



Data-driven Interpolation of Sea Level Anomalies using Analog Data Assimilation

Redouane Lguensat, Phi Huynh Viet, Miao Sun, Ge Chen, Tian Fenglin,
Bertrand Chapron, Ronan Fablet

► To cite this version:

Redouane Lguensat, Phi Huynh Viet, Miao Sun, Ge Chen, Tian Fenglin, et al.. Data-driven Interpolation of Sea Level Anomalies using Analog Data Assimilation. Remote Sensing, 2019, 10.3390/rs11070858 . hal-01609851

HAL Id: hal-01609851

<https://hal.science/hal-01609851>

Submitted on 4 Oct 2017

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.

Data-driven Interpolation of Sea Level Anomalies using Analog Data Assimilation

Redouane Lguensat^{a,*}, Phi Huynh Viet^a, Miao Sun^b, Ge Chen^b, Tian Fenglin^b, Bertrand Chapron^c, Ronan Fablet^a

^a*IMT Atlantique, Lab-STICC, Université Bretagne Loire*

^b*Department of Marine Technology, College of Information Science and Engineering, Ocean University of China, Qingdao, China*

^c*Laboratoire d'Océanographie Physique et Spatiale, Centre de Brest, IFREMER, Plouzané, France*

Abstract

Despite the well-known limitations of Optimal Interpolation (OI), it remains the conventional method to interpolate Sea Level Anomalies (SLA) from altimeter-derived along-track data. In consideration of the recent developments of data-driven methods as a means to better exploit large-scale observation, simulation and reanalysis datasets for solving inverse problems, this study addresses the improvement of the reconstruction of higher-resolution SLA fields using analog strategies. The reconstruction is stated as an analog data assimilation issue, where the analog models rely on patch-based and EOF-based representations to circumvent the curse of dimensionality. We implement an Observation System Simulation Experiment in the South China sea. The reported results show the relevance of the proposed framework with a significant gain in terms of root mean square error for scales below 100km. We further discuss the usefulness of the proposed analog model as a means

*corresponding author, email: redouane.lguensat@imt-atlantique.fr

to exploit high-resolution model simulations for the processing and analysis of current and future satellite-derived altimetric data.

Keywords: Analog Data Assimilation, Sea Level Anomaly, Sea Surface Height, Interpolation, Data-driven methods

1. Introduction

The past twenty years have witnessed a deluge of ocean satellite data, such as sea surface height, sea surface temperature, ocean color, ocean current, sea ice, etc. This has helped building big databases of valuable information and represents a major opportunity for the interplay of ideas between ocean remote sensing community and the data science community. Exploring machine learning methods in general and non-parametric methods in particular is now feasible and is increasingly drawing the attention of many researchers (Zhang et al., 2016; Lary et al., 2016).

More specifically, analog forecasting (Lorenz, 1969) which is among the earliest statistical methods explored in geoscience benefits from recent advances in data science. In short, analog forecasting is based on the assumption that the future state of a system can be predicted throughout the successors of past (or simulated) similar situations (called analogs). The amount of currently available remote sensing and simulation data offers analog methods a great opportunity to catch up their early promises. Several recent works involving applications of analog forecasting methods in geoscience fields contribute in the revival of these methods, recent applications comprise the prediction of soil moisture anomalies (McDermott and Wikle, 2015), the prediction of sea-ice anomalies (Comeau et al., 2017), rainfall nowcasting (Atencia

21 and Zawadzki, 2015), stochastic weather generators (Yiou, 2014), etc. One
 22 may also cite methodological developments such as dynamically-adapted ker-
 23 nels (Zhao and Giannakis, 2014a) and novel parameter estimation schemes
 24 (Horton et al., 2017). Importantly, analog strategies have recently been ex-
 25 tended to address data assimilation issues within the so-called *analog data*
 26 *assimilation* (AnDA) (Lguensat et al., 2017), where the dynamical model
 27 is stated as an analog forecasting model and combined to state-of-the-art
 28 stochastic assimilation procedures such as Ensemble Kalman filters. The re-
 29 cent applications to high-dimensional fields in Fablet et al. (2017) provide
 30 the methodological background for this study.

31 Producing time-continuous and gridded maps of Sea Surface Height (SSH)
 32 is a major challenge in ocean remote sensing with important consequences
 33 on several scientific fields from weather and climate forecasting to opera-
 34 tional needs for fisheries management and marine operations (*e.g.* Hardman-
 35 Mountford et al. (2003)). The reference gridded SSH product commonly used
 36 in the literature is distributed by the Copernicus Marine and Environment
 37 Monitoring Service (CMEMS) (formerly distributed by AVISO). This prod-
 38 uct relies on the interpolation of irregularly-spaced along-track data using an
 39 Optimal Interpolation (OI) method (Le Traon et al., 1998; Bretherton et al.,
 40 1976). While OI is relevant for the retrieval of horizontal scales of SSH fields
 41 greater than $\approx 100km$, its Gaussian assumptions cause the small scales of the
 42 SSH fields to be smoothed. This limitation makes it impossible to resolve
 43 finer-scale processes (typically from a few tens of kilometers to $\approx 100km$)
 44 which may be revealed by along-track altimetric data. This has led to a
 45 variety of research studies to improve the reconstruction of the altimetric

46 fields. One may cite both methodological alternatives to OI, for instance
 47 locally-adapted convolutional models (Fablet et al., 2016) and variational as-
 48 simulation schemes using model-driven dynamical priors (Ubelmann et al.,
 49 2014), as well as studies exploring the synergy between different sea surface
 50 tracers, especially the synergy between SSH and SST (Sea Surface Temper-
 51 ature) fields and Surface Quasi-Geostrophic dynamics (Fablet et al., 2016;
 52 Klein et al., 2009; Isern-Fontanet et al., 2006, 2014; Turiel et al., 2009b,a).

53 In this work, we build upon our recent advances in analog data assimi-
 54 lation and its application to high-dimensional fields (Lguensat et al., 2017;
 55 Fablet et al., 2017). We develop an analog data assimilation model for the
 56 reconstruction of SLA fields from along-track altimeter data. It relies on a
 57 patch-based and EOF-constrained representation of the SLA fields. Using
 58 OFES numerical simulations (Masumoto et al., 2004; Sasaki et al., 2008),
 59 we design an Observation System Simulation Experiment (OSSE) for a case-
 60 study in the South China sea using real along-track sampling patterns of
 61 spaceborne altimeters. Using the resulting groundtruthed dataset, we per-
 62 form a qualitative and quantitative evaluation of the proposed scheme, in-
 63 cluding comparisons to state-of-the-art schemes.

64 The remainder of the paper is organized as follows: Section 2 presents the
 65 different datasets used in this paper to design an OSSE, Section 3 gives in-
 66 sights on the classical methods used for mapping SLA from along track data,
 67 Section 4 introduces the proposed analog data assimilation model. Experi-
 68 mental results for the considered OSSE are shown in Section 5, and Section
 69 6 further discuss the key aspects of this work.

2. Data: OFES (OGCM for the Earth Simulator)

An Observation System Simulation Experiment (OSSE) based on numerical simulations is considered to assess the relevance of the proposed analog assimilation framework. Our OSSE uses these numerical simulations as a groundtruthed dataset from which simulated along-track data are produced. We describe further the data preparation setup in the following sections.

2.1. Model simulation data

The Ocean General Circulation Model (OGCM) for the Earth Simulator (OFES) is considered in this study as the true state of the ocean. The simulation data is described in Masumoto et al. (2004); Sasaki et al. (2008). The coverage of the model is 75°S - 75°N with a horizontal resolution of $1/10^{\circ}$. 34 years (1979-2012) of 3-daily simulation of SSH maps are considered, we proceed to a subtraction of a temporal mean to obtain SLA fields. In this study, our region of interest is located in the South China Sea (105°E to 117°E , 5°N to 25°N). This dataset is split into a training dataset corresponding to the first 33 years (4017 SLA maps) and a test dataset corresponding to the last year of the time series (122 SLA maps).

2.2. Along track data

We consider a realistic situation with a high rate of along track data. More precisely we use along-track data positions registered in 2014 where 4 satellites (Jason2, Cryosat2, Saral/AltiKa, HY-2A) were operating. Data is distributed by Copernicus Marine and Environment Monitoring Service (CMEMS).

