

Assimilation de données par apprentissage pour le suivi de phénomènes océaniques chaotiques

Perrine BAUCHOT¹ Angélique DRÉMEAU¹ Florian SÉVELLEC² Ronan FABLET^{1,3}

¹ Lab-STICC, CNRS ENSTA Bretagne IMT Atlantique UBO ENIB UBS, Brest, France

² Laboratoire d'Océanographie Physique et Spatiale, Univ Brest CNRS IRD Ifremer, Brest, France

³ INRIA team ODYSSEY, Brest, France

Résumé – En géosciences, l'assimilation de données s'est développée pour reconstruire et prédire des processus dynamiques à partir de séries d'observations et d'*a priori* sur la physique des systèmes étudiés. Récemment, l'hybridation de l'assimilation de données et de l'apprentissage profond a ouvert de nouvelles perspectives pour aborder ces relations modèles-données, notamment dans le cas d'observations partielles et peu denses. Dans cet article, nous en étudions l'apport potentiel pour l'analyse de phénomènes océaniques chaotiques, plus particulièrement des événements de Dansgaard-Oeschger qui caractérisent les dynamiques océaniques dans l'Atlantique Nord lors de la dernière ère glaciaire. Nos résultats expérimentaux suggèrent des gains relatifs très importants par rapport à une méthode d'assimilation variationnelle classique lorsque la densité des observations diminue. Ils mettent également en évidence que l'exploitation explicite du modèle dynamique *a priori* ne conduit pas nécessairement aux meilleures performances.

Abstract – In geosciences, data assimilation has been developed to reconstruct and predict dynamical processes from series of observations and from *a priori* on the physics of the studied systems. Recently, the hybridization of data assimilation and deep learning has opened new perspectives to address these model-data interactions, especially in the case of partial and sparse observation. In this paper, we investigate its potential contribution to the analysis of chaotic oceanic phenomena, more specifically to the Dansgaard-Oeschger events which characterise the ocean dynamics in the North Atlantic during the last glacial era. Our experimental results suggest very large relative gains over a classical variational assimilation method when the density of observations decreases. They also show that the explicit exploitation of the *a priori* dynamical model does not necessarily lead to the best performances.

1 Introduction

L'océan est un milieu vaste et imprévisible, où de nombreux phénomènes chaotiques de différentes échelles spatio-temporelles se côtoient et dont les interactions restent encore parfois incomprises. Alors que l'océan occupe plus de 70% de la surface terrestre, seul 10% de ce dernier a été exploré [11]. A la fin du XX^{ème} siècle, l'avènement des satellites d'observations terrestres a permis de faire des progrès considérables en océanographie grâce à une couverture complète et régulière des océans, avec une résolution de plus en plus fine [12]. Cependant, les ondes électromagnétiques des satellites ne pénètrent pas dans la colonne d'eau, qui est observée par mesures *in situ* acquises de manière ponctuelle, notamment au cours de campagnes océanographiques [2]. La rareté des données *in situ* et l'observation partielle des phénomènes océaniques empêchent d'étudier précisément les équilibres et les échanges entre masses d'eau de densités et profondeurs différentes, alors même qu'ils jouent un rôle crucial dans le système Terre. C'est notamment le cas de la Circulation Océanique de Retour-Méridien de l'Atlantique – aussi appelée AMOC – qui correspond au transport de masses d'eau chaudes de surface vers le Pôle Nord et le retour de celles-ci en eaux froides en profondeur, induisant un transport de chaleur dans l'Atlantique Nord. L'AMOC joue un rôle essentiel dans la régulation du climat puisqu'elle capte les excès de chaleur et d'émissions de carbone de la surface terrestre vers le fond des océans [3]. Différentes études ont mis en évidence le caractère chaotique

de la dynamique de l'AMOC, qui a connu des périodes de ralentissement abruptes lors de la dernière ère glaciaire, appelées événements de Dansgaard-Oeschger [17].

La reconstruction de la variabilité des processus océaniques, tels que l'AMOC, exploite classiquement des méthodes d'assimilation de données permettant de combiner observations disponibles et modélisation physique [5]. Deux grandes familles de méthodes de résolution ont émergé : les approches séquentielles telles que le filtre de Kalman, et les approches variationnelles [8, 5]. En combinant les informations apportées par les observations et celles contenues dans le modèle dynamique, l'assimilation variationnelle fournit une estimation de l'état du système par la minimisation des écarts entre les données réelles, le modèle physique et l'état analysé. Au cours des dernières années, différentes études ont exploré les synergies entre apprentissage profond et assimilation de données, encouragées par le développement de l'intelligence artificielle informée par la physique (cf. [6, 14, 1, 4]). Dans ce contexte, la méthode 4DVarNet [9] exploite un modèle neuronal 'bout-en-bout' pour apprendre à représenter la dynamique du système étudié et pour résoudre conjointement une formulation variationnelle, dans le but d'optimiser la reconstruction de l'état vrai du système, sur la base des observations fournies.

