End-to-End Learning of Variational Interpolation Schemes for Satellite-Derived SSH Data

Maxime Beauchamp, Mohamed Mahmoud Amar, Quentin Febvre, Ronan Fablet

END-TO-END LEARNING OF VARIATIONAL INTERPOLATION SCHEMES FOR SATELLITE-DERIVED SSH DATA

Maxime Beauchamp[1,2], Mohamed Mahmoud AMAR[1,2], Quentin Febvre[1,2], Ronan Fablet[1,2]

[1]IMT Atlantique Bretagne Pays-de-la-Loire
Technopole Brest-Iroise - CS 83818, 29238 Brest Cedex 3 (France)
[2] UMR 6285 LabSTICC, TOMS (Statistical Signal Processing and Remote Sensing)

ABSTRACT

The reconstruction of better-resolved sea surface currents is a key challenge in space oceanography. Besides the upcoming SWOT wide-swath altimeter mission, new algorithms are explore to produce improved gap-free gridded products. Based on the recent development of a generic end-to-end deep learning scheme for inverse problems backed on a variational formulation, we investigate how this framework applies to the space-time interpolation of satellite-derived SSH fields. We consider different parameterization of the proposed end-to-end learning scheme, especially regarding the embedded variational solver. Using an Observing System Simulation Experiment based on high-resolution numerical simulations in the Gulf Stream region, we show that the later may significantly outperform the state-of-the-art, including DUACS optimal interpolation product, when jointly considering nadir along-track altimeter data and upcoming SWOT wide-swath data.

Index Terms— satellite altimetry, sea surface dynamics, space-time interpolation, variational models, end-to-end learning

1. INTRODUCTION

In [1], we use a $10^\circ \times 10^\circ$ subdomain of the GULFSTREAM from the North Atlantic NATL60 high resolution deterministic ocean simulation to assess how data-driven methods might help to improve the operational DUACS Optimal Interpolation (OI). Using a so-called Observation System Simulation Experiments (OSSE), two types of pseudo altimetric observational dataset are merged: along-track nadir data for the current capabilities of the observation system and wide-swath SWOT data related to the upcoming SWOT mission. In this further development of this preliminary work, we use the same Ground Truth and observations to evaluate how the connection between 4D-Var variational data assimilation and joint learning of models and solvers bring a new gain to the previous improvements of data-driven methods investigated in [1], namely Analog Data Assimilation (AnDA) and FP-GENN, a simplified version of the new variational proposed framework. First, we briefly introduce the inverse problem related to space-time interpolation of partially observed geophysical fields. Second, we introduce the considered variational model inspired by 4D-Var data assimilation formulation and how to build an iterative update operator based on automatic differential tools to constitute a new avenue in the way of minimizing the variational cost through the end-to-end joint learning of the architectures for both NN-based representations of the dynamics and of the iterative update operator. Then, we propose an experiment based on the NATL60 OSSE to illustrate how the NN-based gradient minimization of the functional leads to better reconstructions of the SSH fields. We also provide the accompanying pytorch code for a reproducible research[1].

2. PROBLEM STATEMENT AND PROPOSED SOLUTION

Let $\mathbf{y}(\Omega) = \{\mathbf{y}_k(\Omega_k)\}$ denotes the partial and potentially noisy observational dataset corresponding to subdomain $\Omega = \{\Omega_k\} \subset \mathcal{D}$, $\overline{\Omega}$ denotes the gappy part of the field and index k refers to time t_k. Using a data assimilation state space formulation, we aim at estimating the hidden space $\mathbf{x} = \{\mathbf{x}_k(\Omega_k)\}$ from the observations $\mathbf{y}$.

