

ÉLÉMENT DE PORTFOLIO 04



Publication

1 DÉFINITION DE CET ÉLÉMENT

Titre de l'élément : Filoche, A., Archambault, T., Charantonis, A. A., and Béréziat, D. (2022). Statistics-free interpolation of ocean observations with deep spatio-temporal prior. In ECML/PKDD Workshop on Machine Learning for Earth Observation and Prediction (MACLEAN).

URL de l'élément : <https://hal.sorbonne-universite.fr/hal-03765735>

2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

L'assimilation de données utilise l'ensemble des informations disponibles pour reconstruire l'état d'un système physique tel que l'atmosphère ou l'océan. En particulier, elle nécessite la connaissance approximative de l'état du système, appelée ébauche, et de la covariance associée aux erreurs induites par cette ébauche (on fait classiquement une hypothèse de loi gaussienne). Cette matrice de covariance, notée B dans la suite, n'est pas toujours connue en pratique mais peut s'obtenir par analyse statistique d'un historique des états passés du système étudié.

Nous proposons une approche alternative aux méthodes utilisées en assimilation de données qui permet de lever l'hypothèse gaussienne faite sur l'erreur de l'ébauche. Cette approche utilise un réseau de neurones profond dont l'architecture permet d'encoder les corrélations spatiales du vecteur d'état décrivant le système. Dans un régime de changement climatique, l'analyse statistique de l'historique perd en partie de sa pertinence, notre approche permettrait d'y répondre.

3 PRÉSENTATION DE CET ÉLÉMENT

En notant X l'état d'un système, Y les observations partielles de ce système, H l'opérateur d'observation (il projette l'état vers l'espace des observations), X_b une ébauche de l'état du système, l'approche standard 3DVar propose de reconstruire X en minimisant la fonction objectif :

$$J(X) = \|Y - HX\|_R^2 + \|X - X_b\|_B^2 \quad (1)$$

où l'on note $\|Z\|_A^2 = \langle Z, A^{-1}Z \rangle$, c'est-à-dire que l'on suppose que les deux termes d'erreur dans J suivent une loi gaussienne dans un formalisme bayésien. Si le bruit sur les observations est connu, il est décrit par la matrice de covariance R , celui sur le terme d'ébauche ne l'est pas. En pratique la matrice B est conçue par des experts ou par analyse statistique.

Nous avons proposé de contrôler la minimisation de J dans l'espace des paramètres d'un réseau de neurones [1] correctement choisi. Ainsi la fonction objectif devient :

$$J(\theta) = \|Y - H \circ g_\theta(z)\|_R^2 \quad \text{s.t.} \quad X = g_\theta(z)$$

et est illustrée par le diagramme de la figure 1. Ainsi, le terme d'ébauche disparaît, et le réseau g_θ est alors

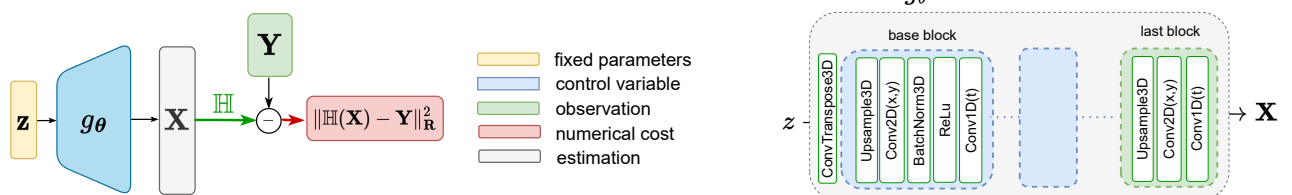


FIGURE 1 – Diagramme de Deep Prior 3DVar

entraîné sur un seul exemple : la donnée que l'on souhaite reconstruire, au contraire des approches habituelles en apprentissage statistique qui sont entraînées sur un corpus d'exemples.

Illustrons cet algorithme sur le problème de reconstruction d'une carte de hauteur d'eau à partir d'observations très partielles qui sont des traces de deux satellites (NADIR et SWOT), tel que montré par la figure 2. La hauteur d'eau est l'une des variables décrivant l'état de l'océan. Avoir une observation de ce champ sur l'ensemble de la grille est nécessaire pour alimenter les modèles océanographiques. Ici on dispose d'une vérité terrain car la méthode est évaluée sur la sortie d'un modèle de simulation numérique de circulation océanographique (NEMO, couramment utilisé par la communauté des océanographes) et une simulation des traces satellites réalisée par un algorithme de l'état de l'art.

