Can SWOT Improve the Reconstruction of Sea Level Anomaly Fields? Insights for Data-driven Approaches in the Western Mediterranean Sea

Manuel Lopez Radcenco, Ananda Pascual, Laura Gomez-Navarro, Abdeldjalil Aissa El Bey, Bertrand Chapron, Ronan Fablet

Can SWOT Improve the Reconstruction of Sea Level Anomaly Fields? Insights for Data-driven Approaches in the Western Mediterranean Sea

M. Lopez-Radcenco, *Student Member, IEEE,* A. Pascual, L. Gomez-Navarro, A. Aissa-El-Bey, *Senior Member, IEEE,* B. Chapron, and R. Fablet, *Senior Member, IEEE*

***Abstract*—Current generation satellite altimetry missions have played a fundamental role in improving our understanding of sea surface dynamics, despite only being able to provide measurements along the satellite track. In this respect, the future SWOT altimetry mission will be the first mission to produce complete two-dimensional wide-swath satellite observations. With a view towards the upcoming SWOT mission launch, we explore the potential of SWOT observations to improve the reconstruction of high-resolution sea level anomaly (SLA) fields from satellite-derived data. Given the ever-increasing availability of multi-source datasets that supports the exploration of data-driven alternatives to classical model-driven formulations, we focus here on recently introduced data-driven models for the interpolation of geophysical fields. Using an Observing System Simulation Experiment (OSSE), we demonstrate the relevance of SWOT observations to better constraint data-driven interpolation models in order to improve the reconstruction of mesoscale features. Reported results suggest that SWOT observations can provide more information than currently available nadir along-track altimetry observations and show an additional SLA reconstruction performance improvement when the the joint assimilation of SWOT and nadir along-track observations is considered.**

***Index Terms*—Sea surface, Remote sensing, Interpolation, Signal sampling, Radar altimetry**

I. Introduction

Current generation satellite altimetry missions have played a substantial role in improving our understanding of sea surface dynamics, despite only being able to provide measurements along the nadir satellite track, with a radar altimeter footprint width of around 2-10 km, as observed in Fig.1. Conventional satellite altimetry observations may then involve a very scarce sampling of the ocean surface with high rates of missing data, big gaps and increased sensibility to orbit characteristics such as track separation and revisit time. Indeed, current generation altimeters have spatial gaps of the order of 100 km between tracks. Moreover, a compromise between track separation and revisit time is to be made, with smaller track separation involving longer revisit times. In this respect, no satellite mission is currently capable of providing high-resolution observations in both space and time. Preliminary studies on the exploitation of nadir along-track satellite-derived data have proven that at least two altimeters are needed to accurately resolve the main spatio-temporal scales of global ocean processes [7], [24]. In [37], Morrow and Le Traon show that at least three altimeters are needed to capture mesoscale signals and processes. Subsequent studies exploiting both numerical models and real satellite data concluded that using up to four altimeters at the same time helps improve the accuracy and reconstruction power of state-of-the-art interpolation techniques [17], [25]–[28], [39]. However, even though the fusion of multiple altimeters allows for the reconstruction of scales that would be unattainable using a single altimeter, no combination of currently airborne altimeters is capable of completely resolving the smaller mesoscale (below 50-70 km) [14], [15] or the sub-mesoscale processes (below 10 km) [6], which are nonetheless crucial to increasing our understanding of a great number of oceanic processes [16].

In this context, it appears relevant to develop novel altimetry techniques that improve on current altimeters and enhance their capabilities to better observe and resolve upper ocean dynamics at smaller scales. In this respect, among recent advancements in remote sensing, the Surface Water and Ocean Topography (SWOT) satellite mission, a joint effort between the US National Aeronautics and Space Administration (NASA), the French Centre Nationale d'Etudes Spatiales (CNES) and the UK and Canada Space Agencies, aims at providing high resolution altimetry maps for both hydrology and oceanography. Specifically, the SWOT mission will be the first mission to exploit wide-swath Ka-band radar interferometry altimeters producing, for the first time, complete two-dimensional wide-swath satellite tracks [9], [16], [18], as seen in Fig. 1. Moreover, the mission's unparalleled high spatial resolution (2 km) should allow us to better capture mesoscale and sub-mesoscale processes [9], [16], which constitutes one of the great challenges within current oceanography research. In preparation for the mission launch in 2021, we use an observing system simulation experiment (OSSE) to explore the potential of exploiting SWOT observations in the context of data-driven methods for the reconstruction of high-resolution

M. Lopez-Radcenco, A. Aissa-El-Bey and R. Fablet are with MT Atlantique, Lab-STICC UMR6285, Brest, France. e-mail: manuel.lopezradcenco@imt-atlantique.fr

A. Pascual is with the Mediterranean Institute for Advanced Studies (IMEDEA) (UIB-CSIC), Esporles, Spain.

L. Gomez-Navarro is with Université Grenoble Alpes, CNRS, IRD, IGE, Grenoble, France and also with the Mediterranean Institute for Advanced Studies (IMEDEA) (UIB-CSIC), Esporles, Spain .

B. Chapron is with the French Research Institute for Exploitation of the Sea (IFREMER), 29280 Plouzané, France.

This work was supported by ANR (grant ANR-13-MONU-0014), Labex Cominlabs project SEACS and OSTST project MANATEE.

altimetry fields from satellite-derived data.

The paper is organized as follows. In section II we formally introduce the problem of geophysical field interpolation from irregularly-sampled data and discus related work and methods. The considered case-study, data and OSSE developed to evaluate the relevance of SWOT data to improve the reconstruction of SLA fields is presented in Section III. A brief overview of the considered data-driven models is presented in Section IV. In Section V we explore the effect of considering different types of observations involving various spatio-temporal sampling patterns, namely nadir along-track and SWOT observations, as well as a fusion of both data sources. Finally, we present our concluding remarks and future work perspectives in Section VI.

II. Problem statement and related work

Given the heterogeneity of currently available remote sensing altimetry observations, the processing and fusion of datasets involving different sampling strategies at multiple spatio-temporal resolutions, irregular sampling patterns and missing data to produce gridded gap-free products is of major interest for both research and industry. The problem of interpolating irregularly-sampled data onto a regular grid belongs to the family of inverse problems, which have been extensively studied [11], [36], [38], [41].

