Neural-Network-based Kalman Filters for the Spatio-Temporal Interpolation of Satellite-derived Sea Surface Temperature

Said Ouala, Ronan Fablet, Cédric Herzet, Bertrand Chapron, Ananda Pascual, Fabrice Collard, Lucile Gaultier

Article

Neural-Network-based Kalman Filters for the Spatio-Temporal Interpolation of Satellite-derived Sea Surface Temperature

Said Ouala [1]*, Ronan Fablet [2], Cédric Herzet [3], Bertrand Chapron [4], Ananda Pascual [5], Fabrice Collard [6], Lucile Gaultier [7]

1 IMT Atlantique, Lab-STICC, UBL, Brest, France; said.ouala@imt-atlantique.fr
2 IMT Atlantique, Lab-STICC, UBL, Brest, France; ronan.fablet@imt-atlantique.fr
3 IMT Atlantique, Lab-STICC, UBL, Brest; INRIA Bretagne-Atlantique, SIMSMART, Rennes, France; cedric.herzet@inria.fr
4 Ifremer, LOPS, Brest, France; Bertrand.Chapron@ifremer.fr
5 IMEDEA, UIB-CSIC, Esporles, Spain; ananda.pascual@imedea.uib-csic.es
6 ODL, Brest, France; dr.fab@oceandatalab.com
7 ODL, Brest, France; lucile.gaultier@oceandatalab.com
* Correspondence: said.ouala@imt-atlantique.fr

Version October 15, 2018 submitted to Remote Sens.

Abstract: In this work we address the reconstruction of gap-free Sea Surface Temperature (SST) fields from irregularly-sampled satellite-derived observations. We develop novel Neural-Network-based (NN-based) Kalman filters for spatio-temporal interpolation issues as an alternative to ensemble Kalman filters (EnKF). The key features of the proposed approach are two-fold: the learning of a probabilistic NN-based representation of 2D geophysical dynamics, the associated parametric Kalman-like filtering scheme for a computationally-efficient spatio-temporal interpolation of Sea Surface Temperature (SST) fields. We illustrate the relevance of our contribution for an OSSE (Observing System Simulation Experiment) in a case-study region off South Africa. Our numerical experiments report significant improvements in terms of reconstruction performance compared with operational and state-of-the-art schemes (e.g., optimal interpolation, Empirical Orthogonal Function (EOF) based interpolation and analog data assimilation).

Keywords: Data assimilation; Dynamical model; Kalman filter; Neural networks; Data-driven models; Interpolation

---

1. Introduction

Satellite sensors and in-situ networks can provide observations of sea surface tracers (e.g. temperature, salinity, ocean colour). However, due to sensors's characteristics (e.g., space-time sampling, sensor type) and their sensitivity to the atmospheric conditions (e.g., rain, clouds), only partial and possibly noisy observations are available. As a consequence, no sensor can provide gap-free high-resolution observations in space and time. A typical example of the missing data pattern for SST is reported in Fig. 3 for an infrared sensor. In some situations, missing data may become very large which makes crucial the development of spatio-temporal interpolation tools.

Within the satellite ocean community, Optimal interpolation (OI) is the standard technique [1–7]. Given a covariance model of spatio-temporal dynamics, the interpolated field results from a linear combination of the observations. The parameters of the linear combination are typically tuned by exploiting some statistical properties of the target field.

In general, stationary covariance hypotheses are considered, which prove relevant for the reconstruction of horizontal scales above 100km. Fine scale components may hardly be retrieved with such approaches and a variety of research studies aim to improve the reconstruction of the high-resolution component of our spatio-temporal fields.

Empirical Orthogonal Function (EOF) based interpolation is an other categorie widely used in geosciences [8–10]. They rely on a Singular Value Decomposition (SVD) to compute the EOF basis, the field is then reconstructed by projecting the observations on the EOF subspace until a convergence criterion is reached [11]. Unfortunately, dealing with high missing data rates decreases the encoded variability in the EOF components witch results in smoothing fine scale components.

Data assimilation is the state-of-the-art framework for the reconstruction of dynamical systems from partial observations based on a given numerical model [12,13]. Statistical data assimilation schemes especially ensemble Kalman filters, have become particularly popular due to their trade-off between computational efficiency and modeling flexibility. Unlike OI and EOF based techniques, these schemes explicitly rely on dynamical priors to address interpolation issues from partial and noisy observations. When dealing with sea surface dynamics, the analytical derivation of these priors involves simplifying assumptions which may not be satisfied by real observations. By contrast, realistic analytical parameterizations may lead to highly computationally-demanding numerical models associated with modeling and inversion uncertainties, which may limit their relevance for an application of the interpolation of a single sea surface tracer.

Recently, data-driven approaches [8,14] 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 dynamical 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 [14]. Analog forecasting operators provide a data-driven formulation of the dynamical operator, which can be used as a plug-and-play operator in Kalman-based assimilation schemes. Combined with patch-based representation, the analog data assimilation was recently proven to be relevant with respect to OI and EOF-based schemes for the spatio-temporal interpolation of sea surface geophysical tracers [15–17].

In this paper, we further investigate data-driven interpolation approaches within a statistical data assimilation framework. We focus on neural network and deep learning models, which have rapidly become the state-of-the-art in machine learning for a wide range of applications, including inverse imaging issues [18]. Recent applications to the assimilation of low-dimensional dynamical systems [19] and to the forecasting of geophysical dynamics [20] have been developed. However, to our knowledge, the design of neural-network-based assimilation models for the spatio-temporal interpolation of geophysical dynamics remain an open challenge, which may greatly benefit from the ability of deep learning models to capture computationally-efficient representations from available ocean observation and simulation datasets. In this study, we address this challenge and propose a novel NN-based Kalman filtering scheme applied to the spatio-temporal interpolation of satellite-derived sea surface temperature. We exploit a ResNet architecture [19,21] and a patch-based decomposition [22] to derive a data-driven representation of spatio-temporal fields. Importantly, this architecture conveys a probabilistic representation through the prediction of a mean component and a covariance pattern. The later may be regarded as a NN-based representation of the covariance patterns issued from Monte Carlo approximations in ensemble assimilation schemes [23]. Overall, the methodological contributions of this work are two-fold: i) we propose a new probabilistic NN-based representation of 2D geophysical dynamics, ii) we derive the associated NN-based Kalman filtering scheme for spatio-temporal interpolation issues. We demonstrate the relevance of these contributions with respect to state-of-the-art approaches [2,8,16] for the spatio-temporal interpolation of satellite-derived SST fields in a case study region off South Africa. This paper is organized as follows. Section 2 reviews data assimilation schemes. Section 3 describes the proposed neural-network-based data assimilation framework. Section 4 presents the results of the numerical experiments. We further discuss our contributions in Section 6.

