ANALOG DATA ASSIMILATION FOR ALONG-TRACK NADIR AND SWOT ALTIMETRY DATA IN THE WESTERN MEDITERRANEAN SEA

M. Lopez-Radcenco[1], *A. Pascual*[2], *L. Gomez-Navarro*[2,3], *A. Aissa-El-Bey*[1], *R. Fablet*[1]

(1) IMT Atlantique, Lab-STICC UMR6285, Brest, France
(2) Mediterranean Institute for Advanced Studies (IMEDEA) (UIB-CSIC), Esporles, Spain
(3) Université Grenoble Alpes, CNRS, IRD, IGE, Grenoble, France

ABSTRACT

The ever increasing availability of in situ, remote sensing and simulation data supports the development of data-driven alternatives to classical model-driven methods for the interpolation of sea surface geophysical fields from partial satellite-derived observations. In this respect, we recently introduced the Analog Data Assimilation (AnDA), which exploits patch-based analog forecasting operators within a classic Kalman-based data assimilation framework. In this work, we consider the application of AnDA to the spatio-temporal interpolation of SLA (Sea Level Anomalies) from two types of satellite altimetry data, namely from along-track nadir data [1] and data from the upcoming wide-swath SWOT mission [2]. We report a sensitivity analysis w.r.t. the main parameters of the proposed AnDA scheme. Overall, the reported benchmarking analysis supports the relevance of the proposed AnDA scheme for an improved reconstruction of mescoscale structures for horizontal scales ranging from ~20km to ~100km, with an gain of 42% (12%) in terms of SLA RMSE (correlation) with respect to Optimal Interpolation (OI) [3]. Results suggest an additional potential improvement from the joint assimilation of SWOT and along-track nadir observations.

***Index Terms*—** Altimetry data, Sea Level Anomaly, Spatio-temporal interpolation, Data-driven model, Data assimilation, Western Mediterranean Sea, SWOT

1. INTRODUCTION

In recent years, developments in remote sensing, in situ measurements and numerical models have led to great advancements in our understanding of ocean dynamics and ocean-atmosphere interactions. In this context, great amounts of data issued from a wide variety of sources are gathered every day. The fusion and processing of such datasets, which usually involve irregular sampling patterns and missing data (due to cloud occlusion, satellite sampling patterns, etc), into high-resolution regularly-gridded gap-free L4 data products is of the utmost importance for operational oceanography and meteorology. The interpolation of gridded fields from irregularly-sampled data belongs to the family of inverse problems, for which multiple model-driven strategies have been proposed [3,4]. Optimal interpolation (OI) [3], a model-driven strategy based on modeling the spatio-temporal covariance of the field of interest, is one of the most commonly used interpolation methods. Given the variety of high-resolution simulation datasets, data-driven strategies have recently been explored as a powerful, computationally-efficient alternative to model-based approaches [5,6].

In [6], we demonstrated the relevance of the data-driven analog data assimilation (AnDA) framework, introduced in [7], for the interpolation of high-dimensional geophysical fields. Here, we develop an application of AnDA to the reconstruction of SLA fields in in the Western Mediterranean Sea from satellite altimetry data. The Western Mediterranean Sea is characterized by relatively small Rosby radii, which makes the reconstruction of mesoscale sea surface dynamics from satellite data particularly challenging. We report a sensitivity analysis of the AnDA framework with respect to its main parameters. Importantly, we explore two different spatial sampling patterns: the one associated with along-track nadir missions [1] and the one from the upcoming wide-swath SWOT mission [2]. Our experiments exploit an OSSE (Observing System Simulation Experiment) with real spatio-temporal sampling patterns. This OSSE results in ground-truthed datasets to benchmark the proposed AnDA scheme w.r.t. state-of-the-art interpolation approaches. Overall, the reported benchmarking analysis supports the relevance of the proposed AnDA scheme for an improved reconstruction of mesoscale structures for horizontal scales ranging from ~20km to ~100km, with considerable gains with respect to Optimal Interpolation (OI) [3].

This paper is organized as follows. In Section 2 we briefly introduce the AnDA framework, the concepts behind it and its implementation. Section 4 presents a sensitivity analysis w.r.t. key AnDA parameters. In Section 5, we report interpolation performance for the two types of satellite data and benchmark the proposed AnDA framework w.r.t. state-of-art

This work was supported by ANR (grant ANR-13-MONU-0014), Labex Cominlabs project SEACS and OSTST project MANATEE. The authors would also like to thank R. Lguensat and P. Viet for their valuable help.

schemes. We present our concluding remarks and future work perspectives in Section 6.

