Statistics-free interpolation of ocean observations with deep spatio-temporal prior

Arthur Filoche, Théo Archambault, Anastase Alexandre Charantonis, Dominique Béréziat

Statistics-free interpolation of ocean observations with deep spatio-temporal prior

Arthur Filoche[1], Théo Archambault[1,2], Anastase Charantonis[3], and Dominique Béréziat[1]

[1] Sorbonne Université, CNRS, LIP6, Paris
[2] Sorbonne Université, LOCEAN, Paris
[3] ENSIIE, LOCEAN, Paris

Abstract. Interpolating sea surface height satellite measurements is a challenging inverse problem as altimeter observation can be very sparse in space and time. Operational methods rely on second-order statistics of ocean evolution which are difficult to estimate due to the high dimensionality of the studied system. In this work, we investigate a statistics-free and unsupervised variational method using a deep spatio-temporal prior, a neural network optimized on only one observational window. Results are aligned with state-of-the-art operational methods.

Keywords: deep prior · space-time interpolation · satellite altimetry, sea surface height · inverse problem

1 Introduction

Monitoring and modeling the ocean is a constant scientific preoccupation whether for global climate understanding or numerical weather prediction. To do so, information from various sensors is processed in order to estimate the state of the ocean. Surface circulation is usually a variable of great interest as it explains the transport of numerous quantities. It can partially be derived from sea surface heights which are observed thanks to altimeter satellites [5,7].

However such data are very sparse in space and time so that interpolating them leads to challenging inverse problems. Even though classical least square methods relying on second-order statistics data [3,9] have a strong operational record and are still getting better thanks to the growing number of available observations [15], deep learning techniques have revolutionized inverse problems solving [11]. But in the Earth science context, ground truth is not available, so that a supervised learning setup is not realistic.

In this work we investigated the deep prior method [17], optimizing a neural architecture on only one spatio-temporal observation of sea surface heights. We show that the designed deep prior provides efficient regularization. The code is available at the following address[4].

[4] `https://github.com/ArFiloche/MACLEAN_deep_spatiotemporal_prior`

2 Optimal interpolation of sea surface height

2.1 Observing System Simulation Experiment

The used dataset and the simulation experiment framework have been introduced in [2] and we use the pre-processing of [6]. The interest here is to estimate the full space-time trajectory of the sea surface height (SSH) variable. The considered ground truth is the result of NATL60 high-resolution ocean simulation [1] re-scaled at $(1/20)°$. We denote the 3D-volume of dimensions (T, n_x, n_y) representing a ground truth space-time trajectory $\mathbf{X}$, an example is displayed in Fig. 1.

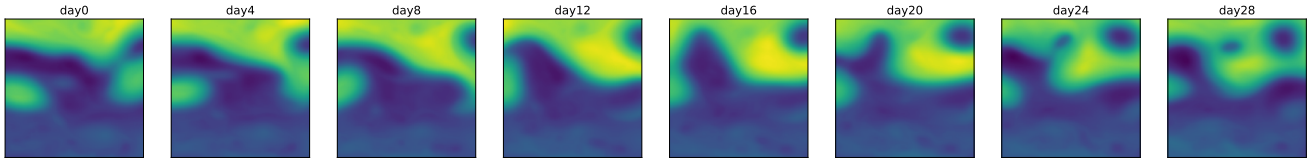

Fig. 1. Example of reference sea surface height trajectory

The observation operator used to create the dataset aims at simulating two satellite sources. The first is a constellation of 4 nadir altimeters [5] with small spatio-temporal coverage. The second is from a wide-swath altimeter, replication the Surface Water and Ocean Topography (SWOT) upcoming mission, and made possible thanks to the observation simulator introduced in [7]. Observation $\mathbf{Y}$ denoted available at regular time-steps, per daily interval and obey the following observation equation $\mathbf{Y} = \mathbb{H}\mathbf{X} + \boldsymbol{\varepsilon}$, where $\mathbb{H}$ is a linear projector associated with satellite tracks and $\boldsymbol{\varepsilon}$ a measurement noise. An example is displayed in Fig. 2, pointing to a significant sparsity in space.