2. Problem statement and related work

Regarding ocean remote sensing data, spatio-temporal interpolation issues can be regarded as the reconstruction of some hidden states from partial and/or noisy observation series, referred to as data assimilation in geoscience [23]. Data assimilation techniques usually involve a state-space evolution model [23]:

$$x_{t+1} = \mathcal{F}(x_t) + \eta_t \tag{1}$$
$$y_{t+1} = \mathcal{H}(x_{t+1}) + \epsilon_t \tag{2}$$

where $t \in \{0, ..., T\}$ represents the temporal resolution of our time series and $\mathcal{F}$ the dynamical model describing the temporal evolution of the physical variables x. The observation model $\mathcal{H}$ links the observation y to the physical variable x. 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.

From a probabilistic point of view, the spatio-temporal interpolation problem can be seen as a Bayesian filtering problem where the main goal is to evaluate the conditional probabilities $p(x_{t+1}|y_1, ..., y_t)$ (prediction distribution of the state x_{t+1} given observations up to time t) and $p(x_{t+1}|y_1, ..., y_t, , y_{t+1})$ (posterior distribution of x_{t+1} given observations up to time $t+1$). Under certain assumptions over the state space model (the dynamical and observation models are linear with Gaussian uncertainties), the prediction and posterior distributions are also Gaussian and can be written as :

$$p(x_{t+1}|y_1, ..., y_t) = \mathcal{N}(x^-_{t+1}, \Sigma^-_{t+1}) \tag{3}$$
$$p(x_{t+1}|y_1, ..., y_{t+1}) = \mathcal{N}(x^+_{t+1}, \Sigma^+_{t+1}) \tag{4}$$

with the means and covariances computed for each time t using the well known Kalman recursion

$$x^-_{t+1} = Fx^+_t \tag{5}$$
$$\Sigma^-_{t+1} = F\Sigma^+_t F^T + Q_t \tag{6}$$
$$x^+_{t+1} = x^-_{t+1} + K_{t+1}[y_{t+1} - H_{t+1}x^-_{t+1}] \tag{7}$$
$$\Sigma^+_{t+1} = \Sigma^-_{t+1} - K_{t+1}H_{t+1}\Sigma^-_{t+1} \tag{8}$$

with

$$K_{t+1} = \Sigma^-_{t+1}H^T_{t+1}[H_{t+1}\Sigma^-_{t+1}H^T_{t+1} + R_t]^{-1}. \tag{9}$$

Here F and H_{t+1} corresponds respectively to some linear dynamical and observation models. The superscript (-) refers to the forecasting of the mean of the state variable x^-_{t+1} and of its covariance matrix Σ^-_{t+1} given observations up to time t but without the new observation at time $t+1$. The superscript (+) refers in the other hand to the mean of the state variable x^+_{t+1} and of the covariance matrix Σ^+_{t+1} given all observations up to time $t+1$. They are referred to as the assimilated mean and covariance. K_{t+1} is the Kalman gain. Kalman filters provide a sequential formulation of the Optimal Interpolation (OI) [24] which may also be solved directly knowing the space-time covariance of processes x and y. For non-linear and high-dimensional dynamical systems, the pdfs are not Gaussian anymore and the above Kalman recursion does define their means and covariances. Ensemble Kalman methods have been proposed to address these issues. The ensemble Kalman filter and smoother [23] are the first sequential filtering techniques used reliably in the reconstruction of geophysical fields. The key idea here is to approximate the forecasting mean x^-_{t+1} and covariance Σ^-_{t+1} by a sample mean and covariance matrix computed by propagating an ensemble of M members, $\{x^{i-}_{t+1}\}_{i=1}^{M}$, using the dynamical model $\mathcal{F}$.

$$x^{i-}_{t+1} = \mathcal{F}(x^{i+}_t, i \in \{0, ..., N\}) \tag{10}$$
$$\Sigma^-_{t+1} = \frac{1}{N-1} D_{t+1} D^t_{t+1} \tag{11}$$
$$D_{t+1} = [x^{1-}_{t+1} - x^-_{t+1}, ...x^{N-}_{t+1} - x^-_{t+1}] \tag{12}$$
$$x^{i+}_{t+1} = x^{i-}_{t+1} + K_{t+1}[y_{t+1} - H_{t+1} x^{i-}_{t+1}] \tag{13}$$
$$K_{t+1} = \Sigma^-_{t+1} H^T_{t+1} [H_{t+1} \Sigma^-_{t+1} H^T_{t+1} + R_t]^{-1} \tag{14}$$
$$\Sigma^+_{t+1} = \Sigma^-_{t+1} - K_{t+1} H_{t+1} \Sigma^-_{t+1} \tag{15}$$

Besides all its advantages, EnKF techniques do not escape the curse of dimensionality. High-dimensional systems require using large ensemble sizes M which may lead to very high-computational complexity. The use of small ensemble sizes in the other hand may result in undersampling the covariance matrix (the considered ensemble is not representative of our systems dynamics) which may in turn result in poor reconstruction performance, including for instance filter divergence and spurious long-range correlations. Proposed solutions such as inflation [25], cross-validation [26] and localization methods[27–29] may require thorough tuning experiments. An alternative strategy based on a model-driven propagation of parametric covariance models [30,31] seems appealing. Using advection priors [32], it propagates parametric covariance structures, which leads to the implementation of the classic Kalman recursion. Accounting for more complex dynamical priors for the covariance structure is an open question, which may limit the applicability of this approach to complex geophysical systems. Inspired by the later parametric framework, we aim to design an efficient sequential filtering technique for the reconstruction of geophysical fields. Rather than considering a model-driven prior to propagate Gaussian states as in [30,31], we investigate NN-based priors, which may be fitted from training data. The resulting NN-based Gaussian representations provide computationally-efficient approximations of the dynamical priors that should prevent undersampling issues within a Kalman recursion.

