



# Interpolation de données manquantes dans des séquences multi-modales d'images géophysiques satellitaires

Sileye Ba, Thomas Corpetti, Fablet Ronan

## ► To cite this version:

Sileye Ba, Thomas Corpetti, Fablet Ronan. Interpolation de données manquantes dans des séquences multi-modales d'images géophysiques satellitaires. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. pp.978-2-9539515-2-3, 2012. <hal-00656503>

**HAL Id: hal-00656503**

**<https://hal.archives-ouvertes.fr/hal-00656503>**

Submitted on 17 Jan 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Interpolation de données manquantes dans des séquences multi-modales d'images géophysiques satellitaires

Silèye O. Ba <sup>†</sup>

Thomas Corpetti <sup>‡</sup>

Ronan Fablet <sup>†</sup>

Lab-STICC<sup>†</sup>, Université Européenne de Bretagne, Plouzané, France

LIAMA-CNRS<sup>‡</sup>, Beijing, Chine

## Résumé

*Cet article étudie l'estimation conjointe de données manquantes et de champs de déplacements dans des séquences multimodales d'observations satellitaires géophysiques. La complexité de la tâche est liée au taux élevé de données manquantes (entre 20% et 90%) pour des observations journalières de haute résolution et la reconstruction de structures fines en accord avec la dynamique sous-jacente. Nous avons développé une méthode basée sur l'assimilation variationnelle de données pour des séries multimodales et multi-résolutions. A l'aide de données synthétiques et de données réelles de la surface océanique, une évaluation numérique et qualitative démontre l'apport de deux composantes clés du modèle proposé : la fusion d'informations multimodales à partir d'une contrainte géométrique basée sur les structures frontales, et la méthode d'assimilation variationnelle utilisant comme à priori dynamique un modèle d'advection-diffusion. Les expérimentations conduites montrent que de bonnes performances de reconstruction sont obtenues pour les observations hautes résolutions en dépit du pourcentage élevé de données manquantes.*

## Mots Clef

Assimilation variationnelle de données, interpolation de données manquantes, inpainting

## Abstract

*In this paper we address the joint estimation of missing data and displacement field estimation from multi-modal geophysical satellite observation series. The complexity of this issue lies in the large percentage of missing data (typically from 20% to 90% for daily high-resolution observations) and the requirement for reconstructing dynamically coherent fine-scale structures in accordance with the underlying turbulent dynamics. We develop a variational data assimilation model for multi-resolution and multi-modal observation series. Using synthetic and real ocean surface observation series, numerical and qualitative evaluations demonstrate the relevance of two key components of the proposed model : the fusion of multi-modal observations through a geometric front-driven constraint and the proposed variational assimilation setting using an advection-diffusion dynamical prior. We show that good reconstruc-*

*tions of high-resolution geophysical observation series can be achieved despite high percentage of missing data.*

## Keywords

variational data assimilation, missing data interpolation, inpainting

## 1 Introduction

De nos jours la surface terrestre et les océans sont couverts par un réseau dense de satellites. De multiples satellites capturent des paramètres terrestres et océaniques à différentes résolutions. La température de la surface océanique (SST) en est un exemple particulier. La radiométrie micro-onde (MO) permet une capture à basse résolution (0.25°) [1] tandis que la radiométrie infrarouge (IR) délivre des mesures hautes résolutions (0.05°) [2].

La mesure de la SST est une activité importante car la SST est une observation clé dans des domaines tels que la prédiction météorologique, l'étude de la circulation de l'océan et des effets du changement climatique [3, 4, 5, 6, 7]. Il est aussi à noter la relation établie entre la pluviométrie en Afrique et la température de surface des océans tropicaux [3] ou la corrélation positive entre le réchauffement de la température de surface océanique et l'augmentation de l'activité cyclonique dans l'océan Atlantique [4, 5]. Ainsi, la SST est un marqueur de la dynamique océanique [6, 7]. La concentration de chlorophylle est aussi un paramètre océanique important. C'est un marqueur effectif de la production biologique de l'océan [8]. En tant que marqueur passif de l'activité océanique, les structures géométriques de la concentration de CHL sont fortement liées à celles exhibées par la SST. Toutes deux (CHL et SST), sont associées à la circulation océanique globale et ainsi qu'aux structures de turbulence aux échelles fines [9]. Ainsi, la concentration de CHL et la SST sont des paramètres en forte interaction, et cette interaction pourrait être prise en compte lors de leur traitement [10, 11].

Les satellites géophysiques exploitent plusieurs modalités (e.g. capteurs IR ou MO) associés à différentes résolutions de capture comme illustré dans la Fig.1. Dans tous les cas, ces capteurs sont sensibles aux conditions atmosphériques telles que les fortes pluies ou la couverture nuageuse. Par conséquent, les séries d'observations comportent de fort pourcentage de données manquantes, les données hautes

résolutions étant plus affectées. Par exemple, dans la région de Malvinas considérée (voir Fig.1), les observations journalières de CHL en comportent entre 20% à 90%. L'interpolation de données manquantes dans les observations géophysiques est alors une étape de pré-traitement nécessaire [12, 13, 14]. Les produits opérationnels sont basés sur les techniques d'interpolation optimale, du filtrage ou du lissage de Kalman, ou des ensembles de filtres de Kalman [13, 14, 15, 16]. Leur limitations principales sont qu'ils requièrent le traitement de très larges matrices de covariance, et font des hypothèses statistiques telles que la stationnarité, la Gaussiannité ou la linéarité sur le modèle dynamique. Ces hypothèses peuvent difficilement être réalisées pour des problèmes réels. En présence de fort taux de données manquantes (voir Fig.1), un gain peut être attendu lors de la fusion d'observations multimodales et multi-résolutions. L'interpolation conjointe d'observations multi-résolution SST pourra améliorer le traitement des données hautes résolutions, étant données que les observations basses résolution comportent beaucoup moins de données manquantes. De plus, dans le cas de données issues de modalités différentes (SST et CHL), les structures géométriques communes aux deux paramètres géophysiques, serviront à mieux contraindre le processus d'interpolation.

