

# Assimilation par Analogues de Données Altimétriques Nadir et SWOT dans la Mer Méditerranée Occidentale

M. Lopez-Radcenco<sup>1</sup> A. Pascual<sup>2</sup> L. Gomez-Navarro<sup>2,3</sup> A. Aissa-El-Bey<sup>1</sup> R. Fablet<sup>1</sup> \*

(1) IMT Atlantique, Lab-STICC, UBL, Brest, France

(2) Mediterranean Institute for Advanced Studies (IMEDEA) (UIB-CSIC), Esporles, Spain

(3) Université Grenoble Alpes, CNRS, IRD, IGE, Grenoble, France

manuel.lopezradcenco@imt-atlantique.fr

## 1 Résumé

La disponibilité toujours croissante de données in situ, de télédétection et de simulation soutient le développement de solutions alternatives aux méthodes classiques d'interpolation des champs géophysiques de la surface de la mer à partir d'observations satellitaires partielles, qui sont, à l'heure actuelle, majoritairement fondées sur des modèles et n'exploitent pas ces nouveaux ensembles de données. À cet égard, nous avons récemment introduit l'Analog Data Assimilation (AnDA), qui exploite des opérateurs de prédiction par analogues basés sur des patches dans un cadre classique d'assimilation de données utilisant un filtre de Kalman. Dans ce travail, nous considérons l'application d'AnDA à l'interpolation spatio-temporelle des données SLA (Sea Level Anomalies) à partir de deux types de données altimétriques satellitaires, à savoir des données au long des traces altimétriques nadir [10] et des données de la future mission altimétrique SWOT [5]. L'analyse comparative rapportée soutient la pertinence d'AnDA pour une reconstruction améliorée des structures de méso-échelle pour des échelles horizontales allant de ~20km à ~100km, avec un gain de 42% (12%) en termes d'EQM (corrélation) pour la SLA, par rapport à l'Interpolation Optimale (OI) [1]. Les résultats suggèrent une amélioration potentielle supplémentaire à partir de l'assimilation conjointe des observations SWOT et des observations nadir au long de la trace.

## 2 Introduction

Au cours des dernières années, des progrès en télédétection, mesures in situ et modèles numériques nous ont permis d'avancer considérablement dans notre compréhension des dynamiques des océans et des interactions entre l'océan et l'atmosphère. Dans ce contexte, de grandes quantités de données provenant d'une grande variété de sources sont recueillies chaque jour. La fusion et le traitement de tels ensembles de données, qui impliquent généralement des schémas d'échantillonnage irrégulier et qui présentent données manquantes (dus à l'occlusion des nuages, les schémas d'échantillonnage des satellites, etc.), pour générer des produits L4 est d'une très grande importance pour l'océanographie et la météorologie. L'interpolation des champs réguliers à partir de données échantillonnées de façon irrégulière appartient à la famille des problèmes inverses, pour lesquels plusieurs stratégies basées sur des modèles ont été proposées [3, 1]. L'interpolation optimale (OI) [1], une stratégie basée sur la modélisation de la covariance spatio-temporelle du champ d'intérêt, est l'une des méthodes d'interpolation les plus couramment utilisées. Tenant compte de l'énorme variété de jeux de données de simulation à haute résolution disponibles aujourd'hui, les stratégies basées sur les données ont récemment été explorées comme une alternative puissante et efficace aux approches basées sur des modèles [11, 4].

Dans [4], nous avons démontré la pertinence de l'assimilation par analogues (AnDA), une approche basée sur les données, introduit dans [8], pour l'interpolation de champs géophysiques de grande dimension. Ici, nous développons une application d'AnDA à la reconstruction des champs de SLA dans la mer Méditerranée Occidentale à partir de données altimétriques satellitaires. La Méditerranée Occidentale est caractérisée par des rayons de Rossby relativement petits, ce qui rend la reconstruction des dynamiques de la surface de la mer à méso-échelle à partir des données satellitaires particulièrement difficile. Nous explorons deux schémas d'échantillonnage spatial différents : celui associé aux missions altimétriques classiques, avec un échantillonnage au long de la trace nadir [10] et celui de la future mission SWOT [5]. Nos expériences exploitent un OSSE (Observing System Simulation Experiment) avec des schémas d'échantillonnage spatio-temporels réels pour comparer AnDA à l'état de l'art pour l'interpolation des champs géophysiques.