A. Model driven approaches

In ocean sciences, sate-of the-art interpolation methods rely on Optimal Interpolation (OI) [3], [8], [29], a model-driven approach that relies on linearity and Gaussianity hypotheses involving spatio-temporally invariant, mean covariance structures. Such hypotheses lead to a lack of accuracy in the representation of smaller scale structures, and it has been verified that it is not possible to recover fine scale structures (between 10-100 km) using OI [6], [14], [15]. This limitation is also associated with the inherent track separation in conventional altimeter tracks, with the length of the considered covariance structure representing a trade-off between model locality, relating to fine scale reconstruction, and observation availability, relating to the size of the gaps to be filled. In this respect, considerable efforts are being made to improve OI [8] or find alternative approaches [47]. Particularly, Escudier et al. [8] propose to improve OI by considering an additional batyhmetry constraint, while Ubelmann et al. [47] developed Dynamical Interpolation, an alternative approach that exploits the QG physical model by imposing a vorticity conservation constraint, and applied it to synthetic SWOT observations. Interestingly, from a physical point of view, Dynamic Interpolation comes to considering variable spatio-temporal covariance structures.

On the other hand, data assimilation techniques [10] have also been regarded as a particularly powerful model-driven tool for the interpolation of high-resolution geophysical fields. Data assimilation refers, in a general way, to methods aimed at combining the equations governing the behaviour of a dynamical system (which ultimately determine a numerical model) with observations conveying information related to such dynamical system. The objective is to improve the reconstruction/forecasting performance of the numerical model by exploiting the additional information introduced by observations. Formally, a state-space formulation is used [10]:

$$\left\{ \begin{array}{rcl} \mathbf{x}(t) & = & \mathcal{M}\left(\mathbf{x}(t-\delta t)\right) \\ \mathbf{y}(t) & = & \mathcal{H}\left(\mathbf{x}(t), \Omega(t)\right) + \eta(t) \end{array} \right. \tag{1}$$

where t is a discrete time index, $\mathbf{x}$ is the hidden state sequence to be reconstructed and $\mathbf{y}$ is the observed data sequence. $\mathcal{M}$ is a dynamical model relating the current state $\mathbf{x}(t)$ to the previous state $\mathbf{x}(t-\delta t)$. $\mathcal{H}$ is an observation operator, where $\Omega(t)$ is a mask accounting for missing data at time t and $\eta(t)$ is a random noise process accounting for observation uncertainties. In classical data assimilation, model $\mathcal{M}$ is applied at each time step to produce an initial forecast $\mathbf{x}^f$. This forecast is then corrected using observation $\mathbf{y}(t)$, by means of either stochastic approaches (such as Kalman filter [23]) aiming at producing the maximum a posteriori estimate (i.e. the most probable state) given the current state and observation sequence, or variational approaches (such as as 3DVAR [34]) aiming at optimizing a cost function that penalizes observation error by means of gradient descent approaches, to produce an improved prediction $\mathbf{x}^a$ (referred to as the analysis). In relation with SWOT, Carrier et al. [5] explore the assimilation of synthetic SWOT observations within a variational data assimilation framework, with promising results for the long-term assimilation of SWOT data.

B. Data-driven approaches

Nowadays, most alternative methods that do not rely on OI or data assimilation remain mostly model-driven, in spite of the growing availability of massive datasets, issued from both real remote sensing or in situ observations as well as numerical simulations or reanalysis issues, which truly supports the exploration of data-driven approaches as a powerful and efficient alternative [13], [41]. Broadly speaking, data-driven approaches aim at exploiting available observation/simulation datasets, which can be considered to accurately depict the spatio-temporal variability of the fields of interest, to emulate the physical model behind the process of interest and better capture the spatio-temporal variabilities that may not be accounted for in purely numerical models.

Initially developed for image processing issues, patch-based and exemplar-based models rely on the representation of images using a dictionary of representative exemplars extracted from the considered dataset. Among these approaches, non-local means and non-local priors [4], [41] have recently known some success in remote sensing applications [11], [36]. Integrating these approaches into data assimilation formulations led to the development of analog data assimilation [20], [30], [46], which relies on replacing the model-based forecasting step of data assimilation issues with a data-driven alternative. Alternatively, projection-based approaches have also recently been applied successfully to the interpolation of geophysicial fields. They involve the data-driven computation of a set of basis functions providing a low-dimensional representation of the considered dataset, and the interpolation of missing

values by exploiting a projection onto the estimated basis functions. Among these approaches, orthogonal projection approaches such as DINEOF [2] and VE-DINEOF [42] have been particularly succesful in remote sensing interpolation issues. However, given that orthogonal decompositions seem unable to recover smaller scale details, alternatively constrains for projections, such as non-negativity, have also been explored in the context of geophysical field interpolation [32].
Here, we follow recent developments [12], [13], [31] that explore data-driven approaches for the interpolation of high-resolution geophysical fields, with a particular focus on the contribution of SWOT observations to improve the reconstruction of SLA fields.

III. Data and case study region

To study the potential of SWOT observations to improve the reconstruction of high-resolution SLA fields, we implement an Observing System Simulation Experiment (OSSE) considering a case-study region in the Western Mediterranean Sea ($36.5°N$ to $40°N$, $1.5°E$ to $8.5°E$). The chosen region of interest is characterized by small Rossby radii, with smaller scales structures whose reconstruction from satellite data appears as considerably challenging. In this respect, the Mediterranean Sea can be regarded as a small scale ocean laboratory, as a wide variety of global mesoscale and sub-mesoscale ocean processes are present, at smaller scale, within both the basin and sub-basin scales [1], [45]. As such, developing efficient strategies to correctly capture and represent smaller scales within the Mediterranean Sea is of the utmost importance for the exploitation of the scientific potential of the region.

A. Data

A synthetic ground-truth dataset of daily high-resolution SLA fields from 2010 to 2013 is generated using the Western Mediterranean Operational Forecasting System (WMOP) numerical model from SOCIB [22]. WMOP is a regional reparametrization of the ROMS model nested in MFS, focused on producing high-resolution realistic ocean dynamics simulations in the Mediterranean Sea. The original grid resolution of the model varies between 1.8 km and 2.2 km ($\sim 1/50°$). We refer interested readers to [22] for an in-depth discussion of implementation details and validation of the WMOP model. For this experiment, however, high-resolution SLA fields produced by WMOP are down-sampled to a $1/20°$ resolution.

