

RAPPORT DE STAGE PROJET DE FIN D'ÉTUDES
ÉCOLE NATIONALE DE LA MÉTÉOROLOGIE

Delphine TECHER

**Détection de la réponse climatique aux
forçages anthropiques par intelligence
artificielle**

Encadré par :

Anne DURIF (CECI, CERFACS)
Laurent TERRAY (CECI, CERFACS)

"All models are wrong but some are useful."
— George E. P. Box

Résumé

La détection du changement climatique à échelle régionale reste aujourd’hui délicate. Le signal du changement climatique peut être masqué ou amplifié par la variabilité interne, difficilement identifiable dans les données d’observations. Des méthodes dites "classiques" ont été élaborées, notamment sur des modèles de climat, afin d’isoler au mieux le bruit que représente la variabilité interne, du signal qu’est le changement climatique.

L’avènement des méthodes d’intelligence artificielle appliquées aux sciences climatiques a mis à disposition de nouveaux outils pour différencier le changement climatique de la variabilité interne. Une méthode de débruitage d’images avec de l’IA appelée "Noise2Noise" est développée dans le cadre de ce stage pour répondre à cette problématique de distinction du bruit (variabilité interne) par rapport à un signal (changement climatique).

L’architecture du réseau est un U-Net s’entraînant sur les données de modèles climatiques afin d’être appliqué aux observations. Plusieurs étapes de tests et d’ajustements du réseau sur les modèles de climat ont abouti au développement de cinq U-Nets. Le réseau de neurones présentant les meilleures performances a ensuite été comparé aux méthodes "classiques" afin d’être évalué.

Aucune méthode ne se démarque clairement des autres. La performance des méthodes étudiées dépend fortement de la période temporelle et de la zone géographique étudiées. Le réseau de neurones présente des résultats équivalents à certaines méthodes classiques dans le cadre idéalisé des modèles. Son application aux observations identifie bien les principaux motifs du changement climatique. Les pistes d’amélioration du réseau de neurones sont nombreuses et sa marge de progression est encourageante pour la poursuite de ce travail.

Mots-clés : *changement climatique, signal, variabilité interne, débruitage, réseau de neurones*

Remerciements

Je tiens à remercier mes encadrants, Anne Durif et Laurent Terray. Anne, j'ai fait mes premiers pas dans le domaine de l'IA avec toi. Merci pour ta patience et ton enthousiasme permanent qui m'ont remotivée à chaque impasse que je pensais rencontrer. Merci pour ta flexibilité à toute épreuve, que ce soit pendant tes vacances ou après la fin de ton contrat, tu ne t'es jamais plainte de mon rythme parfois chaotique. Merci pour tous les fous rires et les récits d'explorations. Laurent, merci tout d'abord pour ce sujet. Tu m'as initiée à cette problématique de détection du changement climatique qui s'est avérée aussi complexe que passionnante. Merci d'aimer autant ce que tu fais, j'espère garder longtemps en mémoire tes conseils sur la méthode scientifique, et plus globalement sur la recherche. Merci pour ta curiosité, ta bonne humeur, et toutes ces discussions sur l'escalade. Enfin, merci à vous deux pour votre soutien infaillible tout au long de ces six mois de stage, il a été précieux.

Je tiens également à remercier toute l'équipe de Globc pour son accueil très chaleureux. Merci à toutes et à tous pour tous ces bons moments passés en salle de pause et au déjeuner. Je retiendrai qu'il faut toujours vérifier que ses simulations ont bien commencé à tourner avant de prendre une heure de pause.

Merci à toute l'équipe de CSG et tout particulièrement à Isabelle d'Ast sans qui aucun de ces UNets n'auraient pu être lancés. Merci pour ta disponibilité et ta réactivité. Merci également de m'avoir mise en contact avec François Courteille de Nvidia que je remercie pour toute l'aide qu'il m'a apportée et pour l'intérêt porté à mon sujet. Vous avez été mes encadrants de l'été.

Je voudrais aussi remercier l'équipe d'organisation de l'IMSC pour m'avoir permis de participer à cette conférence. Le monde du climat est bien vaste et que serait la recherche sans tout le partage et les rencontres qu'elle entraîne...

Merci à Mathis Peyron et Luciano Drozda de l'équipe algo du Cerfacs, mais aussi à Antoine Doury et Enora Cariou du CNRM, et à Céline Bonfils du Lawrence Livermore National Laboratory pour toutes les discussions IA. Merci également à tous les membres du groupe deep learning du lundi soir qui ont facilité mes premiers pas en IA.

Enfin, je remercie Philippe Dandin pour m'avoir conseillé de faire ce stage de fin d'études avec Laurent et sans qui toute cette aventure n'aurait jamais commencé.

Table des matières

Résumé	I
Remerciements	II
Table des matières	III
Introduction	1
1 Introduction à l'intelligence artificielle	4
1.1 Concepts principaux	4
1.1.1 Intelligence artificielle, apprentissage automatique et apprentissage profond	4
1.1.2 Réseau neuronal dense (DNN)	5
1.2 Traitement d'images	7
1.2.1 Réseau neuronal convolutif (CNN)	7
1.2.2 U-Net	9
1.2.3 Débruitage d'images : le Noise2Noise	9
2 Développement du réseau de neurones	11
2.1 Données	11
2.2 Architecture du réseau	11
2.3 Métriques d'évaluation du réseau	12
2.4 Une problématique : cinq U-Nets	13
2.5 Que se passe-t-il du côté de la fonction coût ("loss") ?	18
3 Évaluation de la méthode Noise2Noise	20
3.1 Méthodes "classiques"	20
3.2 Comparaisons des méthodes "classiques" et par apprentissage	21
3.2.1 En moyenne globale	21
3.2.2 Les zones sensibles	21
3.2.3 Et le vainqueur est...	22
3.2.4 Application aux observations	23
4 Conclusion et perspectives	25
4.1 Conclusion	25
4.2 Perspectives	25
Bibliographie	27

Introduction

Un consensus scientifique a été établi concernant l'attribution du changement climatique aux émissions anthropiques. L'augmentation de la température de surface globale a atteint 1.09°C sur la période 2011-2020 par rapport à 1850-1900 d'après le [GIEC \[2023\]](#). Cependant, la détection et l'attribution du changement climatique restent délicates à l'échelle régionale.

Le signal du changement climatique peut être masqué ou amplifié par du bruit : la variabilité interne. La variabilité interne représente les variations observées dans un climat à état stationnaire, sans influence humaine. Ces variations sont causées par des processus physiques naturels non-linéaires, et sont difficilement discernables des variations extrinsèques au système climatique aussi appelées forçages externes anthropiques. On parlera également de réponse forcée (la réponse du système climatique aux forçages externes anthropiques) pour faire référence au changement climatique.

Pour quantifier le changement climatique on dispose des données d'observations qui mesurent l'état actuel de notre climat, mais ces mesures mêlent le signal recherché au bruit qui le masque. La figure 1 (a) montre que ce signal "émerge" difficilement du bruit sur de nombreuses régions, notamment sur l'Arctique. Le signal du réchauffement climatique (b) y est très marqué mais la variabilité interne, également très élevée, masque le changement dans la région. Cela rend la distinction du signal par rapport au bruit moins manifeste sur l'Arctique que sur la zone équatoriale. On considère en général que les températures sont "anormales" lorsque le rapport signal sur bruit dépasse la valeur seuil de 2.

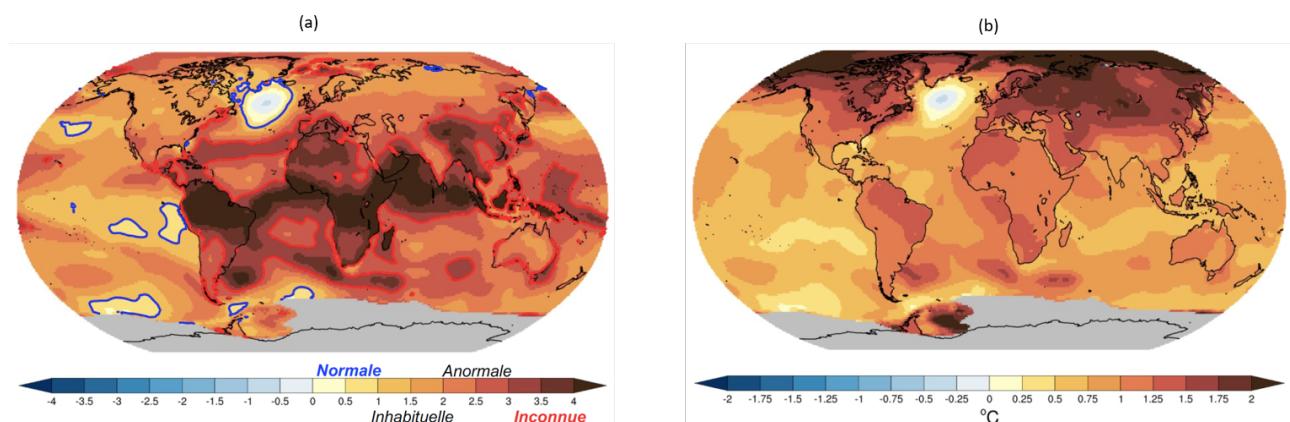


FIGURE 1 – Rapport signal sur bruit (a) des observations de température (Berkeley) avec (b) le signal du changement climatique par rapport à la période 1850-1900 approché par une régression linéaire (équation 1), source : Laurent Terray.

Afin d'isoler le signal du changement climatique (figure 1 (b)), on considère que la température en un point du globe dépend linéairement de la température globale. Le changement est alors le coefficient de la relation et la variabilité interne le résidu (équation 1).

$$T(x, t) = \alpha_x T_{global}(t) + \beta_x \quad (1)$$

avec T la température en un point x au temps t , et T_{global} la température globale.

Néanmoins, cette hypothèse de dépendance linéaire est mise à défaut sur de nombreuses régions où des processus non linéaire pilotent le changement local. Les observations sont également inhomogènes, certaines régions du globe ne disposant pas de séries temporelles de mesures complètes.

Les modèles de climat réalisent alors des simulations afin d'identifier plus précisément la réponse forcée. Ces simulations permettent de mener des "expériences" impossibles à mettre en oeuvre dans le monde réel. De grands ensembles de modèles tels que les SMILEs (Single Model Initial condition Large Ensembles) ont été élaborés. Dans ces ensembles, un modèle de climat est contraint par des forçages externes naturels et anthropiques et sont lancés à partir de conditions initiales. Les différentes conditions initiales imposées dans un modèle de climat permettent d'échantillonner la variabilité interne en générant plusieurs membres, appelés aussi réalisations. Pour un même modèle, la meilleure estimation du changement climatique sera alors la moyenne d'ensemble, la réponse forcée, et la variabilité interne pourra être obtenue via l'écart-type de l'ensemble. On constate sur la figure 2 que la réponse forcée est un signal où les hautes fréquences ont été filtrées des membres de l'ensemble.

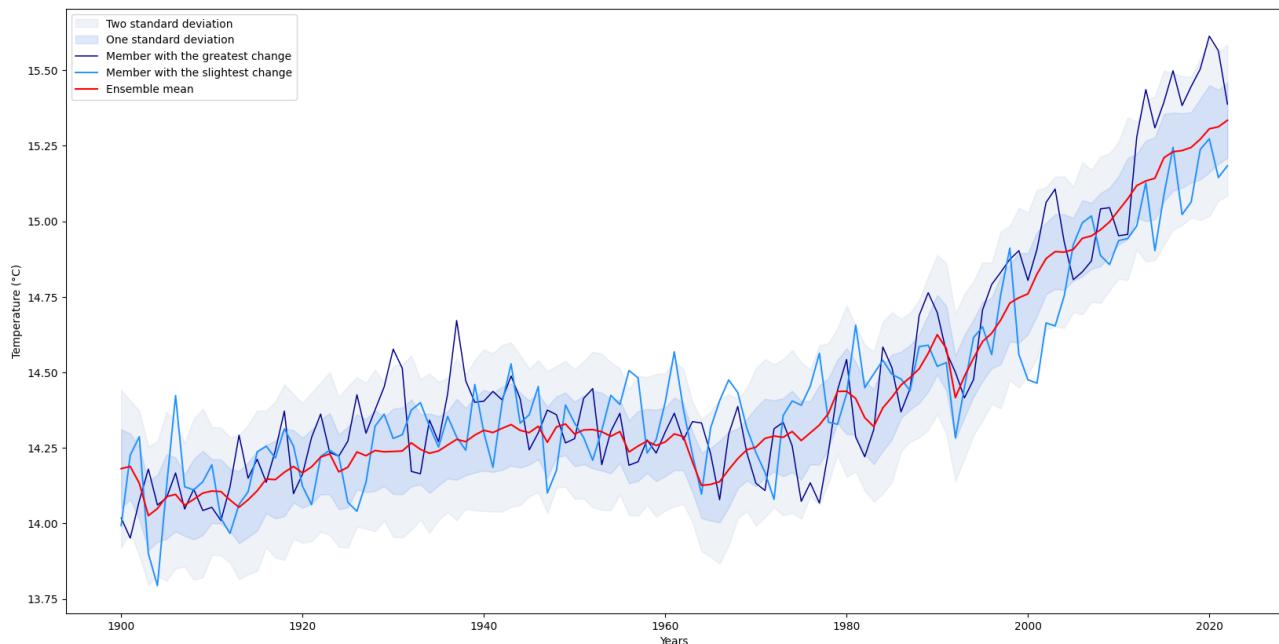


FIGURE 2 – Évolution de la température sur la période 1900-2023 pour la moyenne d'ensemble du modèle CESM2 (rouge), le membre avec un coefficient de tendance linéaire le plus important (bleu foncé), le moins important (bleu clair) sur la période considérée avec en fond gris une fois l'écart-type et en fond bleu deux fois l'écart-type de l'ensemble des membres du modèle.