3. Proposed interpolation model

3.1. Neural-network Gaussian dynamical prior

Our key idea is to exploit neural-network (NN) representations for the time propagation of a Gaussian approximation of the distribution of the state. Compared with dynamical priors in assimilation model (1), which state conditional distribution $x_t|x_{t-1}$, we here consider neural-network representations to extend the prediction step of the Kalman recursion (5-6) to non-linear dynamics. Formally, it comes to define:

$$x^-_{t+1} = \mathcal{F}(x^+_t) \tag{16}$$
$$\Sigma^-_{t+1} = \mathcal{F}_\Sigma(x^+_t, \Sigma^+_t) \tag{17}$$

with x^-_{t+1} and Σ^-_{t+1} the mean and covariance of the prediction of the Gaussian approximation of the state at time $t+1$ given the assimilated mean x^+_t and covariance Σ^+_t at time t. Functions $\mathcal{F}, \mathcal{F}_\Sigma$ are neural networks to be defined with parameter vectors $\theta = (\theta_\mu, \theta_\Sigma)$. It may noted that our parameterization follows (5-6) such that the update of the mean component in (16) only depends on the mean at the previous time step and the update of the covariance depends both on the mean and covariance at the previous time step. Given this NN-based representation of the prediction step of the Kalman filter, we apply the classic Kalman-based filtering under the assumption that the observation model is linear and Gaussian:

$$x^+_{t+1} = x^-_{t+1} + K_{t+1}[y_{t+1} - H_{t+1}x^-_{t+1}] \tag{18}$$
$$K_{t+1} = \Sigma^-_{t+1}H^T_{t+1}[H_{t+1}\Sigma^-_{t+1}H^T_{t+1} + R_t]^{-1} \tag{19}$$

Such a formulation does not require forecasting an ensemble to compute a sample covariance matrix. It results in a significant reduction of the computational complexity. The same holds when compared to the computational complexity of the analog data assimilation which involves ensemble forecasting and repeated nearest-neighbor search.

3.2. Patch-based NN architecture

When considering spatio-temporal fields, the application of the model defined by (16) and (17) should be considered with care to account for the underlying dimensionality, especially for the covariance model in (19). Following our previous works on analog data assimilation [15,16], we consider a patch-based representation[1]. This patch-based representation is fully embedded in the considered NN 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 later involves a reconstruction operator which is learnt from data.

Regarding model $\mathcal{F}$, the proposed architecture proceeds as follows:

- At a given time t, the first layer of the network, which is parameter-free in terms of training, comes to decompose an input field x_t into a collection of N_p $P \times P$ patches $x_{\mathcal{P}_s,t}$, where P is the width and height of each patch and s the patch location in the global field. Each patch is decomposed onto an EOF basis $\mathcal{B}$ according to :

$$z_{\mathcal{P}_s,t} = x_{\mathcal{P}_s,t}\mathcal{B}^T \tag{20}$$

  with $z_{\mathcal{P}_s,t}$ the EOF decomposition of the patch $x_{\mathcal{P}_s,t}$. The EOF decomposition matrix $\mathcal{B}$ is trained offline as preprocessing step;
- The second layer implements a numerical integration scheme (typically, an Euler or 4th-order Runge-Kutta scheme) using a patch-level dynamical model $\mathcal{F}^{\mathcal{P}_s}, s \in [1, ..., N_p]$ to predict $z_{\mathcal{P}_s,t+1}$. For patch-level models $\mathcal{F}^{\mathcal{P}_s}$, we consider residual architectures [21] with a bilinear parameterization [19];
- The third layer is a reconstruction network $\mathcal{F}_r$. It combines the predicted patches $x_{\mathcal{P}_s,t} = z_{\mathcal{P}_s,t}\mathcal{B}, s \in [1, ..., N_p]$ to reconstruct the output field x_t. This reconstruction network $\mathcal{F}_r$ involves a convolution neural network [33].

The details of the considered parameterizations for the second and third layers are given in Section 4. To train mean dynamical model $\mathcal{F}$, we apply a two-step procedure. We first learn the local dynamical models $\mathcal{F}^{\mathcal{P}_s}, s \in [1, ..., N_p]$ based on the minimization of the EOF-patch based forecasting error. The reconstruction network $\mathcal{F}_r$ is then optimized using the same criterion over the global field.

[1] A patch is a $P \times P$ subregion of a 2D field with P the width and the height of the patch.

Regarding covariance model $\mathcal{F}_\Sigma$ we also consider a patch-based representation of the spatial domain $\mathcal{F}_\Sigma^{\mathcal{P}_s}$, more precisely a block-diagonal parameterization of the patch-level covariances in the EOF space. It may be noted that a diagonal parameterization of the covariance in the EOF space forms a full covariance matrix in the original patch space. This block-diagonal covariance model $\mathcal{F}_\Sigma$ is learnt separately for each patch according to a ML (Maximum Likelihood) criterion. The associated training dataset comprises patch-based EOF decompositions of the forecasted states according to the mean model $\mathcal{F}^{\mathcal{P}_s}$ from states of the training dataset corrupted by an additive Gaussian perturbation with a covariance structure Σ_0. Here, Σ_0 is given by the empirical covariance of the EOF patches for the entire training dataset. Overall, for a given patch $\mathcal{P}_s$, we parameterize $\mathcal{F}_\Sigma^{\mathcal{P}_s}$ the restriction of covariance $\mathcal{F}_\Sigma$ onto patch $\mathcal{P}_s$ as:

$$\mathcal{F}_\Sigma^{\mathcal{P}_s}(x_{\mathcal{P}_s,t+1}, \Sigma_{\mathcal{P}_s,t+1}) = \mathcal{B}^t \Psi(\Sigma_{\mathcal{P}_s,t}, \Sigma_0) \cdot \mathcal{F}_D^{\mathcal{P}_s}(z_{\mathcal{P}_s,t}, \Sigma_0) \cdot \mathcal{B} \quad (21)$$

with $\Psi(\Sigma_{\mathcal{P}_s,t-1}, \Sigma_0)$ a scaling function. Among different parameterizations, a constant scaling function $\Psi() = 1$ led to the best performance in our numerical experiments.

To illustrate the relevance of the proposed full covariance matrix parametrization (based on a patch based projection on the EOF space and illustrated for instance by equation 21), we also investigate a diagonal covariance matrix model in the patch space.

3.3. Data assimilation procedure

Given a trained patch-based NN representation as described in the previous section, we derive the associated Kalman-like filtering procedure. As summarized in Algorithm 1, at time step t, given the Gaussian approximation of the posterior likelihood $P(x_{t-1}|y_0, \ldots, y_{t-1})$ with mean x_{t-1}^+ and covariance Σ_{t-1}^+, we first compute the forecasted Gaussian approximation at time t with mean field $\mathcal{F}(x_{t-1}^+)$ and patch-based covariance $\mathcal{F}_\Sigma(x_{t-1}^+, \Sigma_{t-1}^+)$. The assimilation of the new observation y_t is performed at a patch-level. For each patch $\mathcal{P}_s$, we update the patch-level mean $x_{\mathcal{P}_s,t}^+$ and covariance $\Sigma_{\mathcal{P}_s,t}^+$ using Kalman recursion (8) with observation $y_{\mathcal{P}_s,t}$. We then combine these patch-level updates to obtain global mean x_t^+ and covariance Σ_t^+. Whereas we compute global mean x_t^+ using trained reconstruction network $\mathcal{F}_r$, Σ_t^+ just comes to store the collection of patch-level covariances. This procedure is iterated up to the end of the observation sequence.