De façon formelle, la reconstruction d'observations géophysiques peut être posée sous la forme d'interpolation de données manquantes à partir de séquence d'observations multimodales et multi-résolutions. Connues en vision par ordinateur sous la dénomination d'inpainting, les méthodes variationnelles ont été largement exploitées pour l'interpolation de données manquantes. Ces méthodes sont conçues pour supprimer du texte, des craquelures, ou même des objets d'images [18, 17]. Le problème considéré dans cet article admet aussi des liens avec le problème de super-résolution où des images basses résolutions d'une scène sont utilisées pour créer une version haute résolution de la même scène [20, 21, 22]. Contrairement au problème d'inpainting ou de super-résolution, le problème qui nous intéresse comprend une complexité additionnelle :

- Comparé aux applications d'inpainting, les observations comprennent de larges portions de données manquantes ce qui rend la reconstruction de la géométrie locale des structures présentes dans les images géophysiques particulièrement difficile ;
- Comparé aux problèmes classique d'inpainting et de super-résolution de champs multi-valués, dans le problème considéré dans cet article, les séries d'observations multimodales traitées ne partagent que l'organisation spatiale des structures géométriques ;
- Les images géophysiques sont fortement structurées par le champ de déplacement sous jacent [23]. L'estimation jointe du champ de vitesse de la surface océanique apparaît comme un facteur important pour recouvrer un champ dynamiquement cohérent.

Pour résoudre ces problèmes, nous proposons une méthode basée sur l'assimilation variationnelle qui, à partir de don-

nées multimodales et multi-résolution d'observations géophysiques, estime de façon conjointe les données manquantes et un champ de déplacement associé. Les performances des modèles proposées sont évaluées sur des données synthétiques (simulation de modèle turbulence semi quasi-géostrophique [7]), et des données SST et CHL.

La suite de cet article est organisé comme suit. La Section 2 présente le modèle variationnel pour l'estimation de données manquantes et de champs de transport pour une séquence d'observations d'une seule modalité. La Section 3 présente la méthode multimodale d'interpolation de données manquantes et de champs de transport. La Section 4 présente les expérimentations conduites pour l'évaluation des méthodes proposées. Finalement, la Section 5 donne les conclusions.

## 2 Interpolation de données manquantes pour une série uni-modale

Dans cette section, nous supposons disposer d'une séquence d'observations  $I_t, t \in [t_0, t_f]$  correspondant à la séquence d'états  $\theta_t, t \in [t_0, t_f]$  d'une variable géophysique. Dans notre cas, la variable d'état correspond soit à de la SST ou de la CHL. Les observations  $I_t$  comprennent des données manquantes comme dans la Fig.1. La variable d'état est supposée liée à l'observation qui lui correspond par la relation  $I_t = P_t \theta_t + \omega_t$  où  $P_t$  est un opérateur linéaire de projection qui modélise la réduction de dimensionnalité due aux données manquantes<sup>1</sup>.  $\omega_t$  représente un bruit Gaussien centré, indépendant, identiquement distribuée, de matrice covariance  $\Sigma$ .

Etant donné que les variables sont des marqueurs de la dynamique océanique, elles sont liées à la circulation océanique, leur reconstruction doit comprendre comme à priori un modèle dynamique. Supposons que l'évolution temporelle des états soit soumise au modèle dynamique suivant :

$$\begin{cases} \partial_t \theta + \mathbb{M}(\theta, \vartheta) = \eta_t \\ \theta_{t_0} = \theta_0 + \varepsilon \end{cases} \quad (1)$$

où  $\theta_0$  est l'état initial,  $\varepsilon$  et  $\eta_t$ , de matrices de covariances  $B$  and  $Q$ , modélisent l'incertitude à propos de l'état initial et du modèle dynamique<sup>2</sup>. Ici,  $\mathbb{M}(\theta, \vartheta) = \vartheta \nabla \theta - \kappa \Delta \theta$  est un opérateur d'advection-diffusion, défini par le champ d'advection  $\vartheta$  et le paramètre de diffusion  $\kappa$ .  $\nabla$  and  $\Delta$  dénotent l'opérateur gradient et Laplacien. Le modèle d'advection diffusion modélise le transport d'une variable par un champ de vecteurs  $\vartheta$  et un taux de diffusion  $\kappa$ .

Etant donné le modèle dynamique (Eq.1), nous définissons un coût variationnel pour l'interpolation de donnée manquantes et l'estimation de champs de transport par :

$$J(\theta, \vartheta) = \int_{t_0}^{t_f} \|\partial_t \theta + \mathbb{M}(\theta, \vartheta)\|_Q^2 dt + \|\theta_{t_0} - \theta_0\|_B^2 \dots \quad (2)$$

$$+ \int_{t_0}^{t_f} E(\theta_t) dt$$

<sup>1</sup>Les images sont représentées sous une forme vectorielle.

<sup>2</sup> $B$  et  $Q$  sont modélisés par des matrices diagonales et sphériques.

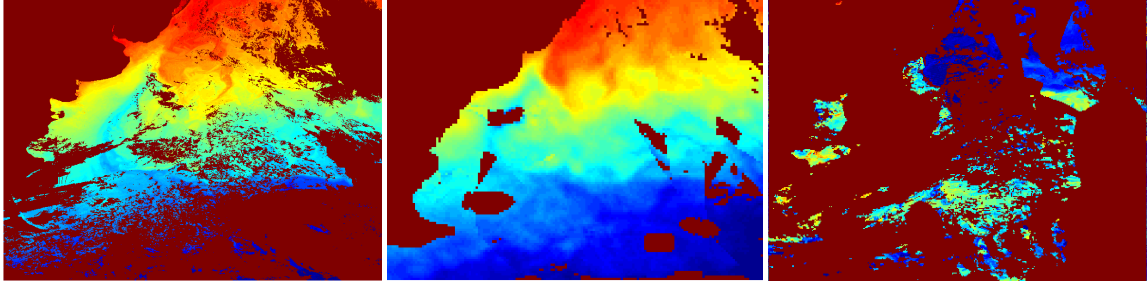


FIG. 1 – Observations géophysiques de l’océan pour la région de Malvinas située au large du Brésil. A droite : observation haute résolution IR (données METOP) ; au centre : observation basse résolution MO (données REMSS) ; à gauche : donnée medium résolution de mesures de concentration de CHL (données MERIS). Les données METOP possèdent une résolution 5 fois plus fine que les données REMSS, et deux fois plus fine que les données MERIS.