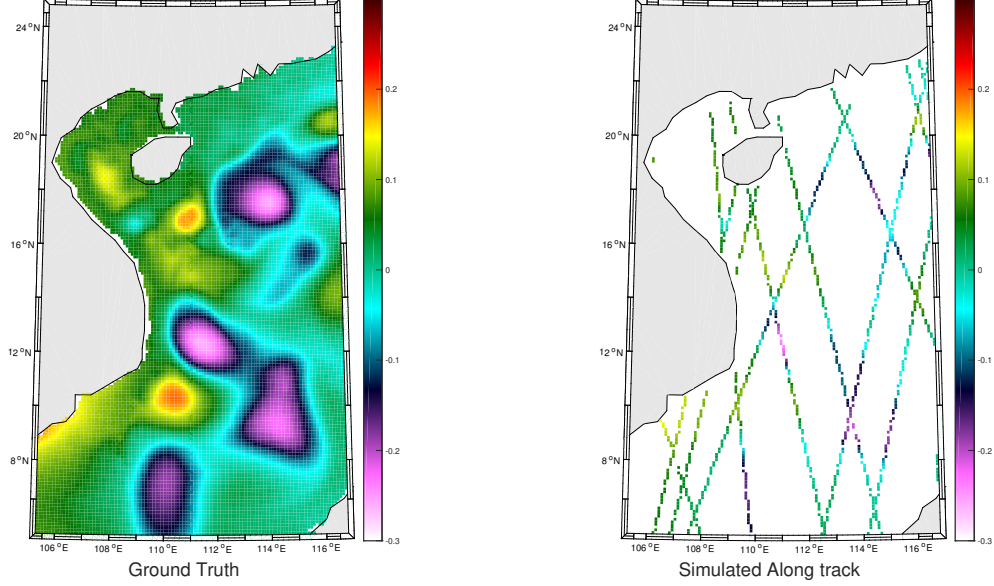


Figure 1: An example of a ground-truth SLA field in the considered region and its associated simulated pseudo-along track.

93 From the reference 3-daily SLA dataset and real along-track data posi-
 94 tions, we generate simulated along-track data from the sampling of a ref-
 95 erence SLA field: more precisely, for a given along-track point, we sample
 96 the closest position of the $1/10^\circ$ regular model grid at the closest time step
 97 of the 3-daily model time series. As we consider a 3-daily assimilation time
 98 step (see Section 2.1 for details), we create a 3-daily pseudo-observation field,
 99 to be fed directly to the assimilation model. As sketched in Figure 2, for a
 100 given time t , we combine all along-track positions for times $t - 1, t$ and $t + 1$
 101 to create an along-track pseudo-observation field at time t . We denote by
 102 $s3dAT$ the simulated 3-daily time series of along-track pseudo-observation
 103 fields. An example of these fields is given in Figure 1.

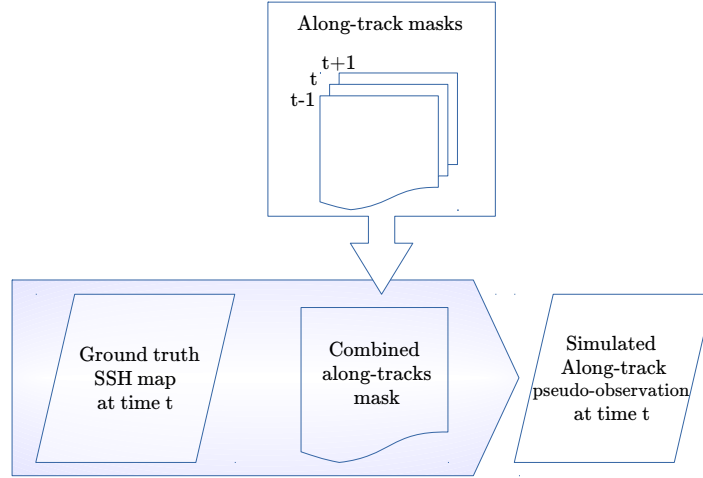


Figure 2: Sketch of the creation of simulated along-track data at a given time t

104 3. Problem statement and related work

105 3.1. Data assimilation and optimal interpolation

106 Data assimilation consists in estimating the true state of a physical vari-
 107 able $\mathbf{x}(t)$ at a specific time t , by combining i) equations governing the dynam-
 108 ics of the variable, ii) available observations $\mathbf{y}(1, \dots, T)$ measuring the variable
 109 and iii) a background or first guess on its initial state \mathbf{x}^b . The estimated state
 110 is generally called the analyzed state and noted by \mathbf{x}^a . Data assimilation is
 111 a typical example of inverse problems, and similar formulations are known
 112 to the statistical signal processing community through optimal control and
 113 estimation theory (Bocquet et al., 2010). We adopt here the unified notation
 114 of Ide et al. (1997) and formulate the problem as a stochastic system in the
 115 following:

$$\begin{cases} \mathbf{x}(t) = \mathcal{M}(\mathbf{x}(t-1)) + \eta(t), \\ \mathbf{y}(t) = \mathcal{H}(\mathbf{x}(t)) + \epsilon(t). \end{cases} \quad (1) \quad (2)$$

Equation 1 represents the dynamical model governing the evolution of state \mathbf{x} through time, while η is a Gaussian centered noise of covariance \mathbf{Q} that models the process error. Equation 2 explains the relationship between the observation $\mathbf{y}(t)$ and the state to be estimated $\mathbf{x}(t)$ through the operator \mathcal{H} . The uncertainty of the observation model is represented by the ϵ error, considered here to be Gaussian centered and of covariance \mathbf{R} . We assume that ϵ and η are independent and that \mathbf{Q} and \mathbf{R} are known. Two main approaches are generally considered for the mathematical resolution of the system (1)-(2), namely, variational data assimilation and stochastic data assimilation. They differ in the way they infer the analyzed state x^a , the first is based on the minimization of a certain cost function while the latter aims to obtain an optimal a posteriori estimate. We encourage the reader to consider the book of Asch et al. (2016) for detailed insights on the various aspects and methods of data assimilation.

A popular data assimilation algorithm that is largely used in the literature to grid sea level anomalies from along-track data is called Optimal Interpolation (OI) (e.g. Le Traon et al. (1998); De Mey and Robinson (1987)), this algorithm is also the method adopted in CMEMS altimetry product. Optimal Interpolation (OI) aims at finding the Best Linear Unbiased Estimator (BLUE) of a field \mathbf{x} given irregularly sampled observations \mathbf{y} in space and time and a background prior \mathbf{x}^b . The multivariate OI equations were derived in Gandin (1966) for meteorology and numerous applications in oceanography have been reported since the early work of Bretherton et al. (1976).

139 Supposing that the background state \mathbf{x}^b has covariance B , and the observa-
 140 tion operator is linear $\mathcal{H} = \mathbf{H}$, the analyzed state \mathbf{x}^a and the analyzed error
 141 covariance \mathbf{P}^a can be calculated using the following OI set of equations:

$$\begin{cases} \mathbf{K} = \mathbf{B}\mathbf{H}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} & \text{called the Kalman gain} & (3) \\ \mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^b) & & (4) \\ \mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} & & (5) \end{cases}$$

142 It worths mentioning that Lorenc (1986) showed that OI is closely related
 143 to the 3D-Var variational data assimilation algorithm which obtains \mathbf{x}^a by
 144 minimizing the following cost function:

$$J(x) = (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}) \quad (6)$$

145 While OI had been shown to successfully retrieve large-scale structures in
 146 the ocean ($\geq 150km$), a well-known limitation of OI is that the Gaussian-
 147 like covariance error matrices smooths out the small-scale information (*e.g.*
 148 mesoscale eddies) (Ubelmann et al., 2014). OI would then underexploit high
 149 resolution altimeter data in the context of future altimetry missions, which
 150 urges to put efforts in trying to improve the method (*e.g.* Escudier et al.
 151 (2013a)) or find other alternatives.

152 3.2. Analog data assimilation

153 Endorsed by the recent development in data-driven methods and data
 154 storage capacities, the Analog Data Assimilation (AnDA) was introduced as
 155 an alternative to classical model-driven data assimilation under one or more
 156 of the following situations (Lguensat et al., 2017):

- 157 • The model is inconsistent with observations
- 158 • The cost of the model integration is high computationally
- 159 • (mandatory) The availability of large datasets of past dynamics of the
 160 variables to be estimated. These datasets are hereinafter called catalogs
 161 and denoted by \mathcal{C} . The catalog is organized in a two-column dictionary
 162 where each state of the system is associated with its successor in time,
 163 forming a set of couples $(\mathcal{A}_i, \mathcal{S}_i)$ where \mathcal{A}_i is called the *analog* and \mathcal{S}_i
 164 its *successor*.

