



**HAL**  
open science

# Time series analysis of medium spatial resolution sensing images: LAI reconstruction, unmixing: application to vegetation monitoring on MODIS data

Xing Gong

► **To cite this version:**

Xing Gong. Time series analysis of medium spatial resolution sensing images: LAI reconstruction, unmixing: application to vegetation monitoring on MODIS data. Signal and Image Processing. Université Rennes 2; Université de Pékin, 2015. English. NNT : 2015REN20021 . tel-01174437

**HAL Id: tel-01174437**

**<https://theses.hal.science/tel-01174437>**

Submitted on 9 Jul 2015

**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.



**THESE / UNIVERSITE RENNES 2 HAUTE BRETAGNE**

*sous le sceau de l'Université européenne de Bretagne*

pour obtenir le titre de :  
**DOCTEUR DE L'UNIVERSITE EUROPEENNE DE BRETAGNE**  
*Mention : Géographie, télédétection, traitement du signal*  
**Ecole doctorale Sciences Humaines et Sociales**

présentée par

**Xing Gong**

Préparée à l'Unité Mixte de Recherche 6554 CNRS  
LETG Rennes COSTEL  
Université de Rennes 2 – Haute Bretagne

**Thèse à soutenir le 30 Janvier 2015**

devant le jury composé de

Analyse de séries temporelles d'images  
à moyenne résolution spatiale :  
reconstruction de profils de LAI,  
démélangeage. Application pour le  
suivi de la végétation sur des images  
MODIS

**Ping TANG**

Professeur CAS-IRSA, Pékin (Chine) / *rapporteur*

**Frédéric BARET**

Directeur de Recherche INRA, Avignon (France) / *rapporteur*

**Bao Gang HU**

Professeur CASIA, Pékin (Chine) / *Examineur*

**Nicolas COURTY**

Maître de Conférences, IRISA, Vannes (France) / *Examineur*

**Thomas CORPETTI**

Directeur de Recherche CNRS, Rennes (France) / *Directeur de thèse*

**Laurence HUBERT-MOY**

Professeur Université de Rennes 2 (France) / *co-directeur de thèse*











# Contents

<b>1</b>	<b>Introduction générale</b>	<b>3</b>
1.1	Contexte scientifique : rôle des variables biophysiques . . . . .	5
1.2	Reconstruction de variables biophysiques . . . . .	7
1.3	Démélangeage de séries temporelles . . . . .	8
1.4	Organisation du document . . . . .	9
<b>I</b>	<b>Première partie : reconstruction de séries temporelles de LAI à partir d'images MODIS</b>	<b>11</b>
<b>2</b>	<b>Biophysical variables reconstruction</b>	<b>13</b>
2.1	Introduction . . . . .	15
2.2	Introduction to the Greenlab model . . . . .	17
2.2.1	Biomass generation . . . . .	17
2.2.2	Biomass repartition . . . . .	18
2.2.3	Parameter Redundancy . . . . .	19
2.2.3.1	Focus main organs . . . . .	19
2.2.3.2	Plant structure simplification . . . . .	20
2.2.3.3	Spoilage and organ aging . . . . .	20
2.3	Data assimilation . . . . .	20
2.3.1	State transition model . . . . .	21
2.3.2	Observation model . . . . .	22
2.3.3	Model inference and parameter estimation techniques . . . . .	22
2.3.4	Application to the recovery of LAI using GreenLab . . . . .	24
2.4	Experimental Results . . . . .	25
2.4.1	Experimental settings . . . . .	25
2.4.2	Data with ground truth . . . . .	26
2.4.2.1	Result on sequence of noisy observations . . . . .	27
2.4.2.2	Sequence with sparse observations . . . . .	28
2.4.2.3	Sequence of missing observations . . . . .	29
2.4.2.4	Overview of these experiments . . . . .	29
2.4.3	Application on large scale study site . . . . .	30
2.5	Conclusion . . . . .	31
<b>II</b>	<b>Deuxième partie : démélangeage non-linéaire : aspects théoriques</b>	<b>35</b>
<b>3</b>	<b>NMF, application to unmixing</b>	<b>37</b>
3.1	Introduction . . . . .	39
3.1.1	Context . . . . .	39
3.1.2	Notations . . . . .	40
3.1.3	Related work . . . . .	40
3.1.4	Proposal . . . . .	41
3.1.5	Contributions . . . . .	42