Ces modèles nous permettent ainsi de « générer plusieurs réalités » afin de pouvoir quantifier au mieux la réponse forcée. Mais ces modèles ne sont pas parfaits et comportent une incertitude épistémique. La connaissance des processus physiques entrant en jeu dans le système climatique est imparfaite (voire absente). L'effet du forçage des gaz à effet de serre est aujourd'hui bien connu mais celui du changement d'utilisation des sols l'est beaucoup moins. Les équations, paramétrisations ou encore la stabilité numérique des modèles ne sont pas parfaites tout comme les conditions initiales utilisées pour lancer les simulations. De plus, la variabilité interne simulée et les réponses aux forçages peuvent être moins fiables à des échelles plus fines qu'à l'échelle mondiale. Le cadre idéalisé des modèles nous permet donc d'avoir une estimation du changement climatique mais celle-ci est imparfaite.

Des méthodes plus travaillées que la moyenne d'ensemble ont été élaborées et permettent de mieux filtrer le signal du changement climatique à partir des modèles de climat mais aussi des observations. Néanmoins, elles présentent toutes des biais, notamment liés aux hypothèses qu'elles impliquent et qui s'ajoutent aux imperfections des différentes sources de données.

Ces dernières années, l'avènement des méthodes d'apprentissage profond en intelligence artificielle (IA) a permis d'étendre et de revisiter des architectures de réseaux de neurones pour les sciences climatiques. Ces méthodes, adaptées à des phénomènes non-linéaires, combinent l'utilisation de simulations climatiques mais aussi des observations afin d'estimer au mieux la réponse forcée en tirant parti des avantages des deux sources de données. [Bône et al. \[2024\]](#) a notamment appliqué une méthode dite de débruitage d'images aux modèles climatiques pour la détection de la réponse climatique aux forçages anthropiques, en considérant le bruit de l'image comme la variabilité interne.

Ce rapport répond à la question suivante : comment appliquer les nouvelles méthodes d'IA, adaptées à des phénomènes non-linéaires aux observations et simulations du système climatique afin de distinguer au mieux le signal du changement climatique par rapport au bruit de la variabilité interne ? Ce rapport s'organise ainsi : la partie 1 est une introduction aux différentes notions portant sur l'intelligence artificielle et sur la méthode de débruitage d'images avec de l'IA. La partie 2 parcourt les différentes étapes de développement d'un réseau de neurones. La partie 3 évalue le réseau développé en le comparant à des méthodes préexistantes pour la distinction de la réponse forcée par rapport à la variabilité interne.

1 | Introduction à l'intelligence artificielle

Ces dernières années, l'intelligence artificielle (IA) a connu un développement sans précédent, notamment dans le domaine de l'IA générative avec l'arrivée de ChatGPT. Les articles portant sur l'intelligence artificielle et ses multiples applications sont maintenant pléthoriques, il faut donc dans un premier temps bien définir l'outil qui a été développé dans le cadre de ce stage et le situer dans cet assortiment de techniques.¹

1.1 Concepts principaux

1.1.1 Intelligence artificielle, apprentissage automatique et apprentissage profond

Le concept d'intelligence artificielle a été développé en 1950 lorsque quelques pionniers du domaine de l'informatique se sont demandé si un ordinateur pourrait « penser ». Une définition concise du domaine serait « l'effort d'automatiser des tâches intellectuelles normalement effectuées par des humains ». L'intelligence artificielle définie ainsi englobe l'apprentissage automatique (ou apprentissage machine) qui lui-même comprend l'apprentissage profond (figure 1.1).

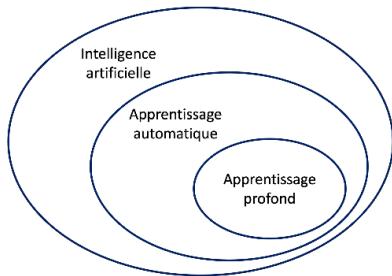


FIGURE 1.1 – Intelligence artificielle, apprentissage automatique, apprentissage profond, source : Chollet [2021]

L'intelligence artificielle s'est ainsi tout d'abord définie comme un ensemble de règles codées et élaborées par des programmeurs qui pensaient qu'une intelligence artificielle de niveau humain ne requerrait qu'un ensemble important de règles explicites. Cette approche, nommée IA symbolique, a rapidement atteint ses limites face à des problèmes plus complexes tels que la classification d'images.

L'apprentissage automatique est ensuite né de l'interrogation suivante : est-ce qu'un ordinateur pourrait aller plus loin « que ce qu'on sait lui dire d'exécuter » et apprendre par lui-même à exécuter une tâche donnée, à nous surprendre ? Cette question a mené à un nouveau paradigme. En programmation classique, l'Homme fournit en entrée des règles (un programme) et des données qui seront traitées selon les règles imposées, puis obtiendra des réponses. La plupart du temps en apprentissage automatique, l'Homme fournit en entrée des données mais aussi les réponses attendues de ces entrées, et le modèle d'IA fournira les règles (figure 1.2). L'idée étant que le modèle d'IA fournisse des réponses initialement inconnues pour de nouvelles données grâce aux règles acquises en amont.

On parlera alors plutôt d'entraînement d'un modèle d'apprentissage automatique plutôt que de programmation. Ce nouveau modèle d'IA est fortement relié aux statistiques mais contrairement à ce domaine des mathématiques, il peut manipuler et assimiler plus facilement d'importantes et complexes

1. Les définitions et figures sont en grande partie issues des manuels sur l'IA de Chollet [2021] et Murphy [2022]

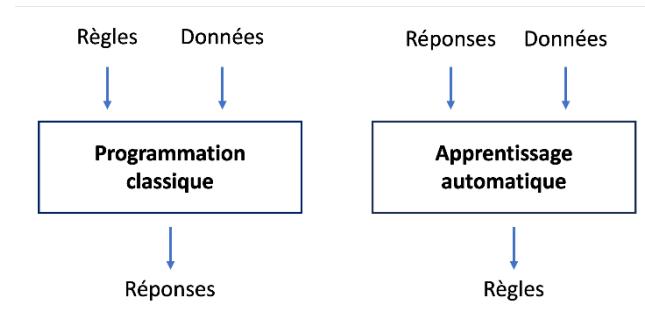


FIGURE 1.2 – Apprentissage automatique, un nouveau paradigme, source : [Chollet \[2021\]](#)

bases de données. Par conséquent, l'apprentissage automatique, et en particulier l'apprentissage profond, présente comparativement peu de théorie mathématique - peut-être trop peu - et est orienté vers l'ingénierie. Il s'agit d'une discipline pratique dans laquelle les idées sont prouvées empiriquement plus souvent que théoriquement.

Quelle est alors la plus-value de l'apprentissage profond par rapport à l'apprentissage automatique ? Le terme « profond » de « l'apprentissage profond » ne fait pas référence à une quelconque compréhension plus profonde obtenue par l'approche, il représente plutôt l'idée de couches successives de représentations qui sont de plus en plus significatives. Le nombre de couches contribuant au modèle d'apprentissage s'appelle la profondeur du modèle d'IA. Cette représentation par couches est presque toujours apprise via des réseaux de neurones. Ce terme fait référence à la neurobiologie, certains concepts centraux de l'apprentissage profond s'inspirant directement de la compréhension actuelle du cerveau. Néanmoins, il serait confus et contre-productif de penser que les réseaux de neurones sont une réplique du fonctionnement de notre cerveau. Dans notre cas, l'apprentissage profond est une structure mathématique permettant d'apprendre des représentations à partir de données au travers de couches successives d'opérations.

Prenons l'exemple de la reconnaissance de chiffres manuscrits et voyons comment le réseau de neurones utilise ses différentes couches pour identifier le bon chiffre. Sur la figure 1.3, le réseau transforme l'image du chiffre manuscrit en de nouvelles images dans les couches 1, 2 et 3. Ces couches lui fournissent de nouvelles informations pour la reconnaissance du chiffre. Le modèle d'apprentissage profond peut être considéré comme un opérateur de distillation d'informations en plusieurs étapes où l'information passe par plusieurs filtres et en ressort « débruitée » vis-à-vis d'une certaine tâche, ici l'identification du chiffre.

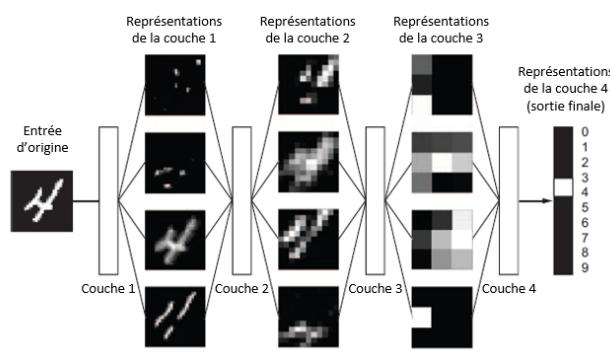


FIGURE 1.3 – Représentations profondes apprises par un modèle d'IA de classification de chiffres, source : [Chollet \[2021\]](#)

1.1.2 Réseau neuronal dense (DNN)

Les différentes couches qui créent la profondeur du réseau transforment la donnée initiale et « stockent » ces opérations de transformations dans des paramètres de la couche appelés poids. Ainsi,

l'apprentissage du réseau passe par la détermination des valeurs de ces poids pour que la sortie du modèle d'IA corresponde bien aux résultats attendus. Trouver la valeur correcte de ces poids est le point central de l'apprentissage, d'autant plus qu'il existe des millions de poids et que la modification de l'un d'entre eux sur une couche modifie la valeur de tous les autres. Pour cela, il est nécessaire d'évaluer l'erreur entre les poids définis et ceux qui permettraient d'avoir la sortie attendue. Cette erreur est quantifiée grâce à une fonction coût ("loss" en anglais). Le score obtenu par cette fonction coût permet au réseau d'ajuster ses poids au cours de son apprentissage ou entraînement. La principale difficulté consiste alors à modifier les poids du réseau afin de diminuer le score de la fonction coût. C'est le but de l'optimiseur qui implémente ce qui est appelé l'algorithme de propagation rétrograde ou de « backpropagation ». Initialement, le réseau attribue aux poids des valeurs aléatoires. Ces valeurs seront ensuite corrigées en passant le jeu de données d'entrée au travers de boucles d'apprentissage, appelées « epochs » qui, répétées un certain nombre de fois, permettent de minimiser la fonction coût. On dira alors que le réseau est entraîné. L'optimiseur, le nombre d'epochs et la fonction coût font partie des hyperparamètres du réseau.

Chaque opération d'un réseau étant différentiable, il est possible de calculer le gradient de la fonction coût d'après les poids du réseau afin de la minimiser grâce à l'algorithme de descente de gradient. On considérera alors que le réseau est entraîné lorsque l'algorithme converge, c'est-à-dire que la fonction coût a atteint un minimum global, le piège résidant dans l'atteinte d'un minimum local. Un taux d'apprentissage ou « learning rate » (LR) est un hyperparamètre utilisé lors de la descente de gradient. A chaque itération, l'algorithme multiplie le gradient par le taux d'apprentissage. On joue ici sur la rapidité de la descente de gradient mais un LR trop grand « sauterait » le minimum global alors qu'un LR trop petit pourrait ralentir la convergence ou atteindre un minimum local.