Compared with the patch-based analog data assimilation [16], it might be noted that we iterate patch-level assimilation steps and global reconstruction steps thanks to the NN-based propagation of the patch-based covariance structure. This procedure potentially allows information propagation from one patch to neighborhing ones after each assimilation step. By contrast, in the patch-based analog data assimilation, each patch is processed independently, such that no such information propagation can occur. This is regarded as a key feature to account for the propagation of geophysical structures (e.g., fronts, eddies, filaments,...).

We refer to the patch-based NNKF reconstruction model using the EOF block-diagonal parameterization of the covariance model $\mathcal{F}_\Sigma$, as model PB-NNKF-EOF. The model using the diagonal parameterization of the covariance model $\mathcal{F}_\Sigma$ in the patch space is referred to as PB-NNKF.

4. Data and experimental setting

As a case-study, we address the spatio-temporal interpolation of satellite-derived SST fields associated with infrared sensors, which may involve high missing data rates (typically from 50% to 90%). We consider the same region and dataset as in [16] to make easier benchmarking analyses.

4.1. Dataset description

As SST time series used here is delivered by the UK Met Office [2] from January 2008 to December 2015. The spatial resolution of our SST field is 0.05° and the temporal resolution $h = 1$ day. The data from 2008 to 2014 were used as training data and we tested our approach on the 2015 data. To perform

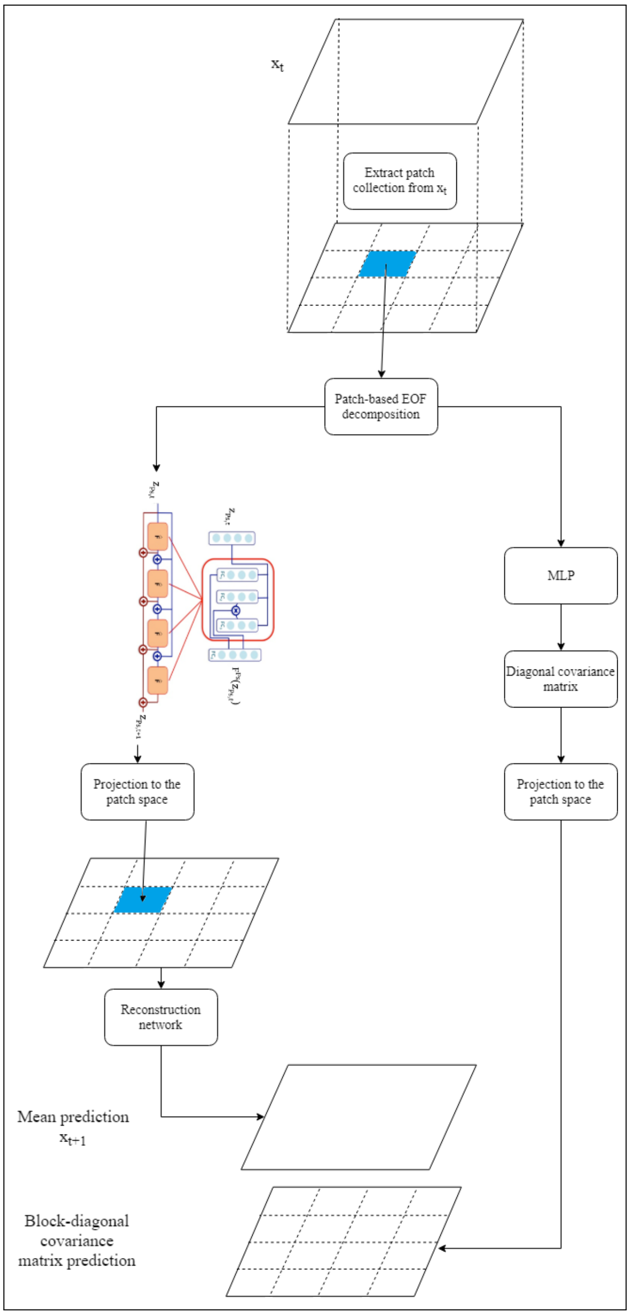

Figure 1. Proposed neural-network-based representation of a spatio-temporal dynamical system. *The input X_t is first decomposed into $P \times P$ patches, each patch is then propagated using its associate local dynamical model. The output X_{t+1} is then reconstructed by injecting the forecasted patches into the reconstruction model $\mathcal{F}_r$.*

Algorithm 1 Patch-based NNKF reconstruction

procedure PB-NNKF($\mathcal{F}$,$\mathcal{F}_\Sigma$,y,R)
for t in $[0, ..., T]$:
  $x_t^- \leftarrow \mathcal{F}(x_{t-1}^+)$
  $[\Sigma_{\mathcal{P}_0,t}^-, ..., \Sigma_{\mathcal{P}_{N_p},t}^-] \leftarrow \mathcal{F}_\Sigma(x_{t-1}^+, \Sigma_{t-1}^+)$
  $[x_{\mathcal{P}_0,t}^-, ..., x_{\mathcal{P}_{N_p},t}^-] \leftarrow ExtractPatches(x_t^-)$
  $[y_{\mathcal{P}_0,t}, ..., y_{\mathcal{P}_{N_p},t}] \leftarrow ExtractPatches(y_t)$
  for s in $[1, ..., N_p]$:
    $K_{\mathcal{P}_s,t} = \Sigma_{\mathcal{P}_s,t}^- H_{\mathcal{P}_s,t}^t [H_{\mathcal{P}_s,t} \Sigma_{\mathcal{P}_s,t}^- H_{\mathcal{P}_s,t}^t + R_t]^{-1}$
    $X_{\mathcal{P}_s,t}^+ = x_{\mathcal{P}_s,t}^- + K_{\mathcal{P}_s,t}[y_{\mathcal{P}_s,t} - H_{\mathcal{P}_s,t} x_{\mathcal{P}_s,t}^-]$
    $\Sigma_{\mathcal{P}_s,t}^+ = \Sigma_{\mathcal{P}_s,t}^- - K_{\mathcal{P}_p,t} H_{\mathcal{P}_p,t} \Sigma_{\mathcal{P}_p,t}^-$
  $x_t^+ \leftarrow Reconstruct([x_{\mathcal{P}_0,t}^+, ..., x_{\mathcal{P}_{N_s},t}^+])$
  $\Sigma_t^+ \leftarrow Reconstruct([\Sigma_{\mathcal{P}_0,t}^+, ..., \Sigma_{\mathcal{P}_{N_s},t}^+])$

a quantitative evaluation, we simulated realistic spatio-temporal cloud patterns using METOP-AVHRR masks. This sensor is highly sensitive to the cloud cover and results in very high missing data rates as illustrated in Fig. 3. As case-study area, we select an area off South Africa (from $2.5°E, 38.75°S$ to $32.5°E, 58.75°S$). This region involves complex fine-scale SST dynamics (e.g., fronts, filaments). It makes it relevant for the considered quantitative evaluation.

