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

To cite this version:
Said Ouala, Ronan Fablet, Cédric Herzet, Bertrand Chapron, Ananda Pascual, et al.. Neural-Network-based Kalman Filters for the Spatio-Temporal Interpolation of Satellite-derived Sea Surface Temperature. 2018. <hal-01896654>

HAL Id: hal-01896654
https://hal-imt-atlantique.archives-ouvertes.fr/hal-01896654
Submitted on 16 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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


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]:

\[
\begin{align*}
    x_{t+1} &= \mathcal{F}(x_t) + \eta_t \quad (1) \\
    y_{t+1} &= \mathcal{H}(x_{t+1}) + \epsilon_t \\
\end{align*}
\]

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 \( x \). \( \eta_t \) and \( \epsilon_t \) 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:

\[
\begin{align*}
    p(x_{t+1}|y_1, ..., y_t) &= \mathcal{N}(x_{t+1}^{-}, \Sigma_{t+1}^{-}) \quad (3) \\
    p(x_{t+1}|y_1, ..., y_{t+1}) &= \mathcal{N}(x_{t+1}^{+}, \Sigma_{t+1}^{+}) \quad (4)
\end{align*}
\]

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

\[
\begin{align*}
    x_{t+1}^{-} &= \mathcal{F} x_t^{-} \quad (5) \\
    \Sigma_{t+1}^{-} &= \mathcal{F} \Sigma_{t}^{-} \mathcal{F}^T + Q_t \quad (6) \\
    x_{t+1}^{+} &= x_{t+1}^{-} + K_{t+1}[y_{t+1} - H_{t+1}x_{t+1}^{-}] \quad (7) \\
    \Sigma_{t+1}^{+} &= \Sigma_{t+1}^{-} - K_{t+1} H_{t+1} \Sigma_{t+1}^{-} \Sigma_{t+1}^{-1} H_{t+1}^T \quad (8)
\end{align*}
\]

with

\[
K_{t+1} = \Sigma_{t+1}^{-} H_{t+1}^T [H_{t+1} \Sigma_{t+1}^{-1} H_{t+1}^T + R_t]^{-1}.
\]

Here \( \mathcal{F} \) 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 \( \Sigma_{t+1}^{-} \) 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 \( \Sigma_{t+1}^{+} \) 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 $\Sigma_{t+1}^-$ by a sample mean and covariance matrix computed by propagating an ensemble of $M$ members, $\{x_{t+1}^-\}_{i=1}^M$, using the dynamical model $F$.

$$x_{t+1}^- = F(x_{t}^+, i \in \{0, ..., N\})$$  \hspace{0.5cm} (10)

$$\Sigma_{t+1} = \frac{1}{N-1}D_{t+1}D_{t+1}^\top$$  \hspace{0.5cm} (11)

$$D_{t+1} = [x_{t+1}^- - x_{t+1}^+ ... x_{t+1}^- - x_{t+1}^-]$$  \hspace{0.5cm} (12)

$$x_{t+1}^+ = x_{t+1}^- + K_{t+1}[y_{t+1} - H_{t+1}x_{t+1}^-]$$  \hspace{0.5cm} (13)

$$K_{t+1} = \Sigma_{t+1}H_{t+1}^\top[H_{t+1}\Sigma_{t+1}H_{t+1}^\top + R_t]^{-1}$$  \hspace{0.5cm} (14)

$$\Sigma_{t+1}^+ = \Sigma_{t+1} - K_{t+1}H_{t+1}\Sigma_{t+1}^-$$  \hspace{0.5cm} (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 $x_t|\chi_t$, 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:

$$x_{t+1}^- = \mathcal{F}(x_t^+)$$  \hspace{0.5cm} (16)

$$\Sigma_{t+1}^- = \mathcal{F}_\Sigma(x_t^+, \Sigma_t^+)$$  \hspace{0.5cm} (17)

with $x_{t+1}^-$ and $\Sigma_{t+1}^-$ 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 $\Sigma_t^+$ 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:

\[ x_{t+1}^+ = x_{t+1}^- + K_{t+1} [y_{t+1} - H_{t+1} x_{t+1}^-] \]  
\[ K_{t+1} = \Sigma_{t+1}^- H_{t+1}^T [H_{t+1} \Sigma_{t+1}^- H_{t+1}^T + R_{t+1}]^{-1} \]

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. 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 \( F \), 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 \( x_{P, s} \), 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 \( B \) according to:

\[ z_{P, s} = x_{P, s} B^T \]  

with \( z_{P, s} \) the EOF decomposition of the patch \( x_{P, s} \). The EOF decomposition matrix \( B \) 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 \( F_{P, s}, s \in [1, \ldots, N_p] \) to predict
\( z_{P, s+1} \). For patch-level models \( F_{P, s} \), we consider residual architectures [21] with a bilinear
parameterization [19];
- The third layer is a reconstruction network \( F_r \). It combines the predicted patches \( x_{P, s} =
\z_{P, s} B, s \in [1, \ldots, N_p] \) to reconstruct the output field \( x_t \). This reconstruction network \( F_r \)
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 \( F \), we apply a two-step procedure. We first learn the local
dynamical models \( F_{P, s}, s \in [1, \ldots, N_p] \) based on the minimization of the EOF-patch based forecasting
error. The reconstruction network \( F_r \) 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 $\mathcal{F}_\Sigma^P$, 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 $\mathcal{F}_\Sigma^P$ from states of the training dataset corrupted by an additive Gaussian perturbation with a covariance structure $\Sigma_0$. Here, $\Sigma_0$ is given by the empirical covariance of the EOF patches for the entire training dataset. Overall, for a given patch $P_s$, we parameterize $\mathcal{F}_\Sigma^P$ the restriction of covariance $\mathcal{F}_\Sigma$ onto patch $P_s$ as:

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

with $\Psi(\Sigma_{P_s,t-1}, \Sigma_0)$ 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 $x^r_{t-1}$ and covariance $\Sigma^r_{t-1}$, we first compute the forecasted Gaussian approximation at time $t$ with mean field $\mathcal{F}(x^r_{t-1})$ and patch-based covariance $\mathcal{F}_\Sigma^P(x^r_{t-1}, \Sigma^r_{t-1})$. The assimilation of the new observation $y_t$ is performed at a patch-level. For each patch $P_s$, we update the patch-level mean $x^r_{P_s,t}$ and covariance $\Sigma^r_{P_s,t}$ using Kalman recursion (8) with observation $y_{P_s,t}$. We then combine these patch-level updates to obtain global mean $x^r_t$ and covariance $\Sigma^r_t$. Whereas we compute global mean $x^r_t$ using trained reconstruction network $\mathcal{F}_r$, $\Sigma^r_t$ 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 neighboring 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 $F_r$. 
We exploit patch-level representations with non-overlapping 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, 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 $\mathcal{F}$ is a convolutional neural network with 3 convolutional layers. The first two layers comprise 64 filters of size $3 \times 3$ with a ReLU activation and the last layer is a linear convolutional layer with one filter. Regarding covariance model $\mathcal{F}_D$, we consider a diagonal covariance model within each patch. Each element of diagonal involves a 3-layer MLP with 4 neurons and ReLU activation functions on the hidden layers and a softmax 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 $20 \times 20$ 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.

4.2. Experimental setting

The proposed neural-network-based Kalman scheme involves the following parameter setting. We exploit patch-level representations with non-overlapping 20 $\times$ 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 $\mathcal{F}_{P}^{\mathcal{P}_s}$, 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 $\mathcal{F}$ is a convolutional neural network with 3 convolutional layers. The first two layers comprise 64 filters of size $3 \times 3$ with a ReLU activation and the last layer is a linear convolutional layer with one filter. Regarding covariance model $\mathcal{F}_D$, we consider a diagonal covariance model within each patch. Each element of diagonal involves a 3-layer MLP with 4 neurons and ReLU activation functions on the hidden layers and a softmax 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 $20 \times 20$ 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.
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 $F^P$. 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 $F$.

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...
<table>
<thead>
<tr>
<th>Assimilation method</th>
<th>Patch1</th>
<th>Patch2</th>
<th>Patch3</th>
<th>Patch4</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAF EnKF</td>
<td>0.50</td>
<td>0.25</td>
<td>0.22</td>
<td>0.39</td>
</tr>
<tr>
<td>Bi-NN-EnKF</td>
<td>0.55</td>
<td>0.23</td>
<td>0.22</td>
<td>0.30</td>
</tr>
<tr>
<td>Bi-NN-NNKF-EOF</td>
<td><strong>0.46</strong></td>
<td><strong>0.20</strong></td>
<td><strong>0.19</strong></td>
<td><strong>0.27</strong></td>
</tr>
</tbody>
</table>

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 \( F^p \) 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 $\mathcal{F}$, 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.

<table>
<thead>
<tr>
<th>Model</th>
<th>Forecasting RMSE ($°C$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t + h$</td>
</tr>
<tr>
<td>PB-NN</td>
<td>0.48</td>
</tr>
<tr>
<td>LAF</td>
<td>0.50</td>
</tr>
<tr>
<td>GAF</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Table 2. Forecasting experiment for several prediction time steps

<table>
<thead>
<tr>
<th>Model</th>
<th>Entire map</th>
<th>Missing data areas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>Correlation</td>
</tr>
<tr>
<td></td>
<td>$SST$($°C$)</td>
<td>$\nabla SST$($°C/°$)</td>
</tr>
<tr>
<td>PB-NNKF-EOF</td>
<td>0.33</td>
<td>0.13</td>
</tr>
<tr>
<td>PB-NNKF</td>
<td>0.51</td>
<td>0.18</td>
</tr>
<tr>
<td>LAF-EnKF</td>
<td>0.43</td>
<td>0.16</td>
</tr>
<tr>
<td>GAF-EnKF</td>
<td>0.48</td>
<td>0.19</td>
</tr>
<tr>
<td>PB-VE-DINEOF</td>
<td>0.54</td>
<td>0.20</td>
</tr>
<tr>
<td>OI</td>
<td>0.76</td>
<td>0.25</td>
</tr>
</tbody>
</table>

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. and B.C. discussed the experiments;
R.L. and R.F. wrote the paper. B.C. 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,
Region 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*
2. Donlon, C.J.; Martin, M.; Stark, J.; Roberts-Jones, J.; Fiedler, E.; Wimmer, W. The Operational Sea Surface
3. Le Traon, P.Y.; Nadal, F.; Ducet, N. An improved mapping method of multisatellite altimeter data. *Journal of
4. Droghei, R.; Buongiorno Nardelli, B.; Santoleri, R. A New Global Sea Surface Salinity and Density Dataset
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
6. Ducet, N.; Le Traon, P.Y.; Reverdin, G. Global high-resolution mapping of ocean circulation from
spatial interpolation of CTD and ADCP data. *Deep Sea Research Part I: Oceanographic Research Papers*
8. Ping, B.; Su, F.; Meng, Y. An Improved DINEOF Algorithm for Filling Missing Values in Spatio-Temporal
10. Alvera-Azcárate, A.; Barth, A.; Parard, G.; Beckers, J.M. Analysis of SMOS sea surface salinity data using
11. Beckers, J.M.; Rixen, M. EOF Calculations and Data Filling from Incomplete
A.M.; Dalby, T.; Li, D.; Payne, T.J.; Saunders, F.W. The Met. Office global three-dimensional variational