Dans cette étude, nous étudions la pertinence de ces approches neuronales pour reconstruire des processus océaniques, tels que l'AMOC, pour lesquels les observations disponibles sont très peu denses. Nous nous focalisons sur la représentation réduite des événements de Dansgaard-Oeschger

[17]. Au-delà de l’intercomparaison entre trois méthodes d’assimilation, nous proposons une première analyse de l’impact de la densité des observations sur la reconstruction de la dynamique de l’AMOC durant la dernière ère glaciaire.

La suite de l’article est structurée en trois parties. Dans une première partie, nous introduisons l’assimilation de données pour la reconstruction de phénomènes océaniques partiellement observés, puis nous transposons cette méthode dans le cadre de l’apprentissage profond. Dans une deuxième partie, nous présentons le système physique sous-tendant les données considérées dans cet article. Enfin, nous montrons les apports de l’apprentissage profond à l’assimilation de données dans le cas des évènements de Dansgaard-Oeschger.

2 Méthodologie

Nous considérons le système d’équations suivant :

$$\begin{cases} \dot{\mathbf{X}}(t) = \mathbf{M}(\mathbf{X}(t)) \\ \mathbf{Y}(t) = \mathbf{H}_t \mathbf{X}(t) + \mathbf{q}_{\text{obs}}(t) \end{cases} \quad (1)$$

où \mathbf{X} représente l’état que nous cherchons à reconstruire à l’instant t considéré. \mathbf{Y} est le vecteur des observations acquises avec un bruit $\mathbf{q}_{\text{obs}}(t) \sim \mathcal{N}(0, \sigma_{\text{obs}}^2)$. \mathbf{M} représente le modèle dynamique, c’est-à-dire les équations physiques modélisant le phénomène étudié, tandis que \mathbf{H}_t représente le modèle d’observation et fait donc référence à l’échantillonnage choisi ou imposé pour observer \mathbf{X} .

2.1 Assimilation variationnelle : 4DVar

La méthode variationnelle 4DVar produit une estimation $\hat{\mathbf{X}}$ de l’état vrai \mathbf{X} selon une approche permettant de propager l’information apportée par les observations au reste du système non observé [10, 5]. Ce problème d’optimisation repose sur la minimisation d’un coût variationnel J sur une fenêtre temporelle d’assimilation ΔT fixée, tel que $\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} J(\mathbf{X})$ où J est défini selon :

$$\begin{aligned} J(\mathbf{X}) = & \frac{1}{\Delta T} \left[\alpha_{\text{obs}} \int_{t_0}^{t_0 + \Delta T} \|\mathbf{Y}(t) - \mathbf{H}_t \mathbf{X}(t)\|_2^2 dt \right. \\ & + \alpha_{\text{dyn}} \int_{t_0}^{t_0 + \Delta T} \|\phi(\mathbf{X}(t)) - \dot{\mathbf{X}}(t)\|_2^2 dt \\ & \left. + \alpha_B \|\mathbf{X}_B - \mathbf{X}(t_0)\|_2^2 \right]. \end{aligned} \quad (2)$$

Le premier terme du coût variationnel J évalue la différence entre les observations et l’état du système, tandis que le deuxième terme calcule la différence entre la sortie du modèle dynamique \mathbf{M} et l’état \mathbf{X} sur la fenêtre temporelle choisie ΔT , sur la base d’une fonction ϕ classiquement implémentant les équations physiques (e.g. (1)). Le dernier terme de ce coût variationnel est une régularisation appliquée pour assurer la stabilité de l’assimilation. Elle fournit un *a priori* sur la condition initiale du système et est communément appelée *background*. Enfin, α_{obs} , α_{dyn} et α_B sont des coefficients de pondération à ajuster manuellement en fonction du cas d’étude.

2.2 Avec l’apprentissage profond : 4DVarNet

Différentes études ont récemment exploré la définition d’architectures de réseaux de neurones dérivées des méthodes

d’assimilation de données [9, 4, 7]. Nous nous intéressons ici à la méthode 4DVarNet qui repose sur une formulation variationnelle de type 4DVar et met en oeuvre un schéma itératif analogue à une descente de gradient.