où  $E(\theta)$  est un terme d’attache aux observations défini par :

$$E(\theta) = \int_{\mathcal{D}(I)} (I - P\theta)^2 dp + \beta \int_{\mathcal{D}(\theta)} \|\nabla\theta\|_q^q dp \quad (3)$$

où  $\mathcal{D}$  dénote le domaine de définition de son argument,  $\beta$  est un poids positif pondérant la contribution du terme de régularité dans le terme d’attache aux observations  $E$ .  $\|\cdot\|_Q$  représente la distance de mahalanobis par rapport à la matrice de covariance  $Q$  et  $\|\cdot\|_q$  est la norme  $L^q$ . Nous avons choisi  $q = 2$ .

La séquence de variables d’état et de champs de transport  $(\hat{\theta}_t, \hat{\vartheta}_t), t \in [t_0, t_f]$  minimisant la fonction coût  $J(\theta, \vartheta)$  possède les propriétés suivantes. En présence de données, l’état  $\hat{\theta}_t$  est similaire à l’observation  $I_t$  à cause du premier terme de l’Eq.3. Sa régularité spatiale est contrainte par le second terme de l’Eq.3. L’évolution temporelle des états est contrainte par le modèle dynamique (Eq.1) à cause du premier terme de l’Eq.2.

En calcul des variations, une méthode standard pour résoudre des problèmes variationnels est le recours aux équations d’Euler-Lagrange. Par exemple, dans un cas statique l’état optimal  $\hat{\theta}$  pour le coût relatif aux observations  $E(\theta)$  peut être trouvé comme la valeur stationnaire de l’équation d’Euler-Lagrange  $\partial_t\theta = -\delta_\theta E$  où  $\delta_\theta E$  est la différentielle du terme d’attache aux observations. Cependant, pour le problème variationnel de l’Eq.2, une utilisation directe des équations d’Euler-Lagrange se révèle inefficace du fait de la grande dimensionnalité de l’espace d’état. Comme proposé par [24], une méthode indirecte consiste à introduire les variables adjointes  $\lambda_t = Q^{-1}(\partial_t\theta + \mathbb{M}(\theta_t, \vartheta_t))$ . Avec ces variables, calculer les variations du coût dynamique de l’Eq.2 par rapport à la variable d’état, le champ de transport, et la valeur initiale, et faire usage de la propriété que les variations s’annulent pour les valeurs optimales donnent le système d’équations :

$$\begin{aligned} \delta_{\theta_t} J &= -\partial_t\lambda + \delta_\theta \mathbb{M}^*(\lambda, \vartheta) + \delta_\theta E(\theta_t) = 0 \\ \delta_{\theta_{t_0}} J &= B^{-1}(\theta_{t_0} - \theta_0) - \lambda_{t_0} = 0 \\ \delta_{\vartheta_t} J &= \lambda_t \nabla\theta_t = 0 \end{aligned} \quad (4)$$

où  $\delta_\theta \mathbb{M}$  est la différentielle du modèle dynamique, et  $\delta_\theta \mathbb{M}^*$  est l’opérateur adjoint. Suivant une procédure itérative, la solution du problème variationnel peut être linéarisé autour de valeurs initiales  $\hat{\theta}_t$  comme suit :

$$\begin{aligned} \theta_t &= \hat{\theta}_t + d\theta_t \\ \hat{\theta}_{t_0} &= \theta_0 \end{aligned} \quad (5)$$

Sachant par définition des variables adjointes que  $\partial_t\theta + \mathbb{M}(\theta_t, \vartheta_t) = Q\lambda_t$ , et en combinant la deuxième ligne de l’Eq.4 et l’Eq.5 donne :

$$\begin{aligned} d\theta_{t_0} &= B\lambda_{t_0} \\ \partial_t d\theta + \delta_\theta \mathbb{M}(d\theta, \vartheta) &= Q\lambda_t \end{aligned} \quad (6)$$

Ainsi, les variations du coût peuvent être estimées de façon itérative par une intégration rétrograde des variables adjointes prenant en compte le coût statique des observations (première ligne de l’Eq.4), suivie par une intégration avant des variations prenant en compte les valeurs des variables adjointes (seconde ligne de l’Eq.6) [25, 26, 27]. Etant donné les variations, les solutions peuvent être mises à jour. Le champ de transport peut être directement mis à jour par descente de gradient à l’aide de la troisième ligne de l’Eq.4. Les étapes de l’algorithme d’assimilation variationnelle sont présentées dans la Fig.2.

### 3 Interpolation pour une série multimodale

La Section 2 présente une méthode pour l’estimation de données manquantes et de champs d’advection associés étant donné une séquence d’observations d’une variable géophysique. Cependant, en général, les variables géophysiques traitées ne sont pas indépendantes les unes des autres. Le champ de transport sous jacent comprend des structures de convergences et de divergences qui induisent la formation de structures frontales dans des variables passives [28]. Ainsi, les variables étudiées (CHL, SST) présentent des structures géométriques similaires dans les régions frontales. Cette propriété peut être utilisée pour in-

