Formulation neuronale en assimilation de données : application à l'interpolation spatio-temporelle de données océanographiques et extensions nouvelles pour l'interpolation optimale

Maxime Beauchamp[1], Ronan Fablet[1], Quentin Febvre[1]

1. IMT Atlantique Bretagne-Pays de la Loire, Brest, France

maxime.beauchamp@imt-atlantique.fr

2022, May 10

Problem statement

- $\mathbf{y}(\Omega) = \{\mathbf{y}_k(\Omega_k)\}$: the partial and potentially noisy observational dataset
- $\Omega = \{\Omega_k\} \subset \mathcal{D}$, $\overline{\Omega}$ denotes the gappy part of the field and index k refers to time t_k.

Problem

Using a data assimilation (DA) state space formulation, we aim at estimating the hidden space

$$\mathbf{x} = \{\mathbf{x}_k(\mathcal{D})\}$$

based on the available observations $\mathbf{y}$

Current solutions

- Covariance-based **Kriging** (Chilès and Delfiner, 2012), **BLUE**, **OI** (Traon et al., 1998) and **SPDE**-based version (Lindgren et al., 2011)

$$\begin{aligned} \mathbf{x}^\star &= \boldsymbol{\Sigma}_{\mathbf{xy}}\boldsymbol{\Sigma}_{\mathbf{yy}}^{-1}\mathbf{y} \\ &= -\mathbf{Q}_{\mathbf{xx}}^{-1}\mathbf{Q}_{\mathbf{xy}}\mathbf{y} \end{aligned}$$

- Model-based DA :

State space formulation $\begin{cases} \mathbf{x}_{k+1} &= \mathcal{M}_{k+1}(\mathbf{x}_k) + \boldsymbol{\eta}_k \\ \mathbf{y}_k &= \mathcal{H}_k(\mathbf{x}_k) + \varepsilon_k \end{cases}$

Sequential assimilation, (En)KF, see e.g. (Evensen, 2009)
Variational assimilation, (3DVar, 4DVar) (Asch et al., 2016)
Hybrid methods

- Data-driven DA :
  - **Analog forecasting operator** embedded in EnKF (AnDA) (Tandeo et al., 2015)
  - **Hybrid DA and machine/deep learning synergy** : inference of unresolved scale parametrizations (Brajard et al., 2021; Bocquet et al., 2019), use of DA-based parameterizations inference in numerical models involving machine learning (O'Gorman and Dwyer, 2018; Rasp et al., 2018)

Variational model (I)

Considering a variational data assimilation scheme Asch et al. (2016), 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 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 Carrassi et al. (2018).

Variational model (II)

For inverse problems with time-related processes, the minimization of functional $\mathcal{J}_\Phi$ usually involves an iterative gradient-based approach :

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

In our case, we are interested in purely data-driven operator Φ :

- we use NN-based bilinear representations, a way of embedding Markovian priors in CNN (Fablet et al., 2019; Beauchamp et al., 2020) with advection schemes.
- NN-based Φ : deep learning automatic differentiation tools to compute $\nabla_{\mathbf{x}} \mathcal{J}_\Phi$ given the architecture of operator Φ

Trainable Gradient-based (Grad) solver architecture

Let us denote by Γ the iterative gradient-based update operator. Following meta-learning schemes Andrychowicz et al. (2016), a residual LSTM-based representation of operator Γ is considered here where the i^{th} iterative update of the solver is given by :

$$\left\{\begin{array}{rcl} g^{(i+1)} & = & LSTM\left[\alpha \cdot \nabla_{\mathbf{x}} \mathcal{J}_{\Phi}(\mathbf{x}^{(i)}, \mathbf{y}, \Omega), h(i), c(i)\right] \\ x^{(i+1)} & = & x^{(i)} - \mathcal{T}\left(g^{(i+1)}\right) \end{array}\right. \quad (1.1)$$

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 Arras et al. (2019), α is a normalization scalar and $\mathcal{T}$ a linear or convolutional mapping.

