End-to-End Kalman Filter in a High Dimensional Linear Embedding of the Observations

Said Ouala, Pierre Tandeo, Bertrand Chapron, Fabrice Collard, and Ronan Fablet

Abstract Data assimilation techniques are the state-of-the-art approaches in the reconstruction of a spatio-temporal geophysical state such as the atmosphere or the ocean. These methods rely on a numerical model that fills the spatial and temporal gaps in the observational network. Unfortunately, limitations regarding the uncertainty of the state estimate may arise when considering the restriction of the data assimilation problems to a small subset of observations, as encountered for instance in ocean surface reconstruction. These limitations motivated the exploration of reconstruction techniques that do not rely on numerical models. In this context, the increasing availability of geophysical observations and model simulations motivates the exploitation of machine learning tools to tackle the reconstruction of ocean surface variables. In this work, we formulate sea surface spatio-temporal reconstruction problems as state space Bayesian smoothing problems with unknown augmented linear dynamics. The solution of the smoothing problem, given by the Kalman smoother, is written in a differentiable framework which allows, given some training data, to optimize the parameters of the state space model.

Keywords Kalman filter · Machine learning · Spatio-temporal interpolation

1 Introduction

Data assimilation in a broad sense can be considered as the inference of a hidden state, based on several sources of information. When considering data assimilation

S. Ouala (✉) · P. Tandeo · R. Fablet
IMT Atlantique, Lab-STICC, Brest, France
e-mail: said.ouala@imt-atlantique.fr

B. Chapron
Ifremer, LOPS, Plouzané, France

F. Collard
ODL, Locmaria-Plouzané, France

in the context of oceanography, these schemes exploit, in addition to some given observations, a dynamical model to perform simulations from given ocean states [1]. Unfortunately, realistic analytic parameterizations of the dynamical model, in the context of sea surface variables reconstruction, lead to computationally demanding representations [2]. Furthermore, when associated to a small subset of observations (as encountered for instance when assimilating sea surface variables with a global ocean model), these realistic models may result in modeling and inversion uncertainties. On the other hand, the analytic derivation of computationally-efficient, low-order models involves theoretical assumptions, which may not be fulfilled by real observations. These limitations motivated the exploration of interpolation techniques that do not require an explicit dynamical representation. Among other methods, Optimal Interpolation (OI) became the state-of-the-art framework [3, 4]. This technique does not need an explicit formulation of the dynamical model and rather relies on the modelization of the covariance of the spatio-temporal fields. Despite the success of OI, this technique tends to smooth the fine scale structures which motivates the development of new spatio-temporal interpolation schemes, mainly based on machine learning representations [5–10].

From the perspective of the machine learning community, state-of-the-art reconstruction techniques are usually formulated as inverse problems, where one searches to maximize the reconstruction performance of an inversion model, given the observed field as an input. Several methods were developed for this purpose in the fields of signal denoising [11, 12] and image inpainting [13] where the inversion model typically relies on a deep learning architecture. This *end-to-end* learning strategy, differs from classical inversion techniques used in geosciences, where the state-space representations (specifically the dynamical models) and the inversion schemes are a priori unrelated. The recent exploration of machine learning representations in the context of sea surface fields reconstruction was inspired by the latter methodological viewpoint, where a data-driven dynamical model is optimized based on the minimization of a forecasting cost. This data-driven prior is then plugged into a data assimilation framework to perform reconstruction based on classical (Kalman based, variational formulations) inversion schemes [7, 14, 8].

Recently, several works investigated *end-to-end* deep learning architectures in the resolution of reconstruction issues in geosciences [15–17, 10]. However, this tools, although relevant, were naturally explored in the context of image denoising and inpainting applications due to the lack of methodological formulation. When considering geosciences applications, a huge effort was carried within the geosciences community to derive reconstruction algorithms that, beyond being efficient with respect to a given metric, are robust and rely on a solid methodological formulation. From this point of view, we believe that *end-to-end* deep learning techniques should build on such methodological knowledge to propose new reconstruction solutions that can achieve both a decent performance score, and remain theoretically relevant which helps the understanding and generalization of these algorithms. From this point of view, we exploit ideas from machine learning and Bayesian filtering to propose a framework that is able to provide a relevant reconstruction of a spatio-temporal state. Specifically, we formulate a new state space model for ocean surface

observations based on an augmented linear dynamical system. Assuming that the model and observation errors are Gaussian, the solution of the filtering/smoothing problem on this new state space model is given by the Kalman filter/smoother. Inspired by deep learning architectures, the Kalman recursion is written in a differentiable framework, which allows for the derivation of the parameters of the new state-space model based on a reconstruction cost of the observations.