2. ANALOG DATA ASSIMILATION FRAMEWORK

The analog data assimilation [7] is a fully data-driven data assimilation framework for dynamical systems. It relies on the assumption that a dynamical model can be built directly from a catalog of realistic simulations of system state dynamics. The general idea is that a forecast can be made by exploiting the similarity between the current state and simulated states within the catalog.
Formally, we use a state-space formulation [4]:

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

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 η is a random noise process accounting for uncertainties.
In practice, at each time step, $\mathcal{M}$ is linearly approximated using the K nearest neighbours, referred as the analogs, of the current state $\mathbf{x}(t)$ within the catalog. The future states of these analogs are used to produce a forecast, thus replacing the classic model-based forecasting step of data assimilation with a data-driven approach. The resulting analog forecasting operator can be used as a plug-in on a classic stochastic assimilation models. Here, we consider an Ensemble Kalman filter/smoother (EnKFS) to assimilate partial observations sampled at a δt time steps. It is important to notice that the assimilated observations can be directly the observations at times t_0, $t_0+\delta t$,... or, alternatively, pseudo-observations built by accumulating observations on a time window $t_0 \pm D$.
It should also be noted that, given the high-dimensional nature of the data to be interpolated, the application of a dimensionality reduction technique is required prior to the nearest-neighbour search for similar states within the catalog. In this respect, following [6], the proposed AnDA scheme involves a patch-based representation [5]. The field to be interpolated is decomposed into overlapping $W_p \times W_p$ patches $\mathcal{P}(s,t)$, at location s and time t, and an EOF-based decomposition is applied for each patch: $\mathcal{P}(s,t) = \sum_{k=1}^{N_{EOF}} \alpha_k(s,t) B_k$, where B_k are the EOFs and $\alpha_k(s,t)$ are the coefficients of the decomposition of patch $\mathcal{P}(s,t)$ onto its EOFs. The interpolation of field $\mathbf{x}$ comes to the independent assimilation of each patch. Overall, the final reconstruction of field $\mathbf{x}$ comes to spatially average overlapping patches.

3. CASE-STUDY AND OSSE

As case-study, we consider a region in the Western Mediterranean Sea ($36.5°N$ to $40°N$, $1.5°E$ to $8.5°E$). We use

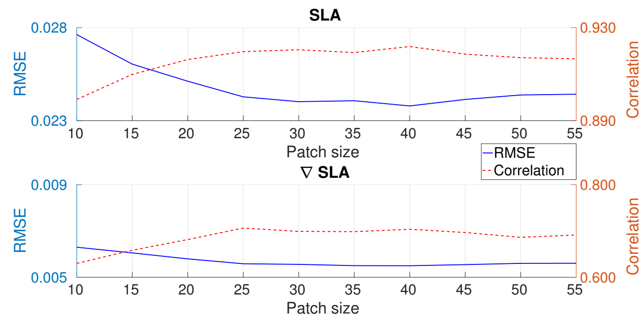

Fig. 1: AnDA sensitivity to patch size W_p

as realistic high-resolution reference daily ROMS numerical simulations [8] from 2010 to 2013 with a $1/20°$ resolution.
We implement Observing System Simulation Experiments (OSSE) to generate synthetic observations of satellite altimeter data. For along-track nadir data, we use real satellite track spatio-temporal locations from a four-altimeter sampling configuration in 2014. A Gaussian white noise of variance σ_n^2 is added to simulate an acquisition noise.
We proceed similarly for the upcoming SWOT mission, using SWOT simulator [9]. In this OSSE, the observation noise embeds different simulated noise processes [9, 10].

4. PARAMETER SENSITIVITY ANALYSIS

We apply the analog data assimilation to SLA anomaly fields w.r.t OI with a $\delta t = 1$ days assimilation lag, using data from 2010-2012 as our catalog, and we evaluate the performance of the assimilation framework on data from 2013. We study the interpolation performance in terms of root mean squared error (RMSE) and correlation coefficient for both the predicted SLA field and the gradient of the predicted field ∇SLA.
All the reported experiments use the following default parameters: $W_p = 35$, $K = 100$, $\sigma_n^2 = 0$, $\delta t = 1$, $D = 0$, $N_{EOF} = 9$. Our sensitivity analysis focuses on the impact of the patch size W_p, the number of neighbours K and the pseudo-observation half-window size D.

4.1. Patch size

Figure 1 presents SLA and ∇SLA RMSE and correlation coefficient as a function of patch size W_p. The optimal patch size appears to be between 125 and 200km (i.e., 25-40 pixels), which matches the smallest horizontal scales resolved by OI. This is in agreement with the expected gain of AnDA to improve the reconstruction of mesoscale structures for horizontal scales ranging from ~10-15 km to ~100 km.

