

# Analyse de séries temporelles massives d'images satellitaires

Applications à la cartographie des écosystèmes

Mathieu Fauvel\*

Stéphane Girard<sup>◊</sup>

Serge Iovleff<sup>◊</sup>

\* Dynafor, University of Toulouse, INRA, INPT, INPT-EI PURPAN, 31326 Castanet Tolosan, France

<sup>◊</sup> Team Mistis, Inria Grenoble Rhône-Alpes, LJK, 38334 Montbonnot, France

<sup>◊</sup> Team Modal, Inria Lille Nord Europe, LPP UMR-8524, Université Lille, 59655 Villeneuve d'Ascq, France

## 1 Contexte

Dans un contexte de changement climatique, la surveillance de l'état et de le contrôle l'évolution des territoires sont des enjeux majeurs pour les années à venir. Ces informations sont à la base des politiques d'aménagement du territoire, de l'échelle régionale aux échelles nationale et internationale et de la mise en place de programmes de lutte contre le réchauffement climatique [PFW<sup>+</sup>13]. De récents accords multilatéraux rendent obligatoires notamment la cartographie des services écosystémiques <sup>1</sup> à l'échelle nationale [Nat15]. Cependant, il existe actuellement trop peu de techniques de surveillance des écosystèmes qui puissent être utilisées et comparées à l'échelle mondiale [KTS<sup>+</sup>15].

La télédétection satellitaire permet d'observer les paysages sur des échelles spatiales importantes, avec une fréquence de revisite régulière et pour un coût financier modéré (en particulier avec les politiques de libre accès aux données mises en place par l'Union Européenne). Dans le cadre du programme Européen *Copernicus*, les constellations *Sentinel-1* et *Sentinel-2* fournissent actuellement une couverture complète du territoire national tous les 5 jours, à une résolution spatiale décamétrique et cela dans plusieurs domaines du spectre électromagnétique. Il est ainsi possible de suivre les évolutions intra- et inter-annuelle des écosystèmes à une résolution spatiale fine à l'échelle du territoire.

Ce flux massif de données d'observation de la terre fournit une description riche et détaillée des écosystèmes, permettant le contrôle de leur état et de leur évolution. Cependant, l'augmentation conjointe du nombre de pixels à traiter et du nombre de variables spectro-temporelles associées à chaque pixel rend complexe le traitement de ces données. En particulier, les méthodes conventionnelles d'apprentissage définies pour des séries d'images à des résolutions spatiales et temporelles moyennes ne sont plus adaptées. On citera en exemple les temps de calculs qui explosent pour les méthodes à noyaux (de type SVM) rendant inexploitable ces méthodes pour des volumes de données importants.

Une autre problématique importante concerne l'échantillonnage temporel irrégulier des pixels à traiter. En raison de l'orbite des satellites *Sentinel-1&2*, couplée à la présence de nuages, les pixels observés n'ont pas les mêmes instants d'acquisition. Cette hétérogénéité d'échantillonnage requiert un pré-traitement important lorsque des méthodes conventionnelles d'analyses sont utilisées [IAT<sup>+</sup>15]. En effet, pour être opérantes ces méthodes sont contraintes de représenter un pixel par un vecteur de taille constante, ou chaque composante correspond à une acquisition

---

1. Services écosystémiques : services rendus par la nature à l'humanité [MEA05].

spectro-temporelle. Il est alors nécessaire de filtrer et de recalculer temporellement tous les pixels à traiter, *e.g.*, environ 6 500 milliards de pixels pour la France.

De manière plus générale, la nature temporelle des données a été peu abordée en télédétection pour la cartographie des écosystèmes [GWW16]. En dehors de l'étape de recalage temporel, les approches conventionnelles ne prennent pas en compte la nature ordonnée des acquisitions temporelles. Par exemple, la distance euclidienne entre vecteurs est classiquement utilisée pour comparer deux pixels entre eux. Or cette distance est indépendante de l'ordre des variables. Par conséquent une partie de l'information contenue dans les séries temporelles d'images n'est pas utilisée dans les traitements.

Pour exploiter complètement l'information fournie par les séries temporelles d'images "nouvelle génération", des développements méthodologiques sont donc nécessaires. Cela concerne à la fois la gestion de grands volumes de données et la prise en compte de la nature temporelle des séries.

## 2 Objectifs de la thèse

Cette thèse en co-encadrement entre l'UMR DYNAFOR, l'équipe Mistis à Inria et l'équipe Probabilités et Statistique du laboratoire Paul Painlevé (LPP, UMR-8524) s'intéresse à la *statistique fonctionnelle* pour la cartographie des écosystèmes [RS05]. Pour cela, nous allons développer des méthodes d'analyses permettant à la fois de s'affranchir de l'étape de recalage temporel et de manipuler de grands volumes de données.