4.2. Experimental setting

The proposed neural-network-based Kalman scheme involves the following parameter setting. The proposed patch-based and NN-based Kalman filter is applied to SST anomaly fields w.r.t. optimally-interpolated SST fields (see below for the parameterization of the optimal interpolation). These optimally-interpolated fields provide a relevant reconstruction of horizontal scales up to ≈100km. We exploit patch-level representations with non-overlapping 20×20 patches. For each patch $\mathcal{P}_s$, we learn an EOF basis from the training data. We keep the first 50 EOF components, which amount on average to 95% of the total variance. For the patch-level NN model $\mathcal{F}^{\mathcal{P}_s}$, we use a bilinear residual neural network architecture as proposed in [34] with 60 linear neurons, 100 bilinear neurons and 10 fully-connected layers with a Relu activation. The reconstruction model $\mathcal{F}_r$ is a convolutional neural network with 3 convolutional layers. The first two layers comprise 64 filters of size 3×3 with a *Relu* activation and the last layer is a linear convolutional layer with one filter. Regarding covariance model $\mathcal{F}_D^{\mathcal{P}_s}$, we consider a diagonal covariance model within each patch. Each element of diagonal involves a a 3-layer MLP with 4 neurons and *Relu* activation functions on the hidden layers and a *softplus* activation in the output layer. With a view to evaluating the EOF-based covariance parameterization, we consider both PB-NNKF-EOF and PB-NNKF schemes.

We perform a quantitative analysis of the interpolation performance of the proposed scheme with respect to an optimal interpolation, the analog data assimilation [16] and the EOF based interpolation method VE-DINEOF. The considered parameter setting is as follows:

- Optimal interpolation (OI) : We use a Gaussian kernel with a spatial correlation length of 100km and a temporal resolution length of 3 days. These parameters were empirically tuned for the considered dataset using a cross-validation experiment.
- Analog data assimilation (LAF-EnKF, GAF-ENKF): We apply both the global and local analog data assimilation schemes, referred to as G-AnDA and L-AnDA [14,16]. Similarly to the proposed scheme, we consider 20×20 patches and 50-dimensional EOF decomposition with an overlapping of 10 pixels. We let the reader refer to [14,16] for a detailed description of this data-driven approach, which relies on nearest-neighbor regression techniques.

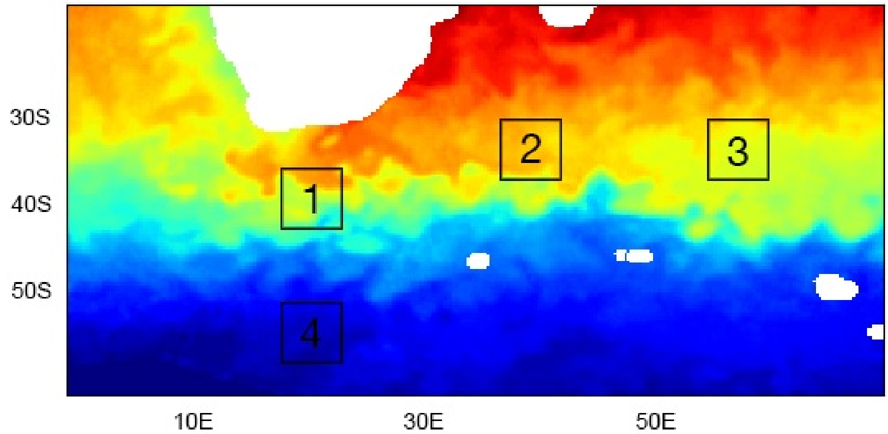

Figure 2. Selected patches on the high resolution component of the SST data. *(The SST map corresponds to July 19, 2015)*

- EOF based reconstruction (PB-VE-DINEOF): We also compare our approach to the state-of-the-art interpolation scheme based on the projection of our observations with missing data on an EOF basis [8]. The SST field is here decomposed as described in the analog data assimilation application into a collection of 20×20 patches with a 10 pixels overlapping. Each patch is then reconstructed using the VE-DINEOF method.

5. Results and discussion

We report in this section the results of the considered numerical experiments. We first focus on patch-level performance as the patch-based representation is at the core of the proposed interpolation model. We then report interpolation performance for the whole case-study region.

5.1. Patch-level interpolation performance

We first evaluate the patch-level interpolation performance of the proposed scheme for four patches corresponding to different dynamical modes as illustrated in Fig. 2 located in the area ($5°E$ to $75°E$ and latitude $25°S$ to $55°S$). In Tab.1, we report the interpolation performance in terms of RMSE (root mean square error) for the proposed EOF NN-based scheme (NNKF-EOF) and include a comparison to the local analog data assimilation (LAF-EnKF). With a view to specifically analyzing the relevance of NN-based parametric covariance model, we also apply an ensemble Kalman filter with the trained dynamical model $\mathcal{F}^{\mathcal{P}_s}$. The reported results clearly illustrate the relevance of the proposed NN-based scheme for the assimilation of a single patch. The proposed NN-based scheme, which combines a NN-based formulation of the mean forecasting operator and of the associated covariance pattern, slightly outperforms the ensemble Kalman filters, while also significantly reducing the computational complexity induced by the generation of ensembles of size 500.

5.2. Global interpolation performance

We further evaluate the performance of the proposed schemes over the considered case-study region. Tab.2 report the mean forecasting RMS error of the proposed NN-based representation compared with local and global forecasting operator [14]. The proposed patch-level NN-based model outperforms the benchmarked approaches by about 5-15% in terms of interpolation RMSE, which stresses the relevance of mean dynamical model $\mathcal{F}$.