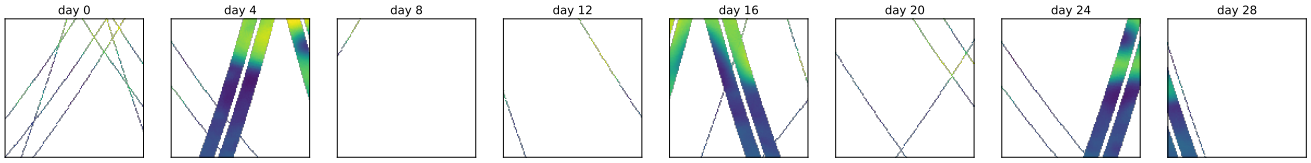

Fig. 2. Example of sea surface height satellite observation along a trajectory

2.2 DUACS analysis

The Data Unification and Altimeter Combination System (DUACS) [15] analysis is a result of a best linear unbiased estimation (BLUE) [9]. This estimation relies on the knowledge of second-order statistics, covariance matrices of state, and

noise that we denote $\mathbf{B}$ and $\mathbf{R}$, respectively. Such statistics are usually hard to estimate for a high dimensional system like the Ocean, but DUACS leverage 25 years of reprocessed sea level altimetry so that this estimation is a strong baseline. The produced estimation $\hat{\mathbf{X}}_{blue} = \mathbf{B}\mathbb{H}^T(\mathbb{H}\mathbf{B}\mathbb{H}^T + \mathbf{R})^{-1}$ can be achieved equivalently in a variational manner [12], minimizing the energy function detailed in Eq. 1 and condensed using the Mahalanobis distance in Eq. 2. This can be seen as a least-square regression with a Thikonov regularizer promoting prior knowledge in the estimation.

$$\mathcal{J}(\mathbf{X}) = (\mathbf{Y} - \mathbb{H}\mathbf{X})^T\mathbf{R}^{-1}(\mathbf{Y} - \mathbb{H}\mathbf{X}) + \mathbf{X}^T\mathbf{B}^{-1}\mathbf{X} \tag{1}$$

$$= \|\mathbf{Y} - \mathbb{H}\mathbf{X}\|^2_{\mathbf{R}} + \|\mathbf{X}\|^2_{\mathbf{B}} \tag{2}$$

2.3 Deep spatio-temporal prior

The idea behind deep prior [17] is that using a well-suited neural network to generate the solution of a variational problem can act as a handcrafted regularization, leveraging spatial and spectral bias induced by the architecture [4,14]. This means that the control parameters are shifted from the system state space to the neural network parameters space. From a practical standpoint, a generator network $g_{\boldsymbol{\theta}}$ outputs the solution from a latent state z such that $g_{\boldsymbol{\theta}}(z) = \hat{\mathbf{X}}$. In our case, we ask the network to output the spatio-temporal system state trajectory on a specified window.

$$\mathcal{J}(\boldsymbol{\theta}) = \|\mathbf{Y} - \mathbb{H} \circ g_{\boldsymbol{\theta}}(z)\|_2^2 \tag{3}$$

Architecture The global design of the used deep prior is largely inspired by generative convolutional architectures introduced in [13]. To avoid checkerboard artifacts, we replaced deconvolution operations as described in [10]. Finally, to ensure spatio-temporal coherence of the generated solution we used (2+1)D convolution [16], which is an alternative to 3D convolutions being less expensive computationally. A schematic view of the architecture is provided in Fig. 3.

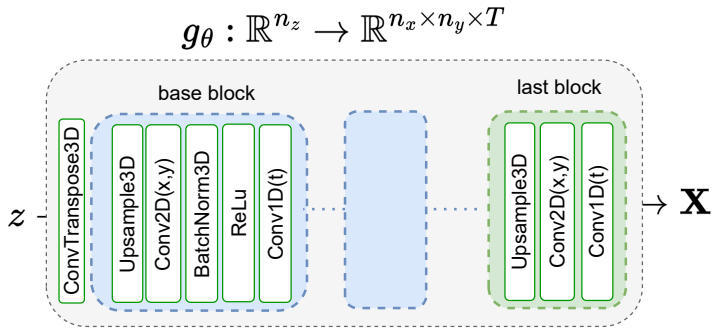

Fig. 3. Schematic view of the deep spatio-temporal prior architecture

3 Experimental results

3.1 Observational window

We tested the method on 32-day windows with 128×128 sized observation. While optimizing deep priors, we observed that reached optimum was significantly different depending on the weights initialization of the generator network. To overcome this issue, we trained multiple deep generators with different initialization and averaged their results, constituting an ensemble.