Fixed-Point (FP) solver

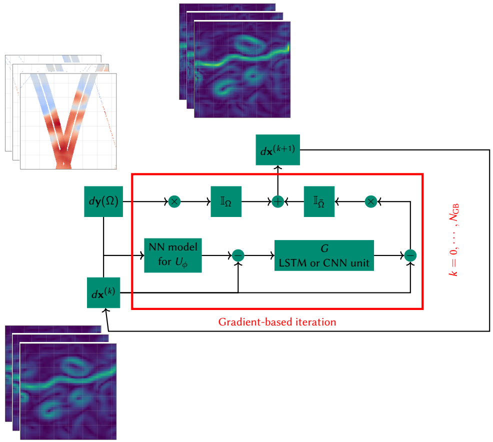
When replacing :

- the LSTM cell by the identity operator
- $\mathcal{J}_{\Phi}(\mathbf{x}, \mathbf{y}, \Omega)$ by its single regularization term $\mathcal{J}_{\Phi}^{b}(\mathbf{x})$,

the gradient-based solver leads to a parameter-free fixed-point version of the algorithm Beauchamp et al. (2020); Fablet et al. (2019).

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, the initialization $\mathbf{x}^{(0)}$ of state $\mathbf{x}$ and the observations $\mathbf{y}$ on domain Ω.



Sketch of the gradient-based algorithm

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

$$\arg\min_{\Phi,\Gamma} \mathcal{L}(\mathbf{x}, \mathbf{x}^\star) \text{ s.t. } \mathbf{x}^\star = \Psi_{\Phi,\Gamma}(\mathbf{x}^{(0)}, \mathbf{y}, \Omega) \tag{1.2}$$

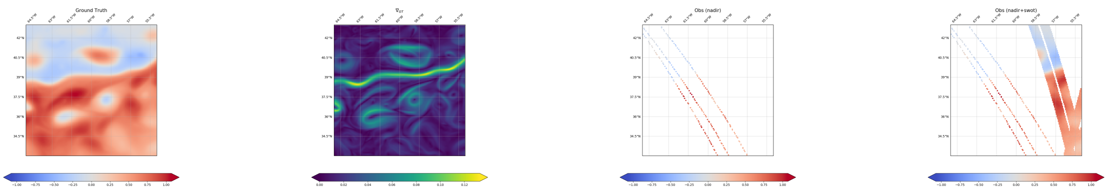
In case of supervised learning, where targets are gap-free, $\mathcal{L}(\mathbf{x}, \mathbf{x}^\star) = ||\mathbf{x} - \mathbf{x}^\star||^2 + ||\nabla_{\mathbf{x}} - \nabla_{\mathbf{x}^\star}||^2$

Observation System Simulation Experiment (I)

- Ground truth dataset $\mathbf{x}$: high-resolution $1/60^\circ$ NATL60 configuration of the NEMO (Nucleus for European Modeling of the Ocean) model
- A $10^\circ \times 10^\circ$ GULFSTREAM region is used with downgraded resolution to $1/20^\circ$, principally led by mesoscale processes
- A $10^\circ \times 8^\circ$ "open-ocean" OSMOSIS region is used with downgraded resolution to $1/20^\circ$

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.

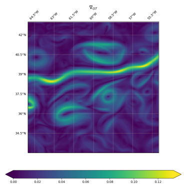


(a) *Ground Truth (SSH)* **(b)** *Ground Truth (∇_{SSH})* **(c)** *Observations (nadir)* **(d)** *Observations (nadir+swot)*

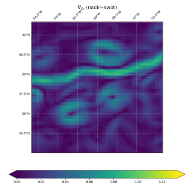
Ground Truth (SSH & ∇_{SSH}) and pseudo-observations (nadir & nadir+swot) on August 4, 2013

Observation System Simulation Experiment (II)

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



(a) *Ground Truth* (∇_{SSH})



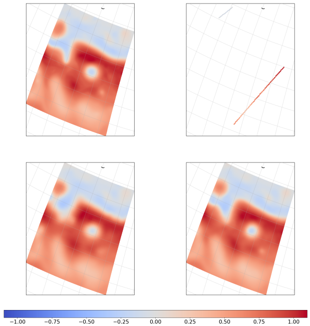
(b) *OI* (∇_{SSH})

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

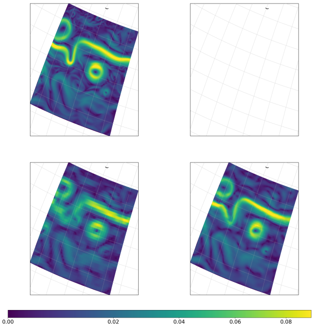
4DVarNet performance on the GULFSTREAM domain compared to DUACS OI and BFN over the period from 2012-10-22 to 2012-12-02 (42 days)

