du modèle ont été déterminés empiriquement par itérations successives. Nous énumérons ci-dessous les principaux paramètres de notre modèle et certains choix d'implémentation, afin de faciliter la reproductibilité de nos résultats :

- l'état de la cellule LSTM a une dimension de 256 ;
- la couche LSTM est mono-directionnelle ;
- nous utilisons une régularisation L2 avec un poids de 10^{-5} ;
- la taille de chaque mini-batch est de 256 éléments ;
- l'entraînement est effectué avec 120 *epochs* et avec un arrêt précoce. L'arrêt précoce consiste à arrêter l'apprentissage du réseau dès que des signes explicites de surapprentissage sont observés.

Le modèle a été développé à l'aide de la bibliothèque PyTorch (v. 1.9) pour Python (PASZKE et al. 2019).

4.2.2.2 Émission d'une alerte

Notre modèle propose des prédictions de Ca sous la forme d'un problème de régression. Cependant, il est souvent plus utile pour un utilisateur final dans un contexte de prise de décision de bénéficier d'un système d'alerte. Un tel système d'alerte (binaire) peut être construit à partir de notre modèle de régression avec la méthode suivante : si nous prédisons que Ca dépassera une valeur seuil donnée au cours des t prochaines heures alors nous émettons une alerte (classe 1), si nous prédisons que nous resterons en dessous de ce seuil alors nous n'émettons aucune alerte (classe 0). La seule difficulté réside dans le choix d'un seuil approprié.

Dans notre exemple, nous choisissons un seuil basé autant que possible sur des critères opérationnels. Le seuil doit être significatif pour l'utilisateur final,

c'est-à-dire que le déclenchement d'une alerte doit correspondre à une situation pour laquelle on attend de l'opérateur une décision ou une action. L'indice Ca représentant l'état de remplissage des ceintures de radiation par des électrons de haute énergie, nous essayons de trouver un seuil de Ca associé à un risque non négligeable de dommages dus à la charge de surface.

La Figure 4.4 montre que l'indice Ca a un coefficient de corrélation assez élevé ($R \approx 0,83$) avec la dynamique des flux d'électrons intégrés d'énergie $E \geq 30$ keV pour $L \approx 6$. De plus, MATÉO-VÉLEZ et al. 2018 montre que le risque de dommages dus à la charge de surface pour un satellite en orbite géostationnaire (c'est-à-dire à $L \approx 6$) est bien corrélé avec le flux d'électrons d'énergie $10 \leq E \leq 50$ keV lorsque ce dernier est supérieur à $10^8 \text{ cm}^{-2}\text{s}^{-1}\text{sr}^{-1}$. Un jour au cours duquel le flux d'électrons d'énergie $10 \leq E \leq 50$ keV est toujours resté supérieur à cette valeur a une fluence quotidienne minimale de $8,64 \times 10^{12} \text{ cm}^{-2}\text{s}^{-1}$. À partir de cette valeur, nous définissons un seuil de fluence égal à $8 \times 10^{12} \text{ cm}^{-2}\text{s}^{-1}$.

Nous avons ensuite essayé de trouver un seuil de Ca qui maximise la corrélation entre les dépassements mensuels de la fluence électronique⁷ et les dépassements mensuels du seuil de Ca (en utilisant le maximum journalier de Ca). Pour les fluences journalières, nous avons utilisé les mesures fournies par l'instrument *Magnetospheric Plasma Analyzer* (MPA) embarqué dans la mission LANL 1991-80 placé en orbite GEO entre 1997 et 2006, dans la gamme d'énergie 35 – 46 keV (MCCOMAS et al. 1993). Nous avons constaté le seuil de Ca maximisant la corrélation entre le nombre de dépassements mensuels de fluence et les dépassements mensuels de Ca est $Ca_{\text{seuil}} = 38$ nT. Ceci est illustré sur la Figure 4.11.

On notera a posteriori que la valeur de $Ca_{\text{seuil}} = 38$ nT correspond approximativement au percentile 0,95 de la distribution des valeurs de Ca , ce qui semble statistiquement satisfaisant. En effet, c'est une valeur qui est donc suffisamment rare pour en faire un seuil d'alerte crédible et utile – un opérateur ne souhaiterait probablement pas recevoir une alerte lorsque la valeur de Ca ne dépasse que la médiane, par exemple. Mais c'est aussi une valeur qui n'est pas trop élevée, ce qui permet un meilleur apprentissage pour le réseau de neurones (en effet, plus le seuil est élevé, moins on a d'échantillons pour entraîner et évaluer le modèle – c'est aussi pour ça que nous n'avons pas utilisé un seuil correspondant à la définition d'un événement extrême). Insistons également sur le fait que cette valeur seuil utilisée pour définir nos classes binaires dans notre étude n'est qu'un exemple, et qu'en fonction de l'effet considéré (charge interne, charge de surface, événements singuliers, etc.), de l'orbite considérée, ou encore du satellite considéré (et donc de sa structure) il serait plus intéressant d'utiliser d'autres seuils, et possiblement d'augmenter le nombre de classes.

Insistons ici sur le fait que nous émettons une alerte pour un horizon de prédiction $t \leq 24$ h si Ca dépassera (selon nos prédictions) la valeur seuil au cours des t prochaines heures, et non pas dans exactement t heures. Il n'aura pas échappé au lecteur que ce système de classification binaire est différent de celui qui avait

7. C'est-à-dire le nombre de jours par mois où la fluence quotidienne a excédé le seuil de fluence de $8 \times 10^{12} \text{ cm}^{-2}\text{s}^{-1}$.

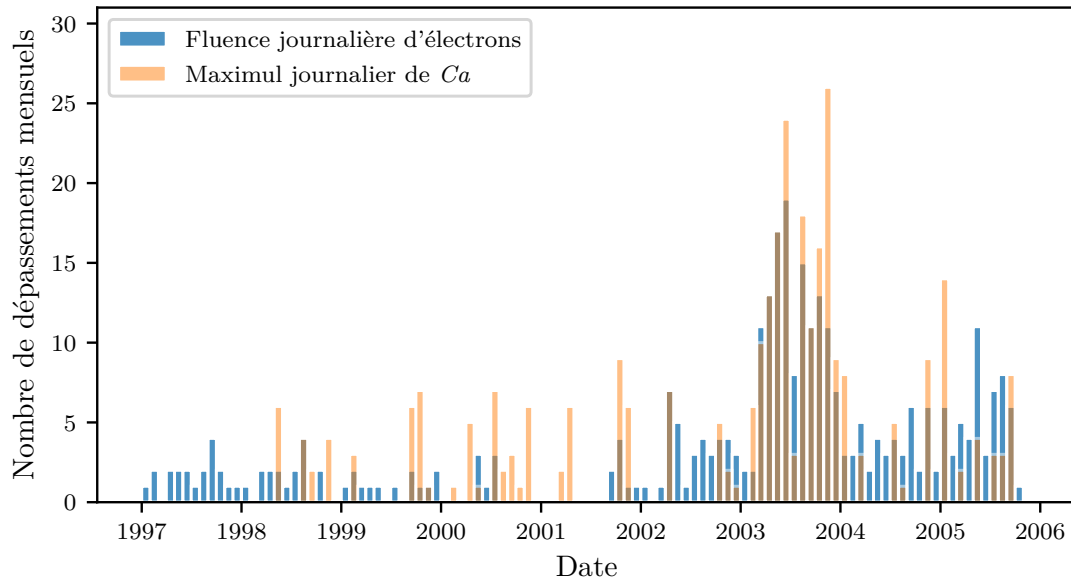


FIGURE 4.11 – Nombre de jours par mois pour lesquels l’instrument LANL 1991-80/MPA a mesuré une fluence électronique quotidienne pour la gamme d’énergie $10 \leq E \leq 50$ keV supérieure à $8 \times 10^{12} \text{ cm}^{-2} \text{ s}^{-1}$ ainsi que le nombre de jours par mois pour lesquels le maximum journalier de Ca quotidien était supérieur à 38 nT. Reproduit de BERNOUX et al. 2021.

été utilisé au Chapitre 3. Nous avons fait ce choix car dans le chapitre précédent, le fait de transformer nos prédictions numériques en classes binaires était un simple moyen d’évaluation plus poussé du modèle. Ici, le système d’alerte a été pensé pour être proche de ce qui pourrait être proposé dans un contexte opérationnel.

4.2.3 Évaluation du modèle

4.2.3.1 Les métriques d’évaluation

On peut distinguer deux classes de métriques d’évaluation : celles pour l’évaluation des prédictions vues comme un problème de régression, et celles évaluant le système d’alerte binaire.

En ce qui concerne l’évaluation des prédictions numériques, nous utiliserons les mêmes outils déjà utilisés dans le chapitre précédent, à savoir :

- la racine de l’erreur quadratique moyenne ou *Root Mean Squared Error* (RMSE) ;
- la racine de l’erreur quadratique moyenne normalisée ou *Normalized RMSE* (NRMSE) ;
- le coefficient de corrélation linéaire R ;
- le *Temporal Distortion Mix* (TDM).

Pour ce qui est du système d'alertes prédictif, nous l'évaluerons également avec des métriques déjà vues précédemment, à savoir :

- la *précision* ;
- le *rappel* ;
- le F_{score} ;
- le *FAR* ;
- le *Threat score* (TS) ;
- le *Heidke Skill Score* (HSS).

4.2.3.2 Catalogues d'événements

Pour évaluer notre modèle de manière encore plus détaillée, nous divisons l'ensemble de test en sous-parties correspondant aux périodes de perturbations induites d'une part par les *ICME* et d'autre part par des *SIR*, (incluant donc des *CIRs*). Pour ce faire, nous utilisons le catalogue d'*ICMEs* fourni par CHI et al. 2016 et la liste de *SIRs* fournie par CHI et al. 2018. Ces bases de données comprennent l'heure de début et l'heure de fin de plusieurs perturbations géomagnétiques induites par des *ICMEs* ou des *SIRs* entre 1995 et 2015 (2016 pour les *SIRs*). Selon ces bases de données, 212 *SIRs* et 204 *ICMEs* ont été observées dans l'environnement proche de la Terre entre 1997 et 2002 inclus. Dans notre étude, nous définissons une période de perturbation induite par une *ICME* (ou par une *SIR*) comme la période pendant laquelle une perturbation géomagnétique induite par une *ICME* (ou par une *SIR*) a une influence sur la dynamique de l'indice Ca . Le début de la période de perturbation est donné par le début de l'orage tel qu'indiqué dans la base de données. La fin de la période de perturbation est donnée en ajoutant $\tau = 4$ jours à la date de fin de l'orage telle qu'indiquée dans la base de données. Nous pouvons donc évaluer nos modèles en utilisant uniquement les périodes de perturbation induites par des *ICMEs* ou des *SIRs* et mieux comprendre la précision de nos prévisions. Le Tableau 4.2 résume le nombre d'échantillons de données dans chaque sous-ensemble et détaille le nombre d'échantillons appartenant aux périodes de perturbation.

Les listes d'événements induits par des *SIRs* et *ICMEs* que nous avons utilisées se terminent respectivement en 2015 et 2016. Par conséquent, le nombre d'échantillons dans chaque période de perturbation pour l'ensemble d'entraînement et l'ensemble complet est en réalité plus important que ceux indiqués dans ce tableau. Cela n'a aucune conséquence dans cette étude puisque nous ne sous-divisons l'ensemble de test en fonction de la nature de la perturbation uniquement à des fins d'évaluation du modèle.

4.2.3.3 Modèle de référence

Comme il s'agit de la première étude qui se focalise sur la prévision de l'indice Ca , il n'y a pas de modèle de référence immédiat auquel nous pouvons comparer

TABLEAU 4.2 – Nombre d'échantillons de données dans chaque ensemble de données, y compris le nombre d'échantillons appartenant à une période de perturbation. Adapté de BERNOUX et al. 2021.

Ensemble de données	Nombre total d'échantillons	Nombre d'échantillons situés dans une période de perturbation		
		induite par SIR uniquement	induite par ICME uniquement	induite par SIR et ICME
Entraînement	139 512	> 77 710	> 28 047	> 4219
Validation	16 801	10 794	2251	888
Évaluation	51 841	27 776	24 058	5407
Complet	208 154	> 116 280	> 54 356	> 10 514

notre modèle. Le modèle de référence habituellement utilisé dans une telle situation est le « modèle de persistance » (également connu sous le nom de « modèle naïf »), qui consiste simplement à supposer que la valeur prédite est la même que la dernière valeur observée. Cependant, ce modèle de référence ne peut pas être utilisé de manière pertinente ici car nous n'incluons pas les valeurs passées de l'indice Ca parmi les entrées de notre modèle, notre postulat étant précisément de ne pas utiliser les valeurs *quick-look* de aa (et donc de Ca). C'est pourquoi nous avons également entraîné un simple modèle de régression linéaire au sens des moindres carrés pour prévoir l'indice Ca à partir des mêmes paramètres de vent solaire qu'avec le réseau de neurones, à l'exception notable que le modèle linéaire de base utilise uniquement la dernière observation de chaque paramètre du vent solaire comme entrée (et non plusieurs valeurs passées comme avec le réseau de neurones).

4.2.4 Analyse des résultats

4.2.4.1 Régression

Les résultats de régression obtenus avec le modèle de référence et le modèle de réseaux de neurones (LSTM-NN) sont présentés dans le tableau 4.3. Tout d'abord, nous pouvons voir que les métriques classiques ($RMSE$, R) donnent de bien meilleures valeurs avec le modèle LSTM-NN qu'avec le modèle de référence. Pour un horizon temporel de 3 h, la $RMSE$ avec le LSTM-NN est environ 3,1 fois plus faible qu'avec le modèle de référence (2,62 au lieu de 8,13), et pour un horizon temporel de 24 h ce rapport est de 2,0 (4,17 au lieu de 8,16). C'est une indication supplémentaire du fait que les réseaux de neurones, en particulier les réseaux récurrents sont efficaces pour capturer le couplage vent solaire–magnétosphère. Les valeurs de $RMSE$ doivent être mises en perspective avec la distribution statistique de l'indice Ca , qui, sur la période de test, a une variance de 8,9 nT et un écart inter-quartile de 10,5 nT. Cette comparaison nous permet d'affirmer que les valeurs de $RMSE$ sont satisfaisantes, surtout pour un modèle qui n'inclut pas l'indice Ca parmi ses entrées. Nous constatons également que les valeurs de corrélation linéaire R sont assez élevées ($\geq 0,9$ pour tous les jeux de test jusqu'à un horizon temporel de 18 h, au lieu de $\leq 0,65$ avec le modèle de référence), ce qui est très satisfaisant.

TABLEAU 4.3 – Évaluation du modèle LSTM-NN et du modèle de référence linéaire dans le contexte du problème de régression. Les modèles ont été évalués avec le jeu de test complet ainsi qu'avec le jeu de test induit par les SIRs et le jeu de test induit par les ICMEs. Adapté de BERNOUX et al. 2021.

Horizon temporel (heures)	RMSE (nT)			R			TDM		
	Complet	SIR	ICME	Complet	SIR	ICME	Complet	SIR	ICME
3	2,62 (8,13)	2,42 (6,37)	3,43 (11,18)	0,96 (0,63)	0,95 (0,64)	0,95 (0,64)	0,13 (-0,42)	0,07 (-0,30)	0,08 (-0,55)
6	2,75 (8,08)	2,52 (6,37)	3,57 (11,10)	0,95 (0,64)	0,94 (0,64)	0,95 (0,65)	0,21 (-0,37)	0,14 (-0,25)	0,16 (-0,51)
9	2,95 (8,05)	2,67 (6,39)	3,79 (11,05)	0,95 (0,64)	0,93 (0,64)	0,94 (0,65)	0,39 (-0,32)	0,21 (-0,21)	0,25 (-0,49)
12	3,17 (8,05)	2,85 (6,43)	4,03 (11,01)	0,94 (0,64)	0,92 (0,64)	0,93 (0,65)	0,39 (-0,28)	0,28 (-0,16)	0,35 (-0,46)
15	3,42 (8,05)	3,04 (6,48)	4,29 (10,97)	0,93 (0,64)	0,91 (0,63)	0,92 (0,64)	0,46 (-0,23)	0,35 (-0,12)	0,43 (-0,41)
18	3,66 (8,08)	3,24 (6,54)	4,53 (10,94)	0,91 (0,63)	0,90 (0,63)	0,91 (0,64)	0,52 (-0,17)	0,40 (-0,06)	0,48 (-0,35)
21	3,92 (8,11)	3,43 (6,59)	4,78 (10,92)	0,90 (0,62)	0,89 (0,62)	0,90 (0,63)	0,56 (-0,13)	0,45 (-0,02)	0,52 (-0,30)
24	4,17 (8,16)	3,62 (6,65)	5,03 (10,90)	0,89 (0,62)	0,88 (0,61)	0,89 (0,63)	0,60 (-0,10)	0,49 (0,03)	0,58 (-0,29)

Note : Les résultats obtenus avec le modèle LSTM-NN sont indiqués en gras. Les résultats obtenus avec le modèle de référence sont indiqués entre parenthèses.

Le TDM donne des valeurs inférieures à 0,2 pour un horizon temporel de 3 h, voire 6 h pour les ensembles de tests basés sur des périodes de perturbation. Cela indique que jusqu'à environ 6 heures, nos prévisions sont bien alignées temporellement avec les valeurs cibles. Au-delà, la valeur de TDM augmente jusqu'à 0,60 pour un horizon de 24 h avec le jeu de test complet, ce qui indique qu'il y a un retard presque systématique entre les valeurs prévues et les valeurs cibles.

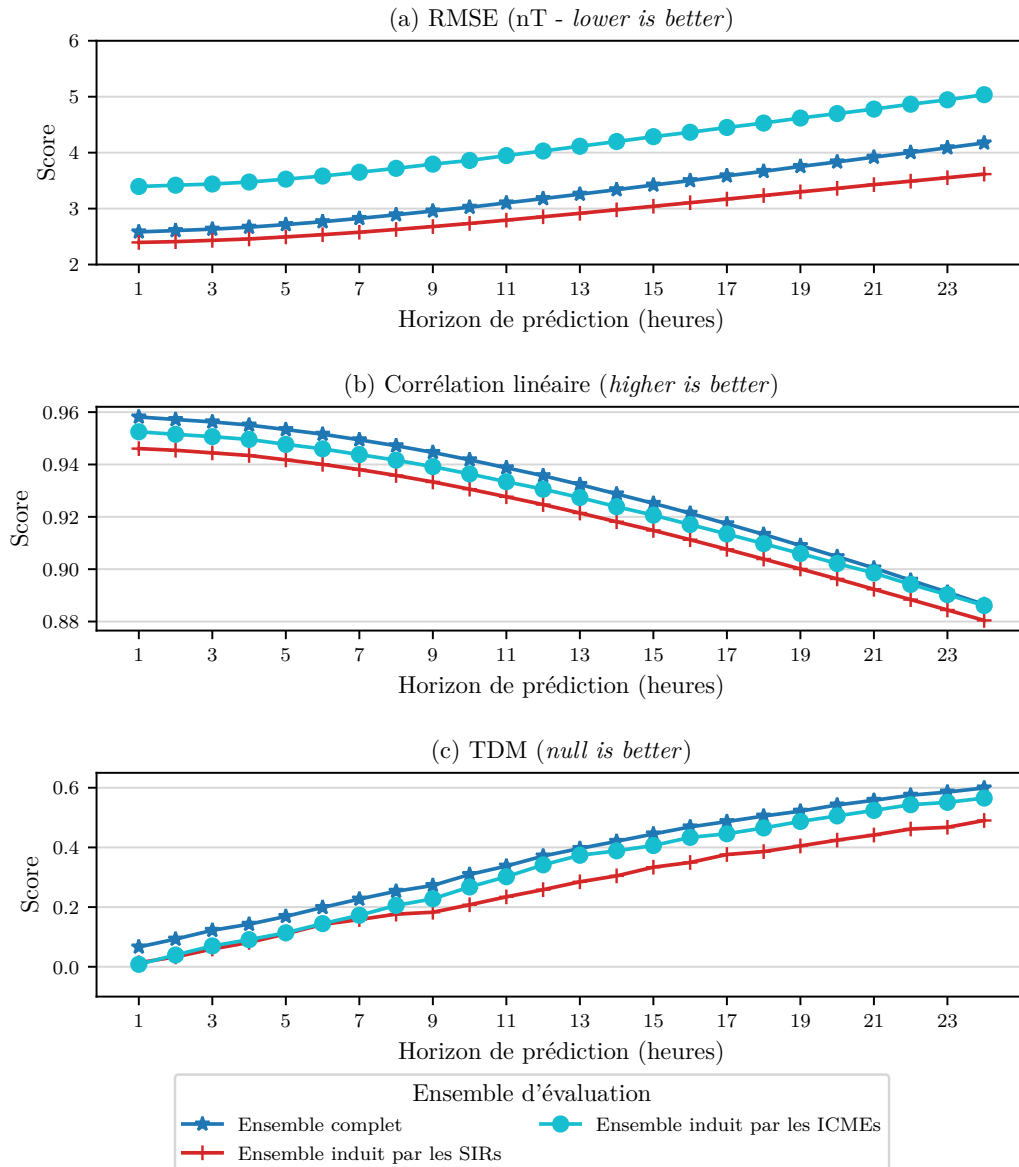


FIGURE 4.12 – Évaluation du modèle LSTM-NN avec trois mesures (RMSE, R et TDM) pour des valeurs d'horizon temporel allant de 1 heure à 24 heures. Trois ensembles d'évaluation (ensemble de test complet, ensemble induit par les SIRs et ensemble induit par les ICMEs) ont été utilisés. Reproduit de BERNOUX et al. 2021.

Sans surprise, les valeurs des métriques conventionnelles se dégradent toutes à mesure que l'horizon temporel augmente. Cette dégradation (augmentation pour

$RMSE$ et TDM , diminution pour R) semble être lente et régulière, comme le montre la Figure 4.12. Cependant, pour cette même raison, il est difficile de dire à partir de ces seules métriques à partir de quel horizon temporel le modèle n'est plus valide d'un point de vue opérationnel.

Nous observons également qu'en général, le modèle LSTM-NN est plus performant pendant les périodes de perturbations induites par les $SIRs$ que pendant les périodes de perturbations induites par les $ICMEs$. Pour un horizon temporel de 3 h, la $RMSE$ est 1,4 fois plus élevée pour la période induite par les $ICMEs$ que pour la période induite par les $SIRs$, ce qui est loin d'être négligeable. La Figure 4.13 montre plusieurs exemples de prévisions pour deux orages géomagnétiques : l'un induit par une $ICME$ et l'autre par une SIR . Cette figure montre les valeurs de prévision pour 4 horizons temporels différents (3, 6, 12 et 24 heures) réalisées avec le modèle LSTM-NN et le modèle linéaire de référence. Nous indiquons également les valeurs de TDM correspondant à chaque prévision (les valeurs pour le modèle de base sont entre parenthèses). Il ressort clairement de cette figure que le modèle LSTM-NN est plus performant que le modèle linéaire, comme l'indiquent déjà les mesures d'évaluation du problème de régression. Dans ces exemples, la dynamique de l'orage semble bien capturée, et les valeurs prédites sont effectivement proches des valeurs observées, comme l'indique la $RMSE$. De plus, il apparaît que les valeurs négatives de TDM mesurées avec le modèle linéaire sont dues au fait que le modèle a des difficultés à modéliser correctement la phase de décroissance de l'orage, qui diminue trop rapidement et apparaît donc « en avance » par rapport à la série réelle.

En outre, le fait que les séries temporelles prédites (avec le modèle LSTM-NN) et observées présentent un décalage temporel à mesure que l'horizon temporel augmente apparaît de façon évidente dans ces exemples. Il semblerait que ce décalage temporel soit plus prononcé au début de la période de perturbation que pendant la phase de décroissance de l'orage, qui dans l'exemple de l'orage causé par une SIR reste bien prédite même 24 heures en avance.

La différence de performance entre les périodes de perturbation induites par les $ICMEs$ et celles induites par les $SIRs$ pourrait donc s'expliquer, au moins en partie, par le fait que Ca augmente plus rapidement pendant les perturbations induites par les $ICMEs$. Comme l'indiquent les valeurs de TDM (et comme nous le verrons plus loin avec les métriques d'évaluation de classification), le modèle LSTM-NN semble être moins performant pendant la phase initiale d'une perturbation. Étant donné que lors des perturbations induites par des $SIRs$, l'augmentation initiale de Ca est plus lente que pendant les perturbations induites par des $ICMEs$, il est logique que la $RMSE$ pendant le début d'une période de perturbation soit plus faible dans le cas d'une SIR . Cela explique sans doute pourquoi la $RMSE$ globale est plus faible pour l'ensemble de test induit par les $SIRs$ que pour l'ensemble de test induit par les $ICMEs$.

La Figure 4.14 montre la $NRMSE$ et la corrélation linéaire R pour les prévisions avec un horizon temporel de 3 heures, après avoir réparti les valeurs

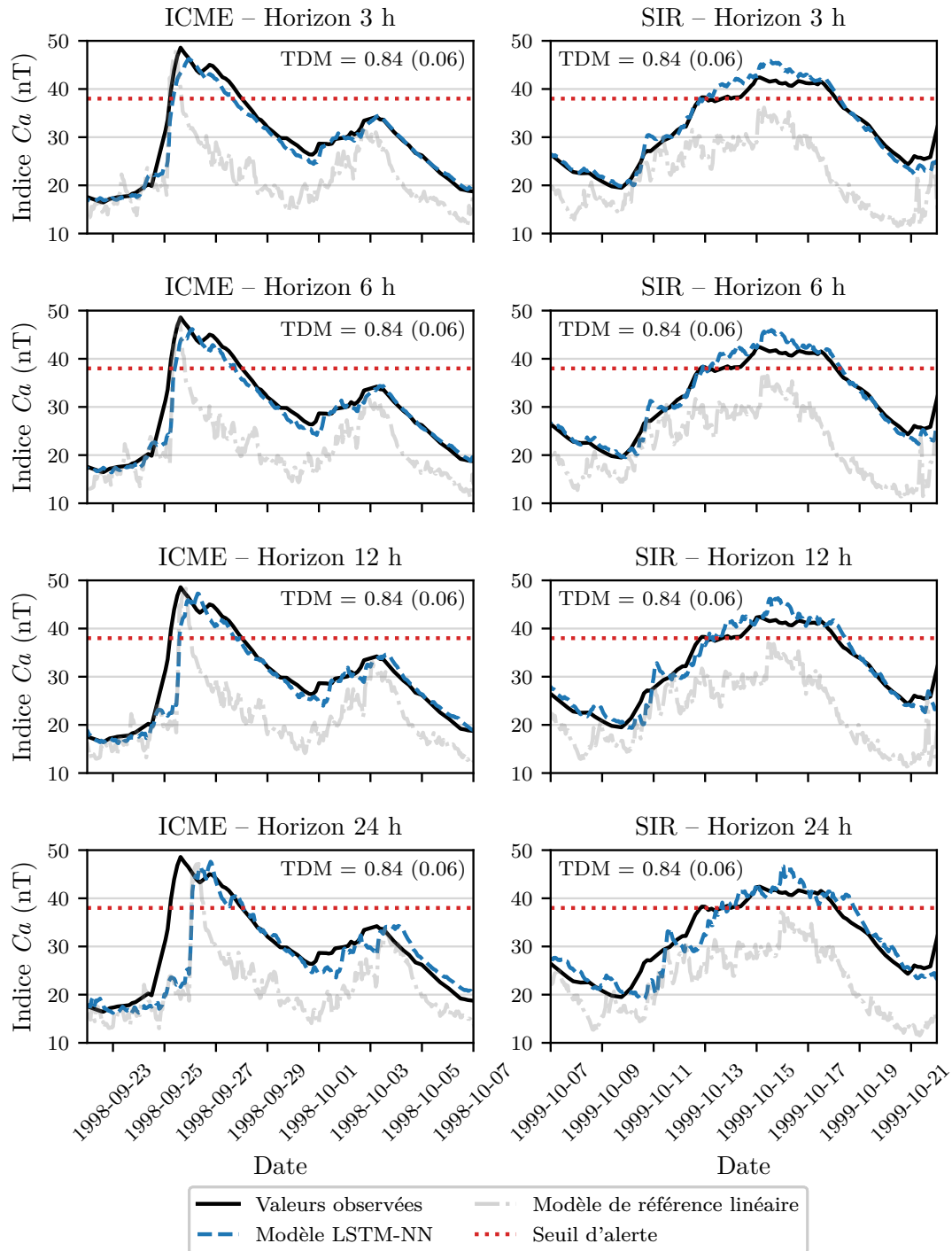


FIGURE 4.13 – Exemple de prévisions obtenues avec le modèle LSTM-NN et le modèle linéaire pendant deux orages géomagnétiques, le premier (à gauche) étant un orage piloté par une ICME et le second étant piloté par une SIR (à droite). Quatre horizons de prévision différents ont été utilisés (3, 6, 12 et 24 heures). La valeur seuil de C_a utilisée pour la classification binaire est indiquée en ligne pointillée bleue. Pour chaque prédiction, la valeur de TDM correspondante est donnée. Reproduit de BERNOUX et al. 2021.

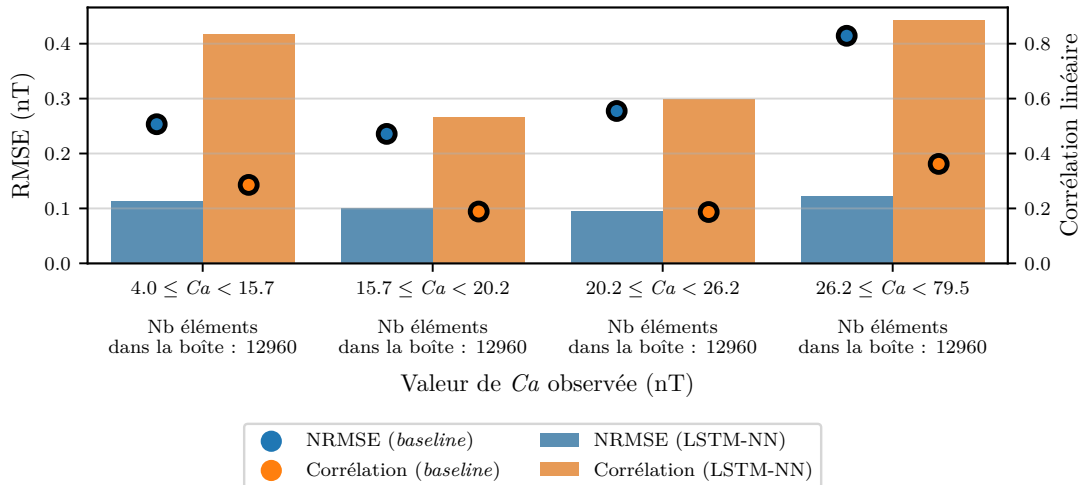


FIGURE 4.14 – NRMSE et corrélation linéaire des valeurs prédites avec 3 heures d’avance par rapport aux valeurs observées, en fonction des quartiles de Ca sur l’ensemble de test. Reproduit de BERNOUX et al. 2021.

de Ca observé en quartiles⁸. Nous utilisons ici la NRMSE car nous comparons les prévisions pour des valeurs de Ca à des échelles différentes, donc utiliser la RMSE pour cette comparaison serait biaisé. Il apparaît que le modèle LSTM-NN donne des valeurs de NRMSE stables lorsque Ca augmente, ce qui montre que le modèle a des performances similaires non seulement lorsque Ca est faible, mais aussi lorsqu’il atteint des valeurs plus élevées, contrairement au modèle de référence. La corrélation linéaire R pour les deux modèles diminue lorsque Ca se situe entre le premier et le troisième quartile. Ceci est très probablement dû au choix des intervalles correspondant aux quartiles de Ca . En effet, l’intervalle des valeurs de Ca dans ces intervalles est inférieur à 6 nT, c’est-à-dire de l’ordre de seulement deux fois la RMSE. Il n’est donc pas surprenant que la dispersion des valeurs prédites sur une si petite plage de valeurs observées rende la corrélation linéaire plus faible dans ces intervalles. Pour résumer, cette figure nous montre que le modèle donne des résultats stables et reste un bien meilleur modèle que le modèle linéaire pour toute la distribution des valeurs de Ca , avec une amélioration notable pour les valeurs élevées de Ca .

4.2.4.2 Classification

Les résultats pour le problème sous forme de système d’alerte binaire sont présentés dans le Tableau 4.4 et la Figure 4.15. Pour un horizon temporel de 3 heures, près de 85% des alertes émises étaient des vrais positifs, tandis que 84% des dépassements de seuil ont été détectés. Pour un horizon temporel de 24 heures, ces chiffres augmentent et diminuent respectivement à 87% et 73%. Le fait que la précision augmente avec l’horizon temporel est dû à la définition de nos classes binaires. En effet, nous essayons de prévoir si le seuil sera dépassé à un

8. Chaque boîte contient donc un quart des observations.

moment donné au cours des prochaines t heures (et non à un moment précis). Dans notre cas, lorsque l’horizon augmente, le modèle prévoit moins souvent des faux positifs et plus de faux négatifs. C’est pourquoi la **précision** augmente (légèrement), de manière assez contre-intuitive. Ceci met en évidence la nécessité de disposer de plusieurs méthodes d’évaluation afin d’obtenir une idée plus exhaustive de la performance réelle du modèle. Notons que le **rappel** diminue assez fortement et, de fait, le F_{score} , qui est la moyenne harmonique de la **précision** et du **rappel**, diminue, passant de 0,84 pour un horizon temporel de 3 h à 0,80 pour un horizon temporel de 24 h. Cela indique donc que le modèle est plus performant pour des horizons temporels plus courts.

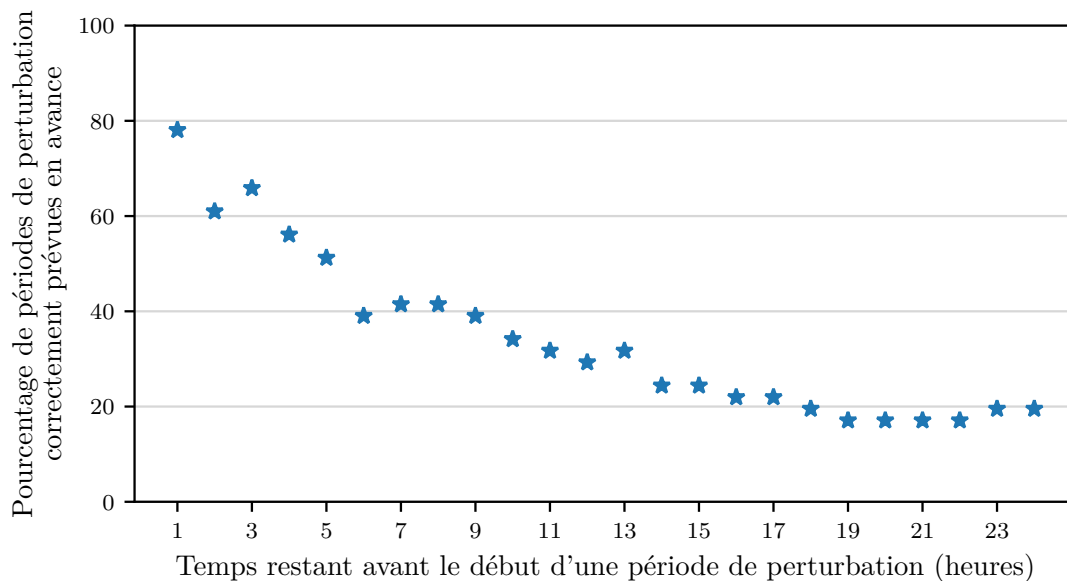


FIGURE 4.15 – Pourcentage de périodes de perturbations correctement prédites jusqu’à 24 h en avance, alors qu’on se situait à l’instant T dans une période de calme, en fonction du temps restant avant le dépassement (entre 1 heure et 24 heures). Reproduit de BERNOUX et al. 2021.

En l’absence de seuils empiriques, il est difficile de dire à partir de quel pourcentage de **précision** et de **rappel** le modèle devient satisfaisant. En termes absolus, prédire correctement plus de deux périodes de perturbation sur trois tout en ne produisant qu’environ 25 % de faux positifs peut sembler être un objectif satisfaisant. Cependant, en fonction des contraintes économiques liées, par exemple, à l’exploitation de systèmes spatiaux que nous cherchons à protéger, cela pourrait être largement insuffisant. Nous ne pouvons conclure définitivement sur la qualité absolue de notre modèle mais seulement sur sa qualité relative à des critères qui seraient définis par un opérateur et qui dépendent de chaque mission spatiale ou de l’objectif visé.

Notons cependant que les valeurs des scores (F_{score} , **TS**, **HSS**) sont assez élevées, notamment pour le **HSS**. En termes absolus, ces valeurs sont assez difficiles à interpréter et devraient surtout servir à l’avenir de point de comparaison pour d’éventuelles études portant sur la prévision de signaux physiques similaires.

TABLEAU 4.4 – Évaluation du modèle LSTM-NN et du modèle de référence linéaire dans le contexte du problème de classification. Adapté de BERNOUX et al. 2021.

Horizon temporel (heures)	précision	rappel	F_{score}	FAR	TS	HSS
3	0,85 (0,64)	0,84 (0,07)	0,84 (0,12)	0,010 (0,002)	0,73 (0,06)	0,83 (0,11)
6	0,85 (0,66)	0,83 (0,07)	0,84 (0,12)	0,010 (0,002)	0,73 (0,07)	0,83 (0,11)
9	0,86 (0,67)	0,82 (0,07)	0,84 (0,12)	0,009 (0,002)	0,72 (0,07)	0,83 (0,11)
12	0,87 (0,68)	0,80 (0,07)	0,83 (0,13)	0,009 (0,002)	0,71 (0,07)	0,82 (0,12)
15	0,87 (0,69)	0,79 (0,07)	0,83 (0,13)	0,009 (0,002)	0,70 (0,07)	0,81 (0,12)
18	0,87 (0,69)	0,77 (0,07)	0,82 (0,13)	0,009 (0,002)	0,69 (0,07)	0,80 (0,12)
21	0,87 (0,70)	0,75 (0,07)	0,81 (0,13)	0,009 (0,002)	0,68 (0,07)	0,79 (0,12)
24	0,87 (0,69)	0,73 (0,07)	0,80 (0,13)	0,009 (0,002)	0,66 (0,07)	0,78 (0,12)

Note : Les résultats obtenus avec le modèle LSTM-NN sont indiqués en gras. Les résultats obtenus avec le modèle de référence sont indiqués entre parenthèses.

Un résultat plus facile à interpréter est le pourcentage de périodes de perturbation prévues à l'avance, donné dans la Figure 4.15. Pour obtenir ces valeurs, nous avons calculé le pourcentage de fois où le modèle a été capable de répondre correctement à la question : « le seuil sera-t-il dépassé au cours des prochaines 24 heures ? » pour chaque pas de temps (entre 1 h et 24 h) précédant le début d'une période de perturbation. Nous nous intéressons donc ici uniquement à la capacité du modèle à prédire le début d'une période de perturbation (sans prendre en compte la poursuite d'une telle période). Il apparaît que le modèle est capable de répondre correctement à cette question un peu moins de 80 % du temps 1 heure avant le dépassement du seuil. Ce pourcentage reste supérieur à 50 % jusqu'à 5 heures avant le dépassement du seuil. Entre 6 et 9 heures avant un dépassement de seuil, environ 40 % des périodes de perturbation ont été correctement prévues. Moins de 25 % des dépassements de seuil ont été prévus au moins 15 heures en avance. Cela montre que même si 73 % du total des dépassements ont été détectés entre 1 heure et 24 heures avant qu'ils ne se produisent, moins d'une période de perturbation sur deux a été détectée 6 heures avant qu'elle ne se produise et moins d'une sur quatre a été détectée 24 heures avant qu'elle ne se produise. Il s'agit là d'une mesure beaucoup plus significative du caractère opérationnel de notre modèle, qui confirme le point que nous avons soulevé précédemment concernant la difficulté à prévoir le début d'un orage géomagnétique.

4.2.5 Bilan sur la prédiction de C_a

Les résultats mentionnés ci-dessus ne sont en réalité pas très surprenants puisque nos modèles reposent sur des paramètres du vent solaire mesurés à proximité de la Terre. Par conséquent, le recul temporel pour prévoir la dynamique des ceintures de radiation est faible. Cela se reflète dans les mesures de TDM qui indiquent que les prévisions sont globalement très bien alignées temporellement avec les observations pour des valeurs d'horizon de prévision inférieures à 6 heures, ce qui correspond approximativement au temps de réaction de la géomagnétosphère interagissant avec une perturbation arrivant près de la Terre. On peut donc en déduire d'une part que notre modèle semble être en accord avec la physique du problème, mais aussi que cette même physique nous empêche, si nous ne changeons pas la nature de nos entrées, d'avoir de bonnes performances opérationnelles pour des horizons de prévision plus grands.

Par ailleurs, il semble délicat de trouver un horizon de prédiction limite pour notre modèle, au-delà duquel il est possible d'affirmer définitivement que le modèle n'est plus opérationnel. Comme mentionné plus haut, cela dépend des besoins d'un opérateur. En l'absence de valeurs seuils qui pourraient servir de repères pour des métriques telles que la précision ou le rappel, nous ne pouvons que deviner. Une façon de le faire serait de considérer le pourcentage d'orages prédits en avance. Si nous prenons un seuil de 50 %, alors la limite de l'horizon de prédiction opérationnel de notre modèle est de 5 heures. Avec un seuil de 75 %, notre horizon de prédiction opérationnel n'est que d'une heure. Une autre méthode consiste à prendre en compte le TDM. Avec un seuil arbitraire de 0,2, la limite de l'horizon de prédiction de notre modèle est de 6 heures, alors qu'avec un seuil de 0,1, la limite de l'horizon

n'est que de 1 heure pour l'ensemble des tests, mais de 4 heures pendant les périodes de perturbation induites par les **SIR** et de 3 heures pendant les périodes de perturbation induites par les **ICME**.

Rappelons qu'une limite de 6 heures a été trouvée dans d'autres articles traitant de la prévision de l'indice *Dst* (LAZZÚS et al. 2017; GRUET et al. 2018). Certaines études qui visent à prévoir l'indice *Kp*, comme AYALA SOLARES et al. 2016; SEXTON et al. 2019, prétendent pouvoir prévoir l'indice *Kp* jusqu'à 24 heures à l'avance. Il serait intéressant d'évaluer les performances opérationnelles des modèles présentés dans ces articles avec des métriques comme le TDM et en évaluant uniquement la capacité à prévoir le début d'un orage, afin d'avoir une compréhension plus complète de leur efficacité réelle dans des contextes opérationnels. Insistons cependant sur le fait que la difficulté pour les longs horizons de prédiction réside dans le début de l'orage et non dans sa continuité, car l'accumulation d'énergie permet de trouver un lien entre les paramètres du vent solaire et les indices géomagnétiques même après 6 heures. C'est notamment le cas avec un indice intégré dans le temps tel que *Ca*, qui permet d'obtenir de bonnes performances globales de prévision jusqu'à 24 heures à l'avance.

Il pourrait être tentant de comparer nos résultats à ceux présentés par exemple dans FORSYTH et al. 2020 où les auteurs présentent un modèle pour prévoir les flux d'électrons d'énergie $E \geq 2$ MeV mesurés par le satellite NOAA GOES-15, à partir des paramètres de vent solaire et évaluent également leur modèle avec des mesures de classification. Par exemple, l'un de leurs modèles pour un horizon temporel de 6 heures donne une précision de 0,75 alors que pour le même horizon temporel, le nôtre donne un taux de réussite plus élevé de 0,87. Cependant, cette comparaison n'est pas valable car nous ne nous concentrons pas sur la même gamme d'énergie et notre modèle n'utilise pas les mêmes seuils et critères de classification. En effet, nous répondons ici à la question : « le seuil sera-t-il dépassé quelque part dans les prochaines t heures ? » Dans le cas de FORSYTH et al. 2020, la question est : « le seuil sera-t-il dépassé dans exactement t heures ? » Nous avons choisi d'aborder le problème de cette manière car nous pensons qu'un système d'alerte défini de cette façon est plus utile, surtout si l'on pose cette question pour plusieurs horizons temporels t . Nous tenions à insister ici sur le fait que, comme le souligne CAMPOREALE 2019, la comparaison des performances d'un modèle par rapport à un autre n'est pas simple et qu'il convient d'être prudent en le faisant.

4.3 Bilan du chapitre

Partant du constat que la prédiction d'indices comme *Dst* ou *Kp* ne permettait qu'une prise en compte limitée du caractère géoeffectif de certains événements issus des interactions Soleil–Terre, et plus particulièrement du couplage vent solaire–magnétosphère, nous avons proposé l'utilisation de l'indice *Ca*. L'indice *Ca* est obtenu à partir de mesures au sol, possède un grand historique de données et est facile à calculer. Sa définition en fait un indice représentatif de l'état de remplissage général des ceintures de radiations terrestres en électrons de haute énergie. Il

permet notamment une meilleure prise en compte de phénomènes causant une accumulation d'énergie dans les ceintures de radiations, comme les **SIRs** ou bien les **ICMEs** multiples.

Nous avons ensuite cherché à prédire cet indice à l'aide de réseaux de neurones (LSTM-NN) ayant une architecture très similaire à ceux ayant déjà démontré de bonnes performances pour prédire des indices comme Kp et Dst . Les choix d'implémentation effectués dans cette étude ont été faits en gardant à l'esprit un contexte opérationnel. Ces choix concernent par exemple les entrées utilisées dans le modèle, ainsi que la méthodologie d'évaluation du modèle. À cette fin, nous avons souligné l'importance de choisir des ensembles d'entraînement et de test statistiquement et physiquement représentatifs, notamment du point de vue des événements extrêmes.

Nous avons également souligné la nécessité d'utiliser des mesures adéquates pour évaluer le modèle, puisque les métriques classiques telles que la **RMSE** ou la corrélation linéaire ne sont pas en mesure de donner une vision exhaustive de la performance du modèle, en particulier pendant les périodes de perturbation. C'est pourquoi nous utilisons le **TDM** pour mesurer la tendance d'une prévision à être en retard ou en avance par rapport aux observations réelles. Nous montrons que les valeurs de **TDM** se dégradent avec l'augmentation de l'horizon de prédiction, rendant nos prévisions de moins en moins compétentes.

Afin de rendre le modèle plus utile opérationnellement nous proposons la définition d'un système d'alertes, transformant nos prévisions numériques en classes binaires. Le choix du seuil utilisé pour définir les classes binaires a été fait en tenant compte du risque pour les satellites en orbite **GEO** de subir des dommages dus à la charge de surface. L'évaluation du problème de classification binaire montre que, même si les mesures de régression semblaient excellentes, le réseau de neurones ne propose pas des performances exceptionnelles lorsqu'il s'agit de prévoir le début d'une période de perturbation. Ceci est très certainement dû à la proximité spatiale (et donc temporelle) entre les paramètres du vent solaire utilisés pour alimenter le modèle et la magnétosphère terrestre. Pour l'instant, même si les métriques sont bonnes et bien meilleures que celles obtenues avec le modèle linéaire de référence, il serait difficile d'affirmer que le modèle LSTM-NN est parfaitement adapté pour une utilisation en situation opérationnelle. Il représente cependant un premier et grand pas vers cet objectif.

Nous étions parvenus à une conclusion similaire lorsque nous cherchions à prédire l'indice Dst dans le chapitre précédent. Il apparaît donc clairement que pour améliorer les résultats des prévisions pour des horizons temporels de 24 heures et au-delà, il pourrait être intéressant d'augmenter la distance entre les entrées de nos modèles et leurs sorties. Une façon d'atteindre cet objectif serait d'utiliser les données provenant de l'imagerie solaire. Cela sera l'objet des chapitres suivants.

Chapitre 5

Prédire la géoefficacité d'événements solaires à partir d'images du Soleil : une nouvelle méthode

Ideas are so beautiful and they're so abstract. And they do exist someplace. I don't know if there's a name for it. And I think they exist, like fish. And I believe that if you sit quietly, like you're fishing, you will catch ideas. The real, you know, beautiful, big ones swim kinda deep down there so you have to be very quiet, and you know, wait for them to come along.

(David Lynch, Catching the Big Fish)

Sommaire

Préambule	170
5.1 Observations solaires et apprentissage automatique : une revue . . .	171
5.1.1 Transformation des données	171
5.1.2 Détection d'objets	173
5.1.3 Prédiction de signaux physiques	174
5.2 Philosophie de notre nouvelle approche	176
5.3 Présentation de l'architecture de SERENADE	177
5.3.1 L'architecture globale	178
5.3.2 Les données	180
5.3.2.1 L'imagerie solaire	180
5.3.2.2 Le pré-traitement des données d'imagerie solaire . .	181
5.3.2.3 L'indice géomagnétique K_p	183
5.3.3 Le réseau de neurones	184
5.3.3.1 Le Feature Extractor	184
5.3.3.2 Le Temporal Encoder	187
5.3.3.3 Le Decoder	188
5.3.3.4 Récapitulatif de l'architecture	190
5.3.4 L'apprentissage du modèle	190
5.3.4.1 Séparation des données pour entraînement, valida-	
tion et test	190
5.3.4.2 Implémentation pratique	196
5.4 Méthodologie de l'évaluation	199
5.4.1 Les modèles de référence	199
5.4.1.1 Climatologie	200
5.4.1.2 Persistance	200
5.4.1.3 Récurrence à 27 jours	201
5.4.2 Le banc d'évaluation	202
5.4.2.1 Banc d'évaluation pour le problème de régression	202
5.4.2.2 Banc d'évaluation pour le problème de classification	203
5.5 Bilan du chapitre	203

Préambule

Comme nous l'avons vu dans les chapitres précédents, l'utilisation de paramètres du vent solaire mesurés (trop) proches de la Terre pour la prédiction de la dynamique des ceintures de radiations n'est plus suffisante dès que nous souhaitons prédire ces perturbations avec plusieurs heures, voire plusieurs jours d'avance. Afin de résoudre ce problème, nous considérons dans ce chapitre l'utilisation d'imagerie solaire comme pilote de nos modèles. En effet, en utilisant directement de l'imagerie solaire, c'est-à-dire en observant directement certaines des structures solaires à l'origine de phénomènes – éruptifs ou non – potentiellement géoeffectifs, nous pouvons espérer gagner un horizon de prédiction supplémentaire équivalent au temps de propagation de ces phénomènes entre le Soleil et la Terre. Dans ce chapitre nous proposons donc d'étudier cette hypothèse en construisant un modèle s'appuyant sur des réseaux de neurones pour la prédiction d'indices géomagnétiques à partir

d'imagerie solaire.