2 Method

Motivation Let us assume the following state-space model

$$\dot{\mathbf{x}}_t = f(\mathbf{x}_t) + \boldsymbol{\eta}_t \tag{1}$$

$$\mathbf{y}_t = \mathcal{H}_t(\mathbf{x}_t) + \boldsymbol{\epsilon}_t \tag{2}$$

where $t \in [0, +\infty]$ is time. The variables $\mathbf{x}_t \in \mathbb{R}^s$ and $\mathbf{y}_t \in \mathbb{R}^n$ represent the state variables and the observations respectively. f and $\mathcal{H}_t$ are the dynamical and observation operators. $\boldsymbol{\eta}_t$ and $\boldsymbol{\epsilon}_t$ are random processes accounting for the uncertainties. They are defined as centered Gaussian processes with covariances $\mathbf{Q}_t$ and $\mathbf{R}_t$ respectively.

In the context of geosciences, and when considering the resolution of filtering and smoothing problems using data assimilation, the dynamical and observation models f and $\mathcal{H}$, the model and observation error covariances $\mathbf{Q}_t$ and $\mathbf{R}_t$ as well as the true state $\mathbf{x}_t$ of Eqs. (1) and (2) are either unavailable or too complicated to handle. In this context, we show in this work how to exploit observations $\mathbf{y}_t$ sampled from time t_1 to time t_f to learn a Bayesian scheme that allows for reconstruction applications given new observations (i.e., at time $t > t_f$).

Definition of a New State Space Model In this work, we consider an embedding of the observations as proposed in [18]. Specifically, we project our observations (or a reduced order version of our observations) into a higher dimensional space where the dynamics of the observations are assumed to be linear. Formally, in order to derive our new state-space model, we first start by writing an augmented state $\mathbf{u}_t$ such as $\mathbf{u}_t^T = [(\mathbf{My}_t)^T, \mathbf{z}_t^T]$ with $\mathbf{z}_t \in \mathbb{R}^l$ is the unobserved component of the augmented state $\mathbf{u}_t$ and $\mathbf{M} \in \mathbb{R}^{r \times n}$ with $r \leq n$ a linear projection operator (that can be used for instance in the context of reduced order modeling). The matrix $\mathbf{M}$ is assumed to have r orthogonal lines so that the matrix $\mathbf{M}^{-1} = \mathbf{M}^T$ verifies $\mathbf{MM}^{-1} = \mathbf{I}$. We used in this work an Empirical Orthogonal Functions (EOF) projection. This constraints $\mathbf{M}$ to be a matrix of orthogonal eigenvectors of the covariance matrix of the centered data. The augmented state $\mathbf{u}_t \in \mathbb{R}^{d_E}$, with $d_E = l + r$, evolves in time according to the following state-space model:

$$\dot{\mathbf{u}}_t = \mathbf{A}_\sigma \mathbf{u}_t + \boldsymbol{\eta}_t \tag{3}$$

$$\mathbf{y}_t = \mathbf{M}^{-1}\mathbf{G}\mathbf{u}_t + \boldsymbol{\epsilon}_t \tag{4}$$

where the dynamical operator $\mathbf{A}_\sigma$ is a $d_E \times d_E$ matrix with coefficients σ. $\mathbf{G}$ is a projection matrix that satisfies $\mathbf{M}\mathbf{y}_t = \mathbf{G}\mathbf{u}_t$. The eigenvalues of the matrix $\mathbf{A}_\sigma$ encode the decaying and oscillating modes of the dynamics that are learned from data. Furthermore, the matrix $\mathbf{A}_\sigma$ can be constrained to be skew-symmetric (simply by imposing $\mathbf{A}_\sigma = 0.5(\mathbf{B}_\sigma - \mathbf{B}_\sigma^T)$ with $\mathbf{B}_\sigma$ a trainable matrix) so the solution of (3) will be written as a weighted sum of $d_E/2$ trainable oscillations, where the corresponding frequencies been encoded in the imaginary parts of the eigenvalues of $\mathbf{A}_\sigma$. This formulation is highly suitable for Hamiltonian (conservative) dynamical systems since the energy of the system is conserved and allows guaranteeing long term boundedness of the model. Furthermore, this formulation differs fundamentally from classical Auto Regressive (AR) models written in the space of the observations. Indeed, simple AR models only have a number of $r < d_E$ eigenvalues, which limits their expressivity.