B. Observing System Simulation Experiment

The generated ground-truth SLA fields dataset is split into a training dataset comprising years 2010-2012, and a test dataset comprising year 2013, which serves as ground-truth for the generation of synthetic observations and the evaluation of reconstruction performance.
Synthetic observations emulating satellite altimetry are generated using the proposed OSSE, which exploits spatio-temporal locations derived from real satellite tracks from a four altimeter configuration (comprising Jason-2, Cryosat-2, Saral-Altika and Hy-2A satellite missions) in 2014. The year 2014 was chosen as reference because it provides the greatest number of spatio-temporal locations from real airborne altimetry missions. Following [39], acquisition noise is simulated by means of an additive Gaussian white noise of variance $\sigma_n^2 = 3\ cm^2$.
SWOT-like observations, on the other hand, are generated from the ground-truth high-resolution SLA fields using the SWOT simulator [18], provided by the SWOT science team. Synthetic observations obtained with the SWOT simulator are produced by linearly interpolating the ground-truth SLA fields onto a synthetically generated SWOT-like 2 km resolution grid computed using the expected orbit characteristics of the SWOT mission. Additionally, the SWOT simulator emulates most noise sources expected to influence the mission once launched. According to the SWOT mission error budget [9], this includes both correlated noise sources, related to the novel radar interferometry altimetry technique, its spatio-temporal sampling geometry and atmospheric perturbations (e.g. roll, phase, timing, baseline dilation and wet troposphere errors), and uncorrelated noise sources related to sensor acquisition error (e.g. KaRIn sensor noise). In [33], we demonstrated the high sensitivity of AnDA to correlated noise sources. Taking these results into account, we consider here synthetic SWOT observations involving uncorrelated noise sources only, under the assumption that a pre-processing scheme can be exploited to either filter correlated noise sources [19] or take them into account into the correlation structure of data assimilation schemes [44].
Finally, both types of synthetic observations are linearly interpolated onto a regular $1/20°$ grid to match the high-resolution ground-truth SLA fields.

C. Pseudo-observations

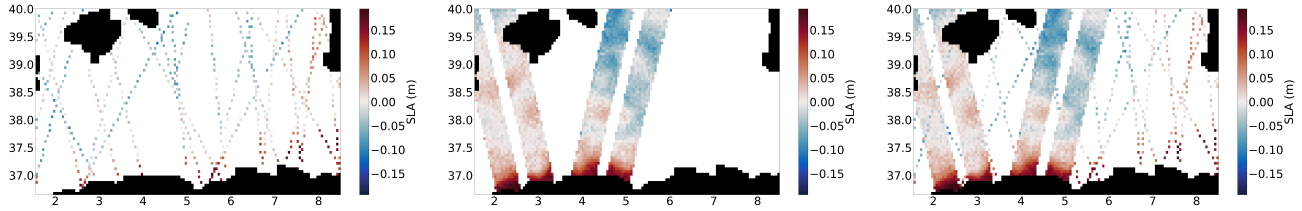
To further investigate alternative sampling strategies, two different synthetic observation generation methods are explored, as illustrated in Fig. 1. The first strategy comes to simply considering observations at times t_0, $t_0 + \delta t$,..., $t_0 + n\delta t$,... etc. directly. Alternatively, the second strategy relies on building pseudo-observations by accumulating multiple daily observations from a time window $t_0 \pm D$ centered at current day t_0. Additionally, nadir along-track and SWOT observations (or pseudo-observations) from the same day t_0 can also be combined, which allows for the seamless fusion of both data sources. An illustration of this principle is presented in Fig. 1.

IV. Data-driven spatio-temporal approaches

In this section, we present three data-driven approaches that we will exploit subsequently to evaluate the relevance of SWOT to improve the reconstruction of SLA fields.

A. Analog data assimilation framework

The analog data assimilation (AnDA) [30] was introduced as a data-driven alternative to classical data assimilation, and is based on the hypothesis that dynamical model $\mathcal{M}$ can be emulated by exploiting a catalog of realistic observations/simulations of the dynamical system. In this way,



(a) AT_5 (b) $SWOT_0$ (c) $AT_5 + SWOT_0$

Fig. 1. Examples of observations and pseudo-observations generated for the 9^{th} March, 2013 from the ground-truth SLA fields by considering either observations at current day t_0 or observations accumulated over a time window $t_0 \pm D$, with $D = 5$ days, for both nadir along-track altimetry observations (noted as AT_D) and pseudo-SWOT observations (noted as $SWOT_D$). *(1a) Nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as AT_5). (1b) Daily SWOT observations (noted as $SWOT_0$). (1c) Fusion of daily nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (noted as $AT_5 + SWOT_0$).*

at each time step, model $\mathcal{M}$ is replaced by a data-driven numerical approximation, the analog forecasting operator $\mathcal{F}$. Specifically, for current system state $\mathbf{x}$, and assuming a catalog of past state dynamics large enough to accurately depict the underlying processes is available, operator $\mathcal{F}$ is built from the K most similar states to $\mathbf{x}$ within the catalog, referred to as analogs. This idea, known as analog forecasting, was introduced by Lorenz in the context of weather forecasting [35], and relates to exemplar-based approaches such as non-local means [4] and non-local regularization [41]. Given that the future state of the analogs, known as successors, is known and exists within the catalog, this information can be used to adjust a linear regression that can be subsequently applied to the current state $\mathbf{x}$ to obtain forecast $\mathbf{x}^f$. In this respect, the procedure relies on the similarity between the current state $\mathbf{x}$ and simulated states stored in the catalog, and linear regression weights are introduced to account for the distance between the current state and its analog within the catalog, so that closer analogs will have a stronger contribution to the estimation of the forecasting linear regression. Once this step is completed, analog data assimilation proceeds similarly to classical data assimilation and exploits a stochastic scheme to combine forecast $\mathbf{x}^f$ with observation $\mathbf{y}$ and produce the analysis $\mathbf{x}^a$. In our current implementation, we consider an Ensemble Kalman filter/smoother (EnKFS) to assimilate partial observations sampled at a δt time step.

The application of the proposed framework to the reconstruction of high-resolution geophysical fields from irregularly-sampled data requires, however, the implementation of a number of strategies to deal with both the two-dimensional geometry as well as the high-dimensional nature of the problem. Such strategies include a multi-scale decomposition, so that larger scales are resolved by means of an OI scheme, while the reconstruction of smaller scales relies on the proposed formulation, and a patch-based EOF representation that allows the assimilation to be performed on a lower-dimensional EOF-based representation of the dynamical system. We refer the interested reader to [13], [31] for a detailed discussion of these aspects.