A chaque itération $i \in \mathbb{N}^*$, l’état analysé est mis à jour selon $\mathbf{X}^{(i)} = \mathbf{X}^{(i-1)} - \delta \text{LSTM}[\nabla J(\mathbf{X}^{(i-1)})]$ où δ pondère la mise à jour de l’état \mathbf{X} . Comme en méta-apprentissage [19], l’utilisation d’une couche Long-Short Term Memory (LSTM) convolutive permet d’optimiser la descente de gradient, en prenant en compte les itérations précédentes.

Dans cette étude, 4DVarNet est testé selon deux configurations : *i*) 4DVarNet-ode considère une fonction ϕ définie selon les équations physiques (1) et l’implémente par intégration numérique selon un schéma de Runge-Kutta 4 ; *ii*) 4DVarNet-unet définit ϕ par apprentissage d’un réseau U-Net [16], capable de capturer les différentes échelles du système considéré.

Le réseau de neurones LSTM est entraîné par apprentissage supervisé reposant sur la minimisation d’un coût d’apprentissage défini comme la somme des erreurs quadratiques moyennes de reconstruction de l’état \mathbf{X} . Une régularisation sur la dérivée temporelle de \mathbf{X} est également appliquée pour limiter le bruit de reconstruction. L’entraînement est réalisé sur des *batches* de taille 128 par un optimiseur d’Adam. 300 époques d’entraînement, suivies par 150 époques permettant d’obtenir des réglages plus précis, sont nécessaires pour atteindre la convergence du coût d’apprentissage. Notons que dans la configuration 4DVarNet-unet, le modèle et le solveur sont appris conjointement lors de l’entraînement, à partir de données générées par les équations physiques [9].

3 Données

3.1 Modèle dynamique considéré

Notre étude expérimentale exploite une représentation réduite de l’AMOC proposée dans [17]. Ce modèle idéalisé permet de reproduire la variabilité centennale à millénaire de l’AMOC observée lors des évènements de Dansgaard-Oeschger de la dernière ère glaciaire [18]. Il est décrit par un état tri-dimensionnel $\mathbf{X}(t) \triangleq [x_1(t), x_2(t), x_3(t)]^T$ régi par le système d’équations différentielles ordinaires suivant :

$$\dot{\mathbf{X}}(t) = \mathbf{M}(\mathbf{X}(t)) = \begin{pmatrix} -\lambda(x_1(t) + x_3(t)) \\ (\omega_0 + x_1(t))x_3(t) - kx_2(t) + f \\ -(\omega_0 + x_1(t))x_2(t) - kx_3(t) \end{pmatrix}$$

où λ est un coefficient de friction linéaire, ω_0 est la composante constante de la circulation océanique de retournement méridien tel que $\omega = \omega_0 + x_1$, k est le coefficient d’amortissement linéaire et f est l’intensité du flux d’eau douce.

3.2 Génération des données

Nous avons simulé le modèle \mathbf{M} sur une période de 100.000 ans, via une intégration numérique avec un pas de temps de 1 an. Cette simulation longue est ensuite découpée en séries temporelles de $\Delta T = 2.500$ ans. Afin de reconstruire avec précision les changements de régime de l’AMOC, nous nous sommes assurés de la présence d’au moins une période d’arrêt de l’AMOC, correspondant à un extrême climatique, dans chaque série temporelle. $\frac{2}{3}$ (resp. $\frac{1}{3}$) de ces périodes composent l’ensemble de données d’entraînement (resp. test). Les

observations $\mathbf{Y}(t)$ sont construites à partir de cette simulation, selon le modèle linéaire (1). Nous nous plaçons dans le cas où seules x_2 et x_3 sont observées, *i.e.* \mathbf{H}_t est une matrice identité dans laquelle le premier élément de la diagonale est remplacé par un 0. Les points d'observations sont d'abord placés à une fréquence d'échantillonnage régulière de $\frac{1}{50} \text{ yr}^{-1}$ pour capturer les oscillations centennales de l'AMOC (de l'ordre de 250 ans), et ainsi respecter le critère de Shannon. Cette fréquence est ensuite modifiée entre $[\frac{1}{10} \text{ yr}^{-1}, \frac{1}{120} \text{ yr}^{-1}]$ pour étudier l'influence de la densité des observations sur les performances de reconstruction. Enfin, un bruit de mesure est ajouté et correspond à 10% de la variance des oscillations centennales de l'AMOC. Ce pourcentage permet d'appliquer le même « ratio » de bruit sur chaque variable observée.