3.1.6	Outline . . . . .	43
3.2	Geometry aware CSS procedure . . . . .	44
3.2.1	CSS as a manifold sampling procedure . . . . .	44
3.2.2	Original Simplex Volume Maximization . . . . .	45
3.2.3	Exact simplex volume maximization . . . . .	46
3.3	Sparse projections onto the RKHS simplex . . . . .	47
3.3.1	Projection on the RKHS simplex . . . . .	47
3.3.2	Convergence of the projector . . . . .	48
3.4	Complete SAGA procedure . . . . .	50
3.4.1	Algorithm . . . . .	50
3.4.2	Computational complexity . . . . .	50
3.5	Experiments and results . . . . .	52
3.5.1	Experiments on toy datasets . . . . .	52
3.5.2	Comparison on subsampling strategies . . . . .	54
3.5.3	Computational complexity . . . . .	56
3.5.4	Application to classification on real datasets . . . . .	56
3.6	Conclusion . . . . .	59
<b>III</b>	<b>Troisième partie : démélangeage de séries temporelles</b>	
	<b>61</b>	
<b>4</b>	<b>Time series unmixing</b>	<b>63</b>
4.1	Introduction . . . . .	65
4.2	Related work . . . . .	68
4.2.1	Endmember extraction introduction . . . . .	68
4.2.2	Abundance estimation . . . . .	69
4.3	Methodology for time series unmixing . . . . .	69
4.3.1	Kernel time series . . . . .	69
4.3.2	Endmember extraction . . . . .	72
4.3.3	Abundance estimation . . . . .	74
4.3.4	Overall procedure . . . . .	75
4.4	Experimental validation . . . . .	76
4.4.1	Experimental settings . . . . .	76
4.4.2	Dataset properties . . . . .	77
4.4.3	Endmember extraction validation . . . . .	78
4.4.4	Pixel unmixing result evaluation . . . . .	79
4.5	Conclusion . . . . .	81
<b>5</b>	<b>Conclusion &amp; Perspectives</b>	<b>85</b>
5.1	Résumé des contributions . . . . .	85
5.2	Perspectives . . . . .	86
<b>IV</b>	<b>Annexe</b>	<b>89</b>
<b>6</b>	<b>GreenLab plant growth model</b>	<b>91</b>
<b>7</b>	<b>Publications</b>	<b>95</b>
		<b>95</b>

<i>CONTENTS</i>	v
<b>Bibliography</b>	<b>97</b>
<b>List of Tables</b>	<b>113</b>
<b>List of Figures</b>	<b>115</b>



## Remerciements

As the end of this doctoral project approaches, there comes the reminiscence of my precious experiences in Beijing and Rennes. During these last years, I met many people, scientific or not, who helped me, in a way, to successfully accomplish this work. I feel like expressing my deepest gratitude to them all for everything.

First, I would like to thank my supervisor, Thomas Corpetti, from whom, I have learned really a lot. I am grateful for his generous support, his precious advices and his valuable friendship during the course of my master's and doctoral years. I also would like to thank my co-supervisors, Baogang Hu and Laurence Hubert-Moy, for their scientific knowledge on machine learning and remote sensing respectively. Without their support, this co-supervision between France and China would never be possible. A special thanks to the COSTEL and OBELIX groups for welcoming me in Brittany during my stay in France.

I want to thank the two referees of my manuscript, Prof. Ping Tang and Prof. Frédéric Baret for the time and energy they put in reading my manuscript. and their insightful comments are really appreciated.

Also I would like to thank Prof. Nicolas Courty, Prof. Mengzhen Kang, Prof. Véronique Prinnet for their valuable guidances on machine learning, computer vision, and plant growth model respectively. Their profound knowledges on respective research area have been a catalyst for my study. I want to express my earnest appreciation to all professors, employees, students from LETG-Rennes COSTEL and Institute of Automation, Chinese Academy of Sciences, they helped me on both the research and the life when I was in France and China, and they are also involved in my doctoral project.