165 Given the considerations above, AnDA resorts to evaluating filtering,
 166 resp. smoothing, posterior likelihood, *i.e.* the distribution of the state to
 167 be estimated $\mathbf{x}(t)$ at time t , given past and current observations $\mathbf{y}(1, \dots, t)$,
 168 resp. given all the available observation $\mathbf{y}(1, \dots, T)$. This evaluation relies on
 169 the following state-space model:

$$\begin{cases} \mathbf{x}(t) = \mathcal{F}(\mathbf{x}(t-1)) + \eta(t), & (7) \\ \mathbf{y}(t) = \mathcal{H}(\mathbf{x}(t)) + \epsilon(t). & (8) \end{cases}$$

170 The difference between AnDA and classical data assimilation resides in
 171 the transition model equation 7. The counterpart of a model-driven operator
 172 \mathcal{M} of Equation 1 is here the operator \mathcal{F} which refers to the considered data-
 173 driven operator, so called, the analog forecasting operator. This operator
 174 makes use of the available catalog \mathcal{C} and assumes that the state forecast can
 175 be inferred from similar situations in the past. Provided the definitions of the
 176 analogs and successors given above, the derivation of this operator resorts
 177 to characterizing the transition distribution *i.e.* $p(\mathbf{x}(t)|\mathbf{x}(t-1))$. Following
 178 Lguensat et al. (2017), a Gaussian conditional distribution is adopted:

$$\mathbf{x}(t)|\mathbf{x}(t-1) \sim \mathcal{N}(\mu_t, \Sigma_t) \quad (9)$$

where $\mathcal{N}(\mu_t, \Sigma_t)$ is a Gaussian distribution of mean μ_t and covariance Σ_t . These parameters of the Gaussian distribution are calculated using the result of a K nearest neighbors search. The K nearest neighbors (analogs) $\mathcal{A}_{k \in (1, \dots, K)}$ of state $\mathbf{x}(t-1)$ and their successors $\mathcal{S}_{k \in (1, \dots, K)}$, along with a weight associated to each pair $(\mathcal{A}_k, \mathcal{S}_k)$ are used to calculate μ_t and Σ_t , the forecast state $\mathbf{x}(t)$ is then sampled from $\mathcal{N}(\mu_t, \Sigma_t)$. The weights are defined using a Gaussian kernel \mathcal{K}_G .

$$\mathcal{K}_G(u, v) = \exp\left(-\frac{\|u - v\|^2}{\sigma}\right), \quad (10)$$

Scale parameter σ is locally-adapted to the median value of the K distances $\|x(t-1) - \mathcal{A}_k\|^2$ to the K analogs. Other types of kernels might be considered (e.g. Zhao and Giannakis (2014b); McDermott and Wikle (2015)), investigating kernel choice is out of the scope of this paper.

The mean and the covariance of the transition distribution might be calculated following several strategies. We consider in this work the three analog forecasting operators introduced in AnDA (Lguensat et al., 2017):

- **Locally-constant operator:** Mean μ_t and covariance Σ_t are given by the weighted mean and covariance of the K successors $\mathcal{S}_{k \in (1, \dots, K)}$.
- **Locally-incremental operator:** Here, the increments between the K analogs and their corresponding successors are calculated $\mathcal{S}_{k \in (1, \dots, K)} - \mathcal{A}_{k \in (1, \dots, K)}$. The weighted mean of the K increments is then added to

198 the $\mathbf{x}(t-1)$ to obtain μ_t . While Σ_t results in the weighted covariance
 199 of these differences.

200 • **Locally-linear operator:** A weighted least-square estimation of the
 201 linear regression of the state at time t given the state at time $t-1$ is
 202 performed based on the K pairs $(\mathcal{A}_k, \mathcal{S}_k)$. The parameters of the linear
 203 regression are then applied to state $\mathbf{x}(t-1)$ to obtain μ_t . Covariance Σ_t
 204 is represented by the covariance of the residuals of the fitted weighted
 205 linear regression.

206 The application of the AnDA framework faces the curse of dimensionality
 207 *i.e.* the search of analogs is highly affected by the dimensionality of the
 208 problem and can be irrelevant for dimensions above 20 (Lguensat et al.,
 209 2017). As proposed in Fablet et al. (2017), the extension of AnDA models to
 210 high-dimensional fields may then rely on turning the global assimilation issue
 211 into a series of lower-dimensional ones. We consider here an approach similar
 212 to Fablet et al. (2017) using a patch-based and EOF-based representation of
 213 the 2D fields, *i.e.* the 2D fields are decomposed into a set of overlapping
 214 patches, each patch being projected onto an EOF space. Analog strategies
 215 then apply to patch-level time series in the EOF space.

216 Overall, as detailed in the following section, the proposed analog data
 217 assimilation model for SLA fields relies on three key components: a patch-
 218 based representation of the SLA fields, the selection of a kernel to retrieve
 219 analogs and the specification of a patch-level analog forecasting operator.

220 4. Analog reconstruction for altimeter-derived SLA

221 4.1. Patch-based state-space formulation

222 As stated above, OI may be considered as an efficient model-based method
 223 to recover large-scale structures of SLA fields. Following Fablet et al. (2017),
 224 this suggests considering the following two-scale additive decomposition:

$$X = \bar{X} + dX + \xi \quad (11)$$

225 where \bar{X} is the large-scale component of the SLA field, typically issued from
 226 an optimal interpolation, dX the fine-scale component of the SLA field we
 227 aim to reconstruct and ξ remaining unresolved scales.

228 The reconstruction of field dX involves a patch-based and EOF-based
 229 representation. It consists in regarding field dX as a set of $P \times P$ overlap-
 230 ping patches (*e.g.* $2^\circ \times 2^\circ$). This set of patches is referred to as \mathcal{P} , and we
 231 denote by \mathcal{P}_s the patch centered on position s . After building a catalog $\mathcal{C}_{\mathcal{P}}$
 232 of patches from the available dataset of residual fields $X - \bar{X}$ (see Section
 233 3.2), we proceed to an EOF decomposition of each patch in the catalog. The
 234 reconstruction of field $dX(\mathcal{P}_s, t)$ at time t is then stated as the analog assim-
 235 ilation of the coefficients of the EOF decomposition in the EOF space given
 236 an observation series in the patch space. Formally, $dX(\mathcal{P}_s, t)$ decomposes as
 237 a linear combination of a number N_E of EOF basis functions:

$$dX(\mathcal{P}_s, t) = \sum_{k=1}^{N_E} \alpha_k(s, t) EOF_k \quad (12)$$

238 with EOF_k the k^{th} EOF basis and $\alpha_k(s, t)$ the corresponding coefficient for
 239 patch \mathcal{P}_s at time t . Let us denote by $\Phi(\mathcal{P}_s, t)$ the vector of the N_E coefficients

240 $\alpha_k(s, t)$. This vector represents the projection of $dX(\mathcal{P}_s, t)$ in the lower-
 241 dimensional EOF space.

242 4.2. Patch-based analog dynamical models

243 Given the considered patch-based representation of field dX , the proposed
 244 patch-based analog assimilation scheme involves a dynamical model stated
 245 in the EOF space. Formally, Equation 9 leads to the following Gaussian
 246 conditional distribution in the EOF space

$$\Phi(\mathcal{P}_s, t) | \Phi(\mathcal{P}_s, t - 1) \sim \mathcal{N}(\mu(s, t), \Sigma(s, t)) \quad (13)$$

247 We consider the three analog forecasting operators presented in Section 3.2,
 248 namely, the locally-constant, the locally incremental and the locally-linear.
 249 The calculation of the weights associated to each analog-successor pair relies
 250 on a Gaussian kernel \mathcal{K}_G (Equation 10). The search for analogs in the N_E -
 251 dimensional patch space (in practice, N_E ranges from 5 to 20) ensures a better
 252 accuracy in the retrieval of relevant analogs compared to a direct search in
 253 the high-dimensional space of state dX . It also reduces the computational
 254 complexity of the proposed scheme.

255 Another important extension of the current study is the possibility of ex-
 256 ploiting auxiliary variables with the state vector Φ in the analog forecasting
 257 models. Such variables may be considered in the search for analogs as well as
 258 regression variables in locally-linear analog setting. Regarding the targeted
 259 application to the reconstruction of SSH fields and the proposed two-scale
 260 decomposition (Equation 11), two types of auxiliary variables seem to be
 261 of interest: the low-resolution component \bar{X} to take into account inter-scale

relationship (Fablet et al., 2016), and Sea Surface Temperature (SST) with respect to the widely acknowledged SST-SSH synergies (Fablet et al., 2016; Klein et al., 2009; Isern-Fontanet et al., 2014). We also apply patch-level EOF-based decompositions to include both types of variables in the considered analog forecasting models (Equation 13).