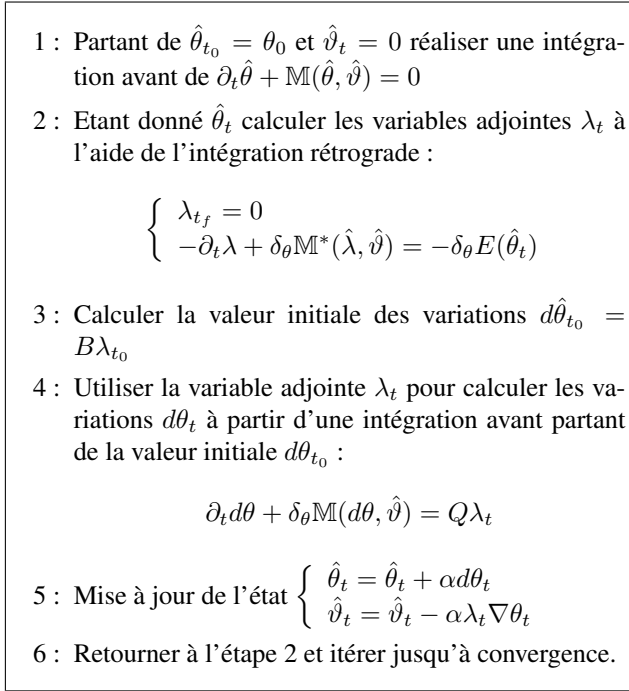


FIG. 2 – Algorithme d'assimilation variationnelle de données.  $\alpha > 0$  est un paramètre de descente de gradient.

roduire une contrainte de similarité géométrique additionnelle entre les différentes variables. En termes de représentation d'images, cette contrainte signifie que dans les différentes variables, les lignes de niveaux supportant les gradients de fortes magnitudes doivent être localement parallèles. Dans la suite, nous donnons les détails à propos du coût qui comprend cette contrainte additionnelle.

Notons par  $I_t^{(i)}, t \in [t_0, t_f], i = 1, 2$  les séries d'observations avec données manquantes correspondant aux séquences de variables  $\theta_t^{(i)}, t \in [t_0, t_f], i = 1, 2$ . Sans perte de généralités, nous supposons que  $\theta_t^{(1)}$  est de plus faible résolution spatiale et  $\Lambda$  est un opérateur linéaire de sous échantillonnage tel que  $\Lambda\theta_t^{(2)}$  possède la même dimensionnalité que  $\theta_t^{(1)}$ . Nous définissons le coût suivant :

$$\begin{aligned} \mathcal{E}(\theta^{(1)}, \theta^{(2)}) &= E(\theta^{(1)}) + E(\theta^{(2)}) \dots \\ &- \gamma \int_{\mathcal{D}(\theta^{(1)})} g_\tau(|\nabla\theta^{(1)}|) \rho_\epsilon \left( \left\langle \frac{\nabla\theta^{(1)}}{|\nabla\theta^{(1)}|}, \frac{\nabla\Lambda\theta^{(2)}}{|\nabla\Lambda\theta^{(2)}|} \right\rangle \right) dp \end{aligned} \quad (7)$$

où  $E(\theta)$  est le terme d'attache aux données défini dans l'Eq.3,  $g_\tau(x) = 1 - e^{-\tau x}$  for  $x > 0$  est une fonction de pondération qui, pour  $x = |\nabla\theta^{(1)}|$ , donne un poids plus fort aux pixels de  $\theta^{(1)}$  où le gradient est fort, c'est à dire aux pixels appartenant aux structures frontales et donne un poids faible aux pixels où la magnitude du gradient est faible. Le paramètre  $\tau > 0$ , de la fonction  $g_\tau$ , fixe la magnitude du gradient définissant les structures frontales. La fonction  $\rho_\epsilon(s) = \sqrt{s^2 + \epsilon}$ , avec  $\epsilon > 0$ , est une approximation différentiable de la valeur absolue.  $\gamma > 0$  pondère la

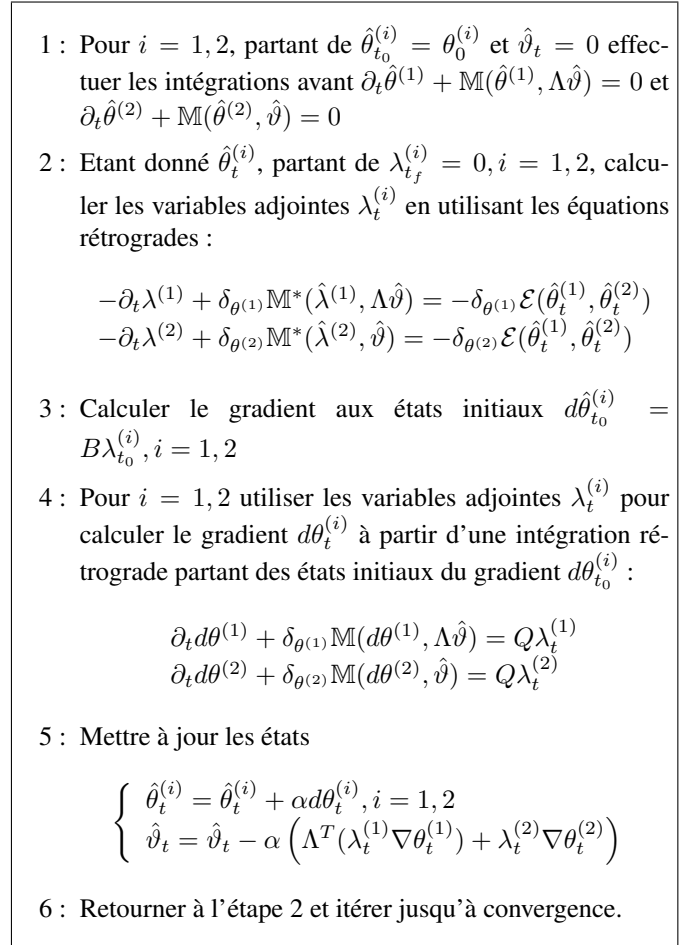


FIG. 3 – Algorithme pour l'estimation multimodale de données manquantes et de champs d'advection.

contribution du coût géométrique dans le terme d'attache aux observations  $\mathcal{E}(\theta^{(1)}, \theta^{(2)})$ . Les estimations  $\hat{\theta}^{(1)}$  et  $\hat{\theta}^{(2)}$  minimisant le coût  $\mathcal{E}(\theta^{(1)}, \theta^{(2)})$  seront similaires aux observations lorsqu'elles seront disponibles à cause des deux premiers termes de l'Eq.7. De plus,  $\hat{\theta}^{(2)}$  va présenter des structures frontales cohérentes aux structures frontales présentes dans la variable  $\hat{\theta}^{(1)}$  à cause du terme d'a priori géométrique qui est minimisée lorsque les deux variables ont leur structures frontales alignées.