My pleasant life experience in France was attributed to many friends: Pauline Dusseux, Xavier Foissard, Julie Betbeder, Igor Sirmik, Jean Nabucet, Alban Thomas, Jingjiao Zhang, Hongquan Wang, Emilie Bourget, Sébastien Rapinel, Clémence Vannier, Solen Leclerch, Véronique Michot, Edwige Motte, Haya Mahmoud, Heloisa de Camargo Tozato, Fernanda Cabre, Amit Kumar Sharma, Renan Le Roux, Mengyuan Liu, Cyril Bonnefoy, Lucelma Nascimento and many others. Special thanks to them for the nice atmosphere and all the good moments in France that we shared together.

Many thanks go to the interesting colleagues in the Sino-French Laboratory for Computer Science, Automation and Applied Mathematics (LIAMA). A variety of academic and social activities have extended my vision and my understanding of both scientific research and personal life. And I would like to thank my colleagues I met since my master's years in this group: Cui Zhang, Dongmin Ma, Fei Yuan, Cyril Cassisa, Antoine Lefebvre, Paul Marcombes, Pascal Zille, Pierre Allain, Chunguo Li, Xiaowan Zhang, Linlin Cao, Yong Zhang, Sichen Shen, Nan Zhang, Guanghui Ma, Zhiyong Ran, Baoyuan Wu, and others.

Huge thanks to all my family, particularly my parents, who provide perpetual support for all my life and encourage me in my choices. The most special thanks are for my girlfriend: Shan Huang, who tolerates numerous flaws in my character.

Thanks to all!



# Introduction générale



# Introduction générale

Les images satellites sont largement exploitées pour l'observation de l'environnement car elles fournissent une vision à différentes échelles d'un site d'étude et permettent ainsi de limiter l'accès au terrain. A l'heure actuelle, l'accumulation d'images acquises fournit une source d'information extrêmement précieuse [Jensen 1996][Sabins Jr 1978] autorisant une meilleure interprétation et compréhension de systèmes environnementaux, et notamment à travers l'étude de l'occupation et de l'usage des sols.

Parmi la variété de satellites disponibles, les satellites défilants à moyenne résolution spatiale, tels que MODIS ou encore le futur SENTINEL-2, possèdent un taux de revisite intéressant (une image par jour est disponible avec MODIS, cf quelques caractéristiques sur le tableau 1.1<sup>1</sup>) qui permet d'analyser finement la dynamique de certains phénomènes, notamment ceux liés à la végétation. C'est précisément sur cet aspect lié à la haute répétitivité temporelle des satellites à moyenne résolution spatiale que cette thèse s'intéresse.

L'analyse des séries temporelles résultantes pose à l'heure actuelle un certain nombre de difficultés, listées ci-dessous :

- Tout d'abord, les données sont souvent bruitées en raison de la couverture nuageuse, des conditions d'acquisition, de l'angle de vue, ..., comme illustré sur la figure 1.1.
- Par ailleurs, la nature temporelle des données fait qu'une forte corrélation entre les différentes valeurs au sein d'un même pixel existe. Il apparaît ainsi intéressant de prendre en considération ce lien pour analyser les séries et cela n'est à l'heure actuelle pas considéré dans la plupart des techniques.

<sup>1</sup>Source: V. Tramutoli, FORMation of Multi-disciplinary Approaches to Training in Earth Observation, Erasmus Intensive Programmes

Table 1.1: Caractéristiques en termes de résolution, nombre de bandes spectrales et fauchée des principaux satellites d'observation de la terre

Sensor	N° Band	Spatial Resol	Temporal Resol	Swath
IKONOS	5	1-4 m	~ 30 days	11 km
QUICKBIRD	5	0.61-2.88 m	~ 30 days	16.5 km
SPOT HRV	4	10-20 m	26 days	117 km
LANDSAT ETM+	8	15-30 m	16 days	185 km
MODIS	36	250-1000 m	1.5 days	2320 km
NOAA AVHRR	5	1100 m	0.25-0.5 days	2600 km
METEOSAT	3	2200-4500 m	0.5 hours	-
NOAA GOES	5	1000-8000 m	0.25 hours	-
MSG	12	1000-3000 m	0.25 hours	-