4.3. Numerical resolution

Given the proposed analog assimilation model, the proposed scheme first relies on the creation of patch-level catalogs from the training dataset. This step requires the computation of a training dataset of fine scale data $dX_{training}$, this is done by subtracting a large-scale component $\bar{X}_{training}$ from the original training dataset. Here, we consider the large-scale component of training data to be the result of a global¹ EOF-based reconstruction using a number of EOF components that retains 95% of the dataset variance, which accounts for horizontal scales up to $\sim 100\text{km}$. This global EOF-based decomposition provides a computationally-efficient means for defining large-scale component \bar{X} . This EOF-based decomposition step is followed by the extraction of overlapping patches for all variables of interest, namely $\bar{X}_{training}$, $dX_{training}$ and potential auxiliary variables, and the identification of the EOF basis functions from the resulting raw patch datasets. This leads to the creation of a patch-level catalog $\mathcal{C}_{\mathcal{P}}$ from the EOF-based representations of each patch.

Given the patch-level catalog, the algorithm applied for the mapping SLA fields from along-track data, referred to PB-AnDA (for Patch-Based AnDA),

¹By global, we mean here an EOF decomposition over the entire case study region, by contrast to the patch-level decomposition considered in the analog assimilation setting.

284 involves the following steps:

- 285 • the computation of the large-scale component \bar{X} , here, we consider
286 the result of optimal interpolation (OI) projected onto the global EOF
287 basis functions.
- 288 • the decomposition of the case study region into overlapping $P \times P$
289 patches, here, 20×20 patches
- 290 • For each patch position s , the application of an analog data assimilation
291 scheme, namely the Analog Ensemble Kalman Smoother (AnEnKS)
292 (Lguensat et al., 2017), for patch \mathcal{P}_s of field dX . As stated in (13), the
293 assimilation is performed in the EOF space, *i.e.* for EOF decomposition
294 $\Phi(\mathcal{P}_s, t)$, using the operator derived from EOF-based reconstruction
295 (12) and decomposition (11) as observation model \mathcal{H} in (8) and the
296 patch-level training catalog described in the previous section. In the
297 analog forecasting setting, The search for analogs is restricted to patch
298 exemplars in the catalog within a local spatial neighborhood (typically
299 a patch-level 8-neighborhood), except for patches along the seashore
300 for which the search for analogs is restricted to patch exemplars at the
301 same location.
- 302 • the reconstruction of fields dX from the set of assimilated patches
303 $\{dX(\mathcal{P}_s, \cdot)\}_s$. This relies on a spatial averaging over overlapping patches
304 (here, a 5-pixel overlapping in both directions). In practice, we do not
305 apply the patch-level assimilation to all grid positions. Consequently,
306 the spatial averaging may result in blocky artifacts. We then apply

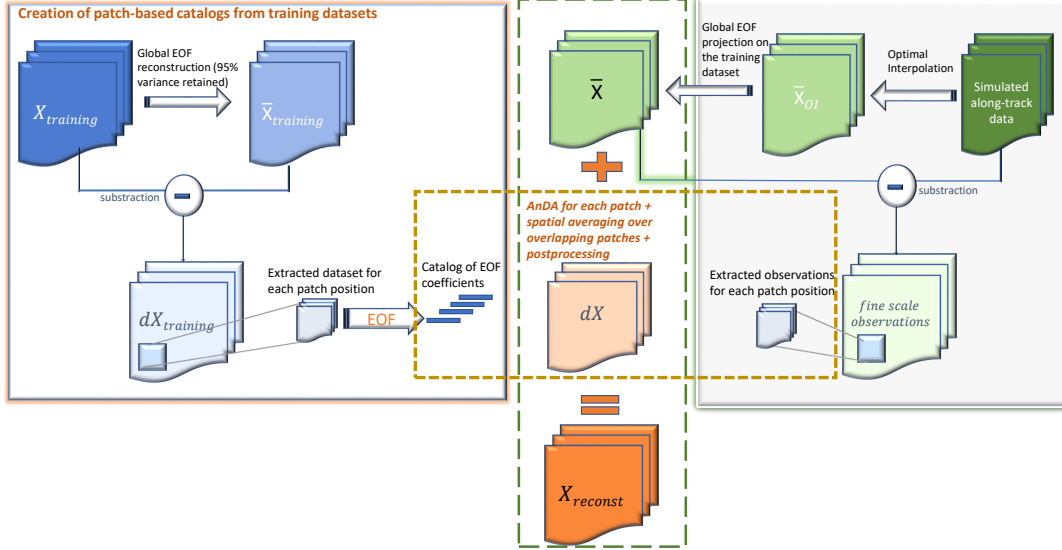


Figure 3: Sketch of the proposed patch-based Analog Data Assimilation (PB-AnDA). The left block details the construction of the patch-based catalogs from the training dataset. The right block illustrates the process of obtaining the large-scale component of the SLA reconstructed field. The orange dashed rectangle represents the application of the AnDA using the catalog and the fine-scale observations. Finally, the green dashed rectangle shows the final addition operation that yields the reconstructed SLA field.

307 a patchwise EOF-based decomposition-reconstruction with a smaller
 308 patch-size (here, 17×17 patches) to remove these blocky artifacts.

- 309 • the reconstruction of fields X as $\bar{X} + dX$.

310 5. Results

311 We evaluate the proposed PB-AnDA approach using the OSSE presented
 312 in Section 2. We perform a qualitative and quantitative comparison to state-
 313 of-the-art approaches. We first describe the parameter setting used for the

314 PB-AnDA as well as benchmarked models, namely OI, an EOF-based ap-
 315 proach (Ping et al., 2016) and a direct application of AnDA at the region
 316 level. We then report numerical experiments for noise-free and noisy ob-
 317 servation data as well the relevance of auxiliary variables in the proposed
 318 PB-AnDA scheme.

319 5.1. Experimental setting

320 We detail below the parameter setting of the models evaluated in the
 321 reported experiments, including the proposed PB-AnDA scheme:

- 322 • *PB-AnDA*: We consider 20×20 patches with 15-dimensional EOF de-
 323 compositions ($N_E = 15$), which typically accounts for 99% of the data
 324 variance for the considered dataset. The postprocessing step exploits
 325 17×17 patches and a 15-dimensional EOF decomposition. Regard-
 326 ing the parametrization of the AnEnKS procedure, we experimentally
 327 cross-validated the number of nearest neighbors K to 50, the number
 328 of ensemble members $n_{ensemble}$ to 100 and the observation covariance
 329 error in Equation 8 to $\mathbf{R} = 0.001$.
- 330 • *Optimal Interpolation*: We apply an Optimal Interpolation to the pro-
 331 cessed along-track data. It provides the low-resolution component for
 332 the proposed PB-AnDA model and a model-driven reference for eval-
 333 uation purposes. The background field is a null field. We use a Gaus-
 334 sian covariance model with a spatial correlation length of 100km and
 335 a temporal correlation length of 15 days (± 5 timesteps since our data
 336 is 3-daily). These choices result from a cross-validation experiment.

- 337 • *VE-DINEOF*: We apply a second state-of-the-art interpolation scheme
338 using a data-driven strategy solely based on EOF decompositions, namely
339 VE-DINEOF (Ping et al., 2016). We implement a patch-based version
340 of VE-DINEOF to make it comparable to the proposed PB-AnDA set-
341 ting. Given the same EOF decomposition as in PB-AnDA, the patch-
342 level VE-DINEOF iterates patchwise EOF projection-reconstruction
343 of the detail field dX . This scheme is initialized from the along-track
344 pseudo-observation field for along-track data positions and \bar{X} for miss-
345 ing data positions. After each projection-reconstruction, we only up-
346 date missing data areas. We run this iterative process until conver-
347 gence.
- 348 • *G-AnDA*: With a view to evaluating the relevance of the patch-based
349 decomposition, we also apply AnDA at the region scale, referred to as
350 G-AnDA. It relies on an EOF-based decomposition of the detail field
351 dX . We use 150 EOF components, which accounts for more than 99%
352 of the total variance of the SSH dataset. From cross-validation ex-
353 periments, the associated AnEnKS procedure relies on a locally-linear
354 analog forecasting model with $K = 500$ analogs, $n_{ensemble} = 100$ en-
355 semble members and an observation covariance error in Equation 8 set
356 to $\mathbf{R} = 0.001$

357 The patch-based experiments were run on Teralab infrastructure using a
358 multi-core virtual machine (30 CPUs, 64G of RAM). We used the Python
359 toolbox for patch-based analog data assimilation (Fablet et al., 2017) (avail-
360 able at github.com/rfablet/PB-ANDA). Optimal Interpolation was imple-
361 mented on Matlab using Escudier et al. (2013b). Throughout the exper-

362 iments, two metrics are used to assess the performance of the considered
 363 interpolation methods: i) daily and mean Root Mean Square Error (RMSE)
 364 series between the reconstructed SLA fields X and the groundtruthed ones,
 365 ii) daily and mean correlation coefficient between the fine-scale component
 366 dX of the reconstructed SLA fields and of the groundtruthed ones.