Dans un premier temps nous passerons en revue et analyserons différentes études utilisant des données provenant d'instruments observant le Soleil (et notamment des missions [SOHO](#) et [SDO](#)) conjointement à des modèles d'apprentissage automatique. Après avoir justifié la pertinence de notre approche, nous détaillerons l'architecture de notre modèle, avant de présenter les méthodes d'évaluation utilisées. Les premiers résultats obtenus avec cette approche seront donnés et étudiés dans le Chapitre 6. Le contenu de ces deux chapitres fait l'objet d'un article publié dans *Journal of Geophysical Research: Space Physics* ([BERNOUX et al. 2022a](#), voir Annexe C.3).

5.1 Observations solaires et apprentissage automatique : une revue

L'utilisation de données provenant d'instruments observant le Soleil – que ce soit des images [EUV](#), des magnétogrammes ou des données qui en sont dérivées – au sein de modèles s'appuyant sur l'apprentissage machine n'est pas inédite, mais est tout de même assez récente. L'essor de l'apprentissage automatique dans ce domaine a non seulement été permis par l'augmentation conjointe de la puissance de calcul et des capacités de stockage des machines contemporaines, comme cela a été et est toujours le cas pour de nombreux autres domaines d'application, mais aussi par la mise à disposition concomitante ces dernières années d'imagerie solaire de plus en plus fiable, avec des résolutions spatio-temporelles plus élevées, comme celles fournies par la mission [SDO](#) (voir le Chapitre 2).

Nous proposons ici une revue (non-exhaustive, mais que nous avons essayé de rendre assez représentative) de différentes études ayant utilisé des données d'observation solaire conjointement à des méthodes d'apprentissage automatique afin d'avoir une idée plus précise de ce qui a déjà été fait avec ces données (et ce qui n'a pas encore été tenté), comment cela a été fait et ce qu'il est probablement possible de faire. À cet effet, nous choisissons de séparer ces études en trois grandes catégories correspondant à la nature de l'objectif visé par l'étude : la transformation des données, la détection d'objets et la prédiction d'une grandeur physique.

5.1.1 Transformation des données

Par « transformation des données » nous entendons les études utilisant des méthodes d'apprentissage automatique pour convertir, améliorer et modifier une caractéristique de ces données, qui sont le plus souvent des images.

L'objectif de ces études est principalement la création d'instruments virtuels. En effet, si on peut transformer des magnétogrammes en images [EUV](#) avec un réseau de neurones, on peut imaginer utiliser ce réseau pour créer l'équivalent d'images [EUV](#) spatiales à partir de magnétogrammes obtenus depuis des instru-

ments terrestres, comme ceux du réseau GONG par exemple. De telles images seraient sans doute de moins bonne qualité, mais pourraient par exemple servir à compléter les autres bases de données lors de périodes présentant des « trous ». Par exemple, PARK et al. 2019 utilise des CNNs pour transformer des magnétogrammes fournis par SDO/HMI en images EUV fournies par SDO/AIA. D'une manière similaire, cette fois en utilisant une architecture de CNN appelée U-Net, SALVATELLI et al. 2019 propose d'obtenir une image de synthèse imitant une image SDO/AIA à 211 Å à partir de trois images à 94, 171 et 193 Å. SZENICER et al. 2019 propose même de convertir des images EUV en courbes d'irradiance solaire de synthèse (on est donc sur une transformation qui change la structure même de la donnée, contrairement aux transformations magnétogramme vers EUV par exemple). Enfin LI et al. 2020 et ZHANG et al. 2020 proposent même des conversions entre des données issues d'instruments spatiaux – en l'occurrence AIA – en images radio (à 6 GHz pour les premiers, 17 GHz pour les seconds) obtenues à partir de radiohéliographes terrestres.

En rendant ces méthodes fiables et robustes (ce qu'elles ne sont pas encore), on pourrait même envisager d'alléger les instruments envoyés dans l'espace (et donc diminuer le coût d'une mission) en utilisant des télescopes spatiaux ayant moins de canaux, mais à partir desquels on serait capable de reconstruire d'autres canaux. On peut également envisager la création de bases de données continues plus étendues dans le temps en créant des instruments virtuels à partir d'instruments embarqués sur des missions se succédant les unes aux autres. Du point de vue de la physique du Soleil, il est intéressant de noter que les réseaux de neurones semblent capables de trouver et exploiter la redondance existant entre des données de nature physique différente, ce qui n'est pas trivial. De là à dire que ces modèles parviennent à « comprendre » une partie de la physique solaire sous-jacente il y a un grand pas que nous ne franchirons pas, mais c'est une piste qui serait intéressante à explorer.

Outre les instruments virtuels, l'autre objectif qui ressort de ce corpus est celui d'améliorer la qualité des données fournies par les instruments. Ainsi, JUNGBLUTH et al. 2019 et GITIAUX et al. 2019 proposent des méthodes reposant sur des réseaux de neurones convolutionnels pour augmenter la résolution spatiale de magnétogrammes. PARK et al. 2020 propose également de rendre les magnétogrammes plus fins en les débruitant à l'aide de réseaux de convolution génératifs. Dans un autre registre, SANTOS et al. 2021 propose un *pipeline* construit autour de CNNs pour automatiquement recalibrer les images EUV d'AIA, par exemple pour compenser la dégradation naturelle de l'instrument au cours de son cycle de vie.

Au-delà de la simple amélioration des données, ces méthodes pourraient également servir à homogénéiser des jeux de données différents, par exemple des images issues de différents imageurs, dont la résolution n'a cessé d'augmenter à chaque mission (ou presque). Étant donné que lorsqu'on utilise une approche pilotée par les données il est très important d'avoir des données « nettoyées » et de qualité homogène, et que le processus de calibration de données peut être très long et fastidieux, le fait de pouvoir automatiser ces étapes serait un véritable atout.

5.1.2 Détection d'objets

Ici nous présentons un corpus d'études utilisant des techniques d'apprentissage automatique pour la détection ou la sélection d'objets – c'est-à-dire des structures ou caractéristiques particulières dignes d'intérêt – sur des images ou des magnétogrammes.

Ici les objectifs des différentes études sont plus diversifiés. Premièrement, certaines s'intéressent à la détection des *CMEs*. C'est par exemple le cas de AL-OMARI et al. 2010, qui propose d'étudier les associations entre *CMEs* et filaments, en analysant de façon croisée le catalogue de filaments US NGDC et le catalogue de *CMEs* observées par SOHO/LASCO, à l'aide d'une SVM. Autre exemple, BYRNE 2015 crée des catalogues automatiques de *CMEs* et utilise un algorithme de *k-means* pour distinguer des *CMEs* multiples spatio-temporellement proches.

Ensuite, d'autres études s'intéressent à la détection de *flares*, avec des méthodes parfois très différentes. ZHANG et al. 2011 utilise divers algorithmes (classifieur bayésien, arbres décisionnels, etc.) pour essayer de détecter des précurseurs d'éruptions solaires à partir de magnétogrammes SOHO/MDI et de catalogues de régions actives. AHMED et al. 2013 propose l'utilisation de réseaux de neurones pour sélectionner les meilleures caractéristiques de régions actives pour la prédiction de *flares* dans le catalogue de régions actives Solar Monitor Active Region Tracker Magnetic Features (SMART MF) – dérivé de SOHO/MDI. Enfin, CAMPI et al. 2019 résume certains des résultats du projet FLARECAST (GEORGOULIS et al. 2021), notamment en ce qui concerne la sélection automatique (à l'aide d'une régression hybride-lasso) des meilleures caractéristiques de régions actives présentes dans le catalogue Space Weather HMI Archive Patch (HMI) pour la prédiction de *flares*.

Enfin, certaines études, comme celle de DELOUILLE et al. 2018, se penchent sur l'utilisation de SVMs pour distinguer les trous coronaux des filaments sur les magnétogrammes, tandis qu'ILLARIONOV et TLATOV 2018 utilisent des réseaux de neurones convolutionnels sur des images SDO/AIA à 193 Å pour détecter et détourner les trous coronaux automatiquement.

On constate que parmi ces études, une majorité utilise des données qui sont dérivées de l'imagerie solaire, plutôt que des images du Soleil directement. De plus, la majorité utilise également des méthodes d'apprentissage automatique autres que des réseaux de neurones. Ces deux aspects sont sans doute liés : il est en effet très difficile pour la plupart des algorithmes d'apprentissage d'extraire de l'information pertinente des images à cause de leur grande dimensionnalité, qu'il faut alors compenser par un très grand nombre d'échantillons : c'est le fléau de la dimension (BELLMAN 2003). Lorsqu'on ne dispose pas d'un nombre d'échantillons suffisant il faut alors réduire la dimension du problème. Utiliser des catalogues pré-existants de caractéristiques des structures qui nous intéressent (régions actives, *CMEs*, trous coronaux, etc) est une bonne façon de réduire la dimension tout en conservant du sens physique. Toutefois il est indéniable qu'on perd alors une grande partie de l'information contenue dans l'image d'origine.

On observe qu'ici aussi de nombreuses études (comme AL-OMARI et al. 2010 ;

ZHANG et al. 2011 ; DELOUILLE et al. 2018) essaient d'exploiter les méthodes d'apprentissage automatique pour parvenir à détecter ou bien distinguer des phénomènes physiques qui le sont difficilement à l'aide de méthodes empiriques ou physiques. On remarque également que deux autres grandes catégories d'objectifs peuvent être définies dans ce corpus : automatiser le processus de détection de structures pour la création de catalogues et l'amélioration des modèles de prédiction (ces deux objectifs ne s'excluant pas) à des fins de météorologie de l'espace, qui sont l'objet de la sous-section suivante. Notons bien que l'utilisation de *machine learning* ne permettra sans doute pas une amélioration notable de la qualité même des catalogues, mais leur automatisation peut rendre immédiate leur exploitation en conditions opérationnelles, ce qui représenterait une forte valeur ajoutée.

5.1.3 Prédiction de signaux physiques

Ici nous nous intéressons aux études portant directement sur la prédiction de signaux physiques à partir d'images et de magnétogrammes du Soleil et s'appuyant au moins partiellement sur des modèles d'apprentissage automatique ou profond. Plus nombreuses que celles portant sur la transformation et la détection, elles témoignent du dynamisme des applications du *machine learning* en météorologie de l'espace. Le corpus présenté ici n'est encore une fois pas exhaustif, même s'il se veut assez représentatif.

Premièrement, de très nombreuses études essaient de prédire l'apparition d'embrillancements et d'éruptions solaires avant qu'elles aient été observées. C'est par exemple le cas de COLAK et QAHWAJI 2009, qui utilise des réseaux de neurones appliqués aux magnétogrammes de SOHO/MDI pour prédire la probabilité d'apparition d'éruption et leur classe jusqu'à 72 heures en avance. C'est aussi le cas de HUANG et al. 2013, qui propose un modèle amélioré prenant également en compte la longitude des régions actives. LIU et al. 2017 étudie l'utilisation de l'algorithme de *Random Forest* (RF) pour résoudre ce même problème. FLORIOS et al. 2018 explore l'utilisation de MLPs, SVMs et de RFs pour la prédiction de *flares* à partir de patches issus de SHARP. D'autres articles, plus récents, utilisent des CNNs pour prédire les éruptions, comme PARK et al. 2018 ou ZHENG et al. 2019. Des approches utilisant des LSTM (LIU et al. 2019a), un algorithme de κ -NN (NISHIZUKA et al. 2017), ou encore des méthodes d'ensemble (LIM et al. 2019) ont également été étudiées. Enfin, TLATOV et al. 2020 étudie l'utilisation de magnétogrammes obtenus à partir d'observatoires terrestres plutôt que spatiaux pour la prédiction de *flares*. Ce qui ressort de ce corpus est que la complexification des méthodes utilisées améliore les résultats au fur et à mesure, et que les *flares* X sont de mieux en mieux prédits. Cependant, les modèles actuels ont encore du mal à distinguer les *flares* de classes intermédiaires (C et M par exemple). Du point de vue de la physique, HUANG et al. 2018 montre que les réseaux de neurones semblent construire leurs prédictions en s'appuyant sur les régions des magnétogrammes présentant une inversion de la polarité magnétique.

D'autres études s'intéressent plutôt aux CMEs. Ainsi, CHOI et al. 2012 propose d'étudier le caractère géoeffectif d'halo-CMEs (le caractère géoeffectif

est déduit de l'indice Dst) en utilisant des SVMs appliqués à des caractéristiques extraites de CMEs observées par SOHO/LASCO et de mesures de flux X par des satellites NOAA-GOES. LIU et al. 2018 propose des prédictions du temps d'arrivée des CMEs sur Terre en utilisant des réseaux de neurones récurrents LSTM. PESNELL 2020 souligne que les prédictions du temps d'arrivée des CMEs s'améliore constamment, notamment grâce à l'accumulation de données provenant des missions STEREO, SOHO et SDO. GOLUB et al. 2020 suggère que l'utilisation d'imagerie EUV pourrait également contribuer à améliorer ces prédictions.

Finalement, très peu d'études publiées se penchent sur la question de la prédiction de paramètres du vent solaire à partir d'imagerie solaire, et de la géoeffectivité du couplage vent solaire–magnétosphère, en utilisant des méthodes d'apprentissage automatique. À notre connaissance, à l'heure actuelle seul UPENDRAN et al. 2020 propose une méthode utilisant des réseaux de neurones complexes pour prédire la vitesse du vent solaire quelques jours en avance à partir d'imagerie solaire aux longueurs d'onde 193 et 211 Å. En particuliers, leurs auteurs essaient de prédire les vents solaires rapides, ce qui en fait une des seules études de ce corpus à se pencher sur cette question. En plus de CHOI et al. 2012 déjà précédemment cité, TASISTRO-HART et al. 2021 s'intéresse également à la géoeffectivité en proposant une prédiction probabiliste 6 heures en avance de l'indice Dst en utilisant des réseaux de neurones alimentés non seulement par des paramètres du vent solaire, mais également par des données issues du catalogue de CMEs de SOHO/LASCO.

Ainsi il apparaît qu'au cours de la dernière décennie, les données issues de l'imagerie solaire ont été largement utilisées pour effectuer différents types de prédictions en s'appuyant sur l'apprentissage automatique. Bien que la majeure partie de la littérature soit consacrée à la prédiction de *flares*, les CMEs sont également largement étudiées. La diversité des horizons de prédictions utilisés (de quelques heures à plusieurs jours), la multiplicité des architectures, et la variété des données utilisées rendent souvent les comparaisons entre deux modèles très difficiles (CINTO et al. 2020). On constate également une accélération du nombre d'études combinant imagerie solaire et apprentissage automatique, avec un grand nombre d'études mentionnées dans ces corpus publiées après le début de cette thèse en octobre 2019. Parmi les grandes tendances observables, on peut citer le passage de modèles déterministes vers des modèles probabilistes, une plus grande attention portée sur l'évaluation juste des modèles, et l'imagerie EUV qui devient de plus en plus utilisée.

Les différentes études présentées dans les dernières sous-sections montrent que le *machine learning*, et plus spécifiquement les réseaux de neurones, semblent capables d'exploiter l'information contenue dans l'imagerie solaire de manière efficace, en améliorant un bon nombre de modèles de références, mais aussi de manière pertinente, en étant capable de distinguer et exploiter les régions du Soleil (trous coronaux, régions actives, etc.), voire même de saisir une partie de la physique solaire sous-jacente. Tout cela nous encourage dans la suite à proposer une nouvelle approche, exploitant l'imagerie solaire EUV, pour la prédiction directe d'indices géomagnétiques, décrite en détail dans les sections qui suivent.

5.2 Philosophie de notre nouvelle approche

Comme nous l'avons rapidement exposé en préambule de ce chapitre, notre objectif ici est de prolonger l'horizon de prédiction efficace des indices géomagnétiques au-delà de la limite de quelques heures établie dans les chapitres précédents. Afin de gagner un horizon de prédiction supplémentaire, il faut « prendre de l'avance » sur la propagation des plasmas (vent solaire, CMEs) potentiellement géoeffectifs. En remontant à leur source, c'est-à-dire au Soleil, on peut espérer gagner en horizon de prédiction un temps équivalent au temps de propagation de ces structures du Soleil à la Terre.

Rappelons que la distance Soleil-Terre moyenne est de :

$$1 \text{ au} \approx 1,5 \times 10^8 \text{ km.}$$

Sachant que la vitesse du vent solaire lent s'établit autour de $300 - 400 \text{ km} \cdot \text{s}^{-1}$ et celle du vent solaire rapide autour de $600 - 800 \text{ km} \cdot \text{s}^{-1}$ (voire plus), on peut estimer à l'ordre 0 que le temps de propagation du vent solaire entre le Soleil et la Terre varie entre un peu plus de 2 jours et 5 – 6 jours. La distribution de la vitesse de propagation des CMEs étant du même ordre de grandeur (YURCHYSHYN et al. 2005), on peut estimer leur temps de propagation du Soleil à la Terre comme étant compris entre 1 à 4 jours (IWAI et al. 2021). On peut donc espérer parvenir à prédire la géoeffectivité d'un événement solaire observé avec une avance allant de 2 jours jusqu'à 6 – 7 jours.

L'approche que nous proposons ici s'appuie sur l'utilisation de réseaux de neurones reliant directement des images du Soleil capturées entre des instants $t - T$ et t à l'indice géomagnétique Kp à un instant $t + \tau$. Le choix d'utiliser directement des images du Soleil et non des paramètres qui en sont dérivés provient de deux raisons principales. La première est le fait que toute l'information contenue dans un catalogue dérivé est contenue dans les observations du Soleil, l'inverse n'étant pas vrai. Étant donné qu'avec les algorithmes et la puissance de calcul actuels nous pouvons traiter des données de très haute dimension comme une série d'images, il devient intéressant d'utiliser directement dans des réseaux de neurones les images du Soleil et non des données dérivées de plus faible dimension. La deuxième raison est qu'en l'absence d'une automatisation fiable de l'extraction des caractéristiques (*features*) constituant un catalogue, il est nécessaire de les extraire à la main, ce qui n'est pas compatible avec la mise en place d'un *pipeline* opérationnel de prédiction automatique de bout en bout. Nous utilisons également uniquement des images EUV et non des magnétogrammes. Nous faisons ce choix car les magnétogrammes montrent essentiellement les régions actives, et pas les autres structures comme les trous coronaux, ou les boucles de champ coronales.

Il pourrait également sembler plus logique, ou en tout cas plus intuitif d'essayer, dans un premier temps, de prédire les caractéristiques du vent solaire, pour ensuite en déduire leur géoeffectivité d'une manière semblable à celle présentée dans les chapitres précédents, en séparant grossièrement les processus physiques de propagation du vent solaire du Soleil à la Terre d'un côté, et le couplage vent solaire–magnétosphère de l'autre. Bien qu'une telle approche ait son intérêt, notre

choix de prédire directement un indice géomagnétique s'explique par plusieurs arguments. En premier lieu, le fait de prédire directement l'indice géomagnétique à l'aide d'un unique modèle nous permet de plus facilement maîtriser l'incertitude du modèle. Si nous entraînaions plusieurs modèles entraînés séparément les uns à la suite des autres, il serait plus difficile de comprendre et estimer la propagation des erreurs. De plus, il ne faut pas perdre de vue qu'un réseau de neurones n'est qu'un estimateur très sophistiqué, s'appuyant sur un ensemble complexe de corrélations statistiques entre nos entrées et nos sorties. Par conséquent, à moins qu'il n'y soit contraint à l'aide de méthodes dédiées, un réseau de neurones ne va pas chercher à reproduire exactement toute la physique sous-jacente aux interactions Soleil-Terre, mais chercher à reproduire les corrélations entre nos jeux de données (corrélations qui sont la conséquence des processus physiques). Il est donc raisonnable de s'attendre à ce qu'un réseau de neurones suffisamment complexe et adapté, alimenté avec les bonnes données, soit capable de connecter directement le Soleil et son activité à la dynamique de la magnétosphère terrestre, en contournant en apparence les processus physiques intermédiaires. Enfin, il est notoirement difficile de prédire la composante B_z du champ magnétique interplanétaire à partir de caractéristiques sur l'activité solaire. Or B_z est un des paramètres du vent solaire les plus essentiels pour la prédiction d'indices géomagnétiques. Ainsi, prédire directement Kp est une façon de contourner le problème de la prédiction de B_z , qui semble au moins aussi complexe. C'est aussi pour cette raison que nous explorons ici une approche fondée uniquement sur les réseaux de neurones, et non sur des modèles purement physiques, car ces derniers montrent, à l'heure actuelle, des limites fortes quant à la prédiction de certains des paramètres du vent solaire à L1, et notamment B_z (voir par exemple MACNEICE et al. 2018 ; SCOLINI et al. 2019 ; ASVESTARI et al. 2021).

Notons cependant que les choix et contraintes adoptés dans notre modèle ne sont pas nécessairement définitifs. Insistons ici sur le fait que notre approche est principalement de chercher à vérifier s'il est possible de relier directement l'activité solaire (par le biais d'imagerie solaire) à l'activité géomagnétique, à l'aide des méthodes statistiques complexes que sont les réseaux de neurones. Ainsi, en fonction des résultats obtenus et des conclusions que l'on peut en tirer, il sera très certainement nécessaire de boucler sur les choix présentés ci-dessus et les adapter. Notre modèle n'est que le premier d'un processus de développement qui se veut itératif.

Maintenant que nos principaux choix méthodologiques ont été explicités, il convient de présenter notre modèle et son *pipeline* tel qu'ils ont été conçus et implémentés.

5.3 Présentation de l'architecture de **SERENADE**

Nous présentons ici l'architecture globale et détaillée de notre modèle de prédiction directe d'indices géomagnétiques à partir d'imagerie solaire, que nous avons baptisé *SERENADE* (Sun-EaRth intEraction dAta-DrivEn model). Nous

l'appliquons à l'indice Kp mais nous verrons dans le Chapitre 6 qu'il peut être utilisé pour la prévision d'autres indices, comme Ca .

Dans un premier temps nous exposons l'architecture globale de **SERENADE**. Ensuite, nous expliquons les données utilisées et leur pré-traitement. Après, nous détaillons chacun des modules de notre réseau de neurones et leur rôle. Enfin nous nous intéressons à la méthode d'apprentissage du modèle.

5.3.1 L'architecture globale

Avant de passer en revue et expliquer dans le détail le rôle et le fonctionnement de chaque module de **SERENADE**, nous présentons ici le modèle dans son ensemble. Un schéma de son architecture est donné en Figure 5.1.

Cette architecture générale est assez linéaire. Tout d'abord une séquence d'images du Soleil, correspondant chacune à un instant différent, passe par une étape de pré-traitement, qui a pour but d'en améliorer l'exploitabilité (par exemple en modifiant le contraste) et de les rendre compatibles avec le réseau de neurones. Toutes ces transformations sont déterministes et sont les mêmes pour toutes les images.

Ces images transformées alimentent ensuite le réseau de neurones. Nous avons conçu ce réseau de neurones de telle sorte à ce qu'il soit capable d'extraire de l'information utile de nos données en prenant en compte leur caractère intrinsèquement spatio-temporel. Pour cela nous utilisons trois modules :

- le Feature Extractor : il s'agit des couches du réseau de neurones en charge de l'extraction de caractéristiques (les *features*) utiles provenant du caractère spatial (ici, bidimensionnel car nos images sont mono-canal) de chacune des images du Soleil ;
- le Temporal Decoder : il s'agit des couches du réseau de neurones qui vont, à partir des caractéristiques spatiales préalablement obtenues, extraire des caractéristiques temporelles, provenant du fait que nous alimentons le réseau avec un ensemble d'images du Soleil correspondant à plusieurs instants. Nous obtenons après cette étape une représentation spatio-temporelle abstraite résumant l'ensemble des caractéristiques utiles contenues dans la séquence d'images du Soleil utilisées ;
- le Decoder : il s'agit des couches du réseau de neurones qui vont décoder la représentation spatio-temporelle obtenue pour la transformer en une prédiction probabiliste compréhensible par l'humain.

Dans la section 5.3.3 nous expliquons pourquoi, au moins dans cette première version de **SERENADE**, le Feature Extractor ne nécessite pas d'apprentissage, à l'inverse des autres sous-modules.

Enfin, nous obtenons donc la prédiction souhaitée. Comme nous travaillons maintenant sur des échelles temporelles qui ne sont plus de l'ordre de l'heure, mais de la journée, ici nous n'essaierons non-pas de prédire la valeur (tri-)horaire de Kp ,

mais son maximum journalier. Pour cela nous entraînerons plusieurs réseaux, un pour chacun des horizons de prédiction suivants : 2, 3, 4, 5, 6 et 7 jours.

Dans les sous-sections suivantes nous expliquons en détails chacune de ces étapes.

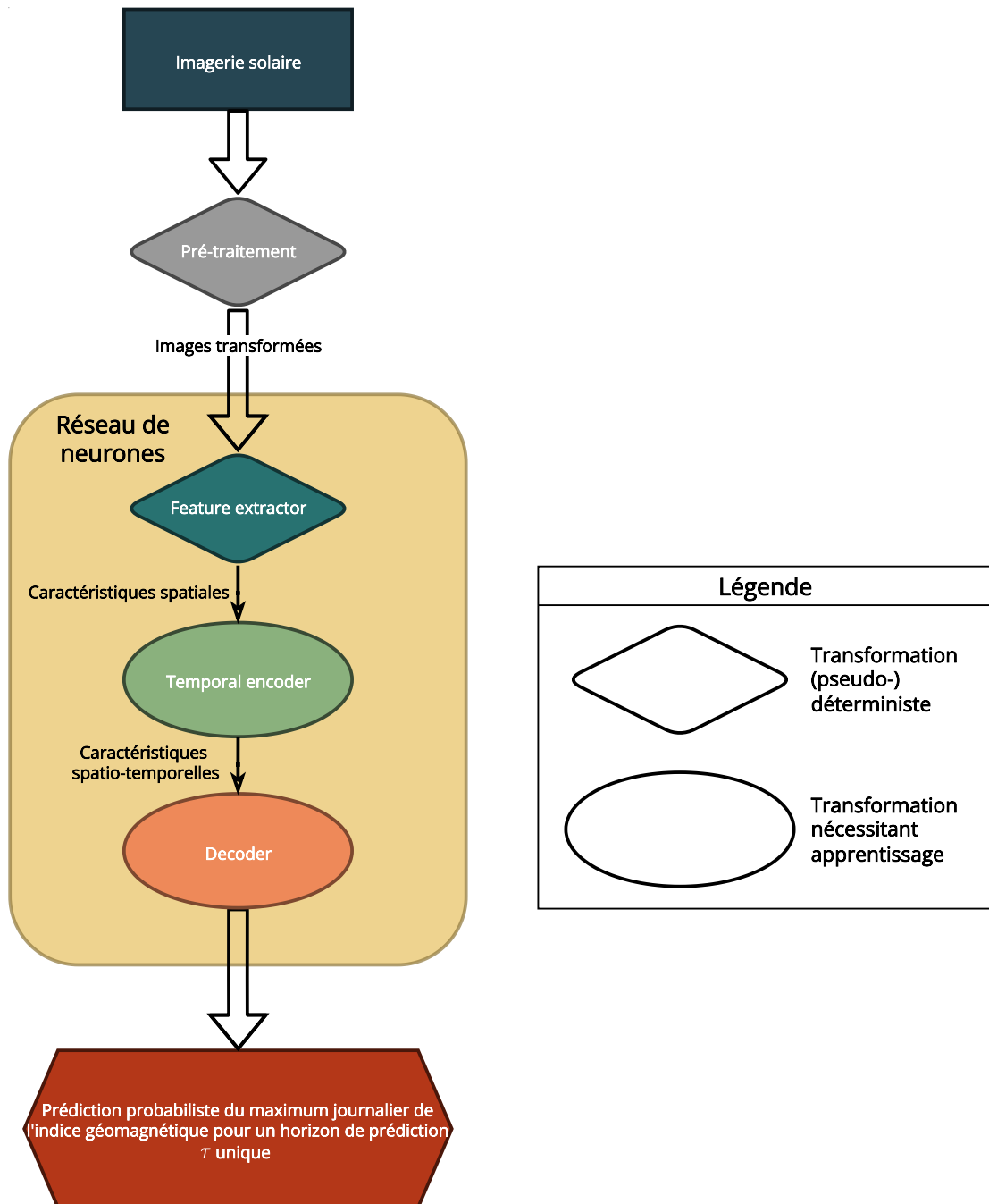


FIGURE 5.1 – Schéma de l'architecture globale de *SERENADE*.

5.3.2 Les données

Nous décrivons tout d'abord les données utilisées, et justifions les choix effectués, avant de présenter les transformations qui leur sont appliquées en amont du réseau de neurones.

5.3.2.1 L'imagerie solaire

Comme nous l'avons déjà mentionné, nous utilisons uniquement des images du Soleil comme paramètres d'entrée du modèle *SERENADE*. Avant cela il nous faut décider quelles images utiliser, et donc choisir au moins un instrument et une longueur d'onde d'observation.

Tout d'abord précisons rapidement que nous privilégions les télescopes spatiaux aux télescopes terrestres, qui, à cause de l'atmosphère terrestre, ont des capacités d'observation du Soleil dans l'*EUV* bien plus limitées. Le choix de l'instrument se joue alors essentiellement entre ceux embarqués dans les missions *SOHO* et *SDO*, deux missions observant le Soleil avec une ligne de visée quasi fixe, et fournissant des images *EUV* et des magnétogrammes avec une haute résolution et une fréquence temporelle élevée (comme nous l'avons déjà vu au Chapitre 2). En plus des performances d'acquisition plus élevées de *SDO* par rapport à *SOHO*, un élément qui a été déterminant dans notre choix a été la publication au début de cette thèse par GALVEZ et al. 2019 d'un jeu de données d'images *SDO* déjà partiellement pré-traité pour des applications d'apprentissage automatique. Ce jeu de données inclut les images fournies par les trois instruments *AIA*, *HMI* et *EVE* à toutes les longueurs d'onde disponibles entre mai 2010 et décembre 2018. La publication de cet ensemble de données nous a définitivement convaincus de commencer cette étude avec les données de *SDO*. Nous détaillons dans la sous-section 5.3.2.2 l'ensemble des transformations déjà effectuées dans les données de GALVEZ et al. 2019.

Les données de GALVEZ et al. 2019 (que nous appellerons dans la suite « données *SDO/AIA – ML-READY* ») contiennent donc toutes les longueurs d'ondes d'acquisition de *SDO/AIA*, au format NPZ, lisible à l'aide de la librairie Python *NumPy* (HARRIS et al. 2020). Nous choisissons d'utiliser dans notre modèle les images obtenues à la longueur d'onde 193 Å. En effet, à cette longueur d'onde on observe bien la couronne solaire, avec ses trous coronaux, ainsi que les régions éruptives, ce qui synthétise bien les origines des phénomènes potentiellement géoeffectifs qui nous intéressent. Notons à posteriori que UPENDRAN et al. 2020 ont utilisé les longueurs d'onde 193 Å et 211 Å pour les mêmes raisons, et trouvé que la première donnait de meilleurs résultats pour la prédiction de V_{sw} sans doute en raison de la mise en valeur des trous coronaux sur cette longueur d'onde.

Enfin, nous devons choisir le nombre d'images et leur étendue temporelle utilisées en entrée du modèle pour obtenir une prédiction de *Kp*. Comme nous avons vu que le vent solaire lent met au plus environ 6 jours à atteindre la Terre depuis le Soleil, il semble peu utile (au moins dans ce premier modèle) de prendre des images largement au-delà de cette fenêtre temporelle. Comme notre horizon

de prédiction le plus court est de 2 jours, prendre des images comprises entre t et $t - 5j$ devrait nous suffire. Finalement, il nous faut choisir le nombre d'images par jour que nous utiliserons. Ce choix est assez important, car si le pas de temps entre deux images est trop grand alors nous pouvons manquer des événements, dont le temps caractéristique d'observation est de l'ordre de quelques minutes à quelques heures. Or, un trop grand nombre d'images n'est pas possible pour des raisons de temps d'apprentissage, et d'espace dans la mémoire de la carte graphique utilisée pour l'apprentissage. Nous faisons donc ici un premier choix radical de ne prendre qu'une image toutes les 12 heures (au plus). Cela va certes probablement nous faire manquer beaucoup d'embrillancements et de *CMEs*. Mais, comme nous le verrons également plus bas avec d'autres choix de design, cela nous permet de garder un temps d'apprentissage correct pour le développement de cette première version de *SERENADE*. À titre d'exemple, nous avons estimé que le temps d'apprentissage du réseau de neurones est multiplié par 4 lorsqu'on utilise 12 images par jour au lieu de 2. Ce facteur est très important (voir la sous-section 5.3.4), et étant donné la nature intrinsèquement itérative du développement d'un modèle de réseau de neurones fonctionnel, un temps à ce point rallongé peut se révéler prohibitif. Ce choix est donc nécessaire et imposé par des contraintes certes non-physiques, mais qu'il nous est impossible d'ignorer.

Toutefois ce choix n'est pas non plus hors de propos. En effet, les structures comme les trous coronaux ont une durée d'observation bien plus grande, de l'ordre de plusieurs jours, voire semaines. Par ailleurs, même si on rate un embrillancement, on peut espérer observer certains de ses précurseurs. Nous utilisons donc des séquences d'images allant de $t - 4,5j$ à t avec un pas de temps de 12 heures entre chaque image, ce qui nous donne des séquences de 10 images. Si une image correspondant à un instant donné que nous voulons utiliser est manquante, nous utilisons l'image qui en est la plus proche, avec un écart maximal de 6 heures par rapport à l'instant initialement voulu. Si jamais une telle image n'existe pas, alors nous défaussons la séquence entière.

5.3.2.2 Le pré-traitement des données d'imagerie solaire

Nous avons établi que nous utilisons ici pour chaque prédiction une séquence de 10 images *EUV* fournies par *SDO/AIA* à la longueur d'onde 193 Å. Avant d'être utilisables au sein d'un modèle composé de réseau de neurones il est nécessaire de les transformer pour plusieurs raisons :

- s'assurer de la bonne qualité des données utilisées ;
- s'assurer que la qualité des données est homogène dans l'ensemble du jeu de données ;
- mettre en valeur certaines caractéristiques que nous considérons à priori comme importantes ;
- respecter certaines contraintes numériques imposées par la machine de calcul et le langage de programmation ;
- respecter certaines contraintes extérieures, comme un temps de calcul limité.

C'est afin de répondre à certains de ces points que nous utilisons le jeu de données [SDO/AIA – ML-READY](#). En effet, les auteurs de cet ensemble de données ont effectué les opérations suivantes :

- les images corrompues (partiellement ou totalement) ont été enlevées ;
- les images ont été recalibrées afin de compenser la dégradation due au vieillissement de l'instrument ;
- le pas de temps entre deux images a été augmenté à 6 minutes ;
- la résolution des images a été réduite à 512×512 ;
- les pôles du Soleil ont été alignés sur toutes les images ;
- le Soleil est centré et occupe la même surface sur chaque image.

La principale valeur ajoutée des données [SDO/AIA – ML-READY](#) est leur homogénéité, due à la recalibration et au ré-alignement des images, et leur « propreté » grâce à la suppression des données corrompues. Néanmoins il ne faut pas négliger les avantages dûs aux autres opérations. En effet, réduire la fréquence d'acquisition et la résolution des images permet également de réduire la taille du jeu de données complet à environ 6,5 To, au lieu de plusieurs pétaoctets (ce qui n'est pas compatible avec nos moyens de calcul scientifique). Certes, nous perdons alors une des qualités de [SDO/AIA](#), qui est de fournir des images hautement résolues (4096×4096 pixels). Toutefois, de telles résolutions sont difficilement compatibles avec des applications de *machine learning* (plus la dimension du problème est grande, plus nous avons besoin d'échantillons, et plus l'apprentissage est long).

Nous avons téléchargé uniquement les images correspondant aux heures entières (ce qui nous donne au plus 24 images par jour). Ainsi, la taille de notre jeu de données pour la longueur d'onde 193 \AA est de 66 Go, ce qui est raisonnable, et correspond à nos besoins, au moins dans un premier temps.

A ces pré-traitements déjà effectués dans l'ensemble de données que nous utilisons, nous rajoutons un ensemble de transformations destinées à rendre les images plus exploitables. Tout d'abord, afin de renforcer le contraste nous appliquons un écrêtage aux valeurs d'intensité de chaque pixel. Cette opération est assez classique et est effectuée dans d'autres études (cf. par exemple AHMADZADEH et al. 2019). Pour déterminer les valeurs minimales et maximales pour l'écrêtage nous nous sommes appuyés sur la statistique des valeurs d'intensité des pixels fournie dans AHMADZADEH et al. 2019 (dans lequel il est trouvé que le 95ème percentile des valeurs d'intensité des pixels pour $\lambda = 193 \text{ \AA}$ est de 3968), à partir de laquelle nous avons fait une recherche empirique. Après quelques tests itératifs, nous avons convergé vers les valeurs suivantes :

$$\begin{aligned} V_{min} &= 100, \\ V_{max} &= 5000. \end{aligned}$$

Après écrêtage, afin de réduire l'étalement de la distribution des valeurs des pixels, nous passons toutes les images en \log_{10} . Pour illustrer notre propos, la

Figure 5.2 montre à quoi ressemble une image à la longueur d'onde 193 Å avant transformation, après écrêtage seul, après écrêtage et passage au logarithme.

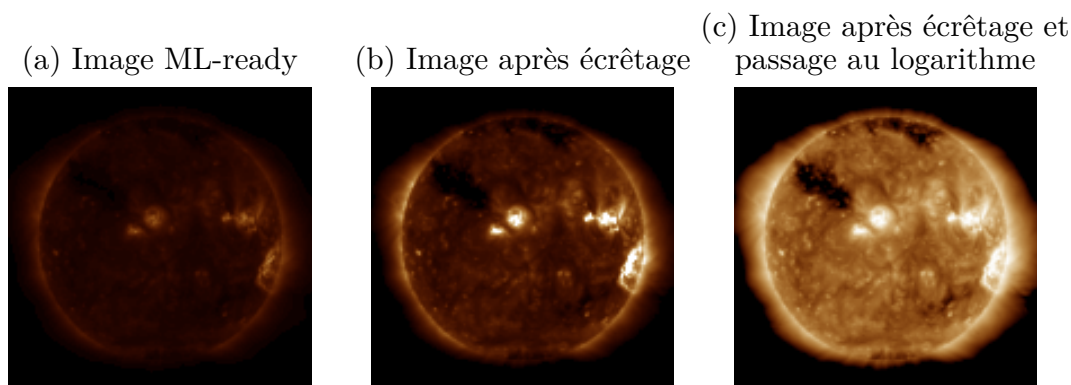


FIGURE 5.2 – Image du Soleil à la longueur d'onde 193 Å capturée le 08 septembre 2017 12h00 par *SDO/AIA*. (a) Image fournie dans GALVEZ et al. 2019 sans pré-traitement supplémentaire. (b) La même image après écrêtage. (c) La même image après écrêtage et passage au logarithme base 10.

Pour assurer la stabilité numérique lors de l'apprentissage du réseau, les valeurs des pixels sont rééchelonnées entre 0 et 1. Enfin, afin d'assurer la compatibilité de nos images avec le module Feature Extractor de notre réseau (cf. section 5.3.3.1), nos images, qui étaient préalablement en échelle de gris et de dimension 512×512 sont transformées en images de type RGB (en recopiant l'unique canal trois fois) de dimension 224×224 (en faisant une interpolation bilinéaire) et centrées et réduites en utilisant des coefficients de valeur $[0.485, 0.456, 0.406]$ pour la moyenne de chaque canal RGB et $[0.229, 0.224, 0.225]$ pour l'écart-type de chaque canal RGB.

Notons que l'ensemble des transformations décrites ici sont déterministes (il n'y a pas de composante aléatoire). À chaque fois que nous utilisons ces images dans notre réseau, nous devons les transformer de la sorte, ce qui représente un temps non négligeable lors de l'étape d'apprentissage du modèle. C'est pourquoi nous avons également créé une copie de notre jeu d'images déjà entièrement pré-transformées et stockées dans un unique fichier au format HDF5 (THE HDF GROUP 2000).

5.3.2.3 L'indice géomagnétique K_p

Comme nous l'avons rapidement précisé en section 5.3.1, nous cherchons à prédire la valeur maximale journalière de K_p . Par valeur maximale journalière nous entendons la valeur maximale atteinte par l'indice K_p dans une fenêtre glissante de longueur 24 h, et non pas la valeur maximale atteinte par K_p au cours d'une date fixe (de 00h00 à 23h59). Ainsi, pour une prévision 2 jours en avance, effectuée à un instant t , la valeur cible de K_p correspondant à l'instant $t + 2j$ que nous voulons prédire est le K_p maximal atteint entre $t + 24$ h et $t + 48$ h. Nous notons cette valeur $K_{p_{\max,24\text{h}}}$.

Outre ce passage de valeurs (tri-)horaires à des valeurs « journalières », nous ne procédons pas à d'autre transformation (rappelons qu'il n'y a pas de trous de données dans la série temporelle de Kp).

5.3.3 Le réseau de neurones

Nous nous intéressons maintenant à l'architecture de notre réseau de neurones, et au détail de chacun de ses sous modules.

5.3.3.1 Le Feature Extractor

Le Feature Extractor est le premier module de notre réseau de neurones. Il s'agit d'un réseau de type **CNN**, dont nous avons expliqué le principe de fonctionnement au Chapitre 2. Ces réseaux ont depuis une dizaine d'années montré leur très grande efficacité dans le traitement de problèmes de vision par ordinateur, notamment pour la classification d'images ou de vidéos, la segmentation d'images et la détection d'objets. Les **CNNs** font partie des types de réseaux de neurones les plus utilisés dans les articles cités en section 5.1.

Une difficulté, déjà évoquée dans la sous-section 5.3.2.1, s'est d'emblée présentée à nous : les temps d'apprentissages très longs. Nous nous sommes très rapidement aperçus qu'un apprentissage classique d'un réseau convolutionnel « à partir de zéro » serait quasiment impossible à cause du temps de calcul que cela implique. En effet, l'utilisation d'environ 10^5 séquences de 10 images de relativement haute dimension (512×512) nécessite un très grand nombre ($\approx 10^6$ par *epoch*) d'accès disque en lecture, qui ajoutés au temps de passage et de rétropropagation du gradient dans le réseau de neurones, nous donnaient des *epochs* d'apprentissage d'une durée unitaire de quelques dizaines de minutes (alors qu'un apprentissage complet nécessite plusieurs dizaines d'*epochs*, ainsi que le recours à une méthode d'évaluation croisée imbriquée – voir section 5.3.4). Un tel temps d'apprentissage étant incompatible avec le développement d'un premier modèle (cette première version de **SERENADE** étant, rappelons-le, une preuve de concept), nous avons dû chercher un moyen de contourner ce problème.

Une partie de la solution a déjà été apportée en sous-section 5.3.2.1, où nous avons mentionné que le fait de ne prendre que deux images par jour était essentiellement dû à des raisons de temps de calcul. Toutefois, la clef pour résoudre ce problème a été d'avoir recours à l'apprentissage par transfert (*Transfer Learning*). Pour rappel, l'apprentissage par transfert est une technique consistant à transférer de la connaissance et de la compétence acquise en résolvant un certain problème vers un autre problème (PAN et YANG 2010).

Plus concrètement, l'idée est d'utiliser un réseau de neurones profond pré-conçu et ayant déjà été entraîné sur certaines données pour résoudre un autre problème à partir d'autres données. En effet, intuitivement, les couches profondes d'un **CNN** ayant été préalablement entraîné sur une base de données pour résoudre, par exemple, un problème de classification d'images, capturent des représentations,

formes et caractéristiques abstraites des images, et que ce sont seulement les toutes dernières couches qui vont ensuite classer les images, en s'appuyant sur ces représentations abstraites. Ainsi, on pourrait utiliser un réseau pré-entraîné soit comme point de départ d'un nouvel entraînement, avec de nouvelles données, soit tel quel, sans apprentissage supplémentaire, en figeant les coefficients du réseau et en enlevant les couches « spécialisées », c'est-à-dire en ne conservant que les couches profondes, qui extraient des représentations abstraites.

L'apprentissage par transfert est de plus en plus populaire, surtout dans le domaine de la classification d'images, dans des cas d'application où la quantité d'échantillons (*i. e.* d'images) disponibles n'est pas très élevée par rapport à la dimension du problème. Surtout, l'apprentissage par transfert a montré son efficacité dans de nombreux problèmes, par exemple dans le domaine de la classification d'images biomédicales (RAZAVIAN et al. 2014; PAWLOWSKI et al. 2016; LOPES et VALIATI 2017), en utilisant des réseaux de neurones pré-entraînés sur la base de données ImageNet (DENG et al. 2009), qui inclut plus de 14 millions d'images portant sur des milliers de concepts.

Le package `torchvision` pour Python, développé par l'équipe de `Pytorch`, propose de nombreux réseaux convolutionnels pré-entraînés sur la base de données ImageNet. Parmi ceux proposés, nous utilisons ici le modèle GoogLeNet (SZEGEDY et al. 2014). Il s'agit du modèle ayant remporté la compétition ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC 2014) portant sur la classification d'images d'ImageNet. Il s'agit d'un réseau profond s'appuyant sur des modules appelés « Inception ». Le module Inception est principalement composé d'un ensemble de couches de convolution avec différentes tailles de noyau (1×1 , 3×3 et 5×5) mises en parallèle, afin d'extraire de l'information de l'image à différentes échelles spatiales. Les couches de noyau de taille 3×3 et 5×5 sont précédées d'un noyau de taille 1×1 qui sert à réduire la dimension des sorties des couches précédentes (selon la dimension des filtres). Un schéma du module inception est donné en Figure 5.3.

Le réseau GoogLeNet est composé de 9 modules Inception superposés (auxquels se rajoutent quelques couches de type MaxPooling afin de diminuer petit à petit la dimension de l'image). Après les 9 modules Inception, une seule couche dense, suivie de la fonction d'activation SoftMax, sert de classifieur.

Dans *SERENADE* nous utilisons donc un réseau GoogLeNet pré-entraîné, auquel nous enlevons les couches de classification. Nos *features* sont donc les sorties obtenues après le dernier module Inception (et la couche de AveragePooling qui suit). Ainsi, à partir d'une image de dimensions $3\times 224\times 224$ nous obtenons un vecteur de taille 1024×1 qui résume de façon abstraite l'information contenue dans l'image. Nous avons choisi d'utiliser le réseau GoogLeNet, et pas une architecture plus récente, comme par exemple ResNet-5050 (HE et al. 2015) car GoogLeNet est un réseau moins profond, comprenant donc moins de paramètres (≈ 7 millions pour GoogLeNet contre ≈ 25 millions pour ResNet-50) pour des performances similaires, et la taille du vecteur de sortie pour GoogLeNet (1024) est deux fois plus petite que chez la plupart des autres modèles pré-entraînés proposés dans `torchvision`

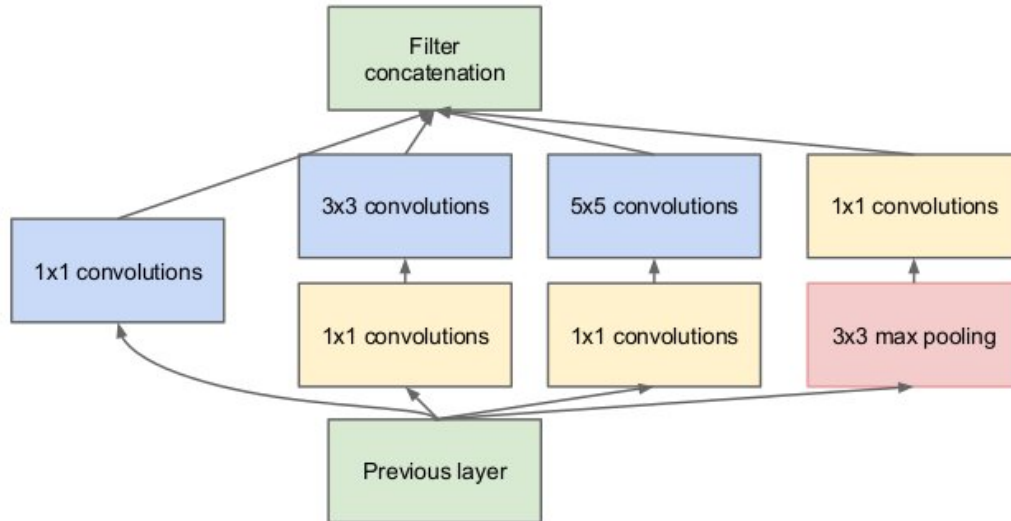


FIGURE 5.3 – Architecture du module Inception. Extrait de SZEGEDY et al. 2014.

(2048). Cela est, dans notre cas, a priori un avantage dans l'optique de connecter ce Feature Extractor à un Temporal Decoder, afin d'éviter que celui-ci n'ait trop de paramètres à entraîner.