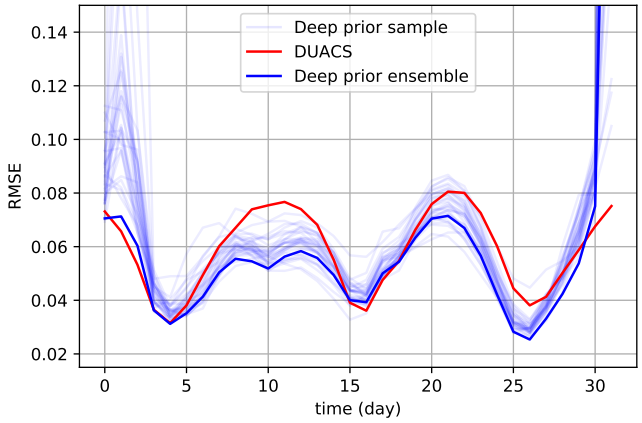

Fig. 4. RMSE comparison of optimal interpolation from DUACS and deep spatio-temporal prior on a single 32-day observational window example

In Fig. 4, estimation from DUACS and deep priors are compared using the root mean square errors (RMSE) metric. We observe that the ensemble is indeed beneficial and performs slightly better than DUACS interpolation. We also notice border effects, such that deep prior estimation deteriorates at the beginning and at the end of the temporal window. Logically, the DUACS optimal interpolation does not suffer from border effects as considered estimation where window-centered.
Looking at the error maps displayed in Fig. 5 we see that both methods have very similar spatial structures. We also notice that error maps for the DUACS optimal interpolation present checkered numerical artifacts while the deep prior ones are smoother. Our interpretation is that various biases induced by the chosen deep architecture emphasize low-frequency patterns avoiding high-frequency artifacts introduced by numerical optimization directly at the pixel level.

3.2 Year-long analysis

We also compared both methods on a year-long analysis. But training an ensemble of deep prior at each window can be computationally cumbersome. To

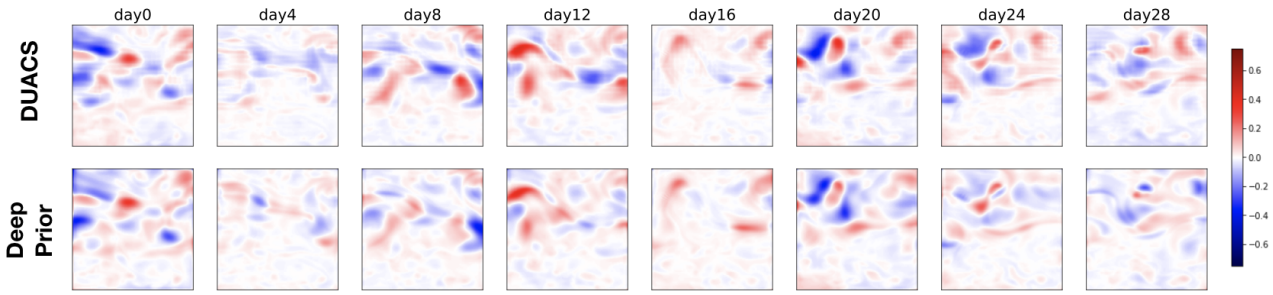

Fig. 5. Error maps of DUACS and deep prior estimation at various times in the same observational window

overcome this issue but still benefit from ensemble performances, we adopted a sliding window along the year and averaged estimation from different windows excluding border estimation. Results are displayed in Fig. 6. As for the single window experiment, RMSE scores are in the same range and slightly better with an ensemble of deep prior.

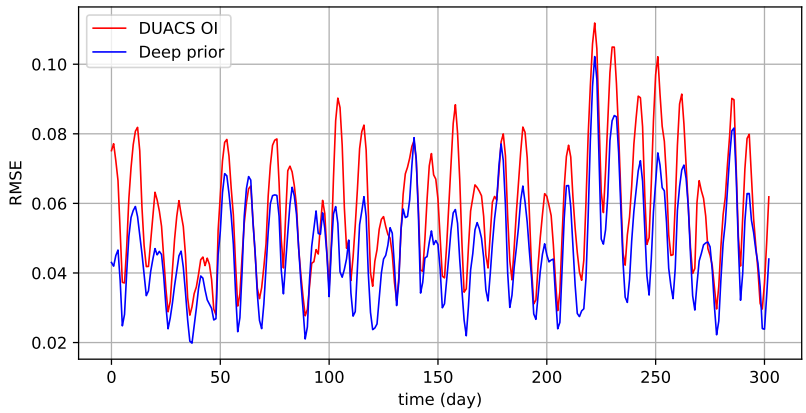

Fig. 6. RMSE comparison of optimal interpolation from DUACS and sliding-averaged deep spatio-temporal prior on a year-long period

3.3 Conv(2+1)D ablation

To justify the use of (2+1)D convolutions, we performed a similar experiment using only 2D convolutions and considering the time as channels. Results dis-

played in Fig. 7 show that such prior lacks temporal coherence and degrades performances, particularly at times where observations are very sparse.

