Intercomparison of data-driven and learning-based interpolations of along-track Nadir and wide-swath SWOT altimetry observations

Maxime Beauchamp[1], Ronan Fablet[1], Clément Ubelmann[2], Maxime Ballarotta[3]
and Bertrand Chapron[4],

1. IMT Atlantique Bretagne-Pays de la Loire, Brest, France
2. Ocean Next, Grenoble, France
3. Collecte Localisation Satellites (CLS), Ramonville St-Agne, France
4. IFREMER, Brest

maxime.beauchamp@imt-atlantique.fr

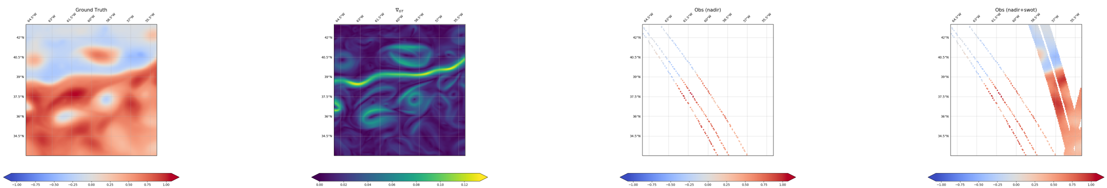
IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

2020, October 13

- Ground truth dataset $\mathbf{x}$: high-resolution 1/60° NATL60 configuration of the NEMO (Nucleus for European Modeling of the Ocean) model
- A $10^\circ \times 10^\circ$ GULFSTREAM region is used with down-graded resolution to 1/20°

GULFSTREAM domain

- OSSE : pseudo-altimetric nadir and SWOT observational datasets $\mathbf{y} = \{\mathbf{y}_k\}$ at time t_k are generated by a realistic sub-sampling satellite constellations on subdomain $\Omega = \{\Omega_k\}$ of the grid.

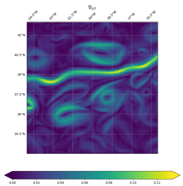


(a) *Ground Truth (SSH)* **(b)** *Ground Truth (∇_{SSH})* **(c)** *Observations (nadir)* **(d)** *Observations (nadir+swot)*

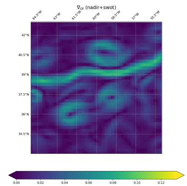
Ground Truth (SSH & ∇_{SSH}) and pseudo-observations (nadir & nadir+swot) on August 4, 2013

Methods

DUACS OI $\overline{x}$ (Taburet et al.) as a baseline : significant smooting, solving spatial scales up to 150km :



(a) *Ground Truth* (∇_{SSH})



(b) *OI* (∇_{SSH})

NATL60 & OI SSH and ∇_{SSH} *on August 4, 2013*

All the interpolations methods used here will work on the anomaly field $d\mathbf{x}$:

$$\mathbf{x} = \overline{\mathbf{x}} + d\mathbf{x} + \epsilon$$

Data-driven and learning-based approaches

- **VE-DINEOF** is a state-of-the-art interpolation approach (Ping et al., 2016) using an EOF-based iterative filling strategy. Typically the large-scale component provided by the OI is used (or 0 values if working on the anomaly) as a first guess to fill in the missing data over Ω;
- **The Analog Data Assimilation (AnDA)** (Lguensat et al., 2017) is a purely data-driven data assimilation method introducing a statistical operator $\mathcal{A}$ as a substitute for the dynamical model $\mathcal{M}$ in a classic state-space formulation;
- **Convolutional Neural Networks (CNN) :** specifically dedicated to spatio-temporal interpolation problems (Fablet et al., 2019), neural DINEOF extensions + an explicit link with variational data assimilation

