Sea Surface Dynamics Reconstruction Using Neural Networks Based Kalman Filter

Ouala Said [1], Fablet Ronan [1], Herzet Cedric [1, 2], Drumetz Lucas [1], Chapron Bertrand [3], Pascual Ananda [4], Collard Fabrice [5], Gaultier Lucile [5]

[1] IMT Atlantique; Lab-STICC, Brest, France
[2] INRIA Bretagne-Atlantique, Fluminance, Rennes, France
[3] Ifremer, LOPS, Brest, France
[4] IMEDEA, UIB-CSIC, Esporles, Spain
[5] ODL, Brest, France

Abstract :

In this work, we propose an alternative to the Ensemble Kalman filter through the implementation of a neural networks filtering scheme based on a parametric stochastic model. From our numerical experiment, we prove the relevance of the proposed architecture in the reconstruction of geophysical fields with respect to the state-of-the-art schemes.

Keywords : Neural networks, Kalman filter, Stochastic models

1. INTRODUCTION

High resolution monitoring of the sea surface geophysical parameters is one of the major challenges in oceanography. Producing high resolution gridded spatio-temporal products of physical variables such as sea surface temperature, sea surface height and sea surface salinity is of key interest for several scientific fields [1].

Observations of these geophysical tracers are provided by satellite remote sensing observations and in-situ monitoring. However, they usually involve irregular sampling patterns due to the sensor's characteristics and we are only given partial and possibly noisy observations. As a result, no sensor can provide high-resolution (in space and time) gap-free observations. Missing data rates may become very large which makes crucial the development of spatio-temporal interpolation tools for end-users.

Optimal interpolation (OI) is the classic technique used by operational products [2]. It relies on the modelization of the covariance of the spatio-temporal fields. In general, stationary covariance hypotheses are considered which prove relevant

for the reconstruction of horizontal scales above 100km. Fine scale components on the other hand may hardly be retrieved with such approaches and a variety of research studies aim to improve the reconstruction quality of the high resolution component of the spatio-temporal fields.

Data assimilation, which includes OI techniques, is the state-of-the-art framework for the reconstruction of dynamical systems from partial observations [3] [4]. Statistical data assimilation has become particularly popular due to their efficient trade-off between computational performance and modeling flexibility. Broadly speaking, data assimilation provides means to combine a prior on the dynamics of the considered system and partial observations. The definition of this dynamical prior is a critical aspect. Regarding sea surface dynamics, the analytic derivation of computationally-efficient models involves theoretical assumptions, which may not be fulfilled by the real observations and reduce their applicability. On the other hand, realistic analytic parameterizations may lead to highly-computationally-demanding numerical models associated with modeling and inversion uncertainties, which reduce their applicability when addressing reconstruction issues to only some variables of a system.

Recently, data-driven approaches [5, 6] have emerged as relevant alternatives to model-driven schemes. They take benefit from the increasing availability of remote sensing observation and simulation data to derive priors from these datasets. Analog methods are one of the first data-driven techniques to develop this data-driven paradigm within a data assimilation framework [5]. Analog forecasting operators provide a data-driven formulation of the dynamical operator, which can be plugged in classic Kalman-based assimilation schemes. Applications to sea surface geophysical fields [7] using a patch-based representation demonstrated the relevance of such approaches compared with OI and Empirical Orthogonal Functions (EOF) based schemes.

In this work, we further investigate data-driven approaches within a statistical data assimilation framework for

the reconstruction of sea surface dynamics. More specifically, we propose to replace the ensemble forecasting formulation in the ensemble Kalman filter by a stochastic forecasting model parametrized by a neural network. We consider two case studies for the reconstruction of Sea Level Anomaly (SLA) field in the Western Mediterranean sea and for the reconstruction of Sea Surface Temperature (SST) dynamics off South Africa. From our numerical experiment this model proves relevant for the reconstruction of high resolution scales below 100km.

2. PROPOSED MODEL

2.1. High resolution SLA component extraction

We aim to reconstruct the high resolution scales unresolved by classical techniques such as OI. Therefore, given a high resolution spatio-temporal field x and the corresponding observations y (possibly noisy with missing data, typically along track observations for an SLA field). A large scale component $\overline{x}$ is derived using these observations based on an optimal interpolation. The high resolution component of our field is then deduced according to :

$$z = x - \overline{x} \tag{1}$$

Here, z corresponds to the high resolution scales unresolved by OI.