B. Multiscale VE-DINEOF

VE-DINEOF [42] is a state-of-the-art data-driven approach for the interpolation of irregularly-sampled geophysical fields that relies on an EOF [21], [40] decomposition. The principle behind this approach is simple, and involves an iterative procedure in which, at each iteration, missing data values are reconstructed by exploiting an EOF decomposition of the field of interest (with missing values being set to zero for the computation of EOFs). While the number of EOFs considered for the interpolation of missing values is increased sequentially at each iteration in classical DINEOF, VE-DINEOF relies on an improved formulation in which the number of EOFs used for missing data interpolation is optimized at each iteration using cross-validation. Furtheremore, in [31] we introduced MS-VE-DINEOF, a multi-scale, patch-based version of VE-DINEOF. This new formulation was developed to ensure a fair comparison between VE-DINEOF and AnDA in the context of the interpolation of geophysical fields. Following these developments, we exploit here MS-VE-DINEOF for the interpolation of SLA fields.

C. Locally-adapted dictionary-based non-negative decomposition of convolutional operators

The third data-driven approach considered, non-negative linear decomposition (NNLD), was introduced in [32] for the interpolation of irregularly-sampled geophysical fields. It relies on a globally fitted dictionary-based convolutional model that is then locally adjusted in overlapping spatial regions, and considers nadir along-track data accumulated on a multiple day window. More precisely, the proposed approach estimates a global convolutional model for each day t_0 by considering along-track observations on a time window centered around t_0. Daily global models are then projected onto a dictionary built using a non-negative data-driven dictionary learning scheme. The proposed non-negative decomposition then allows for the local readjustment of the global model on smaller, overlapping spatial regions, thus obtaining a more local representation.

V. Results

We first perform a preliminary comparative analysis of the considered methods by evaluating their performance when exploiting nadir along-track altimetry observations. For all the reported experiments, performance is measured by means of the root mean square error (in m) and correlation coefficient, for both the predicted SLA field and the gradient of the predicted SLA field (∇SLA). The default parameter settings

TABLE I
ROOT MEAN SQUARE ERROR (CORRELATION) FOR SLA RECONSTRUCTION FROM NADIR ALONG-TRACK OBSERVATIONS FOR THE DIFFERENT ALGORITHMS CONSIDERED, NAMELY OI [8], NNLD [32], MS-VE-DINEOF [30], [42] AND ANDA [30], [31]. BEST RESULT IN **BOLD**.

Setting	SLA	∇SLA
OI	0.02927 (0.8451)	0.006655 (0.6052)
MS-VE-DINEOF	0.02820 (0.9059)	0.006416 (0.6339)
NNLD	0.02115 (0.6965)	0.004506 (0.5893)
AnDA	**0.01978** (**0.9457**)	**0.004699** (**0.7660**)

TABLE II
ROOT MEAN SQUARE ERROR (CORRELATION) FOR ANDA SLA RECONSTRUCTION FROM NADIR ALONG-TRACK OBSERVATIONS (AT_D) AND WIDE-SWATH SWOT OBSERVATIONS ($SWOT_D$). FOR EACH TYPE OF OBSERVATIONS, BOTH DAILY OBSERVATIONS ($D = 0$) AND OBSERVATIONS ACCUMULATED ON A TIME WINDOW $t_0 \pm D$ WITH $D = 5$ DAYS ARE CONSIDERED. BEST RESULT IN **BOLD**.

Setting	SLA	∇SLA
AT_0	0.02395 (0.9186)	0.005507 (0.6989)
AT_5	0.01978 (0.9457)	0.004699 (0.7660)
$\mathbf{SWOT_0}$	**0.01810** (**0.9543**)	**0.004436** (**0.7857**)
$SWOT_5$	0.01920 (0.9502)	0.004345 (0.7913)

TABLE III
ROOT MEAN SQUARE ERROR (CORRELATION) FOR ANDA SLA RECONSTRUCTION FROM THE FUSION OF NADIR ALONG-TRACK OBSERVATIONS (AT_D) AND SWOT OBSERVATIONS ($SWOT_D$). FOR EACH TYPE OF OBSERVATIONS, BOTH DAILY OBSERVATIONS ($D = 0$) AND OBSERVATIONS ACCUMULATED ON A TIME WINDOW $t_0 \pm D$ WITH $D = 5$ DAYS ARE CONSIDERED. BEST RESULT IN **BOLD**.

Setting	SLA	∇SLA
$AT_0 + SWOT_0$	0.01742 (0.9576)	0.004375 (0.7934)
$AT_5 + SWOT_5$	0.01876 (0.9523)	0.004318 (0.7952)
$AT_5 + SWOT_0$	**0.01687** (**0.9607**)	**0.004286** (**0.8051**)

considered for AnDA and MS-VE-DINEOF are: patch size $W_p = 30$ pixels (~ 150 km), $K = 25$ neighbours, $\delta t = 1$ days, $N_{EOF} = 9$ EOF components, pseudo-observation half-window size $D = 5$. The OI SLA reconstruction reported was obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of 10 days and a spatial correlation scale of 100 km, while NNLD exploits a global model estimated from nadir along-track data accumulated on a 21-day window and then locally re-adjusted on $2^\circ \times 2^\circ$ overlapping regions. Table I presents SLA and ∇SLA RMSE and correlation results obtained when considering nadir along-track data for the different methods considered. From Table I, it appears that AnDA considerably outperforms the other proposed approaches. This is to be expected, given that AnDA is formulated in the context of a data assimilation framework, which allows it to explicitly model the temporal dependencies present within the exploited datasets. The other proposed approaches, on the other hand, only take into account this information implicitly within their formulations. Taking these results into account, we subsequently focus on exploring the use of SWOT observations within the AnDA framework.