Nous avons procédé à quelques essais à l'aide d'autres architectures utilisées en tant que Feature Extractor pré-entraîné, comme ResNet-50, ResNet-152, VGG16 (SIMONYAN et ZISSERMAN 2015), et nous avons toujours obtenu des performances au moins égales, si ce n'est meilleures avec GoogLeNet, pour un temps d'inférence plus faible. Il ne s'agit ici pas d'une conclusion définitive quant à l'efficacité de ces modèles pour un apprentissage par transfert pour la prédiction de $Kp_{\max,24h}$ (il nous faudrait à minima un banc de test complet pour cela) mais d'observations empiriques, qui à ce stade nous suffisent.

Enfin, un dernier avantage de l'utilisation d'un Feature Extractor pré-entraîné est qu'il est possible de faire passer toutes nos images du Soleil à travers ce réseau une unique fois, puis de sauvegarder et réutiliser directement les *features* obtenues en entrée du Temporal Decoder. Cela nous permet de gagner un temps précieux d'accès disque en lecture – nos entrées sont alors en effet de dimension 10×1024 au lieu de $10 \times 3 \times 224 \times 224$ (on gagne un facteur 147) – qui en cumul nous permet de diviser le temps d'une *epoch* d'apprentissage par 30.

En dernier mot, notons que comme nous ne contrôlons pas directement le Feature Extractor, il faut s'assurer que les valeurs numériques extraites de celui-ci sont compatibles avec une utilisation dans le reste du réseau de neurones. C'est pourquoi nous ré-échelonnons les valeurs de chacune des 1024 *features* pour qu'elles aient des valeurs entre 0 et 1 (les valeurs maximales et minimales admises par chaque *feature* étant calculées sur l'ensemble d'entraînement).

Un schéma synthétique du Feature Extractor est donné en Figure 5.4.

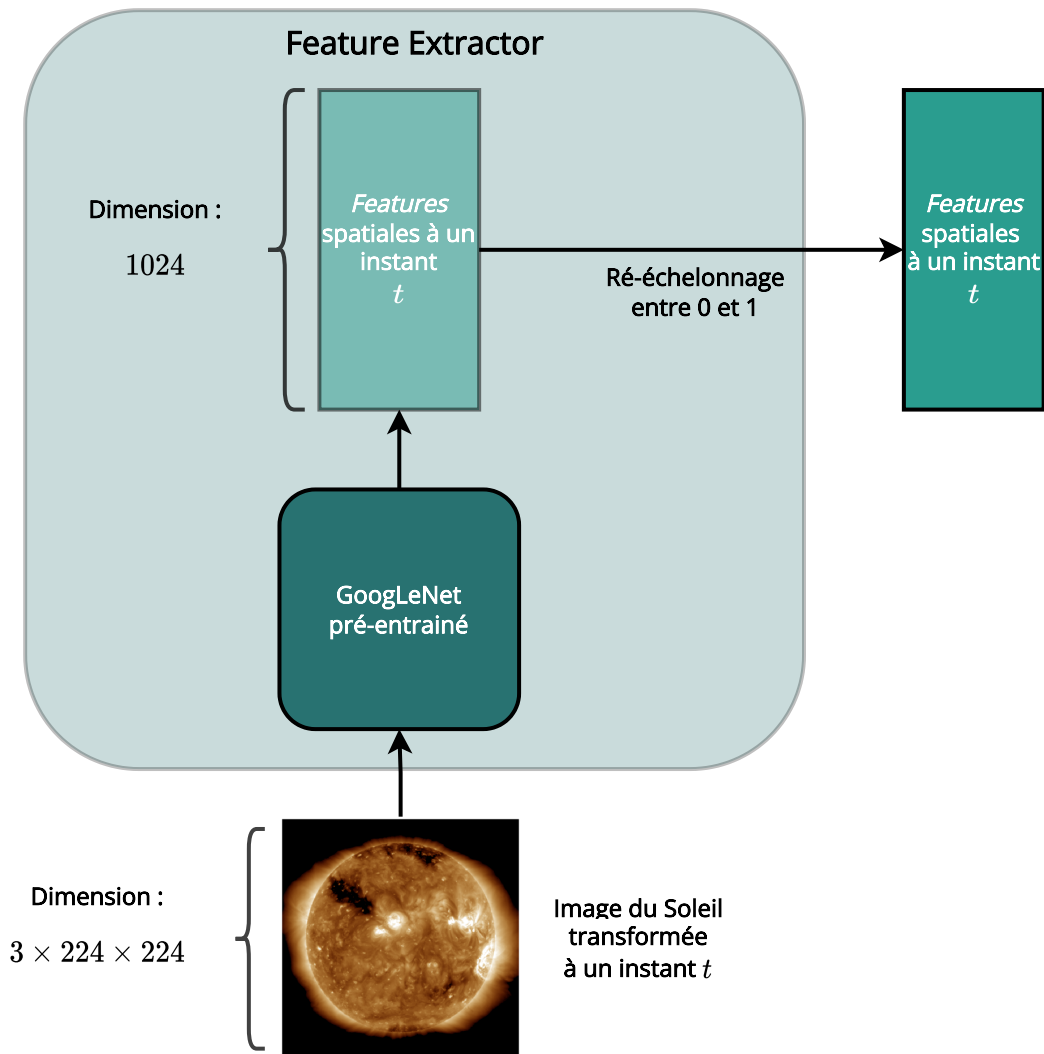


FIGURE 5.4 – Schéma synthétique du Feature Extractor.

5.3.3.2 Le Temporal Encoder

Le Temporal Encoder est le module du réseau de neurones qui est chargé d'extraire de l'information du caractère temporel de nos données. En effet, comme nous l'avons établi, nous alimentons notre réseau avec une séquence de 10 images successives du Soleil, qui sont transformées en 10 vecteurs de *features* après passage dans le Feature Extractor. L'objectif suivant est donc de combiner ces vecteurs pour en extraire de l'information provenant de l'évolution temporelle des caractéristiques extraites des images.

Pour ce faire nous utilisons simplement un réseau de type *LSTM*, la même architecture que nous avons déjà utilisé pour la prédiction des indices géomagnétiques à partir des paramètres du vent solaire mesurés en L1. Le *LSTM* que nous utilisons dans *SERENADE* possède deux couches et ses états cachés (pour les deux couches) sont de dimension 2048. Il s'agit d'un *LSTM* de type mono-directionnel. Un schéma synthétisant l'architecture du Temporal Decoder est donné en Figure 5.5.

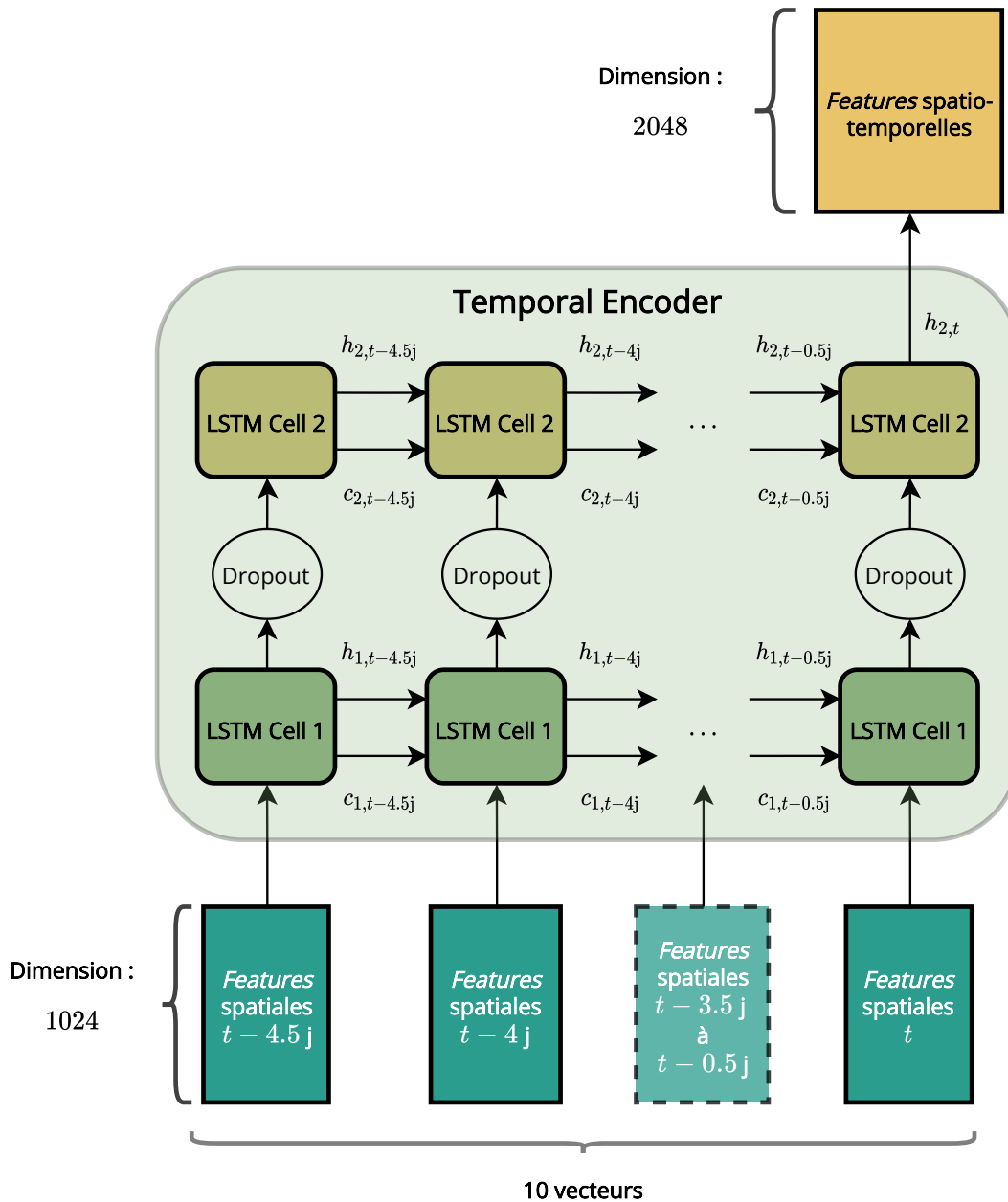


FIGURE 5.5 – Schéma synthétique du Temporal Encoder.

5.3.3.3 Le Decoder

Le Decoder est le module chargé de convertir les caractéristiques spatio-temporelles (extraites suite au passage des images d'abord dans le Feature Extractor, puis dans le Temporal Encoder) en une prédiction compréhensible par l'humain.

Ici, nous souhaitons obtenir des prédictions probabilistes. En effet, comme nous l'avons vu au Chapitre 3, les prédictions probabilistes ont plusieurs avantages. Tout d'abord, une prédiction probabiliste est plus utile qu'une prédiction déterministe d'un point de vue opérationnel car il est souvent plus aisé de prendre une décision lorsqu'on peut quantifier un risque avec un niveau de probabilité. De plus, le fait de faire en sorte que le réseau, dans son apprentissage, prenne en compte cette

incertitude sur la sortie peut également le rendre plus précis, comme nous l'avons vu.

De même que précédemment, nous chercherons à avoir des prédictions sous la forme d'une distribution gaussienne. Il nous faut donc prédire deux valeurs, la moyenne et l'écart-type de la gaussienne. Nous adoptons la même démarche que dans les chapitres précédents, en séparant le Décodeur en deux sous-parties parallèles, en charge du décodage de la moyenne ou de l'écart-type. Chacune des sous-parties est ici composée de trois couches linéaires de taille décroissante. La première a une dimension de sortie égale à 512, la deuxième à 64, la dernière à 1. Une fonction d'activation de type *ELU* et une couche de type *dropout* sont utilisées après les deux premières couches de chacune des branches du Décodeur. Un schéma de l'architecture du Décodeur est donné en Figure 5.6. Notons que les valeurs de ces dimensions (et le nombre de couches) ont été obtenues par convergence après quelques essais, et semblaient nous donner les meilleurs résultats. Cependant, ce ne sont peut-être pas les valeurs optimales, mais une véritable recherche par grille serait trop coûteuse. Comme chaque branche du Décodeur est composée de trois couches, chacune d'entre elles peut être vue comme un *MLP*.

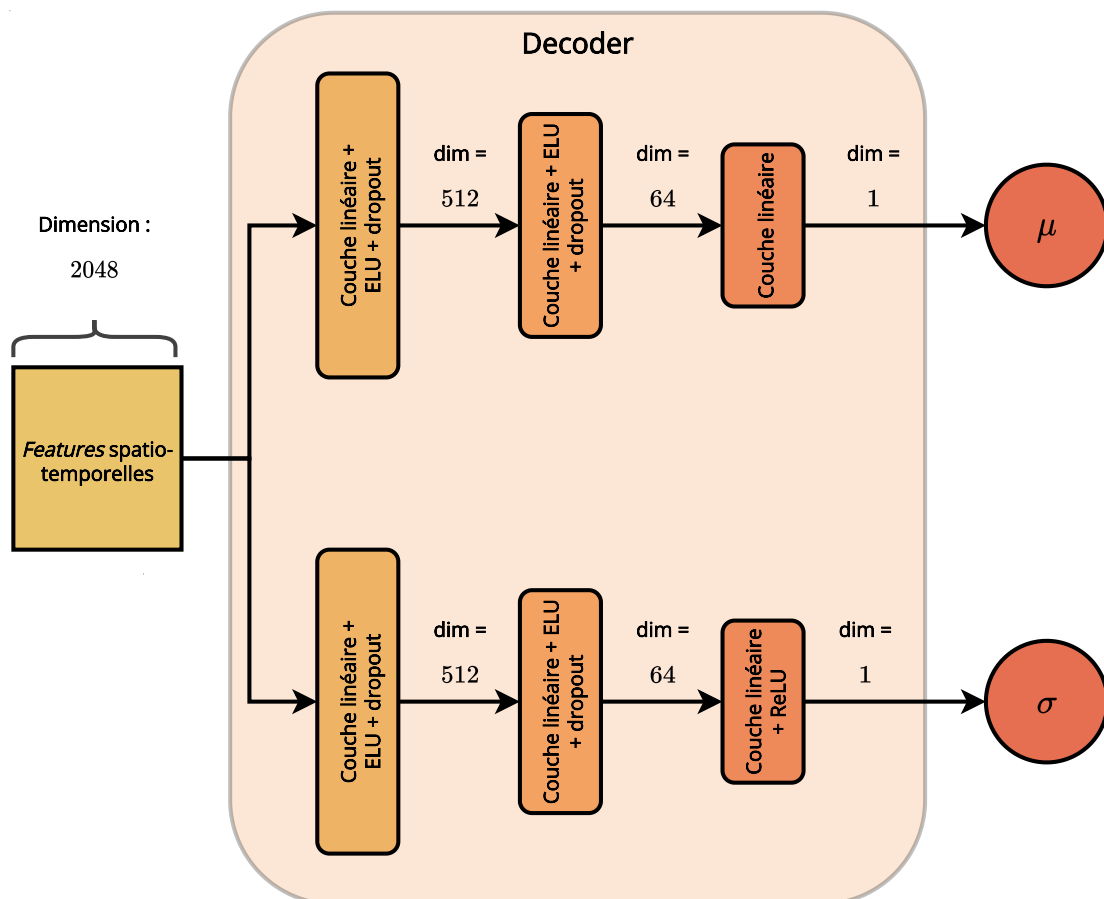


FIGURE 5.6 – Schéma synthétique du Décodeur. μ et σ représentent respectivement la moyenne et l'écart-type d'une distribution gaussienne.

Notons que pour nous assurer que l'écart-type soit positif nous plaçons une fonction d'activation de type *ReLU* en sortie de la branche prédisant sa valeur.

5.3.3.4 Récapitulatif de l'architecture

Récapitulons ici l'architecture globale de **SERENADE**. Tout d'abord 10 images du Soleil prises entre des instants $t-4,5j$ et t , à la longueur d'onde 193 \AA , et issues de l'ensemble **SDO/AIA – ML-READY** sont transformées afin d'en améliorer la qualité, et les rendre compatibles avec la suite. À l'aide d'un réseau de type GoogLeNet pré-entraîné sur la base de données ImageNet, et dont les dernières couches (spécialisées) ont été enlevées, nous extrayons de chaque image un vecteur de *features* spatiales de longueur 1024, ré-échelonné entre 0 et 1. Ces *features* spatiales alimentent un réseau de type LSTM, ayant une seule couche, et de dimension 2048. Le réseau LSTM fournit un unique vecteur de *features* spatio-temporelles, qui sont converties par un Decoder composé de deux MLPs en parallèle, qui prédisent la moyenne et l'écart-type d'une distribution gaussienne. Ces deux valeurs sont combinées pour fournir une prédiction probabiliste de la valeur maximale atteinte par Kp au cours d'une fenêtre de longueur 24 h se terminant à un moment futur $t + \tau$ (notée $Kp_{\max,24h}$). La Figure 5.7 synthétise cette architecture.

5.3.4 L'apprentissage du modèle

Maintenant que l'architecture de **SERENADE** est établie, nous détaillons ici les choix effectués pour la phase d'apprentissage du réseau de neurones, que ce soit en termes d'implémentation pratique, que de choix de méthodes d'optimisation. Toutefois, avant cela il nous faut aborder la question du choix des jeux d'entraînement, de validation, et de test.

5.3.4.1 Séparation des données pour entraînement, validation et test

Le jeu de données **SDO/AIA – ML-READY** inclut les images capturées par **SDO** entre le 13 mai 2010 et le 9 décembre 2018. Cela correspond à une grande partie du cycle solaire 24 (janvier 2008 – décembre 2019). Cependant, comme nous l'avons déjà évoqué, ce cycle solaire a été particulièrement faible. Nous avons déjà évoqué au Chapitre 4 le fait que les périodes récentes pouvaient être moins représentatives du point de vue de l'indice Ca . C'est aussi cette raison qui nous a poussés, aux Chapitres 3 et 4 à choisir notre ensemble de test au cours du cycle solaire 23, bien plus dynamique et intense, incluant des événements extrêmes du point de vue des ceintures de radiations.

Or, il n'y a pas eu d'événement extrême au cours du cycle solaire 24. Pire, celui-ci a été très peu intense comparé au précédent, ce qui se voit dans la distribution des valeurs de Kp , comme le montre la Figure 5.8. Cette Figure montre clairement que la distribution des valeurs prises par Kp au cours des deux cycles est très différente. En effet, la distribution du cycle solaire 24 est bien plus concentrée autour de valeurs basses de Kp . Comme le montre la sous-figure c), il y a, proportionnellement au nombre d'échantillons, environ 2,1 fois plus de chances de voir $Kp = 0$ dans le cycle solaire 24 que dans le précédent. De même, les moyennes et grandes valeurs de Kp (dès que $Kp \geq 2$) sont plus rares dans le cycle 24 par rapport au 23, et cette raréfaction s'intensifie plus la valeur de Kp est grande.

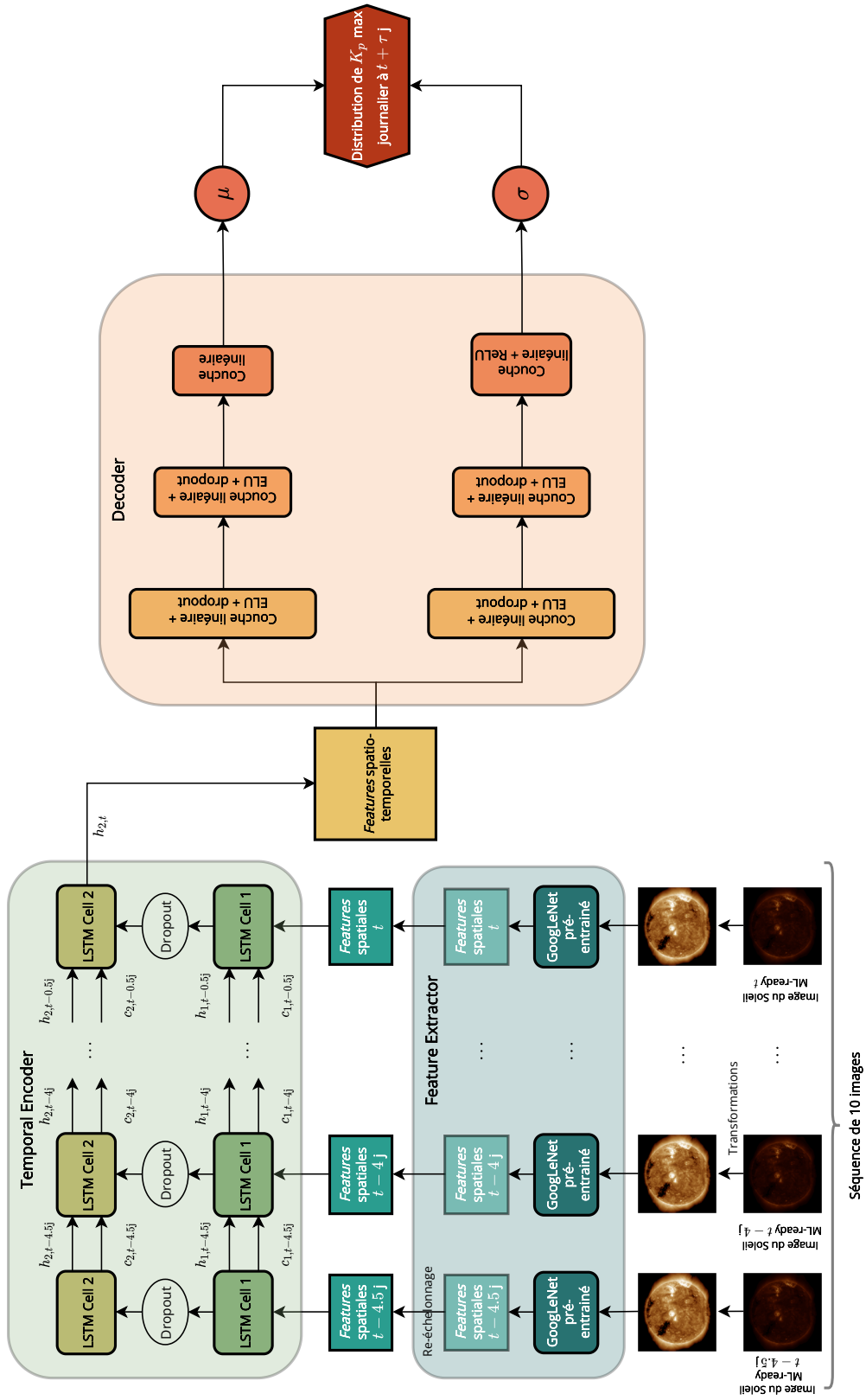


FIGURE 5.7 – Schéma synthétique de SERENADE. μ et σ représentent respectivement la moyenne et l'écart-type d'une distribution gaussienne.

Ainsi notre base de données est encore plus déséquilibrée qu'auparavant (dans les exemples des chapitres précédents) entre les valeurs fortes et les valeurs faibles de l'indice géomagnétique. De plus, le nombre (absolu) d'échantillons correspondant à des périodes d'orage sera faible, ce qui peut être problématique.

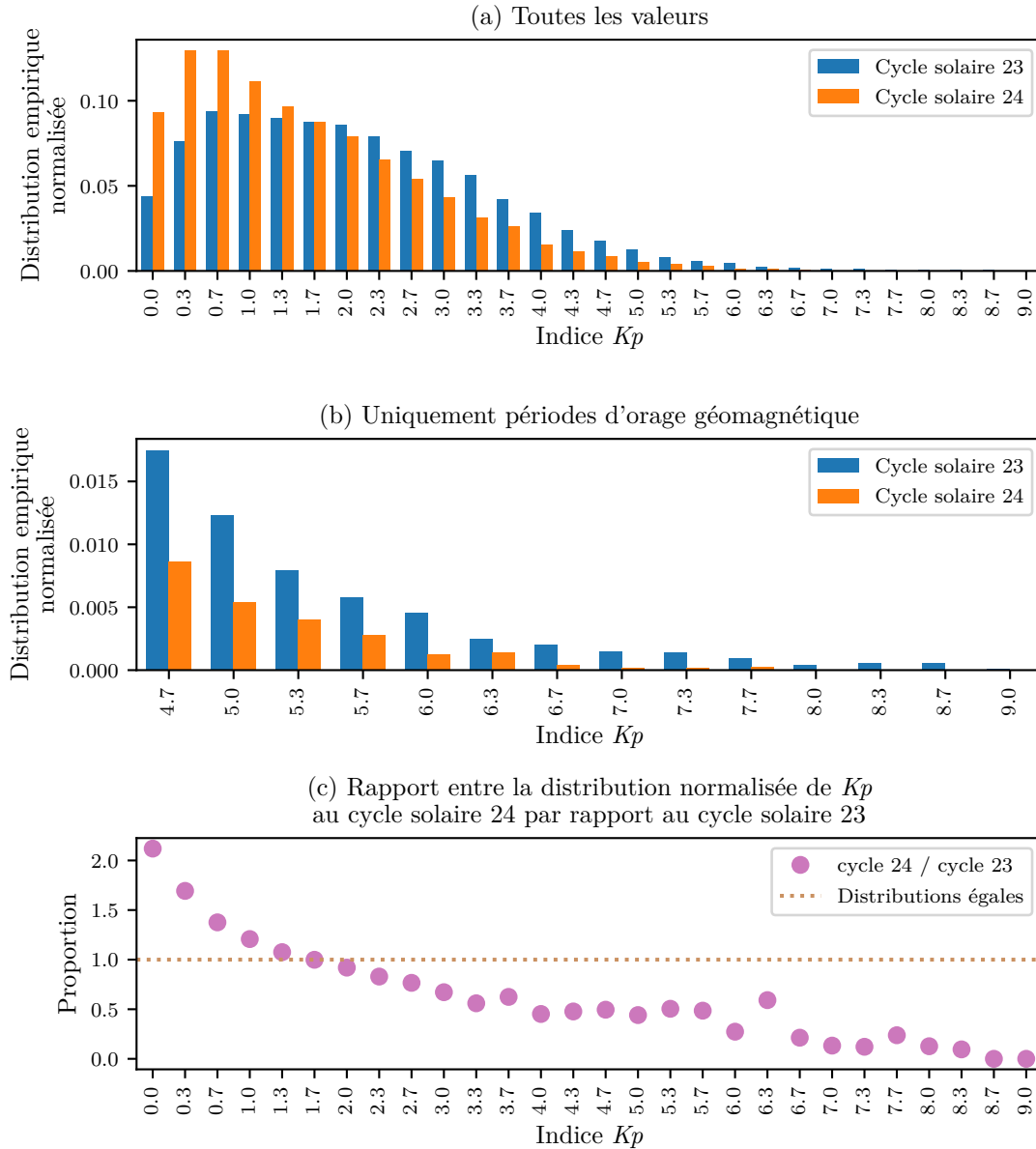


FIGURE 5.8 – Distribution normalisée des valeurs prises par l'indice Kp au cours des cycles solaires 23 et 24, a) pour toutes les valeurs possibles et b) uniquement pour les périodes d'orage. Le rapport entre les distributions obtenues pour le cycle solaire 23 par rapport au cycle solaire 23 est donnée en c).

Un autre problème soulevé par l'utilisation du jeu *SDO/AIA – ML-READY* est sa couverture temporelle relativement restreinte, couvrant seulement environ 8 années et demie. Comme nous l'avons dit, cela représente moins d'un cycle solaire, alors que dans les Chapitres 3 et 4 nous utilisons près de 24 années de données,

soit presque trois fois plus de données. Dans l'absolu avoir moins de données n'est pas forcément un problème, mais ici cela l'est pour deux raisons.

La première raison est que'n travaillant avec des données d'imagerie solaire, la dimension de notre problème est plus élevée. À cause du fléau de la dimension il nous faut travailler avec d'autant plus d'échantillons. Le fait d'utiliser un Feature Extractor pré-entraîné permet de réduire la dimension de notre problème – c'est une raison supplémentaire (en plus de celles déjà mentionnées en sous-section 5.3.3.1) pour laquelle nous avons utilisé cette approche. C'est aussi pour cela que nous avons privilégié le réseau GoogLeNet qui donne en sortie des vecteurs de dimension 1024, contre 2048 pour certains de ses concurrents.

La deuxième raison est que nos données couvrent moins d'un cycle solaire. Or nous savons que chaque phase d'un cycle solaire peut être très différente, avec des dynamiques propres, qui ont des origines solaires différentes (par exemple, une augmentation des trous coronaires dans la phase descendante d'un cycle). En ayant moins d'un cycle solaire de données, le risque est que nous entraînions un réseau sur une période du cycle solaire qui soit peu représentative de ce qui est présent dans les ensembles de validation et de test. De plus, il est certain que peu importe l'ensemble de test choisi, celui-ci ne pourra pas être représentatif de ce qui est observable sur le Soleil durant un cycle solaire. Par conséquent tout banc de test serait voué à être biaisé.

Pour éviter cela, il faudrait que l'ensemble de test soit très grand, ce qui rendrait d'autant plus petit l'ensemble d'entraînement, alors que nous avons déjà peu de données pour assurer un entraînement correct. Afin de résoudre – au moins partiellement – tous les problèmes mentionnés ci-dessus, et pour faire en sorte d'avoir des ensembles d'entraînement et de test assez grands pour assurer à la fois un bon entraînement et une évaluation non-biaisée, nous avons recours ici à une méthode d'évaluation croisée imbriquée.

Afin d'expliquer ce qu'est l'évaluation croisée imbriquée, décrivons d'abord le concept de validation croisée. La validation croisée (*cross-validation*) « simple » consiste à entraîner plusieurs fois le réseau de neurones en faisant varier l'ensemble de validation. Pour cela, l'approche la plus souvent utilisée est de partager l'ensemble de nos données en $k + 1$ blocs distincts. Un de ces blocs est retenu pour l'évaluation finale, et n'est jamais utilisé pendant l'entraînement. Les k blocs restants sont utilisés tour à tour pour l'entraînement et la validation en effectuant des permutations, avec $k - 1$ blocs étant utilisés comme ensemble d'entraînement et 1 bloc comme ensemble de validation. Notons bien qu'à chaque permutation de nos ensembles d'entraînement et de validation nous recommençons l'entraînement de zéro. À la fin du processus d'apprentissage par validation croisée nous avons autant de modèles que de permutations effectuées, c'est-à-dire k . La validation croisée est généralement utilisée dans les *pipelines* de *machine learning* pour choisir de façon optimale les hyperparamètres du modèle, tout en conservant un ensemble d'entraînement aussi grand que possible. En effet, dans certains cas (comme le nôtre), le fait de valider un modèle avec deux ensembles de validation différents peut ensuite mener à des performances lors de l'évaluation très différentes, car les

ensembles de validation ne sont pas tous représentatifs des mêmes dynamiques, et on va donc avoir tendance à en privilégier certaines au détriment d'autres.

L'évaluation croisée imbriquée consiste à rajouter une couche d'évaluation croisée par dessus la validation croisée. Concrètement cela signifie que nous allons faire varier l'ensemble de test parmi l'ensemble de nos données, et utiliser les données restantes pour entraîner le modèle à l'aide d'une validation croisée simple. Cela crée deux boucles d'évaluation et de validation croisées, d'où le nom d'évaluation croisée imbriquée (*nested cross-evaluation*). Le principe de l'évaluation croisée imbriquée est schématisé en Figure 5.9. Le principal avantage de l'évaluation croisée imbriquée est que toutes nos données seront utilisées pour entraîner, valider mais aussi évaluer le modèle, permettant un entraînement complet, tout en conservant une évaluation exhaustive et juste. Son principal inconvénient est que cette démarche multiplie le nombre d'apprentissages nécessaires pour entraîner le modèle. Ici, $(k + 1) \cdot k$ entraînements sont nécessaires, ce qui rallonge d'autant la durée d'apprentissage.

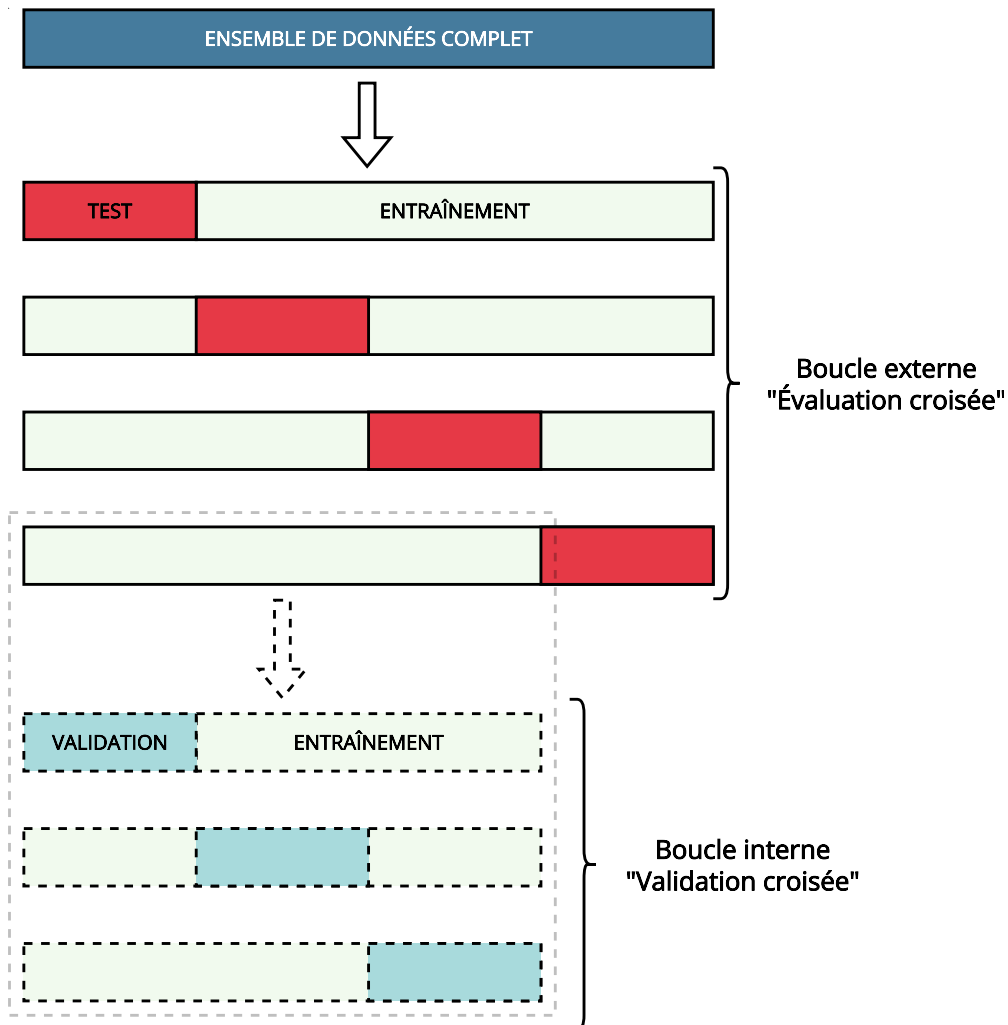


FIGURE 5.9 – Schéma de la méthode d'évaluation croisée imbriquée.

Afin d'implémenter une stratégie d'évaluation croisée imbriquée, il faut d'abord déterminer comment nos données seront partagées en blocs. Ce choix est important, car utiliser trop de blocs peut multiplier la durée d'apprentissage de façon

rédhibitoire. De plus, un trop grand nombre de blocs signifie que chaque bloc est individuellement de petite taille. Dans ce cas, nos ensembles d'évaluation et de validation risquent d'être trop petits pour être statistiquement significatifs. À l'inverse, utiliser trop peu de blocs signifie que chaque bloc est de taille relativement grande, et alors il est possible que ce soit notre ensemble d'entraînement qui soit de trop petite taille, limitant ainsi le potentiel d'apprentissage. Dans notre situation, notre jeu de données peut être naturellement divisé en 9 années calendaires. Nous choisissons des blocs correspondant à une année calendaire. Cela signifie que l'apprentissage complet de notre modèle nécessite $9 * 8 = 72$ entraînements différents.

Afin d'exploiter au mieux l'évaluation croisée imbriquée, il faut également déterminer une stratégie d'utilisation des différents modèles entraînés. Autrement dit, pour chacun des 9 ensembles d'évaluation (c'est-à-dire chacune des 9 années pour lesquelles on a des données), il faut déterminer comment utiliser les 8 modèles entraînés avec chacun des ensembles de validation. Comme chaque ensemble de validation représente une partie différente du cycle solaire, il est difficile de les comparer entre eux et d'utiliser une stratégie se basant sur des métriques de performance (autrement dit, il est difficile de comparer objectivement les résultats obtenus avec chaque ensemble de validation et conclure si un ensemble de validation donne de meilleurs résultats qu'un autre). C'est pourquoi nous choisissons de combiner le résultat de tous les entraînements pour en faire un modèle d'ensemble. Pour chaque jeu d'évaluation, notre prédiction sera une combinaison ensembliste des prédictions obtenues en ayant entraîné et validé le modèle avec chacun des blocs. Nous donnons plus de détails sur l'implémentation de l'aspect ensembliste de *SERENADE* en section 5.3.4.2.

Récapitulons notre démarche. En utilisant l'imagerie solaire, et plus particulièrement l'ensemble *SDO/AIA - ML-READY*, nous sommes confrontés à un double problème : nous avons peu de données, et nos données sont issues d'un cycle solaire plutôt calme. En combinant ces deux assertions, il apparaît à première vue que nous devons choisir entre avoir un ensemble d'entraînement assez grand et représentatif, tout en ne pouvant évaluer notre modèle correctement, ou bien conserver suffisamment de données pour évaluer notre modèle de façon plus juste, au risque de ne pas avoir assez de données pour entraîner correctement notre modèle. Afin de résoudre ce dilemme, nous avons recours à l'évaluation croisée imbriquée. Ainsi, nous ne divisons pas notre ensemble de données complet en sous-ensembles d'entraînement, de validation, et d'évaluation fixes, comme c'était le cas de les chapitres précédents. À la place, nous entraînons plusieurs fois notre modèle en faisant permuer les sous-ensembles d'entraînement, de validation, et d'évaluation, faisant en sorte que chacun de nos échantillons sera utilisé à la fois pour entraîner, valider et évaluer notre problème, ce qui résout en grande partie les problèmes évoqués ci-dessus. Afin d'exploiter au mieux cette approche, pour chaque sous-ensemble d'évaluation, nous combinons les résultats des apprentissages obtenus avec chaque couple de sous-ensembles d'entraînement et de validation pour obtenir une prédiction ensembliste.

Le seul inconvénient de cette approche est majeur, car il multiplie notre

temps d'apprentissage par 72. Cela peut paraître contradictoire avec le fait que dans les sous-sections précédentes, nous tentions au contraire de réduire ce temps d'apprentissage au maximum. Cependant cela n'est pas le cas, car les motivations étaient alors différentes. Ici, nous sommes obligés d'avoir recours à cette méthode si nous voulons obtenir une évaluation de notre modèle qui soit à la fois exhaustive et le moins biaisée possible. Les compromis effectués dans les sous-sections précédentes afin de réduire le temps d'apprentissage vont certainement nous mener à des performances limitées par rapport à ce qu'elles pourraient atteindre, mais cela nous permet justement de pouvoir mener à bien le processus de développement itératif de notre modèle, tout en conservant des temps d'apprentissage raisonnables malgré l'évaluation croisée imbriquée. Notons bien que cela serait inutile d'avoir un modèle potentiellement « meilleur » s'il nous était ensuite impossible de l'évaluer correctement.

Enfin, nous faisons remarquer qu'utiliser l'évaluation croisée nous permettra d'évaluer notre modèle sur l'ensemble de la période 2010 – 2018 couverte par le jeu de données *SDO/AIA – ML-READY*, mais également de procéder (avec précaution) à l'évaluation de la performance du modèle pour chacune des 9 années à notre disposition, ce qui nous permettra de vérifier si notre modèle présente (ou non) des performances homogènes tout au long du cycle solaire.

5.3.4.2 Implémentation pratique

Nous détaillons ici l'ensemble de nos choix concernant l'implémentation de l'apprentissage du modèle, allant des algorithmes d'optimisation au matériel utilisé.

Combinaison ensembliste des prédictions Comme nous venons de l'expliquer dans la section précédente, recourir à l'évaluation croisée imbriquée nous amène à devoir combiner nos prédictions obtenues avec chacun des modèles entraînés en un modèle ensembliste. Pour aider le lecteur à mieux comprendre ce besoin nous donnons un exemple. Considérons que nous utilisons les données de l'année 2018 comme ensemble d'évaluation. Alors nous avons à notre disposition les données des années 2010 – 2017 pour entraîner et valider notre modèle. Nous pouvons donc entraîner notre modèle 8 fois, une fois avec chacune des années disponibles, utilisées en tant qu'ensemble de validation, et les 7 autres comme ensemble d'entraînement (par exemple, 2017 est utilisée pour valider, et 2010 – 2016 pour entraîner ; puis nous utilisons 2016 pour valider et 2010 – 2015 et 2017 pour entraîner, etc.). Ainsi, nous obtenons 8 modèles différents, que nous pouvons évaluer sur l'année 2018. La question est alors de savoir comment utiliser ces 8 modèles conjointement. Nous avons décidé de les combiner, pour en faire un modèle ensembliste.

Comme chacune de nos prédictions est produite sous la forme d'une distribution normale, et que chaque entraînement se fait indépendamment des autres, on peut utiliser la propriété de stabilité par additivité de la loi normale¹, selon laquelle

1. En réalité il n'est pas certain que nos prédictions soient réellement strictement indépendantes. Néanmoins, comme il nous est très difficile de quantifier leur covariance, nous faisons cette

si $\forall i \in [1, n]$, les variables aléatoires X_i sont indépendantes et suivent une loi normale $\mathcal{N}(\mu_i, \sigma_i^2)$ alors la somme $X_1 + X_2 + \dots + X_n$ suit une loi normale $\mathcal{N}(\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2)$. Immédiatement, il vient que la moyenne des X_i , $\frac{X_1 + X_2 + \dots + X_n}{n}$ suit une loi normale $\mathcal{N}(\frac{\mu_1 + \dots + \mu_n}{n}, \frac{\sigma_1^2 + \dots + \sigma_n^2}{n^2})$. Ainsi, à l'aide de nos 8 prédictions gaussiennes, nous pouvons obtenir une unique prédiction ensembliste, également gaussienne.

Fonction de perte Comme pour les modèles probabilistes utilisant les paramètres du vent solaire en L1 développés dans le chapitre 3, nous choisissons comme *loss function* une version modifiée de la log-vraisemblance négative gaussienne, ou **MGNLL**, définie par :

$$\begin{aligned} \text{MGNLL}(\mu, \sigma, y_{\text{vrai}}) &= \text{GNLL}(\mu, \sigma, y_{\text{vrai}}) + \text{MSE}(\mu, y_{\text{vrai}}) \\ &= \frac{1}{2} * (\log(\sigma^2) + \frac{(\mu - y_{\text{vrai}})^2}{\sigma^2}) + (\mu - y_{\text{vrai}})^2, \end{aligned} \quad (5.1)$$

où μ et σ sont la moyenne et l'écart-type de la distribution gaussienne prédite par *SERENADE* et y_{vrai} est la vraie valeur observée de Kp .

Critère de convergence Usuellement, le critère de convergence indiquant que le modèle a été suffisamment entraîné est la valeur de la fonction de perte calculée sur l'ensemble de validation. Or ici nous avons remarqué qu'utiliser la **MGNLL** comme critère de convergence nous amenait considérer un modèle sous-entraîné et prédisant uniquement des faibles valeurs de Kp , sans doute à cause de la distribution des valeurs de Kp très concentrée dans les faibles valeurs – surtout au cours du cycle solaire 24, comme nous l'avons déjà vu (cf. Figure 5.8. Ainsi nous avons cherché à construire un critère de convergence qui minimise l'écart moyen entre les valeurs prédites et les observations, mais qui pénalise des prédictions trop concentrées autour des valeurs faibles. En notant $\text{Crit}_{\text{conv}}$ le critère de convergence, \mathbf{y}_{vrai} l'ensemble des observations dans l'ensemble de données considéré, $\boldsymbol{\mu}$ et $\boldsymbol{\sigma}$ l'ensemble des moyennes et écarts-types des distributions gaussiennes prédites par *SERENADE* pour ce même ensemble de données, on a :

$$\begin{aligned} \text{Crit}_{\text{conv}}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{y}_{\text{vrai}}) &= \overline{\text{MSE}}(\boldsymbol{\mu}, \mathbf{y}_{\text{vrai}}) \\ &+ \overline{\text{CRPS}}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{y}_{\text{vrai}}) + (|\text{QQ}_{\text{slope}}(\boldsymbol{\mu}, \mathbf{y}_{\text{vrai}}) - 1| + 1) \\ &- \text{AUROC}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{y}_{\text{vrai}}) - \frac{\text{Sn}(\boldsymbol{\mu})}{\text{Sn}(\mathbf{y}_{\text{vrai}})}, \end{aligned} \quad (5.2)$$

où :

- $\overline{\text{MSE}}$ (resp. $\overline{\text{CRPS}}$) désigne la valeur moyenne de **MSE** (resp. **CRPS**) calculée sur l'ensemble des données ;
- QQ_{slope} désigne la pente du diagramme Quantile-Quantile², qui compare les distributions de \mathbf{y}_{vrai} et $\boldsymbol{\mu}$ (sa valeur idéale est de 1) ;

hypothèse simplificatrice.

2. Ou plus exactement, la pente de sa régression linéaire.

- **AUROC** est l'aire sous la courbe ROC lorsqu'on choisit un seuil de classification binaire égal à 4^- ;
- **Sn** désigne la statistique Sn, c'est-à-dire un estimateur robuste de l'écart-type d'une série.

Ce critère de convergence peut s'expliquer assez simplement en le décomposant. La composante **MSE** favorise les modèles dont la moyenne de la prédiction proche de l'observation. La composante **CRPS** en fait de même mais en prenant en compte le caractère probabiliste de la prédiction. La composante $(|Q_{\text{slope}} - 1| + 1)$ pénalise les modèles dont la distribution des quantiles est éloignée de la distribution des quantiles des vraies observations. Cela favorise donc les modèles qui prennent plus de « risques » (en opposition à un modèle qui aurait un comportement très conservateur en prédisant tout le temps la valeur moyenne de $Kp_{\text{max},24\text{h}}$ par exemple). La composante $\frac{\text{Sn}(\boldsymbol{\mu})}{\text{Sn}(\mathbf{y}_{\text{vrai}})}$ a un objectif similaire, en favorisant les modèles dont l'ensemble des valeurs moyennes des prédictions a un écart-type proche de l'écart-type de l'ensemble des observations. Enfin, la composante **AUROC** favorise les modèles qui sont plus compétents lorsqu'on transforme les prédictions probabilistes en probabilités de dépassement d'un seuil $Kp_{\text{seuil}} = 4^-$ (voir la section 5.4 pour plus d'informations).

Ce critère a été trouvé de façon empirique et itérative, après de multiples essais, en ayant toujours pour objectif de favoriser l'apprentissage d'un modèle qui soit à la fois « expressif » (*i. e.* qui s'éloigne d'un comportement prudent similaire au modèle de climatologie, cf. section 5.4.1.1), mais commettant le moins d'erreurs possible. Comme les valeurs de toutes les composantes sont dans la pratique comprises dans un intervalle $\approx 0,5 - 2$, nous ne leur appliquons pas de coefficient multiplicatif dédié. Il serait sans doute possible d'améliorer encore plus ce critère en testant d'autres métriques et en utilisant des coefficients relatifs, mais à l'heure actuelle ce critère nous convient.

Optimisation Ici nous utilisons toujours l'algorithme d'optimisation ADAM (KINGMA et BA 2017), avec la valeur des paramètres $\boldsymbol{\beta} = (\beta_1, \beta_2) = (0.9, 0.999)$, un *learning rate* $lr = 10^{-5}$ et un coefficient de *weight decay* fixé à $wd = 10^{-4}$.

Epochs Chaque entraînement (avec chacun des sous-ensembles d'entraînement, de validation et d'évaluation) est mené pendant un maximum de $n_{\text{epochs}} = 50$ *epochs*.

Mini-batch et parallélisation La taille des *mini-batch* d'apprentissage utilisée est $N = 128$. Afin d'accélérer le chargement des données, celui-ci est effectué sur CPU, à l'aide de 4 *threads* parallèles, avant que les données soient envoyées sur GPU.

Dropout La valeur du coefficient de dropout appliqué systématiquement entre chaque couche du réseau est de $d = 0.3$.

Environnement Le réseau de neurones a été développé et codé notamment à l'aide des libraires PyTorch (v. 1.9), torchvision (v. 0.10) pour langage Python (v. 3.9).

Encodage des flottants Afin de réduire encore la durée de l'apprentissage (d'un facteur 2 environ), nous encodons ici nos données (c'est-à-dire nos *inputs*, *outputs* et les couches du réseau de neurones qui peuvent l'être) sous la forme de flottants à demi-précision (des flottants encodés sur 16 bits au lieu de 32, voire 64) à l'aide de l'outil de calcul en précision mixte sur GPU de PyTorch. Intuitivement on pourrait s'attendre à ce que l'utilisation d'entiers encodés en `float16` au lieu de `float32` dégrade notre précision. Toutefois des études (MICIKEVICIUS et al. 2018) et des essais menés par des développeurs chez Nvidia et Facebook (HUANG et al. 2020) ont montré empiriquement que ce n'était pas nécessairement le cas. Au contraire, il apparaît que pour certains problèmes de classification, les performances obtenues avec de la précision mixte (c'est-à-dire en combinant demi-précision et précision entière) pouvaient parfois être supérieures, tout en permettant un apprentissage plus rapide, et occupant moins de mémoire sur le GPU. De notre côté nous n'avons pas observé de dégradation en utilisant l'outil de précision mixte automatique, par conséquent nous nous en servons ici.