4.2. Number of neighbours

Figure 2 presents SLA and ∇SLA RMSE and correlation coefficient as a function of the number of nearest neighbours

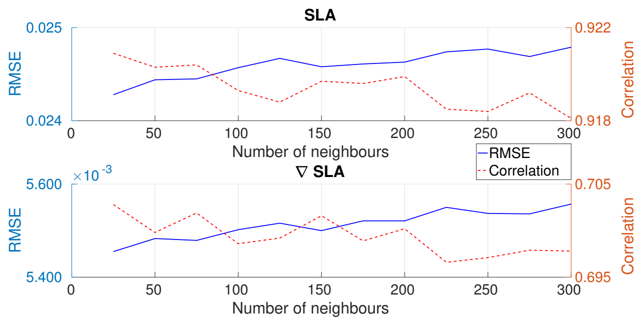

Fig. 2: AnDA sensitivity to number of neighbours K

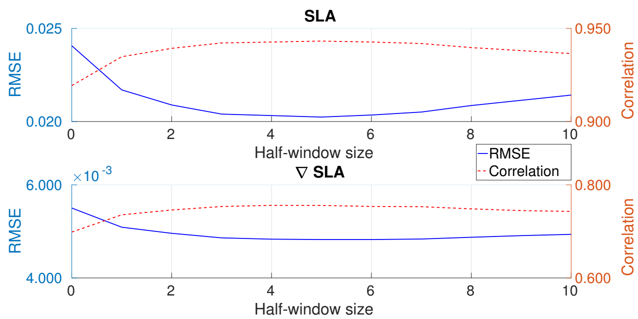

Fig. 3: AnDA sensitivity to observation half-window size D

K used in the analog forecasting step. Interestingly, $K = 25$ neighbours seems enough to obtain a good reconstruction, and using more neighbours seems to have a negative impact on interpolation performance. The lesser the neighbours, the more local the description of the dynamics in the state space. Conversely, considering additional neighbours results in a more global reconstruction, thus hindering performance.

4.3. Pseudo-observations half-window size

Figure 3 presents SLA and ∇SLA RMSE and correlation coefficient as a function of the pseudo-observation half-window size D. The optimal half-window size appears to be $D = 5$, which is in agreement with the temporal correlation scales of smaller mesoscale features.

5. ALONG-TRACK NADIR VS. SWOT DATA

To evaluate the effect of considering different observation sampling patterns, we generate additional pseudo-SWOT wide-swath observations from the simulated high-resolution SLA fields using the SWOT simulator [9]. The experimental procedure and performance metrics considered are identical to those in Section 4. The default parameters considered are: $W_p = 30$ (since our experiments demonstrate this produces the best results for SWOT data), $K = 25$, $\sigma_n^2 = 0$, $\delta t = 1$, $D = 0$, $N_{EOF} = 9$. We include the reconstruction performance of OI [3], and of a data-driven locally-adapted convolution-based non-negative decomposition approach (NNLD), introduced in [11] for the interpolation of irregularly-sampled geophysical fields, as reference.
According to the SWOT mission error budget, SWOT will present both correlated noise (eg. roll, phase and baseline dilation errors) and uncorrelated noise (e.g. KaRIN noise) [10]. In Table 1, we present SLA and ∇SLA RMSE and correlation when pseudo-SWOT observations are used without noise, with an uncorrelated (KaRIN) noise only and with both correlated and uncorrelated noise. Correlated noise has a strong impact on interpolation performance, whereas the AnDA scheme seems relatively robust to KaRIN noise. AnDA uses an EnFKS, which works best when noise is uncorrelated, and is unable to handle correlated noise properly without the use of additional techniques [12]. Our results also indicate that both effects seem to be enhanced when data is accumulated over several days (not shown here).
Table 2 presents SLA and ∇SLA RMSE and correlation results obtained when considering along-track nadir data, SWOT data and a combination of both along-track nadir and SWOT data without noise, for current observations only and for observations accumulated on a time window $t_0 \pm D$, with $D = 5$ days. We can conclude that, even though the best results in terms of SLA reconstruction performance are obtained when considering along-track nadir data accumulated over 5 days, adding SWOT data seems to improve results in terms of ∇SLA reconstruction. However, it should also be noted that accumulating SWOT observations seems to have a negative effect on SLA reconstruction, as wide-swath observations probably increase the sensitivity of AnDA to changes occurring in the SLA field during the days over which observations are accumulated. In this respect, there seems to be a compromise between accumulating data over multiple days, which improves SLA reconstruction performance, and considering SWOT data, which improves results in terms of ∇SLA reconstruction, but is prone to problems arising from inconsistencies between accumulated observations.
Finally, it is worth noting that both accumulating along-track nadir observations over several days and considering wide-swath SWOT observations (alone or combined with along-track nadir observations) seem to be effective strategies to outperform both OI and NNLD. Indeed, compared to OI (NNLD), AnDA presents better (similar) RMSE levels while having higher correlation coefficients, which can be interpreted as mesoscale structures being better recovered.