To evaluate the effect of different observation sampling patterns, we now consider both nadir along-track observations and SWOT wide-swath observations, and compare the performance of AnDA when assimilating both types of observations separately. The default parameter settings considered are: patch size $W_p = 30$ pixels (~ 150 km), $K = 25$ neighbours, $\delta t = 1$ days, $N_{EOF} = 9$ EOF components. Table II presents SLA and ∇SLA RMSE and correlation results obtained when considering the AnDA assimilation of nadir along-track data and SWOT data for current observations only and for observations accumulated on a time window $t_0 \pm D$, with $D = 5$ days. From Table II, a first interesting result is that considering SWOT observations for the current day only (without accumulation) is already enough to outperform OI, MS-VE-DINEOF, NNLD and AnDA when considering only along-track data. Particularly, our results suggest a mean normalized RMSE (nRMSE, computed by normalizing the RMSE by the variance of the ground-truth high-resolution SLA fields) gain of around 2.8% for the whole year 2013 and 6.2% for days when at least one SWOT track is observed (with respect to the assimilation of 5-days accumulated along-track data only), which clearly implies that the performance gain relates to the availability of SWOT observations. This is in agreement with previous studies [5], [43] that suggest that SWOT observations should be able to provide as much information as four conventional altimeters for the reconstruction of mesoscale features. Indeed, for days when a SWOT swath exists within the considered region, the surface coverage of such swath is considerably larger than the surface coverage of the 5-day accumulated nadir along-track observations, which partially explains the increased reconstruction performance. Moreover, our results suggest that days involving a loss in reconstruction performance are, as expected, days where no SWOT observation exist and which thus involve a loss in observation coverage. Another interesting, yet counterintuitive, result is that performance seems to decrease when accumulating SWOT observations over a time window $t_0 \pm D$. We hypothesize that wide-swath observations, given their two-dimensional nature and higher continuous spatial coverage, are prone to capturing moving/changing structures multiple times as observations are accumulated, thus creating inconsistent SLA observations leading to the reconstruction of fictitious structures. Accumulating SWOT observations will thus tend to increase the sensitivity of AnDA to changes in the SLA field during the days over which observations are accumulated. In this respect, accumulating observations over multiple days seems to be an appropriate strategy to improve AnDA SLA reconstruction performance for nadir along-track data, whereas for SWOT data AnDA seems to be prone to issues arising from inconsistencies between accumulated observations.

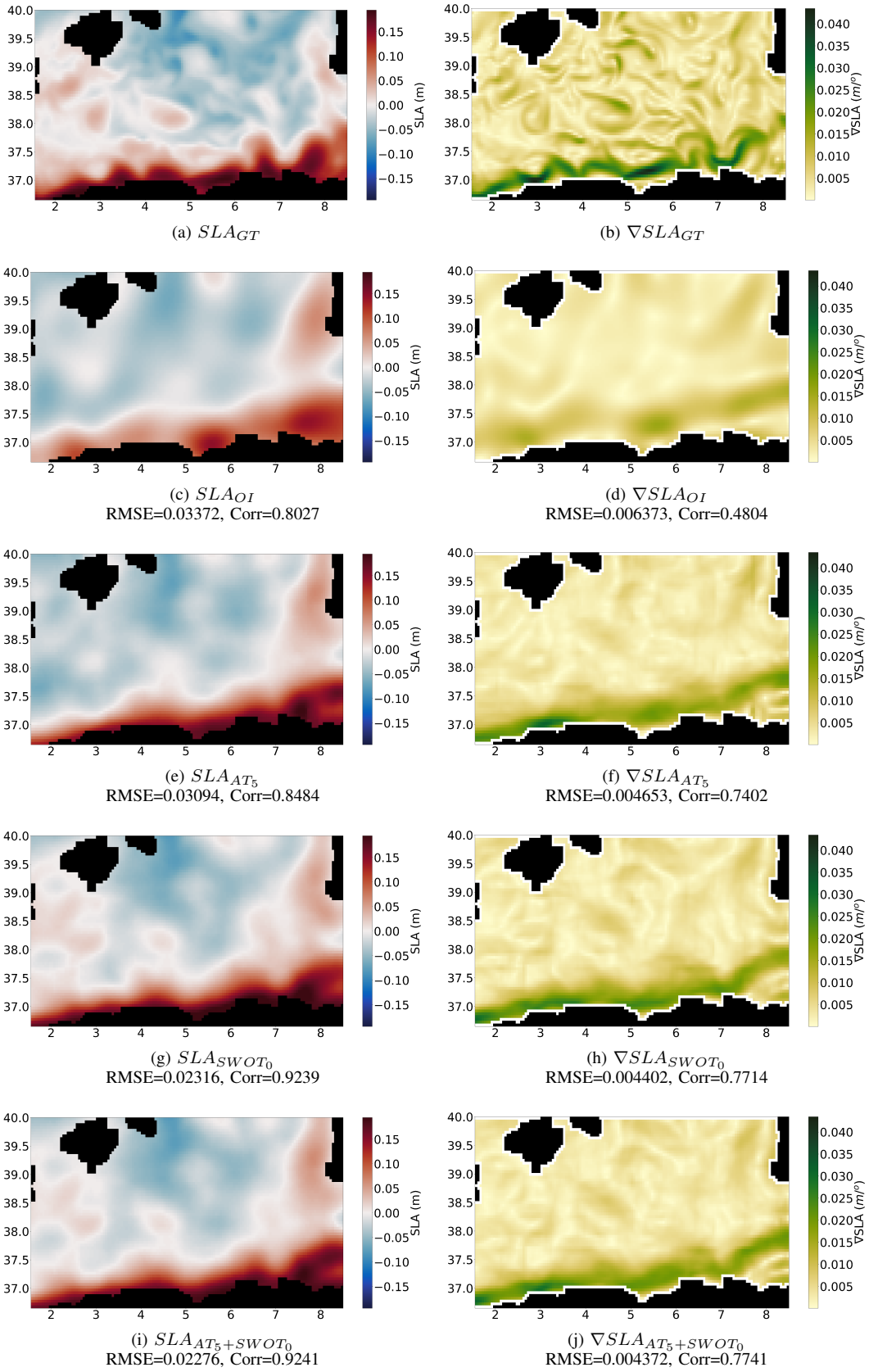
We further evaluate the potential of SWOT to improve the reconstruction of high-resolution SLA fields by considering the AnDA assimilation of a fusion of nadir along-track and SWOT observations. Table III presents SLA and ∇SLA RMSE and correlation results obtained when considering the joint AnDA assimilation of nadir along-track data and SWOT data

for current observations only and for observations accumulated on a time window $t_0 \pm D$, with $D = 5$ days. Not surprisingly, when considering the fusion of nadir along-track and SWOT data, the best strategy involves the fusion of accumulated nadir along-track observations over a time window of $D = 5$ days, thus increasing the spatial coverage of nadir along-track data, and SWOT observations captured on the current day only, thus fully exploiting the two-dimensional nature and added information provided by SWOT without being negatively affected by the increased sensitivity of AnDA to inconsistencies between accumulated SWOT observations.