It is worth noting that this formulation closely relates to the Koopman operator [19] where the augmented state $\mathbf{u}_t$ can be seen as a finite dimensional approximation of the infinite dimensional Hilbert space of measurements of the hidden state $\mathbf{x}_t$. This model takes advantage of a linear formulation of the dynamics in a space of observables, where the resulting model is perfectly linear for a category of dynamical regimes (typically periodic and quasi-periodic ones), and can provide a decent short-term approximation of chaotic regimes. It can also be seen as a generalization of the Dynamic Mode Decomposition (DMD) method, in which $\mathbf{u}_t = \mathbf{M}\mathbf{y}_t$.

Model and Observations Error Covariances The model and observation errors $\boldsymbol{\eta}_t$ and $\boldsymbol{\epsilon}_t$ are assumed to follow Gaussian distributions with zero mean and covariance matrices $\mathbf{Q}_{\lambda,t}$ and $\mathbf{R}_{\phi,t}$, respectively. These covariance models can be parameterized as neural networks with parameter vectors λ and ϕ.

Smoothing Scheme A Kalman smoother, based on the above state-space model, is written in a differentiable framework. The idea is to derive an analytical solution of the posterior distribution $p(\mathbf{u}_t|\mathbf{y}_{t_1:t_f})$, based on the Kalman recursion. Formally, given a regular time discretization $t \in [t_1, \ldots, t_N]$ where N is a positive integer and given the initial moments $\mathbf{u}_{t_1}^a$ and $\mathbf{P}_{t_1}^a$, the mean $\mathbf{u}^s$ and covariance $\mathbf{P}^s$ of the posterior distribution $p(\mathbf{u}_t|\mathbf{y}_{t_1:t_f})$ can be computed as follows:

$$\mathbf{u}_{t+1}^f = \mathbf{F}\mathbf{u}_t^a \tag{5}$$

$$\mathbf{P}_{t+1}^f = \mathbf{F}\mathbf{P}_t^a\mathbf{F}^T + \mathbf{Q}_{\lambda,t} \tag{6}$$

$$\mathbf{K}_{t+1} = \mathbf{P}_{t+1}^f\mathbf{H}^T[\mathbf{H}\mathbf{P}_{t+1}^f(\mathbf{H})^T + \mathbf{R}_{\phi,t}]^{-1} \tag{7}$$

$$\mathbf{u}_{t+1}^a = \mathbf{u}_{t+1}^f + \mathbf{K}_{t+1}[\mathbf{y}_{t+1} - \mathbf{H}\mathbf{u}_{t+1}^f] \tag{8}$$

$$\mathbf{P}_{t+1}^a = \mathbf{P}_{t+1}^f - \mathbf{K}_{t+1}\mathbf{H}\mathbf{P}_{t+1}^f \tag{9}$$

$$\mathbf{K}_{t+1}^s = \mathbf{P}_{t+1}^a \mathbf{F}^T (\mathbf{P}_{t+2}^f)^{-1} \tag{10}$$

$$\mathbf{u}_{t+1}^s = \mathbf{u}_{t+1}^a + \mathbf{K}_{t+1}^s [\mathbf{u}_{t+1}^s - \mathbf{u}_{t+1}^f] \tag{11}$$

$$\mathbf{P}_{t+1}^s = \mathbf{P}_{t+1}^a - \mathbf{K}_{t+1}^s (\mathbf{P}_{t+1}^f - \mathbf{P}_{t+2}^s)(\mathbf{K}_{t+1}^s)^T \tag{12}$$

where $\mathbf{F} = e^{dt\mathbf{A}_\sigma}$ with dt the prediction time step and $\mathbf{H} = \mathbf{M}^{-1}\mathbf{G}$. The smoothing (Eqs. (10), (11) and (12)) is carried backward in time with $\mathbf{P}_{t_f}^s = \mathbf{P}_{t_f}^a$ and $\mathbf{u}_{t_f}^s = \mathbf{u}_{t_f}^a$.