Following our recent study [7], we decompose the spatial scales of our anomaly field z using a patch based EOF decomposition. This representation allows us to significantly reduce the complexity of the proposed model while still keeping a high explained variance ratio. Formally, z is decomposed into N_p $P \times P$ patches $z_{s,t}$ where $t \in \{0, ..., T\}$ is the temporal sampling of our time series and P the width and height of each patch, s represents the patch's position in the global field. Each patch is then projected into an EOF basis $\mathcal{B}_s$ as follow :

$$z_{s,t} = \sum_{k=1}^{N_E} \alpha_{s,k}(t)\beta_{s,k} \tag{2}$$

with $\beta_{s,k}$ the kth EOF basis and $\alpha_{s,k}(t)$ the corresponding EOF coefficients for patch $z_{s,t}$. The projection of the anomaly patch $z_{s,t}$ in the EOF space is the vector of the N_E coefficients $\alpha_{s,k}(t)$ denoted as $v_{s,t}$. We may also note $y_{s,t}$ as the restriction of the observations y in the patch region s at time t.

2.2. Patch-based reconstruction

In oceanography, spatio-temporal data assimilation issues can be regarded as the reconstruction of some hidden states from partial and/or noisy observation series [8]. From this point of view, and based on the above formulation, the patch based reconstruction issue can be formulated as the reconstruction of

the patch based EOF time series $v_{s,t}$ based on given observations $y_{s,t}$. Supposing that the evolution of our EOF states relates to some unknown underlying dynamical model, we can write the following state space model :

$$v_{s,t+1} = \mathcal{F}_s(v_{s,t}) + \eta_t \tag{3}$$
$$y_{s,t+1} = \mathcal{B}_s v_{s,t+1} + \epsilon_t \tag{4}$$

where $\mathcal{F}$ the dynamical model describing the temporal evolution of the EOF states variables v_s. The observation model $\mathcal{B}_s$ is the EOF basis formed by the N_E EOF coefficients $\beta_k, k \in [1, ..., N_E]$. η_t and ϵ_t are random processes accounting for the uncertainties in the dynamical and observation models. They are usually defined as centered Gaussian processes with covariances Q_t and R_t respectively.

Ensemble Kalman filtering schemes [8] are among the state-of-the-art data assimilation techniques in the reconstruction of geophysical dynamics. The key idea of these methods is to approximate the forecasting mean and covariance, which are derived analytically in the linear case of a classical Kalman filter, by a sample mean and covariance matrix computed by propagating an ensemble $v_{s,t}^{i-}, i \in \{1, ..., N\}$ where N represents the ensemble size using the dynamical model $\mathcal{F}_s$.

$$v_{s,t+1}^{i-} = \mathcal{F}_s(v_{s,t}^{i+}), i \in \{1, ..., N\} \tag{5}$$
$$\Sigma_{t+1}^{-} = \frac{1}{N-1} D_{t+1} D_{t+1}^t \tag{6}$$
$$D_{t+1} = [v_{s,t+1}^{1-} - v_{s,t+1}^{-}, ..., v_{s,t+1}^{N-} - v_{s,t+1}^{-}] \tag{7}$$
$$v_{s,t+1}^{i+} = v_{s,t+1}^{i-} + K_{t+1}[y_{s,t+1} - \mathcal{B}_s v_{s,t+1}^{i-}] \tag{8}$$
$$K_{t+1} = \Sigma_{t+1}^{-} \mathcal{B}_s^T [\mathcal{B}_s \Sigma_{t+1}^{-} \mathcal{B}_s^T + R_t]^{-1} \tag{9}$$
$$\Sigma_{t+1}^{+} = \Sigma_{t+1}^{-} - K_{t+1} \mathcal{B}_s \Sigma_{t+1}^{-} \tag{10}$$

Here the superscript (-) refers to the forecasting of the ensemble of state particles given observations up to time t but without the new observation at time $t + 1$. The superscript (+) refers in the other hand to the corrected particles given all observations up to time $t + 1$.

Given a representative dataset of past realizations, we can replace the ensemble forecasting formulation in the ensemble Kalman filter by a data driven parametric stochastic model. Formally, we can replace the equations (5), (6) and (7) in the above ensemble Kalman recursion by the forecasting of a Gaussian mean component v_s^{-} and covariance pattern Σ^{-} as follow :

$$v_{s,t+1}^{-} = \mathcal{F}_s(v_{s,t}^{+}) \tag{11}$$
$$\Sigma_{t+1}^{-} = \mathcal{F}_{\Sigma}^s(v_{s,t}^{+}, \Sigma_t^{+}) \tag{12}$$
$$\mathcal{F}_{\Sigma}^s(v_s(t)^{+}, \Sigma_t^{+}) = \Psi(\Sigma_t^{+}, \Sigma_0) \cdot \mathcal{F}_D^s(v_{s,t}^{+}, \Sigma_0) \tag{13}$$

with $\mathcal{F}_s$ the approximate dynamical model and $\mathcal{F}_D^s(v_s(t), \Sigma_0)$ a diagonal covariance model in the EOF space, $\Psi(\Sigma_t^{+}, \Sigma_0)$ is a scaling function. $\mathcal{F}_s$ and $\mathcal{F}_D^s$ are parametrized by neural networks.