\*Ce travail a été soutenu par l'Agence Nationale de la Recherche (ANR-13-MONU-0014), projet Labex Cominlabs SEACS et projet OSTST MANATEE. Les auteurs tiennent aussi à remercier R. Lguensat et P. Viet pour leur aide.

### 3 Assimilation par analogues

L'assimilation par analogues [8] est une méthode d'assimilation de données pour des systèmes dynamiques, entièrement pilotée par les données. Il repose sur l'hypothèse qu'un modèle dynamique peut être construit directement à partir d'un catalogue de simulations réalistes de la dynamique de l'état du système. L'idée générale est qu'une prédiction peut être faite en exploitant la similarité entre l'état actuel et des états simulés dans le catalogue. Formellement, nous utilisons une formulation de type espace-état [3] :

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

où  $t$  est un index temporel discret,  $\mathbf{x}$  est la séquence d'états cachés à reconstruire et  $\mathbf{y}$  est la séquence de données observées.  $\mathcal{M}$  est un modèle dynamique reliant l'état actuel  $\mathbf{x}(t)$  à l'état précédent  $\mathbf{x}(t - \delta t)$ .  $\mathcal{H}$  est un opérateur d'observation, où  $\Omega(t)$  est un masque représentant les données manquantes à l'instant  $t$  et  $\eta$  est un bruit aléatoire qui tient compte des incertitudes.

En pratique, à chaque pas de temps,  $\mathcal{M}$  est approximé linéairement en utilisant les  $K$  voisins les plus proches de l'état actuel  $\mathbf{x}(t)$  dans le catalogue, appelés les analogues. Les états futurs de ces analogues sont utilisés pour produire une prédiction, remplaçant ainsi l'étape classique de prédiction par modèle par une approche basée sur les données. L'opérateur de prédiction résultant peut être utilisé comme un plug-in sur des schémas classiques d'assimilation stochastique. Ici, nous considérons un filtre/lisseur de Kalman d'Ensemble (EnKFS) pour assimiler des observations partielles échantillonnées à un pas de temps  $\delta t$ . Il est important de noter que les observations assimilées peuvent être des observations aux temps  $t_0, t_0 + \delta t, \dots$  ou, alternativement, des pseudo-observations construites en accumulant des observations sur une fenêtre temporelle  $t_0 \pm D$ .

Il convient également de noter que, compte tenu de la nature hautement dimensionnelle des données à interpoler, l'application d'une technique de réduction de dimensionnalité est nécessaire avant la recherche des plus proches voisins dans le catalogue. À cet égard, suivant [4], AnDA utilise une représentation basée sur des patches [11]. Le champ à interpoler est décomposé en patches  $\mathcal{P}(s, t)$  de taille  $W_p \times W_p$ , à l'emplacement  $s$  et au temps  $t$ , et une décomposition ACP est appliquée pour chaque patch :  $\mathcal{P}(s, t) = \sum_{k=1}^{N_{ACP}} \alpha_k(s, t) B_k$ , où  $B_k$  sont les composants principaux et  $\alpha_k(s, t)$  sont les coefficients de la décomposition du patch  $\mathcal{P}(s, t)$  sur ses composants principaux. L'interpolation du champ  $\mathbf{x}$  se fait en assimilant indépendamment chaque patch, pour ensuite moyenner les patches superposés.