367 5.2. *SLA reconstruction from noise-free along-track data*

368 We first perform an idealized noise-free experiment, where the along-track
 369 observations are noise-free. The observation covariance error in Equation 8
 370 takes the value $\mathbf{R} = 0.001$. The interpolation performances for this ex-
 371 periment are illustrated in Table 1. Our PB-AnDA algorithm significantly
 372 outperforms OI. More specifically, the locally-linear PB-AnDA results in the
 373 best reconstruction among the competing methods. We suggest that this im-
 374 provement comes from the reconstruction of fine-scale features learned from
 375 the archived model simulation data. Figure 4a reports interpolated SSH fields
 376 and their gradient fields which further confirm our intuition. PB-AnDA in-
 377 terpolation shows an enhancement of the gradients and comes out with some
 378 fine-scale eddies that were smoothed out in OI and VE-DINEOF. This is also
 379 confirmed by the Fourier power spectrum of the interpolated SLA fields in
 380 Figure 4b.

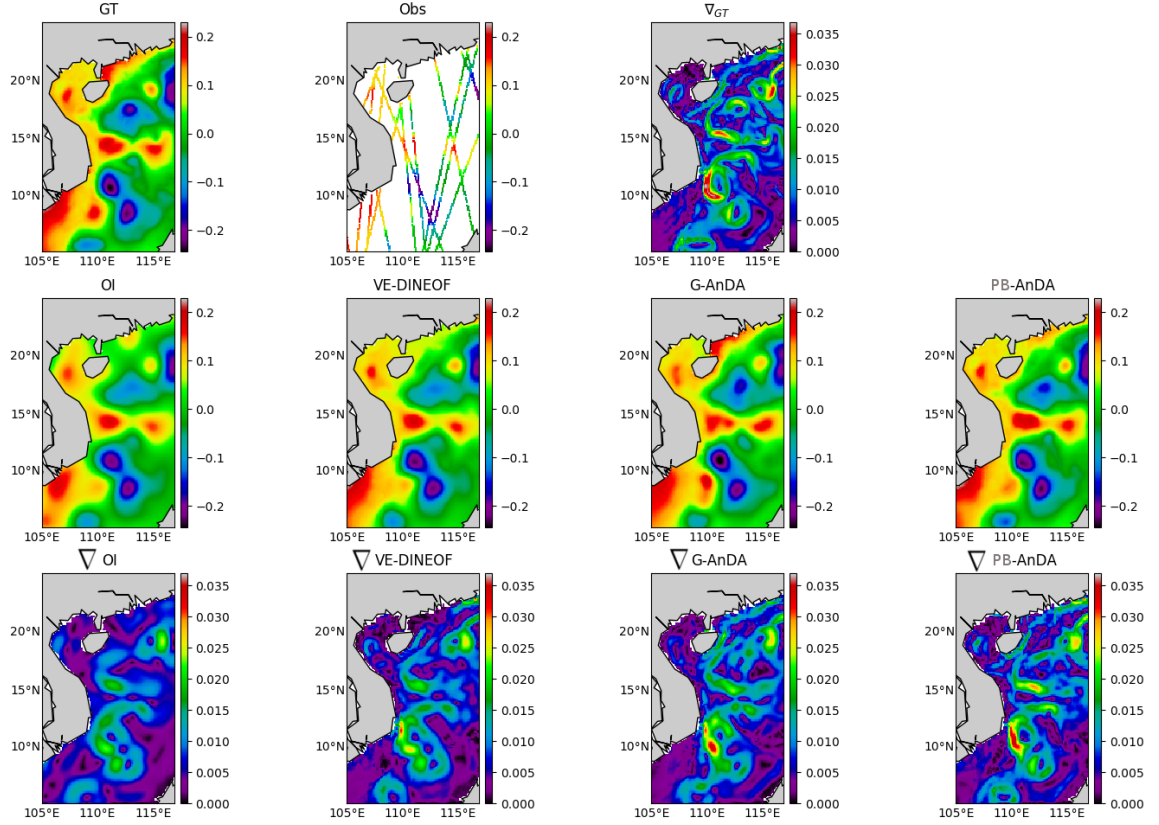
Table 1: SLA Interpolation performance for a noise-free experiment: Root Mean Square Error (RMSE) and correlation statistics for OI, VE-DINEOF, G-AnDA and PB-AnDA w.r.t. the groundtruthed SLA fields. See Section 5.1 for the corresponding parameter settings.

Criterion		RMSE	Correlation
OI		0.026 ± 0.007	0.81 ± 0.08
VE-DINEOF		0.023 ± 0.007	0.85 ± 0.07
G-AnDA		0.020 ± 0.006	0.89 ± 0.04
PB-AnDA	Locally-constant	0.014 ± 0.005	0.95 ± 0.03
	Locally-Increment	0.014 ± 0.005	0.95 ± 0.03
	Locally-Linear	0.013 ± 0.005	0.96 ± 0.02

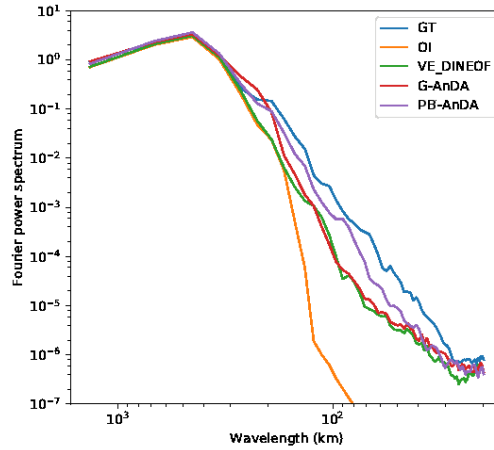
381 5.3. SLA reconstruction from noisy along-track data

382 We also evaluated the proposed approach for noisy along-track data.
383 Here, we run two experiments with an additive zero-mean Gaussian noise
384 applied to the simulated along-track data. We consider a noise covariance of
385 $\mathbf{R} = 0.01$ (Experiment A) and of $\mathbf{R} = 0.03$ (Experiment B) which is more
386 close to the instrumental error of conventional altimeters. Given the resulting
387 noisy along-track dataset, we apply the same methods as for the noise-free
388 case study.

389 We run PB-AnDA using different values for \mathbf{R} . For Experiment A, Table
390 2 shows that the minimum is reached using the true value of the error $\mathbf{R} =$
391 0.01. While for Experiment B, Table 3 shows that the minimum is counter-



(a)



(b)

Figure 4: Reconstructed SLA fields using noise-free along-track observation using OI, DINEOF, G-AnDA, PB-AnDA on the 54th day (February 24th 2012): from left to right, the first row shows the ground truth field, the simulated available along-tracks for that day, the ground truth gradient field. The second and third rows show each of the reconstruction and their corresponding gradient fields, from left to right, OI, VE-DINEOF, G-ANDA and PB-AnDA. The Fourier power spectrum of the competing methods is also included

intuitively reached again using value of the error $\mathbf{R} = 0.01$.

Our algorithm is then compared with the results of the application of the competing algorithms considered in this work. Results are shown in Table 4. PB-AnDA still outperforms OI in terms of RMSE and correlation statistics in both experiments. The locally-linear version of PB-AnDA depicts the best reconstruction performance. We report an example of the reconstruction in Figure 5. Similarly to the noise-free case study, PB-AnDA better recovers finer-scale structures in Fig.5.a compared with OI, VE-DINEOF and G-AnDA. In Fig.5.b, PB-AnDA also better reconstructs a larger-scale North-East structure, poorly sampled by along-track data and hence poorly interpolated by OI.

Table 2: Impact of variance of observation error R in AnDA interpolation performance using noisy along-track data ($R=0.01$): RMSE of AnDA interpolation for different values of parameter R . For the same dataset, OI RMSE is **0.039**.