Learning Scheme The tuning of the trainable parameters vector $\boldsymbol{\theta} = [\sigma, \lambda, \phi]^T$ is carried using the following loss function: $\hat{\theta} = \arg\min_{\theta}\{\gamma_1\mathcal{L}_1 + \gamma_2\mathcal{L}_2\}$ where $\mathcal{L}_1 = \sum_{t=t_0}^{t_N} \|\mathbf{y}_t - \mathbf{H}\mathbf{u}_t^s\|^2$ and $\mathcal{L}_2 = \frac{1}{2}\log(|\mathbf{H}\mathbf{P}_{t+1}^f\mathbf{H}^T + \mathbf{R}_{\phi,t}|)$ $+ \frac{1}{2}\sum_{t=1}^{t=t_N} \|\mathbf{y}_t - \mathbf{H}\mathbf{u}_t^f\|_{\mathbf{H}\mathbf{P}_{t+1}^f\mathbf{H}^T + \mathbf{R}_{\phi,t}}^2$ and γ_1 and γ_2 are weighting parameters. The first term $\mathcal{L}_1$ is simply the quadratic reconstruction error of the observation. The minimization of this error helps to recover an initial guess of the trainable parameters. The second term, $\mathcal{L}_2$ is the negative log likelihood of the observations. This likelihood is derived from the likelihood of the innovation, i.e. $p(\mathbf{y}_{1:T}) = \prod_{t=1}^{t=T} p(\mathbf{y_t}|\mathbf{y_{t-1}})$ [20].

3   Numerical Experiments

3.1   Preliminary Analysis on SST Anomaly Data

As an illustration of the proposed framework, we consider scalar measurements of the anomaly of the Sea Surface Temperature (SST) in the Mediterranean Sea (8.6°N and 43.8°E). The data are computed based on of the annual 99th percentile of Sea Surface Temperature (SST) from model data [21]. The time series consists of daily measurements of the SST anomaly from 1987 to 2019. The training data is composed of a sparse sampling of the original time series, as highlighted in Fig. 1a. The proposed framework is tested with the following configuration: The augmented state space model is built with $\mathbf{M} = I_1$, and $\mathbf{z} \in \mathbb{R}^5$. The model error covariance is a constant matrix of size, $d_E \times d_E$ and the observation error covariance is a scalar parameter that corresponds to the variance of the SST anomaly measurement error. Finally, the training is carried with $\gamma_1 = 0$ and $\gamma_2 = 1$.

Figure 1b highlights the reconstruction performance of the smoothing Probability Density Function (PDF) with respect to the true (unobserved) state. Interestingly, and despite the fact that the observations used to train the parameters of the Kalman filtering scheme were extremely sparse, the proposed framework is able to catch the correct underlying frequencies. Furthermore, the coverage probability of the PDF highlights the effectiveness of the estimated model and observations error covariances.

Fig. 1 Performance of the proposed framework in the reconstruction of the smoothing PDF of SST anomaly data. (**a**) Sparse training data (**b**) Reconstructed smoothing PDF on the test set (We only visualize the standard deviation of the SST anomaly measurements)
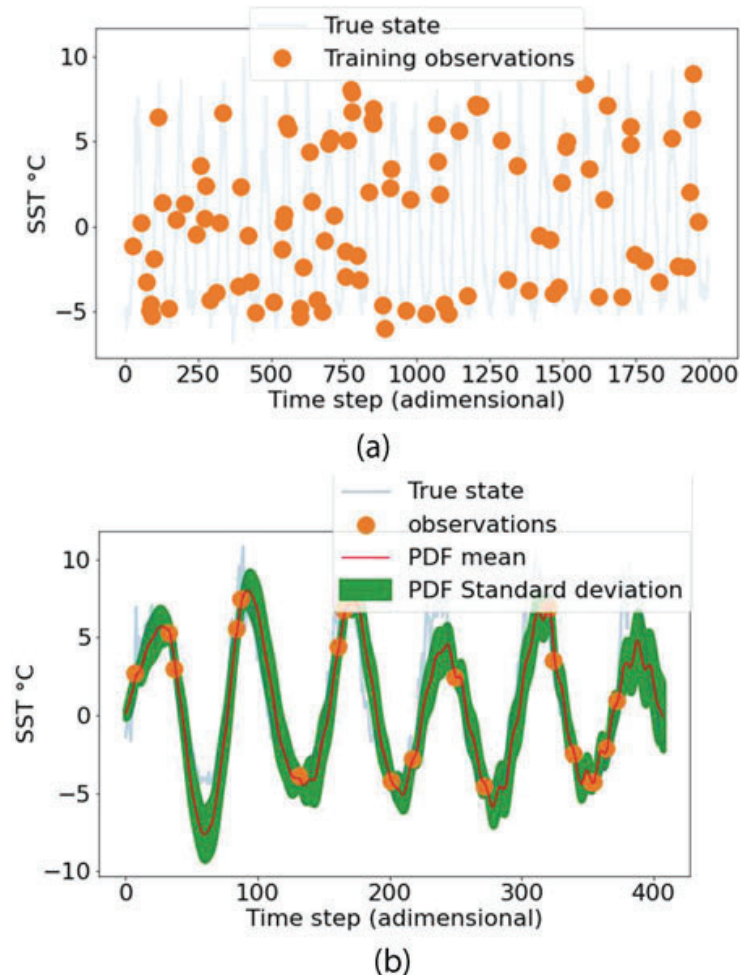