4 Résultats

4.1 Comparaison 4DVar vs. 4DVarNet

Afin d'évaluer les performances de reconstruction des trois méthodes d'assimilation de données implémentées (4DVar, 4DVarNet-ode, 4DVarNet-unet), nous calculons l'erreur quadratique moyenne normalisée de chaque variable reconstruite selon la formule suivante :

$$\text{NMSE} = \frac{1}{N \Delta T \sigma_{\text{GT}}^2} \sum_{n=0}^{N=100} \sum_{t=0}^{\Delta T=2500} (\hat{\mathbf{X}}_n(t) - \mathbf{X}_n^{\text{GT}}(t))^2$$

où \mathbf{X}_n^{GT} représente la vérité terrain, σ_{GT}^2 la variance totale du système \mathbf{M} et N le nombre de trajectoires simulées sur lesquelles nous évaluons les performances de reconstruction.

Dans le tableau ci-dessous, nous constatons que la méthode 4DVarNet-unet permet de diviser par 10, et même par 50 pour x_2 et x_3 les erreurs de reconstruction par rapport à une méthode d'assimilation de données classique. 4DVarNet-ode, méthode pour laquelle seul le problème d'assimilation de données est amélioré par apprentissage, obtient également une reconstruction près de 10 fois meilleure sur x_2 et x_3 que la méthode 4DVar classic. Cela démontrerait d'ores et déjà que l'usage d'un réseau de neurones permettrait d'améliorer la résolution d'un problème d'assimilation de données variationnelle.

Méthode	NMSE x_1	NMSE x_2	NMSE x_3
4DVar classic	9.17×10^{-3}	1.36×10^{-1}	1.44×10^{-1}
4DVarNet-ode	2.85×10^{-3}	1.24×10^{-2}	9.39×10^{-3}
4DVarNet-unet	9.05×10^{-4}	2.58×10^{-3}	2.71×10^{-3}

TABLE 1 : Erreur quadratique moyenne normalisée de reconstruction des trois variables du système AMOC par les trois méthodes d'assimilation de données testées

Remarquons également que les erreurs de reconstruction sont plus faibles sur x_1 alors même que cette variable n'est pas observée. En visualisant l'évolution temporelle et fréquentielle de chacune des variables reconstruites sur une trajectoire, cela nous permet de mieux appréhender la physique du système \mathbf{M} (Fig. 1(a) et (b)). L'évolution temporelle souligne la légère différence d'échelle entre x_1 et (x_2, x_3) , ce qui pourrait expliquer les écarts de performance constatés lors du calcul de la NMSE. Par ailleurs, les variations de plus grandes amplitudes,

correspondant aux événements de ralentissement de l'AMOC (*i.e.* $x_1 > 0.05 \text{ yr}^{-1}$), semblent être reconstruits avec davantage de précision par les deux méthodes d'assimilation par apprentissage que par la méthode 4DVar classic. En modifiant la fonction de coût variationnel, il serait peut-être possible de favoriser la reconstruction de ces extrêmes par 4DVar classic. Cela constitue une des contraintes de la méthode 4DVar, dont la forte sensibilité nécessite de trouver un paramétrage adapté à la situation considérée, alors que l'apprentissage profond permet d'estimer directement les paramètres optimaux lors de l'entraînement du réseau de neurones.

Les spectres en fréquence nous permettent également de retrouver les fréquences caractéristiques du signal que nous cherchons à reconstruire, et notamment la fréquence des événements de Dansgaard-Oeschger. En plus d'obtenir les meilleurs scores en termes de NMSE, nous vérifions ainsi que le 4DVarNet-unet, qui n'utilise pas directement les équations physiques lors de l'assimilation, respecte tout de même les caractéristiques spectrales du système. Ce résultat s'inscrit dans le cadre de la découverte de modèles basés sur les données, un domaine de recherche en plein essor, notamment pour étudier les phénomènes qui n'ont pas encore pu être décrits par des équations basées sur les lois de la physique [15, 13].