We report the mean interpolation performance in Tab. 3 and the time series of interpolation errors illustrated in Fig. 5. The proposed NN-based scheme (PB-NNKF-EOF) leads to very significant improvements with respect to the optimal interpolation in terms of RMSE and correlation coefficients for both the SST and its gradient, which emphasizes fine-scale structures (e.g., relative improvement of

Assimilation method	Considered patch RMSE (°C)			
	*Patch*1	*Patch*2	*Patch*3	*Patch*4
LAF EnKF	0.50	0.25	0.22	0.39
Bi-NN-EnKF	0.55	0.23	0.22	0.30
Bi-NN-NNKF-EOF	**0.46**	**0.20**	**0.19**	**0.27**

Table 1. Patch-level interpolation experiment*: RMSE of the reconstructed anomaly fields for the LAF EnKF (local analog forecasting based ensemble Kalman filter), Bi-NN-EnKF (Bilinear residual neural net model ($\mathcal{F}^{\mathcal{P}_s}$) used in an ensemble Kalman filter), Bi-NN-NNKF (Proposed NNKF based on a bilinear residual neural net dynamical mean model).*

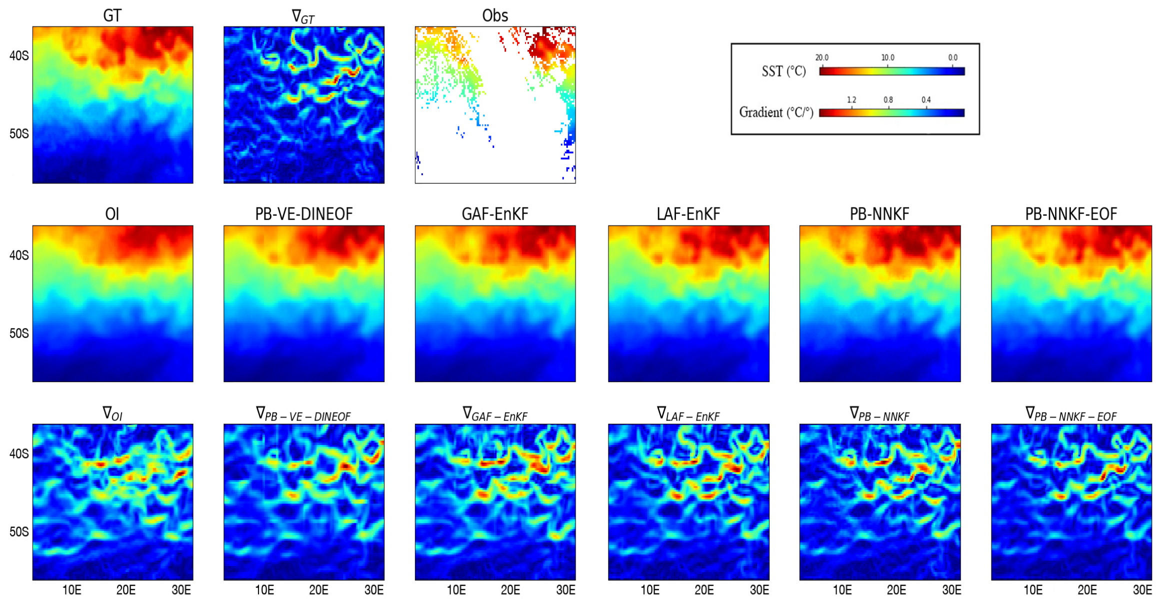

Figure 3. Interpolation of the SST field on July 19 2015: *first row, the reference SST, its gradient and the observation with missing data (here, 82% of missing data); second row, interpolation results using respectively OI, PB-VE-DINEOF, GAN-EnKF, LAN-ENKF, PB-NN-NNKF, PB-NN-NNKF-EOF; third row, gradient of the reconstructed fields.*

the RMSE above 50% for missing data areas for the SST and its gradient). A clear gain is also exhibited w.r.t. analog data assimilation and PB-VE-DINEOF schemes with a relative gain greater than 20% in terms of RMSE for both the SST and its gradient. The same conclusion holds in terms of correlation coefficients close to 90% or above for all parameters for PB-NNKF-EOF scheme, all the other ones depicting correlation coefficients below 85% for SST gradient fields. Although the considered NN-based representation exploits non-overlapping patches, we still come up with significant improvements w.r.t AnDA schemes which involve a 50% overlapping rate between patches. This clearly illustrates the relevance of NN-based representation, which fully embeds the direct and inverse mappings between the SST field and its patch-level representation. Interestingly, Tab.3 also reveals the importance of the EOF-based parameterization of the NN-based covariance model (21) in the improvement of interpolation results w.r.t. AnDA schemes.

We further illustrate these conclusions through interpolation examples in Fig. 3. The visual analysis of the reconstructed SST gradient fields emphasize the relevance of PB-NNKF-EOF scheme to reconstruct fine-scale details. While OI and PB-VE-DINEOF schemes tend to smooth out fine-scale patterns, the analog data assimilation may not account appropriately for patch boundaries. This typically requires an empirical post-processing step [16]. By contrast, the PB-NNKF-EOF scheme fully embeds this post-processing step through reconstruction layer $\mathcal{F}_r$ and learns its parameterization from data, which is shown here to greatly improve patch-based interpolation performance. The analysis of the spectral signatures leads to similar conclusions with the PB-NNKF-EOF scheme being the only one to recover significant energy level up to 50km.

Model	Forecasting RMSE (°C)		
	$t+h$	$t+4h$	$t+8h$
PB-NN	**0.48**	**0.60**	**0.63**
LAF	0.50	0.68	0.76
GAF	0.61	0.74	0.76

Table 2. Forecasting experiment for several prediction time steps

Model	Entire map				Missing data areas			
	RMSE		Correlation		RMSE		Correlation	
	$SST(°C)$	$\nabla SST(°C/°)$	SST	∇SST	$SST(°C)$	$\nabla SST(°C/°)$	SST	∇SST
PB-NNKF-EOF	**0.33**	**0.13**	**99.87**%	**89.30**%	**0.35**	**0.10**	**99.85**%	**93.49**%
PB-NNKF	0.51	0.18	99.75%	81.24%	0.51	0.18	99.71%	81.50%
LAF-EnKF	0.43	0.16	99.79%	84.41%	0.42	0.15	99.77%	86.73%
GAF-EnKF	0.48	0.19	99.74%	79.12%	0.48	0.18	99.72%	80.74%
PB-VE-DINEOF	0.54	0.20	99.68%	75.30%	0.54	0.21	99.66%	74.71%
OI	0.76	0.25	99.37%	60.31%	0.75	0.27	99.37%	55.73%

Table 3. SST interpolation experiment*: Reconstruction correlation coefficient and RMSE over the SST time series and their gradient.*

6. Conclusion

In this work, we addressed neural-network-based models for the spatio-temporal interpolation of satellite-derived SST fields with large missing data rates. We introduced a novel probabilistic