2.1. Considered variational model

Considering a variational data assimilation scheme [2], the state analysis $\mathbf{x}^\star$ is obtained by solving the minimization problem:

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}} \mathcal{J}(\mathbf{x})$$

where the variational cost function $\mathcal{J}(\mathbf{x}) = \mathcal{J}_\Phi(\mathbf{x}, \mathbf{y}, \Omega)$ is generally the sum of an observation term and a regularization term involving an operator Φ which is typically a dynamical

[1]The code of the preprint is available at https://github.com/CIA-Oceanix/4DVARNN-DinAE

prior:

$$\begin{aligned}\mathcal{J}_\Phi(\mathbf{x},\mathbf{y},\Omega) &= \mathcal{J}^o(\mathbf{x},\mathbf{y},\Omega) + \mathcal{J}^b_\Phi(\mathbf{x}) \\ &= \lambda_1||\mathbf{y}-\mathcal{H}(\mathbf{x})||^2_\Omega + \lambda_2||\mathbf{x}-\Phi(\mathbf{x})||^2\end{aligned}$$

with $\mathcal{H}$ the observation operator and $\lambda_{1,2}$ are predefined or learnable scalar weights. This formulation of functional $\mathcal{J}_\Phi(\mathbf{x},\mathbf{y},\Omega)$ directly relates to strong constraint 4D-Var [3]. For inverse problems with time-related processes, the minimization of functional $\mathcal{J}_\Phi$ usually involves iterative gradient-based algorithms and in particular request to consider the adjoint method in classic model-based variational data assimilation schemes [2] where operator Φ identifies to a deterministic model $\mathbf{x}_{k+1} = \mathcal{M}(\mathbf{x}_k)$:

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} - \alpha\nabla_\mathbf{x}\mathcal{J}_\Phi(\mathbf{x}^{(i)},\mathbf{y},\Omega)$$

In our case, we are interested in purely data-driven operator Φ: we consider NN-based Gibbs-Energy (GENN) representations, a way of embedding Markovian priors in CNN [4] which proves to be efficient on SSH altimetric datasets [1]. This enables to use deep learning automatic differentiation tools: the computation of this gradient operator $\nabla_\mathbf{x}\mathcal{J}_\Phi$ given the architecture of operator Φ can be seen as a composition of operators involving tensors, convolutions and activation functions.

2.2. Trainable solver architecture

The proposed end-to-end architecture consists in embedding an iterative gradient-based solver based on the considered variational representation [5]. As inputs, we consider an observation $\mathbf{y}$, the associated observation domain Ω and some initialization $\mathbf{x}^{(0)}$. Let us denote by Γ this iterative update operator. Following meta-learning schemes [6], a residual LSTM-based representation of operator Γ is considered here where the i^{th} iterative update of the solver is given by:

$$\left\{\begin{array}{lcl} g^{(i+1)} &=& LSTM\left[\alpha\cdot\nabla_\mathbf{x}\mathcal{J}_\Phi(\mathbf{x}^{(i)},\mathbf{y},\Omega), h(i), c(i)\right] \\ x^{(i+1)} &=& x^{(i)} - \mathcal{T}\left(g^{(i+1)}\right)\end{array}\right. \quad (1)$$

with $g^{(i+1)}$ is the LSTM output using as input gradient $\nabla_\mathbf{x}\mathcal{J}_\Phi(\mathbf{x}^{(i)},\mathbf{y},\Omega)$, while $h(i)$ and $c(i)$ denotes the internal states of the LSTM [7], α is a normalization scalar and $\mathcal{T}$ a linear or convolutional mapping.
Let note that a CNN architecture could also be used instead of the LSTM representation of Γ and that when replacing both the LSTM cell by the identity operator and the minimization function $\mathcal{J}_\Phi(\mathbf{x},\mathbf{y},\Omega)$ by its single regularization term $\mathcal{J}^b_\Phi(\mathbf{x})$, the gradient-based solver simply leads to a parameter-free fixed-point version of the algorithm, the same used in [1, 4], rather similar to the DINEOF approach, see Section 3.1.

2.3. End-to-end joint learning scheme

Overall, let denote by $\Psi_{\Phi,\Gamma}(\mathbf{x}^{(0)},\mathbf{y},\Omega)$ the output of the end-to-end learning scheme given architectures for both NN-based operators Φ and Γ, see Figure 1, the initialization $\mathbf{x}^{(0)}$ of state $\mathbf{x}$ and the observations $\mathbf{y}$ on domain Ω.