### 4 Simulations et résultats

Comme cas d'étude, nous considérons une région dans la Méditerranée Occidentale ( $36.5^\circ N$  à  $40^\circ N$ ,  $1.5^\circ E$  à  $8.5^\circ E$ ). Nous utilisons des simulations numériques quotidiennes ROMS [7] de 2010 à 2013, avec une résolution  $1/20^\circ$ , comme référence. Nous mettons en œuvre des expériences de simulation d'un système d'observation (OSSE) pour générer des observations synthétiques des données altimétriques satellitaires. Pour les données nadir au long de la trace (AT), nous utilisons des localisations spatio-temporelles de trajectoires satellitaires réelles à partir d'une configuration à quatre altimètres en 2014. Un bruit blanc Gaussien de variance  $\sigma_n^2$  est ajouté pour simuler le bruit d'acquisition. Pour évaluer l'effet de la prise en compte de différents modèles d'échantillonnage d'observation, nous générons aussi des observations pseudo-SWOT à partir des champs de SLA haute résolution simulés en utilisant le simulateur SWOT, ou [6]. le bruit d'observation incorpore différents processus de bruit simulés [6, 2]. Nous appliquons AnDA à l'anomalie entre la SLA et une première prédiction par OI, en utilisant les données de 2010-2012 comme catalogue, et nous évaluons la performance de l'assimilation sur les données de 2013. Nous étudions les performances d'interpolation en termes d'erreur quadratique moyenne (EQM) et de coefficient de corrélation pour le champ SLA prédit et le gradient du champ prédit  $\nabla \text{SLA}$ . Nous rapportons aussi les performances de reconstruction de OI [1], et d'une approche de décomposition non-négative de modèles convolutionnelles pilotée par les données (NNLD), introduite dans [9] pour l'interpolation de champs géophysiques irrégulièrement échantillonnés, comme référence. Toutes les expériences rapportées utilisent les paramètres suivants :  $W_p = 30$ ,  $K = 25$ ,  $\sigma_n^2 = 0$ ,  $\delta t = 1$ ,  $D = 0$ ,  $N_{EOF} = 9$ .

Dans le tableau 1, nous rapportons l'EQM et la corrélation pour les champs de SLA et  $\nabla \text{SLA}$  obtenus en considérant les données nadir au long de la trace, les données SWOT et une combinaison des données nadir et SWOT sans bruit, pour des observations journalières et pour des observations accumulées sur une fenêtre temporelle  $t_0 \pm D$ , avec  $D = 5$  jours. Nous pouvons conclure que, même si les meilleurs résultats en termes de performance de reconstruction de SLA sont obtenus en considérant les données nadir accumulées sur 5 jours, le fait d'ajouter des données SWOT semble améliorer les résultats en termes de reconstruction du  $\nabla \text{SLA}$ . Cependant, il convient également de noter que l'accumulation d'observations SWOT semble avoir un effet négatif sur la reconstruction SLA, du au fait que les observations SWOT à grande fauchée augmentent probablement la sensibilité d'AnDA aux changements survenant dans le champ SLA pendant les jours où les observations sont accumulées. À cet égard, il semble y avoir un compromis entre l'accumulation des données sur plusieurs jours, ce qui améliore les performances de reconstruction de SLA, et l'utilisation des données SWOT, ce qui améliore les résultats en termes de reconstruction de  $\nabla \text{SLA}$ , mais augmente la sensibilité aux incohérences entre les observations accumulées.

Enfin, il convient de noter que l'accumulation d'observations nadir sur plusieurs jours et l'utilisation des observations SWOT à large fauchée (seules ou combinées avec des observations nadir) semblent être toutes les deux des stratégies efficaces

	$D = 0$		$D = 5$	
	$SLA$	$\nabla SLA$	$SLA$	$\nabla SLA$
AT	0,02397 (0,9197)	0,005524 (0,7023)	<b>0.01966 (0.9464)</b>	0.004687 (0.7679)
SWOT	0.02156 (0.9388)	0.004515 (0.7726)	0,02523 (0,9200)	0,004474 (0,7678)
AT+SWOT	0.02043 (0.9445)	0.004448 ( <b>0.7805</b> )	0.02428 (0.9254)	<b>0.004424</b> (0.7745)
OI [1]	0.03389 (0.8447)	0.006661 (0.6050)	-	-
NNLD [9]	0.02127 (0.6957)	0.004513 (0.5892)	-	-

TABLE 1 – EQM (Corrélation) pour AnDA combinant les observations SWOT et AT. Meilleur résultats en **gras**. Résultats pour OI [1] et NNLD [9] donnés comme référence.

pour surpasser OI et NNLD. En effet, comparé à OI (NNLD), AnDA présente une EQM inférieure (similaire) tout en ayant des coefficients de corrélation plus élevés, ce qui peut être interprété comme une meilleure reconstruction des structures à méso-échelle.