Le point de départ de ce travail est la définition d'une méthode d'analyse permettant de classer un pixel quels que soient les instants d'échantillonnage temporel. Nous allons considérer que les échantillons, conditionnellement aux classes, sont distribués selon des *processus Gaussiens*. De plus, nous relâcherons l'hypothèse d'échantillonnage temporel irrégulier en supposant simplement que chaque pixel  $x_i$  est observé  $T_i$  fois durant l'année (ou plus généralement durant la période d'analyse),  $T_i$  dépendant donc de la localisation géographique, de l'orbite du satellite, des nuages et ombres observés pour  $x_i$ .

Le cœur des travaux de thèse s'attachera à spécifier les méthodes et les outils pour apprendre les paramètres des *processus Gaussiens*. En particulier, lors de la création du processus Gaussien, des bases fonctionnelles temporelles vont être définies. Elle devront prendre en compte à la fois l'évolution temporelle (corrélation forte des mesures entre deux instants proches et faible lorsque les instants sont éloignés) et la réponse spectrale à un instant donné (corrélation inter-bandes). Cette étape est importante car l'utilisation d'une base adaptée permet à la fois de s'affranchir du bruit et des données manquantes tout en capturant l'information utile à l'analyse. Précisons toutefois qu'il n'est pas envisagé ici de lisser/projeter les pixels sur cette base, mais d'exprimer la fonction de décision à partir de cette représentation spectro-temporelle de manière à pouvoir l'appliquer sur n'importe quel échantillonnage temporel.

Pour traiter les flux massifs d'images, des algorithmes d'optimisation efficaces utilisant des approches HPC (*high performance computing*) devront être définis. A cet effet, nous allons nous appuyer sur certaines propriétés intéressantes de l'algorithme d'apprentissage des paramètres des *processus Gaussiens* :

1. *Modèles génératifs* : Cela va permettre de traiter les classes indépendamment les unes des autres, l'apprentissage des paramètres pouvant se faire dans des tâches séparées, sur des unités de calculs différentes sans partage de mémoire.
2. *Estimateurs explicites* : Les estimateurs ne nécessitent pas d'étapes d'optimisation complexes (*e.g.*, gradient sous contraintes). De plus, ces estimateurs supportent des modes "on-line", c'est à dire qu'il est possible de répartir les estimations sur différents paquets de données et de mettre à jour l'estimateur final en procédant à la mise en commun. L'approche envisagée

serait basée sur la norme MPI (*Message Passing Interface*), la taille des informations à mettre en commun étant limitée (quelques Mo).

3. La plupart des calculs pour l'apprentissage et pour la prédiction sont basés sur de l'algèbre linéaire (produit matriciel, déterminant, décomposition en valeurs propres/vecteurs propres ...) qui peuvent grandement tirer partie des techniques de parallélisation sur des unités de calculs, soit sur processeurs de type CPU soit de type GPU.

Les développements informatiques seront réalisés à l'aide de l'interface de programmation *Orfeo Toolbox*. Le prototypage pour la mise en œuvre sur une grappe de serveurs sera validée à l'aide des moyens de calculs distribués offert par l'INRA et par Inria. Les codes sources seront distribués sous une licence *copyleft* pour permettre leur distribution et utilisation par la communauté scientifique.

Dans un premier temps, seulement la classification supervisée sera considérée. Une fois les différents objectifs pré-cités remplis, les modèles construits serviront de support pour la définition d'une version non-supervisée de l'algorithme, et à plus long terme, d'une version pour l'estimation de variables (indices de biodiversité ...).

### 3 Présentation de l'équipe encadrante

Ce projet requiert des compétences en télédétection, en mathématiques appliquées et en développement informatique. Les équipes partenaires travaillent dans ces domaines depuis de nombreuses années et sont reconnues (inter-)nationalement comme expertes dans le domaine de la télédétection et de l'analyse fonctionnelle. Les équipes ont déjà collaboré ensemble dans le cadre du projet *CloHé* "*Classification de données hétérogènes avec valeurs manquantes appliquée au traitement des données satellitaires en écologie et cartographie du paysage*" financé par le défi GRANDES MASSES DE DONNÉES SCIENTIFIQUES CNRS MASTODONS. DYNAFOR et l'équipe Mistis ont co-encadré la thèse de Maïlys Lopes intitulée "*Ecological monitoring of semi-natural grasslands : statistical analysis of dense satellite image time series with high spatial resolution*". Serge Iovleff (Modal) et Stéphane Girard (Mistis) sont également impliqués dans l'équipe associée SIMERGE du LIRIMA (Laboratoire International de Recherche en Informatique et Mathématiques Appliquées) via des recherches en classification pour la haute dimension.