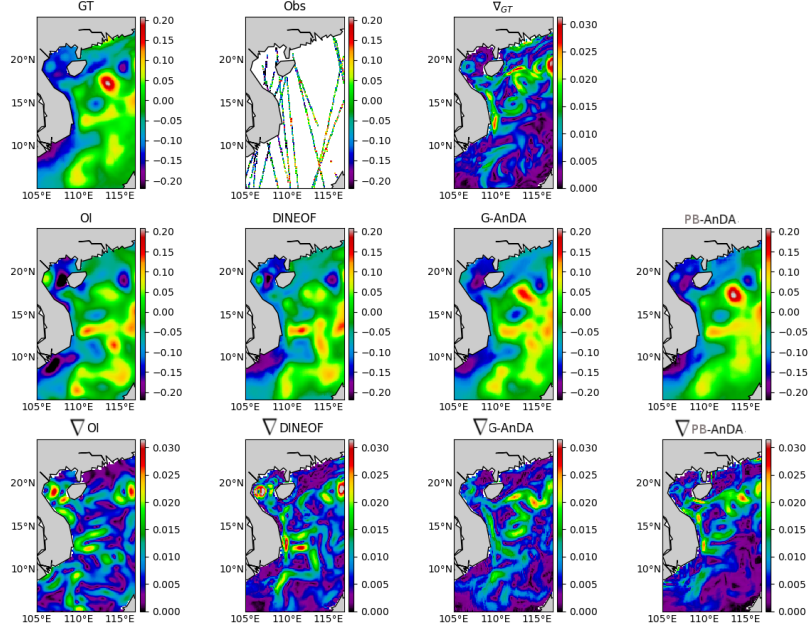
R	0.1	0.05	0.03	0.01	0.005	0.001	0.0001
$rmse_{PB-AnDA}$	0.035	0.030	0.028	0.025	0.025	0.029	0.044

Table 3: Impact of variance of observation error R in AnDA interpolation performance using noisy along-track data ($R=0.03$): RMSE of AnDA interpolation for different values of parameter R . For the same dataset, OI RMSE is **0.066**.

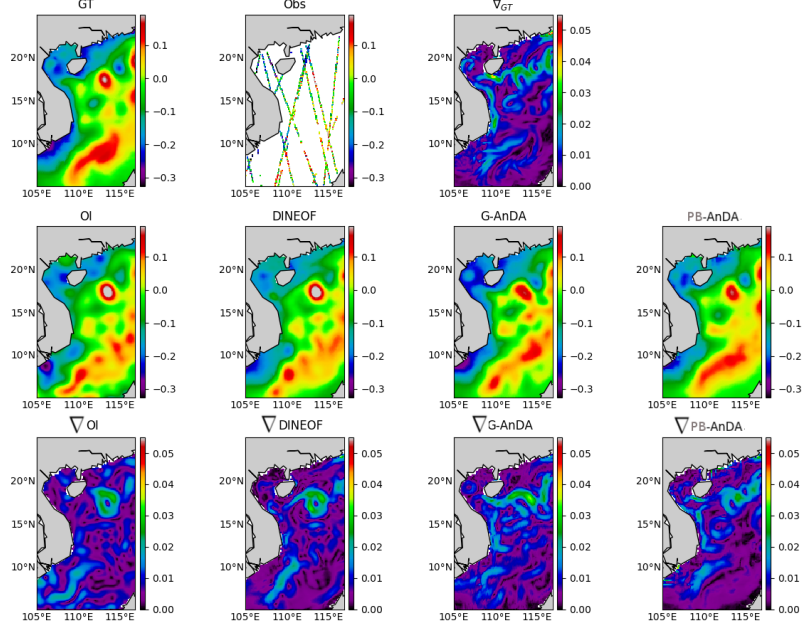
R	0.1	0.05	0.03	0.01	0.005	0.001	0.0001
$rmse_{PB-AnDA}$	0.038	0.036	0.035	0.0349	0.037	0.046	0.076

Table 4: SLA Interpolation performance for noisy along-track data: Root Mean Square Error (RMSE) and correlation statistics for OI, VE-DINEOF, G-AnDA and PB-AnDA w.r.t. the groundtruthed SLA fields. See Section 5.1 for the corresponding parameter settings.

	Criterion	RMSE	Correlation
$R=0.01$	OI	0.039 ± 0.005	0.64 ± 0.09
	VE-DINEOF	0.035 ± 0.005	0.68 ± 0.09
	G-AnDA	0.030 ± 0.005	0.78 ± 0.06
	PB-AnDA	Locally constant	0.026 ± 0.005
		Increment	0.028 ± 0.005
		Local Linear	0.0245 ± 0.005
$R=0.03$	OI	0.066 ± 0.006	0.41 ± 0.09
	VE-DINEOF	0.060 ± 0.006	0.45 ± 0.09
	G-AnDA	0.039 ± 0.006	0.67 ± 0.09
	PB-AnDA	Locally constant	0.035 ± 0.006
		Increment	0.036 ± 0.006
		Local Linear	0.032 ± 0.006



(a)



(b)

Figure 5: Reconstruction of SLA fields from noisy along-track data using OI, DINEOF, G-AnDA & PB-AnDA on day 225th (a) & 228th (b)

403 5.4. PB-AnDA models with auxiliary variables

404 We further explore the flexibility of the analog setting to the use of ad-
 405 ditional geophysical variable information as explained in Section 4.2. In-
 406 tuitively, we expect SLA fields to involve inter-scale dependencies as well
 407 as synergies with other tracers. The use of auxiliary variables provide the
 408 means for evaluating such dependencies and their potential impact on recon-
 409 struction performance. We consider two auxiliary variables that are used in
 410 the locally-linear analog forecasting model (7): i) to account for the rela-
 411 tionship between the large-scale and fine-scale component, we may consider
 412 variable \bar{X} , ii) considering potential SST-SSH synergies, we consider SST
 413 fields. Overall, we consider four parameterization of the regression variables
 414 used in PB-AnDA: the sole use of dX (PB-AnDA- dX); the joint use of dX
 415 and SST fields (PB-AnDA- dX +SST); the joint use of dX and \bar{X} (PB-AnDA-
 416 dX + \bar{X}), the joint use of dX and the groundtruthed version of \bar{X} denoted by
 417 \bar{X}^{GT} , (PB-AnDA- dX + \bar{X}^{GT}). The later provides a lower-bound for the recon-
 418 struction performance, assuming the low-resolution component is perfectly
 419 estimated.

420 We report mean RMSE and correlation statistics for these four PB-
 421 AnDA parameterizations in Table 5 for the noisy case-study. Considering
 422 PB-AnDA- dX as reference, these results show a very slight improvement
 423 when complementing dX with SST information. Though limited, we report
 424 a greater improvement when adding the low-resolution component \bar{X} . In-
 425 terestingly, a significantly greater improvement is obtained when adding the
 426 true low-resolution information. The mean results are in accordance with
 427 Fablet et al. (2016), which reported that large-scale SLA information was

428 more informative than SST to improve the reconstruction of the SLA at
 429 finer scales. Though mean statistics over one year leads to rather limited
 430 improvement, daily RMSE time series (Figure 6) reveal that for some peri-
 431 ods, for instance between day 130 and 150, relative improvements in terms
 432 of RMSE may reach 10% with the additional information brought by the
 433 large-scale component. In this respect, it may noted that PB-AnDA- $dX + \bar{X}$
 434 always perform better than PB-AnDA- dX .

Table 5: PB-AnDA reconstruction performance using noisy along-track data for different choices of the regression variables in the locally-linear analog forecasting model: PB-AnDA- dX using solely dX , PB-AnDA- $dX + \text{SST}$ using both dX and SST, PB-AnDA- $dX + \bar{X}$ using both dX and \bar{X} , and PB-AnDA- $dX + \bar{X}^{GT}$ using dX and the true large-scale component \bar{X}^{GT} .

	PB-AnDA model	RMSE	Correlation
$R=0.01$	PB-AnDA- dX	0.025 ± 0.005	0.83 ± 0.05
	PB-AnDA- $dX + \text{SST}$	0.024 ± 0.005	0.83 ± 0.05
	PB-AnDA- $dX + \bar{X}$	0.023 ± 0.005	0.84 ± 0.05
	PB-AnDA- $dX + \bar{X}^{GT}$	0.021 ± 0.004	0.87 ± 0.04
$R=0.03$	PB-AnDA- dX	0.032 ± 0.006	0.708 ± 0.06
	PB-AnDA- $dX + \text{SST}$	0.031 ± 0.006	0.710 ± 0.06
	PB-AnDA- $dX + \bar{X}$	0.029 ± 0.006	0.717 ± 0.06
	PB-AnDA- $dX + \bar{X}^{GT}$	0.026 ± 0.005	0.730 ± 0.05

