DATA-DRIVEN ASSIMILATION OF IRREGULARLY-SAMPLED IMAGE TIME SERIES

R. Fablet[1] , P. Viet[1], R. Lguensat[1], B. Chapron[2]

(1) IMT Atlantique; Lab-STICC, Brest, France
(2) Ifremer; LOPS, Brest, France

ABSTRACT

We address in this paper the reconstruction of irregurlarly-sampled image time series with an emphasis on geophysical remote sensing data. We develop a data-driven approach, referred to as an analog assimilation and stated as an ensemble Kalman method. Contrary to model-driven assimilation models, we do not exploit a physically-derived dynamic prior but we build a data-driven dynamic prior from a representative dataset of the considered image dynamics. Our contribution is here to extend analog assimilation to images, which involve high-dimensional state space. We combine patch-based representations to a multiscale PCA-constrained decomposition. Numerical experiments for the interpolation of missing data in satellite-derived ocean remote sensing images demonstrate the relevance of the proposed scheme. It outperforms the classical optimal interpolation with a relative RMSE gain of about 50% for the considered case study.

Index Terms— Data assimilation, irregular sampling, image time series, data-driven methods, Kalman methods

1. PROBLEM STATEMENT AND RELATED WORK

The reconstruction of images and image time series from irregularly-sampled observations is a key issue in numerous domains. One may cite satellite-based remote sensing in earth science, where the atmospheric conditions (e.g., clouds, heavy rains) may result in missing data patterns [1, 2]. In-situ monitoring networks also lead to irregularly-sampled observation patterns. Such sampling patterns are for instance very common in geophysics and ecology for the mapping of spatial and spatiotemporal features of interest such as population or habitat dynamics and bio-geophysical properties [3]. Fig.1 illustrates the scarcity of the data for such applications, here a satellite-derived sea surface geophysical field with missing data rates greater than 90% for a particular day and area. This example stresses the non-uniform sampling pattern resulting from the presence of clouds, where large subareas involve almost no observation. This makes the reconstruction of such image time series particularly challenging.

Such interpolation issues are classically addressed using optimal interpolation [1]. Under the assumption that the field is Gaussian and its covariance structure is known, optimal interpolation resorts to the reconstruction with minimal mean-square error. The validity of the Gaussianity assumption and the knowledge or the estimation of the covariance structure clearly affect the quality of the reconstructed fields. A variety of parametric and non-parametric solutions have been investigated [3]. However, for numerous applications, non-Gaussian fine-scale patterns are expected such that optimal interpolation might only be relevant at intermediate or coarse scales. Variational or stochastic state-space models, referred to as data assimilation models in geophysics, address such issues [4, 5]. They rely on the definition of some dynamic model or prior in the considered image space. It may be noted that optimal interpolation may be regarded as a stochastic assimilation method using a direct numerical resolution rather than sequential methods such as Kalman or particle filters [4]. The choice of the dynamic model is again a critical step and generally involves some trade-off between its representativity and its computational complexity.

This makes particularly appealing data-driven schemes, referred to as analog assimilation [6, 7, 8]. The key idea is to replace the explicit definition of the dynamic model by some implicit representation through a dataset of exemplars of the considered process. Analog assimilation shares common features with non-local image and video models [9, 10, 11]. It is however embedded in classical stochastic filtering frameworks, such as particle and Kalman filters, to exploit computationally-efficient sequential methods. By contrast, non-local priors resort to iterative optimization technique, which makes them prohibitive for image time series with sampling patterns as illustrated in Fig.1. Other classical models for image and video denoising and inpainting, such as variational priors [12, 13], appear less relevant to address large missing data areas as in Fig.1.

