

**DEMANDE D'ALLOCATION DE RECHERCHE DE L'ED
SISEO
Année universitaire 2017-2018
SUJET DE THESE**

1. LABORATOIRE <i>Nom ou sigle :LISTIC Statut :EA3703</i>	2. DIRECTION DE THÈSE <i>Directeur de thèse (HDR) :Philippe Bolon Codirecteur éventuel :Yajing Yan</i>
Domaine de compétences de l'ED SISEO : - Environnement X - Organisation - Systèmes	Collaborations éventuelles : -
3. SUJET DE THÈSE Titre : Analyse et reconstruction de données manquantes dans des séries temporelles d'images satellitaires Sentinel-1 (Applications aux mesures de déplacement du glacier d'Argentière)	

4. RESUME

(Français et Anglais)

En français :

Malgré la masse de données (satellitaires et in-situ) disponibles pour diverses applications d'observation de la Terre, l'incomplétude de données reste toujours un problème fréquemment rencontré pour suivre un phénomène. Le développement de méthodes permettant de reconstruire les données manquantes à partir des données observées constitue alors un enjeu majeur en télédétection. Dans cette thèse, nous proposons d'étudier trois méthodes de type analyse objective, en s'appuyant sur des formulations mathématiques 1) l'analyse variationnelle 2) la décomposition en « Empirical Orthogonal Functions » 3) l'interpolation optimale pour des problèmes de reconstruction de données manquantes dans des séries temporelles d'images satellitaires Sentinel-1. Les avantages et les inconvénients de chacune des méthodes vis à vis de la caractéristique spatio-temporelle du signal à reconstruire et de la connaissance a priori du mécanisme physique sous-jacent, seront mis en évidence à travers des jeux de données synthétiques. Dans un second temps, ces méthodes seront appliquées à la mesure de déplacement du glacier d'Argentière, ce qui constitue un véritable défi à cause de la superficie importante de la zone de données manquantes, de la complexité du signal de déplacement à restituer, ainsi que de la connaissance insuffisante du processus physique sous-jacent. Enfin, la modification de certaines méthodes ou la proposition de nouvelles méthodes qui s'adaptent mieux au contexte d'utilisation sera envisagée.

En Anglais:

Despite the mass of data (satellite and in-situ) available for various Earth observation applications, data incompleteness is still a frequently encountered problem in tracking a phenomenon. The development of methods to reconstruct the missing data from observed data is a major challenge in remote sensing. In this thesis, we propose to study three different methods in the family of objective analysis based on mathematical formulations 1) variational analysis 2) decomposition in 'Empirical Orthogonal Functions' 3) optimal interpolation in problems of reconstruction of missing data in time series of Sentinel-1 satellite images. The advantages and disadvantages of each method in terms of spatio-temporal characteristic of the signal to be reconstructed and a priori knowledge of the underlying physical mechanism, will be demonstrated by means of synthetic data sets. In a second step, these methods will be applied to the displacement measurement of the Argentière glacier, which is a real challenge because of the large area of missing data, the complexity of the displacement signal to be restored and insufficient knowledge about the underlying physical process. Finally, the modification of certain methods or the proposal of new methods that better adapt to the context of use will be considered.

5. PROJET DE RECHERCHE DETAILLE

Suite aux lancements successifs de satellites depuis les années 90, les images satellitaires deviennent un outil puissant et prédominant pour l'observation de la Terre, grâce à leur disponibilité régulière et leur grande couverture spatiale. En particulier, les images des satellites radar Sentinel-1 qui couvrent l'Europe tous les 6 jours (tous les 12 jours ailleurs) et sont mises à disposition gratuitement par l'ESA (European Space Agency), fournissent aux scientifiques une grande opportunité pour diverses applications d'observation de la Terre. Malgré la masse de données disponible pour suivre un phénomène, des trous d'observations existent toujours, soit à cause de la disponibilité/qualité de données, soit à cause du traitement qui exploite les données. L'incomplétude de données est donc un problème fréquemment rencontré. En imagerie SAR, les données peuvent ne pas être exploitables dans les zones de faible cohérence, dans les zones d'ombre ou de foldover, etc. De manière générale, les données nécessitent une interprétation quantitative qui permet une caractérisation et une compréhension des mécanismes internes du processus observé. En conséquence, la complétude de données est cruciale pour toutes les recherches et les applications qui les exploitent. Par ailleurs, des données in situ, souvent spatialement dispersées, sont aussi utilisées en complément des données satellitaires. L'interpolation de ces données porte aussi de grand intérêt pour avoir une vision globale du phénomène observé. Pour toutes ces raisons, le développement de méthodes permettant de reconstruire les données manquantes à partir des données observées constitue un enjeu majeur en télédétection.