4.2 Sur l'échantillonnage des observations

Pour terminer cette étude, nous nous intéressons à l'évolution de l'erreur en fonction de la période d'échantillonnage. Comme nous pouvions nous y attendre, l'erreur de reconstruction croît avec la période d'échantillonnage. Cependant, la méthode 4DVar classic est très rapidement affectée par la raréfaction des données, avec une erreur supérieure à 10% dès que la période d'échantillonnage dépasse 40 ans sur la reconstruction de x_2 et x_3 . En revanche, nous notons les performances des méthodes 4DVarNet-ode/unet, qui même avec une période d'échantillonnage supérieure à 100 ans, continuent de produire des résultats très satisfaisants, avec des erreurs de reconstruction de 2% en moyenne pour le 4DVarNet-unet (Fig. 1(c)). 4DVarNet-unet présenterait donc l'avantage d'obtenir de meilleures performances avec moins d'observations, ce qui est particulièrement recherché en océanographie lorsque les données *in situ* ne sont que ponctuelles dans le temps et l'espace.

5 Conclusion

Dans cet article, nous avons proposé d'utiliser deux méthodes d'assimilation de données variationnelle couplées à de l'apprentissage profond : 4DVarNet-ode et 4DVarNet-unet. Ces méthodes ont prouvé leur aptitude à améliorer la résolution du problème de minimisation du coût variationnel. Appliqué au problème de reconstruction de l'AMOC, le 4DVarNet-unet présente en particulier de bonnes performances en reconstruction des extrêmes, grâce à une meilleure capture des différentes échelles du système. Toutefois, une observation peut contenir davantage d'informations sur la dynamique d'un phénomène qu'une autre. Optimiser le choix de la position et du moment d'acquisition d'une observation devrait donc permettre d'affiner notre connaissance des processus océaniques les plus difficiles à étudier.

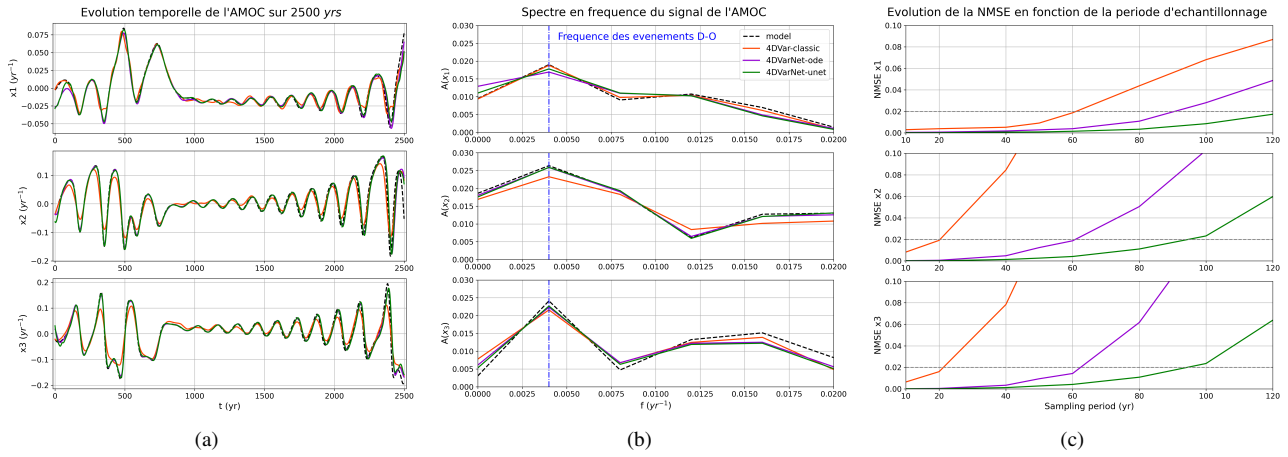


FIGURE 1 : Reconstruction de (x_1, x_2, x_3) par les méthodes d’assimilation considérées : (a) et (b) – Analyse qualitative temporelle et fréquentielle d’une trajectoire selon le setup décrit dans la section 4.1. (c) – Evolution de la NMSE moyenne en fonction de la période d’échantillonnage selon le setup décrit dans la section 4.2.