3.2 Shallow Water Equation (SWE) Case-Study

Dataset Description We consider the SWE without wind stress and bottom friction. The momentum equations are taken to be linear, and the continuity equation is solved in its nonlinear form. The direct numerical simulation is carried using a finite difference method. The size of the domain is set to $1000 \, \text{km} \times 1000 \, \text{km}$ with a corresponding regular discretization of 80×80. The temporal step size was set to satisfy the Courant–Friedrichs–Lewy condition ($h = 40.41 \, \text{s}$). The data were subsampled to $h = 40.41 \times 10$ and 500 time-steps were used as training data. The models were validated on a series of length 100. As observations, we randomly sample 1% of the pixels with a temporal coverage given in Fig. 2.

Parametrization of the Data-Driven Models The application of the above framework in the spatio-temporal reconstruction of sea surface fields should be considered with care to account for the underlying dimensionality. In this context, and following several related works [14, 9], a patch based representations is considered in order to reduce the computational complexity of the model. Specifically, this patch based representations allows a block diagonal modelization of the covariance
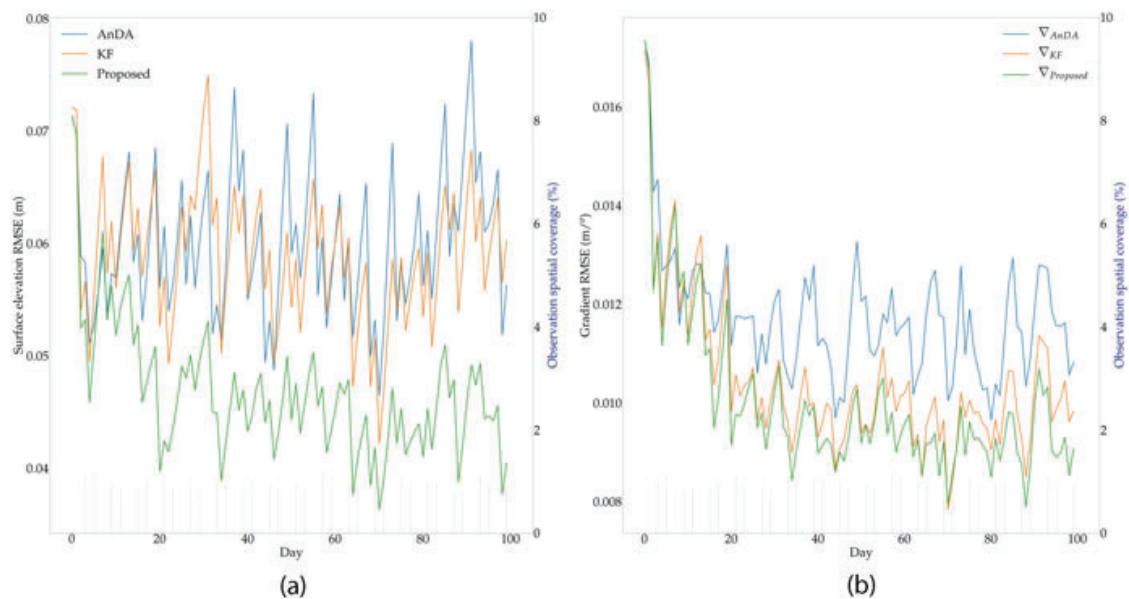
Fig. 2 Daily performance time series: we report the reconstruction performance of the sea surface elevation and its gradient in (**a**) and (**b**) respectively

matrices, which significantly reduces the computational and memory complexity of the model. This patch-based representation is fully embedded in the considered architecture to make explicit both the extraction of the patches from a 2D field and the reconstruction of a 2D field from the collection of patches. The latter involves a reconstruction operator $\mathcal{F}_r$ which is learned from data.

This patch-level representation is carried with a fixed shape of 35×35 pixels and a 10 pixels overlap between neighboring patches, resulting in a total of 16 overlapping patches. For each patch $\mathcal{P}_i$, $i = 1, \ldots, 16$ we learn an EOF basis $\mathbf{M}_{\mathcal{P}_i}$ from the training data. We keep the first 20 EOF components, which amount on average to 95% of the total variance. This patch-based decomposition is shared among all the tested models. The end-to-end Kalman filter architecture (E2EKF) is applied on a patch level with an augmented linear model operating on an embedding of dimension $d_E = 60$. The reconstructed patches are combined through the reconstruction model $\mathcal{F}_r$. This model is implemented as a residual, two blocks, convolutional neural network. The first block of the network contains four layers with 6 filters of size $k \times k$ (with k ranging from 3 to 17). The second block involves 5 layers, the first four containing 24 filters and a similar kernel size distribution as the ones in the first block, the last layer is a linear convolution with a single filter.