L'analyse objective (William and Thomso, 2014) a pour objectif d'utiliser des formulations mathématiques pour en déduire les valeurs à des positions non observées en s'appuyant sur les valeurs à des positions observées. En résumé, 3 catégories de méthodes existent :

1. « Fonction d'ajustement au modèle » : dans cette démarche, les observations sont ajustées à un ensemble d'équations qui décrivent le processus observé. Les méthodes du type moindre carré sont souvent utilisées.
2. « Méthode de correction successive » : il s'agit d'une démarche itérative. On commence par une première estimation, puis on utilise un schéma de pondération linéaire pour ajuster l'estimation initiale et les estimations suivantes en tenant compte des observations à une tolérance spécifiée.
3. « Méthode d'interpolation statistique » (aussi appelée Interpolation optimale) : dans cette approche, la moyenne de l'ensemble de différences entre les données restituées et les observations sont minimisées au sens des moindres carrés.

Ces méthodes ont été largement utilisées dans différentes applications en atmosphère-océanographie. Dans le cas de la mesure du déplacement par imagerie SAR, l'interpolateur Kriage peut être considéré comme l'outil le plus utilisé. Cependant, cette méthode présente des inconvénients et n'est pas toujours opérationnelle à cause de la complexité de la caractéristique du bruit présent dans les signaux de déplacement. Son application est fortement conditionnée par la quantité et la distribution des observations disponibles. De plus, elle ne permet pas de profiter au maximum de l'information temporelle dans le cas d'une série temporelle. L'exploitation des méthodes de type analyse objective semble intéressante et prometteuse, d'une part, pour enrichir la base

de méthodes d'interpolation dans la communauté de mesure de déplacement par imagerie SAR ; d'autre part, pour faire évoluer ces méthodes en proposant des améliorations et/ou en développant de nouvelles variantes dans un nouveau contexte d'utilisation.

Dans le cadre de cette thèse, nous proposons, dans un premier temps, de mettre en place 3 méthodes d'interpolation ou d'approximation de type analyse objective : 1) l'analyse variationnelle (Brasseur et al., 1996) basée sur la définition d'une fonction de coût qui minimise l'écart entre les observations et un ensemble d'équations sous contrainte de régularisation 2) la décomposition EOF (Empirical Orthogonal Functions) (Beckers et al., 2003) qui décompose une série temporelle en modes d'erreur représentant les principales variabilités spatio-temporelles. 3) l'interpolation optimale (Gandin et al., 1965, Brankart and Brasseur, 1996) dans laquelle l'écart entre les données observées et les données restituées est minimisé par la méthode des moindres carrés en s'appuyant sur l'estimation ou la modélisation de la covariance d'erreur.

Des analyses fines sur les 3 familles de méthodes seront abordées afin de mettre en évidence les avantages et les inconvénients de chacune des méthodes. L'applicabilité et la difficulté de chacune des méthodes en fonction du contexte d'utilisation, plus précisément, en fonction de la caractéristique spatio-temporelle du champ de déplacement à reconstruire et de la connaissance sur le mécanisme physique du phénomène observé seront soulignées à travers des simulations synthétiques. Pour ce faire, la performance et la robustesse de chacune des méthodes seront évaluées dans différents cas où l'imperfection plus ou moins importante est présente dans les données, par exemple, le rapport signal sur bruit est faible, la caractéristique spatio-temporelle du signal est compliquée à cause de la présence de multiples processus à différentes échelles spatio-temporelles, la quantité de données observées est insuffisante, la distribution de données est hétérogène, etc. De plus, l'insertion de manière intelligente de l'information issue d'un modèle physique plus ou moins fiable dans les méthodes d'interpolation sera étudiée afin de profiter au mieux du modèle physique en prenant en compte l'incertitude présente dans le modèle.