In this paper, we develop analog assimilation models for irregularly-sampled image time series as in Fig.1. Previous works [6, 7] only apply to low-dimensional spaces and cannot scale up to high-dimensional spaces associated with images. Here, using patch-based image representations and multiscale PCA-constrained decompositions, we demonstrate that analog assimilation models apply to image time series and can

significantly outperform classical model-driven assimilation schemes. This paper is organized as follows. Section 2 introduces the proposed multiscale exemplar-based assimilation model. Section 3 presents experimental results for an application to missing data interpolation associated with satellite-derived sampling patterns. We further discuss our key contributions and future work in Section 4.

2. MULTISCALE ANALOG ASSIMILATION

2.1. Model-driven vs. analog assimilation

Following [4], data assimilation, which aims at reconstructing a state sequence from a given observation sequence, can be stated according to a discretized state-space setting :

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

where t is a discrete time index, $\mathbf{x}$ the hidden state sequence to be reconstructed and $\mathbf{y}$ the observed data sequence. $\mathcal{M}$ states the dynamic prior, that is to say the prior on state $\mathbf{x}(t)$ given previous state $\mathbf{x}(t-1)$. Θ refers to the parameterization of model $\mathcal{M}$. $\mathcal{H}$ is the observation operator associated with the spatiotemporal missing data pattern Ω. η is a random process accounting for noise and calibration uncertainties. In the following, we will consider a Gaussian white noise for the sake of simplicity. Extension to non-Gaussian and/or correlated noise can be considered similarly.

Stochastic model-driven assimilation models [4, 2] resorts to defining parametric priors for model $\mathcal{M}$, often Gaussian priors stated through a covariance structure Γ. By contrast, the analog assimilation builds a non-parametric data-driven prior from a representative dataset of the hidden state dynamics [6]. This dataset is referred to as a *catalog* denoted by $\mathcal{C}$ and formed by pairs $(\tilde{\mathbf{x}}(t_i), \tilde{\mathbf{x}}(t_i + \Delta t))_i$ of consecutive state vectors separated by the same time lag Δt. By convention, Δt is set to 1 without loss of generality. We consider a nearest-neighbor analog strategy. For a given state $\mathbf{x}(t)$ at time t, the data-driven dynamic prior $\mathcal{M}(\mathbf{x}(t), \mathcal{C})$ resorts to :

$$\mathbf{x}(t+1)|\mathbf{x}(t) \propto \sum_i \omega_i\left(\mathbf{x}(t)\right)\delta\left(\mathbf{x}(t+1) - \tilde{\mathbf{x}}(t_i+1)\right)$$

where $\delta(\cdot)$ is the Dirac function. $\omega_i(\mathbf{x}(t))$ is a normalized similarity weight between state $\tilde{\mathbf{x}}(t_i)$ in the catalog and current state $\mathbf{x}(t)$. We here consider a Gaussian similarity measure and a nearest-neighbor setting such that only the K states in the catalog the closest to $\mathbf{x}(t)$ involve a non-null weight.

The simple sampling of the above data-driven dynamic model makes it particularly relevant for a combination with classical stochastic filters, such as ensemble Kalman filters and smoothers, referred to as EnKF and EnKS, as well as particle filters [4, 6]. Ensemble Kalman filters show a greater numerical stability and are considered in this study. We let the reader refer to [6, 8] for further details on analog assimilation.

The dimension of state $\mathbf{x}$ is clearly critical for the relevance of this analog dynamic prior. High-dimensional state makes the search for consistent neighbors much more complex. In addition, the size of the catalog would become extremely large, if not unrealistic, to provide a representative dataset of the spatiotemporal variability of state $\mathbf{x}$. This makes questionable the direct application of the analog assimilation to image time series.

2.2. Data-driven assimilation for image time series

The key contribution of this paper is the development and evaluation of analog assimilation models for image time series. We first exploit a multi-scale decomposition of state $\mathbf{x}$. We assume that we can decompose $\mathbf{x}$ as $\mathbf{x} = \bar{\mathbf{x}} + \sum_l \mathbf{dx_l}$ where $\bar{\mathbf{x}}$ refers to a large-scale (or approximation) component of $\mathbf{x}$ and $\{\mathbf{dx_l}\}_l$ to detail fields at different scales. In the subsequent, we assume $\bar{\mathbf{x}}$ to be known or estimated. As this large-scale component involves some spatial smoothing, an optimal interpolation arises as a natural solution.