6. CONCLUSION

We developed an application of the Analog Data Assimilation (AnDA) to the reconstruction of SLA fields in the Western Mediterranean Sea from satellite altimetry data, namely satellite along-track nadir data and wide-swath SWOT data. The AnDA framework can be regarded as a means to exploit

Table 1: RMSE (Correlation) for AnDA using SWOT observations with different noise settings. Best result in **bold**. Result for OI [3] and NNLD [11] given as reference.

Setting	SLA	∇SLA
No noise	**0.02156** (**0.9388**)	**0.004515** (**0.7726**)
KaRIN noise	0.02166 (0.9383)	0.004530 (0.7724)
All noise sources	0.03926 (0.8042)	0.005999 (0.5736)
OI [3]	0.03389 (0.8447)	0.006661 (0.6050)
NNLD [11]	0.02127 (0.6957)	0.004513 (0.5892)

Table 2: RMSE (Correlation) for AnDA combining SWOT and AT observations. Best result in **bold**. Result for OI [3] and NNLD [11] given as reference.

Setting	SLA	∇SLA
AT ($D = 0$)	0.02397 (0.9197)	0.005524 (0.7023)
AT ($D = 5$)	**0.01966** (**0.9464**)	0.004687 (0.7679)
SWOT ($D = 0$)	0.02156 (0.9388)	0.004515 (0.7726)
SWOT ($D = 5$)	0.02523 (0.9200)	0.004474 (0.7678)
Both ($D = 0$)	0.02043 (0.9445)	0.004448 (**0.7805**)
Both ($D = 5$)	0.02428 (0.9254)	**0.004424** (0.7745)
OI [3]	0.03389 (0.8447)	0.006661 (0.6050)
NNLD [11]	0.02127 (0.6957)	0.004513 (0.5892)

high-resolution numerical simulation datasets for the reconstruction of SLA fields from partial satellite observations. The reported OSSE support the relevance of AnDA with respect to state-of-the-art aproches (OI [1, 3] and NNLD [11]). For instance, we report a clear improvement of 42% (12%) in terms of SLA RMSE (correlation) and 30% (27%) in terms of ∇SLA RMSE (correlation) with respect to OI when considering along-track nadir data. Our experiments also suggest an additional potential improvement from the joint assimilation of SWOT and along-track nadir observations. However, additional preprocessing should be carried out to filter out the correlated noise in SWOT data [12].

Future work will focus on combining such preprocessing strategies with the AnDA framework in order to develop useful tools to process real observations from the future SWOT altimetry mission. Other interesting research avenues include the combination of the different sources of altimetry data as well as considering additional ocean dynamical tracers (e.g. sea surface temperature, sea surface salinity, ocean color, etc.). The exploitation of structural information present in wide-swath observations, for example by means of numerically resolved SLA gradients and/or finite size Liapunov exponents (FSLE), is also an appealing research direction.

7. REFERENCES

[1] A. Pascual, M.-I. Pujol, G. Larnicol, P.-Y. Le Traon, and M.-H. Rio, "Mesoscale mapping capabilities of multi-satellite altimeter missions: First results with real data in the Mediterranean Sea," *J. Mar. Syst.*, vol. 65, no. 1–4, pp. 190–211, Mar. 2007.

[2] L.L. Fu and R. Ferrari, "Observing oceanic submesoscale processes from space," *Eos, Trans. AGU*, vol. 89, no. 48, pp. 488–488, 2008.

[3] 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.*, vol. 40, no. 10, pp. 2148–2153, 2013.

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

[5] G. Peyré, S. Bougleux, and L.D. Cohen, "Non-local Regularization of Inverse Problems," *Inverse Probl. and Imag.*, vol. 5, no. 2, pp. 511–530, 2011.

[6] 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.

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

[8] 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.*, vol. 9, no. sup1, pp. s155–s166, Feb. 2016.

[9] L. Gaultier and C. Ubelmann, "SWOT Simulator Documentation," Tech. Rep., JPL, NASA, 2010.

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

[11] M. López-Radcenco, R. Fablet, A. Aïssa-El-Bey, and P. Ailliot, "Locally-adapted convolution-based super-resolution of irregularly-sampled ocean remote sensing data," *ArXiv e-prints*, Apr. 2017.

[12] G. A. Ruggiero, E. Cosme, J.M. 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.*, vol. 33, no. 12, pp. 2755–2768, 2016.