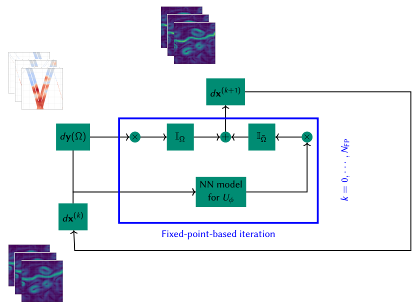
End-to-end learning

- An end-to-end learning representation has recently been introduced in Fablet et al. (2019) to deal with image sequences involving potentially large missing data rates. An energy-based representation $U_\psi = \|d\mathbf{x} - \psi(d\mathbf{x})\|^2_\Omega$ to minimize is introduced where the operator $\psi = \psi_\theta$ denotes a NN-based representation (Convolutional autoencoders **ConvAE** or Gibbs energy related NN **GENN**) of the underlying processes.
- For a specific definition of the hidden state interpolator $\widehat{d\mathbf{x}_k} = I_{U_\psi}(d\mathbf{y}_k(\Omega_k))$ based on the irregular space-time dataset $\{d\mathbf{y}_k(\Omega_k)\}$, the learning problem for optimizing parameters θ of ψ is stated as the minimization of the reconstruction error :

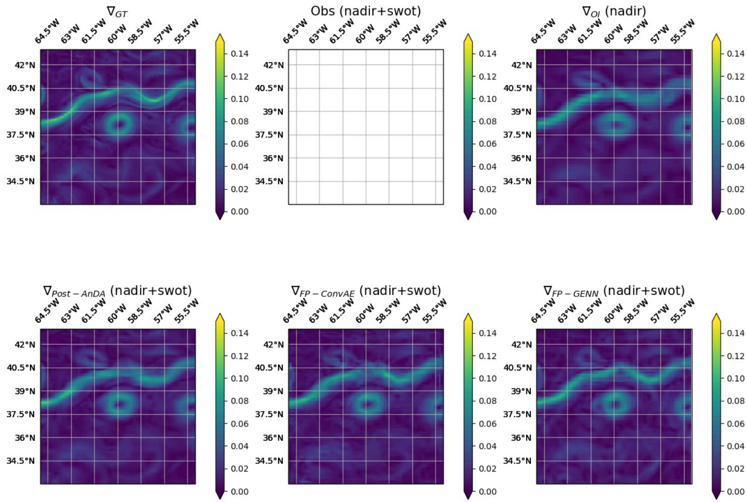
$$\widehat{\theta} = \arg\min_\theta \sum_k \left\| d\mathbf{y}_k(\Omega_k) - I_{U_\psi}(d\mathbf{y}_k(\Omega_k)) \right\|^2_{\Omega_k}$$

An iterative fixed-point (FP) solver is used to optimize parameters θ of the NN-model ψ w.r.t cost U_ψ :

$$\begin{cases} \mathbf{x}^{(i+1)} & = \psi\left(\mathbf{x}^{(i)}\right) \\ \mathbf{x}^{(i+1)}(\Omega) & = \mathbf{y}(\Omega) \\ \mathbf{x}^{(i+1)}(\overline{\Omega}) & = \mathbf{x}^{(i+1)}(\overline{\Omega}) \end{cases}$$



Sketch of the iterative fixed-point algorithm



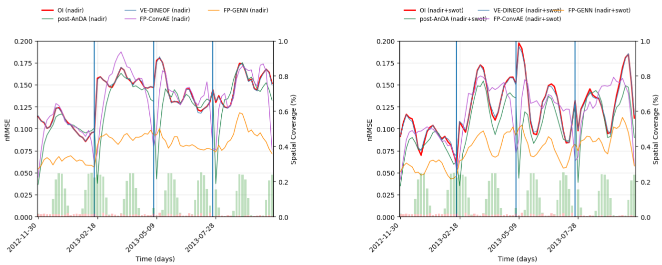
Play/Pause

Global SSH gradient field reconstruction obtained for a joint assimilation/learning of along-track nadir with wide-swath SWOT data

Scores (I)

The scores are computed on four 20-day validation periods over the one-year NATL60 daily dataset :

Up to 40% relative gain on the SSH daily root mean squared error with FP-GENN
Up to 30% relative gain when using 2D SWOT vs 1D along-track nadir