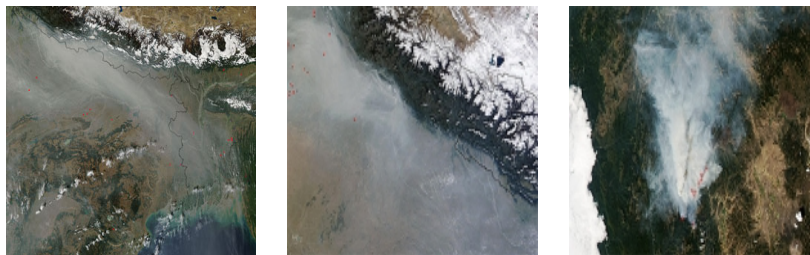


Figure 1.1: **Illustration de données brutes MODIS.** En raison des mauvaises conditions d'acquisition, des nuages, aérosols, ..., les séries temporelles résultantes sont bruitées.

La question de la comparaison des séries est notamment importante. En effet, bien que deux séries soient différentes si on les compare point par point, elles peuvent avoir des comportements dynamiques similaires mais simplement décalés dans le temps (en raison par exemple de phénomènes climatiques), comme illustré dans la figure 1.2. Ainsi, un critère de comparaison classique (basé sur la distance Euclidienne) pourra mettre en évidence des erreurs importantes pour des séries représentant un même comportement dynamique (type de culture, de pratique agricole, ...).

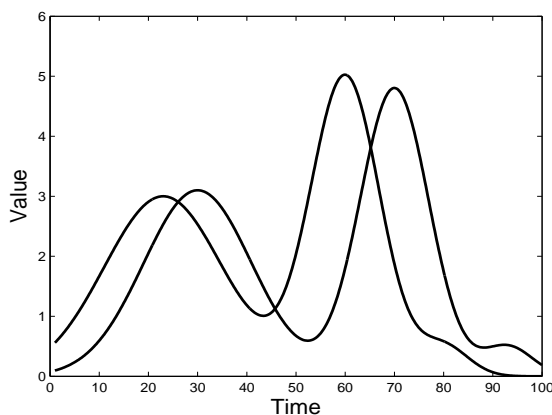


Figure 1.2: **Deux séries temporelles représentant un même comportement dynamique mais décalées dans le temps.** Ici, une comparaison point-par-point des valeurs de chaque série n'est pas optimale.

- Enfin, les satellites possédant une bonne résolution temporelle ont, en contrepartie, une résolution spatiale plus faible. Bien que le futur satellite SENTINEL-2 améliorera nettement la résolution par rapport à MODIS (10m versus 250m), le problème de démélangeage reste important. A l'heure actuelle, des techniques standard et très efficaces de démélangeage existent (cf [Bioucas-Dias 2012][Winter 1999][Iordache 2011]) mais aucune ne prend en compte la spécificité temporelle des séries.

Ce sont ces aspects (débruitage et démelangeage de séries temporelles) liés à la spécificité des données à moyenne résolution spatiale, qui sont au cœur de ce travail de thèse. Ils seront appliqués dans un contexte de suivi de la végétation.

## 1.1 Contexte scientifique : rôle des variables biophysiques

La végétation occupe une part importante des sols à l'échelle de la planète, que ce soit au travers des forêts, des prairies ou des cultures [Townshend 1991]. Ces zones végétalisées sont d'une grande importance aussi bien pour les écosystèmes que pour subvenir aux besoins alimentaires de la population. Le suivi et le contrôle de ces zones est donc primordial et depuis plusieurs décennies, des satellites d'observation de la terre sont lancés en ayant notamment pour objectif de quantifier les ressources disponibles [Richards 2013, Trotter 1991]. Afin de décrire l'état de la végétation, plusieurs paramètres ont été définis par les chercheurs [Carlson 1997][Gamon 1995]. Parmi les plus populaires, on peut notamment citer des indices de végétation comme le NDVI (Normalized Difference Vegetation Index) ou des variables biophysique telles que le fCover (fraction de couvert végétal), le fAPAR (fraction de rayonnement solaire absorbée par les plantes) ou le LAI (Leaf area index), ... La communauté scientifique en télédétection a été très active pour développer des techniques capables de mesurer de telles variables à partir d'images. Détaillons quatre d'entre elles :