Le modèle d'observation couplé et le modèle dynamique d'advection-diffusion (Eq.1) peuvent être combinés pour construire le coût dynamique variationnel couplé

$$\mathcal{J}(\theta^{(1)}, \theta^{(2)}, \vartheta) = \int_{t_0}^{t_f} \|\partial_t \theta^{(1)} + \mathbb{M}(\theta^{(1)}, \Lambda\vartheta)\|_Q^2 dt \dots \quad (8)$$

$$+ \int_{t_0}^{t_f} \|\partial_t \theta^{(2)} + \mathbb{M}(\theta^{(2)}, \vartheta)\|_Q^2 dt + \|\theta_{t_0}^{(1)} - \theta_0^{(1)}\|_B^2 \dots \quad (9)$$

$$+ \|\theta_{t_0}^{(2)} - \theta_0^{(2)}\|_B^2 + \mathcal{E}(\theta^{(1)}, \theta^{(2)}) \quad (10)$$

où nous avons supposé que le champ d'advection est le même pour les deux variables avec  $\vartheta$  étant sa version haute

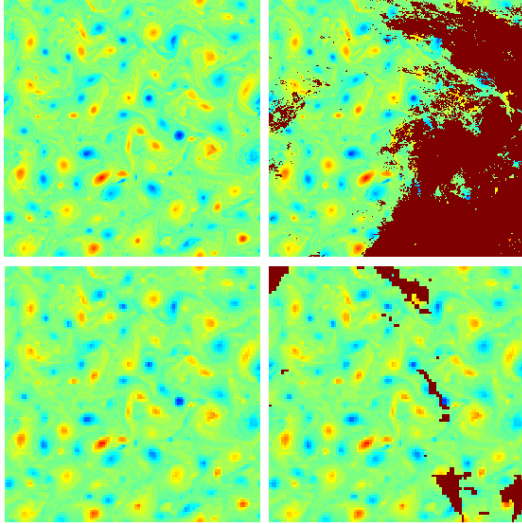


FIG. 4 – Données SQG : (haut ;gauche), données vérité hautes résolutions ; (haut ;droite), observation haute résolution obtenue en masquant les données vérité par un masque issu de la base METOP ; (bas ;gauche), image basse résolution obtenue en sous échantillonnant à un ratio de  $\frac{1}{5}$  la donnée vérité ; (bas ;droite), observation basse résolution obtenue en masquant la donnée basse résolution par un masque issu de la base REMSS.

résolution agissant sur  $\theta^{(2)}$  et  $\Lambda\vartheta$  étant sa version basse résolution agissant sur la variable basse résolution  $\theta^{(1)}$ .

Comme dans la Section 2, le problème dynamique variationnel peut être résolu dans un cadre d'assimilation variationnelle en introduisant les variables adjointes  $\lambda_t^{(1)} = Q^{-1}(\partial_t \theta^{(1)} + \mathbb{M}(\theta^{(1)}, \Lambda\vartheta))$  et  $\lambda_t^{(2)} = Q^{-1}(\partial_t \theta^{(2)} + \mathbb{M}(\theta^{(2)}, \vartheta))$ . L'utilisation des variables adjointes et le calcul des variations du problème dynamique couplé par rapport aux variables à estimer conduisent au système d'équations qui est utilisé pour calculer de manière itérative les solutions optimales. La Fig.3 donne les étapes principales de l'algorithme correspondant.

## 4 Expérimentations

Pour évaluer les méthodes présentées dans la Section 2 et 3, l'analyse qualitative conduite sur l'expérimentation avec des données réelles est complétée par une évaluation quantitative sur des données issues de simulations.

### 4.1 Résultats sur des données SQG

Le premier protocole d'évaluation utilise des données de simulation numérique SQG (surface quasi-géostrophique) [7]. Le choix de ce modèle est motivé par de récents résultats qui montrent que sous certaines conditions, il décrit bien la dynamique océanique [9]. Etant donné des données vérité hautes résolutions, une séquence d'images de basses résolutions est générée par sous échantillonnage. Des masques de données manquantes extraits des bases de

données d'observations réelles sont appliqués pour simuler des données manquantes ayant une configuration réaliste. Pour les données basses résolutions, les masques de données manquantes sont extraits de la base de données REMSS. Les masques pour les données hautes résolutions sont extraits de la base de données METOP (Fig.4). Pour plus de réalisme, un champ de bruit Gaussien est appliqué aux observations. Ce protocole est mis en oeuvre pour obtenir des résultats numériques pour les méthodes proposées et ainsi permettre une comparaison objective de leurs performances.

La Fig.5,a) donne les moyennes et écarts types des erreurs d'interpolation pour chacune des méthodes. Ces résultats montrent que la méthode multi-résolution utilisant les a priori géométriques et temporels (DYNAMIQUE MM) obtient de meilleures performances que les autres méthodes étant donné qu'elle produit des moyennes et écarts types d'erreurs plus faibles. Ensuite, les méthodes qui réalisent les meilleures performances sont dans l'ordre, la méthode dynamique utilisant seulement les observations hautes résolutions (DYNAMIQUE HR), la méthode statique comprenant la contrainte géométrique (STATIQUE MM). Ceci montre que la contrainte dynamique, qui permet de transporter l'information de reconstruction sur des images successives, pourvoit plus d'information que la contrainte géométrique qui permet de transporter l'information de reconstruction le long des structures frontales.

La signification statistique des résultats numériques présentés dans la Fig.5,a) a été mesurée avec un test de signification de Kolmogorov-Smirnov. Ce test a été adopté parce qu'il est adéquat pour des distributions d'erreurs non Gaussiennes. Ce test a montré que la méthode DYNAMIQUE MM est significativement meilleure que les autres méthodes avec un p-value inférieure à 0.001.