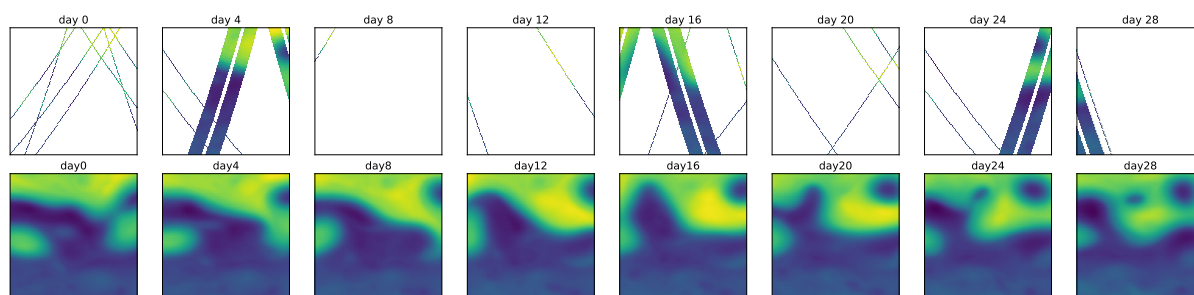


FIGURE 2 – Traces satellites NADIR et SWOT de hauteur d'eau et vérité terrain associée

La méthode de référence est DUACS [4], il s'agit d'une interpolation optimale utilisant 25 années d'observation et d'analyse pour estimer B qui minimise la fonctionnelle de l'équation 1.

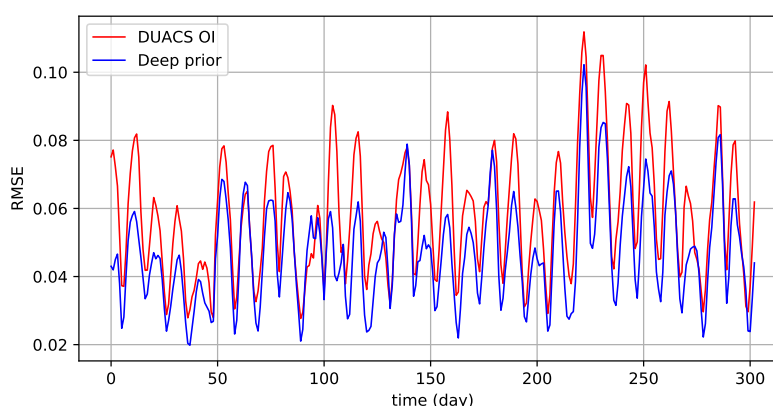



FIGURE 3 – Comparaison entre DUACS et Deep Prior 3DVar sur une année (métrique : erreur quadratique moyenne).

Comme on peut le voir en figure 3, notre méthode bat quasi-systématiquement DUACS sans information sur la matrice B . En conclusion, un réseau génératif profond peut agir comme une matrice d'ébauche B dans 3DVar. Le principe est que le réseau utilise des convolutions 3D pour capturer des corrélations spatio-temporelles compatibles avec la réalité physique. Cette approche a également été utilisée pour retrouver des cartes de circulation qui ne sont jamais observées en observant uniquement des cartes de hauteurs d'eau incomplètes en temps [3], ou pour faire simultanément la reconstruction de la circulation et/ou augmenter significativement (facteur 8 par 8) la résolution spatiale des cartes de hauteurs d'eau [2].

4 RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Arthur Filoche, Théo Archambault, Anastase Alexandre Charantonis, and Dominique Béréziat. Statistics-free interpolation of ocean observations with deep spatio-temporal prior. In *ECML/PKDD Workshop on Machine Learning for Earth Observation and Prediction (MACLEAN)*, September 2022.
- [2] Arthur Filoche and Dominique Béréziat. Simultaneous Assimilation and Downscaling of Simulated Sea Surface Heights with Deep Image Prior. In *RFIAP*, July 2022.

- 
- [3] Arthur Filoche, Dominique Béréziat, and Anastase Alexandre Charantonis. Deep prior in variational assimilation to estimate ocean circulation without explicit regularization. In *Climate Informatics*, Asheville, NC, United States, May 2022.
 - [4] G. Taburet, A. Sanchez-Roman, M. Ballarotta, M.-I. Pujol, J.-F. Legeais, F. Fournier, Y. Faugere, and G. Dibarboure. DUACS DT2018 : 25 years of reprocessed sea level altimetry products. *Ocean Science*, 15(5) :1207–1224, 2019.