1. Le **NDVI** a été utilisé dans de très nombreuses études en raison de la simplicité de son calcul. Cet indice de végétation s'appuie sur la propriété de réflectance spectrale des feuilles vertes (chlorophylle) dans plusieurs longueurs d'onde. Il se calcule en tout point de l'image  $(x, y)$  par la différence normalisée dans les bandes rouge et proches infra-rouge :

$$NDVI = \frac{I_{PIR}(x, y) - I_R(x, y)}{I_{PIR}(x, y) + I_R(x, y)}, \quad (1.1)$$

où  $I_{PIR}$  (respectivement  $I_R$ ) est la réflectance dans le proche infra-rouge (respectivement rouge). Le NDVI, compris entre  $-1$  et  $+1$ , est sensible à vigueur et à la quantité de la végétation. Pour les sols nus, étant donné que les réflectances sont pratiquement identiques dans les bandes rouges et proche infra-rouges, cet indice est proche de 0. A l'inverse, les formations végétales ont une valeur souvent comprise dans l'intervalle  $\sim [0.7, 1]$  en fonction de la densité du couvert. Il a été appliqué avec succès dans de très nombreuses études en analyse de la végétation ([Tucker 2005][DeFries 1994][Pettorelli 2005][Carlson 1994]). Néanmoins, cet indice n'est pas directement lié à une quantité physique et il a été démontré (voir par exemple dans [Gamon 1995] ou dans la thèse de Pauline Dusseux [Dusseux 2014a]) que sa valeur sature à partir d'une certaine densité de végétation.

2. Le **fAPAR** (Fraction of Absorbed Photosynthetically Active Radiation) [Knyazikhin 1998] est la fraction de rayonnement solaire absorbée par les plantes dans le domaine spectral permettant la photosynthèse ( $\sim [400 - 700nm]$ ). Cette variable biophysique est beaucoup utilisée pour modéliser les écosystèmes étant donné qu'elle intervient beaucoup dans les échanges

d'énergie, de vapeur d'eau et de dioxyde de carbone entre la surface terrestre et l'atmosphère. Elle a également été utilisée dans de nombreuses applications ([Ridao 1998][Gobron 2008] [Zhou 2002]).

3. Le **FCover** (Fraction of green vegetation Cover) [Gutman 1998] correspond à la fraction surface couverte par des feuilles vertes (en général, l'élément de surface considéré dans une image est le pixel). Le Fcover est sensible à la quantité de végétation mais indépendant de l'orientation de l'illumination. Cette propriété fait que cet indice est de plus en plus utilisé pour le suivi de la végétation ([Camacho 2013][Gutman 1998][LI 2003]).
4. Le **LAI** (Leaf Area Index) [Tripathi 2014] est un nombre sans dimensions défini comme le ratio entre la quantité de surface verte par unité de surface. Il est directement lié à la quantité de feuilles. C'est une variable clé qui intervient dans divers processus tels que l'absorption de la lumière par la canopée, la photosynthèse, la respiration, l'évapotranspiration, l'absorption d'azote ou encore l'interaction avec les précipitations. C'est un indice très utilisé pour analyser les systèmes végétaux ([Myneni 1997][Chen 1997] [Turner 1999][Pierce 1988]).

Comme il est démontré dans plusieurs études (cf notamment la thèse de Pauline Dusseux [Dusseux 2014a][Dusseux 2014b]), les techniques d'estimation du LAI sont assez efficaces (comme l'inversion du modèle PROSAIL, décrite plus bas) et les valeurs de LAI apparaissent plus robustes aux différentes conditions d'acquisition et moins sensibles aux effets de saturation que le NDVI. Nous avons ainsi fait le choix dans ce travail de s'appuyer sur des séries temporelles de LAI pour capter l'état de la végétation. Ainsi, réussir à dériver des estimations fiables de LAI à partir d'images satellites constitue un défi majeur avec de nombreuses applications potentielles, que ce soit pour l'analyse de l'occupation ou de l'utilisation du sol (en prenant notamment en considération l'évolution temporelles de ces paramètres).