To further illustrate the gain brought forward by considering additional SWOT data, Fig. 2 depicts the ground-truth SLA and ∇SLA fields (Fig. (2a) and Fig. (2b)), as well as the interpolated SLA and ∇SLA fields for the AnDA assimilation of classic nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (Fig. (2e) and Fig. (2f)), for the AnDA assimilation of daily SWOT observations (Fig. (2g) and Fig. (2h)) and for the AnDA assimilation of a combination of daily nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (Fig. (2i) and Fig. (2j)). We also include the SLA and ∇SLA fields interpolated by OI as reference (Fig. (2c) and Fig. (2d)). In agreement with previous results, the SLA field reconstructed from daily SWOT data is closer to the ground-truth SLA field than the reconstruction obtained when considering 5-days accumulated nadir along-track observations. Moreover, the reconstruction improvement associated with considering additional SWOT data appears clearly in the presented results, specially in the western region south off Mallorca, where the sampling provided by SWOT observations (Fig. 1b) helps better reconstruct small-scale structures that are not recovered by OI or AnDA using nadir along-track observations only. A completely opposite behaviour involving the smoothing of details can be observed for all considered cases in the eastern part of the considered region, where the sparse sampling provided by nadir along-track observations (Fig. 1a) only is not enough to accurately reconstruct smaller scale details. As far as gradient fields reconstruction is concerned, it seems clear that AnDA is able to better reconstruct small-scale features and details when the fusion of 5-days accumulated nadir along-track data and daily SWOT data is considered. Fig. (3) supports previous results by depicting the time series of the observation coverage (given as the ratio between the number of observed pixels and the total number of pixels in the considered region, excluding land pixels) for the fusion of daily nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (Fig. (3, top)), and the time series of the nRMSE gain obtained when adding SWOT observations to nadir along-track data accumulated on a time window of size $t_0 \pm D$ with $D = 5$ (given by $nRMSE_{AT_5} - nRMSE_{AT_5+SWOT_0}$) (Fig. (3, bottom)). The red filling indicates periods during which there is a positive nRMSE gain when considering additional SWOT data (i.e. $nRMSE_{AT_5+SWOT_0} \leq nRMSE_{AT_5}$). In Fig. (3, top), one can easily differentiate between the contribution of nadir along-track observations, producing a smaller, more stable low frequency periodic signal oscillating around 0.05, and that of SWOT observations, producing a train of spikes associated with the increased surface coverage of SWOT swaths. This spike pattern is highly correlated to the nRMSE gain time series depicted in Fig. (3, bottom) (R=0.60, p<0.001), where a similar peak pattern can be observed. This indicates that the observed nRMSE gain is mainly associated with the increased surface coverage brought forth by SWOT observations. Indeed, our results suggest a mean nRMSE gain of around 5.0% for the whole year 2013 and 7.0% for days when at least one SWOT track is observed (with respect to the assimilation of 5-days accumulated along-track data only), which supports the conclusion that the observed performance gain relates to the availability of SWOT observations.

VI. Conclusion

In this work, we explored the potential of SWOT observations to overcome the limitations of currently available altimetry products and improve the reconstruction of high-resolution sea level anomaly (SLA) fields from satellite observations. With a view towards improving on current state-of-the-art models, we focused on recently introduced data-driven formulations. Indeed, the ever-increasing availability of remote sensing, in situ and simulation datasets truly supports the exploration of data-driven approaches as a powerful alternative to classical model-driven schemes. We exploited an observing system symulation experiment (OSSE) to evaluate the performance of several data-driven approaches for the interpolation of SLA fields from irregularly-sampled altimetry data involving different sampling patterns, namely nadir along-track and wide-swath SWOT observations. In this respect, the Analog Data Assimilation (AnDA) framework, which may be regarded as a means to exploit high-resolution numerical simulation datasets for the reconstruction of SLA fields from partial satellite observations, appeared as particularly relevant to explicitly model upper ocean dynamics temporal dependencies. We demonstrated the relevance of SWOT observations to better constraint AnDA and improve the reconstruction of mesoscale features, with reported results suggesting that SWOT observations can provide more information than currently available nadir along-track altimetry observations. Moreover, an additional SLA reconstruction performance improvement is observed for the joint assimilation of SWOT and nadir along-track observations. Nevertheless, previous studies suggest that AnDA is considerably sensitive to the correlated noise sources associated with the SWOT mission [33]. In this respect, additional preprocessing should be carried out to filter correlated noises in SWOT data [19] or take it into account into the correlation structure of data assimilation schemes [44]. Future work will focus on combining these strategies with the AnDA framework in order to develop useful tools to process real observations from the future SWOT altimetry mission. In this respect, the joint assimilation of SWOT observation gradients and nadir along-track SLA data should be explored as a possible alternative to deal with the correlated noise sources present in SWOT data. Other interesting research avenues include the combination of different sources of altimetry data, as well as considering additional ocean dynamical



(a) SLA_{GT} (b) ∇SLA_{GT}

(c) SLA_{OI}
RMSE=0.03372, Corr=0.8027

(d) ∇SLA_{OI}
RMSE=0.006373, Corr=0.4804

(e) SLA_{AT_5}
RMSE=0.03094, Corr=0.8484

(f) ∇SLA_{AT_5}
RMSE=0.004653, Corr=0.7402

(g) SLA_{SWOT_0}
RMSE=0.02316, Corr=0.9239

(h) ∇SLA_{SWOT_0}
RMSE=0.004402, Corr=0.7714

(i) $SLA_{AT_5+SWOT_0}$
RMSE=0.02276, Corr=0.9241

(j) $\nabla SLA_{AT_5+SWOT_0}$
RMSE=0.004372, Corr=0.7741

Fig. 2. AnDA SLA reconstruction results for the 9^{th} March, 2013. *[(2a) and (2b)] Real high-resolution ground-truth SLA and SLA gradient fields. [(2c) and (2d)] Optimal Interpolation SLA and SLA gradient reconstruction obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of* 10 *days and a spatial correlation scale of* 100 *km. [(2e) and (2f)] AnDA SLA and SLA gradient reconstruction when considering nadir along-track observations accumulated on a time window* $t_0 \pm D$ *with* $D = 5$ *days (noted as* AT_5 *and depicted in (1a)). [(2g) and (2h)] AnDA SLA and SLA gradient reconstruction when considering daily SWOT observations (noted as* $SWOT_0$ *and depicted in (1b)). [(2i) and (2j)] AnDA SLA and SLA gradient reconstruction when considering the fusion of daily nadir along-track observations accumulated on a time window* $t_0 \pm D$ *with* $D = 5$ *days and daily SWOT observations (noted as* $AT_5 + SWOT_0$ *and depicted in (1c)).* RMSE and correlation values given for each considered case.