Le déroulé d'une boucle d'apprentissage est le suivant (figure 1.4) :

1. Sélection d'un paquet (batch) de données d'entrées X et de sorties attendues Y
2. Transformations des données X par le réseau (dans une étape appelée le « forward pass ») pour obtenir des prédictions de Y notées Y'
3. Calcul de l'erreur du réseau entre Y et Y' sur ce paquet de données avec la fonction coût
4. Mise à jour des poids W du réseau grâce à l'optimiseur afin de réduire le score de la fonction coût sur ce paquet de données

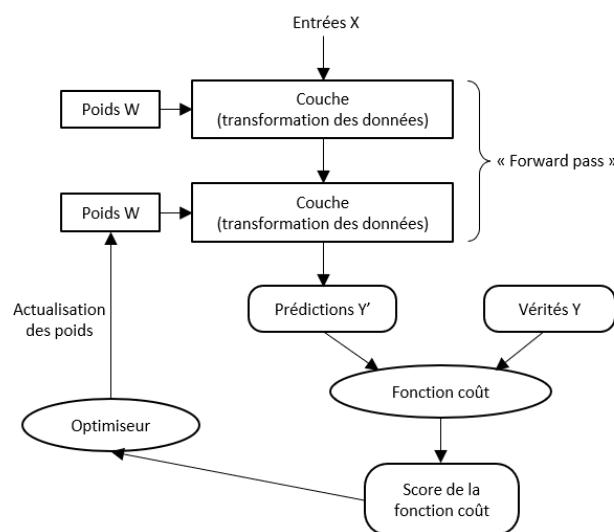


FIGURE 1.4 – Interactions entre les paramètres et hyperparamètres d'un réseau de neurones au cours de l'apprentissage, source : [Chollet \[2021\]](#)

Une des transformations principales des données réalisée par le réseau consiste en l'application d'une fonction d'activation non-linéaire. On aura alors :

$$Y' = \text{fonction d'activation}(W \cdot X + b) \quad (1.1)$$

avec b un biais du réseau, \cdot une multiplication matricielle, W , X et b des tenseurs (respectivement les poids, les données d'entrée et le biais).

Les données fournies en entrée du réseau sont sous la forme de tenseurs. Un tenseur d'ordre 0 est un scalaire, d'ordre 1 un vecteur, d'ordre 2 une forme bilinéaire telle que les matrices. Ce format permet aux données d'avoir plusieurs dimensions qui seront toutes prises en compte et interprétées par le modèle d'IA.

La fonction d'activation permet au réseau d'apprendre des transformations non linéaires des données. Avec des opérateurs linéaires, la plus-value d'une succession de couches serait perdue étant donné que la concaténation d'opérations linéaires est toujours linéaire. Les fonctions Tanh (pour Tangente hyperbolique, équation 1.2) et ReLU (pour Rectified Linear Unit, équation 1.3) sont les plus utilisées, mais il existe d'autres fonctions telles que SiLU (pour Sigmoid Linear Unit, équation 1.4) qui permettent d'introduire de la non-linéarité dans le réseau, et constituent l'avantage principal des réseaux de neurones.

$$\text{Tanh}(x) = \frac{\exp(x) + \exp(-x)}{\exp(x) - \exp(-x)}, \quad x \in \mathbb{R} \quad (1.2)$$

$$\text{ReLU}(x) = (x)^+ = \max(0, x), \quad x \in \mathbb{R} \quad (1.3)$$

$$\text{SiLU}(x) = x\sigma(x) \quad \text{avec} \quad \sigma(x) = \frac{1}{1 + \exp(-x)} \quad \text{la fonction sigmoïde}, \quad x \in \mathbb{R} \quad (1.4)$$

Une fois le réseau entraîné, des jeux de données d'évaluation et de test, qui n'ont pas été vus par le réseau en entraînement, sont utilisés respectivement pour régler les hyperparamètres du réseau (optimiseur, fonction coût, batch size, epochs, ...) et pour évaluer la performance finale du réseau.

1.2 Traitement d'images

Dans le chapitre précédent nous avons découvert le concept d'intelligence artificielle et le mécanisme global d'un réseau de neurones en apprentissage profond. Le traitement des images avec de l'IA fait recours à quelques notions supplémentaires.

1.2.1 Réseau neuronal convolutif (CNN)

On parlera de réseau neuronal convolutif pour les images, que l'on distinguera du réseau neuronal dense précédemment présenté (section 1.1.2), lorsque le produit matriciel de la fonction d'activation (équation 1.1) est remplacé par une opération de convolution. La convolution permet au réseau d'apprendre des motifs locaux dans l'image. Dans l'exemple de la reconnaissance de chiffres manuscrits, les convolutions jouent le rôle des filtres qui créent différentes représentations de l'image au sein d'une couche du réseau (figure 1.3).

Cette caractéristique clé confère aux CNN deux propriétés intéressantes :

- Les motifs qu'ils apprennent sont invariants par translation. Après avoir appris un certain motif dans le coin inférieur droit d'une image, un convnet peut le reconnaître n'importe où. Un réseau dense devrait réapprendre le motif s'il apparaît à un nouvel endroit. Les CNN n'ont besoin que d'une seule image et donc de moins d'échantillons de données pour apprendre des représentations qui peuvent par la suite être généralisées.

- Ils peuvent apprendre des hiérarchies spatiales de motifs (voir figure 1.5). Une première couche de convolution apprend de petits motifs locaux tels que les arêtes, une deuxième couche de convolution apprendra des motifs plus grands composés des caractéristiques des premières couches, et ainsi de suite. Cette méthode permet aux réseaux de convolution d'apprendre efficacement des concepts visuels de plus en plus complexes et abstraits. Ces différentes représentations des données d'entrées constitueront un ensemble appelé espace latent.

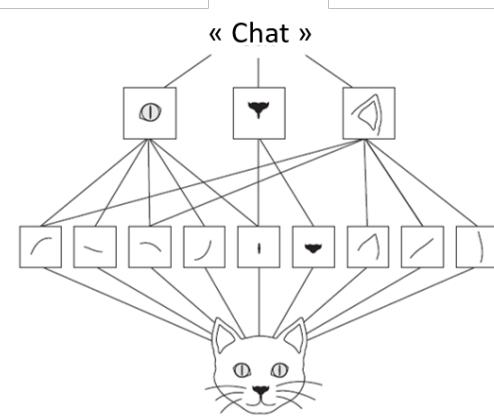


FIGURE 1.5 – Formation d'une hiérarchie spatiale de modules visuels, des traits locaux se combinent à des objets tels que les yeux ou les oreilles pour former un concept : "le chat", source : [Chollet \[2021\]](#)

Une convolution procède de la façon suivante : une tuile « glisse » sur une image en se décalant selon un pas défini et en sommant le produit des valeurs entre la tuile et l'image, selon un produit de Hadamard (figure 1.6). Les convolutions sont définies par trois paramètres clés : la taille des tuiles extraites, le nombre de filtres calculés par la convolution et le pas de la tuile.

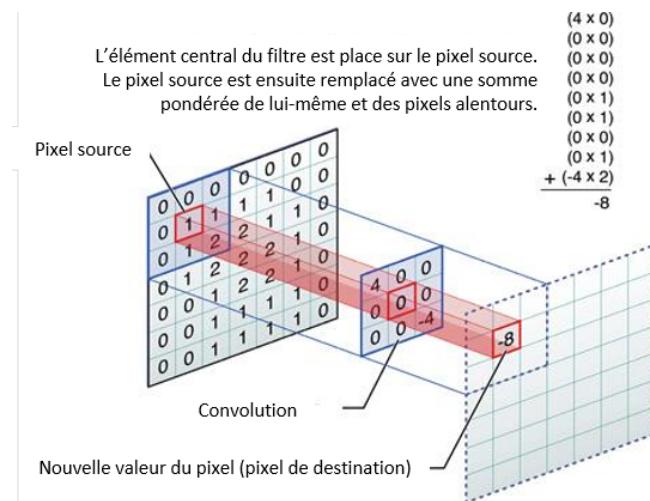


FIGURE 1.6 – Opération de convolution, source : Medium 6 basic things to know about Convolution ³

Il faut noter que la dimension de l'image après le produit de convolution peut différer de sa taille initiale. En effet, le pas de la tuile modifiera les dimensions de l'image (un pas de deux divisera la dimension de l'image par deux) et des effets de bord peuvent apparaître. En effet, sur la figure 1.6 on constate que le filtre est appliqué sur le bord supérieur gauche de l'image. Or le résultat, -8, n'est positionné que sur la deuxième colonne de la deuxième ligne de l'image finale. Selon la dimension du filtre, des bords plus ou moins épais seront coupés sur l'image finale. Afin de compenser cet effet de bord, un padding est appliqué et permet d'attribuer des valeurs à ces bords rognés afin de garder les dimensions d'image souhaitées.

Ces couches de convolutions vont progressivement augmenter le nombre de paramètres du réseau

de neurones. Toutefois, pour éviter que ce nombre ne devienne trop conséquent, une étape dite de « max pooling » suit chaque couche de convolution. Le « max pooling » consiste à faire glisser une fenêtre généralement de taille 2x2 sur l'image, afin de récupérer la valeur maximale des pixels de la fenêtre. Le principe est similaire à la convolution, excepté qu'il n'y a pas de transformation de l'image mais uniquement récupération de la valeur maximale. Outre la réduction du nombre de paramètres, cette étape permet également d'éviter le sur-apprentissage. Le réseau prendra en compte les parties de l'image les plus importantes et ne se dispersera pas dans un grand nombre de petites images. Une taille de 2x2 pixels pour la fenêtre permet d'éviter une trop grande perte d'information.

Ainsi, l'enchaînement des différentes couches dans un réseau neuronal convolutif consiste en une répétition de l'enchaînement : convolution, fonction d'activation (équation 1.1) et max pooling ; mais des structures plus complexes ont été développées afin d'optimiser l'apprentissage.

1.2.2 U-Net

En 2015, Olaf Ronneberger, Philipp Fischer et Thomas Brox développent une nouvelle architecture améliorant le réseau neuronal convolutif pour la segmentation d'images appelé « U-Net » ([Ronneberger and Brox \[2015\]](#)). Dans le domaine du traitement des images numériques et de la vision par ordinateur, la segmentation des images est le processus qui consiste à diviser une image numérique en plusieurs segments d'image (ensembles de pixels). L'objectif de la segmentation est de simplifier et/ou de modifier la représentation d'une image en quelque chose de plus significatif et de plus facile à analyser. La segmentation d'images est généralement utilisée pour localiser des objets et des limites (lignes, courbes, etc.) dans les images.⁴

Le U-Net est un réseau entièrement convolutif dont la spécificité repose sur son architecture en U avec un chemin contractant (ou de descente) et un chemin symétrique d'expansion (ou de montée). Ces chemins sont constitués d'un enchaînement de blocs de sous-échantillonnage, de sur-échantillonnage et de connexion résiduelle.

- Le bloc de **sous-échantillonnage** comprend deux étapes de convolution et une étape de max pooling. On réduit ainsi la taille de l'image tout en augmentant la profondeur des cartes de caractéristiques de l'image, grâce aux filtres de convolution. Cela permet de capturer des informations de plus en plus abstraites et globales de l'image. Ces blocs représentent la partie descendante du U.
- Le bloc de **sur-échantillonnage** inverse ce qui a été effectué par les premiers blocs en augmentant la taille de l'image et en reconstruisant l'image dans sa dimension initiale avec des caractéristiques affinées. C'est la partie montante du U qui comprend des déconvolutions ou « up-convolutions ».
- Le bloc de **connexion résiduelle** permet de conserver les informations importantes tout au long du processus d'apprentissage en combinant les résultats de la branche descendante avec ceux de la branche ascendante. Il permet la fusion des caractéristiques du chemin contractant avec des informations plus détaillées du chemin d'expansion.
- La partie la plus basse du U-Net est un point de transition entre la descente et la montée et ne se compose que de convolutions. Cette étape contient **l'espace latent** le plus vaste.

Cette architecture U-Net a eu un impact majeur, avec plusieurs dizaines de milliers de citations de l'article original. Elle est aujourd'hui une des références en matière de modèles d'IA « image-to-image » que sont les réseaux convertissant une image en une autre de mêmes dimensions.

1.2.3 Débruitage d'images : le Noise2Noise

La segmentation d'image est une étape préliminaire essentielle dans le traitement d'images, notamment pour les méthodes de débruitage d'images corrompues.

4. https://en.wikipedia.org/wiki/Image_segmentation

[Lehtinen et al. \[2018\]](#) a développé en 2018 une méthode d'apprentissage profond de débruitage d'images appelée « Noise2Noise ». Cette méthode reconstruit une image propre en s'entraînant sur un jeu d'images corrompues par du bruit. Les performances de cet algorithme se sont avérées être égales voire supérieures à celles d'un algorithme s'entraînant sur un jeu d'images propres. [Bône et al. \[2024\]](#) a repris cette méthode afin de l'appliquer à des variables climatiques pour distinguer la réponse forcée de la variabilité interne. Comme mentionné dans l'introduction, l'identification de la réponse forcée, qui relève d'un problème de distinction d'un signal par rapport à un bruit, soulève de nombreux défis. Les méthodes d'IA présentent un avantage majeur lié à l'utilisation de méthodes non-linéaires qui permettent de prendre en compte les covariances spatio-temporelles des variables climatiques.