Method	μ(RMSE)	σ(RMSE)	λx (degree)	λt (days)
duacs 4 nadirs	0.92	0.01	1.42	12.0
bfn 4 nadirs	0.92	0.02	1.23	10.6
dymost 4 nadirs	0.91	0.01	1.36	11.79
miost 4 nadirs	0.93	0.01	1.35	10.19
4DVarNet 4 nadirs	**0.94**	**0.01**	**1.06**	**6.42**
duacs 1 swot + 4 nadirs	0.92	0.02	1.22	11.15
bfn 1 swot + 4 nadirs	0.93	0.02	0.8	10.09
dymost 1 swot + 4 nadirs	0.93	0.02	1.2	10.07
miost 1 swot + 4 nadirs	0.94	0.01	1.18	10.14
4DVarNet 1 swot + 4 nadirs	**0.95**	**0.01**	**0.7**	**6.34**

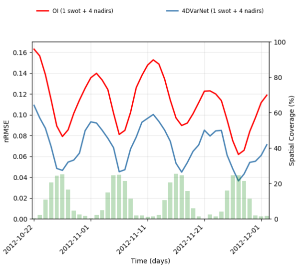
Observation System Simulation Experiment (III) : GULFSTREAM



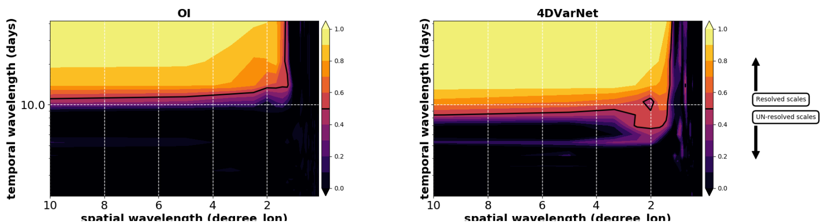
(a) *Raw SSH*



(b) *Gradient SSH*

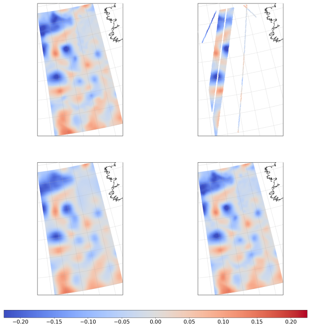


(c) *Time series*

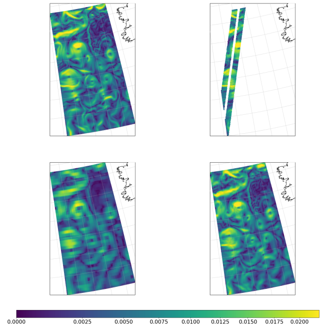


(d) *Spectral performance*

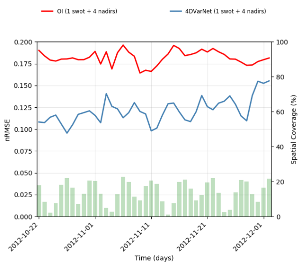
Observation System Simulation Experiment (IV) : OSMOSIS



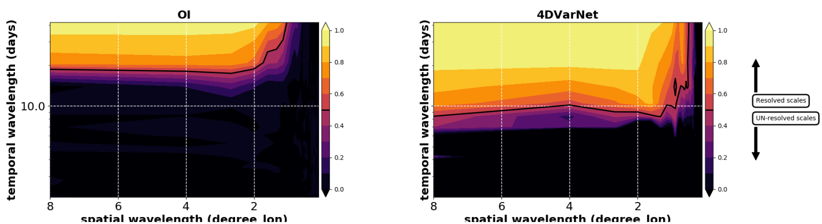
(e) *Raw SSH*



(f) *Gradient SSH*



(g) *Time series*



(h) *Spectral performance*

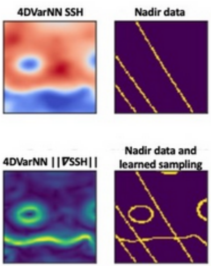
New applications

4DVarNet models with trainable observation models

$$\mathbf{x}^\star = argmin_{\mathbf{x}} ||\mathbf{x}-\mathbf{y}||^2 + \lambda ||\mathbf{x} - \Phi(\mathbf{x})||^2$$

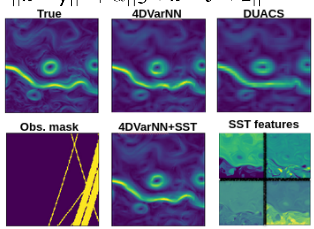
Sparse sampling operator

$$||\mathcal{H}(\mathbf{z}) * (\mathbf{x}-\mathbf{y})|| \text{ with } ||H(\mathbf{z})||_1 < \epsilon$$