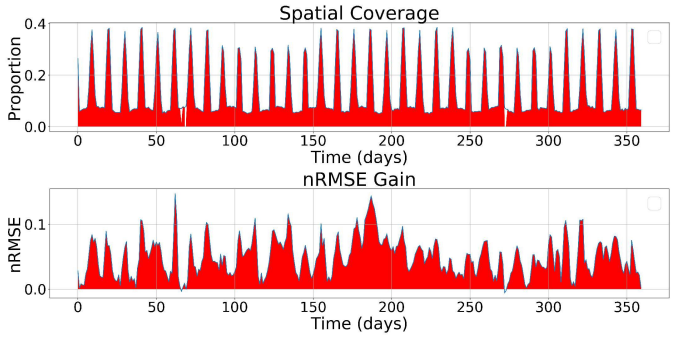


Fig. 3. (Top) Observation coverage (given as the ratio between the number of observed pixels and the total number of pixels in the considered region, excluding land pixels) time series for the fusion of daily nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations. (Bottom) Time series of the nRMSE gain obtained when adding daily SWOT observations to 5-days accumulated along-track data (given by $nRMSE_{AT_5} - nRMSE_{AT_5+SWOT_0}$). The red filling indicates periods during which there is a positive nRMSE gain when considering additional SWOT data (i.e. $nRMSE_{AT_5+SWOT_0} \leq nRMSE_{AT_5}$).

tracers (e.g. sea surface temperature, sea surface salinity, etc.). Further exploring the exploitation of structural information present in wide-swath observations, for example by means of numerically-resolved gradients or finite size Liapunov exponents (FSLE), is also an appealing research direction.

REFERENCES

[1] J. T. Allen, D. A. Smeed, J. Tintoré, and S. Ruiz. Mesoscale subduction at the almeria-oran front: Part 1: Ageostrophic flow. *Journal of Marine Systems*, pages 263–285, 2001.

[2] J. M. Beckers and M. Rixen. EOF Calculations and Data Filling from Incomplete Oceanographic Datasets. *Journal of Atmospheric and Oceanic Technology*, 20(12):1839–1856, Dec. 2003.

[3] F. P. Bretherton, R. E. Davis, and C. B. Fandry. A technique for objective analysis and design of oceanographic experiments applied to mode-73. In *Deep Sea Research and Oceanographic Abstracts*, volume 23, pages 559–582. Elsevier, 1976.

[4] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *IEEE Conf. on Computer Vision and Pattern Recognition, CVPR'05*, volume 2, pages 60–65 vol. 2, June 2005.

[5] M. J. Carrier, H. E. Ngodock, S. R. Smith, I. Souopgui, and B. Bartels. Examining the potential impact of swot observations in an ocean analysis–forecasting system. *Monthly Weather Review*, 144(10):3767–3782, 2016.

[6] D. B. Chelton and M. G. Schlax. The accuracies of smoothed sea surface height fields constructed from tandem satellite altimeter datasets. *Journal of Atmospheric and Oceanic Technology*, 20(9):1276–1302, 2003.

[7] N. Ducet, P.-Y. Le Traon, and G. Reverdin. Global high-resolution mapping of ocean circulation from topex/poseidon and ers-1 and -2. *Journal of Geophysical Research: Oceans*, 105(C8):19477–19498, 2000.

[8] R. Escudier, J. Bouffard, A. Pascual, P.-M. Poulain, and M.-I. Pujol. Improvement of coastal and mesoscale observation from space: Application to the northwestern Mediterranean Sea. *Geophys. Res. Lett.*, 40(10):2148–2153, 2013.

[9] D. Esteban-Fernandez. SWOT project mission performance and error budget document, JPL Doc. JPL D-79084. 2014.

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

[11] R. Fablet and F. Rousseau. Missing data super-resolution using non-local and statistical priors. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 676–680, Sept. 2015.

[12] R. Fablet, J. Verron, B. Mourre, B. Chapron, and A. Pascual. Improving mesoscale altimetric data from a multi-tracer convolutional processing of standard satellite-derived products. *IEEE Transactions on Geoscience and Remote Sensing*, 2017.

[13] R. Fablet, P. H. Viet, and R. Lguensat. Data-driven Models for the Spatio-Temporal Interpolation of satellite-derived SST Fields. *IEEE Trans. Comput. Imag.*, 2017.

[14] J. H. Faghmous, I. Frenger, Y. Yao, R. Warmka, A. Lindell, and V. Kumar. A daily global mesoscale ocean eddy dataset from satellite altimetry. *Scientific Data*, 2:150028 EP –, Jun 2015. Data Descriptor.

[15] L.-L. Fu, D. B. Chelton, P.-Y. Le Traon, and R. Morrow. Eddy dynamics from satellite altimetry. *Oceanography*, 23(4):14–25, 2010.

[16] L.-L. Fu and R. Ferrari. Observing oceanic submesoscale processes from space. *Eos, Trans. AGU*, 89(48):488–488, 2008.

[17] L.-L. Fu, D. Stammer, B. Leben, and D. Chelton. Improved spatial resolution of ocean surface topography from the topex/poseidon - jason-1 tandem altimeter mission. *EOS, Transaction, American Geophysical Union*, 84:247–248, 2003. cited By 2.

[18] L. Gaultier and C. Ubelmann. SWOT Simulator Documentation. Technical report, JPL, NASA, 2010.

[19] L. Gomez-Navarro, R. Fablet, E. Mason, A. Pascual, B. Mourre, E. Cosme, and J. Le Sommer. Swot spatial scales in the western mediterranean sea derived from pseudo-observations and an ad hoc filtering. *Remote Sensing*, 10(4), 2018.