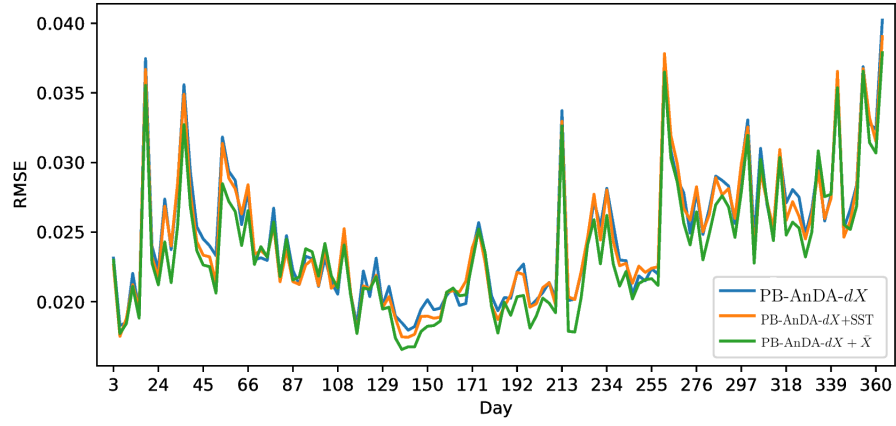
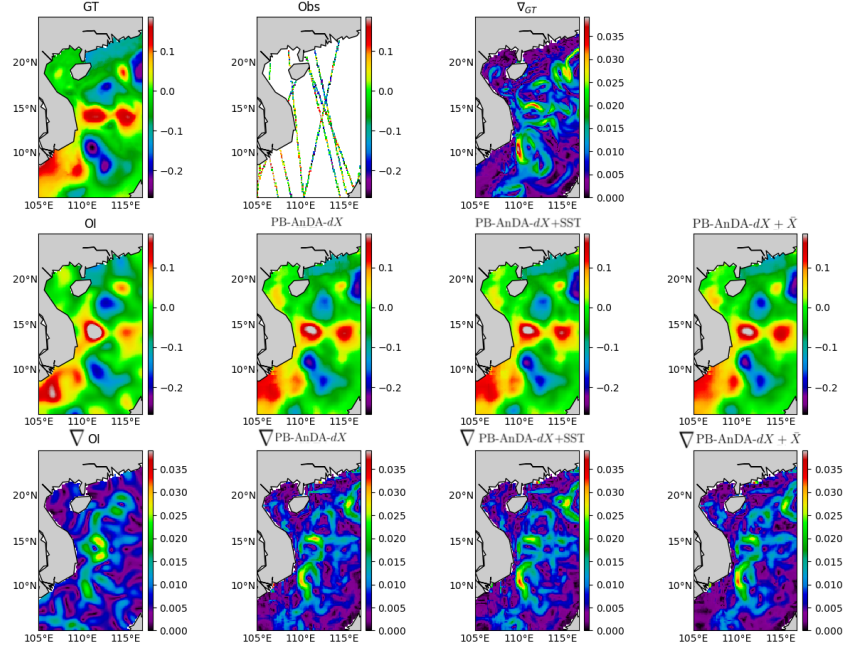
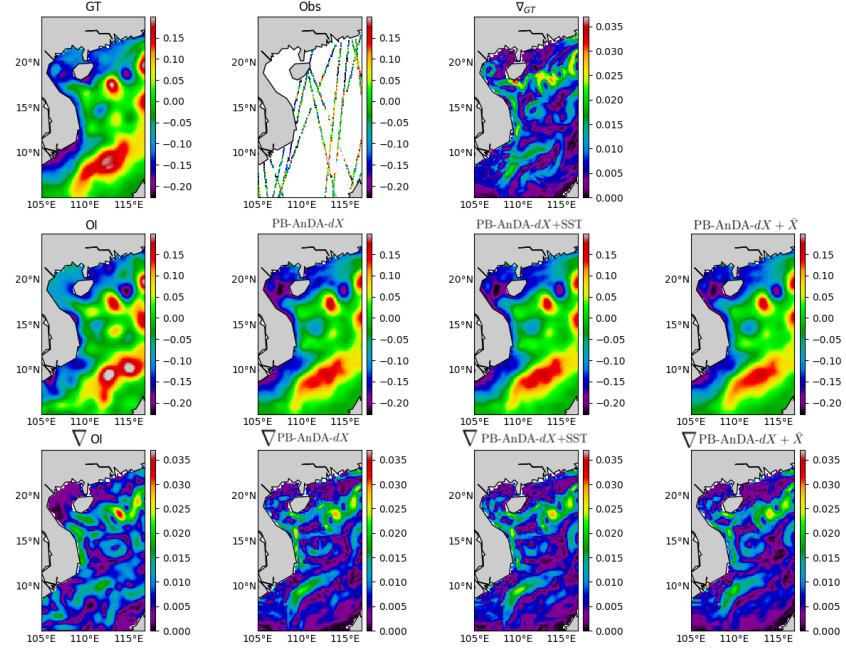


Figure 6: Daily RMSE time series of PB-AnDA SLA reconstructions using noisy along-track data for different choices of the regression variables in the locally-linear analog forecasting model: PB-AnDA- dX (light blue), PB-AnDA- dX +SST (orange) and PB-AnDA- $dX + \bar{X}$ (green).



(a)



(b)

Figure 7: (Noisy observation) Reconstruction of SLA fields using PB-AnDA with different multivariate regression models on day 57th & 237th (b)

435 6. Discussion and conclusion

436 This work sheds light on the opportunities that data science methods are
437 offering to improve altimetry in the era of "Big Data". Assuming the avail-
438 ability of high-resolution numerical simulations, we show that Analog Data
439 Assimilation (AnDA) can outperform the Optimal Interpolation method and
440 retrieve smoothed out structures resulting from the sole use of OI both with
441 idealized noise-free and more realistic noisy observations for the considered
442 case study. Importantly, the reported experiments point out the relevance for
443 combining OI for larger scales (above 100km) whereas the proposed patch-
444 based analog setting successfully applies to the finer-scale range below 100km.
445 This is in agreement with the recent application of the analog data assimila-
446 tion to the reconstruction of cloud-free SST fields (Fablet et al., 2017). We
447 also demonstrate that AnDA can embed complementary variables in a sim-
448 ple manner through the regression variables used in the locally-linear analog
449 forecasting operator. In agreement with our recent analysis (Fablet et al.,
450 2016), we demonstrate that the additional use of local SST and large-scale
451 SLA information may further improve the reconstruction performance for
452 fine-scale structures.

453 Analog data assimilation can be regarded as a means to fuse ocean models
454 and satellite-derived data. We regard this study as a proof-of-concept, which
455 opens research avenues as well as new directions for operational oceanogra-
456 phy. Our results advocate for complementary experiments at the global scale
457 or in different ocean regions for a variety of dynamical situations with a view
458 to further evaluating the relevance of the proposed analog assimilation frame-
459 work. Such experiments should evaluate the sensitivity of the assimilation

460 with respect to the size of the catalog. The scaling up to the global ocean also
461 suggests investigating computationally-efficient implementation of the ana-
462 log data assimilation. In this respect, the proposed patch-based framework
463 intrinsically ensures high parallelization performance. From a methodolog-
464 ical point of view, a relative weakness of the analog forecasting models (9)
465 may be their low physical interpretation compared with physically-derived
466 priors (Ubelmann et al., 2014). The combination of such physically-derived
467 parameterizations to data-driven strategies appear as a promising research
468 direction.

469 Beyond along-track altimeter data as considered in this study, future
470 missions such as SWOT (NASA/CNES) promise an unprecedented coverage
471 around the globe. More specifically, the large swath is expected to provide a
472 large number of data, urging for the inspection of the potential improvements
473 that this new mission will bring compared to classical along-track data. In
474 the context of analog data assimilation, the interest of SWOT data may be
475 two-fold. First, regarding observation model (8), SWOT mission will both
476 significantly increase the number of available observation data and enable
477 the definition of more complex observation models exploiting for instance
478 velocity-based or vorticity-based criterion. Second, SWOT data might also
479 be used to build representative patch-level catalogs of exemplars. Future
480 work should investigate these two directions using simulated SWOT test-
481 beds (Gaultier et al., 2015).

482 Acknowledgments

483 This work was supported by ANR (Agence Nationale de la Recherche,
484 grant ANR-13-MONU-0014), Labex Cominlabs (grant SEACS) and TeraLab
485 (grant TIAMSEA).

486 Asch, M., Bocquet, M., Nodet, M., 2016. Data assimilation: methods, algo-
487 rithms, and applications. Fundamentals of Algorithms. SIAM.

488 URL <https://hal.inria.fr/hal-01402885>

489 Atencia, A., Zawadzki, I., 2015. A comparison of two techniques for gener-
490 ating nowcasting ensembles. part ii: Analogs selection and comparison of
491 techniques. Monthly Weather Review 143 (7), 2890–2908.

492 Bocquet, M., Pires, C. A., Wu, L., 2010. Beyond gaussian statistical modeling
493 in geophysical data assimilation. Monthly Weather Review 138 (8), 2997–
494 3023.

495 Bretherton, F. P., Davis, R. E., Fandry, C., 1976. A technique for objective
496 analysis and design of oceanographic experiments applied to mode-73. In:
497 Deep Sea Research and Oceanographic Abstracts. Vol. 23. Elsevier, pp.
498 559–582.