Within an analog assimilation framework, we build data-driven priors for detail fields $\mathbf{dx_l}$, more specifically patch-based priors [14, 9, 15] to reduce the complexity of the analog prior to the patch size. We resort to the following formulation :

$$\mathbf{x} = \bar{\mathbf{x}} + \sum_{l=1}^{L} \mathbf{dx_l} \text{ with } \mathbf{dx_l}(t+1) = \mathcal{M}(\mathbf{dx_l}(t), \mathcal{C}_l) \quad (2)$$

with $\mathcal{M}(\mathbf{dx_l}(t), \mathcal{C})$ a patch-wise temporal Markovian prior :

$$\forall l, \mathbf{dx_l}(t+1)(\mathcal{P}_{r,l}) = \mathcal{M}(\mathbf{dx_l}(t)(\mathcal{P}_{r,l}), \mathcal{C}_{r,l}) \quad (3)$$

where $\mathcal{P}_{r,l}$ refers to a $W_l \times W_l$ patch around spatial position r at scale l and $\mathcal{C}_{r,l}$ to a catalog of exemplars for location r at scale l. We further constrain this patch-level prior through a PCA-based decomposition with N_l components : $\mathbf{dx_l}(t)(\mathcal{P}_{r,l}) = \sum_{k=1}^{N_l} \alpha_{t,r,k,l} B_k^l$. Using this low-dimensional representation of each patch, we state analog dynamic operator $\mathcal{M}(\mathbf{dx_l}(t)(\mathcal{P}_{r,l}), \mathcal{C}_{r,l})$ as a nearest-neighbor analog operator as introduced in the previous section. This dictionary-based decomposition also reduces the computational complexity of the proposed scheme, especially the forecast and update iterations of the implemented ensemble Kalman filters.

2.3. Numerical resolution

Given its multiscale structure, the resolution of Model (2) for a series of observations $\mathbf{y}$ exploits a classical coarse-to-fine strategy. It sequentially solves for detail components $\{\mathbf{dx_l}\}_l$ from $l = 1$ to $l = L$. At each scale l, a classic iterative stochastic or gradient-based resolution for the resulting spatiotemporal Markovian model [15] would result in an extremely large computational cost. We take advantage of the patch-wide temporal Markovian structure to approximate

the the global optimization as the resolution of independent patch-level reconstructions. The considered algorithm proceed as follows. At a given scale l, we perform a N_L-dimensional analog assimilation issue similar to [8] for each $\mathcal{P}_{r_n,l}$, where position r_n refers to a regular subsampling of the resolution grid with a view to reducing the overall computational complexity. We resort to a sequential ensemble Kalman smoother of the residual observation sequence with respect to the previous scale, i.e. $\mathbf{x}(t) - \mathcal{H}\left(\widehat{\overline{\mathbf{x}}}(t) + \sum_{l'=1}^{l-1} \mathbf{dx}_1', \Omega(t)\right)$. Given PCA basis $\{B_{k,l}\}$, the assimilation applies to decomposition coefficients $\{\alpha_{t,s,k,l}\}$ and observation operator $\mathcal{H}$ in (1) embeds the linear mapping associated with PCA basis $\{B_{k,l}\}$. Given all patch-level reconstructions at scale l, we apply a spatial averaging of the overlapping patches to build the final estimate $\widehat{\mathbf{dx}_1}$ at scale l.