La variabilité interne est donc considérée comme le bruit à retirer de l'image du changement climatique qu'est la réponse forcée. L'objectif ultime du développement de cette méthode consiste à appliquer le réseau entraîné sur des modèles de climat aux données d'observations. En effet, l'hypothèse faite ici consiste à considérer que la réalité est incluse dans la distribution créée par l'ensemble des modèles de climat. Ainsi, le réseau s'entraînera en « apprenant » la variabilité interne et pourra la « reconnaître » et la « supprimer » dans les observations. Une fois le réseau paramétré et entraîné sur le monde idéalisé des modèles, il pourra être appliqué aux observations où certaines variables climatiques, telles que les précipitations, possèdent un signal difficilement distinguable de la variabilité interne avec des méthodes classiques.

Dans cette méthode de débruitage des variables climatiques, les données fournies en entrée sont des paires de membres d'un ensemble de modèle de climat (cf.). En effet, dans le cas des grands ensembles SMILEs, tous les membres d'un modèle climatique partagent la même réponse forcée (le signal), et se distinguent par leur variabilité interne (le bruit). Le réseau s'entraîne alors en identifiant le bruit qui diffère selon les membres. Il extrait ensuite ce bruit de l'image qui est commune à tous les membres de l'ensemble, et qui est le signal que l'on veut récupérer. Cette méthode est dite non supervisée car elle ne compare pas la sortie Y' du réseau à la vérité Y (notations de la figure 1.4), qui est ici la moyenne d'ensemble du modèle de climat, mais la compare à un autre membre qui comporte lui aussi du bruit et sera noté \tilde{Y} . La fonction coût calcule donc l'erreur entre Y' , un membre qui « passe » dans le modèle d'IA, et \tilde{Y} un autre membre de l'ensemble. Cependant, afin d'évaluer les performances du réseau, le calcul du score de la fonction coût sera effectué pour les jeux de données d'évaluation et de test entre la sortie du réseau et la moyenne d'ensemble (la vérité Y).

Maintenant que les différentes notions portant sur l'intelligence artificielle ainsi que la méthode de débruitage d'images utilisée ont été présentées, nous allons parcourir les différentes étapes de développement du réseau.

2 | Développement du réseau de neurones

Le modèle de réseau de neurones développé reprend la méthode de Noise2Noise présentée à la section 2 et est appliqué aux données climatiques en suivant la démarche de [Bône et al. \[2024\]](#). Les données utilisées ainsi que l'architecture du réseau sont présentées dans cette partie. Il nous faut également introduire les différentes métriques nous permettant d'évaluer et d'améliorer notre réseau. Enfin une analyse plus fine de la fonction coût et des poids du réseau sera effectuée.

2.1 Données

Le jeu de données utilisé provient du projet ForceSMIP (Forced Component Estimation Statistical Method Intercomparison Project, [Wills](#)), un hackathon mettant à disposition plusieurs variables de modèles climatiques issus de CMIP 6 (Climate Model Intercomparison Project 6) afin de développer la meilleure méthode d'estimation de la réponse forcée. Les données fournies par ce projet sont mensuelles et recouvrent la période 1880-2022 pour plusieurs variables climatiques, notamment la température de surface, la pression de surface et les précipitations avec une résolution d'environ 2° (72 latitudes et 144 longitudes). Les modèles disponibles sont CanESM5 (25 membres), CESM2 (50 membres), MIROC-ES2L (30 membres), MIROC6 (50 membres) et MPI-ESM1-2-LR (30 membres).

La normalisation des données est une étape importante dans le développement d'un modèle d'IA. Les réseaux de neurones sont très sensibles à cette étape qui peut falsifier les représentations des données faites par le réseau. Chaque normalisation des membres est effectuée ici modèle par modèle : on ne normalise pas en prenant en compte l'ensemble des membres de tous les modèles confondus. On évite ainsi qu'un modèle ayant un biais important ait plus de poids dans l'apprentissage. Par exemple, CanESM5 possède un biais chaud et pourrait, dans le cas des températures, modifier les résultats du réseau en ayant des valeurs plus importantes que celles des autres modèles de climat.

Également, il est préférable de travailler avec des données dont les valeurs sont faibles (généralement comprises entre $[-1,1]$) afin de garantir la stabilité numérique du modèle mais aussi d'avoir un apprentissage plus rapide. Ainsi, pour chaque modèle de climat et pour chaque membre, on calcule les anomalies des variables par rapport à leur moyenne sur 30 ans (1900-1930). Les données sont finalement normalisées entre $[0,1]$ afin de pouvoir utiliser la fonction d'activation SiLU (équations 1.3 et 1.4) sans ramener toutes les valeurs d'anomalies négatives à 0.

2.2 Architecture du réseau

L'architecture du réseau est un U-Net (partie 1.2.2) tel que défini sur la figure 2.1.

Ses hyperparamètres ou caractéristiques sont les suivants :

- **fonctions d'activation** : SiLU pour les couches intermédiaires et Tanh pour les couches d'entrée et de sortie du réseau.
- **batch size** (taille des paquets de données en entrée du réseau) : 8
- **nombre d'epochs** (boucles d'entraînement) : 100
- **fonction coût** : erreur quadratique moyenne (MSE)

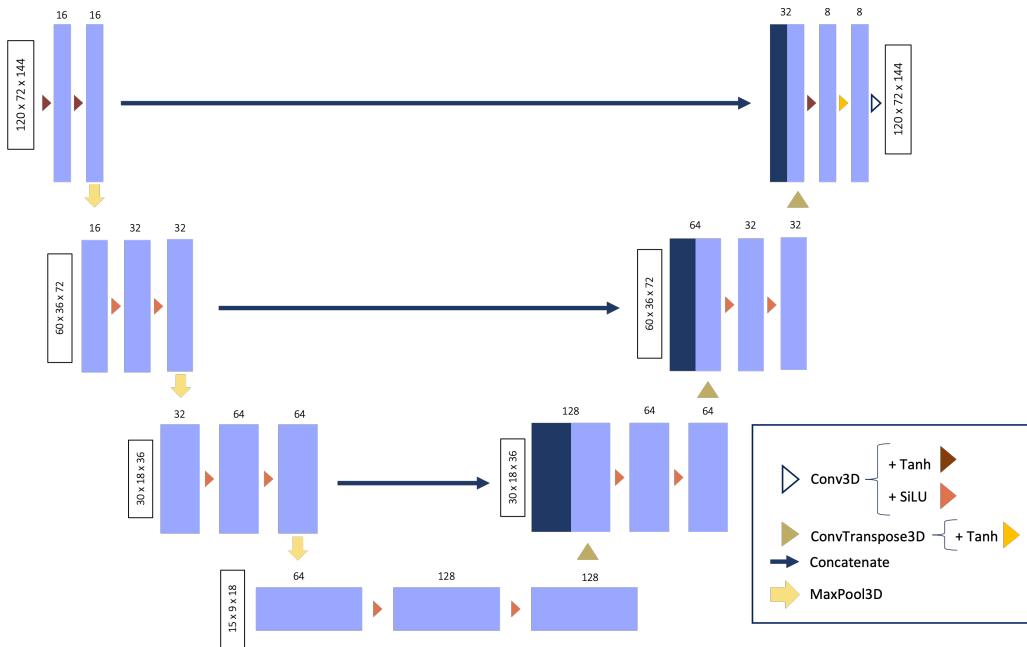


FIGURE 2.1 – Architecture du U-Net développé pour le débruitage d'images

- **optimiseur** : Adam optimizer
- **padding** : bordures supérieures et inférieures : réPLICATION des valeurs des bords ; bordures latérales : valeurs du bord est des cartes complétant les valeurs du bord ouest et inversement (respectivement replication et circular)
- **taux d'apprentissage** : 10^{-5}

2.3 Métriques d'évaluation du réseau

Une fois l'étape de préparation du jeu de données et de construction d'une première architecture de réseau effectuées, une grande partie du développement d'un modèle d'IA est empirique et consiste à effectuer plusieurs tests du réseau afin d'ajuster les hyperparamètres selon les résultats obtenus. Pour cela on utilise plusieurs métriques :

- **le score de la fonction coût** qui permet de suivre l'erreur finale du réseau et informe sur la présence d'une potentielle anomalie lors de l'apprentissage.
- **le pattern correlation** non-centré calcule le coefficient de corrélation linéaire de Pearson entre la tendance (20 dernières années - 20 premières années) de la réponse forcée du modèle de climat et de l'image débruitée par le U-Net. Des valeurs proches de 1 indiquent que les motifs spatiaux de la réponse forcée sont bien reconstruits par le réseau.
- **l'évolution temporelle** de la variable en moyenne globale sur la période d'étude, qui vérifie si le réseau reconstruit le signal de façon cohérente sur la dimension temporelle
- **l'erreur quadratique cumulée** sur une période donnée (RMSE). On considère une variable physique issue d'un modèle de climat z notée $z(t,x)$ avec $x=1,\dots,p$ points de grille aux temps $t=1,\dots,n$. On note $\bar{z}(t,x)$ la valeur de z moyennée sur l'ensemble du modèle de climat étudié (moyenne multi-membres = vérité), on a alors :

$$RMSE(z(x)) = \sqrt{\frac{1}{n} \sum_{t=1}^n (z(t,x) - \bar{z}(t,x))^2} \quad (2.1)$$

Cette métrique quantifie l'erreur sur la période d'étude et montre la distribution spatiale des erreurs. Pour avoir une information plus synthétique, la moyenne globale de la RMSE est aussi

calculée.

Les résultats obtenus avec ces différentes métriques ont mené au développement de plusieurs modèles d'IA. Ces résultats sont tous présentés pour le premier U-Net et l'on se focalisera sur le score de la fonction coût et la RMSE en moyenne globale pour les U-Nets suivants.

2.4 Une problématique : cinq U-Nets

Premier U-Net : 1 variable, 1 modèle climatique, données annuelles Un premier U-Net a été développé sur les données annuelles de température du modèle CESM2 afin de tester la méthode employée de débruitage d'images appliquée à notre problématique.

Quelle est la composition du jeu de données de ce premier réseau ? Sur les 50 membres de l'ensemble, 11 sont aléatoirement réservés à l'étape de validation du réseau et 3 membres sont mis de côté pour le test. Ces 3 membres possèdent une tendance de changement climatique la plus faible, la plus forte et médiane par rapport à son ensemble. 1296 paires (36×36) sont donc fournies en entrée du réseau. Le réseau est lancé plusieurs fois afin d'ajuster les hyperparamètres et d'obtenir le meilleur apprentissage possible.

- Score de la fonction coût lors du test : 7.5×10^{-5}
- Pattern correlation moyen sur les trois membres de test : 0.9
- Evolution temporelle (figure 2.2) : on constate qu'en moyenne globale le U-Net lisse quelques variations, notamment celles liées aux éruptions volcaniques mais suit bien la tendance sur la période d'étude. De même en Sibérie, mais le signal est sous-estimé sur les 40 dernières années.

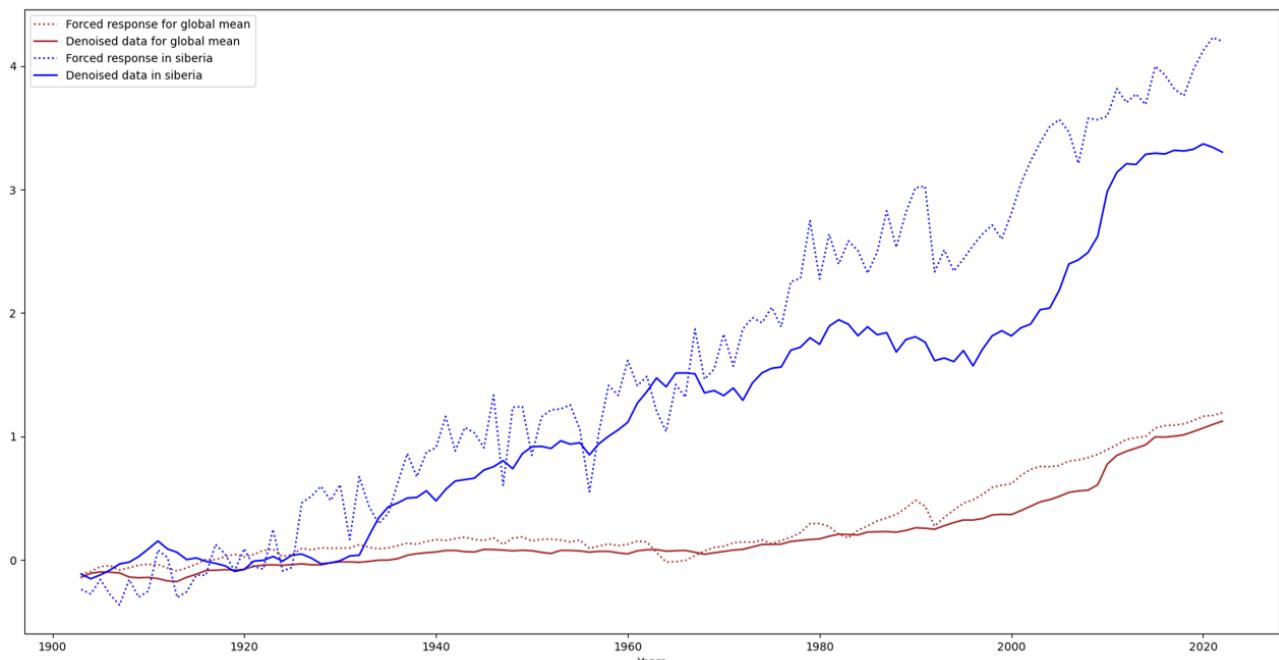


FIGURE 2.2 – Évolution temporelle de la réponse forcée des températures de surface en anomalies par rapport à la période 1900-1930 du membre 26 (membre médian) pour le modèle CESM2 débruité issu du premier réseau (traits pleins) et moyenne d'ensemble (pointillés) en moyenne globale (marron) et pour un point en Sibérie (bleu)

- RMSE moyennée sur les trois membres de test en moyenne globale : 0.19°C
- Pour ce premier U-Net nous décomposons la RMSE (équation 2.1) en plusieurs étapes (figure 2.3) : la réponse forcée (a) présente un gradient de températures nord/sud qui se retrouve bien

sur l'image débruitée (b). La différence entre les 2 cartes (c) nous montre que pour l'année 2022 les zones avec les erreurs les plus élevées se situent au niveau des pôles. L'erreur au carré cumulée sur toute la période (d) confirme cette dernière observation avec des erreurs de 2.6 °C au nord du détroit de Bering.