Matériel Notre réseau a été entraîné sur un serveur de calcul possédant 189 Go de RAM, deux CPU Intel® Xeon® E5-2699 v4 @ 2.20GHz (pour un total de 44 cœurs physiques et 88 cœurs logiques), et un GPU Nvidia Titan V (architecture Volta) possédant 12 Go de mémoire dédiée.

Temps de calcul En combinant l'ensemble des choix mentionnés ci-dessus, et dans les sous-sections précédentes, nous parvenons à une durée d'entraînement pour chaque permutation de l'évaluation croisée imbriquée d'une durée d'environ 18 min, ce qui nous donne un temps d'apprentissage total d'environ 21 – 22 h.

5.4 Méthodologie de l'évaluation

L'architecture et les idées derrière le modèle [SERENADE](#) exposées, nous détaillons ici les modèles de référence utilisés pour évaluer notre modèle, avant de présenter les métriques et mesures d'évaluation utilisées.

5.4.1 Les modèles de référence

Avant de présenter et évaluer les résultats obtenus avec [SERENADE](#), il convient de présenter les modèles de référence (*baselines*) avec lesquels nous allons les comparer. Comme il n'existe pas à l'heure actuelle, à notre connaissance, d'autre modèle de prédiction de $Kp_{\max,24h}$ plusieurs jours en avance, à partir d'imagerie solaire, nous devons nous comparer à des modèles empiriques.

Notons que le *Space Weather Prediction Center* de la NOAA propose sur son site internet³ des prédictions de Kp , avec des valeurs tri-horaires jusqu'à trois jours en avance, et des valeurs maximales journalières jusqu'à 27 jours en avance. Toutefois, ils ne fournissent pas d'information concernant la méthodologie utilisée pour obtenir de telles prédictions, sauf pour la prédiction jusqu'à 27 jours, pour laquelle il est mentionné dans le rapport (NOAA 2012) : « *These 27-day forecasts are based primarily on the persistence of patterns of solar and/or geophysical activity from one 27-day solar rotation to the next* » (« ces prédictions à 27 jours s'appuient en premier lieu sur la persistance de motifs de l'activité solaire et/ou géophysique entre deux rotations solaires de 27 jours »). On peut donc légitimement s'attendre à ce que leur modèle soit simplement une version dérivée du modèle de récurrence à 27 jours présenté ci-dessous (sous-section 5.4.1.3).

5.4.1.1 Climatologie

Le premier modèle de référence, fréquemment utilisé dans les milieux de la météorologie et l'océanographie, est le modèle de climatologie (cf. par exemple le manuel de l'Organisation Mondiale de la Météorologie, WMO 2019). Il s'agit d'un modèle utilisant la distribution passée connue d'une série temporelle comme prédiction pour un instant futur. Cette approche, bien que peu compétente (peu importe l'instant on va toujours prédire la même chose), n'est pas dénuée d'intérêt. En effet, comme nous le montre la distribution des valeurs de Kp (Figure 5.8), celles-ci sont très concentrées autour de valeurs faibles. Cela signifie qu'on est très souvent autour de la valeur moyenne de l'indice, et donc une prédiction climatologique va être très souvent proche de la réalité observée.

Il existe différentes façons de construire un modèle climatologique. Ici, nous adoptons une méthode simple, consistant à prédire une distribution normale, dont la moyenne et l'écart-type sont égaux à la moyenne et l'écart-type calculés sur la période d'entraînement au sens large (*i. e.* incluant entraînement et validation). En notant $\overline{Kp_{\max,24h}}^{\text{climat}}$ la prédiction obtenue avec le modèle de climatologie, et en utilisant les notations antérieures, on a donc :

$$\overline{Kp_{\max,24h}}^{\text{climat}}(t + \tau) = \mathcal{N}(\mu_{Kp_{\max,24h}}^{\text{train}}, \sigma_{Kp_{\max,24h}}^{\text{train}}), \quad (5.3)$$

où $\mu_{Kp_{\max,24h}}^{\text{train}}$ désigne la moyenne et $\sigma_{Kp_{\max,24h}}^{\text{train}}$ désigne l'écart-type de $Kp_{\max,24h}$, calculés sur l'ensemble d'entraînement (au sens large).

5.4.1.2 Persistance

Un autre modèle de référence est le modèle dit de « persistance » (parfois appelé modèle de prédiction naïf). Il s'agit tout simplement d'utiliser la dernière valeur connue de $Kp_{\max,24h}$ comme prédiction pour l'instant $t + \tau$ futur souhaité. En notant $\overline{Kp_{\max,24h}}^{\text{persis}}$ la prédiction obtenue avec le modèle de persistance on

3. <https://www.swpc.noaa.gov/>

obtient dans le cas déterministe :

$$\overline{Kp_{\max,24h}^{\text{persis}}}(t + \tau) = Kp_{\max,24h}(t). \quad (5.4)$$

La Figure 5.10 montre les valeurs d'auto-corrélation linéaire de $Kp_{\max,24h}$ pour différents décalages temporels allant de 0 à 60 jours. On observe que celle-ci chute fortement, et atteint des valeurs proches de 0.1 dès que $\tau \geq 3$ jours. Ainsi ce modèle de référence, qui est difficile à battre lorsqu'on effectue des prévisions à court terme (de l'ordre de quelques heures) devrait devenir moins performant au fur et à mesure que l'horizon de prédiction grandit, jusqu'à une certaine limite (voir ci-dessous).

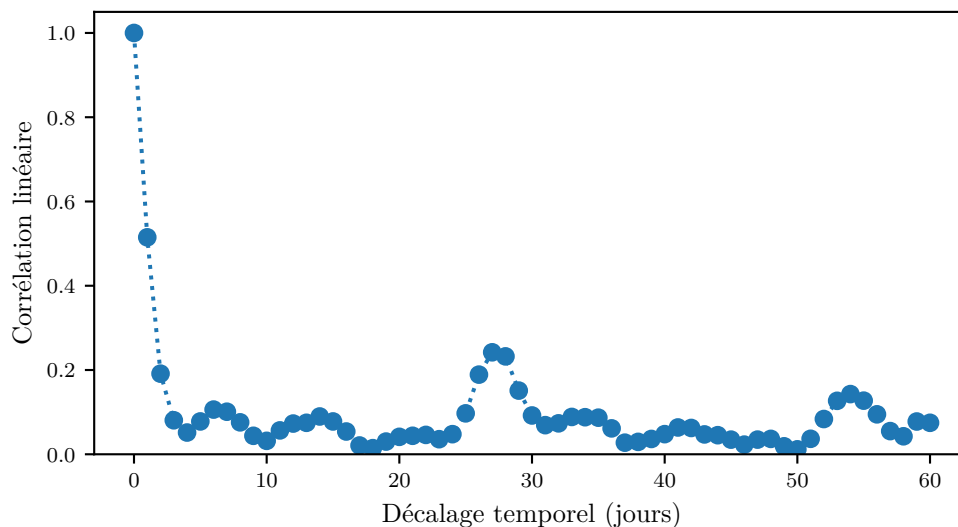


FIGURE 5.10 – Auto-corrélation linéaire de $Kp_{\max,24h}$ pour le cycle solaire 24.

Toutefois SERENADE est un modèle probabiliste, il convient donc de le comparer à d'autres modèles probabilistes. Afin de rendre le modèle de persistance probabiliste, une approche simple consiste à calculer l'écart-type de $Kp_{\max,24h}$ sur l'ensemble d'entraînement (au sens large) et l'utiliser comme écart-type d'une distribution gaussienne pour chacune des prédictions faites sur l'ensemble de test (de la même façon qu'avec le modèle de climatologie). Ainsi, dans le cas probabiliste, on obtient :

$$\overline{Kp_{\max,24h}^{\text{persis}}}(t + \tau) = \mathcal{N}(Kp_{\max,24h}(t), \sigma_{Kp_{\max,24h}}^{\text{train}}), \quad (5.5)$$

où $\mathcal{G}(\mu, \sigma)$ désigne une distribution normale de moyenne μ et d'écart-type σ et $\sigma_{Kp_{\max,24h}}^{\text{train}}$ désigne l'écart-type de $Kp_{\max,24h}$ calculé sur l'ensemble d'entraînement (au sens large).

5.4.1.3 Récurrence à 27 jours

La Figure 5.10 nous montre également que l'auto-corrélation linéaire de $Kp_{\max,24h}$ grandit à nouveau pour atteindre une valeur de 0,32 lorsque le décalage

temporel est de 27 jours. Cela est surtout dû à la persistance de certaines structures sur le Soleil – comme les trous coronaux – entre plusieurs rotations de Carrington, qui ont une durée moyenne de 27 jours. Ainsi le modèle de récurrence à 27 jours est une variante du modèle de persistance dans lequel on va exploiter cette corrélation plus importante pour ces décalages temporels. En notant $\overline{Kp_{\max,24\text{h}}^{\text{rec}-27}}$ la prédiction obtenue avec le modèle de récurrence à 27 jours, on obtient dans le cas déterministe :

$$\overline{Kp_{\max,24\text{h}}^{\text{rec}-27}}(t + \tau) = Kp_{\max,24\text{h}}(t + \tau - 27\text{j}). \quad (5.6)$$

Comme nous l'avons déjà évoqué, il semblerait que les prévisions à moyen terme du SWPC de la NOAA soient dérivées de ce modèle de référence. De plus, des études comme SHPRITS et al. 2019 montrent bien qu'il s'agit actuellement du meilleur modèle pour la prédiction de Kp pour des horizons supérieurs à 24h, malgré l'apparente simplicité du modèle. On peut donc légitimement s'attendre à ce qu'il soit particulièrement dur à battre.

Ici aussi il s'agit d'un modèle déterministe. Nous le rendons probabiliste avec la même méthode que celle employée pour le modèle de persistance, on obtient donc (avec la même notation que précédemment) :

$$\overline{Kp_{\max,24\text{h}}^{\text{rec}-27}}(t + \tau) = \mathcal{N}(Kp_{\max,24\text{h}}(t + \tau - 27\text{j}), \sigma_{Kp_{\max,24\text{h}}}^{\text{train}}). \quad (5.7)$$

5.4.2 Le banc d'évaluation

Nous décrivons ici les métriques avec lesquelles nous allons évaluer nos modèles pour la prédiction de $Kp_{\max,24\text{h}}$. Comme nous l'avons déjà fait dans les chapitres précédents, nous évaluerons nos modèles également sous la forme d'un problème de classification binaire, dont le seuil sera indiqué et justifié.

5.4.2.1 Banc d'évaluation pour le problème de régression

Pour évaluer nos prédictions nous utiliserons les métriques suivantes, que nous avons déjà utilisées dans les chapitres précédents :

- la racine de l'erreur quadratique moyenne ou *Root Mean Squared Error* (RMSE) ;
- le coefficient de corrélation linéaire R ;
- le *Temporal Distortion Mix* (TDM) ;
- le *Continuous Ranked Probability Score* (CRPS).

Nous évaluerons notre modèle sur toute la période disponible (ce qui est rendu possible grâce à l'utilisation de l'évaluation croisée-imbriquée), mais aussi en séparant les périodes, afin d'estimer la performance prédictive de notre modèle en fonction de la phase du cycle solaire.

5.4.2.2 Banc d'évaluation pour le problème de classification

Afin de transformer nos prédictions numériques de $Kp_{\max,24h}$ en un problème de classification binaire, nous effectuons un seuillage simple. Pour cela nous prenons un seuil $Kp_{\text{seuil}} = 4^-$, qui correspond au seuil à partir duquel Kp est considéré comme « actif » selon des centres de météorologie spatiale opérationnelle comme le SWPC⁴ ou le GFZ Potsdam⁵. Nous avons décidé de ne pas prendre une valeur de $Kp_{\text{seuil}} = 5^-$ qui correspondrait uniquement aux périodes d'orage géomagnétique car, comme nous l'avons vu en Figure 5.8, au cours du cycle solaire 24 il n'y a que peu d'échantillons correspondant à cette catégorie (moins de 9,4 %, contre $\approx 24,7\%$ en prenant $Kp_{\text{seuil}} = 4^-$). Notons enfin qu'une valeur de $Kp = 4^-$ est souvent numériquement convertie en $Kp = 3,7$ pour les applications numériques, comme par exemple dans la base OMNIweb.

Ici comme nos prédictions sont probabilistes, nous pouvons également avoir une prédiction de classe probabiliste, correspondant donc à la probabilité que la valeur de $Kp_{\max,24h}$ prédite dépasse Kp_{seuil} , ce qui est facilement calculable comme nos prédictions suivent une distribution gaussienne. Pour évaluer notre modèle nous utiliserons les métriques suivantes :

- la précision ;
- le rappel ;
- le F_{score} ;
- le taux de fausses alertes ou *False Alarm Rate* (FAR) ;
- le *Threat score* (TS) ;
- le *Brier Score* (BS) ;
- le *Heidke Skill Score* (HSS) ;
- la courbe ROC et l'aire sous la courbe ROC (AUROC).

Nous avons déjà utilisé ou mentionné toutes ces métriques, à l'exception du *Brier Score* (BS). Le BS est une métrique permettant d'évaluer des prédictions de classe binaires probabilistes. Notons p_i une prédiction probabiliste ($p_i \in [0, 1]$) et C_i la véritable classe associée ($C_i \in \{0, 1\}$). Alors, si on dispose de N échantillons, on a :

$$\text{BS} = \frac{1}{N} \sum_{i=1}^N (p_i - C_i)^2. \quad (5.8)$$

Le BS est donc la MSE entre les probabilités prédites d'être dans la classe 1 et le véritable label rattaché à une observation. C'est une métrique prenant des valeurs entre 0 et 1, et plus sa valeur est faible, meilleur est notre modèle.

5.5 Bilan du chapitre

L'exploitation d'imagerie solaire au sein de modèles s'appuyant sur l'apprentissage automatique n'est pas un concept nouveau. Néanmoins, à l'heure actuelle

4. <https://www.swpc.noaa.gov/products/planetary-k-index>

5. <https://isdg.gfz-potsdam.de/kp-index/>

aucune étude n'a essayé de connecter directement l'imagerie solaire à l'activité géomagnétique. Nous proposons ici un nouveau modèle, baptisé **SERENADE**, construit dans l'objectif de vérifier s'il est possible de prévoir la géoeffectivité de l'activité solaire (représentée dans un premier temps par l'indice Kp) à partir d'imagerie **EUV** seule. Notre modèle, à l'architecture modulaire combinant plusieurs types de réseaux de neurones, a été construit en faisant des compromis entre sa complexité et la nécessité de le rendre entraînable en un temps raisonnable avec les machines dont nous disposons. Nous avons également conçu un banc d'évaluation complet pour comparer notre modèle à des modèles empiriques de référence de la façon la plus complète possible. Dans le chapitre suivant nous présentons et analysons les performances de notre modèle.

Chapitre 6

Prédire la géoefficacité d'événements solaires à partir d'images du Soleil : premiers résultats

And no one showed us to the land
And no one knows the where's or why's
But something stirs and something tries
And starts to climb towards the light

(extrait de Echoes, Pink Floyd)

Sommaire

Préambule	206
6.1 Analyse des résultats obtenus sur l'ensemble de la période 2010 – 2018	206
6.1.1 Résultats pour le problème de régression	208
6.1.2 Résultats pour le problème de classification	211
6.1.3 Bilan de l'évaluation de SERENADE sur l'ensemble de la période 2010 – 2018	215
6.2 Analyse des performances selon la phase du cycle solaire	215
6.3 Prédiction de l'indice $Ca_{\max,24h}$	222
6.3.1 Modifications du <i>pipeline</i>	222
6.3.2 Évaluation des performances de SERENADE- Ca	223
6.4 Estimation de l'influence de la période utilisée sur les performances	228
6.5 Bilan du chapitre	230

Préambule

Dans le chapitre précédent nous avons présenté l'architecture du modèle **SERENADE**. Dans ce chapitre nous montrons et analysons les résultats obtenus avec cette méthode. Tout d'abord nous étudions successivement les performances obtenues pour les problèmes de régression puis de classification en utilisant l'ensemble des données à notre disposition pour l'évaluation. Ensuite nous décomposons ces résultats et les étudions par le prisme du cycle solaire. Enfin, nous présentons une variante de **SERENADE** destinée à la prédiction de Ca et nous en analysons les performances. Les prévisions fournies par **SERENADE** pour la période 2010 – 2018 sont disponibles publiquement sur [Zenodo](#) (BERNOUX et al. 2022b).

6.1 Analyse des résultats obtenus sur l'ensemble de la période 2010 – 2018

D'emblée, visualisons des exemples de prédictions obtenues avec **SERENADE**. La Figure 6.1 nous montre les prédictions obtenues avec **SERENADE** 2 jours et 5 jours en avance pour deux périodes : de septembre à fin décembre 2012 (fin de la phase ascendante) et de septembre à fin décembre 2015 (fin de la phase ascendante). Dans ces exemples on voit bien que **SERENADE** a du mal à rendre compte de la dynamique de $Kp_{\max,24h}$ en 2012, bien que – et c'est une satisfaction – le modèle propose également peu de faux positifs. D'un autre côté on constate que **SERENADE** parvient à reproduire une plus grande partie de la dynamique en 2015 et que les valeurs observées sont effectivement le plus souvent comprises dans l'intervalle de confiance à 95 %. Notons à ce propos que sur l'ensemble d'évaluation complet, 89 % des valeurs observées se situent dans l'intervalle de confiance à 95 % des prédictions. Ce résultat, déjà plutôt satisfaisant, pourrait sans doute être amélioré par calibration à posteriori.

Cette figure suggère que **SERENADE** a effectivement été capable d'apprendre

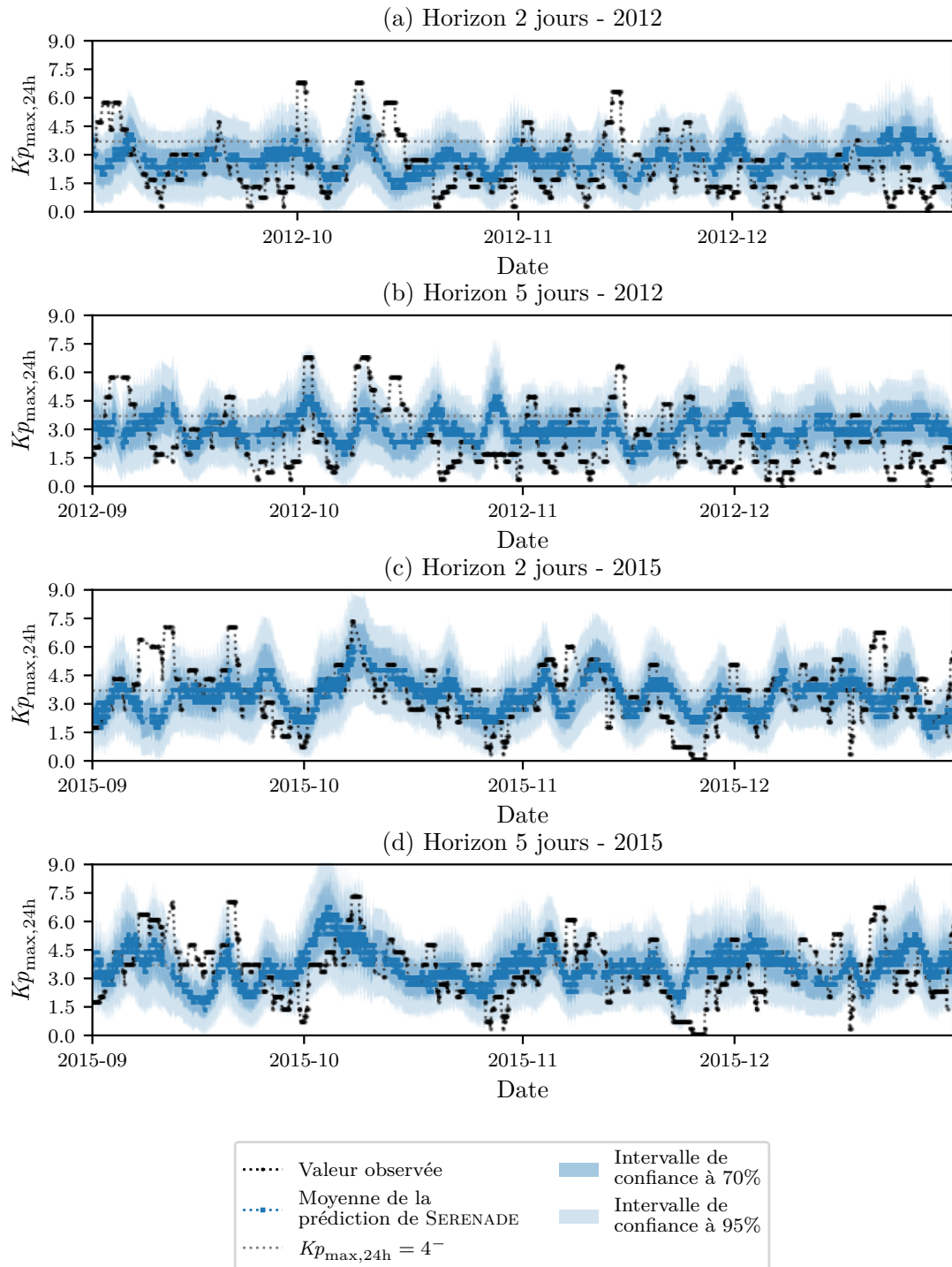


FIGURE 6.1 – Exemples de prédictions obtenues avec SERENADE 2 jours et 5 jours en avance pour deux périodes : de septembre à fin décembre 2012 et de septembre à fin décembre 2015.

au moins partiellement la dynamique de $Kp_{\max,24h}$. Néanmoins nous ne pouvons évidemment pas nous appuyer uniquement sur cette figure pour affirmer cela : un banc d'évaluation complet est nécessaire.

Nous étudions ici les performances du modèle **SERENADE** sur l'intégralité de la période couverte par les données **SDO/AIA – ML-READY** ($\approx 2010 - 2018$), ce qui est rendu possible par l'utilisation de la méthode d'évaluation croisée imbriquée vue dans le chapitre précédent. Nous commençons par analyser les résultats du point de vue du problème de régression probabiliste, puis du point de vue du problème de classification binaire.

6.1.1 Résultats pour le problème de régression

Pour rappel nous avons entraîné six modèles, correspondant à des horizons de prédiction $\tau = 2, 3, 4, 5, 6$ et 7 jours. La Figure 6.2 nous donne les valeurs de **RMSE**, **CRPS**, corrélation linéaire et **TDM** calculés sur l'ensemble de test complet, mais également uniquement pour les instants où $Kp_{\max,24h} \geq 4^-$ (que nous appellerons périodes actives).

La **RMSE**, la corrélation linéaire et le **TDM** sont calculés en utilisant uniquement la moyenne de chaque prédiction gaussienne, tandis que, par définition, le **CRPS** prend aussi en compte l'écart-type prédit.

On observe que la **RMSE** prend des valeurs comprises entre 1,35 et 1,48 nT, c'est-à-dire légèrement plus que l'écart-type de $Kp_{\max,24h}$ sur l'ensemble de la période qui vaut 1,34. Toutefois, le **CRPS** prend des valeurs comprises entre 0,76 et 0,84 nT (contre des valeurs de **MAE** comprises entre 1,07 et 1,51 nT), ce qui suggère que le fait de rendre nos prédictions probabilistes en améliore la compétence (d'environ 28 % systématiquement). On remarque également que la corrélation linéaire prend des valeurs assez faibles en absolu, inférieures à 0,3. En revanche, le **TDM** prend des valeurs proches de 0, ce qui suggère l'absence de décalages temporels prononcés entre les valeurs prédites et les valeurs observées. Les légères variations de **TDM** en fonction de l'horizon et du sous-ensemble considéré sont difficiles à interpréter.

D'une manière générale, et sans grande surprise, les valeurs des métriques (hors **TDM**) se dégradent avec l'augmentation de l'horizon temporel. En revanche, cette dégradation reste assez modérée pour la **RMSE** et le **CRPS**, ce qui est cohérent avec la physique du problème. En effet, comme nous l'avons déjà vu, les différentes composantes du vent solaire se propageant à des vitesses différentes, le temps d'arrivée sur Terre de perturbations détectées au niveau du Soleil peut être plus ou moins long. On voit, de plus, que les métriques sont dégradées lorsqu'on se restreint aux périodes actives uniquement, ce qui n'est également pas surprenant.

Toutefois, ces valeurs sont à mettre en perspective avec celles obtenues avec les modèles de référence. Pour ce faire, nous avons calculé les *Skill Scores* relatifs à (a) la **RMSE**, (b) le **CRPS** et (c) la corrélation linéaire obtenues avec **SERENADE** par rapport à chacun des trois modèles de référence. Un *Skill Score* est une mesure de compétence d'un modèle par rapport à un autre. Il s'obtient par la formule :

$$Skill\ Score = \frac{\text{Score du modèle à tester} - \text{Score du modèle de référence}}{\text{Score idéal} - \text{Score du modèle de référence}}. \quad (6.1)$$

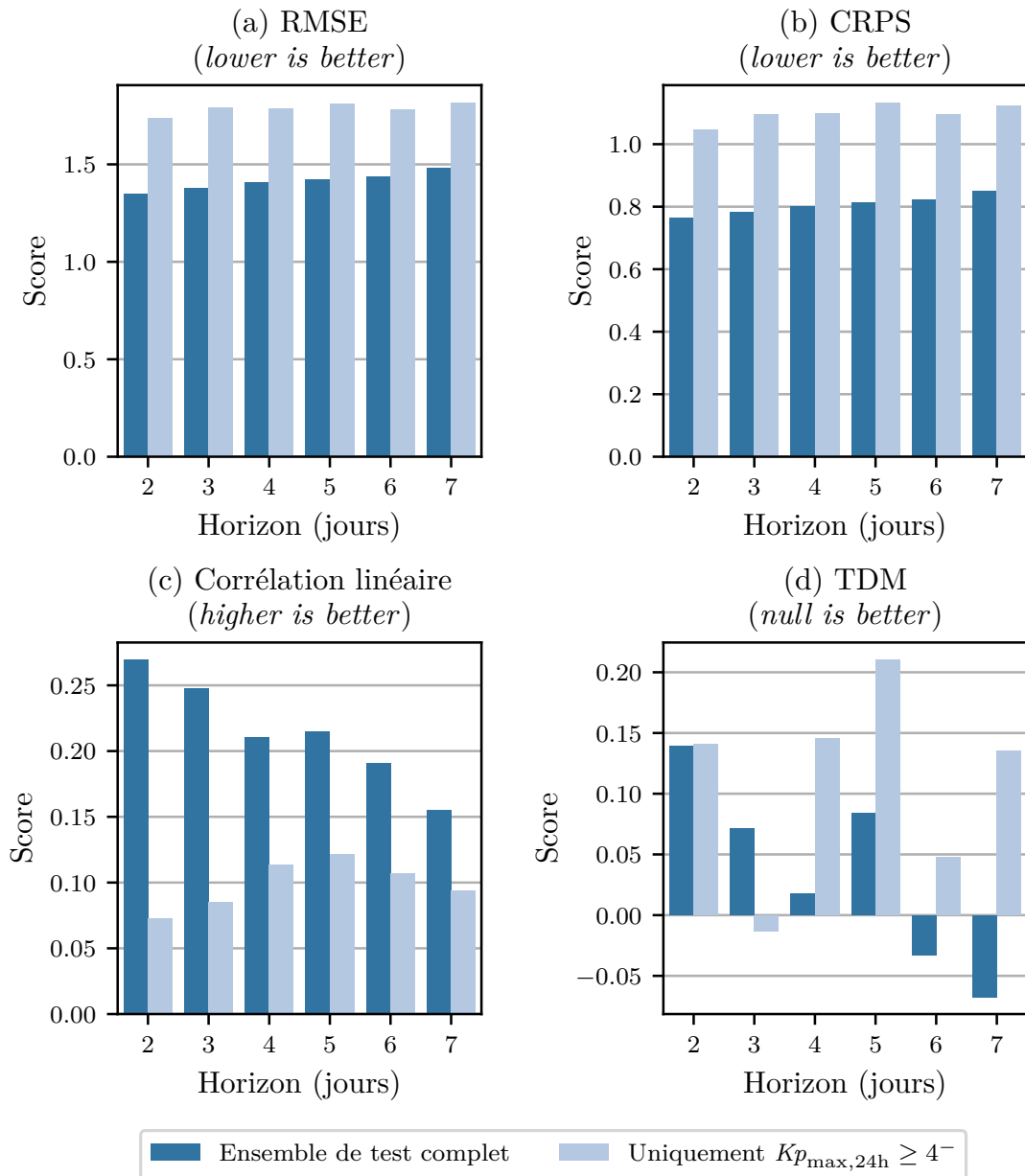


FIGURE 6.2 – Évaluation de **SERENADE** pour six horizons de prédiction différents à l'aide de métriques dédiées au problème de régression.

Il s'agit d'une valeur adimensionnelle comprise entre $-\infty$ et 1. Si le *Skill Score* est négatif, alors le modèle à tester (ici **SERENADE**) est moins bon que le modèle de référence. À l'inverse, si le *Skill Score* est positif, alors **SERENADE** est meilleur que le modèle de référence. Un *Skill Score* nul indique une absence de compétence supplémentaire de **SERENADE** par rapport à la référence, et une valeur de 1 indique que **SERENADE** serait parfait. Les résultats sont donnés en Figure 6.3.

On observe d'emblée que **SERENADE** est plus compétent que l'ensemble des modèles de référence pour les trois métriques sur les périodes actives. Cela montre que **SERENADE** est plus performant que les modèles de persistance, de climatologie mais aussi que le modèle de récurrence à 27 jours sur les périodes les plus

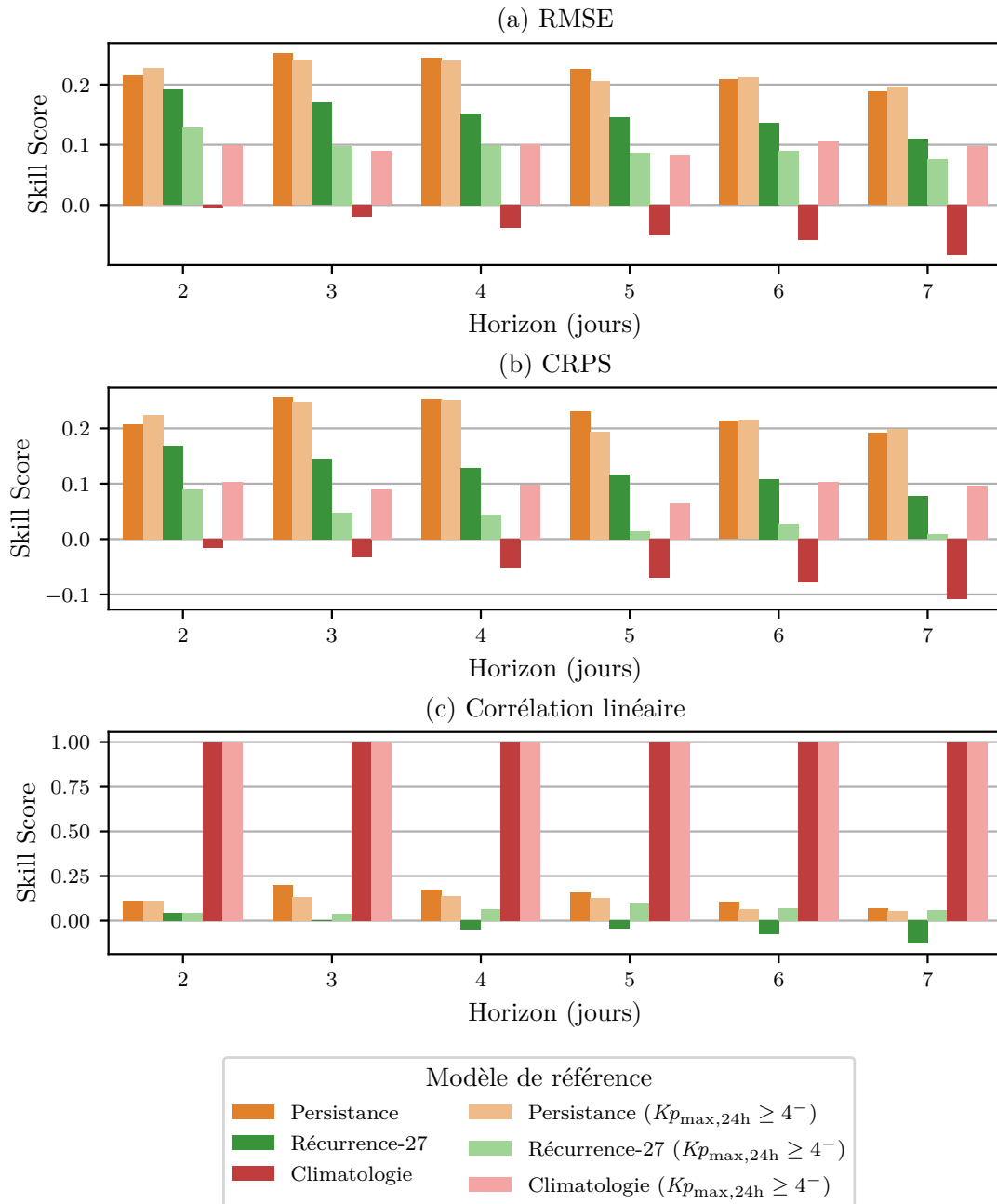


FIGURE 6.3 – Comparaison de SERENADE avec trois modèles de référence à l'aide de *Skill Scores* relatifs à trois métriques dédiées au problème de régression.

Note : Comme nous représentons les *Skill Scores*, plus une barre est haute (en direction des valeurs positives), plus SERENADE est compétent par rapport au modèle de référence utilisé.

dangereuses jusqu'à 7 jours en avance, ce qui est très satisfaisant.

En ce qui concerne les *Skill Scores* sur l'ensemble de test complet il faut distinguer les différents modèles de référence. Premièrement, SERENADE est toujours plus compétent que le modèle de persistance, quelle que soit la métrique considérée ou l'horizon de prédiction. Le modèle de persistance est également le modèle de

référence pour lequel **SERENADE** présente les plus hautes valeurs de *Skill Score*. En second lieu, il apparaît que **SERENADE** est plus compétent que le modèle de récurrence à 27 jours relativement à la **RMSE** et au **CRPS**, mais moins compétent en ce qui concerne la corrélation linéaire pour des horizons de prédiction supérieurs à 4 jours. Rappelons que le modèle de récurrence est considéré actuellement comme le meilleur modèle pour des prédictions de Kp avec un horizon plus grand que 2 jours. Le fait que **SERENADE** apparaisse ici comme étant globalement plus compétent que le modèle de récurrence, y compris sur les périodes actives, nous prouve à posteriori non seulement la pertinence de notre approche, mais aussi que cette première version de **SERENADE** pourrait mener à l'établissement d'un nouvel état de l'art.

Troisièmement, on remarque que le modèle de climatologie est plus compétent en termes de **RMSE** et de **CRPS** que **SERENADE** sur l'ensemble de test complet. Ce résultat est à mettre en perspective avec le fait, déjà discuté en Section 5.3.4 (et illustré par la Figure 5.8), que le cycle solaire 24 a été un cycle de très faible activité, et que la distribution des valeurs de Kp était donc très concentrée autour des très faibles valeurs de Kp . Il est ainsi peu étonnant que le modèle de climatologie produise de bons à très bons résultats pour les métriques de **RMSE** et de **CRPS** sur l'ensemble de la période. Toutefois, **SERENADE** reste bien plus compétent dès qu'on évalue les modèles uniquement sur les périodes actives, ce qui corrobore notre analyse. Enfin, **SERENADE** est infiniment plus compétent que le modèle de climatologie relativement à la corrélation linéaire. En effet, le modèle de climatologie produisant toujours la même valeur, il produit une valeur de corrélation linéaire nulle avec les observations de $Kp_{\max,24h}$. Cela nous rappelle à nouveau l'importance d'utiliser plusieurs métriques d'évaluation de nature différentes pour comparer des modèles divers.

6.1.2 Résultats pour le problème de classification

Afin d'avoir une meilleure compréhension de la performance de **SERENADE** à distinguer les périodes calmes des périodes dites actives (pour rappel, $Kp_{\max,24h} \geq 4^-$), nous évaluons également notre modèle à l'aide de métriques dédiées à l'évaluation de problèmes sous forme de classification binaire. Comme notre modèle est probabiliste, nous pouvons avoir deux types d'alertes : des alertes probabilistes (correspondant à la probabilité d'être en période active) et des alertes déterministes. Ici, l'alerte déterministe est calculée à partir de l'alerte probabiliste en suivant une méthode simple : si on a une probabilité $p \geq p_{\text{seuil}}$ d'être en période active, alors on émet une alerte, sinon non. Il faut simplement déterminer quelle est la probabilité seuil p_{seuil} optimale. Pour ce faire on peut utiliser la courbe de la fonction d'efficacité du récepteur ou *Receiver Operating Characteristic*. Notons qu'ici on utilise la courbe **ROC** et non la courbe **PRC** (comme au Chapitre 3) car nos classes sont moins déséquilibrées : la classe minoritaire représente ici 25% des échantillons (alors qu'au Chapitre 3 cette proportion était de 7,4%). Rappelons que cette courbe s'obtient simplement en comparant les valeurs de **rappel** et de **FAR** obtenues avec plusieurs seuils de probabilité. Nous la donnons pour chaque horizon de prédiction en Figure 6.4

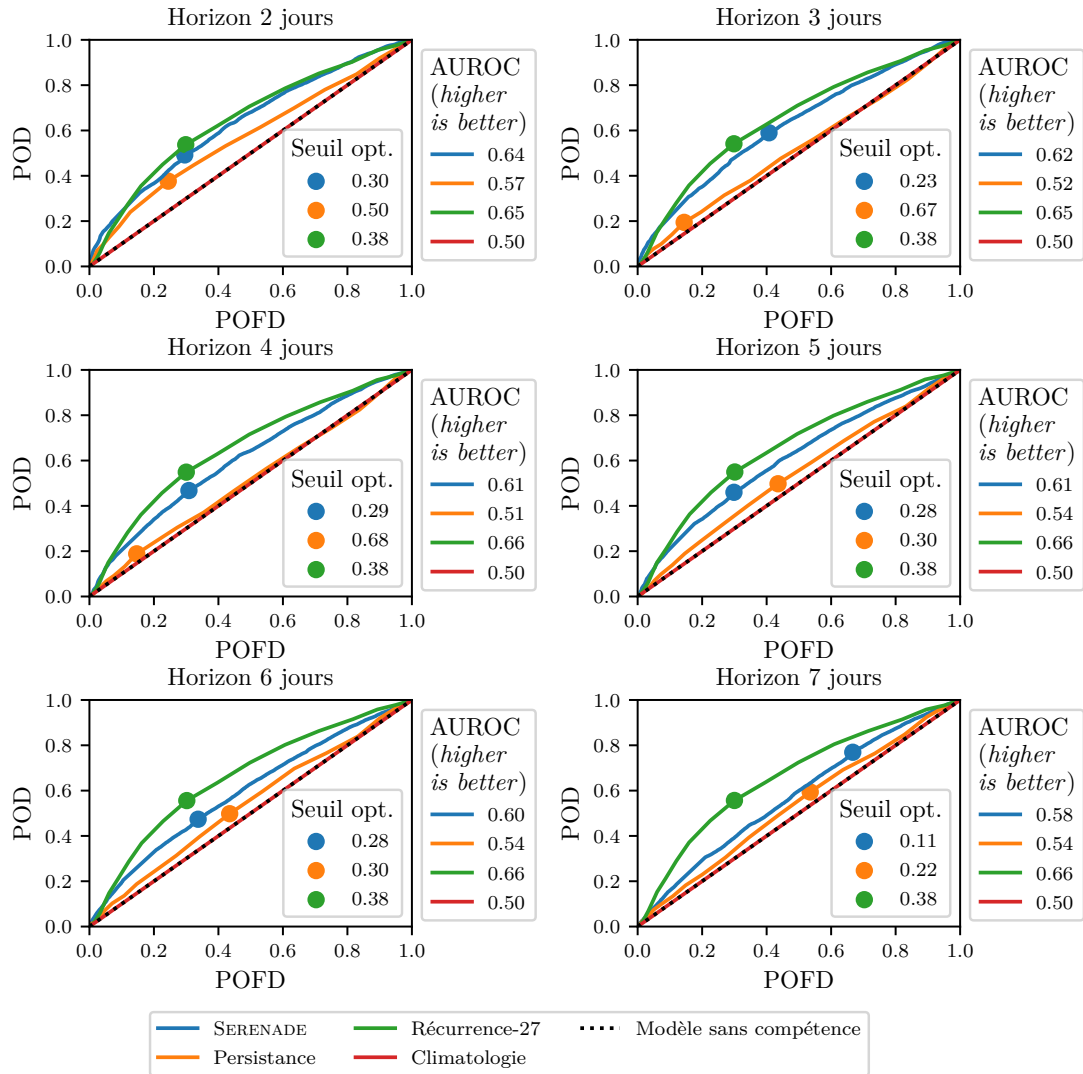


FIGURE 6.4 – Courbes ROC de SERENADE et des modèles de référence pour chaque horizon de prédiction. Les valeurs d’AUROC ainsi que les probabilités p_{seuil} optimales pour chacun des modèles sont données.

Dans cette figure, pour chaque horizon de prédiction nous donnons l’aire sous la courbe ROC (ROC) correspondant à chaque modèle, ainsi que la probabilité seuil p_{seuil} optimale, calculée en prenant le point qui maximise l’écart entre la POD et la POFD. Plusieurs éléments notables ressortent de cette figure.

Tout d’abord, on note, sans surprise, que le modèle de climatologie est équivalent à un modèle sans compétence. Nous avons déjà évoqué cette caractéristique dans la section précédente. Pour rappel, cela est dû au fait que nous prédisons toujours la même valeur avec le modèle de climatologie. Ainsi la probabilité d’être en période active est toujours la même. C’est pourquoi il n’y a pas de probabilité seuil optimale pour le modèle de climatologie.

Notons ensuite que selon la courbe ROC, SERENADE est quasiment équivalent au modèle de récurrence pour un horizon de 2 jours mais la performance de

SERENADE se dégrade lorsque l'horizon de prédiction augmente, avec une valeur d'**AUROC** diminuant progressivement de 0,64 à 0,58 pour un horizon de prédiction de 7 jours (alors que l'**AUROC** du modèle de récurrence reste fixe, par définition). Toutefois **SERENADE** reste toujours meilleur que le modèle de persistance dont l'**AUROC** chute en dessous de 0,55 – atteignant même 0,51, soit un modèle quasiment sans compétence – dès que l'horizon de prédiction est supérieur ou égal à 3 jours.

On remarque enfin que la probabilité seuil optimale pour **SERENADE** oscille entre 0,23 et 0,30 (sauf pour l'horizon de prédiction de 7 jours, où la probabilité seuil chute à 0,11). Nous utiliserons ces probabilités seuil dans la suite pour définir nos alertes déterministes.

Pour résumer, la Figure 6.4 indique que le modèle de récurrence semble mieux distinguer les périodes actives des périodes calmes. **SERENADE** présente des performances qui s'en approchent de près, mais qui se dégradent lorsque l'horizon de prédiction augmente. Les autres modèles de référence sont en revanche beaucoup moins performants.

Il nous faut maintenant affiner ces conclusions à l'aide des autres métriques d'évaluation, données pour chacun des quatre modèles et pour tous les horizons de prédiction en Figure 6.5.

Parmi ces six métriques, seul le *Brier Score* est une métrique dont la valeur optimale est 0. Pour les autres, ce sont des métriques qui ont comme valeur optimale 1. De plus, seul le **BS** prend en compte le caractère probabiliste de la prédiction, les autres métriques utilisant les alertes déterministes définies ci-dessus. Pour le modèle de climatologie, comme il n'y a pas de probabilité seuil optimale, nous avons pris $p_{\text{seuil}} = 0.$, ce qui explique la valeur de **rappel** égale à 1 par exemple.

D'une manière générale on constate que le modèle **SERENADE** est uniquement battu par le modèle de récurrence mais de manière quasiment systématique en ce qui concerne les métriques déterministes. Pour certains horizons de prédiction **SERENADE** peut obtenir de meilleurs scores que le modèle de récurrence (par exemple le **rappel** pour des horizons de 3 et 7 jours), mais généralement le modèle de récurrence reste devant. Cela est notamment vrai pour le **TS** et le **HSS**, qui sont des métriques se voulant assez globales.

En revanche, jusqu'à un horizon de prédiction de 5 jours (inclus), **SERENADE** bat le modèle de récurrence en termes de **BS**. Cela conforte les observations faites lors de l'évaluation du modèle du point de vue de la régression, selon lesquelles une partie non-négligeable de la compétence de **SERENADE** provient de son caractère probabiliste.

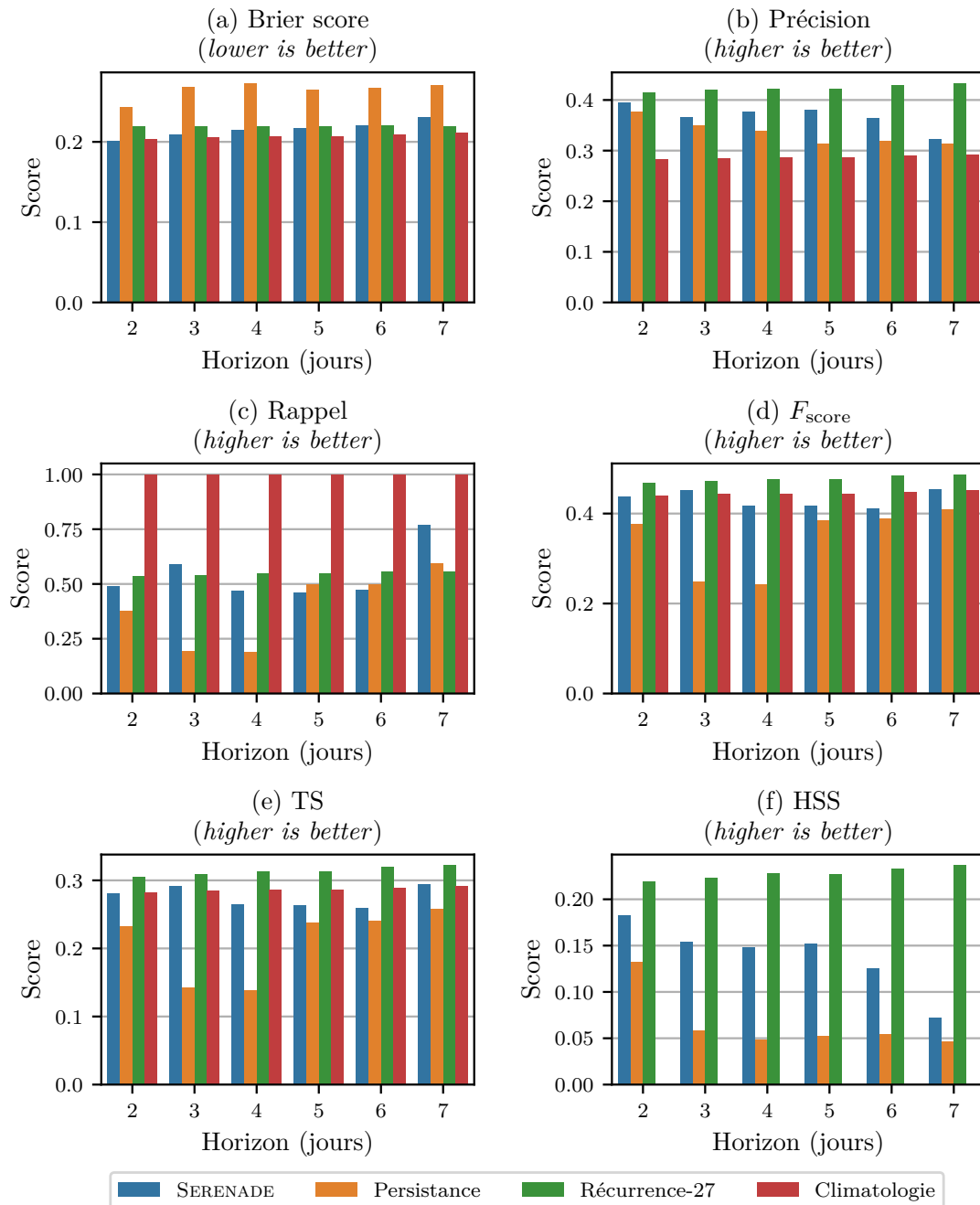


FIGURE 6.5 – Métriques d'évaluation du système d'alerte dérivé de SERENADE et des trois modèles de référence pour tous les horizons de prédiction considérés. À l'exception du *Brier Score*, plus la valeur de la métrique est élevée, plus le modèle est performant.

Note : Dans cette figure, on constate que les résultats pour le modèle de Récurrence-27 varient légèrement en fonction de l'horizon, alors que cela ne devrait pas être le cas. Cela est dû aux trous de données dans le jeu de données SDO/AIA – ML-READY, à cause desquels nous n'avons pas de prédictions pour tous les instants. Cela crée des petites différences sur les bords de notre jeu d'évaluation en fonction de l'horizon, suffisantes pour qu'on observe les petites variations de valeurs que nous venons de mentionner.

6.1.3 Bilan de l'évaluation de SERENADE sur l'ensemble de la période 2010 – 2018

L'évaluation du modèle **SERENADE** et sa comparaison avec les modèles de référence nous montrent plusieurs choses. Tout d'abord il apparaît clairement que **SERENADE** propose de bien meilleures performances que les modèles de persistance et de climatologie, pour des horizons de prédiction allant de 2 à 7 jours.