3. NUMERICAL EXPERIMENTS

3.1. Case-study

To perform a qualitative and quantitative evaluation of the proposed data-driven framework, we consider an application to the reconstruction of satellite-derived SST (sea surface temperature) images. We focus on a region off South Africa, which involves active ocean dynamics and complex turbulence-related structures (fronts, filaments, ...). As illustrated in Fig.1, the considered satellite-derived SST images involve very large and non-uniform missing data patterns, what makes it challenging for our study.

With a view to numerically evaluating reconstruction performance, we proceed as follows. We use a reference gap-free SST time series, namely OSTIA product delivered daily by UK Met Office[1] with a $0.05°$ spatial resolution (approx. 5km) from January 2007 to December 2015. The OSTIA analysis combines satellite data from infrared and microwave sensors along with in situ data. Using the missing data patterns from METOP infrared sensor, we generate realistic high-resolution SST image series associated with a cloud-related irregular space-time sampling. Overall, we are provided with a 8-year daily time series of 600×300 images, i.e. a $\sim$150-000-dimensional state-space when removing land pixels of the case-study area.

3.2. Experimental setting

In our experiments, We evaluate for three methods.
Optimal interpolation (OI) : as baseline approach, we consider an optimal interpolation using a Gaussian prior. OI refers to the operational state-of-the-art procedure for SST interpolation [1]. Empirically, We select an exponential covariance structure [16]. A spatial correlation length of 100km and a temporal correlation length of 3 days were empirically proven optimal in terms of RMSE values.

Global analog assimilation (G-AnEnKS) : we also consider a direct application of the analog assimilation to state $\mathbf{x}$. We apply a spatial PCA-based decomposition of state $\mathbf{x}$ using 200 components, which accounts for more than 99% of the total variance of $\mathbf{x}$. The PCA is trained with gap-free SST data from 2007 to 2014. We build the catalog of exemplars from both the gap-free dataset from 2007 to 2014 and the optimally-interpolated fields for year 2015. $K = 50$ neighbors in the analog forecasting led to the best reconstruction performance.

Multi-scale analog assimilation (MS-AnEnKS) : We consider a two-scale model ($L = 2$ in (Eq.2)). For the coarse-scale component $\overline{\mathbf{x}}$, we consider an optimal interpolation with the same parameterization as described above. The intermediate scale ($l = 1$) involves 40×40 patches and the finest one ($l = 2$) 20×20 patches. We use PCA decompositions with 10 components for both scales. They are trained from the gap-free SST data from 2007 to 2014. We build catalogs of exemplars from both the gap-free SST data from 2007 to 2014 and the optimally-interpolated SST data for year 2015. We consider position-specific catalogs such that for a given patch center r only patches from the same location are considered. $K = 50$ neighbors in the analog forecasting led to the best reconstruction performance.

For each method, we compute the daily time series of the RMSE values between the reconstructed SST field and the reference one for year 2015 as well as the mean RMSE values.

3.3. Results

We report in Table 1 the reconstruction performance of the considered approaches. The proposed multiscale analog assimilation clearly outperforms the optimal interpolation (0.20 vs. 0.42 in terms of relative RMSE), with a mean relative gain of about 50%. By contrast, the direct application of a global analog assimilation only leads to a marginal RMSE gain (0.38 vs. 0.42 in terms of relative RMSE). These results point out the relevance of the multiscale decomposition to define consistent analog priors. The analysis of the time series of the RMSE values in Fig.2 further illustrates these results. The RMSE values of the multiscale analog assimilation depict a

Table 1. Reconstruction performance of the optimal interpolation (OI), the global (G-AnEnKS) and multi-scale (MS-AnEnKS) analog assimilation. We report the mean relative MSE (Mean Square Error) for a 1-year daily time series.