[20] F. Hamilton, T. Berry, and T. Sauer. Ensemble Kalman Filtering without a Model. *Physical Review X*, 6(1):011021, Mar. 2016.

[21] A. Hannachi, I. T. Jolliffe, and D. B. Stephenson. Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology*, 27(9):1119–1152, 2007.

[22] M. Juza, B. Mourre, L. Renault, S. Gómara, K. Sebastián, S. Lora, J. P. Beltran, B. Frontera, B. Garau, C. Troupin, M. Torner, E. Heslop, B. Casas, R. Escudier, G. Vizoso, and J. Tintoré. SOCIB operational ocean forecasting system and multi-platform validation in the Western Mediterranean Sea. *J. Oper. Oceanogr.*, 9(sup1):s155–s166, Feb. 2016.

[23] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, Mar 1960.

[24] C. J. Koblinsky, P. Gaspar, and G. Lagerloef. The future of spaceborne altimetry oceans and climate change: A long-term strategy. Technical report, United States, 1992. N–92-26121.

[25] P.-Y. Le Traon and G. Dibarboure. Velocity mapping capabilities of present and future altimeter missions: The role of high-frequency signals. *Journal of Atmospheric and Oceanic Technology*, 19(12):2077–2088, 2002. cited By 39.

[26] P.-Y. Le Traon and G. Dibarboure. An illustration of the contribution of the topex/poseidon - jason-1 tandem mission to mesoscale variability studies. *Marine Geodesy*, 27(1-2):3–13, 2004. cited By 39.

[27] P.-Y. Le Traon, G. Dibarboure, and N. Ducet. Use of a high-resolution model to analyze the mapping capabilities of multiple-altimeter missions. *Journal of Atmospheric and Oceanic Technology*, 18(7):1277–1288, 2001.

[28] P.-Y. Le Traon, Y. Faugère, F. Hernandez, J. Dorandeu, F. Mertz, and M. Ablain. Can we merge geosat follow-on with topex/poseidon and ers-2 for an improved description of the ocean circulation? *Journal of Atmospheric and Oceanic Technology*, 20(6):889–895, 2003.

[29] P.-Y. Le Traon, F. Nadal, and N. Ducet. An improved mapping method of multisatellite altimeter data. *Journal of Atmospheric and Oceanic Technology*, 15(2):522–534, 1998.

[30] R. Lguensat, P. Tandeo, P. Aillot, and R. Fablet. The Analog Data Assimilation. *Mon. Weather Rev.*, 2017.

[31] R. Lguensat, P. Viet, M. Sun, G. Chen, T. Fenglin, B. Chapron, and R. Fablet. Data-driven Interpolation of Sea Level Anomalies using Analog Data Assimilation. Technical report, Oct. 2017.

[32] M. Lopez-Radcenco, R. Fablet, A. Aissa-El-Bey, and P. Ailliot. Locally-adapted convolution-based super-resolution of irregularly-sampled ocean remote sensing data. In *ICIP 2017 : IEEE International Conference on Image Processing*, Beijing, China, September 2017.

[33] M. Lopez-Radcenco, A. Pascual, L. Gomez-Navarro, A. Aissa-El-Bey, and R. Fablet. Analog data assimilation for along-track nadir and SWOT altimetry data in the Western Mediterranean Sea. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Valencia, Spain, July 2018.

[34] A. C. Lorenc, S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, and F. W. Saunders. The met. office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126(570):2991–3012, 2000.

[35] E. N. Lorenz. Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric sciences*, 26(4):636–646, 1969.

[36] L. Lorenzi, F. Melgani, and G. Mercier. Inpainting strategies for reconstruction of missing data in VHR images. *IEEE Geoscience and Remote Sensing Letters*, 8(5):914–918, 2011.

[37] R. Morrow and P.-Y. Le Traon. Recent advances in observing mesoscale ocean dynamics with satellite altimetry. *Advances In Space Research*, 50(8):1062–1076, Oct 2012.

[38] A. Newson, A. Almansa, M. Fradet, Y. Gousseau, and P. Pérez. Video Inpainting of Complex Scenes. *SIAM Journal on Imaging Sciences*, 7(4):1993–2019, Jan. 2014.

[39] A. Pascual, M.-I. Pujol, G. Larnicol, P.-Y. Le Traon, and M.-H. Rio. Mesoscale mapping capabilities of multisatellite altimeter missions: First results with real data in the Mediterranean Sea. *J. Mar. Syst.*, 65(1–4):190–211, Mar. 2007.

[40] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572, 1901.

[41] G. Peyré, S. Bougleux, and L. Cohen. Non-local Regularization of Inverse Problems. *Inverse Probl. and Imag.*, 5(2):511–530, 2011.

[42] B. Ping, F. Su, and Y. Meng. An Improved DINEOF Algorithm for Filling Missing Values in Spatio-Temporal Sea Surface Temperature Data. *PLOS ONE*, 11(5):e0155928, May 2016.

[43] M.-I. Pujol, G. Dibarboure, P.-Y. Le Traon, and P. Klein. Using high-resolution altimetry to observe mesoscale signals. *Journal Of Atmospheric And Oceanic Technology*, 29(9):1409–1416, 2012.

[44] G. A. Ruggiero, E. Cosme, J. Brankart, J. Le Sommer, and C. Ubelmann. An efficient way to account for observation error correlations in the assimilation of data from the future swot high-resolution altimeter mission. *J. Atmos. Oceanic Tech.*, 33(12):2755–2768, 2016.

[45] S. Ruiz, A. Pascual, B. Garau, I. Pujol, and J. Tintoré. Vertical motion in the upper ocean from glider and altimetry data. *Geophysical Research Letters*, 36(14), 2009.

[46] P. Tandeo, P. Ailliot, J. Ruiz, A. Hannart, B. Chapron, A. Cuzol, V. Monbet, R. Easton, and R. Fablet. Combining Analog Method and Ensemble Data Assimilation: Application to the Lorenz-63 Chaotic System. In V. Lakshmanan, E. Gilleland, A. McGovern, and M. Tingley, editors, *Machine Learning and Data Mining Approaches to Climate Science*, pages 3–12. Springer, 2015.

[47] C. Ubelmann, P. Klein, and L.-L. Fu. Dynamic Interpolation of Sea Surface Height and Potential Applications for Future High-Resolution Altimetry Mapping. *Journal of Atmospheric and Oceanic Technology*, 32(1):177–184, Oct. 2014.