La Fig.5,b) donne les distributions cumulées des probabilités d'erreurs produites par chaque méthode. Ces résultats confirment les conclusions tirées de la Fig.5,a). La contrainte de régularité temporelle dérivant du modèle dynamique est la caractéristique la plus importante, suivie par la contrainte géométrique.

Pour mesurer la qualité de reconstruction des données manquantes, spécialement la reconstruction des structures géométriques, nous reportons dans la Fig.7 les résultats d'interpolation obtenues pour chaque méthode pour l'échantillon d'observation de la Fig.6. Ces exemples illustrent le gain de reconstruction lié à la combinaison entre les contraintes dynamique et géométrique. Il est à noter que les structures géométriques sont mieux localisées par les méthodes imposant une contrainte de cohérence géométrique entre les champs SQG multi-résolutions.

### 4.2 Résultats sur des données réelles

Le second ensemble d'expérimentations concerne les données géophysiques réelles. Nous utilisons les séquences d'images de SST brutes REMSS et METOP (voir Fig.1) pour une évaluation des méthodes sur des données réelles



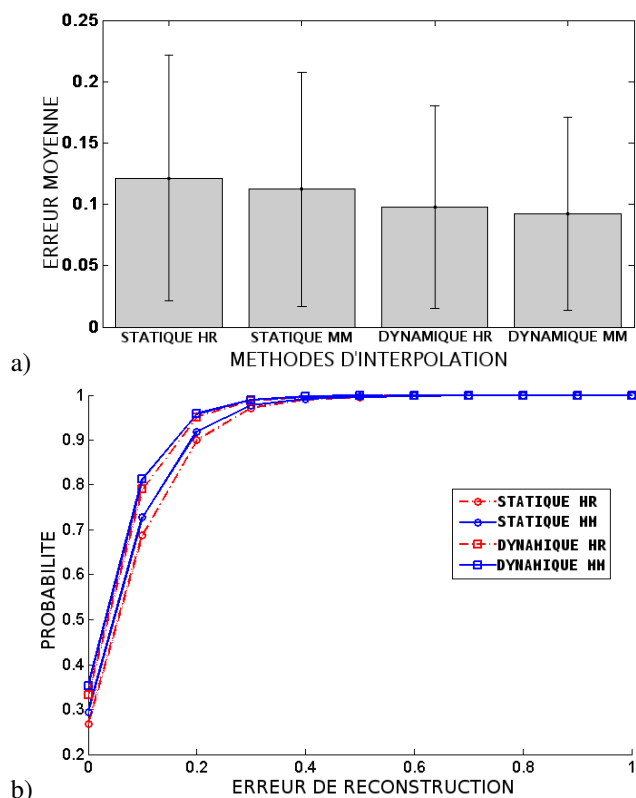


FIG. 5 – Résultats du protocole SQG : distributions cumulées des probabilités d’erreurs pour chaque méthode (en bas). La méthode utilisant le modèle dynamique et l’information multimodale réalise de meilleures performances.

de SST multi-résolution, et des séquences d’images de la base de données REMSS et des séquences d’images de CHL de la base de données MERIS pour une évaluation des méthodes multimodales. Les séquences d’images SST et CHL ont une fréquence d’échantillonnage d’une image par jour. Le but de ces expérimentations et d’évaluer qualitativement les méthodes proposées sur des données réelles. Les Fig.8 et Fig.9 montrent les résultats d’interpolation : la méthode dynamique utilise seulement des données hautes résolutions SST (METOP) ou les images CHL, ou utilise conjointement les observations basses résolutions SST REMSS avec les observations hautes résolution pour exploiter le terme de cohérence géométrique. Ces résultats montrent le bénéfice de la méthode multimodale pour la résolution de structures aux échelles fines (voir Fig.8). Le modèle d’advection-diffusion à lui seul ne suffit pas à reconstruire les structures filamenteuses et les résultats montrent une tendance au lissage de la composante texturée des structures de la chlorophylle. Par opposition, l’interpolation conjointe de la SST basse résolution et des observations CHL permet de reconstruire de manière consistantes les structures filamenteuses. Pour l’estimation de données manquantes pour les observations SST, les résultats reportés dans la Fig.9 ne montrent

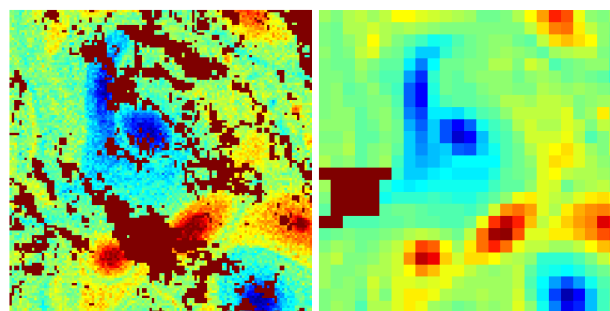


FIG. 6 – Observations SQG : à gauche, image haute résolution contenant 17% de données manquantes ; droite, image basse résolution.

pas les améliorations significatives qui ont été obtenues pour le traitement des observations SQG et CHL. Ceci pourrait être expliqué par un taux plus faible de données manquantes dans les SST hautes résolutions de la base METOP comparée aux observations CHL. En moyenne, les données METOP comprennent 58% de données manquantes, tandis que le taux moyen de données manquantes des données CHL MERIS est de 76%.