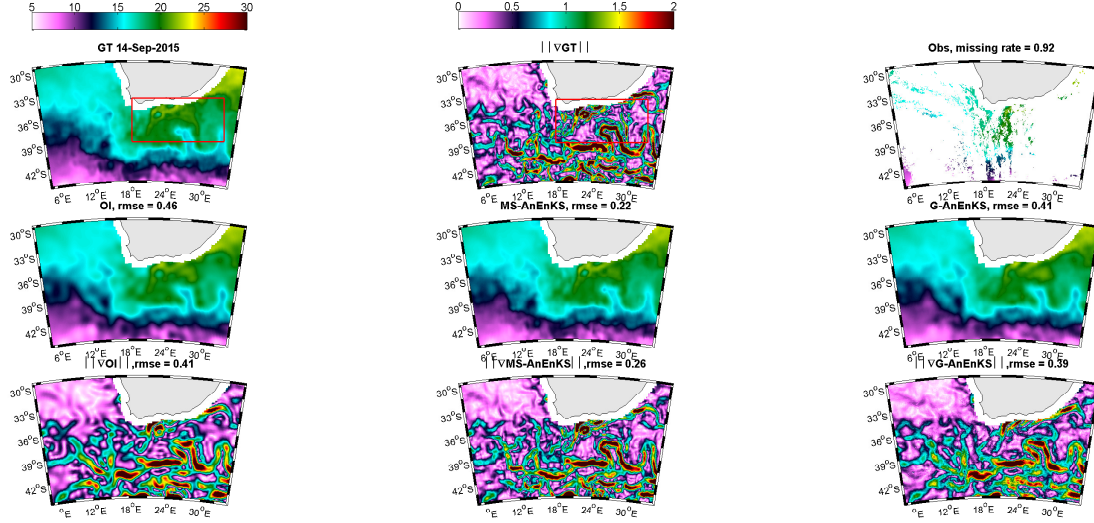| Method	Relative MSE
OI	0.42 ± 0.13
G-AnEnKS	0.38 ± 0.09
MS-AnEnKS	$\mathbf{0.20 \pm 0.06}$

Fig. 1. Reconstruction of a SST image (September 14, 2015) involving a high missing data rate (92%) : first row, reference SST field (groundtruth (GT)), associated gradient magnitude, observed missing data pattern; second row, interpolated images using OI, MS-AnEnKS and G-AnEnKS; third row, gradient magnitude fields of the three reconstructions.
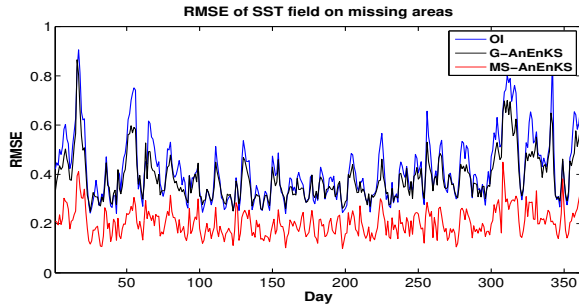

Fig. 2. Time series of the reconstruction error : relative MSE of OI (blue,-), G-AnEnKS (black,-) and MS-AnEnKS (red,-).

lower standard deviation (0.06 vs. 0.13 for OI and 0.09 for G-AnEnKS). Besides, whereas OI and G-AnEnKS reconstructions depict greater RMSE values during the austral summer, this seasonal feature of the RMSE is much reduced using MS-AnEnKS. This suggests a greater robustness of MS-AnEnKS to the temporal variability of the missing data patterns.

We further illustrate in Fig.1 the relevance of the reconstructed fields for a day corresponding to a missing data rate of 92%. Visual differences can be for instance observed in the region South of the African coast highlighted by a red box. In this area, MS-AnEnKS clearly retrieves thinner structures compared with GS-AnEnKS and OI (for instance, the observed light blue core). The reconstruction of the gradient magnitude is also clearly improved. Overall, this example illustrates the greater ability of the proposed approach to reconstruct fine-scale details from very partial observations using the prior knowledge brought by the multi-scale catalogs of exemplars. By contrast, neither the OI nor the global analog

assimilation model can reconstruct such fine-scale details in large missing data areas. They involve a very similar oversmoothing of the local gradients. The analysis of the spectral properties of the reconstructed fields (not shown due to space limitation) further illustrates this point.