Dans le cadre du projet *CloHé*, les équipes ont produit un démonstrateur en R et C++ (disponible à l'adresse suivante [https://modal.lille.inria.fr/CloHe/packages/CloHe\\_2.0.0.tar.gz](https://modal.lille.inria.fr/CloHe/packages/CloHe_2.0.0.tar.gz)) montrant la faisabilité théorique de l'approche proposée dans ce sujet de thèse.

#### 3.1 Mathieu Fauvel

Depuis Septembre 2011, Mathieu Fauvel est Maître de Conférences à l'INP Toulouse et associé à l'UMR 1201 INRA Dynafor. Ses domaines de recherches sont le traitement des images et des signaux à grandes dimensions, l'apprentissage automatique, les méthodes à noyaux ainsi que la fusion de données. Il est impliqué dans plusieurs projets à l'interface entre télédétection et l'écologie du paysage. Pour le projet MUESLI (*MULTiscale mapping of Ecosystem Services by very high spatial resolution hyperspectral and LiDAR remote sensing Imagery*), il a reçu le prix INP INNOV dans la catégorie *perspectives scientifiques*. En 2013, il a reçu le prix du meilleur papier de la revue IEEE TRANS. GEOSCIENCE AND REMOTE SENSING. En 2012, il a reçu le prix du meilleur reviewer pour la même revue.

Actuellement, Mathieu Fauvel est impliqué dans plusieurs projets mêlant l'analyse de données massives d'observation de la terre et problématique environnementale : un projet TOSCA OSO (*Occupations de SOIs*) et un projet MASTODONS (la qualité des données dans le BigData) "*Classification de données hétérogènes avec valeurs manquantes appliquée au traitement des données*

satellites en écologie et cartographie du paysage" (CloHé).

De 2013 à 2016, il a été président du Chapitre France de la IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY. Le Chapitre a reçu le prix du meilleur Chapitre en 2016 de la part de la IEEE GRSS et de la Section France IEEE. Depuis 2016, il est en charge de la coordination des activités des Chapitres Européens de la GRSS.

Il a été un membre du comité scientifique du IEEE WORKSHOP ON THE ANALYSIS OF MULTITEMPORAL REMOTE SENSING IMAGES, 2015, Annecy, France et du IEEE WORKSHOP ON HYPERSPECTRAL IMAGE AND SIGNAL PROCESSING : EVOLUTIONS IN REMOTE SENSING, 2009, Grenoble. Il a été *Editeur Invité* de deux numéros spéciaux de IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (*Information extraction from high-spatial-resolution optical remotely sensed imagery* et *Analysis of multitemporal data and applications*). Depuis 2016, il est *Editeur Associé* de la revue IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING pour les domaines *hyperspectral remote sensing* et *time series data*.

Mathieu Fauvel a co-supervisé ou supervisé 5 Thèses et 10 Masters. Il est co-auteur de plus de 30 articles dans des revues internationales, 3 chapitres de livres et plus de 40 papiers dans des conférences internationales et nationales.

### 3.2 Stéphane Girard

Stéphane Girard est directeur de recherche à Inria Grenoble Rhône-Alpes et également membre du Laboratoire Jean Kuntzmann - LJK (UMR 5224 du CNRS). De 2012 à 2016, il a été responsable du département Probabilités-Statistique du LJK. Son domaine de recherches est la statistique, il y développe trois thèmes : l'apprentissage statistique en grande dimension, l'estimation fonctionnelle et la statistique des valeurs extrêmes. Il est actuellement impliqué dans deux projets promouvant les interactions entre statistique et traitement informatique des grands volumes de données : un projet TelluS-Insmi (promotion des thématiques aux interfaces entre Insmi et Insu) "Classification non-supervisée en grande dimension" et un projet Mastodons (la qualité des données dans le BigData) "Classification de données hétérogènes avec valeurs manquantes appliquée au traitement des données satellitaires en écologie et cartographie du paysage".

Stéphane Girard est éditeur associé des revues JOURNAL OF MULTIVARIATE ANALYSIS (depuis 2016), *Statistics and Computing* (depuis 2012) et membre du Advisory Board de *Dependence Modelling* (depuis 2015). Il a également été co-éditeur de *Statistics for astrophysics, clustering and classification*, vol. 77, EDP sciences, 2016, et éditeur invité des numéros 152(3) et 154(2) du JOURNAL DE LA SOCIÉTÉ FRANÇAISE DE STATISTIQUE. Il a participé à l'organisation des trois premières éditions des écoles d'astrostatistique STAT4ASTRO visant à promouvoir l'usage des méthodes statistiques récentes par les astrophysiciens. Il a en outre été membre du comité d'organisation ou du comité scientifique de 17 conférences.