La comparaison avec le modèle de récurrence est plus nuancée. En effet, l'évaluation du modèle à l'aide des métriques dédiées à la régression montre que **SERENADE** est généralement plus performant que le modèle de récurrence, même lorsqu'on ne considère que les périodes pour lesquelles $Kp_{\max,24h} \geq 4^-$. En revanche, les métriques dédiées à la classification montrent que **SERENADE** a des performances légèrement en deçà du modèle de récurrence lorsqu'il s'agit de distinguer les périodes actives des périodes calmes de façon déterministe. Lorsqu'on conserve le caractère probabiliste des prédictions alors il devient difficile de distinguer les deux modèles, le modèle de récurrence donnant de meilleures valeurs d'**AUROC**, et **SERENADE** de meilleures valeurs de **BS** (jusqu'à 5 jours en avance). Nous concluons donc que cette première version de **SERENADE** est généralement plus précise que le modèle de récurrence, mais que son utilisation en tant que système d'alertes doit se faire en conservant le caractère probabiliste des prédictions.

Enfin, nous pouvons d'ores et déjà répondre à notre problématique posée au début du chapitre précédent, à savoir : est-il possible de relier directement l'activité solaire (par le biais d'imagerie solaire) à l'activité géomagnétique, à l'aide des méthodes statistiques complexes que sont les réseaux de neurones ? Nous venons de montrer que cela était possible, avec des résultats au moins aussi bons que l'état de l'art actuel, obtenus avec ce premier modèle pour lequel beaucoup de compromis ont été faits. Insistons donc sur le fait que **SERENADE**, contrairement au modèle de récurrence, estime réellement la géoefficacité d'un événement solaire. En effet, le modèle de récurrence est un modèle qui s'appuie simplement sur l'auto-corrélation de Kp , sans prendre en compte l'activité solaire. À l'inverse, **SERENADE** s'appuie uniquement sur l'imagerie solaire. Cette différence est fondamentale et contribue à la valeur de notre modèle.

6.2 Analyse des performances selon la phase du cycle solaire

Afin de mieux comprendre le fonctionnement et les performances obtenues avec notre modèle, nous les analysons en distinguant les différentes phases du cycle solaire. Cela nous permettra de mieux caractériser **SERENADE** et apporter des éléments d'explication quant au fonctionnement du modèle.

Tout d'abord nous devons diviser nos données en fonction des phases du cycle solaire. Nos données, représentant une période allant de mi-2010 à fin 2018, ne couvrent pas le cycle solaire 24 en entier. C'est pourquoi nous séparons notre jeu

de données en 3 phases du cycle solaire, correspondant approximativement aux phases ascendante (mi-2010 à fin 2012), maximum (début 2013 à fin 2015) et descendante (début 2016 à fin 2018) – le minimum solaire n'est donc pas représenté. La Figure 6.6 montre l'évolution du cycle solaire (représenté par le nombre de taches solaires) et de $Kp_{\max,24h}$ sur cette période, avec le découpage du cycle solaire que nous proposons.

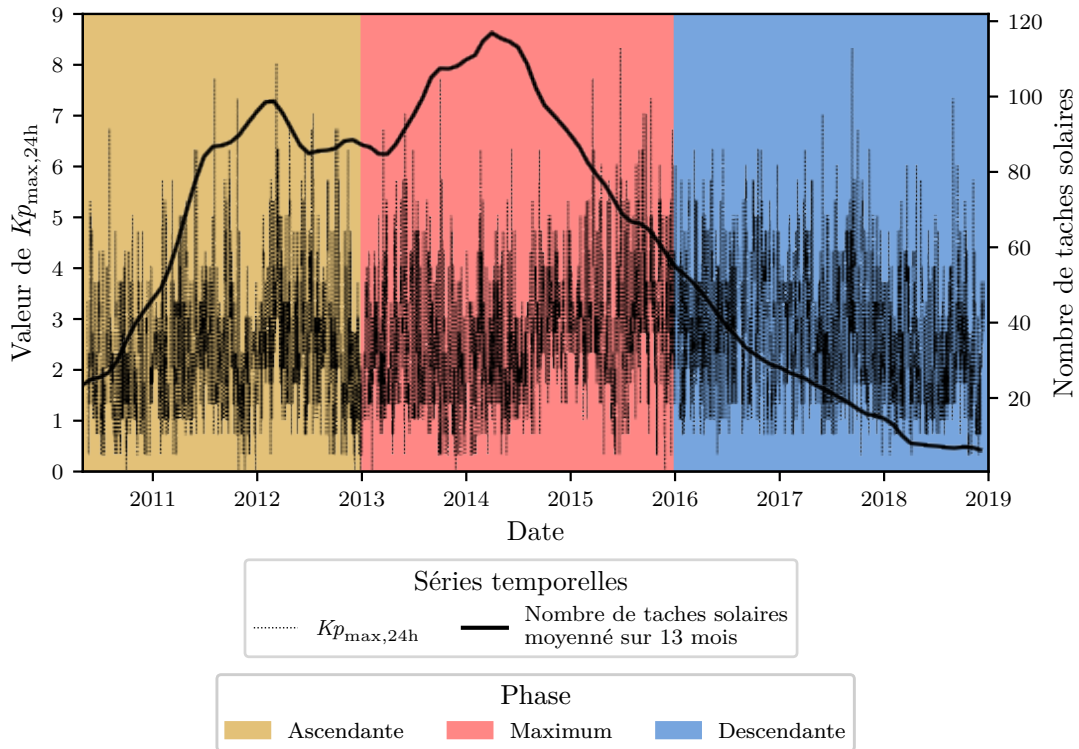


FIGURE 6.6 – Répartition de nos données en fonction des phases du cycle solaire.

Il aurait été possible de proposer un découpage plus fin du cycle solaire, mais celui-ci présente l'avantage de proposer un nombre d'échantillons dans chaque phase du même ordre de grandeur, ce qui nous permet d'effectuer des comparaisons statistiquement significatives. Il y a en effet 17165 échantillons dans la phase ascendante (dont $\approx 21,7\%$ en période active), 20894 au cours du maximum solaire ($\approx 29,6\%$ en période active) et 21403 pendant la phase descendante ($\approx 30,1\%$ en période active).

Nous étudions donc les performances de **SERENADE** pour chacune de ces phases du cycle solaire, en utilisant, comme dans la section précédente, des métriques d'évaluation dédiées d'un côté à la régression et de l'autre à la classification. La Figure 6.7 montre les valeurs des métriques dédiées à la régression obtenues avec chacun des quatre modèles, pour un horizon de prédiction de 2 jours.

Tout d'abord on remarque que nos observations effectuées pour l'ensemble de test complet restent généralement valables ici, avec le modèle **SERENADE** proposant de meilleures performances que les modèles de référence, surtout lorsqu'on ne considère que les périodes actives.

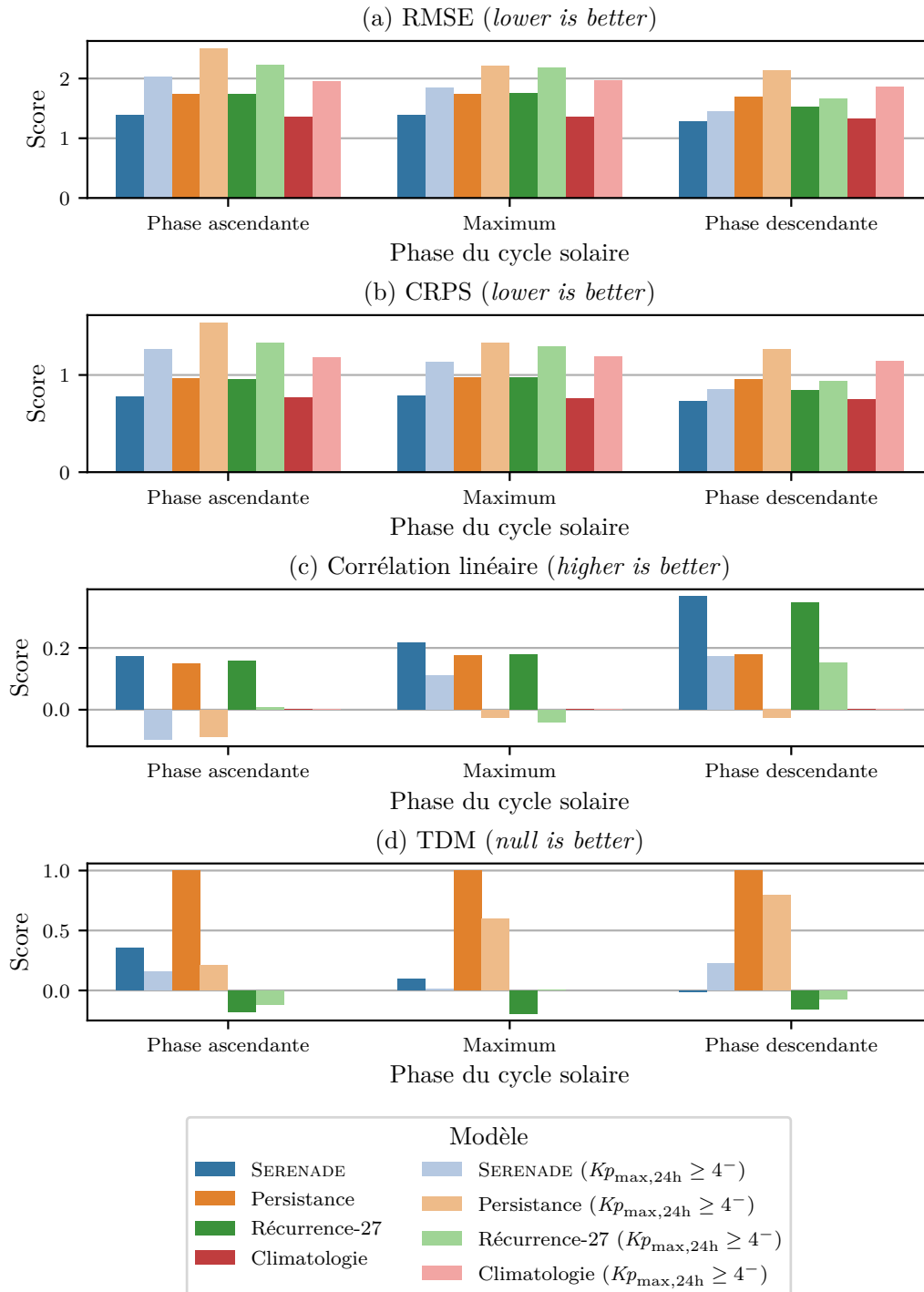


FIGURE 6.7 – Évaluation de **SERENADE** et des modèles de référence en fonction des phases du cycle solaire pour un horizon de prédiction de 2 jours à l'aide de métriques dédiées à l'évaluation du problème de régression.

Dans un second temps, on voit que **SERENADE** est plus performant en phase descendante du cycle solaire qu'en phase ascendante ou en maximum solaire. Par exemple la **RMSE** de **SERENADE** est égale à 1,39 en phase ascendante, 1,38 en maximum et à 1,28 en phase descendante. En ne considérant que les périodes actives la différence est encore plus grande : la RMSE vaut 2,04 en phase ascendante,

1,84 en maximum et 1,45 en phase descendante. **SERENADE** est donc quasiment aussi performant en période active lors de la phase de descente que pour toutes les périodes (calmes et actives) en phases ascendante et en maximum.

Le fait que **SERENADE** soit plus performant en phase descendante, surtout pour les périodes actives est un élément très important pour expliquer le fonctionnement du modèle. En effet, comme nous l'avons vu au Chapitre 1 (voir les Figures 1.10 et 1.14), on observe un plus grand nombre de perturbations magnétosphériques liées à l'activité éruptive du Soleil lorsqu'on se situe en maximum d'activité, tandis qu'en phase descendante on mesure un plus grand nombre de perturbations liées à des vents solaires rapides provenant des trous coronaux.

Or, rappelons que **SERENADE** est un modèle alimenté par des images du Soleil temporellement espacées de 12 heures. De plus, la plupart des éruptions solaires et des **CMEs** sont observables seulement l'espace de quelques minutes à quelques dizaines de minutes au plus. C'est pourquoi avec une telle fréquence d'images on pouvait s'attendre à ce que **SERENADE** ne soit pas capable d'apprendre correctement à prédire les événements ayant pour origine un phénomène éruptif comme une **CME**. En effet, même si, par chance, nous parvenions parfois à avoir dans notre base de données d'apprentissage des images dans lesquelles une **CME** est bien visible, il n'y aurait certainement pas assez d'échantillons de cette sorte et nous avons déjà établi au Chapitre 5 qu'une des limites de notre approche était la quantité limitée de données à notre disposition. En outre, les trous coronaux sont des structures coronales de plus longue durée (de l'ordre de quelques jours à quelques semaines, voire mois). Nos résultats suggèrent donc que **SERENADE** a été capable de mieux apprendre à prédire les événements géomagnétiques ayant pour origine les trous coronaux mais non ceux ayant pour origine une **CME**, en accord avec la physique et la statistique du problème.

La Figure 6.8 montre les valeurs de quatre métriques dédiées à l'évaluation du système d'alerte prédictif sur chacune des phases du cycle solaire. La comparaison de **SERENADE** avec les modèles de référence devient ici plus contrastée qu'en section 6.1. En effet, certes **SERENADE** est ici aussi plus performant en phase descendante, mais notre modèle surpasse même (pour le **BS**, le F_{score} et le **TS**) le modèle de récurrence à 27 jours. En revanche **SERENADE** est moins bon sur les phases ascendantes et maximum, avec des performances parfois proches de celles des modèles de climatologie ou de persistance.

Il apparaît que l'évaluation du système d'alerte accentue les différences entre phases ascendantes et descendantes du cycle solaire. Comme ces métriques servent à mesurer la capacité de **SERENADE** à distinguer les périodes actives de périodes calmes, cela tend à confirmer notre interprétation selon laquelle **SERENADE** serait moins performant en périodes ascendante et maximum car il n'a pas correctement appris à prédire les perturbations induites par des **CMEs**. Toutefois, sa capacité à correctement prédire les événements induits par **SIR** est également validée par le fait que **SERENADE** est alors plus performant que le modèle de récurrence, qui est lui-même un modèle tirant sa compétence prédictive des structures récurrentes à la surface du Soleil comme les trous coronaux.

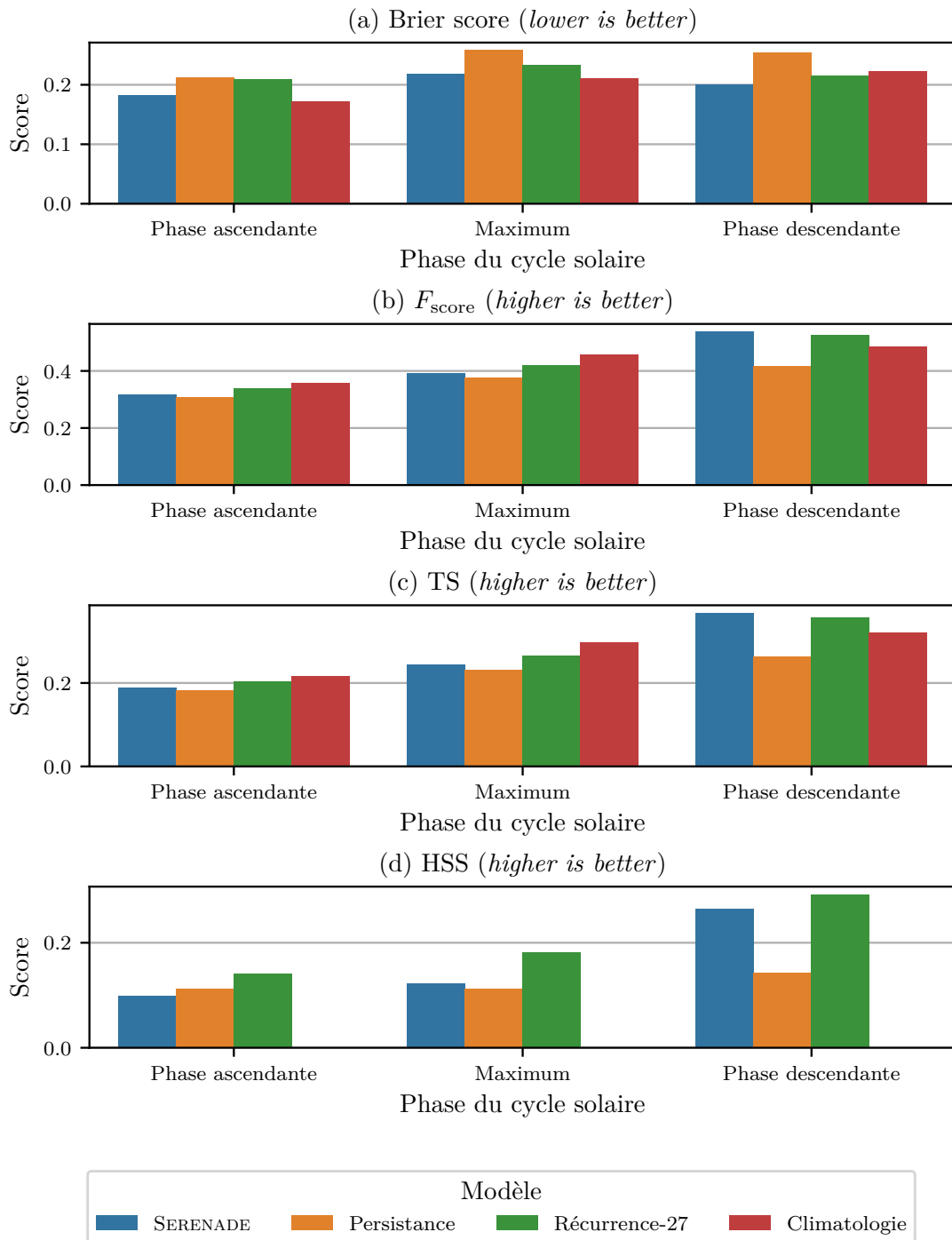


FIGURE 6.8 – Évaluation de SERENADE et des modèles de référence en fonction des phases du cycle solaire pour un horizon de prédiction de 2 jours à l'aide de métriques dédiées à l'évaluation du problème de classification binaire.

Afin de vérifier ces hypothèses, nous avons également eu recours à l'utilisation de catalogues d'événements. En utilisant le catalogue d'ICMEs proposé par NGUYEN 2021 et le catalogue de SIRs fourni par CHI et al. 2018, nous avons reproduit en Figure 6.9 la Figure 6.1 en y faisant apparaître les événements recen-

sés dans ces catalogues. Ainsi, les régions rouge correspondent aux périodes où des ICMEs ont été recensées et les périodes vertes aux SIRs. On observe bien que les ICMEs sont parfaitement « ignorées » par le modèle, tandis que lors des périodes de perturbations dues aux SIRs la dynamique de $Kp_{\max,24h}$ semble bien mieux reproduite. Ainsi, cette simple analyse visuelle s'appuyant sur des catalogues d'événements semble renforcer nos hypothèses sur le fonctionnement et la nature des performances de SERENADE.

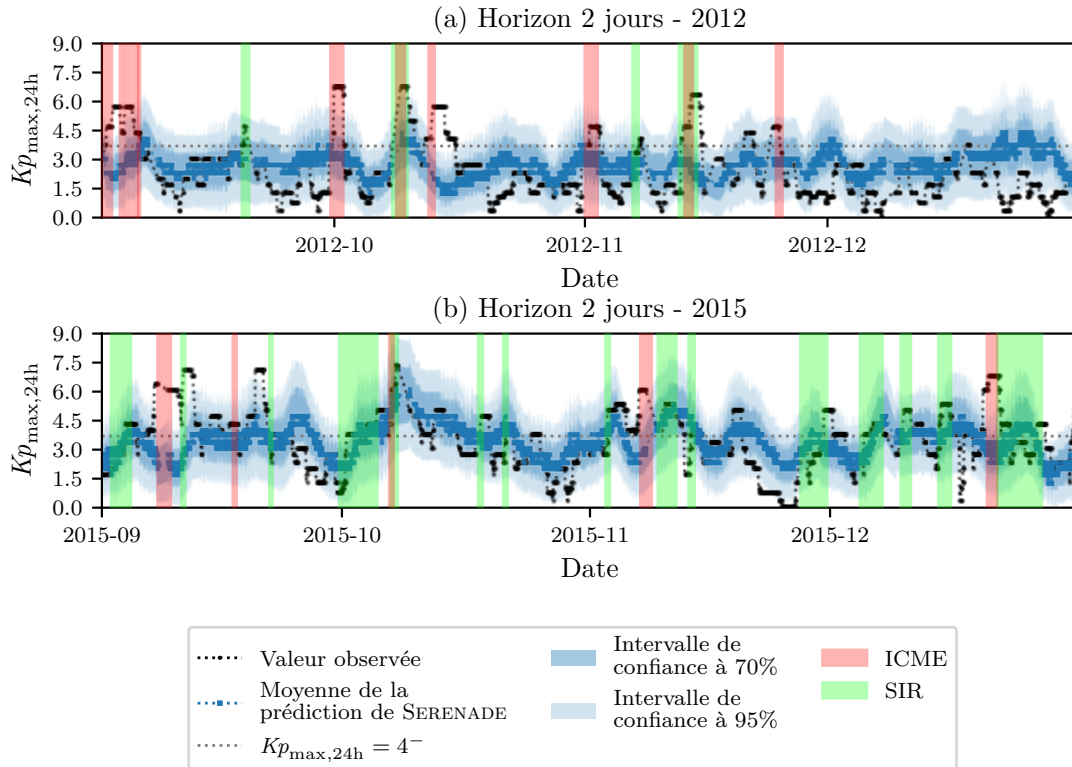


FIGURE 6.9 – Exemples de prédictions obtenues avec SERENADE 2 jours en avance pour deux périodes : de septembre à fin décembre 2012 et de septembre à fin décembre 2015. Les périodes d'ICMEs (resp. de SIRs) sont mises en valeur par une zone rouge (resp. verte).

Notons enfin que la performance (en classification binaire) de SERENADE en phase descendante se dégrade lorsque l'horizon de prédiction augmente, comme le montre la Figure 6.10, qui indique la performance des différents modèles lors de la phase descendante seule, en fonction de l'horizon de prédiction. Lorsque l'horizon de prédiction est supérieur à 3 jours alors SERENADE devient à nouveau moins performant que le modèle de récurrence (pour toutes les métriques de classification) et cette chute de performance est progressive avec l'augmentation de l'horizon de prédiction.

Pour résumer, l'étude des performances de SERENADE en distinguant les différentes phases du cycle solaire montre que notre modèle, pour un horizon de prédiction de 2 jours, est plus performant (y compris, et surtout, en termes de classification binaire) que le modèle de récurrence en phase descendante du cycle solaire, dominée par les perturbations dues aux trous coronaux. Cependant,

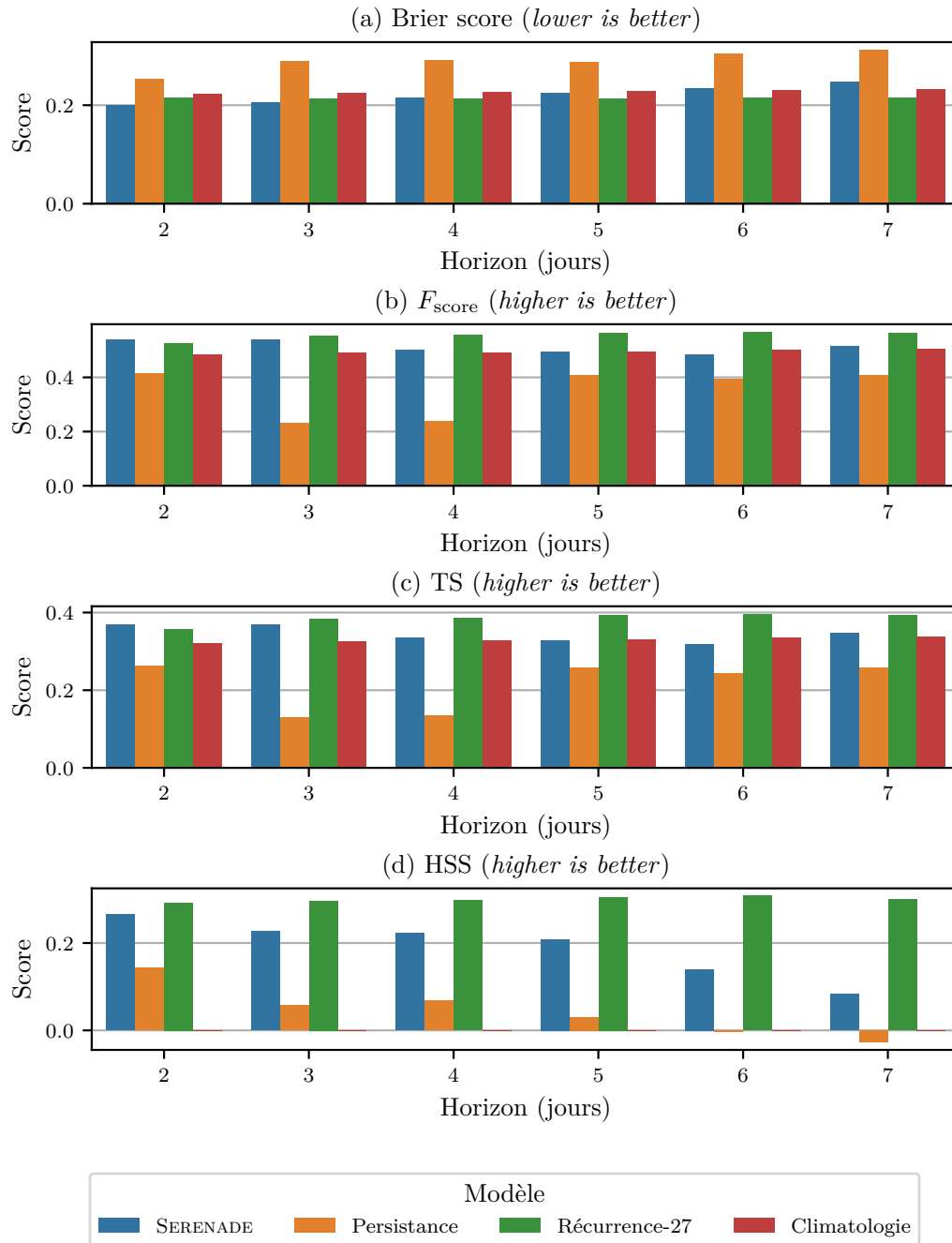


FIGURE 6.10 – Évaluation de **SERENADE** et des modèles de référence en fonction de l'horizon de prédiction lors de la phase descendante du cycle solaire, à l'aide de métriques dédiées à l'évaluation du problème de classification binaire.

les moins bonnes performances de **SERENADE** lors des autres phases du cycle solaire suggèrent que le modèle n'a pas été en mesure d'apprendre à prédire les perturbations dues aux **CMEs**, ce qui n'est pas surprenant étant donné la résolution temporelle de 12 heures des séquences d'images que nous utilisons pour alimenter le modèle. Cette étude nous a donc permis non seulement de mieux caractériser **SERENADE**, mais aussi de pouvoir interpréter partiellement son fonctionnement.

6.3 Prédiction de l'indice $Ca_{\max,24h}$

Nous proposons ici d'appliquer la même méthodologie que précédemment, mais pour prédire cette fois l'indice $Ca_{\max,24h}$, défini à partir de Ca selon la même méthode que $Kp_{\max,24h}$. Les avantages de l'utilisation de $Ca_{\max,24h}$ sont les mêmes que ceux déjà exposés au Chapitre 4, à savoir qu'il s'agit d'un indice plus représentatif de l'état de remplissage des ceintures de radiations.

6.3.1 Modifications du *pipeline*

Pour prédire $Ca_{\max,24h}$ nous adaptons le *pipeline* de SERENADE afin de l'optimiser pour la prédiction de $Ca_{\max,24h}$. Nous appelons le modèle qui en résulte SERENADE- Ca . Ses hyperparamètres optimaux ont été trouvés en faisant une nouvelle recherche empirique par itérations successives. Les différences entre SERENADE et SERENADE- Ca , en dehors de l'indice géomagnétique utilisé, sont les suivantes :

- la dimension des états de chacune des deux couches du LSTM est de 512 (au lieu de 2048) ;
- la méthode d'optimisation est toujours Adam, mais avec un *learning rate* valant $lr = 5 \times 10^{-6}$ (au lieu de 10^{-5}) et un coefficient de *weight decay* fixé à $wd = 10^{-5}$ (au lieu de 10^{-4}) ;
- l'indice $Ca_{\max,24h}$ est re-normalisé pour l'apprentissage, en ramenant sa valeur moyenne à 0 et son écart-type à 1 ($Kp_{\max,24h}$ n'était pas re-normalisé) ;
- le critère de convergence inclut la GNLL, ce qui donne (en reprenant les notations du Chapitre 5) :

$$\begin{aligned} \text{Crit}_{\text{conv}}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{y}_{\text{vrai}}) = & \overline{\text{GNLL}}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{y}_{\text{vrai}}) + \overline{\text{MSE}}(\boldsymbol{\mu}, \mathbf{y}_{\text{vrai}}) \\ & + \overline{\text{CRPS}}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{y}_{\text{vrai}}) + (|\text{QQ}_{\text{slope}}(\boldsymbol{\mu}, \mathbf{y}_{\text{vrai}}) - 1| + 1) \\ & - \text{AUROC}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{y}_{\text{vrai}}) - \frac{\text{Sn}(\boldsymbol{\mu})}{\text{Sn}(\mathbf{y}_{\text{vrai}})}. \end{aligned} \quad (6.2)$$

En plus de ces modifications il nous a fallu choisir un seuil d'alerte pour le problème de classification binaire. Une solution naturelle aurait été de prendre la même valeur qu'au Chapitre 4, à savoir 38 nT. Mais ici, outre le fait que nous prédisons $Ca_{\max,24h}$ et non pas Ca , nous n'avons que trop peu de d'échantillons où $Ca_{\max,24h} \geq 38$ nT (moins de 2% de l'ensemble de nos données). Cela est dû au fait, déjà établi, que le cycle solaire 24 était d'une intensité remarquablement faible. C'est pourquoi nous choisissons un seuil correspondant plutôt au 3^{ème} quartile de la distribution de $Ca_{\max,24h}$ sur la période 2010 – 2018 (ce qui correspond au positionnement de la valeur 4^- dans la distribution de $Kp_{\max,24h}$), et nous prenons donc un seuil égal à 22,0 nT.

Précisons enfin que nous avons entraîné SERENADE- Ca avec des horizons allant jusqu'à 9 jours d'avance, afin de vérifier si le fait que Ca soit un indice intégré dans le temps ne nous permet pas d'obtenir une avance supplémentaire.

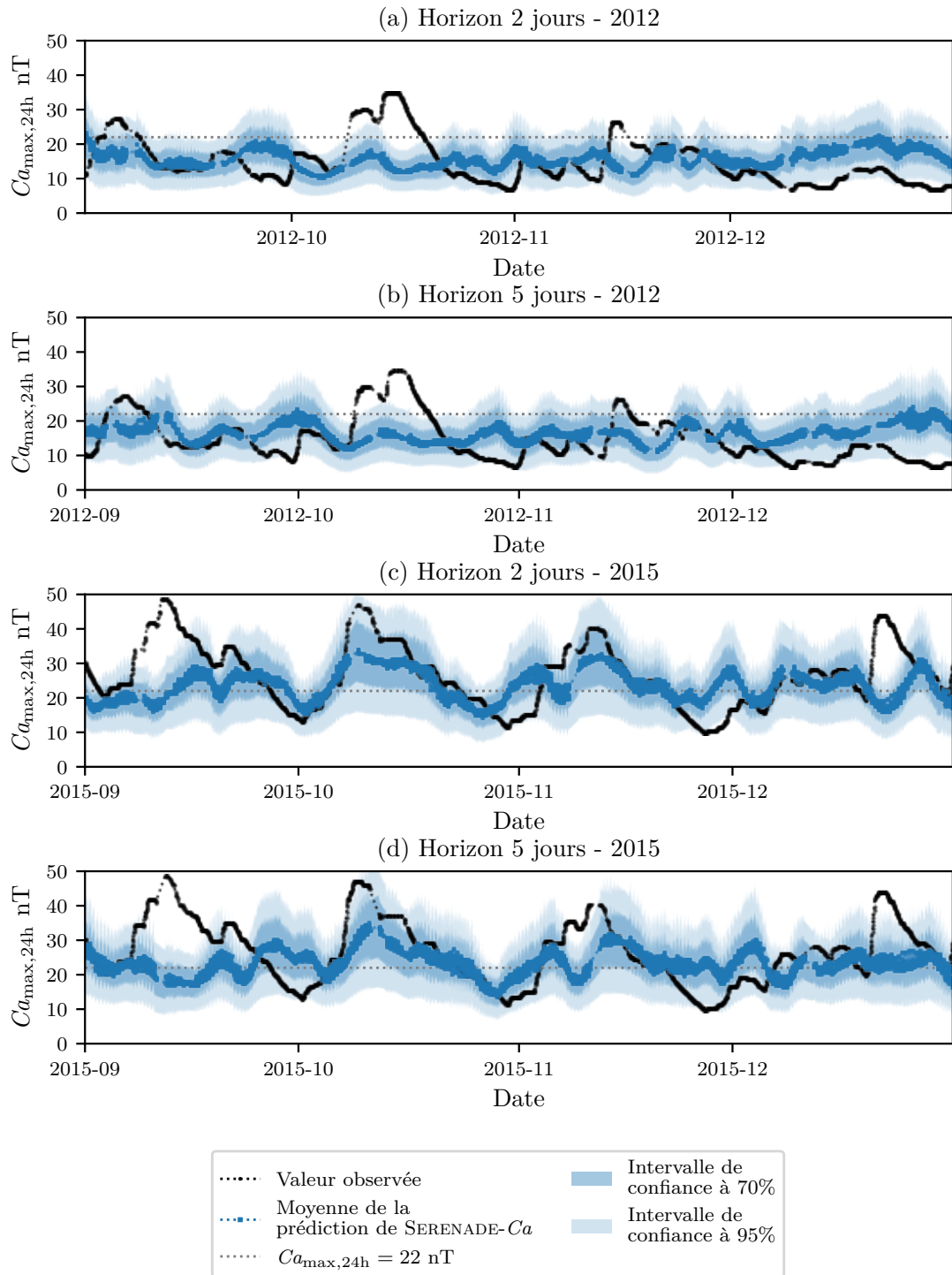
6.3.2 Évaluation des performances de SERENADE- Ca 

FIGURE 6.11 – Exemples de prédictions obtenues avec SERENADE- Ca 2 jours et 5 jours en avance pour deux périodes : de septembre à fin décembre 2012 et de septembre à fin décembre 2015.

Commençons ici aussi par montrer des exemples de prédiction de Ca . La Figure 6.11 montre les prédictions obtenues avec SERENADE- Ca 2 jours et 5 jours en avance pour les deux mêmes périodes qu'en Figure 6.1 : de septembre à fin

décembre 2012 et de septembre à fin décembre 2015. De même qu'avec **SERENADE**, on remarque que **SERENADE-Ca** semble être capable de reproduire certains éléments de la dynamique de $Ca_{\max,24h}$, surtout en 2015, pour un horizon de 2 jours. Le fait de réussir à entraîner un modèle pour la prédiction de deux indices aux dynamiques différentes est plutôt satisfaisant puisque cela montre que notre méthodologie est reproductible et, au prix de quelques ajustements, pourrait éventuellement s'appliquer à encore d'autres indices.

Nous résumons ici les principaux résultats obtenus avec **SERENADE-Ca**. La Figure 6.12 donne les valeurs des métriques d'évaluation du problème de régression pour des horizons de 2, 5, 7 et 9 jours. On distingue les résultats obtenus avec l'ensemble d'évaluation complet et ceux obtenus uniquement sur les périodes actives. Outre les valeurs, qui ne sont numériquement pas les mêmes, on aboutit ici aussi aux mêmes conclusions qu'avec **SERENADE** : la performance du modèle se dégrade avec l'augmentation de l'horizon de prédiction. Toutefois, il nous faut surtout comparer notre modèle aux modèles de référence, ce qui est fait en Figure 6.13.

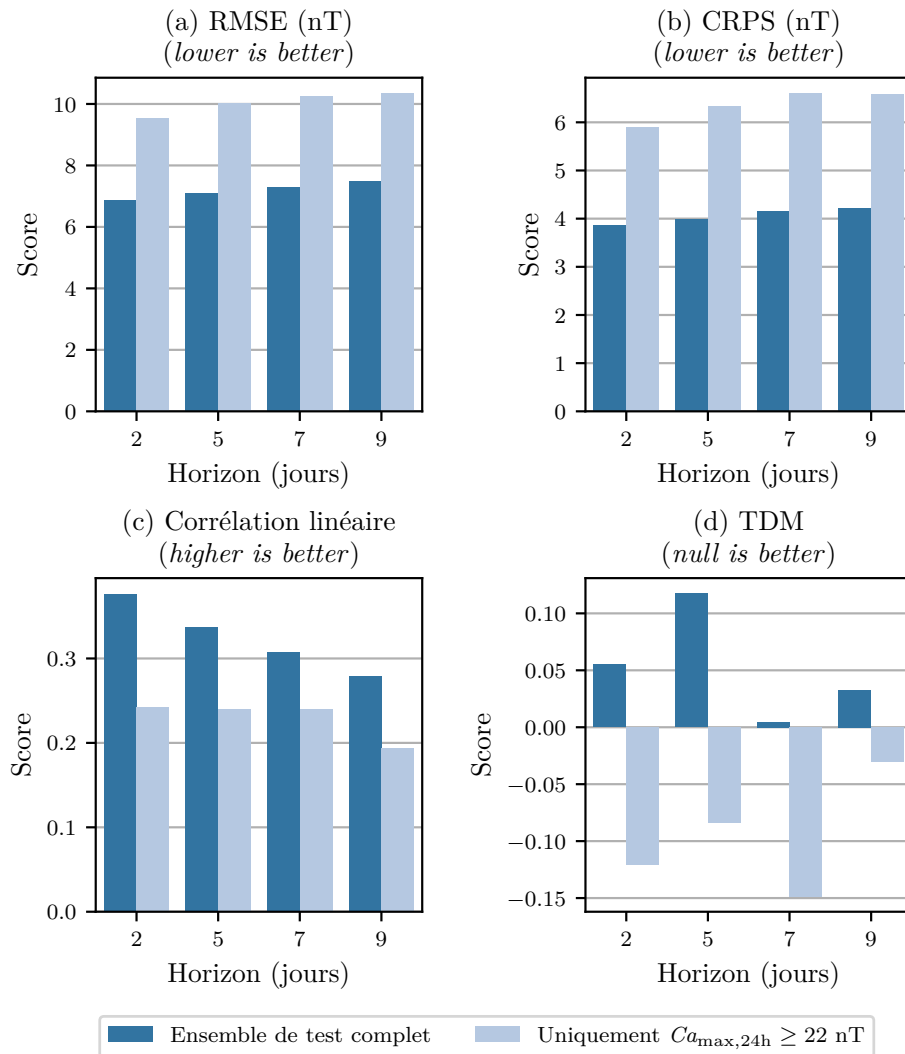


FIGURE 6.12 – Évaluation de **SERENADE-Ca** pour quatre horizons de prédiction différents à l'aide de métriques dédiées au problème de régression.

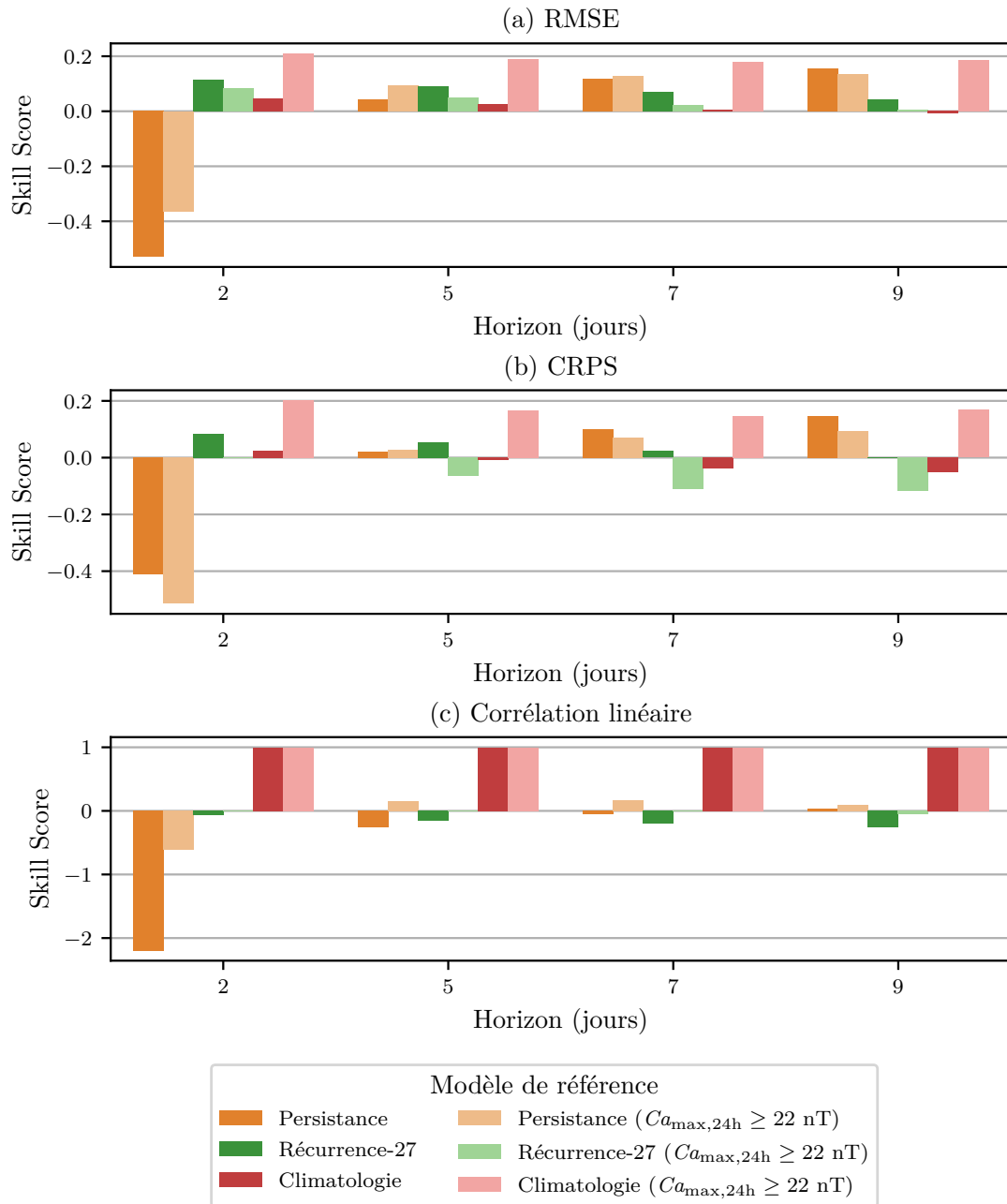


FIGURE 6.13 – Comparaison de *SERENADE-Ca* avec trois modèles de référence à l'aide de *Skill Scores* relatifs à trois métriques dédiées au problème de régression.

Note : Comme nous représentons les *Skill Scores*, plus une barre est haute (en direction des valeurs positives), plus *SERENADE* est compétent par rapport au modèle de référence utilisé.

La Figure 6.13 nous donne les *Skill Score* (relatifs à la RMSE, au CRPS et à la corrélation linéaire) de *SERENADE-Ca* par rapport aux trois mêmes modèles de référence déjà utilisés. On a ici notre première grande différence avec *SERENADE* : pour un horizon de 2 jours, *SERENADE-Ca* est bien moins performant que le modèle de persistance, quels que soient la métrique et l'ensemble d'évaluation considérés. Puis pour des horizons plus grands, *SERENADE-Ca* devient meilleur que le modèle de persistance, mais équivalent (RMSE, R) ou moins bon (CRPS) que le modèle

de récurrence à 27 jours sur les périodes actives.

La grande efficacité du modèle de persistance pour les horizons les plus courts est causée par la très bonne auto-corrélation temporelle de Ca , conséquence du fait que ce soit un indice intégré dans le temps (comme nous l'avons vu au Chapitre 4). Cela est encore plus marqué lorsqu'on évalue $SERENADE-Ca$ en tant que système d'alerte binaire, par exemple avec les courbes ROC, données en Figure 6.14. On voit clairement que le modèle de persistance est très performant pour un horizon de 2 jours, avec une valeur d'AUROC de 0,88, contre 0,68 pour $SERENADE-Ca$ et 0,71 pour le modèle de récurrence, qui ont des courbes ROC très proches. On voit aussi très bien que les performances du modèle de persistance se dégradent fortement avec l'augmentation de l'horizon de prédiction, si bien qu'à 9 jours $SERENADE-Ca$ possède une meilleure AUROC (0,65) que le modèle de persistance (0,63). En revanche, le modèle de récurrence reste stable, et donc à partir du moment où le modèle de persistance perd de sa compétence, il est « relayé » par le modèle de récurrence dans le rôle de meilleur modèle.

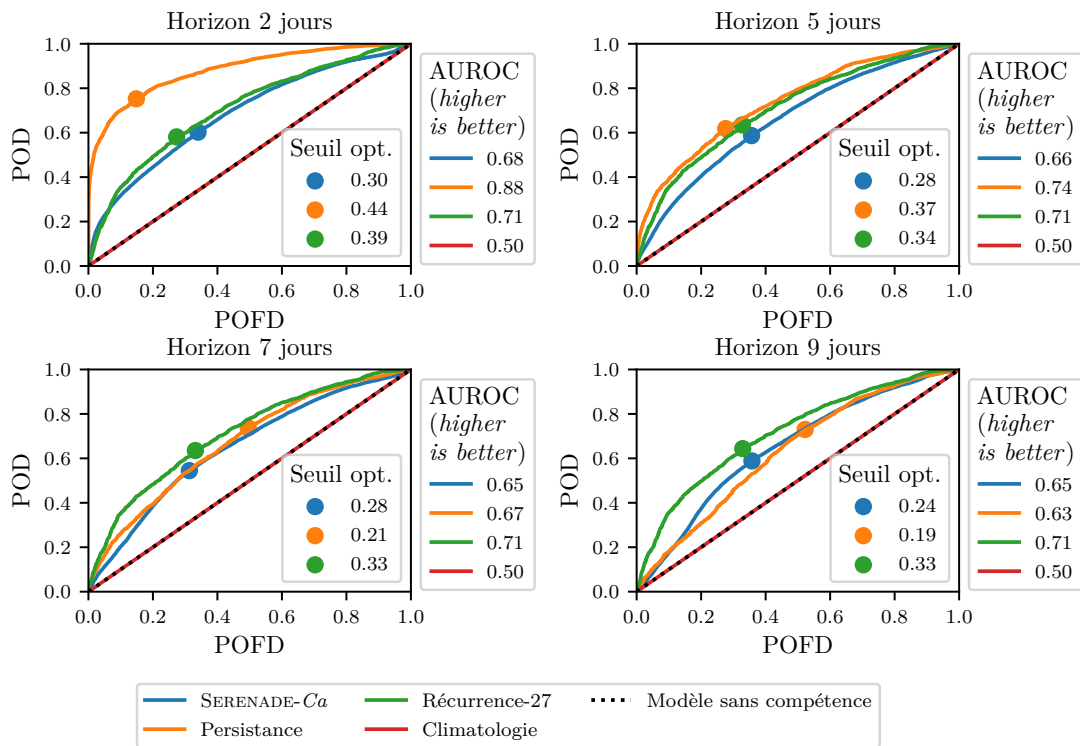


FIGURE 6.14 – Courbes ROC de $SERENADE-Ca$ et des modèles de référence pour les horizons de prédiction 2, 5, 7 et 9 jours. Les valeurs d'AUROC ainsi que les probabilités p_{seuil} optimales pour chacun des modèles sont données.