4. CONCLUSION

In this paper, we developed a novel framework towards model-free and data-driven reconstruction of irregularly-sampled image time series with very large missing data rates (between 50% and 100%). It significantly outperformed the classical model-driven approach, namely optimal interpolation. Whereas previous data-driven models [6, 7] were limited to low-dimensional problems (typically, up to 100 dimensions), our key contribution lies in the design of a multiscale patch-based and PCA-constrained setting to make feasible and efficient the application of such data-driven assimilation models to image spaces (here, 300×600 images). Our model may be regarded as an extension of non-local patch-based models [9] to spatiotemporal fields, where the proposed patch-wise temporal Markovian prior results in an efficient sequential optimization using Kalman methods.

Our results open new research avenues for other applications of data-driven computational models to the reconstruction of spatiotemporal dynamics, especially in the field of earth science where large-scale observation and simulation datasets are available. Future work will explore such applications, especially the synergy between observation and simulation data, as well as the theoretical analysis of analog assimilation performance for non-Gaussian dynamical systems.

5. REFERENCES

[1] C. J. Donlon, M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Xindong, "The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system," *Remote Sensing of Environment*, vol. 116, pp. 140–158, Jan. 2012.

[2] P. Tandeo, P. Ailliot, and E. Autret, "Linear Gaussian state-space model with irregular sampling : application to sea surface temperature," *Stochastic Environmental Research and Risk Assessment*, vol. 25, no. 6, pp. 793–804, Nov. 2010.

[3] J. Li and A. D. Heap, "A review of comparative studies of spatial interpolation methods in environmental sciences : Performance and impact factors," *Ecological Informatics*, vol. 6, no. 3–4, pp. 228–241, July 2011.

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

[5] F.X. Le-Dimet and O. Talagrand, "Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects," *Tellus*, pp. 97–110, 1986.

[6] 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 *Machine Learning and Data Mining Approaches to Climate Science*, V. Lakshmanan, E. Gilleland, A. McGovern, and M. Tingley, Eds., pp. 3–12. Springer International Publishing, 2015, DOI : 10.1007/978-3-319-17220-0_1.

[7] F. Hamilton, T. Berry, and T. Sauer, "Ensemble Kalman Filtering without a Model," *Physical Review X*, vol. 6, no. 1, pp. 011021, Mar. 2016.

[8] R. Lguensat, P. Tandeo, P. Aillot, and R. Fablet, "The Analog Data Assimilation," *Submitted.*

[9] Gabriel Peyré, Sébastien Bougleux, and Laurent D. Cohen, "Non-local Regularization of Inverse Problems," *Inverse Problems and Imaging*, vol. 5, no. 2, pp. 511–530, 2011.

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

[11] A. Newson, A. Almansa, M. Fradet, Y. Gousseau, and P. Pérez, "Video Inpainting of Complex Scenes," *SIAM Journal on Imaging Sciences*, vol. 7, no. 4, pp. 1993–2019, Jan. 2014.

[12] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, pp. 969–985, 2007.

[13] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel, "Axioms and fundamental equations of image processing," *Archive for rational mechanics and analysis*, vol. 123, no. 3, pp. 199–257, 1993.

[14] 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*, June 2005, vol. 2, pp. 60–65 vol. 2.

[15] W. T. Freeman and Liu, "Markov Random Fields for Super-Resolution," in *Advances in Markov Random Fields for Vision and Image Processing*. MIT Press, a. blake, p. kohli, and c. rother, eds. edition, 2011.

[16] P. Tandeo, E. Autret, B. Chapron, R. Fablet, and R. Garello, "SST spatial anisotropic covariances from METOP-AVHRR data," *Remote Sensing of Environment*, vol. 141, pp. 144–148, Feb. 2014.