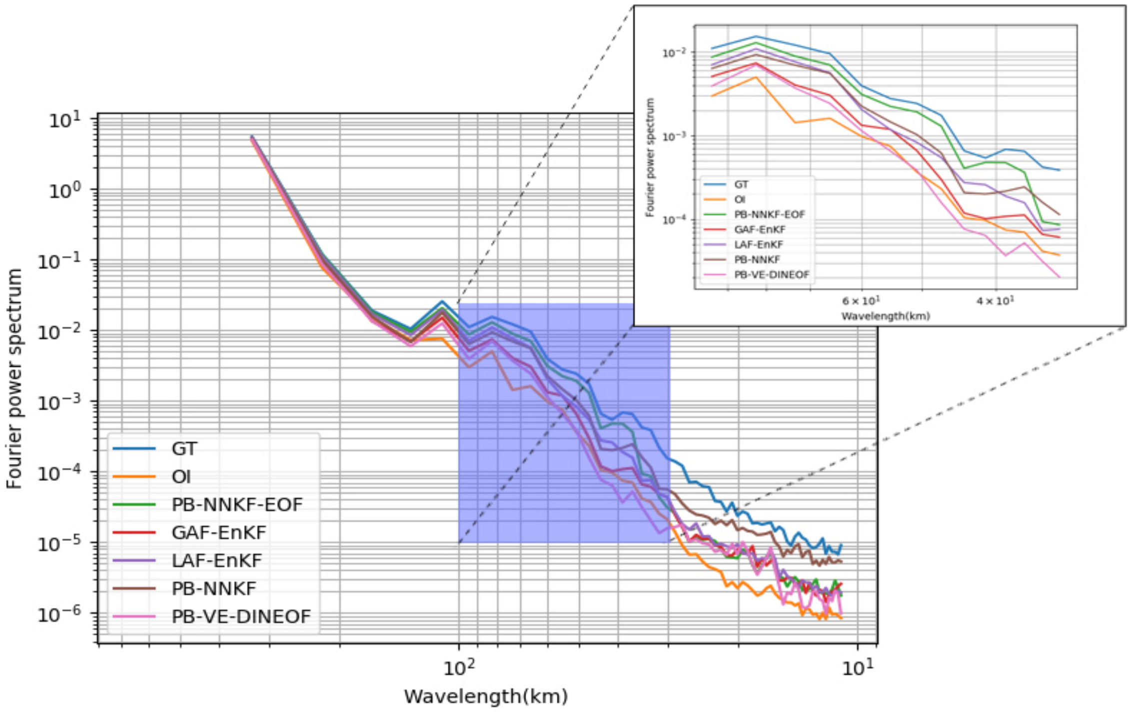

Figure 4. Radially averaged power spectral density of the interpolated SST fields with respect to the reference SST.

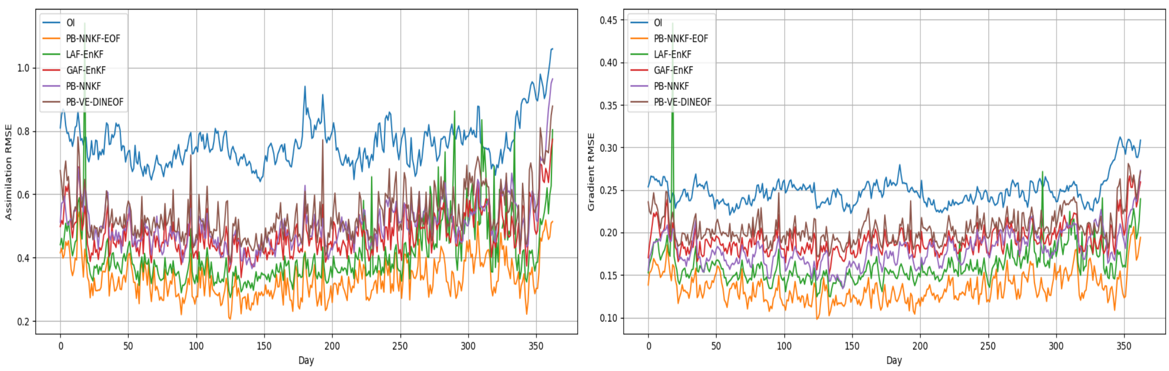

Figure 5. Interpolation RMSE times series for the selected models.

NN-based representation of geophysical dynamics. This representation, which relies on a patch-level and EOF-based representation, allows us to propagate in time a mean component and the covariance of the SST field. It makes direct the derivation of an associated Kalman filter for the spatio-temporal interpolation of SST fields. Our numerical experiments stress a significant gain in interpolation performance w.r.t. optimal interpolation and other state-of-the-art data-driven schemes, such DINEOF [8] and analog data assimilation [14,16].

Further work could explore the application of the proposed framework to other sea surface geophysical tracers, including multi-source and multi-modal interpolation issues. SLA (Sea Level Anomaly) fields could provide an interesting case-study as the associated space-time sampling is particularly scarce and multi-source strategies are of key interest [35].

Author Contributions: S.A, R.F. and C.H. stated the methodology; S.A., R.F., L.G., F.C., B.C and A.P. conceived and designed the experiments; S.A. performed the experiments; L.G., F.C., B.C. and A.P. discussed the experiments; R.L. and R.F. wrote the paper. B.C, A.P. and C.H. proofread the paper.

Funding: This work was supported by GERONIMO project (ANR-13-JS03-0002), Labex Cominlabs (grant SEACS), Region Bretagne, CNES (grant OSTST-MANATEE), Microsoft (AI EU Ocean awards) and by MESR, FEDER, Région Bretagne, Conseil Général du Finistère, Brest Métropole and Institut Mines Télécom in the framework of the VIGISAT program managed by "Groupement Bretagne Télédétection" (BreTel).