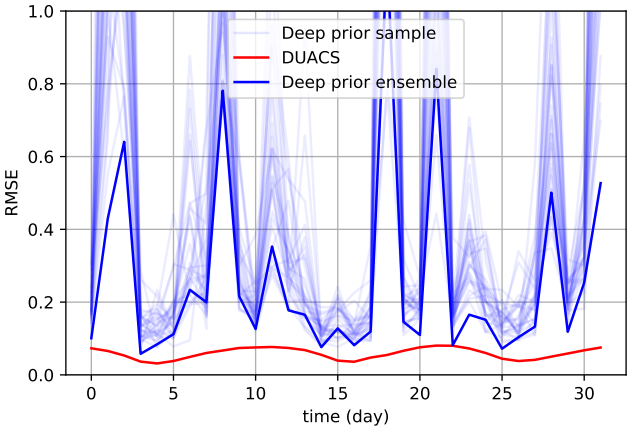

Fig. 7. RMSE comparison of optimal interpolation from DUACS and deep prior with vanilla convolutional architecture, on a single 32-day observational window example

4 Perspective

We extrapolated an idea from the image processing community to interpolate sea surface height observation from altimeter data. We highlighted in a preliminary experiment that a well-suited deep architecture has a strong regularizing effect and can substitute prior knowledge, in our case statistics of high-dimensional physical state. Finally, we give exploratory perspectives.

Automate convergence criteria. Automating the convergence when using deep prior is still an active research field, whether using an early stopping approach [18] or specific architectures [8]. If ground truth is needed to find such criteria, the method loses its appeal.

Retro-engineered the prior. As a deep prior seems to be able to replace second-order statistics, we would be interested in retrieving such statistics from a trained architecture. At the moment we did not succeed in doing so.

Refine the loss function. Observational noise statistics are usually known from measurement devices. We could use such statistics to weigh deep prior costs in a variational data assimilation fashion, for example knowing that nadir and SWOT measurement come with different noises.

References

1. A. Ajayi, J. Le Sommer, E. Chassignet, J.-M. Molines, X. Xu, A. Albert, and E. Cosme. Spatial and temporal variability of north atlantic eddy field at scale less than 100 km. *Earth and Space Science Open Archive*, page 28, 2019.
2. M. Ballarotta, E. Cosme, and A. Albert. ocean-data-challenges/2020a_SSH_mapping_NATL60: Material for SSH mapping data challenge, September 2020. `https://doi.org/10.5281/zenodo.4045400`.
3. F. Bretherton, E. Russ Davis, and C. Fandry. A technique for objective analysis and design of oceanographic experiments applied to mode-73. *Deep Sea Research and Oceanographic Abstracts*, 23(7):559–582, 1976.
4. N. Cohen and A. Shashua. Inductive bias of deep convolutional networks through pooling geometry. In *ICLR*, 2017.
5. V. Enjolras, P. Vincent, J.-C. Souyris, E. Rodriguez, L. Phalippou, and A. Cazenave. Performances study of interferometric radar altimeters: from the instrument to the global mission definition. *Sensors*, 6(3):164–192, 2006.
6. R. Fablet, M. M. Amar, Q. Febvre, M. Beauchamp, and B. Chapron. End-to-end physics-informed representation learning for satellite ocean remote sensing data: Applications to satellite altimetry and sea surface currents. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2021.
7. L. Gaultier, C. Ubelmann, and L.-L. Fu. The challenge of using future swot data for oceanic field reconstruction. *Journal of Atmospheric and Oceanic Technology*, 33(1), 2016.
8. R. Heckel and P. Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. In *ICLR*, 2019.
9. C. Henderson. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31(2):423–447, 1975.
10. A. Odena, V. Dumoulin, and C. Olah. Deconvolution and checkerboard artifacts. *Distill*, 2016.
11. G. Ongie, A. Jalal, C. Metzler, R. Baraniuk, A. Dimakis, and R. Willett. Deep learning techniques for inverse problems in imaging. *Journal on Selected Areas in Information Theory*, 1(1):39–56, 2020.
12. S. Puntanen and G. Styan. The equality of the ordinary least squares estimator and the best linear unbiased estimator. *The American Statistician*, 43(3):153–161, 1989.
13. A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *ICLR*, 2016.
14. N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville. On the spectral bias of neural networks. In *PMLR*, volume 97, pages 5301–5310, 2019.
15. 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.
16. D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri. A closer look at spatiotemporal convolutions for action recognition. In *CVPR*, 2018.
17. D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *CVPR*, June 2018.
18. H. Wang, T. Li, Z. Zhuang, T. Chen, H. Liang, and J. Sun. Early stopping for Deep Image Prior. arXiv, 2022.