499 Comeau, D., Giannakis, D., Zhao, Z., Majda, A. J., 2017. Predicting re-
500 gional and pan-arctic sea ice anomalies with kernel analog forecasting.
501 arXiv preprint arXiv:1705.05228.

502 De Mey, P., Robinson, A. R., 1987. Assimilation of altimeter eddy fields in
503 a limited-area quasi-geostrophic model. Journal of physical oceanography
504 17 (12), 2280–2293.

505 Escudier, R., Bouffard, J., Pascual, A., Poulain, P.-M., Pujol, M.-I., 2013a.
 506 Improvement of coastal and mesoscale observation from space: Applica-
 507 tion to the northwestern mediterranean sea. *Geophysical Research Letters*
 508 40 (10), 2148–2153.

509 Escudier, R., Bouffard, J., Pascual, A., Poulain, P.-M., Pujol, M.-I.,
 510 2013b. Improvement of coastal and mesoscale observation from space:
 511 Application to the northwestern Mediterranean Sea. *Geophysical Research*
 512 *Letters* 40 (10), 2148–2153.
 513 URL <http://onlinelibrary.wiley.com/doi/10.1002/grl.50324/full>

514 Fablet, R., Verron, J., Mourre, B., Chapron, B., Pascual, A., Oct. 2016. Im-
 515 proving mesoscale altimetric data from a multi-tracer convolutional pro-
 516 cessing of standard satellite-derived products, working paper or preprint.
 517 URL <https://hal.archives-ouvertes.fr/hal-01365761>

518 Fablet, R., Viet, P. H., Lguensat, R., 2017. Data-driven Models for the
 519 Spatio-Temporal Interpolation of satellite-derived SST Fields. *IEEE Trans-*
 520 *actions on Computational Imaging*.

521 Gandin, L., 1966. Objective analysis of meteorological fields. by L. S. Gandin.
 522 translated from the russian. jerusalem (israel program for scientific trans-
 523 lations), 1965. pp. vi, 242: 53 figures; 28 tables. 4 ls. 0d. *Quarterly Journal*
 524 *of the Royal Meteorological Society* 92 (393), 447–447.
 525 URL <http://dx.doi.org/10.1002/qj.49709239320>

526 Gaultier, L., Ubelmann, C., Fu, L.-L., Nov. 2015. The Challenge of Using

527 Future SWOT Data for Oceanic Field Reconstruction. *Journal of Atmo-*
528 *spheric and Oceanic Technology* 33 (1), 119–126.

529 Hardman-Mountford, N., Richardson, A., Boyer, D., Kreiner, A., Boyer,
530 H., 2003. Relating sardine recruitment in the northern benguela to
531 satellite-derived sea surface height using a neural network pattern recog-
532 nition approach. *Progress in Oceanography* 59 (2), 241 – 255, eNVIFISH:
533 Investigating environmental causes of pelagic fisheries variability in the
534 SE Atlantic.
535 URL <http://www.sciencedirect.com/science/article/pii/S0079661103001721>

536 Horton, P., Jaboyedoff, M., Obled, C., 2017. Global optimization of an analog
537 method by means of genetic algorithms. *Monthly Weather Review* 145 (4),
538 1275–1294.

539 Ide, K., Courtier, P., Ghil, M., Lorenc, A., 1997. Unified notation for data as-
540 simulation: operational, sequential and variational. *Practice* 75 (1B), 181–
541 189.

542 Isern-Fontanet, J., Chapron, B., Lapeyre, G., Klein, P., 2006. Potential use
543 of microwave sea surface temperatures for the estimation of ocean currents.
544 *GEOPHYSICAL RESEARCH LETTERS* 33, l24608.
545 URL [doi:10.1029/2006GL027801](https://doi.org/10.1029/2006GL027801)

546 Isern-Fontanet, J., Shinde, M., Andersson, C., Mar. 2014. On the Transfer
547 Function between Surface Fields and the Geostrophic Stream Function
548 in the Mediterranean Sea. *Journal of Physical Oceanography* 44 (5),

1406–1423.

URL <http://journals.ametsoc.org/doi/abs/10.1175/JP0-D-13-0186.1>

Klein, P., Isern-Fontanet, J., Lapeyre, G., Rouillet, G., Danioux, E.,
 Chapron, B., Le Gentil, S., Sasaki, H., Jun. 2009. Diagnosis of vertical
 velocities in the upper ocean from high resolution sea surface height.
 Geophysical Research Letters 36 (12), L12603.

URL <http://onlinelibrary.wiley.com/doi/10.1029/2009GL038359/abstract>

Lary, D. J., Alavi, A. H., Gandomi, A. H., Walker, A. L., 2016. Machine
 learning in geosciences and remote sensing. Geoscience Frontiers 7 (1),
 3–10.

Le Traon, P., Nadal, F., Ducet, N., 1998. An improved mapping method of
 multisatellite altimeter data. Journal of Atmospheric and Oceanic Tech-
 nology 15 (2), 522–534.

Lguensat, R., Tandeo, P., Ailliot, P., Pulido, M., Fablet, R., 2017. The analog
 data assimilation. Monthly Weather Review 0 (0), null.

URL <https://doi.org/10.1175/MWR-D-16-0441.1>

Lorenc, A. C., 1986. Analysis methods for numerical weather prediction.
 Quarterly Journal of the Royal Meteorological Society 112 (474), 1177–
 1194.

Lorenz, E. N., 1969. Atmospheric predictability as revealed by naturally oc-
 curring analogues. Journal of the Atmospheric sciences 26 (4), 636–646.

Masumoto, Y., Sasaki, H., Kagimoto, T., Komori, N., Ishida, A., Sasai, Y.,

- 571 Miyama, T., Motoi, T., Mitsudera, H., Takahashi, K., et al., 2004. A fifty-
572 year eddy-resolving simulation of the world ocean: Preliminary outcomes
573 of ofes (ogcm for the earth simulator). *J. Earth Simulator* 1, 35–56.
- 574 McDermott, P. L., Wikle, C. K., 2015. A model-based approach for analog
575 spatio-temporal dynamic forecasting. *Environmetrics*, n/a–n/a.
576 URL <http://dx.doi.org/10.1002/env.2374>
- 577 Ping, B., Su, F., Meng, Y., May 2016. An Improved DINEOF Algorithm
578 for Filling Missing Values in Spatio-Temporal Sea Surface Temperature
579 Data. *PLOS ONE* 11 (5), e0155928.
580 URL <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155928>
- 581 Sasaki, H., Nonaka, M., Masumoto, Y., Sasai, Y., Uehara, H., Sakuma, H.,
582 2008. An eddy-resolving hindcast simulation of the quasi-global ocean from
583 1950 to 2003 on the earth simulator.
- 584 Turiel, A., Nieves, V., Garcia-Ladona, E., Font, J., Rio, M.-H., Larnicol,
585 G., 2009a. The multifractal structure of satellite sea surface temperature
586 maps can be used to obtain global maps of streamlines. *Ocean Science*
587 5 (4), 447–460.
588 URL <https://www.ocean-sci.net/5/447/2009/>
- 589 Turiel, A., Sole, J., Nieves, V., Ballabrera-Poy, J., Garcia-Ladona, E.,
590 2009b. Tracking oceanic currents by singularity analysis of Microwave
591 Sea Surface Temperature images. *Remote Sensing of Environment* In Press.
592 URL <http://www.sciencedirect.com/science/article/B6V6V-4RFD07Y-1/2/9a77ceb1b616>

593 Ubelmann, C., Klein, P., Fu, L.-L., Oct. 2014. Dynamic Interpolation of Sea
 594 Surface Height and Potential Applications for Future High-Resolution
 595 Altimetry Mapping. *Journal of Atmospheric and Oceanic Technology*
 596 32 (1), 177–184.
 597 URL <http://journals.ametsoc.org/doi/abs/10.1175/JTECH-D-14-00152.1>

598 Yiou, P., 2014. Anawege: a weather generator based on analogues of atmo-
 599 spheric circulation. *Geoscientific Model Development* 7 (2), 531–543.

600 Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensing data:
 601 A technical tutorial on the state of the art. *IEEE Geoscience and Remote*
 602 *Sensing Magazine* 4 (2), 22–40.

603 Zhao, Z., Giannakis, D., Dec. 2014a. Analog Forecasting with Dynamics-
 604 Adapted Kernels. arXiv:1412.3831 [physics]ArXiv: 1412.3831.
 605 URL <http://arxiv.org/abs/1412.3831>

606 Zhao, Z., Giannakis, D., 2014b. Analog forecasting with dynamics-adapted
 607 kernels. arXiv preprint arXiv:1412.3831.