The proposed technique is compared in this work to the following schemes:

– Data-driven plug-and-play Kalman filter (KF): In order to show the relevance of the proposed end-to-end architecture, its plug-and-play counterpart is also tested. This model exploits the same patch based augmented linear formulation as the end-to-end one, however, the parameters of the dynamical model are trained based on a forecasting criterion and plugged into a Kalman filtering scheme.

Table 1 Surface elevation (η) interpolation experiment: reconstruction correlation coefficient and root mean squared error (RMSE) over the elevation time series and their gradient. Bold values denote smallest RMSE and highest percentage correlation

| Model	Entire map				Missing data areas			
	RMSE		Correlation		RMSE		Correlation	
	$\eta(m)$	$\nabla\eta(m/°)$	η	$\nabla\eta$	$\eta(m)$	$\nabla\eta(m/°)$	η	$\nabla\eta$
Proposed, E2EKF	**0.046**	**0.009**	**73.10%**	**41.89%**	**0.047**	**0.010**	**73.80%**	**41.90%**
AnDA	0.058	0.011	52.74%	35.91%	0.060	0.011	52.82%	21.25%
KF	0.060	0.010	64.57%	21.21%	0.059	0.010	64.68%	36.06%

– Analog data assimilation (AnDA): We apply the analog data assimilation framework [14, 7] with a locally linear dynamical kernel and an ensemble Kalman filter scheme. Please refer to [14, 7] for a detailed description of this data-driven approach, which relies on nearest-neighbor regression techniques.

Following [14], an EOF based post-processing step is applied to all the reconstructions. Furthermore, in this experiment, we only report the reconstruction performance of the mean component as a relevant benchmark of the uncertainty of the above data-driven models would be out of the scope of this paper. Thus, the model and observation error covariances are assumed to be known matrices with appropriate dimensions, and the training of the proposed model is carried with $\gamma_1 = 1$ and $\gamma_2 = 0$.

Reconstructing Performance of the Proposed Data-Driven Models A quantitative analysis of the benchmark is given in Table 1 based on (i) a mean RMSE criterion and (ii) a mean correlation coefficient criterion of the interpolated fields as well as their gradients. The RMSE and correlation coefficient time series, as well as the spatial coverage of the observations are also reported in Fig. 2. Overall, the proposed end-to-end architecture leads to very significant improvements with respect to the state-of-the-art AnDA technique, as well as to its plug-and-play counterpart both in terms of RMSE and correlation coefficients. These results emphasize the importance of the end-to-end methodology with respect to classical plug-and-play techniques since, when considering data-assimilation applications, and as shown by [16, 10], the reconstruction performance depends, in addition to the quality of the dynamical prior, on the provided measurements and their sampling. Classical plug-and-play techniques, in the opposite to end-to-end strategies, ignore the latter source of information which explains the performance of our framework.