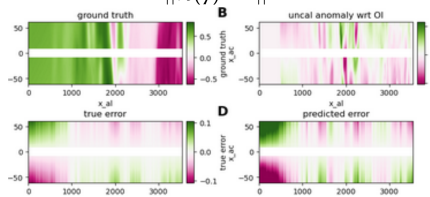
Multimodal observation

$$||\mathbf{x}-\mathbf{y}||^2 + \alpha ||\mathcal{G} * \mathbf{x} - \mathcal{F} * \mathbf{z}||^2$$



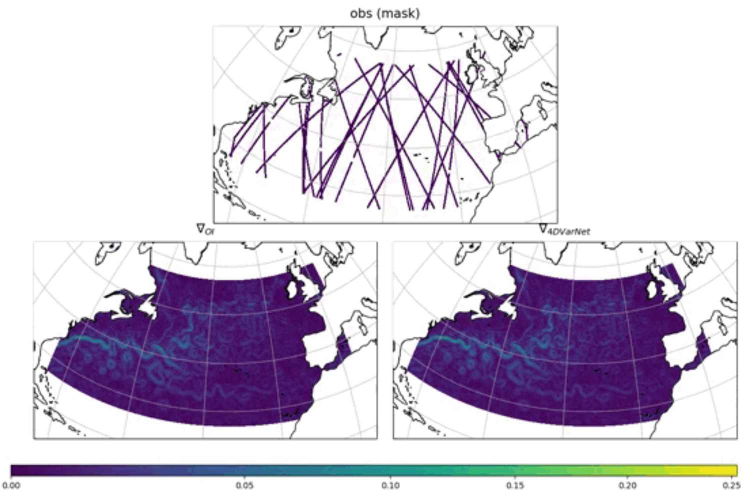
Calibration operator

$$||\mathcal{H}(\mathbf{y}) - \mathbf{x}||^2$$



Observation System Experiment (I)

- Satellite observations are from real 2017 sea surface height data from altimeter (SARAL/Altika, Jason 2, Jason 3, Sentinel 3A, Haiyang-2A and Cryosat-2 altimeter data)
- Validation on the GULFSTREAM domain for the Cryosat-2 altimeter data (not considered in the mapping)



North Atlantic SSH daily interpolation (2017)

Observation System Experiment (II)

The state space formulation includes :

- a low-resolution SSH
- a high-resolution SSH anomaly

while the observation vector includes :

- OI (for the low-resolution SSH)
- the altimetric observations
- the SST (model-based in idealistic experiments)

4DVarNet performance on the GULFSTREAM domain compared to DUACS OI and BFN over the whole year 2017

Method	μ(RMSE)	σ(RMSE)	λx (km)
DUACS	0.88	0.07	152
MIOST	0.89	0.08	139
DYMOST	0.89	0.06	129
BFN	0.88	0.06	122
4DVarNet (SSH only)	0.88	0.08	132
4DVarNet (SSH+SST)	0.88	0.06	122

The variational cost becomes :

$$\begin{aligned}\mathcal{J}_{\Phi}(\mathbf{x},\mathbf{y},\Omega) = {} & \lambda_1 ||\overline{\mathbf{y}}_{\text{SSH}} - \overline{\mathbf{x}}_{\text{SSH}}||^2 + \lambda_2 ||\mathbf{y}_{\text{SSH}} - \mathbf{x}_{\text{SSH}}||^2_{\Omega} \\ & + \sum_i \lambda_{3,i} ||\mathcal{G}_i * \mathbf{y}_{\text{SST}} - \mathcal{F}_i * \mathbf{x}_{\text{SSH}}||^2 \\ & + \lambda_4 ||\mathbf{x} - \Phi(\mathbf{x})||^2\end{aligned}$$

This may be seen as a learning-based generalisation of the SQG formulation :

$$\overline{\mathbf{x}}_{\text{SSH}} + \alpha \Delta^{-1/2} \mathbf{y}_{\text{SST}} + \varepsilon$$

where the high resolution component of the SSH is explained by a convolutional step of the SST with a fractional Laplacian operator. Here two CNN-based operators are applied to SST observations and SSH state space. They are both optimized during the training phase (with Φ and Γ) to satisfy at best the new loss function