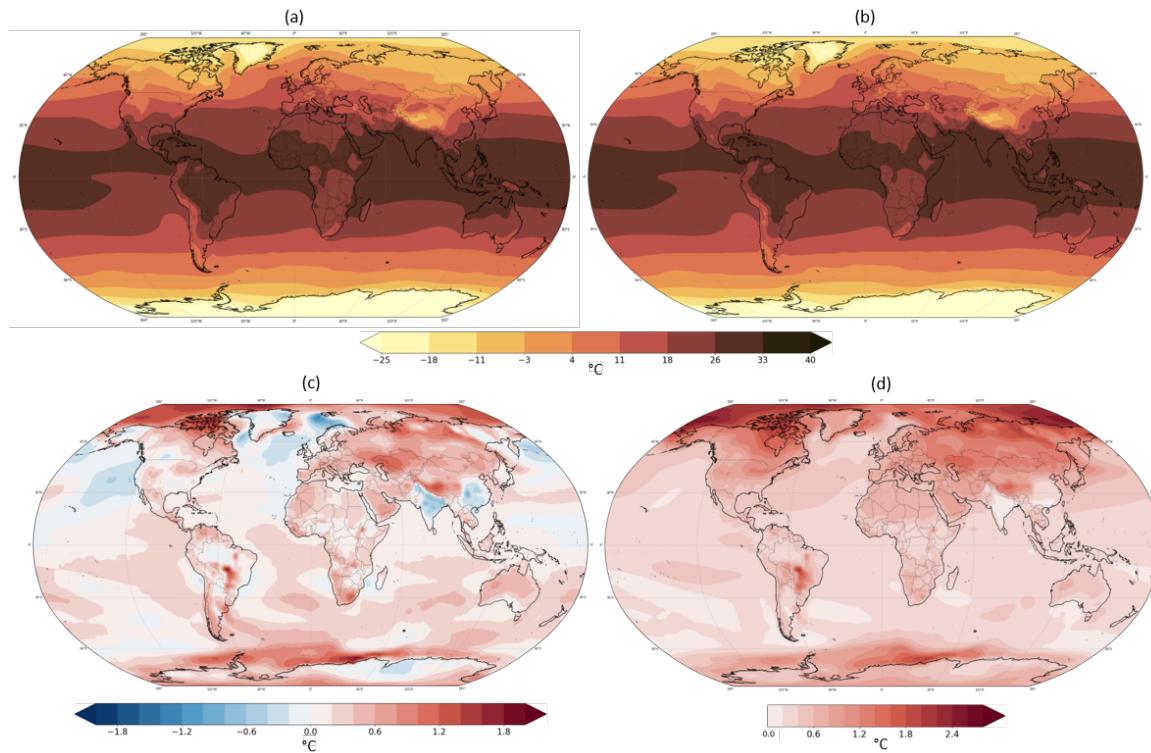


FIGURE 2.3 – Décomposition du calcul de RMSE (équation 2.1) pour les températures du membre 26 du modèle CESM2 : (a) réponse forcée du modèle pour l'année 2022 (\bar{z}), (b) image obtenue en débruitant le membre 26 du réseau pour l'année 2022, (c) différence entre les cartes (a) et (b), (d) RMSE du membre 26 sur la période 1903-2022.

Une fois les hyperparamètres fixés sur le premier U-Net, on étend son entraînement à plusieurs modèles climatiques dont la diversité permet d'avoir une meilleure estimation de la variabilité interne

Deuxième U-Net : 1 variable, 5 modèles climatiques, données annuelles Les paires de membres fournies en entrée sont constituées au sein d'un même modèle pour éviter de donner plus de poids aux modèles biaisés. Quatre modèles sont utilisés pour l'entraînement et un modèle est gardé de façon aléatoire pour l'étape de test en suivant la méthode de K-Fold afin d'évaluer par la même occasion la robustesse de la méthode selon le jeu de données d'apprentissage.

L'évaluation croisée avec la méthode de K-Fold (figure 2.4) consiste à diviser son jeu de données en K partitions en entraînant son réseau sur K-1 partitions et en le validant sur la partition restante. Le score moyen sur chaque fragmentation du jeu de données permet d'évaluer la dépendance du réseau à son jeu d'entraînement et d'estimer si certaines données sont sur-représentées ou sous-représentées dans le jeu d'entraînement.

Le U-Net est alors entraîné sur un jeu de données comportant de 5000 à plus de 7000 paires de membres en entrée et est testé sur 600 à 2500 paires de membres selon le modèle tiré pour le test. L'étape d'évaluation du réseau est dès lors supprimée afin de garder le plus de membres possibles pour l'entraînement du U-Net.

Les scores obtenus pour chaque réseau de test sur les 5 partitions (tableau 2.1 et figure 2.4) sont du même ordre, on considère donc qu'il n'y a pas d'anomalies au sein du jeu de données. Ces valeurs sont plus élevées que celles obtenues pour le U-Net 1 ce qui est cohérent avec la plus grande diversité

du jeu de données. Les valeurs de pattern correlation et de RMSE en moyenne globale sont présentées, respectivement, sur le tableau 2.1 et sur la figure 2.5.

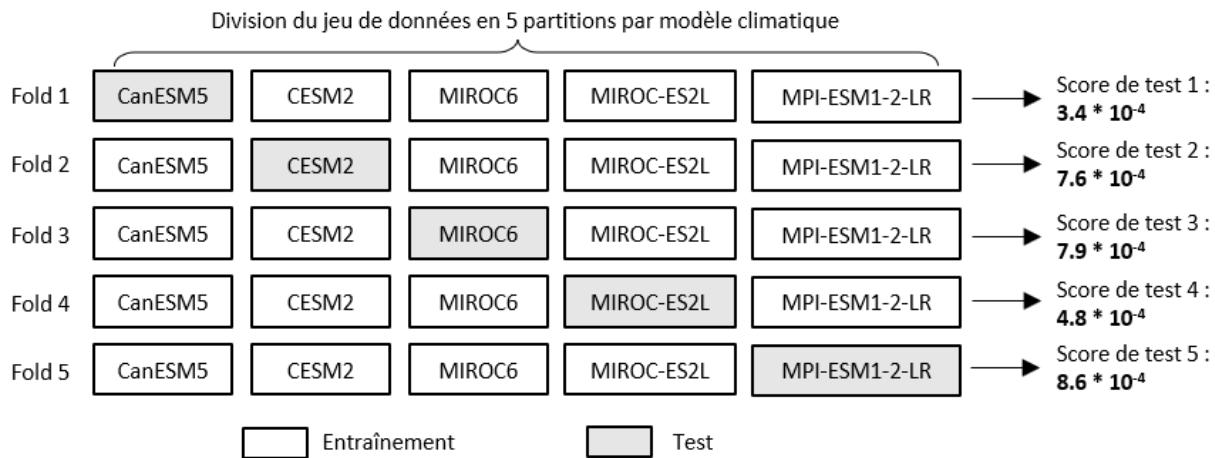


FIGURE 2.4 – Validation croisée et scores du deuxième U-Net sur le modèle de test pour chaque partition

Troisième U-Net : 1 variable, 5 modèles climatiques, données mensuelles Le réseau multi-modèles climatique (U-Net 2) est adapté en remplaçant les données annuelles par des données mensuelles afin d'améliorer les scores. Le réseau apprend ainsi le cycle saisonnier, lissé par la moyenne annuelle. Néanmoins, la multiplication par un facteur douze de la taille des entrées du réseau requiert une mémoire ainsi qu'un temps de calcul plus importants. Afin de ne pas mobiliser l'ensemble des capacités de calculs d'IA du CERFACS, quelques méthodes d'optimisation ont été développées pour pouvoir poursuivre l'amélioration de notre U-Net.

Optimisation

Les modèles d'apprentissage profond tels que ceux utilisés pour la segmentation d'images requièrent une puissance de calcul qui dépasse amplement celle fournie par un ordinateur portable. Depuis le début des années 2000, des sociétés telles que NVIDIA et AMD ont très rapidement développé des puces de parallélisation massive de calculs, les GPU (unités de traitement graphique).

Dans un premier temps, ces puces avaient pour but de répondre au besoin croissant de puissance dans les jeux vidéo afin d'obtenir des scènes, calculées en temps réel, de plus en plus réalistes. En 2007, ces investissements massifs ont bénéficié à la communauté scientifique grâce au développement de CUDA ^{a.}, une interface de programmation pour GPUs. Quelques GPUs ont alors pu remplacer d'importants clusters de CPU (unité centrale de calcul) avec de plus grandes capacités de parallélisation. Les réseaux de neurones profonds, qui techniquement peuvent se résumer à des multiplications matricielles - opérations hautement parallélisables - ont commencé à être implémentés grâce à CUDA en 2011. Les réseaux récemment développés, tel que ChatGPT, n'auraient pas pu voir le jour sans l'existence des GPUs.

La machine utilisée pour entraîner et tester les réseaux développés dans le cadre de ce stage est le processeur NVIDIA GH200 Grace Hopper muni du GPU NVIDIA H100, spécialement conçu pour l'IA. L'architecture GPU et le processeur disponibles au CERFACS ont rendu possible une première méthode d'optimisation du code : la précision mixte. La bibliothèque Python d'apprentissage automatique PyTorch, utilisée dans ce stage, permet d'adapter la précision des opérations GPU afin d'améliorer les performances du réseau sans perdre de leur justesse. En effet, certaines opérations, telles que les convolutions, sont plus rapides avec une précision en float16 alors que les opérations de réduction se font en float32. Cette méthode de précision mixte est couramment associée au « gradient scaling », ou calibration du gradient, pour éviter que des valeurs de gradient de faible amplitude ne soient pas représentables en float16 (perte de précision lors du passage de float32 à float16). Ainsi, les valeurs ne sont pas ramenées à 0 mais multipliées par un facteur d'échelle ou « scaling factor » qui leur donnent plus d'amplitude. Avec cette méthode, les temps de calcul du réseau ont pu être divisés par 2,6.

Des modules développés par Nvidia, tel que Nvidia Nsight Profiler, permettent également de suivre la répartition des opérations sur les CPU et GPU afin d'éviter une sous-exploitation des processeurs, ce qui ralentit et dégrade les performances du réseau, et d'identifier éventuellement une mauvaise implémentation du code. Aucune anomalie n'a été détectée sur ces profils.

a. <https://developer.nvidia.com/about-cuda>

Revenons à notre troisième U-Net. Les valeurs de pattern correlation tableau 2.1 sont, en moyenne, proches voire très proches de 1. Les motifs globaux de la réponse forcée sont bien reconstruits. Toutefois, le réseau présente des valeurs de RMSE (figure 2.5) plus élevées que le deuxième réseau implémenté même si les scores de la fonction coût sont plus faibles (tableau 2.1). Plusieurs hypothèses sont faites :

- le réseau a besoin d'un nombre plus important de boucles d'apprentissage car les dimensions des entrées sont plus conséquentes
- le taux d'apprentissage défini n'est plus adapté et le réseau atteint un minimum local

Plusieurs modifications des hyperparamètres sont effectuées mais fournissent des scores équivalents. Une nouvelle normalisation est alors envisagée : c'est le quatrième U-Net.

Quatrième U-Net : 1 variable, 5 modèles climatiques, données mensuelles, normalisation par rapport au cycle saisonnier Les données mensuelles sont normalisées différemment. On ne retire plus aux données leur moyenne sur 30 ans mais leur cycle saisonnier. On ramène ensuite les valeurs entre [0,1].