Qualitative Analysis of the Proposed Schemes the conclusions of the quantitative analysis are also illustrated through the visual analysis of the reconstructed surface elevation and its gradient in Fig. 3. Interestingly, this visual analysis reveals that the AnDA technique tend to smooth out fine-scale patterns. By contrast, the Kalman filter based schemes (in both its end-to-end and plug and play versions) achieve a better reproduction of fine scale structures, illustrated for instance by the gradients
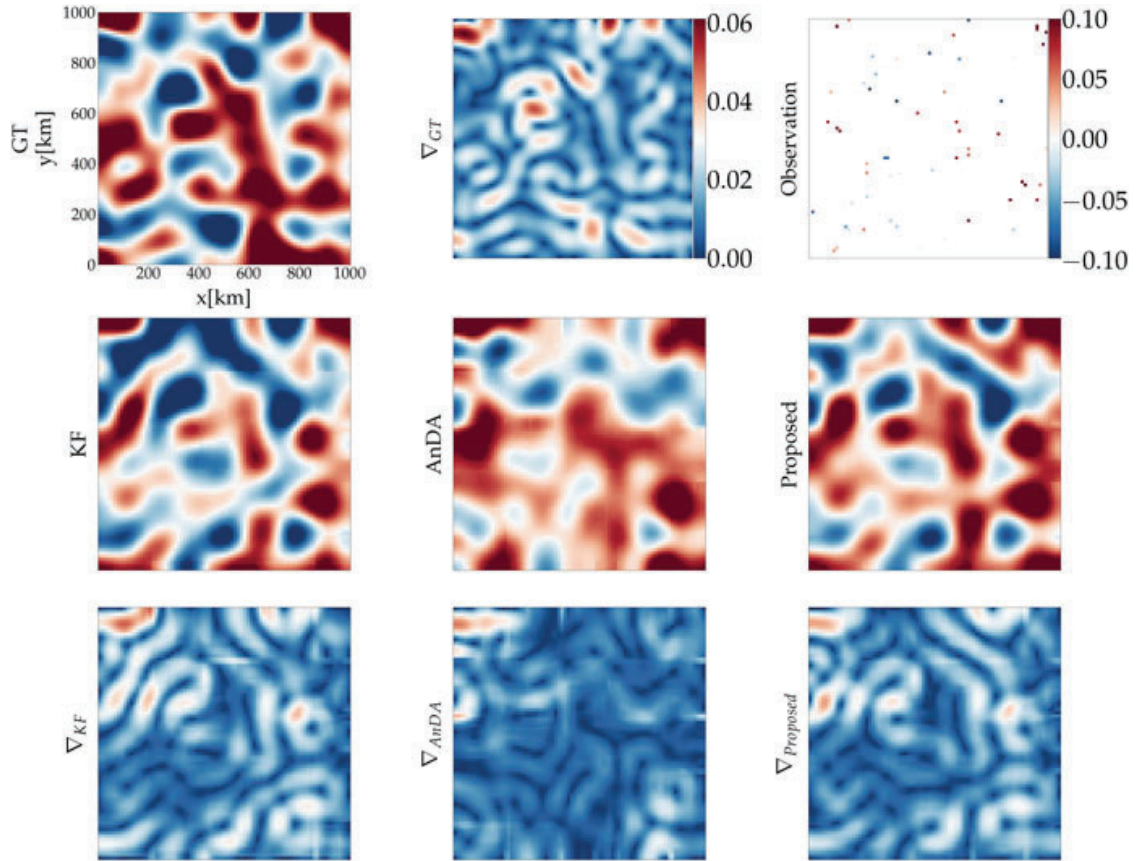
Fig. 3 Interpolation example of the surface elevation field: first row, the reference surface elevation, its gradient and the observation with missing data; second row, interpolation results using respectively the plug-and-play Augmented Koopman Kalman filter, AnDA, and the proposed E2EKF; third row, gradient of the reconstructed fields

of the field. The analysis of the spectral signatures in Fig. 4 leads to similar conclusions since, when compared to the state-of-the-art AnDA technique, as well as to its plug and play counterpart, the proposed end-to-end architecture leads to significant improvements especially regarding the reproduction of the gradient energy-level.

4   Conclusion

Spatio-temporal interpolation applications are important in the context of ocean surface modeling. For this reason, deriving new data assimilation architectures that can perfectly exploit the observations and the current advances in signal processing, modeling and artificial intelligence is crucial. In this context, this work investigated the ability of augmented linear state space models in solving smoothing issues of ocean surface observations using the Kalman filter.
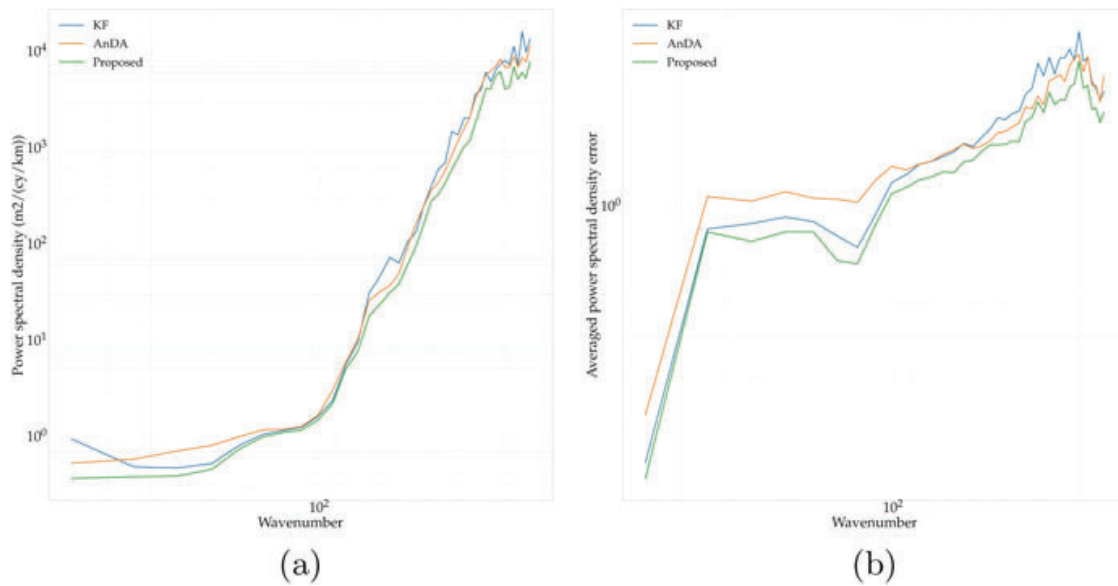
Fig. 4 Spectral comparison of the tested models: the averaged power spectral densities and their error with respect to the ground truth are given in (**a**) and (**b**) respectively