C'est sur ce point précis que se focalise la première partie de ce document. On s'intéresse à la reconstruction de séries temporelles de LAI à partir d'images satellites à moyenne résolution spatiale, notamment en intégrant le modèle de croissance de plante GreenLab [YAN 2004] en utilisant des techniques d'assimilation de données. Les observations, qui permettent d'estimer le LAI à partir des différentes bandes spectrales de MODIS, sont quant à elle issues de l'inversion du modèle de réflectance PROSAIL [Jacquemoud 2009]. La seconde partie se focalise quant à elle sur le problème du démixage spatial des séries temporelles MODIS. En raison de la résolution spatiale modérée du capteur, la surface couverte par un pixel (250m x 250m avec MODIS, 10m x 10m avec SENTINEL-2) est susceptible de contenir plusieurs parcelles et l'information résultante correspond alors à un mélange. Le but du démixage est de recouvrer l'information associée à chaque matériel pur contenu dans un pixel [Bioucas-Dias 2009]. Le problème du démixage a beaucoup été abordé dans la communauté d'analyse d'images hyperspectrales mais à notre connaissance, il n'a pas encore été abordé pour le cas de profils temporels de variables biophysiques. Cela sera étudié dans cette thèse, notamment via des techniques de factorisation non-négative (NMF pour Non Négative Matrix Factorization en anglais) [Courty 2014] et en particulier la méthode SAGA (Sparse and Geometry Aware Matrix Factorization) qui sera introduite dans un premier temps.

Ensuite, nous l'adapterons au cas des séries temporelles, notamment en exploitant des noyaux élastiques [Cuturi 2007][Cuturi 2011] qui sont adaptés à la manipulation de séries.

Détaillons à présent ces deux problématiques.

## 1.2 Reconstruction de variables biophysiques

L'estimation de variables biophysiques à partir d'images satellites est une tâche délicate qui a été abordée par plusieurs auteurs. Un certain nombre de modèles, basés sur l'absorption et la réflectance de la lumière par la végétation en fonction de variables biophysiques, ont été proposés. Ceux-ci permettent de représenter les signaux observés dans les images et l'inversion de ces modèles autorise alors l'estimation de variables à partir d'images [Price 1995][Gower 1999]. Parmi les modèles les plus populaires, PROSAIL [Jacquemoud 2009], qui est la combinaison du modèle de réflectance bidimensionnel SAIL (Scattering by Arbitrary Inclined Leaves) [Verhoef 1984] avec le modèle PROSPECT [Jacquemoud 1990] se focalisant sur les propriétés des feuilles, a été largement utilisé [Jacquemoud 2009, Lecerf 2008, González-Sanpedro 2008]. C'est l'inversion de ce modèle qui sera utilisée pour accéder au LAI à partir d'images MODIS [Lecerf 2008].

Cependant, au sein d'une image satellite, en raison de l'angle de prise de vue, des effets d'aérosols, de la couverture nuageuse, ..., l'estimation directe des variables biophysiques dans les images est susceptible d'être bruitée et incohérente, les corrections radiométriques et atmosphériques appliquées sur les images ne permettant pas de s'affranchir de tous ces effets (cf par exemple les images MODIS de la figure 1.1). Du point de vue temporel, les séries résultantes sont alors inexploitable. Bien que des techniques simples d'interpolation (linéaire, cubiques, ...) peuvent en partie débruiter le signal [Jonsson 2002], ici il semble intéressant de prendre en considération le fait que nous manipulons des variables physiques qui respectent certaines propriétés. En effet, dans le domaine de l'agriculture et/ou de l'écologie, un certain nombre de connaissances sur l'évolution du LAI ont été acquises et il semble alors pertinent de les prendre en considération pour débruiter les estimations réalisées au sein de chaque image. A partir des connaissances acquises, des modèles de croissance de plantes, développés notamment pour optimiser la gestion des cultures [Brisson 2003], décrivent les flux internes et externes de ressources disponibles pour une plante (biomasse, ensoleillement, ...) et modélisent sa croissance. Le modèle GreenLab [YAN 2004], qui fait partie de cette famille, représente l'évolution d'une plante comme une succession de cycles au sein desquels de la biomasse est générée (en fonction de l'ensoleillement, de la quantité de feuilles, ...) puis répartie dans les différents organes (fruits, branches, feuilles, ...).