## 5 Conclusions

Dans cet article, nous avons proposé une méthode multimodale multi-résolution basée sur l’assimilation variationnelle de données pour l’estimation de données manquantes et de champs de déplacements dans des séquences d’images géophysiques. La difficulté de la tâche est liée à une présence importantes de données manquantes, typiquement entre 20% et 90% pour les données hautes résolutions. Ce fort taux de données manquantes rend difficile la reconstruction des structures fines échelles (filaments, vortex). Pour le modèle proposé, nous avons rapporté de bonnes performances de reconstruction pour des données réelles et synthétiques. Le modèle proposé comprend deux caractéristiques principales :

- une contrainte géométrique reliant la géométrie des images hautes résolutions à celle des images basses résolutions. Etant donné que, du fait de leur mode d’acquisition, les images basses résolutions comportent moins de données manquantes, cette contrainte géométrique est un bon moyen pour contraindre la structure géométrique des images hautes résolutions.
- un cadre d’assimilation variationnel comprenant un modèle dynamique d’advection-diffusion permettant l’estimation conjointe des données manquantes et des champs de déplacements. Ce modèle garanti la cohérence spatio-temporelle des variables reconstruites.

En vision par ordinateur, plusieurs approches variationnelles ont été proposées pour résoudre l’*inpainting* de séquences vidéos [17, 18]. Contrairement à la tâche en vue dans cet article, ces applications d’*inpainting* n’ont considérés que des problèmes avec un taux de données manquantes

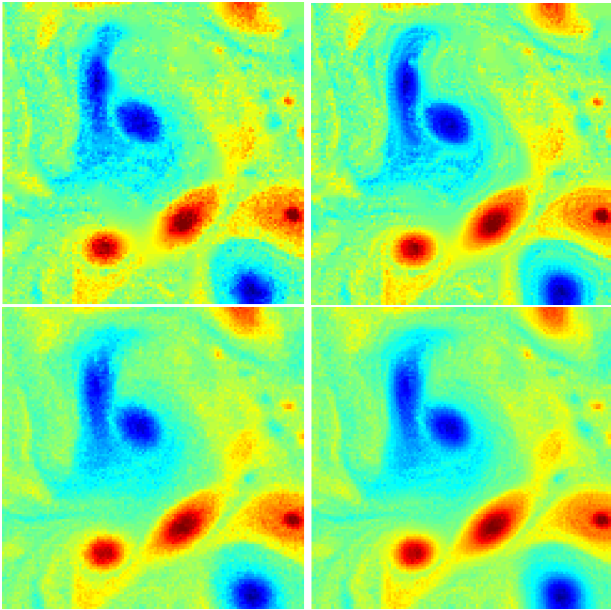


FIG. 7 – Résultats sur les données SQG pour les méthodes STATIQUE HR (haut ;gauche), STATIQUE MM (haut ;droite), DYNAMIQUE HR (bas ;gauche), et DYNAMIQUE MM (bas ; droite). La méthode DYNAMIQUE MM reconstruit des images spatialement plus cohérentes avec des structures géométriques plus fines (voir Fig.6 pour les données brutes).

plus faible. Récemment, d'importantes avancées ont été rapportées pour la super-résolution de séquences vidéo, notamment des méthodes basées sur des approches par exemplaires [20]. Ces méthodes possèdent la capacité de retrouver les structures fines dans des séquences d'images issues de la vidéo-surveillance [20]. Cependant, ils semblent difficile à appliquer à des observations de flot de turbulence. Alors que des contraintes dynamique basées sur des modèles géophysiques, comme celles proposées dans cet article, permettent d'introduire l'information à priori requise pour la super-résolution de ce type de données. Il est à noter que la contrainte géométrique (Eq.7) partage des similitudes avec les termes de diffusion anisotropique utilisé dans l'inpainting pour la restauration d'images [31]. Il impose aux lignes de niveaux supportant les gradients de forte magnitude d'être localement parallèle. Dans la méthode d'assimilation variationnelle proposée, cette diffusion orientée est contrainte par les champs de déplacements relatifs au modèle dynamique. Ceci permet d'obtenir une diffusion anisotropique spatio-temporellement consistente.

Avant nous, d'autres travaux avaient utilisé le cadre de l'assimilation variationnelle pour le traitement de données géophysiques [25, 26, 29]. En particulier, des applications d'estimation de champs de déplacements de mouvement fluide ont été considérées par [25, 26]. Bien que la tâche principale dans cet article soit l'interpolation de données

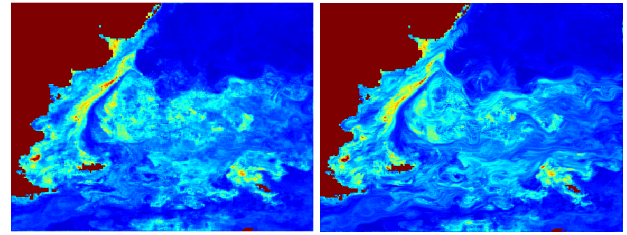


FIG. 8 – Résultats d'interpolation pour les données CHL. La première image donne les résultats pour la méthode DYNAMIQUE HR, la seconde image donne les résultats pour la méthode DYNAMIQUE MM. Les observations brutes correspondantes sont présentées dans la Fig.1.

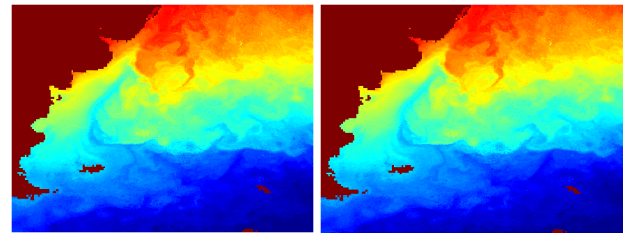


FIG. 9 – Résultats d'interpolation pour les données SST METOP. La première image donne les résultats pour la méthode DYNAMIQUE HR, la seconde image donne les résultats pour la méthode DYNAMIQUE MM. Les observations brutes correspondantes sont présentées dans la Fig.1.

manquantes, l'estimation conjointe des champs de déplacements sous jacent est une caractéristique importante de notre approche. Comparé à [25, 26], notre approche ne comprend pas l'assimilation du champ de déplacement. La dynamique de la surface océanique montre des caractéristiques d'une dynamique de turbulence [30]. Selon nous, le modèle dynamique considéré dans [25, 26] n'est pas approprié à ce type de dynamique. Par conséquent, durant la reconstruction, nous n'avons pas appliqué ce type de modèle pour éviter un lissage excessif des variables géophysiques. De futurs travaux étudieront l'utilisation de modèles d'évolution de turbulence permettant une reconstruction préservant les structures aux échelles fines [7, 9].