Beyond filtering and smoothing applications, we believe that the proposed framework provides an initial playground for learning approximate linear state space models of real observations. Given a sequence of sparse observations, the proposed framework may be able to unfold large scale frequencies that are useful for prediction. Interesting case studies include sea level rise and the increase of the anomaly of the sea surface temperature.

References

1. C. Gordon, C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 16(2–3):147–168, feb 2000.
2. van Leeuwen P. J. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136(653):1991–1999, dec 2010.
3. L. Gandin. Objective analysis of meteorological fields. 1963.
4. F Bouttier and P Courtier. Data assimilation concepts and methods march 1999. *Meteorological training course lecture series. ECMWF*, 718:59, 2002.
5. Aïda ALVERA-AZCÁRATE, Alexander Barth, Damien Sirjacobs, Fabian Lenartz, and Jean-Marie Beckers. Data interpolating empirical orthogonal functions (dineof): a tool for geophysical data analyses. *Mediterranean Marine Science*, 12(3):5–11, 2011.
6. Bo Ping, Fenzhen Su, and Yunshan Meng. An Improved DINEOF Algorithm for Filling Missing Values in Spatio-Temporal Sea Surface Temperature Data. *PLOS ONE*, 11(5):e0155928, may 2016.
7. Redouane Lguensat, Pierre Tandeo, Pierre Ailliot, Manuel Pulido, and Ronan Fablet. The Analog Data Assimilation. *Monthly Weather Review*, aug 2017.

8. Said OUALA, Cédric Herzet, and Ronan Fablet. Sea surface temperature prediction and reconstruction using patch-level neural network representations. In *IGARSS*, Italy, 2018. IEEE.

9. Said Ouala, Ronan Fablet, Cédric Herzet, Bertrand Chapron, Ananda Pascual, Fabrice Collard, and Lucile Gaultier. Neural network based kalman filters for the spatio-temporal interpolation of satellite-derived sea surface temperature. *Remote Sens.*, 10(12):1864, Nov 2018.

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

11. Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017.

12. Chunwei Tian, Yong Xu, and Wangmeng Zuo. Image denoising using deep cnn with batch renormalization. *Neural Networks*, 121:461–473, 2020.

13. Zhen Qin, Qingliang Zeng, Yixin Zong, and Fan Xu. Image inpainting based on deep learning: A review. *Displays*, page 102028, 2021.

14. R. Fablet, P. H. Viet, and R. Lguensat. Data-Driven Models for the Spatio-Temporal Interpolation of Satellite-Derived SST Fields. *IEEE Transactions on Computational Imaging*, 3(4):647–657, dec 2017.

15. Qiang Zhang, Qiangqiang Yuan, Chao Zeng, Xinghua Li, and Yancong Wei. Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8):4274–4288, 2018.

16. Ronan Fablet, Lucas Drumetz, and François Rousseau. End-to-end learning of energy-based representations for irregularly-sampled signals and images, 2019.

17. Bo Ping, Fenzhen Su, Xingxing Han, and Yunshan Meng. Applications of deep learning-based super-resolution for sea surface temperature reconstruction. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:887–896, 2020.

18. S Ouala, D Nguyen, L Drumetz, B Chapron, A Pascual, F Collard, L Gaultier, and R Fablet. Learning latent dynamics for partially observed chaotic systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(10):103121, 2020.

19. Bernard O Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the national academy of sciences of the united states of america*, 17(5):315, 1931.

20. Alberto Carrassi, Marc Bocquet, Alexis Hannart, and Michael Ghil. Estimating model evidence using data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143(703):866–880, 2017.

21. Karina von Schuckmann et al. Copernicus marine service ocean state report, issue 3. *Journal of Operational Oceanography*, 12(sup1):S1–S123, 2019.