Stéphane Girard a encadré ou co-encadré 15 thèses. Il est co-auteur de plus de 80 articles dans des revues internationales, 14 chapitres de livres et plus de 150 communications dans des conférences internationales et nationales.

### 3.3 Serge Iovleff

Serge Iovleff est maître de conférence hors classe à l'IUT "A" -Département Informatique de l'université de Lille 1, membre de l'équipe Probabilités et Statistiques du LPP (UMR 8524 du CNRS) et membre de l'équipe Inria Modal (Model for Data Analysis and Learning). Son domaine de recherche est l'analyse de donnée avec un focus sur la réduction de dimension ainsi que la classification non-supervisée. Il est responsable du défi CNRS Mastodons "Classification de données hétérogènes avec valeurs manquantes appliquée au traitement des données satellitaires en écologie et cartographie du paysage" (CloHé) depuis février 2016. Il a co-organisé à Lille

la conférence "Learning with functional data" du 7 octobre 2016 à Lille (<https://functional-data.univ-lille1.fr>).

Il est très actif dans le domaine du logiciel libre, responsable de la librairie de calcul orientée vers les statistiques STK++ (<http://stkpp.org>) et mainteneur des packages R : MixAll (<https://CRAN.R-project.org/package=MixAll>) dédié à la classification non-supervisée des données hétérogènes et blockcluster (<https://CRAN.R-project.org/package=blockcluster>) dédié au co-clustering des tableaux de données.

## 4 Candidature

Pour candidater, prendre contact avec les co-encadrants par courriel (joindre un CV détaillé) :

— [mathieu.fauvel@ensat.fr](mailto:mathieu.fauvel@ensat.fr)

— [stephane.girard@inria.fr](mailto:stephane.girard@inria.fr)

— [serge.iovleff@inria.fr](mailto:serge.iovleff@inria.fr)

Les candidatures sont ouvertes jusqu'à mi-Mars 2018.

## Références

- [GWW16] Cristina Gómez, Joanne C. White, and Michael A. Wulder. Optical remotely sensed time series data for land cover classification : A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116 :55 – 72, 2016.
- [IAT<sup>+</sup>15] Jordi Inglada, Marcela Arias, Benjamin Tardy, Olivier Hagolle, Silvia Valero, David Morin, Gérard Dedieu, Guadalupe Sepulcre, Sophie Bontemps, Pierre Defourny, and Benjamin Koetz. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing*, 7(9) :12356–12379, 2015.
- [KTS<sup>+</sup>15] Daniel S. Karp, Heather Tallis, René Sachse, Ben Halpern, Kirsten Thonicke, Wolfgang Cramer, Harold Mooney, Stephen Polasky, Britta Tietjen, Katharina Waha, Ariane Walz, and Stacie Wolny. National indicators for observing ecosystem service change. *Global Environmental Change*, 35 :12 – 21, 2015.
- [MEA05] Millennium Ecosystem Assessment. <http://www.millenniumassessment.org/en/Multiscale.html>, 2005.
- [MTCA17] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2) :645–657, Feb 2017.
- [Nat15] United Nations. Transforming our world : The 2030 agenda for sustainable development. Technical report, United Nations, 2015. <https://sustainabledevelopment.un.org/post2015/transformingourworld/publication>.
- [PFW<sup>+</sup>13] H. M. Pereira, S. Ferrier, M. Walters, G. N. Geller, R. H. G. Jongman, R. J. Scholes, M. W. Bruford, N. Brummitt, S. H. M. Butchart, A. C. Cardoso, N. C. Coops, E. Dullo, D. P. Faith, J. Freyhof, R. D. Gregory, C. Heip, R. Höft, G. Hurtt, W. Jetz, D. S. Karp, M. A. McGeoch, D. Obura, Y. Onoda, N. Pettoelli, B. Reyers, R. Sayre, J. P. W. Scharlemann, S. N. Stuart, E. Turak, M. Walpole, and M. Wegmann. Essential biodiversity variables. *Science*, 339(6117) :277–278, 2013.
- [PIG12] F. Petitjean, J. Inglada, and P. Gancarski. Satellite image time series analysis under time warping. *IEEE Transactions on Geoscience and Remote Sensing*, 50(8) :3081–3095, Aug 2012.
- [PLSM17] T. Postadjian, A. Le Bris, H. Sahbi, and C. Mallet. Investigating the Potential of Deep Neural Networks for Large-Scale Classification of Very High Resolution Satellite Images. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 183–190, May 2017.
- [RS05] J. O Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, New York, 2005.