The proposed mean dynamical model $\mathcal{F}_s$ is learnt based on the minimization of the forecasting error. The covariance model $\mathcal{F}_D^s$ model is optimized using an ensemble likelihood maximization to emulate the ensemble forecasting used in classical ensemble filtering techniques.

3. NUMERICAL EXPERIMENTS

3.1. SLA case study

Dataset description : The dataset used in our experiments is a gap-free SLA time series obtained using the WMOP product [9]. The spatial resolution of our data is a $0.05°$ and the temporal resolution $h = 1$ day. We use the data from January 2009 to December 2014 as training data and we tested our approach on the first 347 days of the year 2015. We also implement an Observing system simulation experiment (OSSE) to generate synthetic observations of satellite altimeter data from real satellite track spatio-temporal locations from a four-altimeter sampling configuration in 2015. We test our approach on two different patches located on south Mallorca ($2.5°E$ to $4.25°E$, $37.25°N$ to $39.5°N$) and north Algeria ($2.5°E$ to $4.25°E$, $36.5°N$ to $38.25°N$). These two patches were selected to be representative of different dynamical behaviors. The patch size used in this work is $P = 35$ and the EOF space dimension $N_E = 18$, which amounts to capture 95% of the total variance.

Experimental setting : Regarding the proposed Neural Networks Kalman Filter (NNKF), the following configuration was adopted. For the patch-level NN model $\mathcal{F}$, we use a bilinear residual neural network architecture with 16 linear neurons, 20 bilinear neurons and 6 fully-connected layers with a *Relu* activation.

Regarding the covariance model $\mathcal{F}_\Sigma$, we consider an MLP with 6 layers, 5 hidden layers with 100 neurones and *Relu* activations and an output layer with a *softplus* activation.

We perform a quantitative analysis of the interpolation performance of the proposed scheme with respect to an optimal interpolation, and the analog data assimilation. Overall, the considered parameter setting is as follows:

- Optimal interpolation (OI) : We use a Gaussian kernel with a spatial correlation length of 100 km and a temporal resolution length of 20 days. These parameters were empirically tuned for the considered dataset using a cross-validation experiment. The OI reconstruction was held into the Balearic region ($1.5°E$ to $8.5°E$, $36.5°N$ to $40°N$) since patch based OI reconstruction leads to very poor reconstructions due to the big rate of missing data.

- Analog data assimilation (LAF-EnKF, GAF-EnKF): We apply both the global and local analog data assimi-

lation schemes, referred to as GAF-EnKF, LAF-EnKF respectively [7, 5]. This technique is based on a locally linear analog forecasting operator in a classical Ensemble Kalman filter with 100 particles.

Finally, regarding the assimilation experiment with the proposed model and the analog data assimilation, we concatenate the observations given at day t with a half window size of 3 days.

Interpolation performances : The analysis of the assimilation results presented in the table 1 illustrates clearly the advantage of the proposed framework. Despite the fact that the OI reconstruction was performed through a much larger region which provides more observations, especially in the patches boundaries, Our method still gives better results. Also, our model gives similar results comparing to the local analogs based assimilation technique within a huge gain in computational complexity. This is principally due to the parametric data-driven formulation of the Gaussian transition function.

	Model		SLA	Gradient	Exec time
Patch 1	OI	RMSE	0.031	0.0053	
		Correlation	82.66%	52.42%	
	GAF-EnKF	RMSE	0.027	0.0048	$\sim 40sec$
		Correlation	87.28%	63.58%	
	LAF-EnKF	RMSE	0.025	0.0041	$\sim 390sec$
		Correlation	89.21%	68.75%	
	NNKF	RMSE	0.026	0.0041	$\sim 2sec$
		Correlation	88.35%	66.72%	
Patch 2	OI	RMSE	0.040	0.0072	
		Correlation	65.05%	30.21%	
	GAF-EnKF	RMSE	0.036	0.0066	$\sim 40sec$
		Correlation	82.07%	42.19%	
	LAF-EnKF	RMSE	0.034	0.0060	$\sim 390sec$
		Correlation	84.98%	45.94%	
	NNKF	RMSE	0.036	0.0061	$\sim 2sec$
		Correlation	84.34%	45.71%	

Table 1: *SLA interpolation experiment: Mean reconstruction correlation coefficient and RMSE over the SLA time series and their gradients.*

3.2. SST case study

Dataset description : We use as SST data the OSTIA product [2]. The spatial resolution of the SST field is $0.05°$ and the temporal resolution $h = 1$ day. We use the data from January 2008 to December 2014 as training data and we tested our approach on the 2015 data. As case-study region, we consider a region off south Africa (from $2.5°E$, $38.75°S$ to $32.5°E$, $58.75°S$). For the data assimilation experiment, We also simulated realistic spatio-temporal cloud patterns over the test set using the METOP-AVHRR masks.