Link with optimal interpolation

To ease the link with classic OI/SK equations, let rewrite the matrix formulation of the prior term as :

$$\begin{aligned} & \mathbf{x}^\mathsf{T}\mathbf{P}^{-1}\mathbf{x} \\ =& \mathbf{x}^\mathsf{T}\mathbf{Q}\mathbf{x} \\ =& \mathbf{x}^\mathsf{T}\mathbf{S}^\mathsf{T}\mathbf{S}\mathbf{x} \\ =& ||\mathbf{x}\mathbf{S}||^2 = ||\mathbf{x} - \Phi(\mathbf{x})||^2 \end{aligned}$$

with $\Phi = (1 - \mathbf{S})$, and $\mathbf{S}$ is the square root of the precision matrix.
When operator Φ is linear, as it is the case when discretizing the fractional operator of a linear SPDE in its matrix formulation, equating the gradient of the SK cost function

$$\mathcal{J}(\mathbf{x}) = ||\mathbf{y} - \mathbf{x}||^2_\Omega + \lambda||\mathbf{x} - \Phi(\mathbf{x})||^2$$

to zero leads to :

$$\nabla\mathcal{J}(\mathbf{x}) = (\mathbf{x} - \mathbf{y})\mathbf{1}_\Omega + \lambda\Phi^\mathsf{T}(\mathbf{x} - \Phi\mathbf{x}) = 0$$

with optimal solution :

$$\mathbf{x}^\star = \left(\mathbf{1}_\Omega + \lambda\Phi^\mathsf{T} + \lambda\Phi^\mathsf{T}\Phi\right)^{-1}\mathbf{y}$$

Link with optimal interpolation

Using for instance an SPDE-based evolution equation, we can try to :

- speed up the inversion with an LSTM solver based on a metalearning scheme for the gradient descent of the variational OI cost instead of inverting large (even if sparse) prior precision matrices
- learn the parameters of the SPDE and fill in the precision matrix
- identify the SPDE and its parameter with PINN-based approaches and use it as the evolution model in 4DVarNet

SPDE-based model

$$\frac{\partial \mathbf{x}}{\partial t} + \{\kappa^2(\mathbf{s},t) + \mathbf{m}(\mathbf{s},t)\cdot\nabla - \nabla\cdot\mathbf{H}(\mathbf{s},t)\nabla\}^{\alpha/2}\,\mathbf{x}(\mathbf{s},t) = \tau\mathbf{z}(\mathbf{s},t)$$

$\mathbf{Q_{xx}}$ given
with $\Phi = (1 - \mathbf{Q_{xx}}^{1/2})$

$\mathbf{x}^\star = -\mathbf{Q_{xx}}^{-1}\mathbf{Q_{xy}}\mathbf{y}$ LSTM solver

$\mathbf{Q_{xx}}$ structure given
κ, $\mathbf{H}$, $\mathbf{m}$ unknown
with $\Phi = (1 - \mathbf{Q_{xx}}^{1/2})$

NN-based parameter estimation

$\mathbf{x}^\star = -\mathbf{Q_{xx}}^{-1}\mathbf{Q_{xy}}\mathbf{y}$ LSTM solver

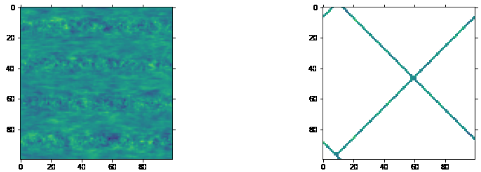
$\mathbf{\Phi}$: SPDE evolution equation

(s)PINN SPDE identification

$\mathbf{x}^\star = -\mathbf{Q_{xx}}^{-1}\mathbf{Q_{xy}}\mathbf{y}$ LSTM solver

Data challenges

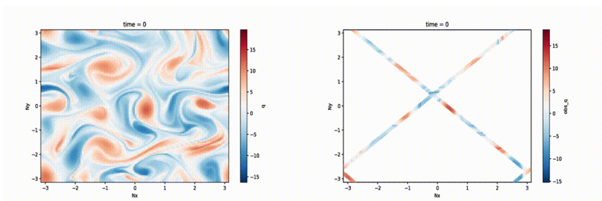
https://github.com/maxbeauchamp/2022a_SPDE_GP_mapping



(a) *simulation* **(b)** *observations*

SPDE-based GP data challenge

https://github.com/jejjohnson/2022b_qg_mapping



(a) *simulation* **(b)** *observations*

Quasi-Geostrophic data challenge

Take-home messages

Done

- We can bridge DNN and variational models to solve inverse problems
- Learning jointly variational priors, observation models and solvers
- Optimal sampling as a learning issue under sparsity constraint
- Multimodal DA as a learning issue with a trainable feature extraction operator

In progress

- **Stochastic implementation** in progress

References I

M. Andrychowicz, M. Denil, S. Gomez, M. W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, and N. De Freitas. Learning to learn by gradient descent by gradient descent. In *Advances in neural information processing systems*, pages 3981–3989, 2016.