References

1. Escudier, R.; Bouffard, J.; Pascual, A.; Poulain, P.M.; Pujol, M.I. Improvement of coastal and mesoscale observation from space: Application to the northwestern Mediterranean Sea. *Geophysical Research Letters* **2013**, *40*, 2148–2153. doi:10.1002/grl.50324.
2. Donlon, C.J.; Martin, M.; Stark, J.; Roberts-Jones, J.; Fiedler, E.; Wimmer, W. The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sensing of Environment* **2012**, *116*, 140–158. doi:10.1016/j.rse.2010.10.017.
3. Le Traon, P.Y.; Nadal, F.; Ducet, N. An improved mapping method of multisatellite altimeter data. *Journal of atmospheric and oceanic technology* **1998**, *15*, 522–534.
4. Droghei, R.; Buongiorno Nardelli, B.; Santoleri, R. A New Global Sea Surface Salinity and Density Dataset From Multivariate Observations (1993–2016). *Frontiers in Marine Science* **2018**, *5*, 84.
5. Nardelli, B.B.; Pisano, A.; Tronconi, C.; Santoleri, R. Evaluation of different covariance models for the operational interpolation of high resolution satellite Sea Surface Temperature data over the Mediterranean Sea. *Remote Sensing of Environment* **2015**, *164*, 334–343.
6. Ducet, N.; Le Traon, P.Y.; Reverdin, G. Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and-2. *Journal of Geophysical Research: Oceans* **2000**, *105*, 19477–19498.
7. Gomis, D.; Ruiz, S.; Pedder, M.A. Diagnostic analysis of the 3D ageostrophic circulation from a multivariate spatial interpolation of CTD and ADCP data. *Deep Sea Research Part I: Oceanographic Research Papers* **2001**, *48*, 269–295. doi:10.1016/S0967-0637(00)00060-1.
8. Ping, B.; Su, F.; Meng, Y. An Improved DINEOF Algorithm for Filling Missing Values in Spatio-Temporal Sea Surface Temperature Data. *PLOS ONE* **2016**, *11*, e0155928. doi:10.1371/journal.pone.0155928.
9. Olmedo, E.; Taupier-Letage, I.; Turiel, A.; Alvera-Azcárate, A. Improving SMOS Sea Surface Salinity in the Western Mediterranean Sea through Multivariate and Multifractal Analysis. *Remote Sensing* **2018**, *10*, 485.
10. Alvera-Azcárate, A.; Barth, A.; Parard, G.; Beckers, J.M. Analysis of SMOS sea surface salinity data using DINEOF. *Remote sensing of environment* **2016**, *180*, 137–145.
11. Beckers, J.M.; Rixen, M. EOF Calculations and Data Filling from Incomplete Oceanographic Datasets. *Journal of Atmospheric and Oceanic Technology* **2003**, *20*, 1839–1856. doi:10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2.
12. Bertino, L.; Evensen, G.; Wackernagel, H. Sequential Data Assimilation Techniques in Oceanography. *International Statistical Review* **2003**, *71*, 223–241. doi:10.1111/j.1751-5823.2003.tb00194.x.
13. Lorenc, A.C.; Ballard, S.P.; Bell, R.S.; Ingleby, N.B.; Andrews, P.L.F.; Barker, D.M.; Bray, J.R.; Clayton, A.M.; Dalby, T.; Li, D.; Payne, T.J.; Saunders, F.W. The Met. Office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society* **2000**, *126*, 2991–3012. doi:10.1002/qj.49712657002.

14. Lguensat, R.; Tandeo, P.; Ailliot, P.; Pulido, M.; Fablet, R. The Analog Data Assimilation. *Monthly Weather Review* **2017**. doi:10.1175/MWR-D-16-0441.1.
15. Lguensat, R.; Huynh Viet, P.; Sun, M.; Chen, G.; Fenglin, T.; Chapron, B.; FABLET, R. Data-driven Interpolation of Sea Level Anomalies using Analog Data Assimilation.
16. Fablet, R.; Viet, P.H.; Lguensat, R. Data-Driven Models for the Spatio-Temporal Interpolation of Satellite-Derived SST Fields. *IEEE Transactions on Computational Imaging* **2017**, *3*, 647–657. doi:10.1109/TCI.2017.2749184.
17. Tandeo, P.; Ailliot, P.; Chapron, B.; Lguensat, R.; Fablet, R. The analog data assimilation: application to 20 years of altimetric data. International Workshop on Climate Informatics; , 2015; pp. 1 – 2. doi:10.13140/RG.2.1.4030.5681.
18. Egmont-Petersen, M.; de Ridder, D.; Handels, H. Image processing with neural networks—a review. *Pattern Recognition* **2002**, *35*, 2279–2301. doi:10.1016/S0031-3203(01)00178-9.
19. Fablet, R.; Ouala, S.; Herzet, C. Bilinear residual Neural Network for the identification and forecasting of dynamical systems. *SciRate* **2017**.
20. Braakmann-Folgmann, A.; Roscher, R.; Wenzel, S.; Uebbing, B.; Kusche, J. Sea Level Anomaly Prediction using Recurrent Neural Networks. *arXiv:1710.07099 [cs]* **2017**. arXiv: 1710.07099.
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]* **2015**. arXiv: 1512.03385.
22. Mairal, J.; Sapiro, G.; Elad, M. Learning multiscale sparse representations for image and video restoration. *Multiscale Modeling & Simulation* **2008**, *7*, 214–241.
23. Evensen, G. *Data Assimilation*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009. doi:10.1007/978-3-642-03711-5.
24. Bertino, L.; Evensen, G.; Wackernagel, H. Sequential Data Assimilation Techniques in Oceanography. *International Statistical Review* **2007**, *71*, 223–241. doi:10.1111/j.1751-5823.2003.tb00194.x.
25. Anderson, J.L.; Anderson, S.L. A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. *Monthly Weather Review* **1999**, *127*, 2741–2758. doi:10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.
26. Houtekamer, P.L.; Mitchell, H.L. Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review* **1998**, *126*, 796–811. doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.
27. Gaspari, G.; Cohn, S.E. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society* **1999**, *125*, 723–757. doi:10.1002/qj.49712555417.
28. Houtekamer, P.L.; Mitchell, H.L. A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review* **2001**, *129*, 123–137. doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.
29. Bocquet, M. Localization and the iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society* **2016**, *142*, 1075–1089. doi:10.1002/qj.2711.
30. Pannekoucke, O.; Emili, E.; Thual, O. Modelling of local length-scale dynamics and isotropizing deformations. *Quarterly Journal of the Royal Meteorological Society* **2013**, *140*, 1387–1398. doi:10.1002/qj.2204.
31. Pannekoucke, O.; Ricci, S.; Barthelemy, S.; Ménard, R.; Thual, O. Parametric Kalman filter for chemical transport models. *Tellus A: Dynamic Meteorology and Oceanography* **2016**, *68*, 31547. doi:10.3402/tellusa.v68.31547.
32. Cohn, S.E. Dynamics of Short-Term Univariate Forecast Error Covariances. *Monthly Weather Review* **1993**, *121*, 3123–3149. doi:10.1175/1520-0493(1993)121<3123:DOSTUF>2.0.CO;2.
33. LeCun, Y.; Haffner, P.; Bottou, L.; Bengio, Y. Object Recognition with Gradient-Based Learning. In *Shape, Contour and Grouping in Computer Vision*; Forsyth, D.A.; Mundy, J.L.; di Gesú, V.; Cipolla, R., Eds.; Lecture Notes in Computer Science, Springer Berlin Heidelberg: Berlin, Heidelberg, 1999; pp. 319–345. doi:10.1007/3-540-46805-6_19.
34. Ouala, S.; Herzet, C.; Fablet, R. Sea surface temperature prediction and reconstruction using patch-level neural network representations. *arXiv:1806.00144 [cs, stat]* **2018**. arXiv: 1806.00144.
35. Fablet, R.; Verron, J.; Mourre, B.; Chapron, B.; Pascual, A. Improving Mesoscale Altimetric Data From a Multitracer Convolutional Processing of Standard Satellite-Derived Products. *IEEE Transactions on Geoscience and Remote Sensing* **2018**, *56*, 2518–2525. doi:10.1109/TGRS.2017.2750491.