## 5 Conclusion

Nous avons développé une application de l’assimilation par analogues (AnDA) à la reconstruction des champs de SLA dans la Méditerranée Occidentale à partir des données altimétriques satellitaires, à savoir les données satellites nadir au long de la trace et les données SWOT à large fauchée. AnDA peut être considéré comme un moyen d’exploiter des ensembles de données de simulation numérique à haute résolution pour la reconstruction des champs de SLA à partir d’observations satellitaires partielles. Les expériences rapportés supportent la pertinence d’AnDA par rapport aux approches de l’état de l’art (OI [1, 10] et NNLD [9]). En effet, nous rapportons une amélioration de 42% (12%) en termes de SLA RMSE (corrélation) et 30% (27%) en termes de  $\nabla$  SLA RMSE (corrélation) par rapport à OI en considérant des données nadir le long de la trace. Nos expériences suggèrent également une amélioration potentielle supplémentaire à partir de l’assimilation conjointe des observations SWOT et des observations nadir au long de la trace.

Les travaux futurs se concentreront sur la combinaison d’AnDA avec des stratégies de pré-traitement des données SWOT, afin de développer des outils efficaces pour le traitement des observations réelles de la future mission altimétrique SWOT. D’autres pistes de recherche intéressantes comprennent la combinaison des différentes sources de données altimétriques ainsi que d’autres traceurs dynamiques océaniques (par ex. température/salinité de la surface de la mer, couleur de l’océan, etc.). L’exploitation des informations structurelles présentes dans les observations à fauchée large, par exemple au moyen de gradients de SLA et/ou des exposants de Liapunov à taille finie (FSLE), est également une direction de recherche attirante.

## Références

- [1] R. ESCUDIER et al. “Improvement of coastal and mesoscale observation from space : Application to the northwestern Mediterranean Sea”. In : *Geophys. Res. Lett.* 40.10 (2013), p. 2148–2153.
- [2] D. ESTEBAN-FERNANDEZ. *SWOT project mission performance and error budget document*, JPL Doc. JPL D-79084. Rapp. tech. JPL, NASA, 2014.
- [3] G. EVENSEN. *Data Assimilation*. en. Berlin, Heidelberg : Springer Berlin Heidelberg, 2009.
- [4] R. FABLET, P. H. VIET et R. LGUENSAT. “Data-driven Models for the Spatio-Temporal Interpolation of satellite-derived SST Fields”. In : *IEEE Trans. Comput. Imag.* (2017).
- [5] L.L. FU et R. FERRARI. “Observing Oceanic Submesoscale Processes From Space”. In : *Eos, Trans. AGU* 89.48 (2008), p. 488–488.
- [6] L. GAULTIER et C. UBELMANN. *SWOT Simulator Documentation*. Rapp. tech. JPL, NASA, 2010.
- [7] M. JUZA et al. “SOCIB operational ocean forecasting system and multi-platform validation in the Western Mediterranean Sea”. In : *J. Oper. Oceanogr.* 9.sup1 (fév. 2016), s155–s166.
- [8] R. LGUENSAT et al. “The Analog Data Assimilation”. In : *Mon. Weather Rev.* (2017).
- [9] M. LÓPEZ-RADCENCO et al. “Locally-adapted convolution-based super-resolution of irregularly-sampled ocean remote sensing data”. In : *ArXiv e-prints* (avr. 2017). arXiv : 1704.02162 [stat.ML].
- [10] A. PASCUAL et al. “Mesoscale mapping capabilities of multisatellite altimeter missions : First results with real data in the Mediterranean Sea”. In : *J. Mar. Syst.* 65.1–4 (mar. 2007), p. 190–211.
- [11] G. PEYRÉ, S. BOUGLEUX et L.D. COHEN. “Non-local Regularization of Inverse Problems”. In : *Inverse Probl. and Imag.* 5.2 (2011), p. 511–530.