Références

- [1] Rossella ARCUCCI, Jiangcheng ZHU, Shuang HU et Yi-Ke GUO : Deep Data Assimilation : Integrating Deep Learning with Data Assimilation. *Applied Sciences*, 11(3):1114, janvier 2021.
- [2] Bàrbara BARCELÓ-LLULL, Ananda PASCUAL, Antonio SÁNCHEZ-ROMÁN, Eugenio CUTOLO, Francesco D’OVIDIO, Gina FIFANI, Enrico SER-GIACOMI, Simón RUIZ, Evan MASON, Frédéric CYR, Andrea DOGLIOLI, Baptiste MOURRE, John T. ALLEN, Eva ALOU-FONT, Benjamín CASAS, Lara DÍAZ-BARROSO, Franck DUMAS, Laura GÓMEZ-NAVARRO et Cristian MUÑOZ : Fine-Scale Ocean Currents Derived From in situ Observations in Anticipation of the Upcoming SWOT Altimetric Mission. *Frontiers in Marine Science*, 8, 2021.
- [3] Niklas BOERS : Observation-based early-warning signals for a collapse of the Atlantic Meridional Overturning Circulation. *Nature Climate Change*, 11(8):680–688, août 2021.
- [4] Julien BRAJARD, Alberto CARASSI, Marc BOCQUET et Laurent BERTINO : Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations : a case study with the Lorenz 96 model. *Journal of Computational Science*, 44:101171, juillet 2020. arXiv :2001.01520.
- [5] A. CARRASSI, Marc BOCQUET, Laurent BERTINO et Geir EVENSEN : Data Assimilation in the Geosciences - An overview on methods, issues and perspectives. *Wiley Interdisciplinary Reviews : Climate Change*, 9, septembre 2017.
- [6] Ricky T. Q. CHEN, Yulia RUBANOVA, Jesse BETTENCOURT et David DUVENAUD : Neural Ordinary Differential Equations, décembre 2019. arXiv :1806.07366.
- [7] Emmanuel de BEZENAC, Arthur PAJOT et Patrick GALLINARI : Deep Learning for Physical Processes : Incorporating Prior Scientific Knowledge, janvier 2018. arXiv :1711.07970.
- [8] Geir EVENSEN : *Data Assimilation*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [9] Ronan FABLET, Bertrand CHAPRON, L. DRUMETZ, E. MÉMIN, O. PANNEKOUCKE et F. ROUSSEAU : Learning Variational Data Assimilation Models and Solvers. *Journal Of Advances In Modeling Earth Systems*, 13(10), octobre 2021.
- [10] C. JOHNSON, N. K. NICHOLS et B. J. HOSKINS : Very large inverse problems in atmosphere and ocean modelling. *International Journal for Numerical Methods in Fluids*, 47(8-9):759–771, mars 2005.
- [11] Eonjoo KIM et Mae L. SETO : An Approach to Choose Observation Systems to Observe Ocean Phenomena. In *2022 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 306–313, septembre 2022. ISSN : 2576-7046.
- [12] Walter MUNK : Chapter 1 Oceanography before, and after, the advent of satellites. In David HALPERN, éditeur : *Elsevier Oceanography Series*, volume 63 de *Satellites, oceanography and society*, pages 1–4. Elsevier, janvier 2000.
- [13] Said OUALA, Duong NGUYEN, Lucas DRUMETZ, Bertrand CHAPRON, Ananda PASCUAL, Fabrice COLLARD, Lucile GAULTIER et Ronan FABLET : Learning Latent Dynamics for Partially-Observed Chaotic Systems. *Chaos : An Interdisciplinary Journal of Nonlinear Science*, 30(10):103121, octobre 2020. arXiv :1907.02452.
- [14] M. RAISSI, P. PERDIKARIS et G.E. KARNIADAKIS : Physics-informed neural networks : A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, février 2019.
- [15] Maziar RAISSI, Paris PERDIKARIS et George Em KARNIADAKIS : Multistep Neural Networks for Data-driven Discovery of Nonlinear Dynamical Systems, janvier 2018. arXiv :1801.01236.
- [16] Olaf RONNEBERGER, Philipp FISCHER et Thomas BROX : U-Net : Convolutional Networks for Biomedical Image Segmentation. In Nassir NAVAB, Joachim HORNEGGER, William M. WELLS et Alejandro F. FRANGI, éditeurs : *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Lecture Notes in Computer Science, pages 234–241, Cham, 2015. Springer International Publishing.
- [17] Florian SÉVELLEC et Alexey V. FEDOROV : Millennial Variability in an Idealized Ocean Model : Predicting the AMOC Regime Shifts. *Journal of Climate*, 27(10):3551–3564, mai 2014.
- [18] Florian SÉVELLEC et Alexey V. FEDOROV : Unstable AMOC during glacial intervals and millennial variability : The role of mean sea ice extent. *Earth and Planetary Science Letters*, 429:60–68, novembre 2015.
- [19] Joaquin VANSCHOREN : Meta-Learning : A Survey, octobre 2018. arXiv :1810.03548.