Experimental setting : Regarding the proposed neural net-

works Kalman filter (NNKF), we used the following configuration. The case study region was decomposed into a collection of non-overlapping 20×20 patches. Each patch was then projected into an EOF basis with 50 components (witch amounts to 95 % of the total variance). Each patch based EOF decomposition was then forecasted using the described patch based stochastic model (a bilinear residual neural network with 60 linear neurons, 100 bilinear neurons and 10 fully-connected layers with a *Relu* activation for the mean model $\mathcal{F}_s$ and an MLP with 3 layers, 2 hidden layers with 200 neurones and *Relu* activations and an output layer with a *softplus* activation for the covariance model $\mathcal{F}_D^s$). Finally and to avoid a post-processing step, the forecasted EOF states are propagated back into the patch space and fed into a convolutional neural network (with 3 layers). This network is trained to combine the forecasted patches to form the global region.

We compare our approach to the same methods mentioned above. For the optimal interpolation, we use a Gaussian kernel with a spatial correlation length of 100km and a temporal resolution length of 3 days. For the analog data assimilation, we decompose the region into 20×20 patches with 50% overlapping. Each patch is propagated into an EOF basis with 50 components and reconstructed using a locally linear analog forecasting operator in an ensemble Kalman filter.

Interpolation performances : We report the data assimilation performance in Tab.2. The proposed NNKF leads to significant improvements with respect to OI and to the analog based schemes. From a methodological point of view, the interpolation improvement with respect to OI is principally due to the use of a data driven dynamical prior. Concerning the analog data assimilation, the mapping between the SST field and its patch-level representation plays a significant role in the improvement of our reconstruction. The analog data assimilation relies on a post-processing filtering step to remove patches boundaries. This post-processing step decreases the quality of the reconstruction as some fine scale details could be filtered.

Model		SST	Gradient
OI	RMSE	0.76	0.25
	Correlation	99.35%	60.31%
GAF-EnKF	RMSE	0.48	0.19
	Correlation	99.74%	79.12%
LAF-EnKF	RMSE	0.43	0.16
	Correlation	99.79%	84.41%
PB-NNKF-EOF	RMSE	**0.33**	**0.13**
	Correlation	**99.87%**	**89.87%**

Table 2: *SST interpolation experiment*: *Mean reconstruction correlation coefficient and RMSE over the SST time series.*

4. CONCLUSION

In this work, propose an alternative to the ensemble Kalman filter for the spatio-temporal reconstruction of sea surface geophysical tracers. Our architecture relies on the definition of a Gaussian dynamical model that will replace the ensemble forecasting step in the Kalman recursion. From our numerical experiment, our model gives at least similar results to the ensemble Kalman filter in the reconstruction of patch based SLA and SST dynamics while drastically decreasing the numerical complexity of the interpolation routine. This gain in numerical complexity is due to the propagation of a parametric Gaussian distribution instead of an ensemble.

5. REFERENCES

[1] N. J. Hardman-Mountford, A. J. Richardson, D. C. Boyer, A. Kreiner, and H. J. Boyer, "Relating sardine recruitment in the Northern Benguela to satellite-derived sea surface height using a neural network pattern recognition approach," *Progress in Oceanography*, vol. 59, no. 2, pp. 241–255, oct 2003.

[2] C. J. Donlon, M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer, "The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system," *Remote Sensing of Environment*, vol. 116, no. Supplement C, pp. 140–158, jan 2012.

[3] L. Bertino, G. Evensen, and H. Wackernagel, "Sequential Data Assimilation Techniques in Oceanography," *International Statistical Review*, vol. 71, no. 2, pp. 223–241, 2003.

[4] A. C. Lorenc, S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, and F. W. Saunders, "The Met. Office global three-dimensional variational data assimilation scheme," *Quarterly Journal of the Royal Meteorological Society*, vol. 126, no. 570, pp. 2991–3012, oct 2000.

[5] R. Lguensat, P. Tandeo, P. Ailliot, M. Pulido, and R. Fablet, "The Analog Data Assimilation," *Monthly Weather Review*, aug 2017.

[6] 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, may 2016.

[7] 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*, vol. 3, no. 4, pp. 647–657, dec 2017.

[8] G. Evensen, *Data Assimilation*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[9] M. Juza, B. Mourre, L. Renault, S. Gómara, K. Sebastián, S. Lora, J. P. Beltran, B. Frontera, B. Garau, C. Troupin, M. Torner, E. Heslop, B. Casas, R. Escudier, G. Vizoso, and J. Tintoré, "Socib operational ocean forecasting system and multi-platform validation in the western mediterranean sea," *Journal of Operational Oceanography*, vol. 9, no. sup1, pp. s155–s166, 2016.