On constate que les valeurs de RMSE de ce réseau sont plus élevées que ceux du U-Net 2 (figure 2.5) alors que les scores de la fonction coût sont plus faibles (tableau 2.1) mais équivalents à ceux du U-Net 3. Les valeurs de pattern correlation sont proches de 1, comme pour le U-Net précédent.

Néanmoins, ce réseau est meilleur que le précédent avec de bien meilleures valeurs de RMSE pour les modèles CESM2, MIROC6 et MIROC-ES2L. Le temps de calcul de ce réseau étant très élevé (9 jours) les tests sont poursuivis sur les U-Net travaillant avec des données annuelles normalisées par rapport à la moyenne sur 30 ans.

Cinquième U-Net : 2 variables, 5 modèles climatiques, données annuelles Le U-Net 2 est amélioré en ajoutant une variable, la pression de surface, aux températures. En effet, la représentation des données sous forme de tenseurs (section 1.1.2) permet d'ajouter une nouvelle dimension (la pression). Cet ajout permet au réseau d'apprendre les corrélations physiques entre les deux variables.

On constate que les scores de la fonction coût des modèles CanESM5, MIROC-ES2L et MPI-ESM1-2-LR sont très élevés (tableau 2.1). Les valeurs de RMSE de ces modèles (figure 2.5) sont plus élevés que pour le U-Net 2 excepté pour MPI-ESM1-2-LR qui présente des performances équivalentes. Concernant CESM2 l'erreur quadratique est légèrement dégradée et les résultats sont meilleurs pour MIROC6. Les valeurs de pattern correlation sont équivalents à celles du U-Net 2, et moins bonnes que pour les réseaux travaillant avec des données mensuelles.

	CanESM5	CESM2	MIROC6	MIROC-ES2L	MPI-ESM1-2-LR
U-Net 1	-	0.8 <u>0.9</u>	-	-	-
U-Net 2	3.4 <u>0.5</u>	7.6 <u>0.8</u>	7.9 <u>0.7</u>	4.8 <u>0.7</u>	8.6 <u>0.7</u>
U-Net 3	2.5 <u>0.9</u>	2.6 <u>0.9</u>	2 <u>~1</u>	2.4 <u>0.9</u>	2.4 <u>0.9</u>
U-Net 4	2.6 <u>0.9</u>	2.1 <u>0.9</u>	2 <u>0.9</u>	3.1 <u>0.9</u>	2 <u>0.9</u>
U-Net 5	230 <u>0.6</u>	3.1 <u>0.7</u>	1.5 <u>0.6</u>	190 <u>0.7</u>	270 <u>0.6</u>

TABLE 2.1 – Scores de la fonction coût (en 10^{-4}) et du pattern correlation moyen sur l'ensemble des membres du modèle (souligné) des cinq U-Net développés pour chaque modèle en partition test

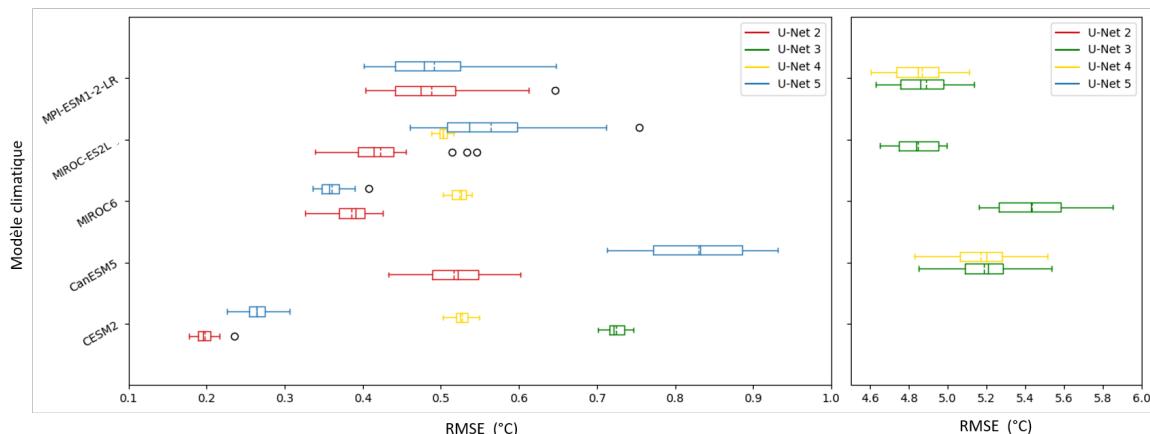


FIGURE 2.5 – Valeurs de RMSE en moyenne globale sur la période 1903-2022 pour les 25 premiers membres de chaque modèle climatique en partition test pour les U-Nets 2 (rouge), 3 (vert), 4 (jaune) et 5 (bleu)

Après analyse des différents scores obtenus pour chaque U-Net et prise en considération de la robustesse et des temps de calcul de chacun des réseaux, le U-Net 2 a été retenu pour la suite des analyses.

2.5 Que se passe-t-il du côté de la fonction coût ("loss") ?

Tout au long du développement des différents U-Nets, plusieurs tests ont été effectués afin d'obtenir une décroissance puis convergence de la fonction coût et donc un réseau bien entraîné. La figure 2.6 montre que les U-Nets 2 et 3 présentent une fonction coût qui décroît sans converger à la fin des 100 epochs. Les autres réseaux semblent converger très rapidement mais les courbes zoomées révèlent des courbes similaires aux autres U-Net.

Différentes modifications des hyperparamètres sont testées : taux d'apprentissage adaptatif, écrêtage de gradient ou "clipping gradient" afin d'éviter d'avoir des valeurs de gradients trop grandes menant à une divergence de l'algorithme de descente de gradient, augmentation du nombre d'epochs. Malheureusement, ces tests n'ont pas mené à une convergence des réseaux sur les dernières epochs.

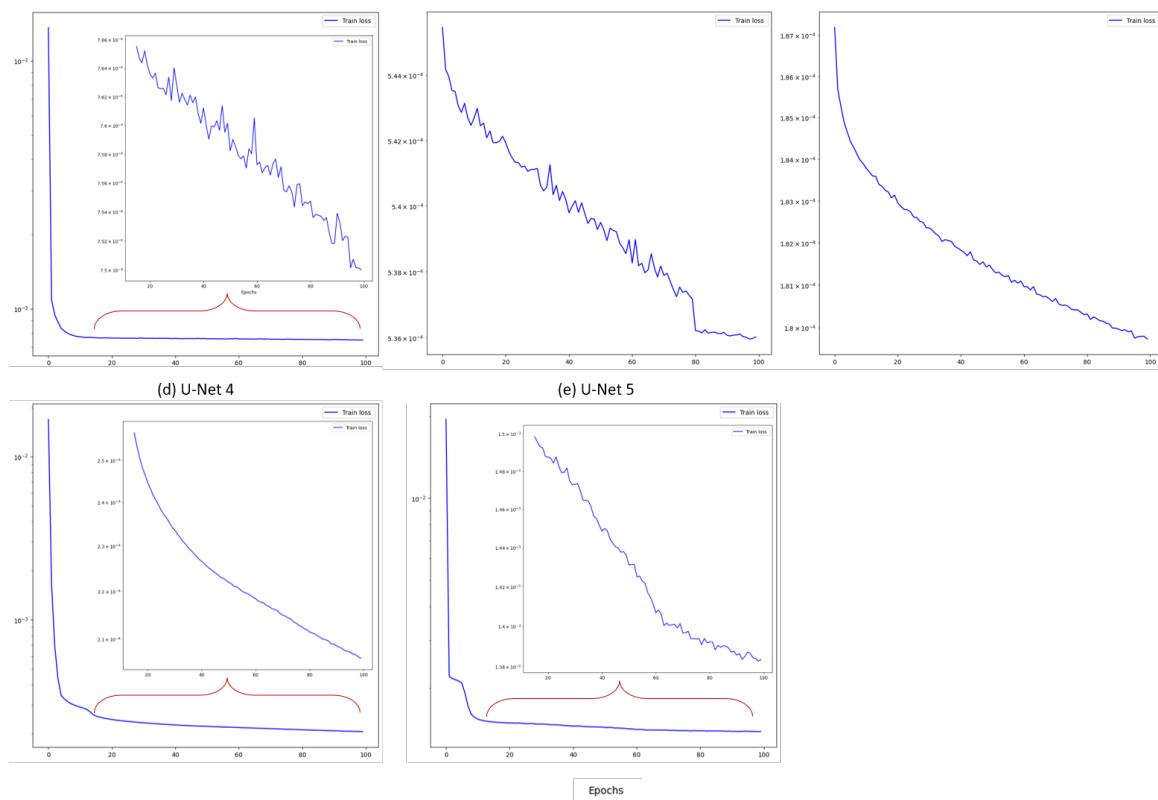


FIGURE 2.6 – Fonction coût pour la partition ayant pour modèle de test CESM2 pour les cinq U-Net développés avec un zoom sur les 85 dernières epochs pour les U-Net 1, 4 et 5.

L'hypothèse faite face à cette anomalie est la suivante : la méthode de Noise2Noise utilise en entrée une paire de membres d'un modèle de climat. Ces membres comportent tous les deux du bruit lié à la variabilité interne. Le réseau ne « voit » donc jamais la vérité vers laquelle il doit tendre mais uniquement des images entachées d'erreur. La particularité de cette méthode pourrait expliquer "l'infinie" décroissance de la fonction coût et sa convergence très rapide dès le début de l'apprentissage.

Afin d'obtenir une analyse plus fine de la convergence du réseau, les poids du réseau sont enregistrés à chaque couche du U-Net tout au long de l'apprentissage. La norme L2 des matrices de poids est calculée. La figure 2.7 montre que les poids des gradients diminuent puis semblent converger sur les trois premières epochs. Lorsqu'on affiche les poids sur toute la durée de l'entraînement (100 epochs), la courbe oscille très rapidement autour de 0.0005. Il semblerait donc que les poids du réseau se stabilisent très rapidement autour d'une valeur limite et donc qu'il y a convergence.

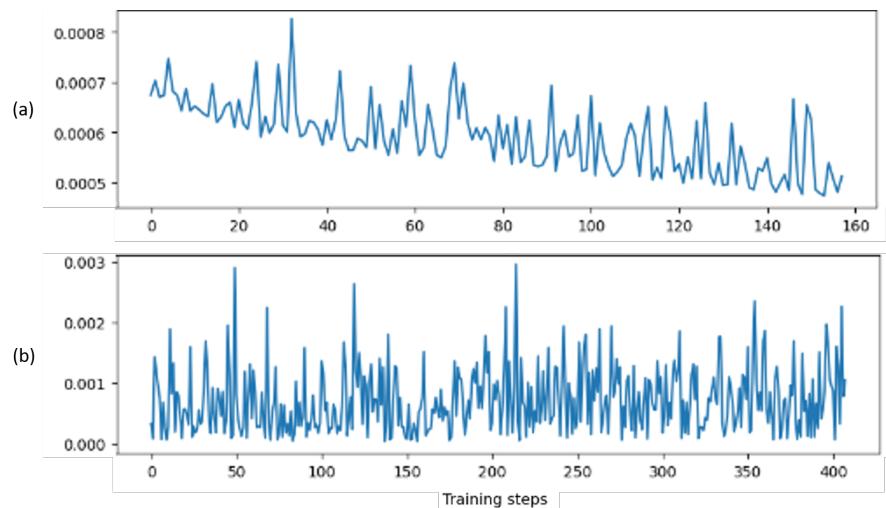


FIGURE 2.7 – Poids en sortie de l'avant dernière couche de la branche ascendante du U-Net 2 pour 3 epochs (a) et 100 epochs (b)

3 | Évaluation de la méthode Noise2Noise

La méthode d'apprentissage profond qui a été codée fait appel à des méthodes d'IA développées récemment, mais d'autres méthodes plus « classiques » permettent également d'avoir une estimation de la réponse forcée. L'évaluation de la plus-value du réseau développé par rapport à celles préexistantes nécessite de la comparer à ces méthodes classiques.

3.1 Méthodes "classiques"

Les méthodes classiques utilisées sont les suivantes :

Tendances La méthode la plus simple consiste à calculer la tendance du changement climatique sur la période en ajustant les données à une fonction linéaire, quadratique ou cubique sur la période d'étude (1900-2023).

Régression linéaire Une seconde méthode consiste à régresser la valeur de la température moyenne globale en chaque point de grille en s'appuyant sur l'hypothèse qui affirme que la température à fine échelle dépend linéairement de la température globale.

Analyse en composantes principales Une troisième méthode s'appuie sur la méthode d'analyse en composantes principales (ACP). L'ACP étudie la variabilité spatiale et temporelle des champs physiques. Elle identifie les principales structures de covariances temporelle et spatiale dans les données afin de les redéfinir dans un nouvel espace dont les « axes » sont appelés fonctions orthogonales empiriques (EOF). Ces fonctions sont classées par fraction de la variance totale capturée. Les données définies dans cette nouvelle base sont appelées composantes principales (PC). Dans notre étude, on se limite à la première composante qui représente le premier mode de variabilité des données : la réponse forcée. Les modes suivants échantillonnent des phénomènes à plus courte échelle temporelle tels que l'ENSO (El Niño-Southern Oscillation). Cette méthode permet de réduire la complexité multi-dimensionnelle des données en ne gardant que les dimensions les plus importantes d'après leur variabilité.