En outre, lorsqu'on évalue $SERENADE-Ca$ en fonction des phases du cycle solaire, il est difficile d'aboutir à une conclusion comme avec $SERENADE$. Les Figures 6.15 et la Figure 6.16 donnent les résultats en régression et en classification pour un horizon de 2 jours lorsqu'on distingue les phases du cycle solaire. Il semblerait que $SERENADE-Ca$ soit sensiblement équivalent en maximum solaire et en phase

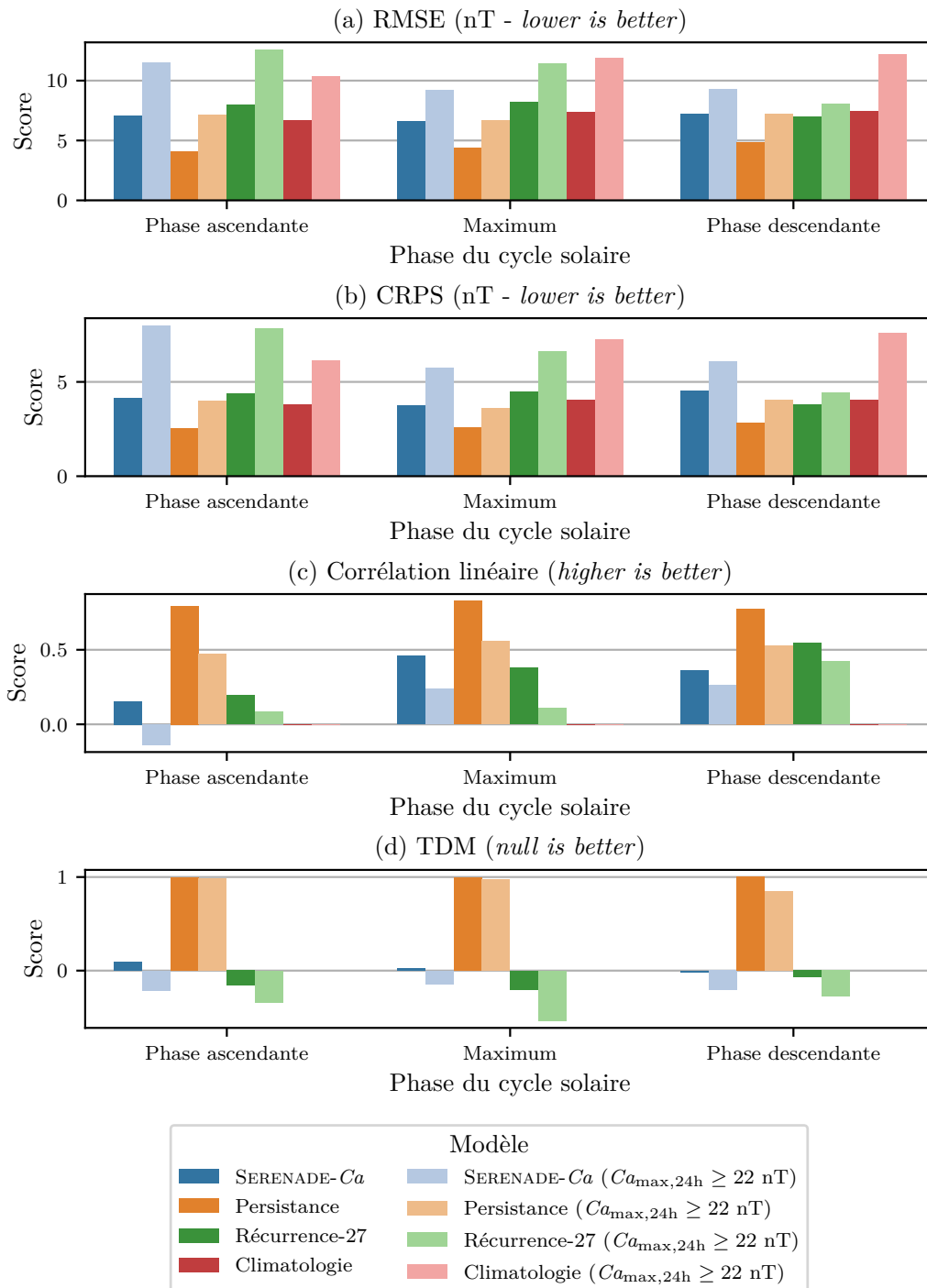


FIGURE 6.15 – Évaluation de **SERENADE- Ca** et des modèles de référence en fonction des phases du cycle solaire pour un horizon de prédiction de 2 jours à l'aide de métriques dédiées à l'évaluation du problème de régression.

descendante (et moins performant en phase ascendante, ce que semblait déjà montrer la Figure 6.11). Dans tous les cas, pour un horizon de 2 jours, *SERENADE-Ca* est dépassé par au moins le modèle de persistance pour toutes les phases du cycle solaire et pour toutes les métriques, sauf le *TDM*, qui nous rappelle que le modèle de persistance sera toujours en retard, ce qui est un grand défaut en conditions opérationnelles.

Les figures montrant les valeurs des métriques de classification sur l'ensemble d'évaluation complet (Figure B.1) et uniquement en phase descendante pour les quatre horizons testés ici (Figure B.2) sont données à titre indicatif en Annexe B. Nous ne les commentons pas ici car elles ne nous apprennent rien de plus que ce qui vient d'être établi, à savoir que *SERENADE-Ca* se comporte de façon similaire à *SERENADE*, mais que pour des horizons plus courts le modèle de persistance est le meilleur.

6.4 Estimation de l'influence de la période utilisée sur les performances

Comme nous l'avons mentionné à maintes reprises, le cycle solaire 24 a été d'intensité très faible comparé au cycle solaire précédent. Ce fait était une des raisons qui nous a poussés à avoir recours à la méthode d'évaluation croisée imbriquée. Cela nous a également amenés à changer les valeurs de seuil utilisées pour la classification binaire de $Ca_{\max,24h}$. Intuitivement, le fait d'entraîner un modèle avec des données peu représentatives de l'ensemble de la dynamique observable au niveau des activités solaires et géomagnétiques devrait être un des principaux facteurs limitant la capacité de généralisation de notre modèle à des nouvelles données (dégradant ainsi la performance qu'on pourrait attendre). C'est pourquoi il serait intéressant d'avoir une première estimation quantitative simple de la dégradation de la performance imputable à la période utilisée pour entraîner le modèle.

Pour ce faire, nous avons repris le modèle proposé dans le Chapitre 3 dédié à la prédiction à court terme de l'indice *Dst*. Nous l'avons entraîné à nouveau, mais en changeant les périodes utilisées pour l'entraînement et la validation. Ainsi, nous l'avons entraîné (pour un horizon de 1 et 3 h) en utilisant uniquement les données mesurées entre 2010 et 2017 inclus (contre 2003 à 2018 avant) et validés avec les données mesurées en 2018 (contre 1995 et 1996 avant). Nous avons ensuite évalué ces modèles en utilisant la même période d'évaluation que celle déjà utilisée dans les chapitres précédents, à savoir la période 1997 – 2002 (inclus). Nous avons mesuré les valeurs de *NRMSE* et de corrélation linéaire obtenues avec les deux ensembles d'entraînement, en distinguant les périodes calmes ($Dst > -50$ nT), les périodes d'orages modérés (-150 nT $< Dst \leq -50$ nT) et les périodes d'orages intenses ($Dst \leq -150$ nT). Les écarts relatifs de performances obtenues avec le nouvel entraînement par rapport à l'ancien sont donnés dans le Tableau 6.1.

On constate d'emblée que le fait d'entraîner le modèle uniquement avec des

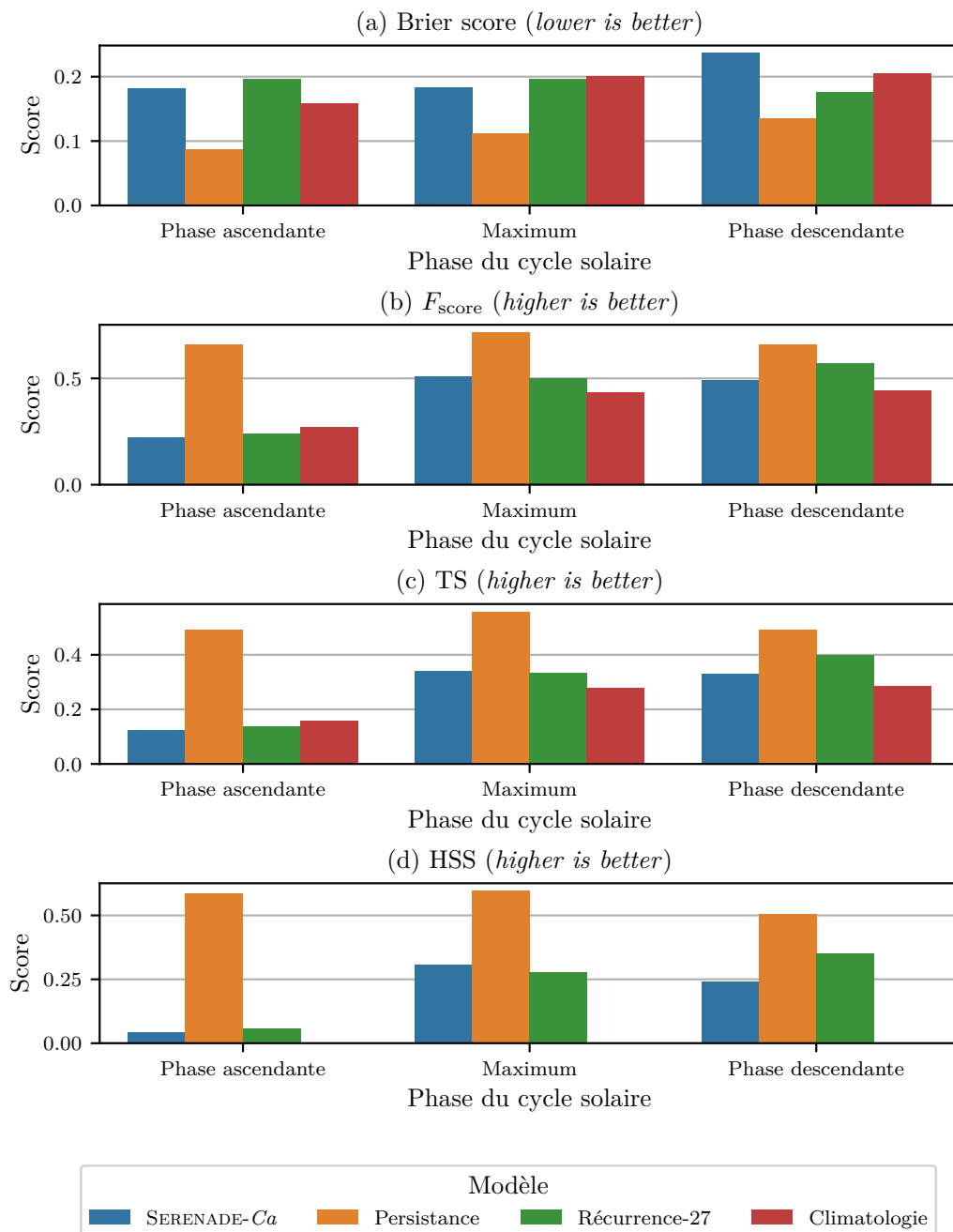


FIGURE 6.16 – Évaluation de *SERENADE- Ca* et des modèles de référence en fonction des phases du cycle solaire pour un horizon de prédiction de 2 jours à l'aide de métriques dédiées à l'évaluation du problème de classification binaire.

TABLEAU 6.1 – Comparaison des performances obtenues par le modèle de prédiction à 1 h de Dst lorsqu'on entraîne le modèle uniquement avec les données observées entre 2010 et 2018, par rapport aux performances obtenues lorsqu'on entraîne avec les données observées entre 2003 et 2018. On distingue les périodes calmes des périodes d'orages (modérés et intenses). Un pourcentage négatif indique une dégradation de la valeur de la métrique (augmentation pour $NRMSE$ et baisse pour R).

Horizon (h)	Écart relatif de $NRMSE$			Écart relatif de R		
	Période calme	Orages modérés	Orages intenses	Période calme	Orages modérés	Orages intenses
1	0 %	-3,5 %	-90,0 %	0 %	-0,2 %	-13,8 %
3	-0,2 %	+0,8 %	-27,4 %	0 %	-1,5 %	-37,6 %

données mesurées au cours de la période 2010 – 2018 ne change presque rien en période calme ou d'orage modéré, mais dégrade très fortement les performances pour les périodes d'orages intenses. Par exemple, pour un horizon de 1 h, la $NRMSE$ passe de 0,461 à 0,876 (ce qui représente une dégradation d'environ -90,0 %). Ces écarts relatifs, uniquement imputables à l'ensemble de données utilisé pour entraîner et valider le modèle, montre bien qu'un modèle qui a été entraîné uniquement sur des périodes peu intenses aura du mal à généraliser son apprentissage à des périodes de perturbations plus fortes.

Ainsi, bien que ces chiffres ne soient pas transposables directement au cas de $SERENADE$ et de $SERENADE-Ca$, on peut décemment émettre l'hypothèse que la période utilisée pour les entraîner est un des principaux facteurs limitant leurs performances.

6.5 Bilan du chapitre

Nous avons montré que la méthodologie déployée avec $SERENADE$ permet effectivement d'obtenir des prévisions de la géoeffectivité de l'activité solaire. Nous avons notamment démontré qu'une approche utilisant uniquement de l'imagerie solaire et des réseaux de neurones pour la prédiction d'indices géomagnétiques est viable, ce qui n'avait jamais été publié. Nos résultats sont d'autant plus encourageants que ce modèle est une première version, pour laquelle de nombreux compromis ont été faits. Pourtant notre modèle parvient déjà, dans le cas de l'indice Kp , à rivaliser, voire à dépasser le modèle empirique de récurrence à 27 jours, qui était jusqu'ici le modèle proposant les meilleures prédictions d'indices géomagnétiques pour des horizons supérieurs à une journée.

En analysant les résultats en fonction de la phase du cycle solaire, un faisceau concordant d'indices nous amène à penser que notre modèle a principalement

appris à modéliser les perturbations prenant leurs origines dans les trous coronaux. Afin d'améliorer le modèle, un effort spécifique sera nécessaire pour assurer une meilleure prise en compte des CMEs. Enfin, si on souhaite améliorer le modèle adapté à la prédiction de Ca , il faudra porter une attention particulière à mieux gérer la question de la forte auto-corrélation temporelle de cet indice.

Notre modèle est limité par deux principaux aspects, à savoir son architecture non-optimisée pour la performance pure (à cause des compromis mentionnés dans le chapitre précédent) et les données utilisées pour l'entraîner (ce qui inclut la faible intensité du cycle solaire 24). Nous proposons, en les justifiant, de nombreuses pistes d'amélioration de SERENADE dans les conclusions de ce manuscrit.

Conclusions

Bilan général de nos travaux

Au cours de nos travaux nous avons étudié divers axes pour l'amélioration des capacités de prévision des perturbations géomagnétiques d'origine solaire, en nous appuyant principalement sur des méthodes relevant de l'apprentissage profond.

Tout d'abord, nous avons étudié la question de la prédiction d'indices géomagnétiques à partir de paramètres du vent solaire mesurés en L1. En partant de l'état de l'art proposé par l'étude de GRUET et al. 2018, nous avons développé un nouveau modèle de prédiction de l'indice *Dst*. Ce modèle s'appuie notamment sur un réseau de neurones possédant des couches récurrentes de type LSTM. Nous l'avons construit en cherchant à optimiser chaque composante du modèle de base, en y incluant des connaissances issues de la physique du couplage vent solaire–magnétosphère et en exploitant les plus récentes avancées dans le domaine de l'apprentissage profond – dont l'exploitation a également été rendue possible grâce à l'augmentation de nos moyens de calcul. À l'aide d'un banc d'évaluation (que nous avons conçu pour être plus complet que la simple utilisation de quelques métriques classiques comme la RMSE et la corrélation linéaire) nous montrons que notre modèle est plus précis, surtout en périodes d'orages intenses. Nous montrons également que notre modèle parvient à mieux faire la distinction entre les périodes calmes et les périodes d'orages. En particulier, avec un horizon d'une heure, nous prédisons correctement 20 % d'orages en plus par rapport au modèle de base.

Nous avons également proposé une modification de notre modèle permettant de rendre ses prévisions probabilistes : au lieu de prédire une valeur scalaire, nous prédisons une distribution de probabilité gaussienne. D'autres approches utilisant notamment des processus gaussiens ou *Gaussian Processes* avaient déjà été proposées pour atteindre cet objectif mais notre méthode a l'avantage de présenter un surcoût informatique négligeable. À l'aide de métriques comme le CRPS nous montrons que notre méthode permet de rendre encore plus compétent notre modèle déterministe. Ainsi, nous avons pu améliorer la précision du modèle de prévision de *Dst* par rapport à l'état de l'art. Néanmoins nous avons également montré avec le *Temporal Distortion Mix* (TDM) que notre modèle (et le modèle de base) avait tendance à présenter systématiquement un retard par rapport aux observations. Les performances de notre modèle se dégradent notablement avec l'augmentation de l'horizon de prévision, si bien qu'il nous est difficile de dire qu'il reste utilisable dans un contexte opérationnel pour les horizons de prédiction supérieurs à 6 h (voire même en deçà de cette limite, en fonction des critères de performance qu'un opérateur pourrait imposer). Notre modèle possède alors deux

principales limitations. La première réside dans la nature de l'indice géomagnétique utilisé, *Dst*, dont la corrélation avec la géoeffectivité des événements du point de vue des ceintures de radiation est faible. La seconde est la limite de l'horizon de prévision, qui réduit fortement l'utilité du modèle.

Afin de rendre notre modèle plus utile dans un contexte d'étude de la dynamique des ceintures de radiations électroniques terrestres, nous avons conçu (d'après une idée préalablement proposée par ROCHEL et al. 2016) et étudié un nouvel indice géomagnétique, *Ca*. Il s'agit d'un indice géomagnétique intégré dans le temps fondé sur l'indice *aa*. Cela permet notamment une meilleure prise en compte de phénomènes causant une accumulation d'énergie dans les ceintures de radiations, comme les *SIRs* ou bien les *ICMEs* multiples. En l'étudiant statistiquement nous avons montré que *Ca* est généralement bien corrélé avec les flux d'électrons dans les ceintures de radiations. De plus, en exploitant son historique de 150 ans de données, nous l'avons étudié par le prisme de l'analyse des événements extrêmes. Cela nous a permis de caractériser sa queue de distribution et de montrer que nous avons déjà connu des événements extrêmes ($Ca \geq 67$ nT) depuis le début de l'ère spatiale, incluant des événements cinquantennaux et centennaux. En revanche, aucun événement extrême n'a été observé au cours du dernier cycle solaire (BERNOUX et MAGET 2020).

Nous avons adapté le modèle de prédiction de *Dst* pour la prédiction de *Ca* en contexte opérationnel. Nous avons montré qu'une approche fondée sur l'utilisation exclusive de paramètres du vent solaire donne de bons résultats pour la prévision de *Ca*, qui pourrait être utilisé à l'avenir pour constituer un système d'alertes relatif à l'état de remplissage des ceintures de radiations. En revanche, à l'aide d'un banc d'évaluation complet, nous avons montré que l'horizon de prévision efficace reste cantonné à des valeurs inférieures à 24 h (BERNOUX et al. 2021).

Pour résoudre ce problème nous avons étudié la faisabilité d'une approche exploitant uniquement et conjointement de l'imagerie solaire et des réseaux de neurones. Nous avons conçu un modèle, *SERENADE*, dont l'architecture est construite autour de trois modules constitués de couches de réseaux de neurones de différentes natures (*CNNs*, *LSTMs*, couches linéaires, etc.), afin d'extraire l'information contenue dans des séquences de 10 images du Soleil (espacées d'environ 12 h) à la longueur d'onde 193 Å. Ce modèle a été conçu en prenant en compte les contraintes imposées par notre environnement de développement (puissance de calcul, capacité de stockage, etc.). Nous avons entraîné ce modèle avec un *pipeline* dédié à la prévision des indices $Kp_{\max,24h}$ et $Ca_{\max,24h}$, pour des horizons de prévision allant de 2 à 7 jours (9 jours pour $Ca_{\max,24h}$) d'avance. Afin de l'entraîner et l'évaluer de la manière la plus complète et juste possible, nous avons eu recours à la méthode d'évaluation croisée imbriquée. Nous avons évalué encore une fois notre modèle avec un banc de test complet, adapté de celui de nos précédents modèles, et nous avons comparé les performances obtenues à celles produites par trois modèles de référence empiriques, parmi lesquels le modèle de récurrence à 27 jours, actuellement réputé être le meilleur modèle pour la prévision de *Kp* avec un horizon supérieur à 24 h.

Nous avons montré que *SERENADE* parvient à apprendre – au moins partielle-

ment – à prédire la géoefficacité des interactions Soleil–Terre uniquement à partir d’images EUV à 193 Å, ce qui n’avait jamais été démontré auparavant. Cela révèle qu’une approche fondée sur des réseaux de neurones est viable et peut être un complément efficace aux modèles physiques. De plus, nous avons observé que dans le cas de $Kp_{\max,24h}$, SERENADE propose des performances meilleures que deux des trois modèles de référence, et des performances équivalentes au modèle de récurrence. En particulier, nous avons montré que SERENADE est meilleur que le modèle de récurrence lors de la phase descendante du cycle solaire, au cours desquelles une majorité d’événements géoeffectifs sont causés par des SIRs. Lorsqu’on prend en compte le fait que nous utilisons des images du Soleil espacées de 12 h, ce résultat n’est pas surprenant et suggère qu’une amélioration de SERENADE passerait notamment par l’utilisation de nouvelles données permettant une meilleure prise en compte des CMEs.

Ainsi, en nous penchant sur les trois axes d’études identifiés en introduction de ce manuscrit, nous avons proposé différentes améliorations ou pistes d’améliorations de nos capacités de prédiction de l’activité géomagnétique à l’aide de méthodes d’apprentissage profond. Nos résultats confirment les conclusions d’études précédentes et apportent de nouveaux éléments montrant la pertinence de ces approches, en complément des approches empiriques et physiques utilisées actuellement en météorologie de l’espace. Néanmoins, nous avons identifié des limites à nos modèles, qui pourraient faire l’objet d’études subséquentes. Nous en évoquons quelques-unes ci-dessous.

Perspectives scientifiques et techniques

À propos des modèles s’appuyant sur les paramètres du vent solaire

En premier lieu, il serait très certainement possible d’améliorer encore les modèles de prévision d’indices géomagnétiques à partir des paramètres du vent solaire. Une des pistes les plus évidentes consiste à ajouter des nouvelles sources de données en entrée du modèle, qui viennent compléter l’information déjà contenue dans les paramètres du vent solaire. C’est une proposition faite par exemple par GRUET et al. 2018 (données GPS), CHAKRABORTY et MORLEY 2020 (rayons X mesurés par les satellites GOES) ou encore TASISTRO-HART et al. 2021 (caractéristiques de CMEs), qui montrent que l’utilisation de ces données supplémentaires est effectivement pertinente. Nous n’avons pas abordé cette question au cours de nos travaux car nous avons préféré nous concentrer sur la question de l’élargissement de l’horizon de prédiction.

Une autre piste d’amélioration de ces modèles consiste en l’utilisation de nouvelles architectures de réseaux de neurones, comme les réseaux de type *Transformers*, ayant émergé au cours des cinq dernières années (VASWANI et al. 2017), qui démontrent leur grande efficacité à résoudre des problèmes pour lesquels les réseaux de type LSTMs et CNNs étaient jusqu’ici préférés. Néanmoins ces réseaux

nécessitent une grande quantité de données pour être correctement entraînés, il n'est donc pas garanti que leur utilisation soit pertinente dans notre cas : une étude dédiée pourrait répondre à cette question.

À propos de SERENADE

Les possibilités d'amélioration du modèle [SERENADE](#) sont sans doute encore plus nombreuses et de natures multiples. Tout d'abord, maintenant qu'un grand effort a déjà été mené pour optimiser l'architecture et ajuster les hyperparamètres du modèle, on peut envisager l'utilisation d'un plus grand nombre d'images, avec un écart temporel bien plus resserré (inférieur à 1 h contre 12 h actuellement). Pour que l'apprentissage se fasse efficacement, cela nécessiterait sans doute l'utilisation d'un plus grand nombre d'échantillons, provenant d'un cycle solaire de plus forte intensité. Pour cela, il est envisagé d'utiliser des images provenant de [SOHO](#) en complément ou en remplacement des données [SDO](#).

En restant dans le domaine des améliorations liées aux données utilisées, on peut envisager l'utilisation d'autres types d'images, comme des images [EUV](#) à d'autres longueurs d'onde (en supplément des images à 193 Å), ou encore des images de nature différente, comme des magnétogrammes ou des coronographes. Par ailleurs, il est également envisageable de rajouter aux données d'imagerie solaire d'autres données, à commencer par l'historique des valeurs des indices géomagnétiques (ce qui pourrait nous permettre d'exploiter l'auto-corrélation persistante et/ou récurrente des indices dans notre modèle).

Ces modifications étant difficilement faisables avec le Feature Extractor actuel (un réseau GoogLeNet pré-entraîné), il faudrait sans doute également le modifier, en construisant un module *ad-hoc* que nous entraînerions, ou bien en utilisant un réseau pré-entraîné, mais sur des images du Soleil et non des images de la base de données ImageNet (on peut imaginer exploiter partiellement ou entièrement un modèle pré-entraîné pour créer des instruments virtuels parmi ceux mentionnés en Section 5.1).

Enfin, une des principales pistes à explorer pour mieux comprendre (et donc améliorer) les prévisions fournies par le modèle réside dans l'utilisation de méthodes pour interpréter et expliquer le fonctionnement des réseaux de neurones. En effet, le principal défaut de l'utilisation des réseaux de neurones est leur fonctionnement en « boîte noire », qui rend difficile, voire impossible, pour l'utilisateur le fait de contrôler et de comprendre les décisions effectuées par le modèle. Pour résoudre ce problème, diverses méthodes dites d'[IA explicable](#) ou *explainable AI* (IA) ont été proposées au cours des dernières années ([RIBEIRO et al. 2016](#) ; [ARRIETA et al. 2019](#)). Certaines méthodes sont dédiées à l'explication de décisions prises par un classifieur à partir d'une ou plusieurs image ([MANTTARI et al. 2020](#)), ou par un régresseur à partir de séries temporelles de faible dimension ([ROJAT et al. 2021](#)). Néanmoins, peu de solutions « sur étagère » existent à ce jour pour expliquer des régresseurs dont les décisions sont prises à partir de plusieurs images (comme c'est le cas de [SERENADE](#)). C'est pourquoi un effort substantiel – mais nécessaire

– serait nécessaire pour interpréter et expliquer plus en détail le fonctionnement de **SERENADE**, afin notamment de vérifier que ses décisions sont fondées sur des critères pertinents du point de vue de la physique solaire.

À propos du déploiement en contexte opérationnel et des approches « hybrides »

Outre l'amélioration des performances brutes de ces modèles, nous pensons qu'il faudrait maintenant également porter l'effort sur l'intégration de ces modèles (utilisant vent solaire et imagerie solaire) dans des environnements opérationnels (comme ceux proposés par des agences spatiales comme l'**ESA** avec son *Space Weather Service Network*).

Pour les modèles s'appuyant sur les paramètres du vent solaire cela passe en priorité par une meilleure quantification de leur comportement et de leur capacité à généraliser, notamment en cas d'événement extrême en dehors des seuils avec lesquels le modèle a été entraîné. Une étude de l'intégration de ce type de modèles au sein d'architectures plus globales incluant plusieurs modèles serait également à considérer avec attention. La question du « chaînage » de modèles physiques (**MHD**) de propagation du vent solaire avec des réseaux de neurones pour la prédiction d'indices géomagnétiques a déjà été l'objet d'études, par exemple dans le cadre du projet européen **SAFE SPACE**¹. Des premiers résultats montrent que la gestion des incertitudes (notamment sur les paramètres d'entrée) est un des facteurs limitant le plus cette approche « hybride ». Ces méthodes nécessitent également de faire fonctionner les modèles **MHD** plusieurs fois pour obtenir des prévisions d'ensemble (sans lesquelles ils ne peuvent fournir que des prédictions déterministes), ce qui peut se révéler assez long, surtout dans un contexte opérationnel. On imagine aisément que, dans ce cadre, **SERENADE** pourrait servir de modèle de remplacement (*surrogate model*) permettant d'obtenir une première estimation de la prédiction (probabiliste) bien plus rapidement.

Néanmoins, il est sans aucun doute pertinent de continuer à explorer la question de l'utilisation de méthodes « hybrides » physique-**IA**, dans l'optique d'une prise en compte accrue de nos connaissances sur la physique du système Soleil-Terre au sein de modèles exploitant la puissance inférentielle des réseaux de neurones. Au cours de cette thèse nous avons cherché à étudier la faisabilité d'un modèle constitué d'un réseau de neurones entraîné pour prédire des indices géomagnétiques non-pas à partir de paramètres du vent solaire mesurés en L1, mais à partir de paramètres du vent solaire modélisés par le modèle **SWIFT** proposé par l'**Institut de Recherche en Astrophysique et Planétologie (IRAP)**. Notons bien qu'ici notre souhait était d'entraîner le réseau sur les sorties du modèle **SWIFT** (ce qui pourrait permettre au modèle d'apprendre à corriger certains biais complexes, mais systématiques, du modèle **SWIFT**), et non pas uniquement de les utiliser pour produire des prévisions une fois l'entraînement fait sur les données de la base **OMNIWeb** (comme évoqué dans le paragraphe

1. <https://www.safespace-h2020.eu/>

précédent). Malheureusement cette étude n'a pu aboutir au cours de nos travaux, pour des questions de calendrier (beaucoup de données issues des simulations **SWIFT** étaient nécessaires). Néanmoins cette piste étant prometteuse, il serait intéressant de la poursuivre à l'avenir et de comparer les performances atteignables avec un tel modèle hybride avec les performances de **SERENADE**. Dans tous les cas, parvenir à introduire plus de connaissances physiques préalables au sein des modèles d'apprentissage automatique sera certainement une des clefs permettant d'en améliorer les performances prédictives.

Bibliographie

- ACERO, F. J. et al. (2018). « A Limit for the Values of the Dst Geomagnetic Index ». In : *Geophysical Research Letters* 45.18, p. 9435-9440. ISSN : 1944-8007. DOI : [10.1029/2018GL079676](https://doi.org/10.1029/2018GL079676) (cf. p. 95).
- AHMADZADEH, A., D. J. KEMPTON et R. A. ANGRYK (2019). « A Curated Image Parameter Data Set from the Solar Dynamics Observatory Mission ». In : *The Astrophysical Journal Supplement Series* 243.1, p. 18. ISSN : 0067-0049. DOI : [10.3847/1538-4365/ab253a](https://doi.org/10.3847/1538-4365/ab253a) (cf. p. 182).
- AHMED, O. et al. (2013). « Solar Flare Prediction Using Advanced Feature Extraction, Machine Learning, and Feature Selection ». In : *Solar Physics* 283.1, p. 157-175. DOI : [10.1007/s11207-011-9896-1](https://doi.org/10.1007/s11207-011-9896-1) (cf. p. 173).
- AKASOFU, S. I. (1981). « Energy coupling between the solar wind and the magnetosphere ». In : *Space Science Reviews* 28.2, p. 121-190. ISSN : 1572-9672. DOI : [10.1007/BF00218810](https://doi.org/10.1007/BF00218810) (cf. p. 110).
- ALKEN, P. et al. (2021). « International Geomagnetic Reference Field : the thirteenth generation ». In : *Earth, Planets and Space* 73.1, p. 49. ISSN : 1880-5981. DOI : [10.1186/s40623-020-01288-x](https://doi.org/10.1186/s40623-020-01288-x) (cf. p. 22).
- AMODEI, D. et al. (2015). « Deep Speech 2 : End-to-End Speech Recognition in English and Mandarin ». In : *arXiv :1512.02595 [cs]*. arXiv : [1512.02595](https://arxiv.org/abs/1512.02595) (cf. p. 59).
- ARGE, C. N. et V. J. PIZZO (2000). « Improvement in the prediction of solar wind conditions using near-real time solar magnetic field updates ». In : *Journal of Geophysical Research : Space Physics* 105 (A5), p. 10465-10479. ISSN : 2156-2202. DOI : [10.1029/1999JA000262](https://doi.org/10.1029/1999JA000262) (cf. p. 86).
- ARGE, C. N. et al. (2010). « Air Force Data Assimilative Photospheric Flux Transport (ADAPT) Model ». In : *AIP Conference Proceedings* 1216.1, p. 343-346. ISSN : 0094-243X. DOI : [10.1063/1.3395870](https://doi.org/10.1063/1.3395870) (cf. p. 76).
- ARRIETA, A. B. et al. (2019). « Explainable Artificial Intelligence (XAI) : Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI ». In : *arXiv :1910.10045 [cs]*. arXiv : [1910.10045](https://arxiv.org/abs/1910.10045) (cf. p. 236).
- ASVESTARI, E. et al. (2021). « Modelling a multi-spacecraft coronal mass ejection encounter with EUHFORIA ». In : *Astronomy & Astrophysics* 652, A27. ISSN : 0004-6361, 1432-0746. DOI : [10.1051/0004-6361/202140315](https://doi.org/10.1051/0004-6361/202140315) (cf. p. 177).
- AULANIER, G. (2013). « The physical mechanisms that initiate and drive solar eruptions ». In : *Proceedings of the International Astronomical Union* 8 (S300), p. 184-196. ISSN : 1743-9213, 1743-9221. DOI : [10.1017/S1743921313010958](https://doi.org/10.1017/S1743921313010958) (cf. p. 17).
- AYALA SOLARES, J. R. et al. (2016). « Modeling and prediction of global magnetic disturbance in near-Earth space : A case study for Kp index using NARX models ». In : *Space Weather* 14.10, p. 899-916. ISSN : 1542-7390. DOI : [10.1002/2016SW001463](https://doi.org/10.1002/2016SW001463) (cf. p. 167).
- BACCOUCHE, M. et al. (2011). « Sequential Deep Learning for Human Action Recognition ». In : *Human Behavior Understanding*. Sous la dir. d'A. A. SALAH et B. LEPRI.

- Lecture Notes in Computer Science. Berlin, Heidelberg : Springer, p. 29-39. ISBN : 978-3-642-25446-8. DOI : [10.1007/978-3-642-25446-8_4](https://doi.org/10.1007/978-3-642-25446-8_4) (cf. p. 61).
- BAKER, D. N. (1998). « What is space weather? » In : *Advances in Space Research. Solar-Terrestrial Relations : Predicting the Effects on the Near-Earth Environment* 22.1, p. 7-16. ISSN : 0273-1177. DOI : [10.1016/S0273-1177\(97\)01095-8](https://doi.org/10.1016/S0273-1177(97)01095-8) (cf. p. 35).
- BAKER, D. N. et al. (1996). « Neutral line model of substorms : Past results and present view ». In : *Journal of Geophysical Research : Space Physics* 101 (A6), p. 12975-13010. ISSN : 2156-2202. DOI : [10.1029/95JA03753](https://doi.org/10.1029/95JA03753) (cf. p. 31).
- BAKER, D. N. et M. I. PANASYUK (2017). « Discovering Earth's radiation belts ». In : *Physics Today* 70.12, p. 46-51. ISSN : 0031-9228. DOI : [10.1063/PT.3.3791](https://doi.org/10.1063/PT.3.3791) (cf. p. 27).
- BALA, R. et P. REIFF (2012). « Improvements in short-term forecasting of geomagnetic activity ». In : *Space Weather* 10.6. ISSN : 1542-7390. DOI : [10.1029/2012SW000779](https://doi.org/10.1029/2012SW000779) (cf. p. 88, 95, 111).
- BALDI, P. et P. J. SADOWSKI (2013). « Understanding dropout ». In : *Advances in neural information processing systems* 26, p. 2814-2822 (cf. p. 64, 108).
- BARTELS, J. (1949). « The standardized index, Ks, and the planetary index, Kp ». In : *IATME Bull* 97.12, p. 0001 (cf. p. 80).
- BARTELS, J., N. H. HECK et H. F. JOHNSTON (1939). « The three-hour-range index measuring geomagnetic activity ». In : *Terrestrial Magnetism and Atmospheric Electricity* 44.4, p. 411-454. ISSN : 0096-8013. DOI : [10.1029/TE044i004p00411](https://doi.org/10.1029/TE044i004p00411) (cf. p. 79).
- BAUDIN, M. et al. (2015). « Open TURNS : An industrial software for uncertainty quantification in simulation ». In : *arXiv :1501.05242 [math, stat]*. arXiv : [1501.05242](https://arxiv.org/abs/1501.05242) (cf. p. 264).
- BELLMAN, R. E. (2003). *Dynamic Programming*. Courier Corporation. 388 p. ISBN : 978-0-486-42809-3 (cf. p. 173).
- BELMONT, G. et al. (2013). *Collisionless Plasmas in Astrophysics*. ISBN : 978-3-527-41074-3. Wiley (cf. p. 28).
- BENACQUISTA, R., S. ROCHEL et G. ROLLAND (2017). « Understanding the variability of magnetic storms caused by ICMEs ». In : *Annales Geophysicae* 35.1, p. 147-159. ISSN : 0992-7689. DOI : <https://doi.org/10.5194/angeo-35-147-2017> (cf. p. 19, 141).
- BENGIO, Y. (2012). « Practical recommendations for gradient-based training of deep architectures ». In : *arXiv :1206.5533 [cs]*. arXiv : [1206.5533](https://arxiv.org/abs/1206.5533) (cf. p. 50).
- BERGHMANS, D. et al. (2006). « SWAP onboard PROBA 2, a new EUV imager for solar monitoring ». In : *Advances in Space Research. Magnetospheric dynamics and the international living with a star program* 38.8, p. 1807-1811. ISSN : 0273-1177. DOI : [10.1016/j.asr.2005.03.070](https://doi.org/10.1016/j.asr.2005.03.070) (cf. p. 71).
- BERKSON, J. (1944). « Application of the Logistic Function to Bio-Assay ». In : *Journal of the American Statistical Association* 39.227, p. 357-365. ISSN : 0162-1459. DOI : [10.1080/01621459.1944.10500699](https://doi.org/10.1080/01621459.1944.10500699) (cf. p. 41).
- BERNOUX, G. et V. MAGET (2020). « Characterizing Extreme Geomagnetic Storms Using Extreme Value Analysis : A Discussion on the Representativeness of Short Data Sets ». In : *Space Weather* 18.6, e2020SW002450. ISSN : 1542-7390. DOI : [10.1029/2020SW002450](https://doi.org/10.1029/2020SW002450) (cf. p. 138, 142-144, 147, 149, 151, 234, 267).

-
- BERNOUX, G. et al. (2021). « An operational approach to forecast the Earth's radiation belts dynamics ». In : *Journal of Space Weather and Space Climate* 11, p. 60. ISSN : 2115-7251. DOI : [10.1051/swsc/2021045](https://doi.org/10.1051/swsc/2021045) (cf. p. 138, 156, 158-160, 162-165, 234, 267).
- BERNOUX, G. et al. (2022a). « Forecasting the Geomagnetic Activity Several Days in Advance Using Neural Networks Driven by Solar EUV Imaging ». In : *Journal of Geophysical Research : Space Physics* 127.10, e2022JA030868. ISSN : 2169-9402. DOI : [10.1029/2022JA030868](https://doi.org/10.1029/2022JA030868) (cf. p. 171, 268).
- BERNOUX, G. et al. (2022b). *Probabilistic Forecasts of the Daily Maximum of the Kp Index Produced by the SERENADE Prototype Model and Three Empirical Models for the Period 2010-2018*. DOI : [10.5281/zenodo.6873690](https://doi.org/10.5281/zenodo.6873690) (cf. p. 206).
- BEUTIER, T. (1993). « Modélisation tridimensionnelle pour l'étude de la dynamique des ceintures de radiation ». These de doctorat. École nationale supérieure de l'aéronautique et de l'espace (Toulouse ; 1972-2007) (cf. p. 89).
- BLAKE, J. B. et al. (2013). « The Magnetic Electron Ion Spectrometer (MagEIS) Instruments Aboard the Radiation Belt Storm Probes (RBSP) Spacecraft ». In : *Space Science Reviews* 179.1, p. 383-421. ISSN : 1572-9672. DOI : [10.1007/s11214-013-9991-8](https://doi.org/10.1007/s11214-013-9991-8) (cf. p. 141).
- BOBERG, F., P. WINTOFT et H. LUNDSTEDT (2000). « Real time Kp predictions from solar wind data using neural networks ». In : *Physics and Chemistry of the Earth, Part C : Solar, Terrestrial & Planetary Science* 25.4, p. 275-280. ISSN : 1464-1917. DOI : [10.1016/S1464-1917\(00\)00016-7](https://doi.org/10.1016/S1464-1917(00)00016-7) (cf. p. 87).
- BOLDUC, L. (2002). « GIC observations and studies in the Hydro-Québec power system ». In : *Journal of Atmospheric and Solar-Terrestrial Physics. Space Weather Effects on Technological Systems* 64.16, p. 1793-1802. ISSN : 1364-6826. DOI : [10.1016/S1364-6826\(02\)00128-1](https://doi.org/10.1016/S1364-6826(02)00128-1) (cf. p. 33).
- BOROVSKY, J. E. (2020a). « A survey of geomagnetic and plasma time lags in the solar-wind-driven magnetosphere of earth ». In : *Journal of Atmospheric and Solar-Terrestrial Physics* 208, p. 105376. ISSN : 1364-6826. DOI : [10.1016/j.jastp.2020.105376](https://doi.org/10.1016/j.jastp.2020.105376) (cf. p. 81, 109).
- BOROVSKY, J. E. (2020b). « What magnetospheric and ionospheric researchers should know about the solar wind ». In : *Journal of Atmospheric and Solar-Terrestrial Physics* 204, p. 105271. ISSN : 1364-6826. DOI : [10.1016/j.jastp.2020.105271](https://doi.org/10.1016/j.jastp.2020.105271) (cf. p. 19, 20).
- BOROVSKY, J. E. et M. H. DENTON (2006). « Differences between CME-driven storms and CIR-driven storms ». In : *Journal of Geophysical Research : Space Physics* 111 (A7). ISSN : 2156-2202. DOI : [10.1029/2005JA011447](https://doi.org/10.1029/2005JA011447) (cf. p. 31, 95).
- BOROVSKY, J. E. et M. H. DENTON (2010). « Solar wind turbulence and shear : A superposed-epoch analysis of corotating interaction regions at 1 AU ». In : *Journal of Geophysical Research : Space Physics* 115 (A10). ISSN : 2156-2202. DOI : [10.1029/2009JA014966](https://doi.org/10.1029/2009JA014966) (cf. p. 19).
- BOROVSKY, J. E. et Y. Y. SHPRITS (2017). « Is the Dst Index Sufficient to Define All Geospace Storms? ». In : *Journal of Geophysical Research : Space Physics* 122.11, p. 11, 543-11, 547. ISSN : 2169-9402. DOI : [10.1002/2017JA024679](https://doi.org/10.1002/2017JA024679) (cf. p. 81, 139).
- BOTHMER, V. et I. A. DAGLIS (2007). *Space Weather : Physics and Effects*. Springer Science & Business Media. 517 p. ISBN : 978-3-540-34578-7 (cf. p. 35).
- BOURDARIE, S. A. et V. F. MAGET (2012). « Electron radiation belt data assimilation with an ensemble Kalman filter relying on the Salammbô code ». In : *Annales*

- Geophysicae* 30.6, p. 929-943. ISSN : 1432-0576. DOI : [10.5194/angeo-30-929-2012](https://doi.org/10.5194/angeo-30-929-2012) (cf. p. 82, 139).
- BOURDARIE, S. (1996). « Modélisation de la dynamique des ceintures de radiations à l'échelle de l'orage magnétique ». These de doctorat. Toulouse, ENSAE (cf. p. 21, 89).
- BRADLEY, A. P. (1997). « The use of the area under the ROC curve in the evaluation of machine learning algorithms ». In : *Pattern Recognition* 30.7, p. 1145-1159. ISSN : 0031-3203. DOI : [10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2) (cf. p. 133).
- BRUECKNER, G. E. et al. (1995). « The Large Angle Spectroscopic Coronagraph (LASCO) ». In : *Solar Physics* 162.1, p. 357-402. ISSN : 1573-093X. DOI : [10.1007/BF00733434](https://doi.org/10.1007/BF00733434) (cf. p. 68, 69).
- BRUNET, A. et al. (2021). « OMEP-EOR : A MeV proton flux specification model for electric orbit raising missions ». In : *Journal of Space Weather and Space Climate* 11, p. 55. ISSN : 2115-7251. DOI : [10.1051/swsc/2021038](https://doi.org/10.1051/swsc/2021038) (cf. p. 89).
- BUDUMA, N. et N. LOCASCIO (2017). *Fundamentals of Deep Learning : Designing Next-generation Machine Intelligence Algorithms*. O'Reilly Media. 283 p. ISBN : 978-1-4919-2561-4 (cf. p. 43).
- BURTON, R. K., R. L. MCPHERRON et C. T. RUSSELL (1975). « An empirical relationship between interplanetary conditions and Dst ». In : *Journal of Geophysical Research (1896-1977)* 80.31, p. 4204-4214. ISSN : 2156-2202. DOI : [10.1029/JA080i031p04204](https://doi.org/10.1029/JA080i031p04204) (cf. p. 87).
- BUSSY-VIRAT, C. D. et A. J. RIDLEY (2014). « Predictions of the solar wind speed by the probability distribution function model ». In : *Space Weather* 12.6, p. 337-353. ISSN : 1542-7390. DOI : [10.1002/2014SW001051](https://doi.org/10.1002/2014SW001051) (cf. p. 86).
- BYRNE, J. (2015). « Investigating the kinematics of coronal mass ejections with the automated CORIMP catalog ». In : *Journal of Space Weather and Space Climate* 5. DOI : [10.1051/swsc/2015020](https://doi.org/10.1051/swsc/2015020) (cf. p. 173).
- CADE III, W. B. et C. CHAN-PARK (2015). « The Origin of "Space Weather" ». In : *Space Weather* 13.2, p. 99-103. ISSN : 1542-7390. DOI : [10.1002/2014SW001141](https://doi.org/10.1002/2014SW001141) (cf. p. 35).
- CALDWELL, B., E. MCCARRON et S. JONAS (2017). « An abridged history of federal involvement in space weather forecasting ». In : *Space Weather* 15.10, p. 1222-1237. ISSN : 1542-7390. DOI : [10.1002/2017SW001626](https://doi.org/10.1002/2017SW001626) (cf. p. 35).
- CAMPI, C. et al. (2019). « Feature Ranking of Active Region Source Properties in Solar Flare Forecasting and the Uncompromised Stochasticity of Flare Occurrence ». In : *Astrophysical Journal* 883.2. DOI : [10.3847/1538-4357/ab3c26](https://doi.org/10.3847/1538-4357/ab3c26) (cf. p. 173).
- CAMPOREALE, E. (2019). « The Challenge of Machine Learning in Space Weather : Nowcasting and Forecasting ». In : *Space Weather* 17.8, p. 1166-1207. ISSN : 1542-7390. DOI : [10.1029/2018SW002061](https://doi.org/10.1029/2018SW002061) (cf. p. 167).
- CAMPOREALE, E. et al. (2019). « On the Generation of Probabilistic Forecasts From Deterministic Models ». In : *Space Weather* 17.3, p. 455-475. DOI : [10.1029/2018SW002026](https://doi.org/10.1029/2018SW002026) (cf. p. 88, 128).
- CAMPOREALE, E., S. WING et J. JOHNSON (2018). *Machine Learning Techniques for Space Weather*. Elsevier. 454 p. ISBN : 978-0-12-811789-7 (cf. p. 42).
- CAUCHY, A. (1847). « Méthode générale pour la résolution des systemes d'équations simultanées ». In : 25.1847, p. 536-538 (cf. p. 48).

-
- CAWLEY, G. C. et N. L. C. TALBOT (2010). « On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation ». In : *Journal of Machine Learning Research* 11.70, p. 2079-2107. ISSN : 1533-7928 (cf. p. 62).
- CHAKRABORTY, S. et S. K. MORLEY (2020). « Probabilistic prediction of geomagnetic storms and the Kp index ». In : *Journal of Space Weather and Space Climate* 10, p. 36. ISSN : 2115-7251. DOI : [10.1051/swsc/2020037](https://doi.org/10.1051/swsc/2020037) (cf. p. 43, 88, 112, 133, 135, 136, 235).
- CHANDORKAR, M., E. CAMPOREALE et S. WING (2017). « Probabilistic forecasting of the disturbance storm time index : An autoregressive Gaussian process approach ». In : *Space Weather* 15.8, p. 1004-1019. ISSN : 1542-7390. DOI : [10.1002/2017SW001627](https://doi.org/10.1002/2017SW001627) (cf. p. 88, 111, 135).
- CHEN, Y. et al. (2019). « PreMevE : New Predictive Model for Megaelectron-Volt Electrons Inside Earth's Outer Radiation Belt ». In : *Space Weather* 17.3, p. 438-454. ISSN : 1542-7390. DOI : [10.1029/2018SW002095](https://doi.org/10.1029/2018SW002095) (cf. p. 139).
- CHI, Y. et al. (2016). « Statistical Study of the Interplanetary Coronal Mass Ejections from 1995 to 2015 ». In : *Solar Physics* 291.8, p. 2419-2439. ISSN : 1573-093X. DOI : [10.1007/s11207-016-0971-5](https://doi.org/10.1007/s11207-016-0971-5) (cf. p. 82, 157).
- CHI, Y. et al. (2018). « Geoeffectiveness of Stream Interaction Regions From 1995 to 2016 ». In : *Space Weather* 16.12, p. 1960-1971. ISSN : 1542-7390. DOI : <https://doi.org/10.1029/2018SW001894> (cf. p. 82, 157, 219).
- CHIMMULA, V. K. R. et L. ZHANG (2020). « Time series forecasting of COVID-19 transmission in Canada using LSTM networks ». In : *Chaos, Solitons & Fractals* 135, p. 109864. ISSN : 0960-0779. DOI : [10.1016/j.chaos.2020.109864](https://doi.org/10.1016/j.chaos.2020.109864) (cf. p. 59).
- CHOI, S. et al. (2012). « Application of support vector machine to the prediction of geo-effective halo CMEs ». In : *Journal of the Korean Astronomical Society* 45.2, p. 31-38. DOI : [10.5303/JKAS.2012.45.2.31](https://doi.org/10.5303/JKAS.2012.45.2.31) (cf. p. 174, 175).
- CHOLLET, F. (2021). *Deep learning with python*. Simon et Schuster (cf. p. 41, 43, 47).
- CINTO, T. et al. (2020). « A framework for designing and evaluating solar flare forecasting systems ». In : *Monthly Notices of the Royal Astronomical Society* 495.3, p. 3332-3349. DOI : [10.1093/mnras/staa1257](https://doi.org/10.1093/mnras/staa1257) (cf. p. 175).
- CLEVERT, D.-A., T. UNTERTHINER et S. HOCHREITER (2016). « Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs) ». In : *arXiv :1511.07289 [cs]*. arXiv : [1511.07289](https://arxiv.org/abs/1511.07289) (cf. p. 52).
- COLAK, T. et R. QAHWAJI (2009). « Automated solar activity prediction : A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares ». In : *Space Weather* 7.6. DOI : [10.1029/2008SW000401](https://doi.org/10.1029/2008SW000401) (cf. p. 174).
- COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. London : Springer London. DOI : [10.1007/978-1-4471-3675-0](https://doi.org/10.1007/978-1-4471-3675-0) (cf. p. 147).
- COLLADO-VILLAVERDE, A., P. MUÑOZ et C. CID (2021). « Deep Neural Networks With Convolutional and LSTM Layers for SYM-H and ASY-H Forecasting ». In : *Space Weather* 19.6, e2021SW002748. ISSN : 1542-7390. DOI : [10.1029/2021SW002748](https://doi.org/10.1029/2021SW002748) (cf. p. 88, 112).
- COMMUNITY, T. S. et al. (2020). « The SunPy Project : Open Source Development and Status of the Version 1.0 Core Package ». In : *The Astrophysical Journal* 890.1, p. 68. ISSN : 0004-637X. DOI : [10.3847/1538-4357/ab4f7a](https://doi.org/10.3847/1538-4357/ab4f7a) (cf. p. 77).