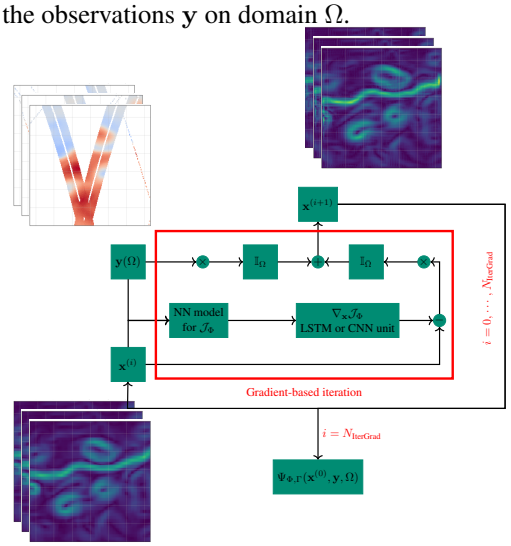


Fig. 1: Sketch of the gradient-based algorithm

Then, the joint learning of operators $\{\Phi,\Gamma\}$ is stated as the minimization of a reconstruction cost:

$$\arg\min_{\Phi,\Gamma}\mathcal{L}(\mathbf{x},\mathbf{x}^\star) \text{ s.t. } \mathbf{x}^\star = \Psi_{\Phi,\Gamma}(\mathbf{x}^{(0)},\mathbf{y},\Omega) \quad (2)$$

In case of supervised learning, where targets are gap-free, $\mathcal{L}(\mathbf{x},\mathbf{x}^\star) = ||\mathbf{x}-\mathbf{x}^\star||^2 + ||\nabla_\mathbf{x} - \nabla_{\mathbf{x}^\star}||^2$, i.e. the L2-norm of the difference between state $\mathbf{x}$ and reconstruction $\mathbf{x}^\star$ with an additional term related to the gradient of state $\mathbf{x}$. In case of unspervised learning, given the observations $\mathbf{y}$ on domain Ω and hidden state $\mathbf{x}$, the 4DVar cost function may be used $\mathcal{L}(\mathbf{x},\mathbf{x}^\star) = \lambda_1||\mathbf{y}-\mathcal{H}(\mathbf{x})||^2_\Omega + \lambda_2||\mathbf{x}-\Phi(\mathbf{x})||^2$ with weights λ_1 and λ_2 to adapt according to the reliability of the observations.

3. EXPERIMENTS

In this section, we propose an intercomparison exercise of several data-driven and learning-based approaches to help for the reconstruction of altimetric fields. As a baseline the DUACS operational processing tool based on well established optimal interpolation (OI) techniques will be considered. The other data-driven approaches used in the intercomparison are 1) AnDA [8], a purely data-driven data assimilation scheme combining a patch-based analog forecasting operator with Kalman-based ensemble data assimilation, 2) VE-DINEOF [9], an EOF-based iterative method to interpolate in space and time the missing data, and 3) (Fixed-Point) FP-GENN and Grad-GENN, the proposed end-to-end learning framework to learn jointly the NN-based representation Φ of the

dynamics coupled with a NN-based solver Γ of the targeted minimization problem. In the proposed setup, FP-GENN uses a number of 5 projections while Grad-GENN uses a number of 2 iterations. We give a detailed evaluation of the results obtained over a small region [33° N, 43° N ; -65° W, -55° W], part of the GULFSTREAM and mainly driven by energetic mesoscale dynamics. Four 20 days long validation periods are used along the one year NATL60 dataset.

3.1. Experimental setting

The Nature Run (NR) is the NATL60 high-resolution (1/60°) configuration [10] of the NEMO (Nucleus for European Modeling of the Ocean) model, covering a one year period from October 1st, 2012 to September 29th, 2013. The resolution of the nature run is downgraded to $1/20°$. Two types of altimetric datasets are combined as pseudo-observational data: their generation is fully detailed in [1]. Last, The DUACS OI [11] with 0.25° resolution is used as baseline.

3.2. Results

The daily nRMSE of AnDA, VE-DINEOF, FP-GENN and Grad-GENN are displayed on Figure 2: Grad-GENN outperforms the other methods. In particular, the gain of the iterative gradient-based solver over the fixed point algorithm used in the previous work is significant. Let note that AnDA exhibits better scores at both start and end of the four validation period, which is explained by the strong persistence, thus good analog forecasting operator, of the SSH mesoscale dynamics over the region.