## Références

- [1] F.J. Wentz and C.L. Gentemann and D.K. Smith and al, Satellite measurements of sea surface temperature through clouds, *Science*, Vol. 288, 2000.
- [2] P. Le Borgne and G. Legendre and A Marsouin, Operational SST Retrieval from MetOp/AVHRR, *EUMETSAT Conference*, 2007.
- [3] P. Camberlin and S. Janicot and I. Pocard, Seasonality and atmospheric dynamics of the teleconnection between African rainfall and the tropical sea surface



- temperature : Atlantic vs. ENSO, *International Journal of Climatology*, Vol. 21, 2001.
- [4] H. Giordani and G. Caniaux, Sensitivity of Cyclogenesis to Sea Surface Temperature in the Northwestern Atlantic, *Monthly Weather Review*, Vol. 129(6), 2001.
- [5] M. A. Saunders and A. S. Lea, Large contribution of sea surface warming to recent increase in Atlantic hurricane activity, *Nature*, Vol. 451(7178), 2008.
- [6] D. S. Ullman and P. C. Cornillon, Evaluation of front detection methods for satellite-derived SST data using in situ observations, *Journal of Atmospheric and Oceanic Technology*, 2000.
- [7] J. Sukhtame and R. Pierrehumbert, Surface quasigeostrophic turbulence : The study of an active scalar, *Chaos*, 2002.
- [8] X. Zheng and H. Wei, Analysis of Chlorophyll Concentration during the Phytoplankton Spring Bloom in the Yellow Sea Based on the MODIS Data, *Life System Modeling and Intelligent Computing*, 2010.
- [9] J. Isern-Fontanet and B. Chapron and G. Lapyere and P. Klein, Potential use of microwave sea surface temperatures for the estimation of ocean currents, *Geophysical Research Letter*, 2006.
- [10] I. M. Belkina and J.E O'Reilly, An algorithm for oceanic front detection in chlorophyll and SST satellite imagery, *Journal of Marine Systems*, 2009.
- [11] H. U. Solanki and R. M. Dwivedi and S.R. Nayak and V. S. Somvanshi and D. K. Gulati and S. K. Pattnayak, Fishery forecast using OCM chlorophyll concentration and AVHRR SST : validation results off Gujarat coast, India, *International Journal of Remote Sensing*, 2003.
- [12] S. O. Ba and T. Corpetti and B. Chapron and R. Fablet, Variational data assimilation for missing data interpolation in SST images, *EEE IGARSS*, 2010.
- [13] J. L. Hoyer and J. Shea, Optimal interpolation of sea surface temperature for the North Sea and Baltic Sea, *Journal of Marine Systems*, Vol. 65(1-4), 2007.
- [14] R.W. Reynolds and T. M. Smith, Improved global sea surface temperature analyses using optimum interpolation, *Journal of Climate*, Vol. 7, 1994.
- [15] L. Bertino and G. Evensen and H. Wackernagel, Sequential Data Assimilation Techniques in Oceanography, *International Statistical Review*, 2003.
- [16] D. Youzhuan and F. Dongyang W. Zhihui and M. Zhihua and Z. Juhong, Reconstruction of Incomplete Satellite Oceanographic Data Sets Based on EOF and Kriging Methods, *Image and Signal Processing for Remote Sensing*, 2008.
- [17] J.-F. Aujol and S. Ladjja and S. Masnou, Exemplar-based inpainting from a variational point of view, *SIAM Journal on Mathematical Analysis*, 2010.
- [18] P. Arias and V. Casselles and G. Sapiro, A Variational Framework for Non-local Image Inpainting, *Int. Conf. on Energy Minimization Methods in CVPR*, 2009.
- [19] K. A. Patwardhan and G. Sapiro and M. Bertalmio, Video Inpainting Under Constrained Camera Motion, *Science*, 2010.
- [20] V. Cheung and B. J. Frey and N. Jojic, Video epitomes, *Int. Journal on Computer Vision*, 2007.
- [21] S. Babacan and R. Molina and Katsaggelos, Variational Bayesian Super Resolution, *IEEE transaction on Image Processing*, 2009.
- [22] D. Glasner and S. Bagon and M. Irani, Super-Resolution from a Single Image, *Int. Conference on Computer Vision*, 2009.
- [23] N. Verbrugge and G. Reverdin, Contribution of horizontal advection to the interannual variability of sea surface temperature in the North Atlantic, *Journal of Physical Oceanography*, 2003.
- [24] A. Bennett and M. Thorburn, The generalized inverse of a nonlinear quasigeostrophic ocean circulation model, *Journal of Physical and Oceanogr.*, Vol. 22, 1990.
- [25] N. Papadakis and T. Corpetti and E. Memin, Dynamically consistent optical flow estimation, *Inter. Conference on Computer Vision*, 2007.
- [26] N. Papadakis and P. Heas and E. Memin, Image assimilation for motion estimation of atmospheric layers with shallow-water model, *Asian Conference on Computer Vision*, 2007.
- [27] O. Talagrand and P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation, *Quarterly Journal of the Royal Meteorological Society*, Vol 113(478), 1987.
- [28] G. Gawarkiewicz and D.C. Chapman, Formation and maintenance of shelfbreak fronts in an unstratified flow, *Journal of physical oceanography*, 1991.
- [29] I. Gejadze and M. Honnorat and F.-X. Le Dimet and J. Monnier, On variational data assimilation for 1D and 2D fluvial hydraulics, *European Conference on Mathematics for Industry*, 2006.
- [30] G. Lapeyre and P. Klein and B. L. Hua, Oceanic restratification forced by surface frontogenesis, *Journal of Physical Oceanography*, Vol. 36, 2006.
- [31] D. Tschumperle and R. Deriche, Vector-Valued Image Regularization with PDE's : A Common Framework for Different Applications, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.