- COSTELLO, K. A. (1998). « Moving the Rice MSFM into a real-time forecast mode using solar wind driven forecast modules ». Thesis. Rice University (cf. p. 87).
- CRANMER, S. R. (2009). « Coronal Holes ». In : *Living Reviews in Solar Physics* 6.1, p. 3. ISSN : 1614-4961. DOI : [10.12942/lrsp-2009-3](https://doi.org/10.12942/lrsp-2009-3) (cf. p. 13).
- CROOKER, N. U. et G. L. SISCOE (1986). « The Effect of the Solar Wind on the Terrestrial Environment ». In : *Physics of the Sun : Volume III : Astrophysics and Solar-Terrestrial Relations*. Sous la dir. de P. A. STURROCK et al. Geophysics and Astrophysics Monographs. Dordrecht : Springer Netherlands, p. 193-249. ISBN : 978-94-009-5255-3. DOI : [10.1007/978-94-009-5255-3_6](https://doi.org/10.1007/978-94-009-5255-3_6) (cf. p. 14).
- DAHMEN, N., F. ROGIER et V. MAGET (2020). « On the modelling of highly anisotropic diffusion for electron radiation belt dynamic codes ». In : *Computer Physics Communications* 254, p. 107342. ISSN : 0010-4655. DOI : [10.1016/j.cpc.2020.107342](https://doi.org/10.1016/j.cpc.2020.107342) (cf. p. 89).
- DAHMEN, N. A. (2020). « Développement d'un nouveau coeur numérique pour le code de calcul Salammbô de modélisation des ceintures de radiation terrestres ». These de doctorat. Toulouse, ISAE (cf. p. 21, 22, 89).
- DAVIS, J. et M. GOADRICH (2006). « The relationship between Precision-Recall and ROC curves ». In : *Proceedings of the 23rd international conference on Machine learning - ICML '06*. the 23rd international conference. Pittsburgh, Pennsylvania : ACM Press, p. 233-240. ISBN : 978-1-59593-383-6. DOI : [10.1145/1143844.1143874](https://doi.org/10.1145/1143844.1143874) (cf. p. 133).
- DAVIS, T. N. et M. SUGIURA (1966). « Auroral electrojet activity index AE and its universal time variations ». In : *Journal of Geophysical Research (1896-1977)* 71.3, p. 785-801. ISSN : 2156-2202. DOI : [10.1029/JZ071i003p00785](https://doi.org/10.1029/JZ071i003p00785) (cf. p. 81).
- DELABOUDINIÈRE, J.-P. et al. (1995). « EIT : Extreme-UltraViolet Imaging Telescope for the SOHO Mission ». In : *The SOHO Mission*. Sous la dir. de B. FLECK, V. DOMINGO et A. POLAND. Dordrecht : Springer Netherlands, p. 291-312. ISBN : 978-94-009-0191-9. DOI : [10.1007/978-94-009-0191-9_8](https://doi.org/10.1007/978-94-009-0191-9_8) (cf. p. 67).
- DELOUILLE, V. et al. (2018). « Coronal holes detection using supervised classification ». In : *Machine Learning Techniques for Space Weather*, p. 365-395. DOI : [10.1016/B978-0-12-811788-0.00015-9](https://doi.org/10.1016/B978-0-12-811788-0.00015-9) (cf. p. 173, 174).
- DENG, J. et al. (2009). « ImageNet : A large-scale hierarchical image database ». In : *CVPR09* (cf. p. 185).
- DHAR, V. (2013). « Data science and prediction ». In : *Communications of the ACM* 56.12, p. 64-73. ISSN : 0001-0782. DOI : [10.1145/2500499](https://doi.org/10.1145/2500499) (cf. p. 40).
- DIMITROVA, S. et E. S. BABAYEV (2021). « Space Weather Effects on Human Health ». In : *Variability of the Sun and Sun-like Stars : from Asteroseismology to Space Weather*. EDP Sciences, p. 177-186. ISBN : 978-2-7598-2196-9. DOI : [10.1051/978-2-7598-2196-9.c012](https://doi.org/10.1051/978-2-7598-2196-9.c012) (cf. p. 35).
- DODSON, H. W. et E. R. HEDEMAN (1970). « Major H α flares in centers of activity with very small or no spots ». In : *Solar Physics* 13.2, p. 401-419. ISSN : 1573-093X. DOI : [10.1007/BF00153560](https://doi.org/10.1007/BF00153560) (cf. p. 16).
- DOMINGO, V., B. FLECK et A. I. POLAND (1995). « The SOHO mission : An overview ». In : *Solar Physics* 162.1, p. 1-37. ISSN : 1573-093X. DOI : [10.1007/BF00733425](https://doi.org/10.1007/BF00733425) (cf. p. 67).
- DUNGEY, J. W. (1961). « Interplanetary Magnetic Field and the Auroral Zones ». In : *Physical Review Letters* 6.2, p. 47-48. DOI : [10.1103/PhysRevLett.6.47](https://doi.org/10.1103/PhysRevLett.6.47) (cf. p. 30).

-
- EASTWOOD, J. P. et al. (2017). « The Scientific Foundations of Forecasting Magnetospheric Space Weather ». In : *Space Science Reviews* 212.3, p. 1221-1252. ISSN : 1572-9672. DOI : [10.1007/s11214-017-0399-8](https://doi.org/10.1007/s11214-017-0399-8) (cf. p. 30, 31, 33).
- EBERT, R. W. et al. (2009). « Bulk properties of the slow and fast solar wind and interplanetary coronal mass ejections measured by Ulysses : Three polar orbits of observations ». In : *Journal of Geophysical Research : Space Physics* 114 (A1). ISSN : 2156-2202. DOI : [10.1029/2008JA013631](https://doi.org/10.1029/2008JA013631) (cf. p. 19, 20).
- ELMAN, J. L. (1990). « Finding Structure in Time ». In : *Cognitive Science* 14.2, p. 179-211. ISSN : 1551-6709. DOI : [10.1207/s15516709cog1402_1](https://doi.org/10.1207/s15516709cog1402_1) (cf. p. 55).
- FAN, Y. (2021). « Magnetic fields in the solar convection zone ». In : *Living Reviews in Solar Physics* 18.1, p. 5. ISSN : 1614-4961. DOI : [10.1007/s41116-021-00031-2](https://doi.org/10.1007/s41116-021-00031-2) (cf. p. 17).
- FLETCHER, L. et al. (2011). « An Observational Overview of Solar Flares ». In : *Space Science Reviews* 159.1, p. 19. ISSN : 1572-9672. DOI : [10.1007/s11214-010-9701-8](https://doi.org/10.1007/s11214-010-9701-8) (cf. p. 16, 18).
- FLORIOS, K. et al. (2018). « Forecasting Solar Flares Using Magnetogram-based Predictors and Machine Learning ». In : *Solar Physics* 293.2. DOI : [10.1007/s11207-018-1250-4](https://doi.org/10.1007/s11207-018-1250-4) (cf. p. 174).
- FORSYTH, C. et al. (2020). « Forecasting GOES 15 >2 MeV Electron Fluxes From Solar Wind Data and Geomagnetic Indices ». In : *Space Weather* 18.8, e2019SW002416. ISSN : 1542-7390. DOI : [10.1029/2019SW002416](https://doi.org/10.1029/2019SW002416) (cf. p. 167).
- FOX, N. J. et al. (2016). « The Solar Probe Plus Mission : Humanity's First Visit to Our Star ». In : *Space Science Reviews* 204.1, p. 7-48. ISSN : 1572-9672. DOI : [10.1007/s11214-015-0211-6](https://doi.org/10.1007/s11214-015-0211-6) (cf. p. 75).
- FRETZ, K. A., E. Y. ADAMS et K. W. KIRBY (2016). « The Van Allen Probes Observatories : Overview and Operation to Date ». In : *Johns Hopkins APL Technical Digest* 33.3, p. 14 (cf. p. 27).
- FRÍAS-PAREDES, L. et al. (2016). « Introducing the Temporal Distortion Index to perform a bidimensional analysis of renewable energy forecast ». In : *Energy* 94, p. 180-194. ISSN : 0360-5442. DOI : [10.1016/j.energy.2015.10.093](https://doi.org/10.1016/j.energy.2015.10.093) (cf. p. 121).
- FRY, C. D. et al. (2001). « Improvements to the HAF solar wind model for space weather predictions ». In : *Journal of Geophysical Research : Space Physics* 106 (A10), p. 20985-21001. ISSN : 2156-2202. DOI : [10.1029/2000JA000220](https://doi.org/10.1029/2000JA000220) (cf. p. 86).
- GALSGAARD, K. et A. W. LONGBOTTOM (1999). « Formation of Solar Prominences by Flux Convergence ». In : *The Astrophysical Journal* 510.1, p. 444. ISSN : 0004-637X. DOI : [10.1086/306559](https://doi.org/10.1086/306559) (cf. p. 17).
- GALVEZ, R. et al. (2019). « A Machine-learning Data Set Prepared from the NASA Solar Dynamics Observatory Mission ». In : *The Astrophysical Journal Supplement Series* 242.1, p. 7. ISSN : 0067-0049. DOI : [10.3847/1538-4365/ab1005](https://doi.org/10.3847/1538-4365/ab1005) (cf. p. xxxv, 180, 183).
- GEORGOULIS, M. K. et al. (2021). « The flare likelihood and region eruption forecasting (FLARECAST) project : flare forecasting in the big data & machine learning era ». In : *Journal of Space Weather and Space Climate* 11, p. 39. ISSN : 2115-7251. DOI : [10.1051/swsc/2021023](https://doi.org/10.1051/swsc/2021023) (cf. p. 173).
- GÉRON, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow : Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media (cf. p. 43).

- GHIL, M. et al. (2002). « Advanced Spectral Methods for Climatic Time Series ». In : *Reviews of Geophysics* 40.1, p. 3–1–3-41. ISSN : 1944-9208. DOI : [10.1029/2000RG000092](https://doi.org/10.1029/2000RG000092) (cf. p. 112, 263).
- GITIAUX, X. et al. (2019). « Probabilistic Super-Resolution of Solar Magnetograms : Generating Many Explanations and Measuring Uncertainties ». In : *arXiv :1911.01486 [astro-ph, stat]*. arXiv : [1911.01486](https://arxiv.org/abs/1911.01486) (cf. p. 172).
- GLAUERT, S. A., R. B. HORNE et N. P. MEREDITH (2014). « Three-dimensional electron radiation belt simulations using the BAS Radiation Belt Model with new diffusion models for chorus, plasmaspheric hiss, and lightning-generated whistlers ». In : *Journal of Geophysical Research : Space Physics* 119.1, p. 268-289. ISSN : 2169-9402. DOI : [10.1002/2013JA019281](https://doi.org/10.1002/2013JA019281) (cf. p. 89).
- GLEISNER, H., H. LUNDSTEDT et P. WINTOFT (1996). « Predicting geomagnetic storms from solar-wind data using time-delay neural networks ». In : *Annales Geophysicae* 14.7, p. 679-686. ISSN : 1432-0576. DOI : [10.1007/s00585-996-0679-1](https://doi.org/10.1007/s00585-996-0679-1) (cf. p. 87).
- GOLDBERGER, J. et al. (2005). « Neighbourhood components analysis ». In : *Advances in neural information processing systems*, p. 513-520 (cf. p. 41).
- GOLUB, L. et al. (2020). « EUV imaging and spectroscopy for improved space weather forecasting ». In : *Journal of Space Weather and Space Climate* 10, p. 37. ISSN : 2115-7251. DOI : [10.1051/swsc/2020040](https://doi.org/10.1051/swsc/2020040) (cf. p. 175).
- GONZALEZ, W. D. et al. (2011). « Interplanetary Origin of Intense, Superintense and Extreme Geomagnetic Storms ». en. In : *Space Science Reviews* 158.1, p. 69-89. ISSN : 1572-9672. DOI : [10.1007/s11214-010-9715-2](https://doi.org/10.1007/s11214-010-9715-2) (cf. p. 31).
- GOODFELLOW, I., Y. BENGIO et A. COURVILLE (2016). *Deep Learning*. Illustrated edition. Cambridge, Massachusetts : The MIT Press. 800 p. ISBN : 978-0-262-03561-3 (cf. p. 43).
- GRANDIN, M. et al. (2015). « Effects of solar wind high-speed streams on the high-latitude ionosphere : Superposed epoch study ». In : *Journal of Geophysical Research : Space Physics* 120.12, p. 10, 669-10, 687. ISSN : 2169-9402. DOI : [10.1002/2015JA021785](https://doi.org/10.1002/2015JA021785) (cf. p. 31).
- GRANDIN, M., A. T. AIKIO et A. KOZLOVSKY (2019). « Properties and Geoeffectiveness of Solar Wind High-Speed Streams and Stream Interaction Regions During Solar Cycles 23 and 24 ». In : *Journal of Geophysical Research : Space Physics* 124.6, p. 3871-3892. ISSN : 2169-9402. DOI : [10.1029/2018JA026396](https://doi.org/10.1029/2018JA026396) (cf. p. 15, 18, 19).
- GRAVES, A., A.-r. MOHAMED et G. HINTON (2013). « Speech recognition with deep recurrent neural networks ». In : *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, p. 6645-6649. DOI : [10.1109/ICASSP.2013.6638947](https://doi.org/10.1109/ICASSP.2013.6638947) (cf. p. 59).
- GRISON, B. (2006). « Les cornets polaires : étude d'une région clef de l'interface vent solaire-magnétosphère à l'aide des données Cluster. » Thèse de doct. Université Pierre et Marie Curie - Paris VI (cf. p. 24).
- GRUET, M. (2018). « Intelligence artificielle et prévision de l'impact de l'activité solaire sur l'environnement magnétique terrestre ». thesis. Toulouse, ISAE (cf. p. 43, 88, 89, 95).
- GRUET, M. et al. (2018). « Multiple-Hour-Ahead Forecast of the Dst Index Using a Combination of Long Short-Term Memory Neural Network and Gaussian Process ». In : *Space Weather* 16.11, p. 1882-1896. DOI : [10.1029/2018SW001898](https://doi.org/10.1029/2018SW001898) (cf. p. 3, 88, 91, 94, 95, 97-103, 106, 111, 124, 126, 133, 135, 136, 167, 233, 235).

-
- GUERRA, J. A., S. A. MURRAY et E. DOORNBOS (2020). « The Use of Ensembles in Space Weather Forecasting ». In : *Space Weather* 18.2, e2020SW002443. ISSN : 1542-7390. DOI : [10.1029/2020SW002443](https://doi.org/10.1029/2020SW002443) (cf. p. 90).
- HANDY, B. et al. (1999). « The transition region and coronal explorer ». In : *Solar Physics* 187.2, p. 229-260. ISSN : 1573-093X. DOI : [10.1023/A:1005166902804](https://doi.org/10.1023/A:1005166902804) (cf. p. 74).
- HAPGOOD, M. et al. (2021). « Development of Space Weather Reasonable Worst-Case Scenarios for the UK National Risk Assessment ». In : *Space Weather* 19.4, e2020SW002593. ISSN : 1542-7390. DOI : [10.1029/2020SW002593](https://doi.org/10.1029/2020SW002593) (cf. p. 32, 35).
- HARRIS, C. R. et al. (2020). « Array programming with NumPy ». In : *Nature* 585.7825, p. 357-362. DOI : [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2) (cf. p. 100, 180).
- HARTIGAN, J. A. et M. A. WONG (1979). « Algorithm AS 136 : A K-Means Clustering Algorithm ». In : *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28.1, p. 100-108. ISSN : 0035-9254. DOI : [10.2307/2346830](https://doi.org/10.2307/2346830) (cf. p. 42).
- HARVEY, J. W. et al. (1996). « The Global Oscillation Network Group (GONG) Project ». In : *Science* 272.5266, p. 1284-1286. ISSN : 0036-8075, 1095-9203. DOI : [10.1126/science.272.5266.1284](https://doi.org/10.1126/science.272.5266.1284) (cf. p. 75).
- HATHAWAY, D. H. (2010). « The Solar Cycle ». In : *Living Reviews in Solar Physics* 7.1, p. 1. ISSN : 1614-4961. DOI : [10.12942/lrsp-2010-1](https://doi.org/10.12942/lrsp-2010-1) (cf. p. 9).
- HE, K. et al. (2015). « Deep Residual Learning for Image Recognition ». In : (cf. p. 61, 185).
- HENSMAN, J., A. MATTHEWS et Z. GHAHRAMANI (2014). « Scalable Variational Gaussian Process Classification ». In : *arXiv :1411.2005 [stat]*. arXiv : [1411.2005](https://arxiv.org/abs/1411.2005) (cf. p. 126).
- HERRERA, D., V. F. MAGET et A. SICARD-PIET (2016). « Characterizing magnetopause shadowing effects in the outer electron radiation belt during geomagnetic storms ». In : *Journal of Geophysical Research : Space Physics* 121.10, p. 9517-9530. ISSN : 2169-9402. DOI : [10.1002/2016JA022825](https://doi.org/10.1002/2016JA022825) (cf. p. 32, 89).
- HERRERA, D. (2017). « Prise en compte du temps local dans la modélisation des ceintures de radiation terrestres ». thesis. Toulouse, ISAE (cf. p. 21, 28).
- HEY, T., S. TANSLEY et K. TOLLE (2009). *The Fourth Paradigm : Data-Intensive Scientific Discovery* (cf. p. 40).
- HINTON, G. E., S. OSINDERO et Y.-W. TEH (2006). « A Fast Learning Algorithm for Deep Belief Nets ». In : *Neural Computation* 18.7, p. 1527-1554. ISSN : 0899-7667, 1530-888X. DOI : [10.1162/neco.2006.18.7.1527](https://doi.org/10.1162/neco.2006.18.7.1527) (cf. p. 47).
- HOCHREITER, S. (1991). « Untersuchungen zu dynamischen neuronalen Netzen ». In : *Diploma, Technische Universität München* 91.1 (cf. p. 56).
- HOCHREITER, S. et J. SCHMIDHUBER (1997). « Long Short-Term Memory ». In : *Neural Computation* 9.8, p. 1735-1780. ISSN : 0899-7667. DOI : [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735) (cf. p. 57).
- HORNE, R. B. et al. (2013). « Space weather impacts on satellites and forecasting the Earth's electron radiation belts with SPACECAST ». In : *Space Weather* 11.4, p. 169-186. ISSN : 1542-7390. DOI : [10.1002/swe.20023](https://doi.org/10.1002/swe.20023) (cf. p. 34).
- HORNIK, K. (1991). « Approximation capabilities of multilayer feedforward networks ». In : *Neural Networks* 4.2, p. 251-257. ISSN : 0893-6080. DOI : [10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T) (cf. p. 51).

- HORNIK, K., M. STINCHCOMBE et H. WHITE (1989). « Multilayer feedforward networks are universal approximators ». In : *Neural Networks* 2.5, p. 359-366. ISSN : 0893-6080. DOI : [10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8) (cf. p. 51).
- HOWARD, R. A. et al. (2008). « Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI) ». In : *Space Science Reviews* 136.1, p. 67-115. ISSN : 1572-9672. DOI : [10.1007/s11214-008-9341-4](https://doi.org/10.1007/s11214-008-9341-4) (cf. p. 70).
- HUANG, M., C. TEKUR et M. CARILLI (2020). *Introducing native PyTorch automatic mixed precision for faster training on NVIDIA GPUs*. PyTorch Blog. URL : <https://pytorch.org/blog/accelerating-training-on-nvidia-gpus-with-pytorch-automatic-mixed-precision/> (visité le 25/02/2022) (cf. p. 199).
- HUANG, X. et al. (2013). « Improving the performance of solar flare prediction using active longitudes information ». In : *Astronomy and Astrophysics* 549. DOI : [10.1051/0004-6361/201219742](https://doi.org/10.1051/0004-6361/201219742) (cf. p. 174).
- HUANG, X. et al. (2018). « Deep Learning Based Solar Flare Forecasting Model. I. Results for Line-of-sight Magnetograms ». In : *Astrophysical Journal* 856.1. DOI : [10.3847/1538-4357/aaae00](https://doi.org/10.3847/1538-4357/aaae00) (cf. p. 174).
- HUDSON, H. S. (1991). « Solar flares, microflares, nanoflares, and coronal heating ». In : *Solar Physics* 133.2, p. 357-369. ISSN : 1573-093X. DOI : [10.1007/BF00149894](https://doi.org/10.1007/BF00149894) (cf. p. 16).
- HUDSON, H. S. (2011). « Global Properties of Solar Flares ». In : *Space Science Reviews* 158.1, p. 5-41. ISSN : 1572-9672. DOI : [10.1007/s11214-010-9721-4](https://doi.org/10.1007/s11214-010-9721-4) (cf. p. 16).
- HURLBURT, N. et al. (2012). « Heliophysics Event Knowledgebase for the Solar Dynamics Observatory (SDO) and Beyond ». In : *Solar Physics* 275, p. 67-78. ISSN : 0038-0938. DOI : [10.1007/s11207-010-9624-2](https://doi.org/10.1007/s11207-010-9624-2) (cf. p. 76).
- ILLARIONOV, E. A. et A. G. TLATOV (2018). « Segmentation of coronal holes in solar disc images with a convolutional neural network ». In : *Monthly Notices of the Royal Astronomical Society* 481.4, p. 5014-5021. ISSN : 0035-8711. DOI : [10.1093/mnras/sty2628](https://doi.org/10.1093/mnras/sty2628) (cf. p. 173).
- INGUIMBERT, C. (2018). « Les effets des radiations de l'environnement spatial sur les matériaux et systèmes spatiaux ». Habilitation à diriger des recherches. UNIVERSITE DE TOULOUSE 3 PAUL SABATIER (cf. p. 34).
- IOFFE, S. et C. SZEGEDY (2015). « Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift ». In : *arXiv :1502.03167 [cs]*. arXiv : [1502.03167](https://arxiv.org/abs/1502.03167) (cf. p. 64).
- IWAI, K. et al. (2021). « Validation of coronal mass ejection arrival-time forecasts by magnetohydrodynamic simulations based on interplanetary scintillation observations ». In : *Earth, Planets and Space* 73.1, p. 9. ISSN : 1880-5981. DOI : [10.1186/s40623-020-01345-5](https://doi.org/10.1186/s40623-020-01345-5) (cf. p. 176).
- IYEMORI, T., H. MAEDA et T. KAMEI (1979). « Impulse response of geomagnetic indices to interplanetary magnetic field ». In : *Journal of Geomagnetism and Geoelectricity* 31, p. 1-9. ISSN : 0022-1392. DOI : [10.5636/jgg.31.1](https://doi.org/10.5636/jgg.31.1) (cf. p. 87).
- JANVIER, M. (2017). « Three-dimensional magnetic reconnection and its application to solar flares ». In : *Journal of Plasma Physics* 83.1. ISSN : 0022-3778, 1469-7807. DOI : [10.1017/S0022377817000034](https://doi.org/10.1017/S0022377817000034) (cf. p. 16).
- JIAN, L. et al. (2006a). « Properties of Interplanetary Coronal Mass Ejections at One AU During 1995 – 2004 ». In : *Solar Physics* 239.1, p. 393-436. ISSN : 1573-093X. DOI : [10.1007/s11207-006-0133-2](https://doi.org/10.1007/s11207-006-0133-2) (cf. p. 82).

-
- JIAN, L. et al. (2006b). « Properties of Stream Interactions at One AU During 1995 – 2004 ». In : *Solar Physics* 239.1, p. 337-392. ISSN : 1573-093X. DOI : [10.1007/s11207-006-0132-3](https://doi.org/10.1007/s11207-006-0132-3) (cf. p. 82).
- JIN, M. et al. (2017). « Data-constrained coronal mass ejections in a global magnetohydrodynamics model ». In : *The Astrophysical Journal* 834.2, p. 173. ISSN : 0004-637X. DOI : [10.3847/1538-4357/834/2/173](https://doi.org/10.3847/1538-4357/834/2/173) (cf. p. 87).
- JORDAN, M. I. (1990). « Attractor dynamics and parallelism in a connectionist sequential machine ». In : *Artificial neural networks : concept learning*, p. 112-127 (cf. p. 55).
- JUMPER, J. et al. (2021). « Highly accurate protein structure prediction with AlphaFold ». In : *Nature* 596.7873, p. 583-589. ISSN : 1476-4687. DOI : [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2) (cf. p. 64).
- JUNGBLUTH, A. et al. (2019). « Single-Frame Super-Resolution of Solar Magnetograms : Investigating Physics-Based Metrics & Losses ». In : *arXiv :1911.01490 [astro-ph]*. arXiv : [1911.01490](https://arxiv.org/abs/1911.01490) (cf. p. 172).
- KATSAVRIAS, C. et al. (2021). « On the semi-annual variation of relativistic electrons in the outer radiation belt ». In : *Annales Geophysicae* 39.3, p. 413-425. ISSN : 0992-7689. DOI : [10.5194/angeo-39-413-2021](https://doi.org/10.5194/angeo-39-413-2021) (cf. p. 104).
- KAYALIBAY, B., G. JENSEN et P. van der SMAGT (2017). « CNN-based Segmentation of Medical Imaging Data ». In : *arXiv :1701.03056 [cs]*. arXiv : [1701.03056](https://arxiv.org/abs/1701.03056) (cf. p. 61).
- KELLER, C. U., J. W. HARVEY et M. S. GIAMPAPA (2003). « SOLIS : an innovative suite of synoptic solar instruments ». In : *Innovative Telescopes and Instrumentation for Solar Astrophysics*. Innovative Telescopes and Instrumentation for Solar Astrophysics. T. 4853. International Society for Optics et Photonics, p. 194-204. DOI : [10.1117/12.460373](https://doi.org/10.1117/12.460373) (cf. p. 75).
- KIDGER, P. et T. LYONS (2020). « Universal Approximation with Deep Narrow Networks ». In : *arXiv :1905.08539 [cs, math, stat]*. arXiv : [1905.08539](https://arxiv.org/abs/1905.08539) (cf. p. 51).
- KIEFER, J. et J. WOLFOWITZ (1952). « Stochastic estimation of the maximum of a regression function ». In : *The Annals of Mathematical Statistics*, p. 462-466 (cf. p. 50).
- KILPUA, E. K. J. et al. (2015). « Unraveling the drivers of the storm time radiation belt response ». In : *Geophysical Research Letters* 42.9, p. 3076-3084. ISSN : 1944-8007. DOI : [10.1002/2015GL063542](https://doi.org/10.1002/2015GL063542) (cf. p. 82).
- KILPUA, E. K. J. et al. (2017). « Geoeffective Properties of Solar Transients and Stream Interaction Regions ». In : *Space Science Reviews* 212.3, p. 1271-1314. ISSN : 1572-9672. DOI : [10.1007/s11214-017-0411-3](https://doi.org/10.1007/s11214-017-0411-3) (cf. p. 21).
- KIM, S.-J. et al. (2007). « An Interior-Point Method for Large-Scale l1-Regularized Least Squares ». In : *IEEE Journal of Selected Topics in Signal Processing* 1.4, p. 606-617. ISSN : 1941-0484. DOI : [10.1109/JSTSP.2007.910971](https://doi.org/10.1109/JSTSP.2007.910971) (cf. p. 41).
- KING, J. H. et N. E. PAPITASHVILI (2005). « Solar wind spatial scales in and comparisons of hourly Wind and ACE plasma and magnetic field data ». In : *Journal of Geophysical Research : Space Physics* 110 (A2). ISSN : 2156-2202. DOI : [10.1029/2004JA010649](https://doi.org/10.1029/2004JA010649) (cf. p. 78).
- KINGMA, D. P. et J. BA (2017). « Adam : A Method for Stochastic Optimization ». In : *arXiv :1412.6980 [cs]*. arXiv : [1412.6980](https://arxiv.org/abs/1412.6980) (cf. p. 108, 153, 198).
- KIVELSON, M. G. et C. T. RUSSEL (1995). *Introduction to Space Physics*. Cambridge University Press. 594 p. ISBN : 978-0-521-45714-9 (cf. p. 7, 9).

- KONDRASHOV, D., Y. SHPRITS et M. GHIL (2010). « Gap filling of solar wind data by singular spectrum analysis ». In : *Geophysical Research Letters* 37.15. ISSN : 1944-8007. DOI : [10.1029/2010GL044138](https://doi.org/10.1029/2010GL044138) (cf. p. 112, 263).
- KONDRASHOV, D. et al. (2014). « Reconstruction of gaps in the past history of solar wind parameters ». In : *Geophysical Research Letters* 41.8, p. 2702-2707. ISSN : 1944-8007. DOI : [10.1002/2014GL059741](https://doi.org/10.1002/2014GL059741) (cf. p. 112, 264).
- KOONS, H. C. (2001). « Statistical analysis of extreme values in space science ». In : *Journal of Geophysical Research* 106, p. 10915-10921. ISSN : 0148-0227. DOI : [10.1029/2000JA000234](https://doi.org/10.1029/2000JA000234) (cf. p. 147).
- KOSKINEN, H. (2011a). *Physics of Space Storms : From the Solar Surface to the Earth*. Springer Science & Business Media. 431 p. ISBN : 978-3-642-00319-6 (cf. p. 95).
- KOSKINEN, H. E. J. (2011b). « Stormy Tour from the Sun to the Earth ». In : *Physics of Space Storms : From the Solar Surface to the Earth*. Sous la dir. de H. KOSKINEN. Springer Praxis Books. Berlin, Heidelberg : Springer, p. 1-58. ISBN : 978-3-642-00319-6. DOI : [10.1007/978-3-642-00319-6_1](https://doi.org/10.1007/978-3-642-00319-6_1) (cf. p. 25).
- KOSKINEN, H. E. J. et E. K. J. KILPUA (2022). *Physics of Earth's Radiation Belts : Theory and Observations*. Astronomy and Astrophysics Library. Cham : Springer International Publishing. ISBN : 978-3-030-82166-1. DOI : [10.1007/978-3-030-82167-8](https://doi.org/10.1007/978-3-030-82167-8) (cf. p. 7, 21, 22, 26).
- KOSUGI, T. et al. (2007). « The Hinode (Solar-B) Mission : An Overview ». In : *Solar Physics* 243, p. 3-17. DOI : [10.1007/s11207-007-9014-6](https://doi.org/10.1007/s11207-007-9014-6) (cf. p. 66).
- KRIZHEVSKY, A., I. SUTSKEVER et G. E. HINTON (2012). « Imagenet classification with deep convolutional neural networks ». In : *Advances in neural information processing systems* 25, p. 1097-1105 (cf. p. 41, 61).
- KUHN, T. S. et L. MEYER (1983). *La structure des révolutions scientifiques*. Paris : Flammarion. 256 p. ISBN : 978-2-08-210181-3 (cf. p. 40).
- LANTOS, P. (1997). *Le soleil en face : le soleil et les relations Soleil-Terre*. Culture scientifique. Paris : Masson. 193 p. ISBN : 2-225-83054-1 (cf. p. 14).
- LANZEROTTI, L. J. et D. N. BAKER (2017). « Space weather research : Earth's radiation belts ». In : *Space Weather* 15.6, p. 742-745. ISSN : 1542-7390. DOI : [10.1002/2017SW001654](https://doi.org/10.1002/2017SW001654) (cf. p. 33).
- LAPERRE, B., J. AMAYA et G. LAPENTA (2020). « Dynamic Time Warping as a New Evaluation for Dst Forecast With Machine Learning ». In : *Frontiers in Astronomy and Space Sciences* 7.39. ISSN : 2296-987X. DOI : [10.3389/fspas.2020.00039](https://doi.org/10.3389/fspas.2020.00039) (cf. p. 88, 112, 122).
- LAZARO, D. et al. (2013). « Radiation Belt Activity Indices and Solar Proton Event Alarm on the CRATERRE project web site ». In : *2013 14th European Conference on Radiation and Its Effects on Components and Systems (RADECS)*. 2013 14th European Conference on Radiation and Its Effects on Components and Systems (RADECS), p. 1-7. DOI : [10.1109/RADECS.2013.6937420](https://doi.org/10.1109/RADECS.2013.6937420) (cf. p. 83).
- LAZARO, D. et al. (2014). « Radiation Belt Activity Indices and Solar Proton Event Alarm ». In : *IEEE Transactions on Nuclear Science* 61.4, p. 1671-1678. ISSN : 1558-1578. DOI : [10.1109/TNS.2013.2296115](https://doi.org/10.1109/TNS.2013.2296115) (cf. p. 82).
- LAZZÚS, J. A. et al. (2017). « Forecasting the Dst index using a swarm-optimized neural network ». In : *Space Weather* 15.8, p. 1068-1089. ISSN : 1542-7390. DOI : [10.1002/2017SW001608](https://doi.org/10.1002/2017SW001608) (cf. p. 42, 95, 111, 167).

-
- LECUN, Y. et al. (1989). « Backpropagation Applied to Handwritten Zip Code Recognition ». In : *Neural Computation* 1.4, p. 541-551. ISSN : 0899-7667. DOI : [10.1162/neco.1989.1.4.541](https://doi.org/10.1162/neco.1989.1.4.541) (cf. p. 47, 61).
- LEJOSNE, S. et P. KOLLMANN (2020). « Radiation Belt Radial Diffusion at Earth and Beyond ». In : *Space Science Reviews* 216.1, p. 19. ISSN : 1572-9672. DOI : [10.1007/s11214-020-0642-6](https://doi.org/10.1007/s11214-020-0642-6) (cf. p. 32).
- LEMEN, J. R. et al. (2012). « The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO) ». In : *The Solar Dynamics Observatory*. Sous la dir. de P. CHAMBERLIN, W. D. PESNELL et B. THOMPSON. New York, NY : Springer US, p. 17-40. ISBN : 978-1-4614-3673-7. DOI : [10.1007/978-1-4614-3673-7_3](https://doi.org/10.1007/978-1-4614-3673-7_3) (cf. p. 71, 72).
- LETHY, A. et al. (2018). « Prediction of the Dst Index and Analysis of Its Dependence on Solar Wind Parameters Using Neural Network ». In : *Space Weather* 16.9, p. 1277-1290. ISSN : 1542-7390. DOI : [10.1029/2018SW001863](https://doi.org/10.1029/2018SW001863) (cf. p. 89, 111).
- LHOTKA, C. et Y. NARITA (2019). « Kinematic models of the interplanetary magnetic field ». In : *Annales Geophysicae* 37.3, p. 299-314. ISSN : 0992-7689. DOI : [10.5194/angeo-37-299-2019](https://doi.org/10.5194/angeo-37-299-2019) (cf. p. 12).
- LI, Y. et al. (2019). *Similarity Grouping-Guided Neural Network Modeling for Maritime Time Series Prediction* (cf. p. 121, 122).
- LI, Z.-F. et al. (2020). « Synthesising solar radio images from Atmospheric Imaging Assembly extreme-ultraviolet data ». In : *Research in Astronomy and Astrophysics* 20.2. DOI : [10.1088/1674-4527/20/2/18](https://doi.org/10.1088/1674-4527/20/2/18) (cf. p. 172).
- LIEMOHN, M. W. et al. (2018). « Model Evaluation Guidelines for Geomagnetic Index Predictions ». In : *Space Weather* 16.12, p. 2079-2102. ISSN : 1542-7390. DOI : [10.1029/2018SW002067](https://doi.org/10.1029/2018SW002067) (cf. p. 115).
- LILENSTEN, J. et P.-L. BLELLY (1999). *Du soleil à la terre : aéronomie et météorologie de l'espace*. EDP Sciences. Presses universitaires de Grenoble. 430 p. ISBN : 978-2-7061-0834-1 (cf. p. 7, 25).
- LILENSTEN, J. et J. BORNAREL (2001). *Sous les feux du soleil : vers une météorologie de l'espace*. EDP sciences (cf. p. 24).
- LIM, D. et al. (2019). « Ensemble Forecasting of Major Solar Flares with Short-, Mid-, and Long-term Active Region Properties ». In : *Astrophysical Journal* 885.1. DOI : [10.3847/1538-4357/ab45e7](https://doi.org/10.3847/1538-4357/ab45e7) (cf. p. 174).
- LIU, C. et al. (2017). « Predicting Solar Flares Using SDO/HMI Vector Magnetic Data Products and the Random Forest Algorithm ». In : *Astrophysical Journal* 843.2. DOI : [10.3847/1538-4357/aa789b](https://doi.org/10.3847/1538-4357/aa789b) (cf. p. 174).
- LIU, H. et al. (2019a). « Predicting Solar Flares Using a Long Short-term Memory Network ». In : *Astrophysical Journal* 877.2. DOI : [10.3847/1538-4357/ab1b3c](https://doi.org/10.3847/1538-4357/ab1b3c) (cf. p. 174).
- LIU, H. et al. (2019b). « When Gaussian Process Meets Big Data : A Review of Scalable GPs ». In : *arXiv :1807.01065 [cs, stat]*. arXiv : [1807.01065](https://arxiv.org/abs/1807.01065) (cf. p. 126).
- LIU, J. et al. (2018). « A New Tool for CME Arrival Time Prediction using Machine Learning Algorithms : CAT-PUMA ». In : *Astrophysical Journal* 855.2. DOI : [10.3847/1538-4357/aaae69](https://doi.org/10.3847/1538-4357/aaae69) (cf. p. 175).
- LOPES, U. K. et J. F. VALIATI (2017). « Pre-trained convolutional neural networks as feature extractors for tuberculosis detection ». In : *Computers in Biology and Medicine*

- 89, p. 135-143. ISSN : 0010-4825. DOI : [10.1016/j.combiomed.2017.08.001](https://doi.org/10.1016/j.combiomed.2017.08.001) (cf. p. 185).
- LOUPPE, G. (2014). « Understanding random forests : From theory to practice ». Thèse de doct. (cf. p. 41).
- LU, Z. et al. (2017). « The Expressive Power of Neural Networks : A View from the Width ». In : *arXiv :1709.02540 [cs]*. arXiv : [1709.02540](https://arxiv.org/abs/1709.02540) (cf. p. 51).
- LUGAZ, N. et al. (2015). « Extreme geomagnetic disturbances due to shocks within CMEs ». In : *Geophysical Research Letters* 42.12, p. 4694-4701. ISSN : 1944-8007. DOI : [10.1002/2015GL064530](https://doi.org/10.1002/2015GL064530) (cf. p. 141).
- LUNDSTEDT, H., H. GLEISNER et P. WINTOFT (2002). « Operational forecasts of the geomagnetic Dst index ». In : *Geophysical Research Letters* 29.24, p. 34–1–34-4. ISSN : 1944-8007. DOI : [10.1029/2002GL016151](https://doi.org/10.1029/2002GL016151) (cf. p. 87).
- LUNDSTEDT, H. et P. WINTOFT (1994). « Prediction of geomagnetic storms from solar wind data with the use of a neural network ». In : *Annales Geophysicae* 12.1, p. 19-24. ISSN : 1432-0576. DOI : [10.1007/s00585-994-0019-2](https://doi.org/10.1007/s00585-994-0019-2) (cf. p. 87).
- LUO, B., S. LIU et J. GONG (2017). « Two empirical models for short-term forecast of Kp ». In : *Space Weather* 15.3, p. 503-516. ISSN : 1542-7390. DOI : [10.1002/2016SW001585](https://doi.org/10.1002/2016SW001585) (cf. p. 88).
- MACNEICE, P. et al. (2018). « Assessing the Quality of Models of the Ambient Solar Wind ». In : *Space Weather* 16.11, p. 1644-1667. ISSN : 1542-7390. DOI : [10.1029/2018SW002040](https://doi.org/10.1029/2018SW002040) (cf. p. 86, 177).
- MAGET, V., S. BOURDARIE et D. BOSCHER (2008). « Direct Data Assimilation Over Solar Cycle Time-Scales to Improve Proton Radiation Belt Models ». In : *IEEE Transactions on Nuclear Science* 55.4, p. 2188-2196. ISSN : 1558-1578. DOI : [10.1109/TNS.2008.921928](https://doi.org/10.1109/TNS.2008.921928) (cf. p. 90).
- MAGET, V. (2007). « Développement et comparaison de méthodes d'assimilation de données appliquées à la restitution de la dynamique des ceintures de radiation terrestres ». Thèse de doctorat. Toulouse, ISAE (cf. p. 21, 29, 90).
- MANTTARI, J. et al. (2020). « Interpreting Video Features : A Comparison of 3D Convolutional Networks and Convolutional LSTM Networks ». In : *Proceedings of the Asian Conference on Computer Vision* (cf. p. 236).
- MARCHAUDON, A. (2018). « Observation et modélisation des processus de couplage entre la magnétosphère et l'ionosphère terrestres ». Habilitation à diriger des recherches. Université Paul Sabatier, Toulouse (cf. p. 26).
- MARSCH, E. (2006). « Kinetic Physics of the Solar Corona and Solar Wind ». In : *Living Reviews in Solar Physics* 3.1, p. 1. ISSN : 1614-4961. DOI : [10.12942/lrsp-2006-1](https://doi.org/10.12942/lrsp-2006-1) (cf. p. 10).
- MASTERS, D. et C. LUSCHI (2018). « Revisiting Small Batch Training for Deep Neural Networks ». In : *arXiv :1804.07612 [cs, stat]*. arXiv : [1804.07612](https://arxiv.org/abs/1804.07612) (cf. p. 50).
- MATÉO-VÉLEZ, J.-C. et al. (2018). « Spacecraft surface charging induced by severe environments at geosynchronous orbit ». In : *Space Weather* 16.1, p. 89-106. ISSN : 1542-7390. DOI : [10.1002/2017SW001689](https://doi.org/10.1002/2017SW001689) (cf. p. 155).
- MATHESON, J. E. et R. L. WINKLER (1976). « Scoring Rules for Continuous Probability Distributions ». In : *Management Science* 22.10, p. 1087-1096. ISSN : 0025-1909 (cf. p. 130).

-
- MATZKA, J. et al. (2021). « The Geomagnetic Kp Index and Derived Indices of Geomagnetic Activity ». In : *Space Weather* 19.5, e2020SW002641. ISSN : 1542-7390. DOI : [10.1029/2020SW002641](https://doi.org/10.1029/2020SW002641) (cf. p. 80, 82).
- MAYAUD, P.-N. (1968). *Indices Kn, Ks Et Km, 1964-1967 : Texte Français-anglais*. Éditions du Centre national de la recherche scientifique. 182 p. (cf. p. 80, 140).
- MAYAUD, P.-N. (1971). « Une mesure planétaire d'activité magnétique basée sur deux observatoires antipodaux ». In : *Annales Geophysicae* 27, p. 67-70 (cf. p. 80, 140).
- MAYAUD, P.-N. (1972). « The aa indices : A 100-year series characterizing the magnetic activity ». In : *Journal of Geophysical Research* 77.34, p. 6870-6874. ISSN : 2156-2202. DOI : [10.1029/JA077i034p06870](https://doi.org/10.1029/JA077i034p06870) (cf. p. 140).
- MAYAUD, P.-N. (1980). *Derivation, meaning, and use of geomagnetic indices*. Geophysical monograph ; 22. Washington : American Geophysical Union. 154 p. ISBN : 978-0-87590-022-3 (cf. p. 79, 140).
- MCCOMAS, D. J. et al. (1993). « Magnetospheric plasma analyzer : Initial three-spacecraft observations from geosynchronous orbit ». In : *Journal of Geophysical Research : Space Physics* 98 (A8), p. 13453-13465. ISSN : 2156-2202. DOI : <https://doi.org/10.1029/93JA00726> (cf. p. 155).
- MCCOMAS, D. J. et al. (2000). « Solar wind observations over Ulysses' first full polar orbit ». In : *Journal of Geophysical Research : Space Physics* 105 (A5), p. 10419-10433. ISSN : 2156-2202. DOI : [10.1029/1999JA000383](https://doi.org/10.1029/1999JA000383) (cf. p. 12).
- MCCULLOCH, W. S. et W. PITTS (1943). « A logical calculus of the ideas immanent in nervous activity ». In : *The bulletin of mathematical biophysics* 5.4, p. 115-133 (cf. p. 41, 43).
- MCILWAIN, C. E. (1961). « Coordinates for mapping the distribution of magnetically trapped particles ». In : *Journal of Geophysical Research (1896-1977)* 66.11, p. 3681-3691. ISSN : 2156-2202. DOI : [10.1029/JZ066i011p03681](https://doi.org/10.1029/JZ066i011p03681) (cf. p. 28).
- MCLACHLAN, G. J., S. X. LEE et S. I. RATHNAYAKE (2019). « Finite Mixture Models ». In : *Annual Review of Statistics and Its Application* 6.1, p. 355-378. DOI : [10.1146/annurev-statistics-031017-100325](https://doi.org/10.1146/annurev-statistics-031017-100325) (cf. p. 42).
- MEFTAH, M. et al. (2014). « Picard SODISM, a Space Telescope to Study the Sun from the Middle Ultraviolet to the Near Infrared ». In : *Solar Physics* 289.3, p. 1043-1076. ISSN : 1573-093X. DOI : [10.1007/s11207-013-0373-x](https://doi.org/10.1007/s11207-013-0373-x) (cf. p. 74).
- MENVIELLE, M. et al. (2011). « Geomagnetic Indices ». In : *Geomagnetic Observations and Models*. Sous la dir. de M. MANDEA et M. KORTE. IAGA Special Sopron Book Series. Dordrecht : Springer Netherlands, p. 183-228. ISBN : 978-90-481-9858-0. DOI : [10.1007/978-90-481-9858-0_8](https://doi.org/10.1007/978-90-481-9858-0_8) (cf. p. 79).
- MEREDITH, N. P. et al. (2006). « Energetic outer zone electron loss timescales during low geomagnetic activity ». In : *Journal of Geophysical Research : Space Physics* 111 (A5). ISSN : 2156-2202. DOI : [10.1029/2005JA011516](https://doi.org/10.1029/2005JA011516) (cf. p. 140).
- MEREDITH, N. P. et al. (2015). « Extreme relativistic electron fluxes at geosynchronous orbit : Analysis of GOES E > 2 MeV electrons ». In : *Space Weather* 13.3. ISSN : 1542-7390. DOI : [10.1002/2014SW001143](https://doi.org/10.1002/2014SW001143) (cf. p. 147).
- MICIKEVICIUS, P. et al. (2018). « Mixed Precision Training ». In : *arXiv :1710.03740 [cs, stat]*. arXiv : [1710.03740](https://arxiv.org/abs/1710.03740) (cf. p. 199).
- MONTÁNS, F. J. et al. (2019). « Data-driven modeling and learning in science and engineering ». In : *Comptes Rendus Mécanique*. Data-Based Engineering Science and

- Technology 347.11, p. 845-855. ISSN : 1631-0721. DOI : [10.1016/j.crme.2019.11.009](https://doi.org/10.1016/j.crme.2019.11.009) (cf. p. 40).
- MORADI, R., R. BERANGI et B. MINAEI (2020). « A survey of regularization strategies for deep models ». In : *Artificial Intelligence Review* 53.6, p. 3947-3986. ISSN : 1573-7462. DOI : [10.1007/s10462-019-09784-7](https://doi.org/10.1007/s10462-019-09784-7) (cf. p. 64, 108).
- MÖSTL, C. et al. (2018). « Forward Modeling of Coronal Mass Ejection Flux Ropes in the Inner Heliosphere with 3DCORE ». In : *Space Weather* 16.3, p. 216-229. ISSN : 1542-7390. DOI : [10.1002/2017SW001735](https://doi.org/10.1002/2017SW001735) (cf. p. 87).
- MÜLLER, D. et al. (2013). « Solar Orbiter ». In : *Solar Physics* 285.1, p. 25-70. ISSN : 1573-093X. DOI : [10.1007/s11207-012-0085-7](https://doi.org/10.1007/s11207-012-0085-7) (cf. p. 75).
- MURSULA, K., I. G. USOSKIN et G. A. KOVALTISOV (2002). « A 22-year cycle in sunspot activity ». In : *Advances in Space Research* 29.12, p. 1979-1984. ISSN : 0273-1177. DOI : [10.1016/S0273-1177\(02\)00244-2](https://doi.org/10.1016/S0273-1177(02)00244-2) (cf. p. 9).
- NADEAU, R. (1999). *Vocabulaire technique et analytique de l'épistémologie*. Paris : Presses Universitaires de France - PUF. 904 p. ISBN : 978-2-13-049109-5 (cf. p. 4).
- NÉNON, Q. (2018). « Etude et modélisation des ceintures de radiation de Jupiter ». Theses. Doctorat de l'Université de Toulouse délivré par l'Institut Supérieur de l'Aéronautique et de l'Espace (ISAE) (cf. p. 32).
- NGUYEN, G. (2021). « Solar wind / magnetosphere coupling inferred from machine-learning methods ». Thèse de doct. Université Paris-Saclay (cf. p. 219).
- NISHIZUKA, N. et al. (2017). « Solar Flare Prediction Model with Three Machine-learning Algorithms using Ultraviolet Brightening and Vector Magnetograms ». In : *Astrophysical Journal* 835.2. DOI : [10.3847/1538-4357/835/2/156](https://doi.org/10.3847/1538-4357/835/2/156) (cf. p. 174).
- NOAA (2012). *Users Guide to The Preliminary Report and Forecast of Solar Geophysical Data* (cf. p. 200).
- NOLTE, J. T. et al. (1976). « Coronal holes as sources of solar wind. » In : *Solar Physics* 46, p. 303-322. ISSN : 0038-0938. DOI : [10.1007/BF00149859](https://doi.org/10.1007/BF00149859) (cf. p. 13).
- NOSE, M. et al. (2015). « Geomagnetic AE index ». In : *World Data Center for Geomagnetism, Kyoto* 10, p. 15-031 (cf. p. 81).
- ODSTRCIL, D. (2003). « Modeling 3-D solar wind structure ». In : *Advances in Space Research. Heliosphere at Solar Maximum* 32.4, p. 497-506. ISSN : 0273-1177. DOI : [10.1016/S0273-1177\(03\)00332-6](https://doi.org/10.1016/S0273-1177(03)00332-6) (cf. p. 86).
- OFMAN, L. (2004). « The origin of the slow solar wind in coronal streamers ». In : *Advances in Space Research. Streamers, Slow Solar Wind, and the Dynamics of the Magnetosphere* 33.5, p. 681-688. ISSN : 0273-1177. DOI : [10.1016/S0273-1177\(03\)00235-7](https://doi.org/10.1016/S0273-1177(03)00235-7) (cf. p. 12).
- OGAWARA, Y. et al. (1991). « The SOLAR-A Mission : an Overview ». In : *The Yohkoh (Solar-A) Mission*. Sous la dir. de Z. ŠVESTKA et Y. UCHIDA. Dordrecht : Springer Netherlands, p. 1-16. ISBN : 978-94-011-2626-7. DOI : [10.1007/978-94-011-2626-7_1](https://doi.org/10.1007/978-94-011-2626-7_1) (cf. p. 66).
- AL-OMARI, M. et al. (2010). « Machine leaning-based investigation of the associations between CMEs and filaments ». In : *Solar Physics* 262.2, p. 511-539. DOI : [10.1007/s11207-010-9516-5](https://doi.org/10.1007/s11207-010-9516-5) (cf. p. 173).
- OWENS, M. J., P. RILEY et T. S. HORBURY (2017). « Probabilistic Solar Wind and Geomagnetic Forecasting Using an Analogue Ensemble or “Similar Day” Approach ». In : *Solar Physics* 292.5, p. 69. ISSN : 1573-093X. DOI : [10.1007/s11207-017-1090-7](https://doi.org/10.1007/s11207-017-1090-7) (cf. p. 86, 88).