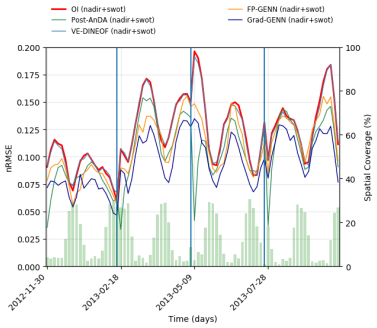


Fig. 2: Daily spatial nRMSE computed on the 80-days non-continuous validation period for OI, (post-)AnDA, VE-DINEOF, FP-ConvAE and FP-GENN. The spatial coverage of along-track nadir and wide-swath SWOT data are provided by the green-colored barplots

Table 1 displays global reconstruction score (R-score) for the known SSH field areas (Ω), the interpolation performance (I-score) for the missing data areas ($\overline{\Omega}$), and the reconstruction performance of the trained NN-based representation of the SSH dynamics for FP-GENN and Grad-GENN when applied to gap-free SSH fields (AE-score). These scores are computed for SSH (after application of a retrieving high-pass filter to keep only the small scales information) and its gradient, using only the ten days long center window of each four validation periods to exclude artificial overestimated performance of AnDA. Once again, Grad-GENN brings something new since AnDA and VE-DINEOF used to be better than FP-GENN in terms of both SSH R-scores [1] and Gradient SSH R/I-scores. Clearly, Grad-GENN motivates the use of the end-to-end learning scheme because of its plug-and-play implementation and high level of performance.

	Model type	R-score	I-score	AE-score
nadir + SWOT	OI	96.39	88.72	-
	AnDA	96.23	90.85	-
	VE-DINEOF	96.49	88.79	-
	FP-GENN	96.69	90.91	99.27
	Grad-GENN	**97.68**	**92.90**	99.42
nadir + SWOT	∇_{OI}	74.94	57.85	-
	∇_{AnDA}	80.25	64.46	-
	$\nabla_{\mathrm{VE-DINEOF}}$	81.62	55.21	-
	$\nabla_{\mathrm{FP-GENN}}$	78.90	63.56	91.27
	$\nabla_{\mathrm{Grad-GENN}}$	**86.51**	**70.18**	93.54

Table 1: SSH and SSH gradient field R/I/AE-scores computed on the four 20-days non-continuous validation period for OI, (post-)AnDA, VE-DINEOF, FP-GENN and Grad-GENN for joint assimilation/learning of along-track nadir with wide-swath SWOT data

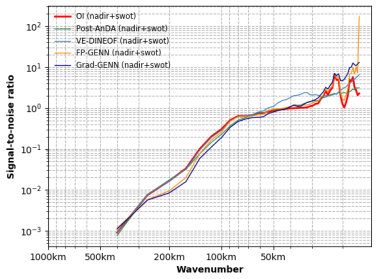
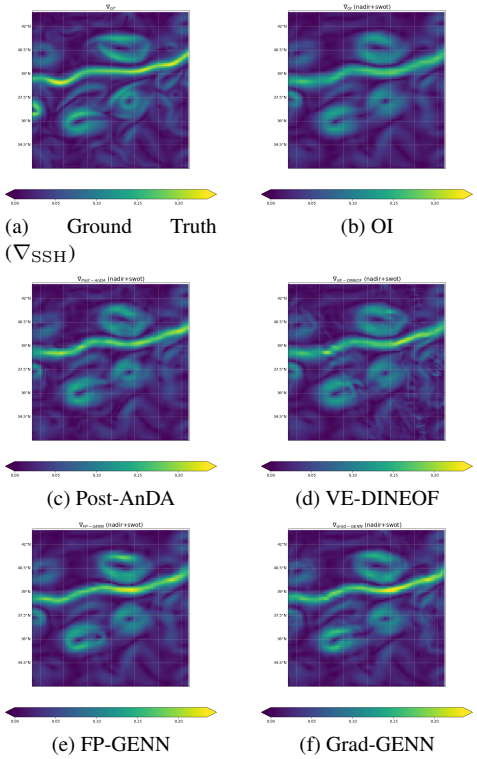


Fig. 3: Signal-to-noise ratio computed on the four 20-days non-continuous validation period for OI, (post-)AnDA, VE-DINEOF, FP-ConvAE and FP-GENN computed for joint assimilation/learning with along-track nadir and wide-swath SWOT data