L. Arras, J. Arjona-Medina, M. Widrich, G. Montavon, M. Gillhofer, K.-R. Müller, S. Hochreiter, and W. Samek. *Explaining and Interpreting LSTMs*, pages 211–238. Springer International Publishing, Cham, 2019. ISBN 978-3-030-28954-6. doi : 10.1007/978-3-030-28954-6_11. URL https://doi.org/10.1007/978-3-030-28954-6_11.

M. Asch, M. Bocquet, and M. Nodet. *Data Assimilation*. Fundamentals of Algorithms. Society for Industrial and Applied Mathematics, Dec. 2016. ISBN 978-1-61197-453-9. doi : 10.1137/1.9781611974546. URL https://doi.org/10.1137/1.9781611974546.

M. Beauchamp, R. Fablet, C. Ubelmann, M. Ballarotta, and B. Chapron. Intercomparison of data-driven and learning-based interpolations of along-track nadir and wide-swath swot altimetry observations. *Remote Sensing*, 12(22), 2020. ISSN 2072-4292. doi : 10.3390/rs12223806. URL https://www.mdpi.com/2072-4292/12/22/3806.

M. Bocquet, J. Brajard, A. Carrassi, and L. Bertino. Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models. *Nonlinear Processes in Geophysics*, 26(3) :143–162, 2019. doi : 10.5194/npg-26-143-2019. URL https://npg.copernicus.org/articles/26/143/2019/.

J. Brajard, A. Carrassi, M. Bocquet, and L. Bertino. Combining data assimilation and machine learning to infer unresolved scale parametrization, 2021. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2020.0086.

References II

A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen. Data assimilation in the geosciences : An overview of methods, issues, and perspectives. *WIREs Climate Change*, 9(5) :e535, 2018. doi : https://doi.org/10.1002/wcc.535. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.535`.

J. Chilès and P. Delfiner. *Geostatistics : modeling spatial uncertainty*. Wiley, New-York, second edition, 2012.

G. Evensen. *Data Assimilation*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 9783642037108 9783642037115. URL `http://link.springer.com/10.1007/978-3-642-03711-5`.

R. Fablet, L. Drumetz, F. Rousseau, and M. Beauchamp. Joint interpolation and representation learning for irregularly-sampled satellite-derived geophysical fields. 2019.

F. Lindgren, H. Rue, and J. LindstrÃ¶m. An explicit link between gaussian fields and gaussian markov random fields : the stochastic partial differential equation approach. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 73(4) :423–498, 2011. doi : 10.1111/j.1467-9868.2011.00777.x. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.00777.x`.

P. A. O'Gorman and J. G. Dwyer. Using machine learning to parameterize moist convection : Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, 10(10) :2548–2563, 2018. doi : https://doi.org/10.1029/2018MS001351. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001351`.

S. Rasp, M. S. Pritchard, and P. Gentine. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39) :9684–9689, 2018. ISSN 0027-8424. doi : 10.1073/pnas.1810286115. URL `https://www.pnas.org/content/115/39/9684`.

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.

References III

P. Tandeo, P. Ailliot, J. Ruiz, A. Hannart, B. Chapron, A. Cuzol, V. Monbet, R. Easton, and R. Fablet. Combining Analog Method and Ensemble Data Assimilation : Application to the Lorenz-63 Chaotic System. In V. Lakshmanan, E. Gilleland, A. McGovern, and M. Tingley, editors, *Machine Learning and Data Mining Approaches to Climate Science*, pages 3–12. Springer, 2015.

P.-Y. Traon, F. Nadal, and N. Ducet. An improved mapping method of multisatellite altimeter data. *Journal of Atmospheric and Oceanic Technology - J ATMOS OCEAN TECHNOL*, 15 :522–534, 04 1998. doi : 10.1175/1520-0426(1998)015⟨0522:AIMMOM⟩2.0.CO;2.