-
- PAN, S. J. et Q. YANG (2010). « A Survey on Transfer Learning ». In : *IEEE Transactions on Knowledge and Data Engineering* 22.10, p. 1345-1359. ISSN : 1558-2191. DOI : [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191) (cf. p. 42, 184).
- PARK, E. et al. (2018). « Application of the Deep Convolutional Neural Network to the Forecast of Solar Flare Occurrence Using Full-disk Solar Magnetograms ». In : *Astrophysical Journal* 869.2. DOI : [10.3847/1538-4357/aaed40](https://doi.org/10.3847/1538-4357/aaed40) (cf. p. 174).
- PARK, E. et al. (2019). « Generation of Solar UV and EUV Images from SDO/HMI Magnetograms by Deep Learning ». In : *Astrophysical Journal Letters* 884.1. DOI : [10.3847/2041-8213/ab46bb](https://doi.org/10.3847/2041-8213/ab46bb) (cf. p. 172).
- PARK, E. et al. (2020). « De-noising SDO/HMI Solar Magnetograms by Image Translation Method Based on Deep Learning ». In : *Astrophysical Journal Letters* 891.1. DOI : [10.3847/2041-8213/ab74d2](https://doi.org/10.3847/2041-8213/ab74d2) (cf. p. 172).
- PARKER, E. N. (1958). « Dynamics of the Interplanetary Gas and Magnetic Fields. » In : *The Astrophysical Journal* 128, p. 664. ISSN : 0004-637X. DOI : [10.1086/146579](https://doi.org/10.1086/146579) (cf. p. 12).
- PASZKE, A. et al. (2017). « Automatic differentiation in PyTorch ». In : 31st Conference on Neural Information Processing Systems (NIPS 2017). Long Beach, CA, USA, p. 4 (cf. p. 50).
- PASZKE, A. et al. (2019). « PyTorch : An imperative style, high-performance deep learning library ». In : *Advances in neural information processing systems 32*. Sous la dir. de H. WALLACH et al. Curran Associates, Inc., p. 8024-8035 (cf. p. 100, 154).
- PAWLOWSKI, N. et al. (2016). « Automating Morphological Profiling with Generic Deep Convolutional Networks ». In : *bioRxiv*, p. 085118. DOI : [10.1101/085118](https://doi.org/10.1101/085118) (cf. p. 185).
- PEARSON, K. (1901). « LIII. On lines and planes of closest fit to systems of points in space ». In : *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11, p. 559-572. ISSN : 1941-5982. DOI : [10.1080/14786440109462720](https://doi.org/10.1080/14786440109462720) (cf. p. 42).
- PEDREGOSA, F. et al. (2011). « Scikit-learn : Machine learning in Python ». In : *Journal of Machine Learning Research* 12, p. 2825-2830 (cf. p. 43, 100).
- PESNELL, W. D., B. J. THOMPSON et P. C. CHAMBERLIN (2012). « The Solar Dynamics Observatory (SDO) ». In : *Solar Physics* 275.1, p. 3-15. ISSN : 1573-093X. DOI : [10.1007/s11207-011-9841-3](https://doi.org/10.1007/s11207-011-9841-3) (cf. p. 71).
- PESNELL, W. (2020). « Lessons learned from predictions of Solar Cycle 24 ». In : *Journal of Space Weather and Space Climate* 10. DOI : [10.1051/swsc/2020060](https://doi.org/10.1051/swsc/2020060) (cf. p. 175).
- PICKANDS, J. (1975). « Statistical Inference Using Extreme Order Statistics ». In : *The Annals of Statistics* 3.1, p. 119-131. ISSN : 0090-5364, 2168-8966. DOI : [10.1214/aos/1176343003](https://doi.org/10.1214/aos/1176343003) (cf. p. 148).
- PINTO, R. F. et A. P. ROUILLARD (2017). « A Multiple Flux-tube Solar Wind Model ». In : 838.2, p. 89. ISSN : 0004-637X. DOI : [10.3847/1538-4357/aa6398](https://doi.org/10.3847/1538-4357/aa6398) (cf. p. 86).
- PIZZO, V. (1978). « A three-dimensional model of corotating streams in the solar wind, 1. Theoretical foundations ». In : *Journal of Geophysical Research : Space Physics* 83 (A12), p. 5563-5572. ISSN : 2156-2202. DOI : [10.1029/JA083iA12p05563](https://doi.org/10.1029/JA083iA12p05563) (cf. p. 14).
- POEDTS, S. et al. (2020). « European Heliospheric FORecasting Information Asset 2.0 ». In : *Journal of Space Weather and Space Climate* 10, p. 57. ISSN : 2115-7251. DOI : [10.1051/swsc/2020055](https://doi.org/10.1051/swsc/2020055) (cf. p. 87).

- POMOELL, J. et S. POEDTS (2018). « EUHFORIA : European heliospheric forecasting information asset ». In : *Journal of Space Weather and Space Climate* 8, A35. ISSN : 2115-7251. DOI : [10.1051/swsc/2018020](https://doi.org/10.1051/swsc/2018020) (cf. p. 86).
- POTGIETER, M. S. (2008). « Solar cycle variations and cosmic rays ». In : *Journal of Atmospheric and Solar-Terrestrial Physics. Recent Observations and Simulations of the Sun-Earth System* 70.2, p. 207-218. ISSN : 1364-6826. DOI : [10.1016/j.jastp.2007.08.023](https://doi.org/10.1016/j.jastp.2007.08.023) (cf. p. 32).
- PRIEST, E. (2014). « A description of the sun ». In : *Magnetohydrodynamics of the sun*. Cambridge : Cambridge University Press, p. 1-73. DOI : [10.1017/CB09781139020732.002](https://doi.org/10.1017/CB09781139020732.002) (cf. p. 7).
- RAINA, R., A. MADHAVAN et A. Y. NG (2009). « Large-scale deep unsupervised learning using graphics processors ». In : *Proceedings of the 26th annual international conference on machine learning*, p. 873-880 (cf. p. 64).
- RAZAVIAN, A. S. et al. (2014). « CNN Features off-the-shelf : an Astounding Baseline for Recognition ». In : *arXiv :1403.6382 [cs]*. arXiv : [1403.6382](https://arxiv.org/abs/1403.6382) (cf. p. 185).
- REGNAULT, F. et al. (2020). *20 Years of ACE Data : How Superposed Epoch Analyses Reveal Generic Features in Interplanetary CME Profiles* (cf. p. 19).
- REISS, M. A. et al. (2021). « Machine Learning for Predicting the Bz Magnetic Field Component From Upstream in Situ Observations of Solar Coronal Mass Ejections ». In : *Space Weather* 19.12, e2021SW002859. ISSN : 1542-7390. DOI : [10.1029/2021SW002859](https://doi.org/10.1029/2021SW002859) (cf. p. 87).
- RÉVILLE, V. et A. S. BRUN (2017). « Global Solar Magnetic Field Organization in the Outer Corona : Influence on the Solar Wind Speed and Mass Flux Over the Cycle ». In : *The Astrophysical Journal* 850.1, p. 45. ISSN : 0004-637X. DOI : [10.3847/1538-4357/aa9218](https://doi.org/10.3847/1538-4357/aa9218) (cf. p. 13).
- RIBEIRO, M. T., S. SINGH et C. GUESTRIN (2016). « "Why Should I Trust You?" : Explaining the Predictions of Any Classifier ». In : *arXiv :1602.04938 [cs, stat]*. arXiv : [1602.04938](https://arxiv.org/abs/1602.04938) (cf. p. 236).
- RIFKIN, R. M. et R. A. LIPPERT (2007). « Notes on Regularized Least Squares ». In : (cf. p. 41).
- RILEY, P. et al. (2017). « Forecasting the properties of the solar wind using simple pattern recognition ». In : *Space Weather* 15.3, p. 526-540. ISSN : 1542-7390. DOI : [10.1002/2016SW001589](https://doi.org/10.1002/2016SW001589) (cf. p. 86).
- ROCHEL, S. et al. (2016). « A radiation belt disturbance study from the space weather point of view ». In : *Acta Astronautica* 128, p. 650-656. ISSN : 0094-5765. DOI : [10.1016/j.actaastro.2016.07.012](https://doi.org/10.1016/j.actaastro.2016.07.012) (cf. p. 140, 141, 234).
- ROEDERER, J. G. (1970). *Dynamics of geomagnetically trapped radiation*. Physics and Chemistry in Space. Berlin, Heidelberg : Springer. 166 p. ISBN : 978-3-642-49300-3 (cf. p. 28).
- ROEDERER, J. G. et S. LEJOSNE (2018). « Coordinates for Representing Radiation Belt Particle Flux ». In : *Journal of Geophysical Research : Space Physics* 123.2, p. 1381-1387. ISSN : 2169-9402. DOI : [10.1002/2017JA025053](https://doi.org/10.1002/2017JA025053) (cf. p. 28).
- ROJAT, T. et al. (2021). « Explainable Artificial Intelligence (XAI) on TimeSeries Data : A Survey ». In : *arXiv :2104.00950 [cs]*. arXiv : [2104.00950](https://arxiv.org/abs/2104.00950) (cf. p. 236).
- ROSENBLATT, F. (1958). « The perceptron : a probabilistic model for information storage and organization in the brain. » In : *Psychological review* 65.6, p. 386 (cf. p. 41, 45).

-
- RUMELHART, D. E., G. E. HINTON et R. J. WILLIAMS (1986). « Learning representations by back-propagating errors ». In : *Nature* 323.6088, p. 533-536. ISSN : 1476-4687. DOI : [10.1038/323533a0](https://doi.org/10.1038/323533a0) (cf. p. 47, 153).
- RUSSAKOVSKY, O. et al. (2015). « ImageNet large scale visual recognition challenge ». In : *International Journal of Computer Vision (IJCV)* 115.3, p. 211-252. DOI : [10.1007/s11263-015-0816-y](https://doi.org/10.1007/s11263-015-0816-y) (cf. p. 41).
- SAGHEER, A. et M. KOTB (2019). « Time series forecasting of petroleum production using deep LSTM recurrent networks ». In : *Neurocomputing* 323, p. 203-213. ISSN : 0925-2312. DOI : [10.1016/j.neucom.2018.09.082](https://doi.org/10.1016/j.neucom.2018.09.082) (cf. p. 59).
- SAITO, T. et M. REHMSMEIER (2015). « The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets ». In : *PLoS ONE* 10.3, e0118432. ISSN : 1932-6203. DOI : [10.1371/journal.pone.0118432](https://doi.org/10.1371/journal.pone.0118432) (cf. p. 133).
- SAKOE, H. et S. CHIBA (1978). « Dynamic programming algorithm optimization for spoken word recognition ». In : *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26.1, p. 43-49. ISSN : 0096-3518. DOI : [10.1109/TASSP.1978.1163055](https://doi.org/10.1109/TASSP.1978.1163055) (cf. p. 121).
- SALMAN, T. M. et al. (2018). « Forecasting Periods of Strong Southward Magnetic Field Following Interplanetary Shocks ». In : *Space Weather* 16.12, p. 2004-2021. ISSN : 1542-7390. DOI : [10.1029/2018SW002056](https://doi.org/10.1029/2018SW002056) (cf. p. 87).
- SALVATELLI, V. et al. (2019). « Using U-Nets to Create High-Fidelity Virtual Observations of the Solar Corona ». In : *arXiv :1911.04006 [astro-ph, physics :physics]*. arXiv : [1911.04006](https://arxiv.org/abs/1911.04006) (cf. p. 172).
- SAMUEL, A. L. (1959). « Some studies in machine learning using the game of checkers ». In : *IBM Journal of research and development* 3.3, p. 210-229 (cf. p. 41).
- SANTOS, L. F. G. d. et al. (2021). « Multi-Channel Auto-Calibration for the Atmospheric Imaging Assembly using Machine Learning ». In : *arXiv :2012.14023 [astro-ph, physics :physics]*. arXiv : [2012.14023](https://arxiv.org/abs/2012.14023) (cf. p. 172).
- SAUVAUD, J.-A. (2002). « Earth Magnetosphere ». In : *Space environment : prevention of risks related to spacecraft charging*. Toulouse : Cépaduès. ISBN : 978-2-85428-579-6 (cf. p. 24).
- SCHERER, D., A. MÜLLER et S. BEHNKE (2010). « Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition ». In : *Artificial Neural Networks – ICANN 2010*. Sous la dir. de K. DIAMANTARAS, W. DUCH et L. S. ILLIADIS. Lecture Notes in Computer Science. Berlin, Heidelberg : Springer, p. 92-101. ISBN : 978-3-642-15825-4. DOI : [10.1007/978-3-642-15825-4_10](https://doi.org/10.1007/978-3-642-15825-4_10) (cf. p. 60).
- SCHERRER, P. H. et al. (1995). « The Solar Oscillations Investigation — Michelson Doppler Imager ». In : *The SOHO Mission*. Sous la dir. de B. FLECK, V. DOMINGO et A. POLAND. Dordrecht : Springer Netherlands, p. 129-188. ISBN : 978-94-009-0191-9. DOI : [10.1007/978-94-009-0191-9_5](https://doi.org/10.1007/978-94-009-0191-9_5) (cf. p. 69).
- SCHMIDHUBER, J. (2015). « Deep learning in neural networks : An overview ». In : *Neural Networks* 61, p. 85-117. ISSN : 0893-6080. DOI : [10.1016/j.neunet.2014.09.003](https://doi.org/10.1016/j.neunet.2014.09.003) (cf. p. 47).
- SCHRIJVER, C. J. et al. (2012). « Estimating the frequency of extremely energetic solar events, based on solar, stellar, lunar, and terrestrial records ». In : *Journal of Geophysical Research : Space Physics* 117 (A8). ISSN : 2156-2202. DOI : [10.1029/2012JA017706](https://doi.org/10.1029/2012JA017706) (cf. p. 16).

- SCOLINI, C. et al. (2019). « Observation-based modelling of magnetised coronal mass ejections with EUHFORIA ». In : *Astronomy & Astrophysics* 626, A122. ISSN : 0004-6361, 1432-0746. DOI : [10.1051/0004-6361/201935053](https://doi.org/10.1051/0004-6361/201935053) (cf. p. 177).
- SEXTON, E. S., K. NYKYRI et X. MA (2019). « Kp forecasting with a recurrent neural network ». In : *Journal of Space Weather and Space Climate* 9, A19. ISSN : 2115-7251. DOI : [10.1051/swsc/2019020](https://doi.org/10.1051/swsc/2019020) (cf. p. 89, 111, 167).
- SHERSTINSKY, A. (2020). « Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network ». In : *Physica D : Nonlinear Phenomena* 404, p. 132306. ISSN : 0167-2789. DOI : [10.1016/j.physd.2019.132306](https://doi.org/10.1016/j.physd.2019.132306) (cf. p. 100).
- SHIOTA, D. et R. KATAOKA (2016). « Magnetohydrodynamic simulation of interplanetary propagation of multiple coronal mass ejections with internal magnetic flux rope (SUSANOO-CME) ». In : *Space Weather* 14.2, p. 56-75. ISSN : 1542-7390. DOI : [10.1002/2015SW001308](https://doi.org/10.1002/2015SW001308) (cf. p. 87).
- SHPRITS, Y. Y., R. VASILE et I. S. ZHELAVSKAYA (2019). « Nowcasting and Predicting the Kp Index Using Historical Values and Real-Time Observations ». In : *Space Weather* 17.8, p. 1219-1229. ISSN : 1542-7390. DOI : [10.1029/2018SW002141](https://doi.org/10.1029/2018SW002141) (cf. p. 89, 111, 136, 202).
- SHUE, J.-H. et al. (1998). « Magnetopause location under extreme solar wind conditions ». In : *Journal of Geophysical Research : Space Physics* 103 (A8), p. 17691-17700. ISSN : 2156-2202. DOI : [10.1029/98JA01103](https://doi.org/10.1029/98JA01103) (cf. p. 32).
- SICARD-PIET, A. et al. (2014). « Effect of plasma density on diffusion rates due to wave particle interactions with chorus and plasmaspheric hiss : extreme event analysis ». In : *Annales Geophysicae* 32.8, p. 1059-1071. ISSN : 0992-7689. DOI : [10.5194/angeo-32-1059-2014](https://doi.org/10.5194/angeo-32-1059-2014) (cf. p. 32, 89).
- SILVER, D. et al. (2016). « Mastering the game of Go with deep neural networks and tree search ». In : *Nature* 529.7587, p. 484-489. ISSN : 1476-4687. DOI : [10.1038/nature16961](https://doi.org/10.1038/nature16961) (cf. p. 42).
- SILVER, D. et al. (2017). « Mastering the game of Go without human knowledge ». In : *Nature* 550.7676, p. 354-359. ISSN : 1476-4687. DOI : [10.1038/nature24270](https://doi.org/10.1038/nature24270) (cf. p. 42).
- SIMONYAN, K. et A. ZISSERMAN (2015). « Very Deep Convolutional Networks for Large-Scale Image Recognition ». In : *arXiv :1409.1556 [cs]*. arXiv : [1409.1556](https://arxiv.org/abs/1409.1556) (cf. p. 186).
- SISCOE, G. L., N. U. CROOKER et C. R. CLAUER (2006). « Dst of the Carrington storm of 1859 ». In : *Advances in Space Research. The Great Historical Geomagnetic Storm of 1859 : A Modern Look* 38.2, p. 173-179. ISSN : 0273-1177. DOI : [10.1016/j.asr.2005.02.102](https://doi.org/10.1016/j.asr.2005.02.102) (cf. p. 33, 147).
- SMIRNOV, A. G. et al. (2020). « Medium Energy Electron Flux in Earth's Outer Radiation Belt (MERLIN) : A Machine Learning Model ». In : *Space Weather* 18.11, e2020SW002532. ISSN : 1542-7390. DOI : <https://doi.org/10.1029/2020SW002532> (cf. p. 90).
- SMOLA, A. J. et B. SCHÖLKOPF (2004). « A tutorial on support vector regression ». In : *Statistics and Computing* 14.3, p. 199-222. ISSN : 1573-1375. DOI : [10.1023/B:STCO.0000035301.49549.88](https://doi.org/10.1023/B:STCO.0000035301.49549.88) (cf. p. 41).
- SOBOL', I. M. (1967). « On the distribution of points in a cube and the approximate evaluation of integrals ». In : *USSR Computational Mathematics and Mathematical Physics* 7.4, p. 86-112. ISSN : 0041-5553. DOI : [10.1016/0041-5553\(67\)90144-9](https://doi.org/10.1016/0041-5553(67)90144-9) (cf. p. 264).

-
- SUGIURA, M. (1964). « Hourly values of equatorial Dst for the IGY ». In : *Ann. Int. Geophys. Yr.* Vol : 35 (cf. p. 80).
- SUTTON, R. S. (1988). « Learning to predict by the methods of temporal differences ». In : *Machine Learning* 3.1, p. 9-44. ISSN : 1573-0565. DOI : [10.1007/BF00115009](https://doi.org/10.1007/BF00115009) (cf. p. 42).
- SUTTON, R. S. et A. G. BARTO (2018). *Reinforcement Learning, second edition : An Introduction.* second edition. Cambridge, Massachusetts : Bradford Books. 552 p. ISBN : 978-0-262-03924-6 (cf. p. 43).
- SZEGEDY, C. et al. (2014). « Going Deeper with Convolutions ». In : *arXiv :1409.4842 [cs]*. arXiv : [1409.4842](https://arxiv.org/abs/1409.4842) (cf. p. 61, 185, 186).
- SZENICER, A. et al. (2019). « A deep learning virtual instrument for monitoring extreme UV solar spectral irradiance ». In : *Science Advances* 5.10. DOI : [10.1126/sciadv.aaw6548](https://doi.org/10.1126/sciadv.aaw6548) (cf. p. 172).
- TAN, Y. et al. (2018). « Geomagnetic Index Kp Forecasting With LSTM ». In : *Space Weather* 16.4, p. 406-416. ISSN : 1542-7390. DOI : [10.1002/2017SW001764](https://doi.org/10.1002/2017SW001764) (cf. p. 88, 112, 136).
- TASISTRO-HART, A., A. GRAYVER et A. KUVSHINOV (2021). « Probabilistic Geomagnetic Storm Forecasting via Deep Learning ». In : *Journal of Geophysical Research : Space Physics* 126.1, e2020JA028228. ISSN : 2169-9402. DOI : [10.1029/2020JA028228](https://doi.org/10.1029/2020JA028228) (cf. p. 128, 175, 235).
- TAVENARD, R. et al. (2020). « Tsllearn, a machine learning toolkit for time series data ». In : *Journal of Machine Learning Research* 21.118, p. 1-6 (cf. p. 122).
- THE HDF GROUP (2000). *Hierarchical data format version 5.* URL : <http://www.hdfgroup.org/HDF5> (cf. p. 183).
- THOMSON, A. W. P., E. B. DAWSON et S. J. REAY (2011). « Quantifying extreme behavior in geomagnetic activity : EXTREME GEOMAGNETIC STORMS ». In : *Space Weather* 9.10. ISSN : 15427390. DOI : [10.1029/2011SW000696](https://doi.org/10.1029/2011SW000696) (cf. p. 147).
- TLATOV, A. et al. (2020). « Prediction of Solar Flares and Background Fluxes of X-Ray Radiation According to Synoptic Ground-Based Observations Using Machine-Learning Models ». In : *Cosmic Research* 58.6, p. 444-449. DOI : [10.1134/S0010952520060106](https://doi.org/10.1134/S0010952520060106) (cf. p. 174).
- TÖRÖK, T. et al. (2018). « Sun-to-Earth MHD Simulation of the 2000 July 14 “Bastille Day” Eruption ». In : *The Astrophysical Journal* 856.1, p. 75. ISSN : 0004-637X. DOI : [10.3847/1538-4357/aab36d](https://doi.org/10.3847/1538-4357/aab36d) (cf. p. 87).
- TROSHICHEV, O. A., N. P. DMITRIEVA et B. M. KUZNETSOV (1979). « Polar cap magnetic activity as a signature of substorm development ». In : *Planetary and Space Science* 27.3, p. 217-221. ISSN : 0032-0633. DOI : [10.1016/0032-0633\(79\)90063-1](https://doi.org/10.1016/0032-0633(79)90063-1) (cf. p. 81).
- TU, W. et al. (2013). « Modeling radiation belt electron dynamics during GEM challenge intervals with the DREAM3D diffusion model ». In : *Journal of Geophysical Research : Space Physics* 118.10, p. 6197-6211. ISSN : 2169-9402. DOI : [10.1002/jgra.50560](https://doi.org/10.1002/jgra.50560) (cf. p. 89).
- TURC, L. (2014). « Interaction des nuages magnétiques éjectés par le Soleil avec l’environnement terrestre ». Thèse de doct. Ecole Polytechnique (cf. p. 24).
- TURING, A. M. (1950). « Computing Machinery and Intelligence ». In : *Mind* LIX.236, p. 433-460. ISSN : 0026-4423. DOI : [10.1093/mind/LIX.236.433](https://doi.org/10.1093/mind/LIX.236.433) (cf. p. 41).

- UNGER, S. (2019). « The Impact of Space Weather on Human Health ». In : *Biomedical Journal of Scientific & Technical Research* 22.1, p. 16442-16443. ISSN : 2574-1241. DOI : [10.26717/BJSTR.2019.22.003709](https://doi.org/10.26717/BJSTR.2019.22.003709) (cf. p. 35).
- UPENDRAN, V. et al. (2020). « Solar Wind Prediction Using Deep Learning ». In : *Space Weather* 18.9, e2020SW002478. ISSN : 1542-7390. DOI : [10.1029/2020SW002478](https://doi.org/10.1029/2020SW002478) (cf. p. 86, 175, 180).
- VALLANCE, L. et al. (2017). « Towards a standardized procedure to assess solar forecast accuracy : A new ramp and time alignment metric ». In : *Solar Energy* 150, p. 408-422. ISSN : 0038-092X. DOI : [10.1016/j.solener.2017.04.064](https://doi.org/10.1016/j.solener.2017.04.064) (cf. p. 121).
- VASWANI, A. et al. (2017). « Attention Is All You Need ». In : *arXiv :1706.03762 [cs]*. arXiv : [1706.03762](https://arxiv.org/abs/1706.03762) (cf. p. 235).
- VAUTARD, R., P. YIOU et M. GHIL (1992). « Singular-spectrum analysis : A toolkit for short, noisy chaotic signals ». In : *Physica D : Nonlinear Phenomena* 58.1, p. 95-126. ISSN : 0167-2789. DOI : [10.1016/0167-2789\(92\)90103-T](https://doi.org/10.1016/0167-2789(92)90103-T) (cf. p. 263).
- WES MCKINNEY (2010). « Data Structures for Statistical Computing in Python ». In : *Proceedings of the 9th Python in Science Conference*. Sous la dir. de S. van der WALT et JARROD MILLMAN, p. 56-61. DOI : [10.25080/Majora-92bf1922-00a](https://doi.org/10.25080/Majora-92bf1922-00a) (cf. p. 100).
- WILLARD, J. et al. (2021). « Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems ». In : *arXiv :2003.04919 [physics, stat]*. arXiv : [2003.04919](https://arxiv.org/abs/2003.04919) (cf. p. 40).
- WILLIAMS, C. K. et C. E. RASMUSSEN (2006). *Gaussian processes for machine learning*. T. 2. 3. MIT (cf. p. 41, 126).
- WILSON, A. G. et al. (2015). « Deep Kernel Learning ». In : *arXiv :1511.02222 [cs, stat]*. arXiv : [1511.02222](https://arxiv.org/abs/1511.02222) (cf. p. 126).
- WILSON, A. G. et al. (2016). « Stochastic Variational Deep Kernel Learning ». In : *arXiv :1611.00336 [cs, stat]*. arXiv : [1611.00336](https://arxiv.org/abs/1611.00336) (cf. p. 126).
- WING, S. et al. (2005). « Kp forecast models ». In : *Journal of Geophysical Research : Space Physics* 110 (A4). ISSN : 2156-2202. DOI : [10.1029/2004JA010500](https://doi.org/10.1029/2004JA010500) (cf. p. 87).
- WINTOFT, P. et al. (2017). « Forecasting Kp from solar wind data : input parameter study using 3-hour averages and 3-hour range values ». In : *Journal of Space Weather and Space Climate* 7, A29. ISSN : 2115-7251. DOI : [10.1051/swsc/2017027](https://doi.org/10.1051/swsc/2017027) (cf. p. 42, 88).
- WMO (2019). *Manual on the Global Data-processing and Forecasting System : Annex IV to the WMO Technical Regulations*. 2019 edition. WMO. Geneva : WMO. 148 p. ISBN : 978-92-63-10485-4 (cf. p. 200).
- WOODS, T. N. et al. (2012). « Extreme Ultraviolet Variability Experiment (EVE) on the Solar Dynamics Observatory (SDO) : Overview of Science Objectives, Instrument Design, Data Products, and Model Developments ». In : *The Solar Dynamics Observatory*. Sous la dir. de P. CHAMBERLIN, W. D. PESNELL et B. THOMPSON. New York, NY : Springer US, p. 115-143. ISBN : 978-1-4614-3673-7. DOI : [10.1007/978-1-4614-3673-7_7](https://doi.org/10.1007/978-1-4614-3673-7_7) (cf. p. 73).
- WORDEN, J. et J. HARVEY (2000). « An Evolving Synoptic Magnetic Flux map and Implications for the Distribution of Photospheric Magnetic Flux ». In : *Solar Physics* 195.2, p. 247-268. ISSN : 1573-093X. DOI : [10.1023/A:1005272502885](https://doi.org/10.1023/A:1005272502885) (cf. p. 76).
- WU, J.-G. et H. LUNDSTEDT (1997). « Geomagnetic storm predictions from solar wind data with the use of dynamic neural networks ». In : *Journal of Geophysical Research :*

-
- Space Physics* 102 (A7), p. 14255-14268. ISSN : 2156-2202. DOI : [10.1029/97JA00975](https://doi.org/10.1029/97JA00975) (cf. p. 87).
- WUELSE, J.-P. et al. (2004). « EUVI : the STEREO-SECCHI extreme ultraviolet imager ». In : *Telescopes and Instrumentation for Solar Astrophysics*. Telescopes and Instrumentation for Solar Astrophysics. T. 5171. International Society for Optics et Photonics, p. 111-122. DOI : [10.1117/12.506877](https://doi.org/10.1117/12.506877) (cf. p. 70).
- XU, J. et al. (2019). « Characteristics of High-Energy Proton Responses to Geomagnetic Activities in the Inner Radiation Belt Observed by the RBSP Satellite ». In : *Journal of Geophysical Research : Space Physics* 124.9, p. 7581-7591. ISSN : 2169-9402. DOI : [10.1029/2018JA026205](https://doi.org/10.1029/2018JA026205) (cf. p. 81).
- YURCHYSHYN, V. et al. (2005). « Statistical Distributions of Speeds of Coronal Mass Ejections ». In : *The Astrophysical Journal* 619.1, p. 599. ISSN : 0004-637X. DOI : [10.1086/426129](https://doi.org/10.1086/426129) (cf. p. 176).
- ZENCHENKO, T. A. et T. K. BREUS (2021). « The Possible Effect of Space Weather Factors on Various Physiological Systems of the Human Organism ». In : *Atmosphere* 12.3, p. 346. ISSN : 2073-4433. DOI : [10.3390/atmos12030346](https://doi.org/10.3390/atmos12030346) (cf. p. 35).
- ZHANG, P.-J., C.-B. WANG et G.-S. PU (2020). « Generating a radioheliograph image from SDO/AIA data with the machine learning method ». In : *Research in Astronomy and Astrophysics* 20.12. DOI : [10.1088/1674-4527/20/12/204](https://doi.org/10.1088/1674-4527/20/12/204) (cf. p. 172).
- ZHANG, X., J. LIU et Q. WANG (2011). « Image feature extraction for solar flare prediction ». In : Proceedings - 4th International Congress on Image and Signal Processing, CISP 2011. T. 2, p. 910-914. DOI : [10.1109/CISP.2011.6100295](https://doi.org/10.1109/CISP.2011.6100295) (cf. p. 173, 174).
- ZHAO, H. et X. LI (2013). « Inward shift of outer radiation belt electrons as a function of Dst index and the influence of the solar wind on electron injections into the slot region ». In : *Journal of Geophysical Research : Space Physics* 118.2, p. 756-764. ISSN : 2169-9402. DOI : [10.1029/2012JA018179](https://doi.org/10.1029/2012JA018179) (cf. p. 81).
- ZHENG, Y., X. LI et X. WANG (2019). « Solar Flare Prediction with the Hybrid Deep Convolutional Neural Network ». In : *Astrophysical Journal* 885.1. DOI : [10.3847/1538-4357/ab46bd](https://doi.org/10.3847/1538-4357/ab46bd) (cf. p. 174).

Annexes

A Description de la méthode de reconstruction des données manquantes par analyse spectrale singulière

Pour remplacer les données manquantes dans nos jeux de données de paramètres du vent solaire, nous avons utilisé la méthode de reconstruction de données par analyse spectrale singulière ou *Singular Spectrum Analysis* (SSA) décrite par KONDRASHOV et al. 2010. La SSA est, dans ses généralités, très bien décrite dans GHIL et al. 2002. Il s'agit d'une méthode d'estimation spectrale non-paramétrique. Il s'agit de reconstruire les trous temporels de séries de données en utilisant les corrélations spatio-temporelles pour mettre en avant des modes cohérents, tout en rejetant le bruit.

Pour appliquer la méthode concrètement on doit suivre les étapes suivantes :

- on sélectionne la série temporelle que l'on souhaite reconstituer, ainsi qu'un ensemble d'autres séries temporelles que l'on sait bien corrélées avec la première et qui ne possèdent pas de trous (séries de référence) ;
- on remplace les données manquantes par des 0 ;
- on calcule la matrice de covariance de la matrice composée de l'ensemble des séries de données pour des décalages temporels inférieurs à une fenêtre M ;
- on calcule les valeurs propres et vecteurs propres de la matrice de covariance ;
- on ne garde que K composantes principales, celles qui expliquent donc la plus grande partie de la variance des données ;
- on projette les composantes principales sur les vecteurs propres pour reconstruire les séries temporelles ;
- dans la matrice initiale des séries temporelles trouées, on remplace les 0 des données manquantes par les valeurs reconstituées ;
- on recommence jusqu'à ce qu'un critère de convergence des valeurs reconstituées ait été atteint.

Ici nous avons utilisé les indices géomagnétiques Kp et Dst comme séries de référence. Toutes les étapes de la méthode sont appliquées automatiquement par un logiciel appelé SSA-MTM toolkit (VAUTARD et al. 1992), mis à disposition publiquement et gratuitement en ligne¹. Mais nous devons manuellement trouver les valeurs des paramètres M et K qui optimisent la reconstruction, ce qui n'est

1. <https://dept.atmos.ucla.edu/tcd/download>

pas trivial. (KONDRASHOV et al. 2014) propose quelques valeurs pour certains paramètres du vent solaire, mais pas tous ceux dont nous avons besoin. Nous avons donc mené une nouvelle recherche de paramètres optimaux, en incluant des données plus récentes que (KONDRASHOV et al. 2014).

Pour trouver les paramètres M et K optimaux pour chaque paramètre du vent solaire, nous avons créé des trous temporels artificiels dans chaque série temporelle pour la période 2008 – 2018. Les lacunes artificielles sont des répliques des lacunes réelles trouvées dans les mêmes séries temporelles, mais au cours de la période 1984 – 1994, de sorte que la distribution et la longueur des lacunes artificielles soient plausibles. Ensuite, nous avons cherché les valeurs optimales de M et de K qui permettaient la meilleure reconstruction des écarts du point de vue de la RMSE et de la corrélation linéaire. Les paramètres optimaux ont été trouvés en effectuant une recherche itérative sur une grille à l’aide de séquences de Sobol quasi-aléatoires à faible discrédance (SOBOL’ 1967) générées par le package pour langage Python `OpenTURNS` (BAUDIN et al. 2015). L’ensemble des paramètres optimaux (à savoir M^* et K^*) trouvés pour chaque série temporelle est indiqué dans le Tableau A.1.

TABLEAU A.1 – Valeurs optimales de M^* et K^* trouvées pour reconstruire les données manquantes dans les séries temporelles de cinq paramètres du vent solaire.

Paramètre du vent solaire	M^*	K^*
V_{sw}	110	29
ρ_{sw}	12	30
T_{sw}	12	29
B_z	9	17
$ B $	7	19

B Éléments supplémentaires d'évaluation du modèle *SERENADE-Ca*

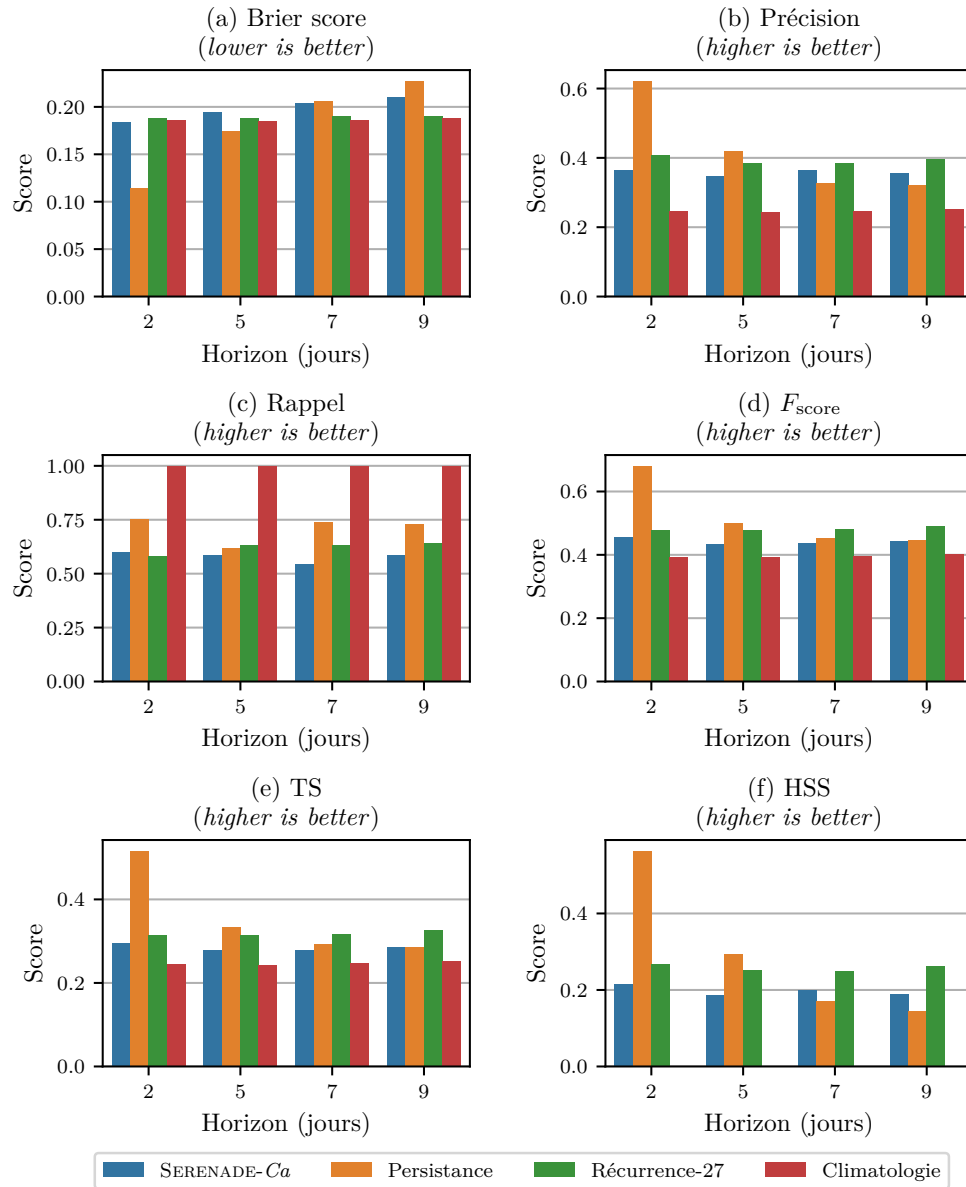


FIGURE B.1 – Métriques d'évaluation du système d'alerte dérivé de *SERENADE-Ca* et des trois modèles de référence pour tous les horizons de prédiction considérés. À l'exception du *Brier Score*, plus la valeur de la métrique est élevée, plus le modèle est performant.

Note : Dans cette figure, on constate que les résultats pour le modèle de Récurrence-27 varient légèrement en fonction de l'horizon, alors que cela ne devrait pas être le cas. Cela est dû aux trous de données dans le jeu de données *SDO/AIA – ML-READY*, à cause desquels nous n'avons pas de prédictions pour tous les instants. Cela crée des petites différences sur les bords de notre jeu d'évaluation en fonction de l'horizon, suffisantes pour qu'on observe les petites variations de valeurs que nous venons de mentionner.

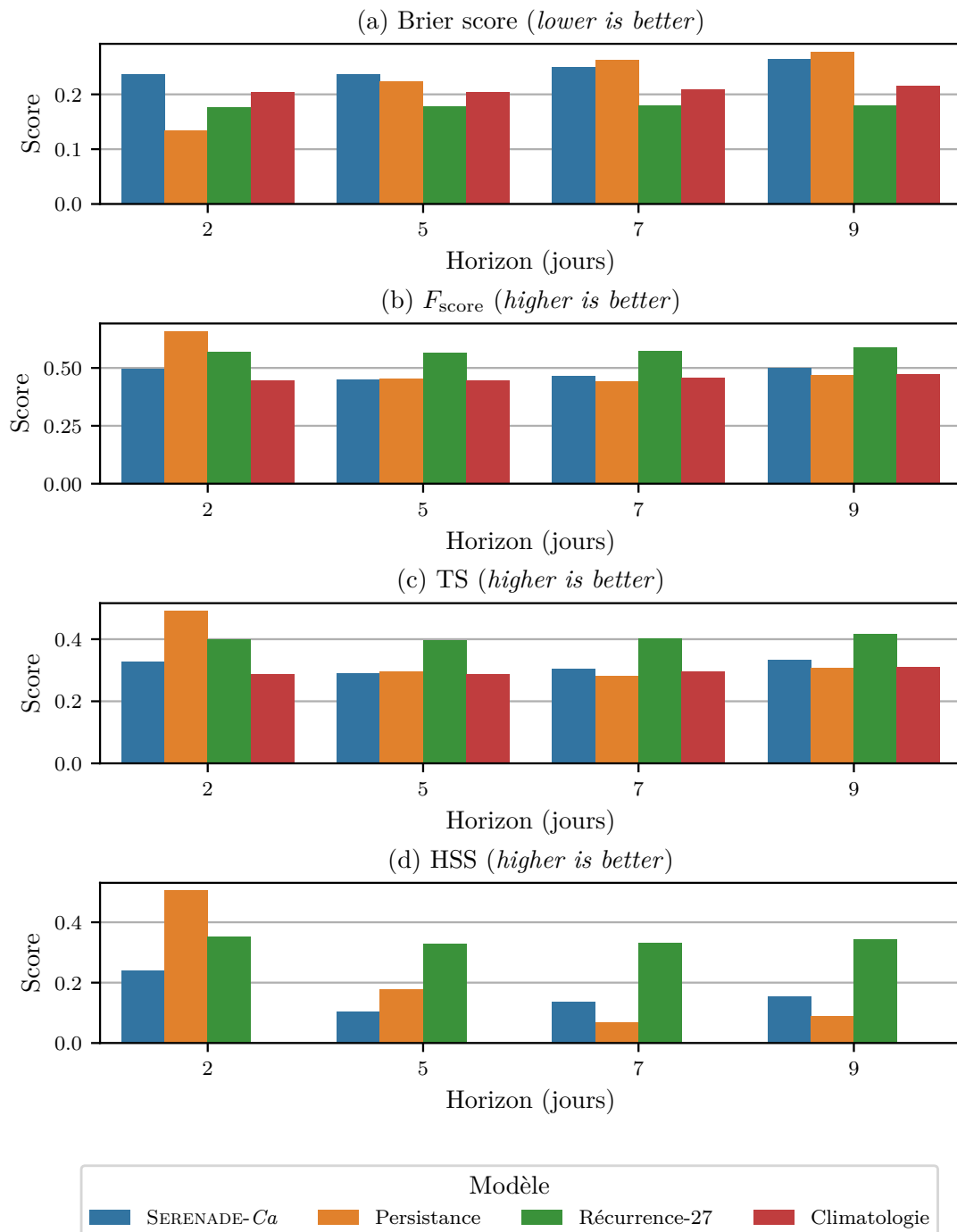


FIGURE B.2 – Évaluation de *SERENADE-Ca* et des modèles de référence en fonction de l’horizon de prédiction lors de la phase descendante du cycle solaire, à l’aide de métriques dédiées à l’évaluation du problème de classification binaire.

C Articles publiés dans des revues à comité de lecture

C.1 Bernoux et Maget, 2020

G. Bernoux and V. Maget (2020). “Characterizing Extreme Geomagnetic Storms Using Extreme Value Analysis: A Discussion on the Representativeness of Short Data Sets”. In: *Space Weather* 18.6, e2020SW002450. ISSN: 1542-7390. DOI: [10.1029/2020SW002450](https://doi.org/10.1029/2020SW002450)

Abstract – One of the main goals when studying Space Weather is to characterize extreme events occurrences and related characteristics. To do so, dedicated statistical methods from the so-called extreme value analysis (EVA) field have been developed. In this study we used Ca index, derived from aa, in order to characterize geoeffectiveness from the radiation belts point of view with a 150-year-long data set. The analysis performed in this study thus focuses on this newsworthy index to provide clues on the reliability of EVA methods. The first main result we present here is that the 1-in-10-, 1-in-50-, and 1-in-100-year events, respectively, match Ca values of 100.39, 131.39, and 142.84 nT. Consequently, the only 1-in-100 event observed during the Space Era would be the “Halloween Storm” in 2003 that reached a Ca value of 147.6 nT. The second main result highlighted in this work is that performing the same analysis with shorter subsets (20 years long) can give significantly different results for two reasons. The first reason is that some short time periods do not display the same distribution of events as the full period. The second reason is that the choice of the correct threshold (when using a Peaks Over Threshold approach) is made difficult with a short data set and leads to inaccurate results. This is a strong result as for accurate estimation of the induced effects of extreme events in radiation belts, we may only rely on short flux data sets from one or another mission (mostly shorter than 20 years).

C.2 Bernoux et al. 2021

G. Bernoux, A. Brunet, É. Buchlin, M. Janvier, and A. Sicard (2021). “An operational approach to forecast the Earth’s radiation belts dynamics”. In: *Journal of Space Weather and Space Climate* 11. Publisher: EDP Sciences, p. 60. ISSN: 2115-7251. DOI: [10.1051/swsc/2021045](https://doi.org/10.1051/swsc/2021045)

Abstract – The Ca index is a time-integrated geomagnetic index that correlates well with the dynamics of high-energy electron fluxes in the outer radiation belts. Therefore, Ca can be used as an indicator for the state of filling of the radiation belts for those electrons. Ca also has the advantage of being a ground-based measurement with extensive historical records. In this work, we propose a data-driven model to forecast Ca up to 24 h in advance from near-Earth solar wind parameters. Our model relies mainly on a recurrent neural network architecture called Long Short Term Memory that has shown good performances in forecasting other geomagnetic indices in previous papers. Most implementation choices in this study were arbitrated from the point of view of a space system operator, including

the data selection and split, the definition of a binary classification threshold, and the evaluation methodology. We evaluate our model (against a linear baseline) using both classical and novel (in the space weather field) measures. In particular, we use the Temporal Distortion Mix (TDM) to assess the propensity of two time series to exhibit time lags. We also evaluate the ability of our model to detect storm onsets during quiet periods. It is shown that our model has high overall accuracy, with evaluation measures deteriorating in a smooth and slow trend over time. However, using the TDM and binary classification forecast evaluation metrics, we show that the forecasts lose some of their usefulness in an operational context even for time horizons shorter than 6 h. This behaviour was not observable when evaluating the model only with metrics such as the root-mean-square error or the Pearson linear correlation. Considering the physics of the problem, this result is not surprising and suggests that the use of more spatially remote data (such as solar imaging) could improve space weather forecasts.

C.3 Bernoux et al. 2022

G. Bernoux, A. Brunet, É. Buchlin, M. Janvier, and A. Sicard (2022a). “Forecasting the Geomagnetic Activity Several Days in Advance Using Neural Networks Driven by Solar EUV Imaging”. In: *Journal of Geophysical Research: Space Physics* 127.10, e2022JA030868. ISSN: 2169-9402. DOI: [10.1029/2022JA030868](https://doi.org/10.1029/2022JA030868)

Abstract – Many models of the near-Earth’s space environment (radiation belts, ionosphere, upper atmosphere, etc.) are driven by geomagnetic indices, representing the state of disturbance of the Earth’s magnetosphere. Over the past decade, machine learning-based methods for forecasting geomagnetic indices from near-Earth solar wind parameters have become popular in the space weather community. These methods often prove to be very accurate and skilled. However, these approaches have the notable drawback of being effective in an operational context only for limited forecasting horizons (often up to a couple of hours ahead at best). In order to increase this prediction horizon, we introduce SERENADE, a novel deep learning-based proof-of-concept model using images delivered by the Atmospheric Imaging Assembly instrument onboard the Solar Dynamics Observatory spacecraft to directly provide probabilistic forecasts of the daily maximum of the geomagnetic index Kp up to a few days ahead. We show in particular that SERENADE is able to capture information on the geomagnetic dynamics from solar imaging alone. In addition, despite it being a prototypical model, our model is more accurate in most situations than three empirical baseline models. However, the model still shows some strong limitations inherent to its structure and the used data set, which could be the focus of future works. This opens the way to a better mid-to-long term data-driven magnetospheric modeling within space weather and geophysical pipelines.