Analyse en composantes basses fréquences Enfin, une dernière méthode reprend celle de l'ACP en la complexifiant : l'analyse en composantes basses fréquences (LFCA). L'ACP maximise la variance capturée par la première EOF et peut regrouper ainsi de multiples processus en suggérant des connexions fallacieuses qui ne sont pas retrouvées dans les mécanismes physiques. [Wills et al. \[2018\]](#) a montré que le premier mode d'une ACP sur les anomalies de température de surface de la mer mensuelles sur le Pacifique confond le signal du changement climatique et de l'ENSO. Le LFCA identifie les modes de variabilité basse fréquence en résolvant une combinaison linéaire d'EOFs qui maximise le ratio variance basse fréquence sur variance totale. La variance basse fréquence est définie comme la variance restante après l'application d'un filtre passe-bas de Lanczos aux données avec pour fréquence de coupure 14 ans. En considérant aussi bien la variance haute-fréquence que celle basse fréquence, cette méthode tire parti de la structure de covariance spatiale de la variabilité à haute fréquence et peut ainsi mieux la filtrer. La méthode de Wills identifie alors des motifs basse fréquence (LFPs) et leurs composantes basse fréquence décorrélées qui sont classées d'après leur ratio variance basse-fréquence

sur variance totale. Ainsi, les modes de variabilité sont triés par rapport à l'échelle temporelle dominante au sein des données et les modes de variabilité basse fréquence sont isolés sans occulter la variabilité rapide qui peut contribuer à leur évolution.

3.2 Comparaisons des méthodes "classiques" et par apprentissage

Les analyses de cette section seront menées uniquement sur le modèle CESM2.

3.2.1 En moyenne globale

La tendance cubique est la meilleure méthode. Sa valeur de RMSE moyennée sur l'ensemble de ses membres est inférieure à celle des autres méthodes et au moins 50% de ses membres ont une valeur de RMSE inférieure au "meilleur" membre de toutes les autres méthodes (figure 3.1).

Viennent ensuite les méthodes de tendance quadratique et le LFCA. Cette seconde méthode présente une plus grande dispersion de ses membres comparativement à la première mais est meilleure en moyenne.

Les méthodes de régression linéaire et le U-Net sont équivalentes mais quelques membres isolés présentent des valeurs supérieures à 0.23 °C pour le réseau de neurones. L'EOF présente une moyenne légèrement supérieure aux deux méthodes précédentes avec une distribution équivalente à celle du réseau de neurones.

Enfin, la tendance linéaire est la pire en moyenne mais la variance de ses membres est la plus faible.

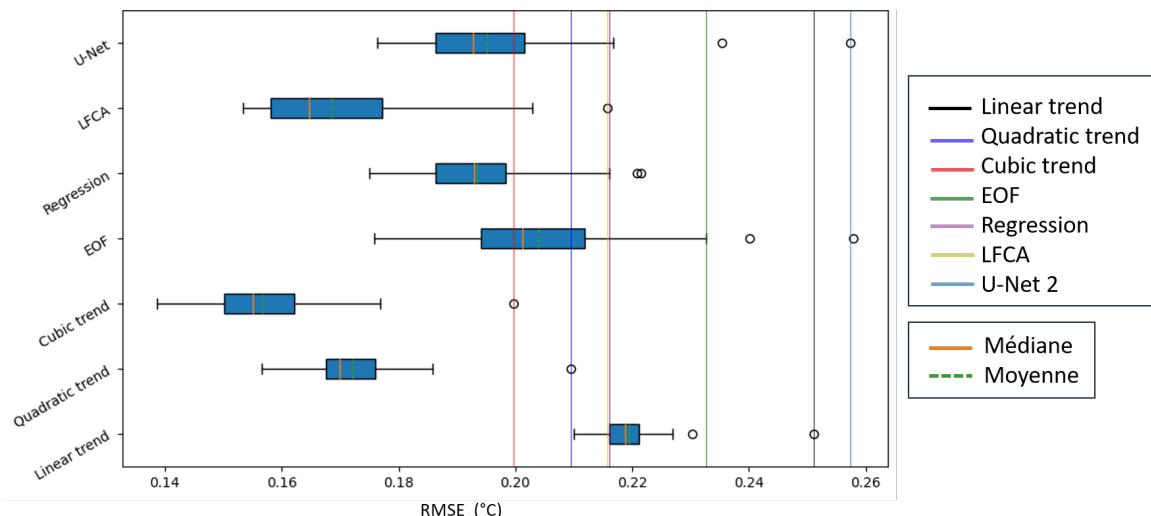


FIGURE 3.1 – Valeurs de RMSE en moyenne globale sur la période 1903-2022 des 50 membres du modèle CESM2 pour chaque méthode d'évaluation de la réponse forcée. Les lignes verticales représentent la position du "pire" membre de la méthode de tendance quadratique pour chacune des autres méthodes.

On constate que le pire membre de la méthode cubique (membre 39) est aussi le pire membre de la méthode linéaire, quadratique, LFCA et du U-Net. Il est positionné à 1.5 fois l'écart inter-quartile au-dessus du 3^{ème} quartile pour l'EOF et la régression.

3.2.2 Les zones sensibles

On constate (figure 3.2) que les zones avec les valeurs de RMSE les plus élevées en moyenne se situent au niveau des pôles et plus particulièrement sur l'Arctique mais aussi en majorité sur les continents.

Les membres du modèle CESM2 et les méthodes ont une variance plus élevée au niveau de la mer du Labrador, du Groenland et jusqu'à la mer de Barents. L'incertitude est également plus élevée sur le bassin Pacifique là où l'on retrouve le phénomène ENSO.

Les zones géographiques où l'on observe de fortes valeurs de RMSE et une incertitude élevée quant à ces valeurs sont celles où les changements de températures sont les plus importants tels que les continents ou le cercle Arctique (amplification polaire).

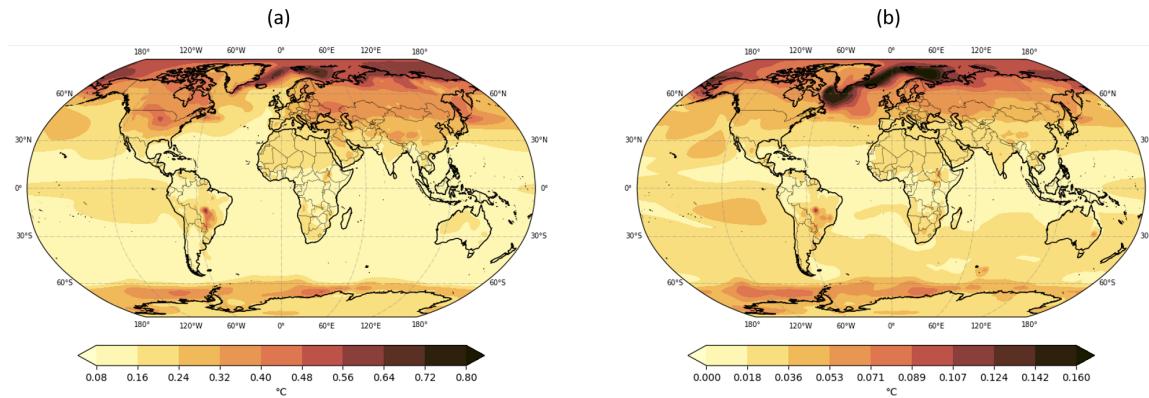


FIGURE 3.2 – Valeurs de RMSE sur la période 1903-2022 avec la moyenne sur les 50 membres et les 7 méthodes étudiées (a), l'écart-type sur les 50 membres et les 7 méthodes (b)

3.2.3 Et le vainqueur est...

Le LFCA arrive au sommet du classement (figure 3.2) en étant la meilleure méthode d'après ses valeurs de RMSE sur près de 43% du globe (les océans en grande partie). La méthode cubique est à la deuxième place du podium en occupant 33% de la surface terrestre. Enfin le U-Net se positionne en 3^{ème} place avec 14%. La régression et l'EOF sont quasiment hors-jeu dans cette compétition avec des valeurs proches de 0%. On constate que le LFCA et le U-Net occupent les zones où l'incertitude sur la valeur de RMSE est la plus grande. La méthode développée par [Wills et al. \[2018\]](#) est bien la meilleure sur la zone d'El Niño/la Niña.

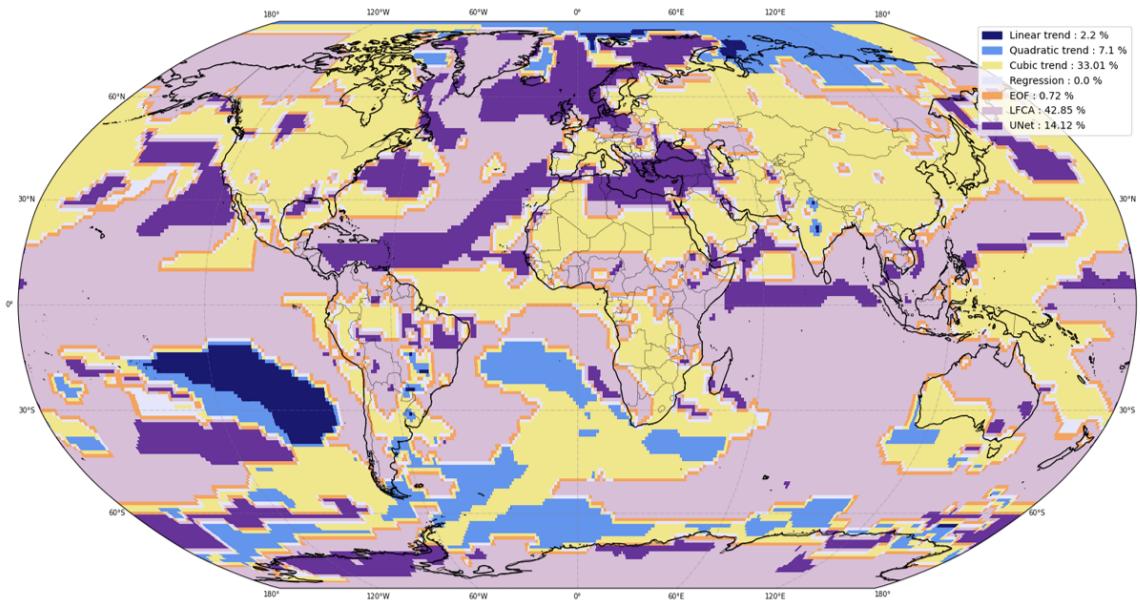


FIGURE 3.3 – Méthode qui a la plus faible valeur de RMSE sur la période 1903-2022 pour chaque point de grille avec la superficie totale occupée par chaque méthode en %

Toutefois, si l'on étudie une période temporelle plus récente où le signal du changement climatique s'est mieux démarqué de la variabilité interne, le classement est différent.

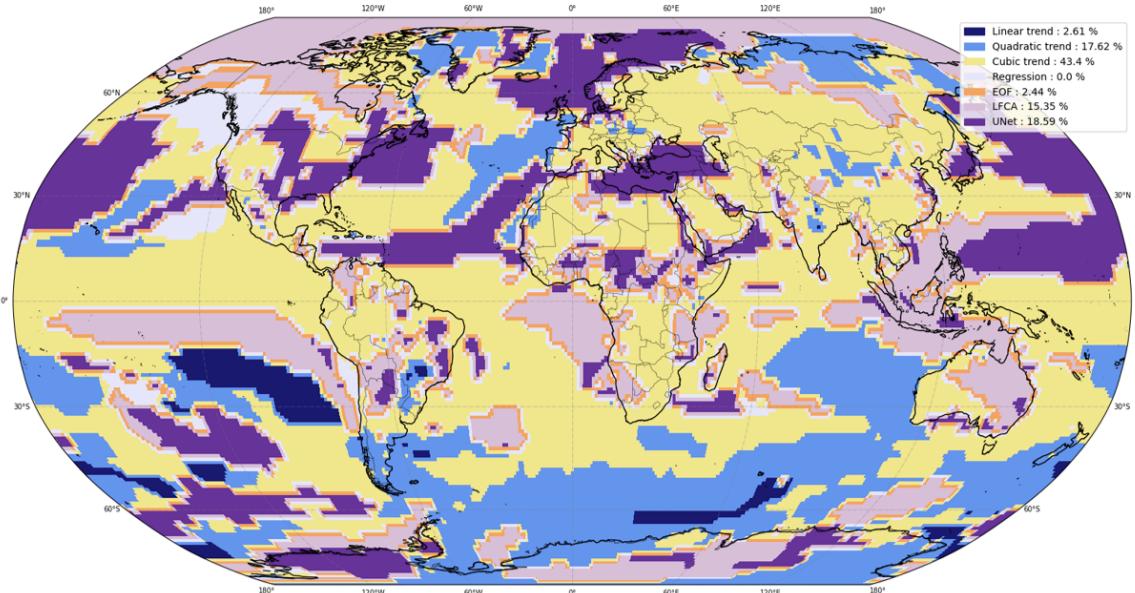


FIGURE 3.4 – Méthode qui a la plus faible valeur de RMSE sur la période 1980-2022 pour chaque point de grille avec la superficie totale occupée par chaque méthode en %

En effet, la tendance cubique prend la première place (43 %) puis vient le U-Net qui devance très légèrement la tendance quadratique (19% contre 18%). Le LFCA est alors évincé du podium en ne se plaçant qu'à la quatrième place mais sans en être très éloigné avec 15% de la surface du globe occupée.