Regarding the signal-to-noise score in the spectral domain (Figure 3), Grad-GENN and AnDA have a closer spectrum w.r.t the ground truth real spectrum, by catching up the sub-mesoscale range up to 70km, if a threshold of 0.5 is used. In addition, Grad-GENN seems to maintain a plateau up to 55km, which demonstrates the importance of the solver in our architecture where the simplified fixed-point version does not enable to improve the OI spectrum on this particular region.
To further enhance the vizualisation of the improvements brought by the different interpolators, Figure 4 depict the spatial SSH Gradient ground truth as well as its global reconstruction based on OI, (post-)AnDA, VE-DINEOF, FP-

GENN and Grad-GENN with joint use of along-track nadir and wide-swath pseudo-observations on August 4, 2013.



(a) Ground Truth (∇_{SSH})
(b) OI
(c) Post-AnDA
(d) VE-DINEOF
(e) FP-GENN
(f) Grad-GENN

Fig. 4: Global SSH gradient field reconstruction (August 4, 2013) obtained by OI, AnDA, VE-DINEOF, FP-GENN and Grad-GENN for a joint assimilation/learning of along-track nadir with wide-swath SWOT data

4. CONCLUSION

In this paper, we demonstrate how the new connection established between variational data assimilation and end-to-end joint learning of NN-based dynamical prior and solver leads to a better reconstruction performance of SSH fields from multi-mission altimeter datasets compared to the operational product and a selection of other data-driven methodologies.

Acknowledgements

This work was supported by LEFE program (LEFE MANU project IA-OAC), CNES (grant SWOT-DIEGO) and ANR Projects Melody and OceaniX. It benefited from HPC and GPU resources from Azure (Microsoft EU Ocean awards) and from GENCI-IDRIS (Grant 2020-101030).

5. REFERENCES

[1] Maxime Beauchamp, Ronan Fablet, Clément Ubelmann, Maxime Ballarotta, and Bertrand Chapron, "Intercomparison of data-driven and learning-based interpolations of along-track nadir and wide-swath swot altimetry observations," *Remote Sensing*, vol. 12, no. 22, 2020.

[2] M. Asch, M. Bocquet, and M. Nodet, *Data Assimilation*, Fundamentals of Algorithms. Society for Industrial and Applied Mathematics, Dec. 2016.

[3] Alberto Carrassi, Marc Bocquet, Laurent Bertino, and Geir Evensen, "Data assimilation in the geosciences: An overview of methods, issues, and perspectives," *WIREs Climate Change*, vol. 9, no. 5, pp. e535, 2018.

[4] Ronan Fablet, Lucas Drumetz, Francois Rousseau, and Maxime Beauchamp, "Joint interpolation and representation learning for irregularly-sampled satellite-derived geophysical fields," 2019.

[5] Ronan Fablet, Lucas Drumetz, and Francois Rousseau, "Joint learning of variational representations and solvers for inverse problems with partially-observed data," 2020.

[6] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas, "Learning to learn by gradient descent by gradient descent," in *Advances in neural information processing systems*, 2016, pp. 3981–3989.

[7] Leila Arras, José Arjona-Medina, Michael Widrich, Grégoire Montavon, Michael Gillhofer, Klaus-Robert Müller, Sepp Hochreiter, and Wojciech Samek, *Explaining and Interpreting LSTMs*, pp. 211–238, Springer International Publishing, Cham, 2019.

[8] R. Lguensat, P. Tandeo, P. Aillot, and R. Fablet, "The Analog Data Assimilation," *Monthly Weather Review*, 2017.

[9] B. Ping, F. Su, and Y. Meng, "An Improved DINEOF Algorithm for Filling Missing Values in Spatio-Temporal Sea Surface Temperature Data," *PLOS ONE*, vol. 11, no. 5, pp. e0155928, 2016.

[10] Jean-Marc Molines, "meom-configurations/NATL60-CJM165: NATL60 code used for CJM165 experiment," .

[11] Guillaume Taburet, Antonio Sanchez-Roman, Maxime Ballarotta, Marie-Isabelle Pujol, Jean-Franois Legeais, Florent Fournier, Yannice Faugere, and Gerald Dibarboure, "DUACS DT2018: 25 years of reprocessed sea level altimetry products," vol. 15, no. 5, pp. 1207–1224, Publisher: Copernicus GmbH.