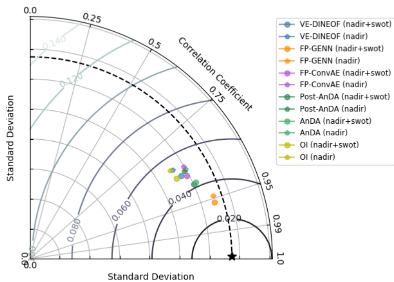
(a) *nadir* **(b)** *nadir+swot*

Daily spatial nRMSE computed on the 80-days non-continuous validation period. The spatial coverage of 0-days accumulated along-track nadir and wide-swath SWOT data are given by the red and green-colored barplots

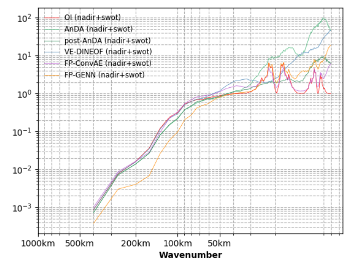
Scores (II)

Reconstruction (R-)score (over Ω) and Interpolation (I-)score (over $\overline{\Omega}$)
FP-GENN always better on I-scores

Reconstruction of the spatial scales up to 50km which is an important improvement compared to the scales that OI is handling by now



(a) *Taylor diagram*



(b) *Signal-to-noise ratio*

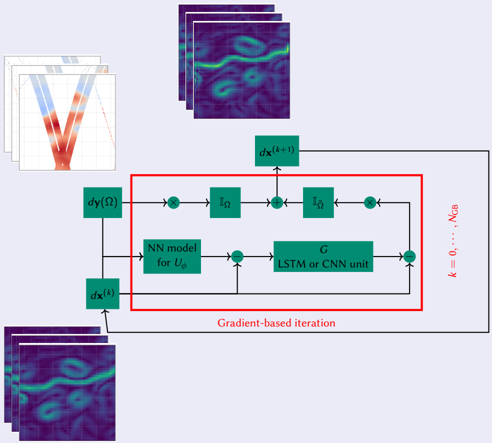
Taylor diagram and signal-to-noise ratio computed on the 80-days non-continuous validation period for a joint assimilation/learning with wide-swath SWOT data

Perspectives

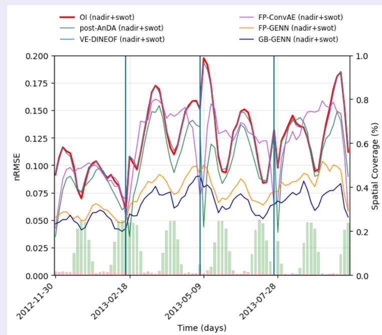
Replace the fixed-point solver by an iterative NN gradient-based descent to optimize parameters θ of the NN-model (ConvAE or GENN) ψ w.r.t cost U_ψ :

$$J_{U_\psi} = J_\psi(\mathbf{x})(\mathbf{x} - \psi(\mathbf{x})) \tag{4.1}$$

where J_{U_ψ}, the gradient of U_ψ, is finally replaced by a ConvNet or LSTM unit $G(\mathbf{x}-\psi(\mathbf{x}))$, thus enabling to solve jointly for the parametrization of ψ and G :



(a) *Sketch of the iterative gradient-based algorithm*



(b) *nRMSE*

Daily spatial nRMSE computed on the 80-days non-continuous validation period with gradient-based solver

References I

R. Fablet, L. Drumetz, and F. Rousseau. End-to-end learning of optimal interpolators for geophysical dynamics. 2019.

R. Lguensat, P. Huynh Viet, M. Sun, G. Chen, T. Fenglin, B. Chapron, and R. Fablet. Data-driven Interpolation of Sea Level Anomalies using Analog Data Assimilation. Oct. 2017. URL https://hal.archives-ouvertes.fr/hal-01609851.

B. Ping, F. Su, and Y. Meng. An Improved DINEOF Algorithm for Filling Missing Values in Spatio-Temporal Sea Surface Temperature Data. *PLOS ONE*, 11(5) :e0155928, 2016. ISSN 1932-6203. doi : 10.1371/journal.pone.0155928. URL http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155928.

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. 15(5) :1207–1224. ISSN 1812-0784. doi : https://doi.org/10.5194/os-15-1207-2019. URL https://www.ocean-sci.net/15/1207/2019/. Publisher : Copernicus GmbH.