Ainsi, la tendance cubique permet de mieux suivre le réchauffement plus important de ces dernières années. On constate que le LFCA est très largement dominant au niveau de l'Arctique mais perd son monopole sur le bassin Pacifique équatorial.

3.2.4 Application aux observations

Le U-Net 2 est finalement appliqué aux observations (figure 3.5).

On retrouve bien de plus fortes anomalies positives de température sur la région Arctique avec un gradient sud/nord sur les pôles plutôt bien représenté. Néanmoins, l'amplification polaire est sous-estimée avec des valeurs inférieures à 2°C aux très hautes latitudes.

Les anomalies ont également tendance à être plus élevées sur continents que sur océans mais le différentiel de températures entre les deux surfaces est sous-estimé et les délimitations manquent de finesse.

Le phénomène de "l'Atlantic low" où l'on retrouve des anomalies négatives de température sur le nord de l'océan Atlantique est sous-estimé. On constate tout de même la présence de valeurs moins élevées sur cette zone mais sans être négatives.

Les anomalies négatives ou proches de zéros habituellement observées sur les zones de reliefs sont ici absentes.

Ainsi, le débruitage effectué par le U-Net est très largement perfectible mais présente déjà les principales structures du changement climatique identifiées à ce jour.

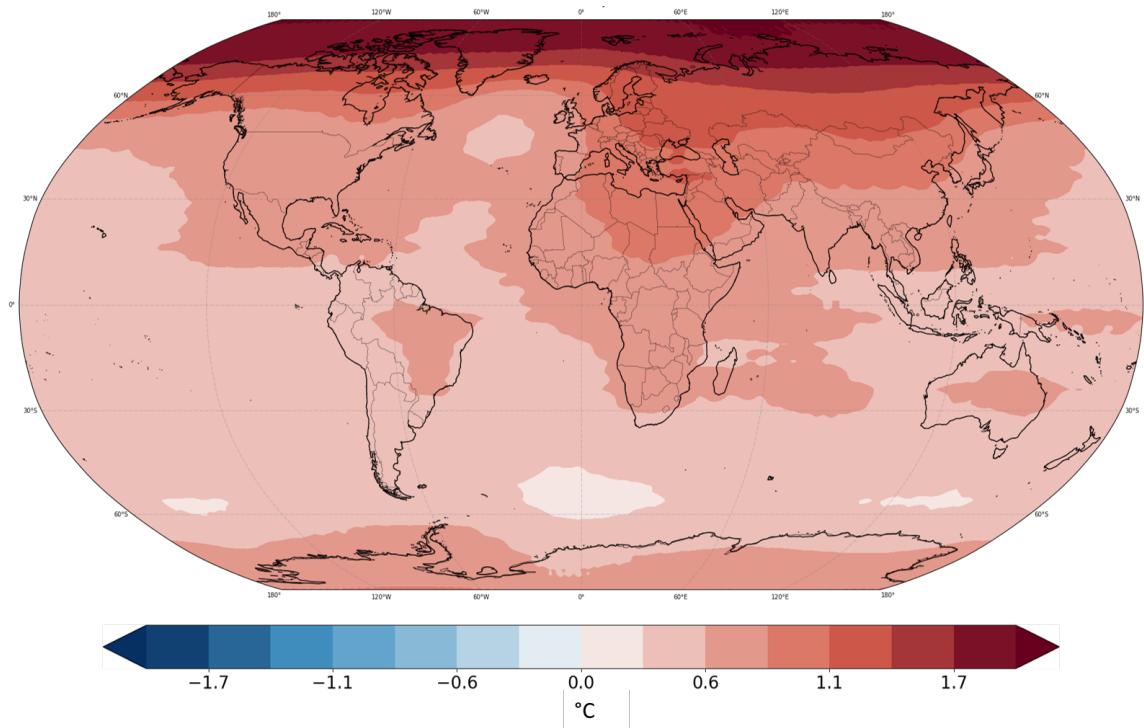


FIGURE 3.5 – Anomalies de température (par rapport à la période 1961-1990) des observations HadCrut5 complétées avec une reconstruction par IA des données manquantes, [Kadow et al. \[2020\]](#), débruitées par le U-Net 2 pour l'année 2020.

4 | Conclusion et perspectives

4.1 Conclusion

Durant ce Projet de Fin d'Études un réseau de neurones basé sur la méthode Noise2Noise de débruitage d'images a été développé afin de distinguer la réponse forcée de la variabilité interne.

Une première étape consistait à se familiariser avec le monde de l'intelligence artificielle. Plusieurs modèles d'IA basés sur l'architecture U-Net ont ensuite été développés afin d'améliorer les performances du réseau, elles-mêmes évaluées selon des métriques telles que l'erreur quadratique moyenne. Après plusieurs étapes d'ajustement des hyperparamètres et d'optimisation du réseau, un U-Net a été sélectionné et comparé à d'autres méthodes plus classiquement utilisées dans le domaine de la climatologie. Cette étape d'évaluation a mis en évidence que la meilleure méthode d'estimation de la réponse forcée dépend fortement de la zone géographique et de la période temporelle étudiée. En moyenne globale sur la période 1903-2022, la tendance cubique est la meilleure. Néanmoins, c'est le LFCA qui occupe la plus grande superficie du globe en tant que meilleure méthode. Si on étudie uniquement les 40 dernières années, la tendance cubique prend alors la première place du podium. Il n'y a donc pas une seule et meilleure méthode, la meilleure solution étant toujours d'en étudier plusieurs.

On constate ainsi que la méthode d'apprentissage profond fournit quelques premiers résultats encourageants. Le réseau de neurones présente plusieurs pistes d'améliorations et les avancées qui seront faites dans le domaine de l'intelligence artificielle lui confère une marge de progression très encourageante pour la poursuite de ce travail.

4.2 Perspectives

Deux axes de poursuite de ce travail pourraient être explorés.

Le premier concerne l'amélioration du réseau de neurones.

- La durée d'entraînement des réseaux utilisant des données mensuelles n'a pas permis le lancement de plusieurs tests pour ces réseaux. Des modifications des hyperparamètres pourraient mener à l'amélioration des scores de ces U-Nets.
- L'architecture en U-Net du réseau semblait être la meilleure option qui s'offrait au début du stage pour le débruitage d'image. Toutefois, cette architecture pourrait être complexifiée en s'inspirant des travaux réalisés dans le domaine des prévisions météorologiques qui traitent déjà de façon plus spécifique les variables de températures, pression, ...
- Une normalisation différente pourrait être envisagée. On étudierait en amont la distribution des valeurs des différentes variables afin de créer la normalisation la plus adaptée à nos données.
- Un plus grand nombre de modèles de climat pourrait être utilisé lors de l'entraînement afin d'avoir un meilleur échantillonnage de la variabilité interne et des résultats plus robustes pour une application aux observations.
- Le travail effectué ici sur les températures pourrait être appliqué à d'autres variables climatiques telles que les précipitations. Une réPLICATION du U-Net 5 a été entraîné sur les données de pression et de précipitations mais aucun ajustement des hyperparamètres à cette nouvelle variable n'a pu être réalisé. Les valeurs de RMSE les plus élevées (figure 4.1) se retrouvent au niveau de la ZCIT

(zone de convergence intertropicale) où l'on retrouve de forts cumuls de précipitations dans les observations.

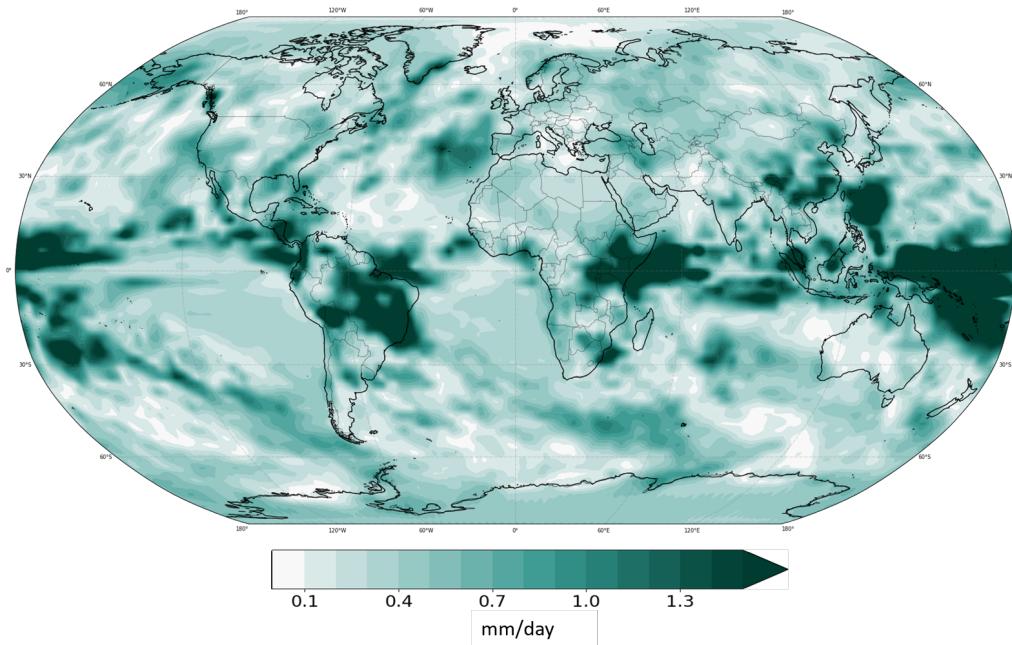


FIGURE 4.1 – RMSE cumulée sur la période 1903-2022 pour les précipitations débruitées avec de l'IA du membre 26 du modèle CESM2

Le second concerne l'étude des différentes méthodes d'évaluation de la réponse forcée.

- Les métriques et méthodes développées pourraient toutes se décliner sur des périodes temporelles et des régions différentes.
- Des liens entre la meilleure méthode sur une période et une région donnée, et la dynamique physique des variables climatiques sur cette même période et région serait très intéressants à établir et permettraient de mieux comprendre pourquoi une méthode en particulier s'avère être la plus adaptée pour un cadre d'étude donné.
- Des méthodes développées plus récemment telles que celle des empreintes optimales régularisées ([Ribes et al. \[2013\]](#)) pourrait être ajoutées aux comparaisons effectuées.

Bibliographie

C. Bône, G. Gastineau, S. Thiria, P. Gallinari, and C. Mejia. Separation of internal and forced variability of climate using a u-net. *Journal of Advances in Modeling Earth Systems*, 2024. doi : <https://doi.org/10.1029/2023MS003964>.

Francois Chollet. *Deep Learning with Python*. 2021.

GIEC. Ipcc, 2023 : Climate change 2023 : Synthesis report. contribution of working groups i, ii and iii to the sixth assessment report of the intergovernmental panel on climate change. Technical report, IPCC, 2023.

C. Kadow, D.M. Hall, and U. Ulbrich. Artificial intelligence reconstructs missing climate information. *Nature geoscience*, 2020. doi : <https://doi.org/10.1038/s41561-020-0582-5>.

Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise : Learning image restoration without clean data, 2018. URL <https://arxiv.org/abs/1803.04189>.

Kevin P. Murphy. *Probabilistic Machine Learning : An introduction*. MIT Press, 2022. URL probml.ai.

A. Ribes, S. Planton, and L. Terray. Application of regularised optimal fingerprinting to attribution. part i : method, properties and idealised analysis. *Climate dynamics*, 2013. doi : <https://doi.org/10.1007/s00382-013-1735-7>.

P. Ronneberger, O.and Fischer and T. Brox. U-net : Convolutional networks for biomedical image segmentation. 2015. doi : https://doi.org/10.1007/978-3-319-24574-4_28.

R. C. Wills, T. Schneider, J. M. Wallace, D. S. Battisti, and D. L Hartmann. Disentangling global warming, multidecadal variability, and el niño in pacific temperatures. *Geophysical Research Letters*, 2018.

Robert Jnglin Wills. Forcesmip. URL <https://sites.google.com/ethz.ch/forcesmip/about?authuser=0>.