Dans un second temps, ces méthodes seront appliquées aux mesures de déplacement du glacier d'Argentière issues d'une série d'images Sentinel-1, ce qui constitue un véritable défi en raison de la grande superficie de la zone de données manquantes (pour certaines zones, aucune technique ne parvient à mesurer les déplacements) et de la complexité du signal de déplacement à cause de la connaissance insuffisante du processus physique sous-jacent et du bruit important présent dans les mesures de déplacement (Jauvin, 2018, Dehecq, 2015, Fallourd, 2012). Pour ceci, la technique de la corrélation d'amplitude sera appliquée à une série temporelle d'images Sentinel-1 pour obtenir une série de mesure de déplacement. Les incertitudes présentes dans les mesures de déplacement seront également estimées et analysées. Ensuite, les 3 méthodes d'interpolation seront utilisées pour reconstruire les données manquantes dans la série temporelle de mesure de déplacement. D'autres types de mesures telles que les mesures issues de l'interférométrie différentielle et les mesures GPS, lorsqu'elles sont disponibles, seront utilisées pour valider les résultats. Dans les zones où aucune mesure d'autre type n'est disponible pour la validation, la technique de la validation croisée (Craven, 1979), qui consiste à enlever les valeurs à des positions observées choisies de manière intelligente,

puis à comparer les valeurs observées et les valeurs restituées à ces positions, sera utilisée afin d'évaluer les résultats de manière la plus objective possible.

En plus de cette application, la proposition des variantes de méthode d'interpolation (modification de certaines méthodes ou combinaison de plusieurs méthodes) qui s'adaptent au mieux au contexte applicatif sera envisagée. Par ailleurs, d'autres approches basées sur l'usage des réseaux de neurones profonds (Deep Learning) qui illustre suffisamment la variabilité des données du contexte d'étude à partir des entraînements sur des bases d'apprentissage de taille conséquente (Pathak et al., 2016), pourront également être envisagées.

Références :

Beckers, J.-M. and M. Rixen, 2003: EOF calculation and data filling from incomplete oceanographic datasets. *Journal of Atmospheric and Oceanic Technology*, 20, 1839-1856

Brankart, J.-M. and P. Brasseur, 1996: Optimal analysis of in situ data in the Western Mediterranean using statistics and cross-validation. *Journal of Atmospheric and Oceanic Technology*, 13, 477-491

Brasseur, P., J.-M. Beckers, J.-M. Brankart, and R. Schoenauen, 1996: Seasonal temperature and salinity fields in the Mediterranean Sea: Climatological analyses of an historical data set. *Deep-Sea Research*, 43, 159-192.

Craven, P. and G. Wahba, 1979: Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31, 377-403.

Dehecq A., 2015, Analyse de la dynamique des glaciers Himalayens et Alpains à partir de 40 ans de données d'observation de la Terre. Thèse de doctorat. Université de Grenoble - Alpes.

Emery W. J. and R.E. Thomson, 2014, *Data analysis methods in physical oceanography*. Third Edition, Elsevier.

Fallourd R., 2012, Suivi des glaciers alpins par combinaison d'informations hétérogènes: images SAR haute résolution et mesures terrain. Thèse de doctorat. Université de Grenoble.

Gandin, L. S., 1965: Objective analysis of meteorological fields. Israel Program for Scientific Translation, Jerusalem, 242 pp

Jauvin M. 2018, Mesure des déformations de surface par imagerie radar satellitaire : Application à la surveillance des territoires de montagne et de l'impact de grands chantiers. Thèse en cours.

Pathak J.D., Krähenbühl P., Donahue J., Darrell T. and Efros A., 2016, Context Encoders: Feature Learning by Inpainting, *Computer Vision and Pattern Recognition (CVPR)*.

6. CANDIDAT RECHERCHE :

Le candidat devra avoir des bonnes connaissances et des compétences solides en mathématique et en statistique. La compétence en programmation (python, C ou matlab) est également nécessaire. Des connaissances en imagerie SAR sont appréciées.

7. FINANCEMENT DE LA THESE : *Le contrat doctoral fixe une rémunération minimale, indexée sur l'évolution des rémunérations de la fonction publique : depuis le 1er février 2017, elle s'élève à **1768,55 euros** bruts mensuels pour une activité de recherche seule. Un avenant attributif d'une mission complémentaire d'enseignement est possible pour une durée de 2 ans. Sous réserve de la publication de l'arrêté fixant le taux de rémunération des heures complémentaires, la rémunération mensuelle sera de 220, 80 euros bruts pour 64 heures ETD par année universitaire.*

8. CONTACT :

Nom prénom : YAN Yajing

Tél : 0450096536

Email : yajing.yan@univ-smb.fr