Dans cette thèse, nous nous sommes appuyés sur le modèle GreenLab pour générer des connaissances a priori sur la dynamique du LAI. Une première étape a été réalisée en vue d'adapter le modèle à l'échelle MODIS (simplification, extraction des paramètres les plus pertinents, ...) et ensuite, ce modèle simplifié a été appliqué à une technique d'assimilation de données pour recouvrer des séries temporelles cohérentes de LAI.

### 1.3 Démélangeage de séries temporelles

Un pixel est dit mixte lorsqu'il est la combinaison de plusieurs signaux issus de matériaux dits purs, ces matériaux étant situés dans la surface couverte par le pixel. Les éléments purs sont souvent appelés *endmembers* en anglais et la fraction qu'ils occupent se nomme l'*abondance*. Le problème de démélangeage consiste donc à estimer les endmembers (nombre, signature) et leur abondance en chaque pixel.

Ce problème a été largement abordé, notamment par la communauté scientifique en imagerie hyperspectrale [Camps-Valls 2005][Chan 2009]. Dans ce type d'imagerie, les rayonnements sont captés dans une centaine (minimum) de bandes spectrales [Bioucas-Dias 2012][Keshava 2002] et cette richesse de longueurs d'ondes permet d'isoler les éléments purs (eau, bâti, végétation, ...), ouvrant l'accès à un ensemble de techniques de démélangeage. La majorité d'entre elles s'appuie sur le fait que la signature spectrale d'un élément pur est invariante au cours du temps. Un grand nombre de méthodes robustes et stables a été proposé sur ce principe (cf le papier d'overview [Bioucas-Dias 2012]). On peut par exemple citer l'approche de Adam et al qui analyse les roches en utilisant un modèle de mélange spectral [Adams 1986], ou bien celle de [Smith 1985] qui s'intéresse aux ressources minérales par analyse en composantes principales. Des spectres de référence acquis en laboratoire et correspondant aux éléments purs sont utilisés comme endmembers, tandis que l'abondance est déterminée par le ratio de la distance entre les pixels mixtes et les endmembers.

Malgré leur efficacité, de telles approches ne peuvent s'appliquer pour le démélangeage de paramètres biophysiques. En effet, bien que l'hypothèse de stabilité des signatures spectrales soit valide pour les minéraux, sols, surfaces en eau, ..., cela n'est plus vrai dans un contexte agricole [Goenaga 2013]. Les états de végétation dépendent en effet de paramètres climatiques et saisonniers [Walthall 1992] et des motifs temporels particuliers, liés à cette évolution, apparaissent. Ainsi, si l'on s'intéresse au démélangeage dans un contexte de suivi de la végétation, l'usage de séries temporelles s'avère important [Dusseux 2014b]. C'est l'axe privilégié dans ce document de thèse. La technique d'extraction d'éléments purs que nous mettrons en œuvre s'appuie sur les noyaux reproduisants et notamment sur le RKHS (Reproducing Kernel Hilbert Space, [Courty 2014]) pour représenter l'information de manière optimale. Pour l'extraction des endmembers, une technique du simplex sera utilisée, chaque élément pur sera alors un des sommets de ce simplexe et l'abondance sera estimée par projection de chaque coordonnée dans le RKHS sur ce simplexe. Comme nous le verrons par la suite, il se posera alors la question de la comparaison de profils temporels : deux séries peuvent représenter le même type de végétation mais peuvent avoir tout de même des différences en termes d'amplitude et de décalage temporel en raison des variabilités climatiques. Ainsi, les distances usuelles ne sont pas optimales. Nous porterons une attention particulière aux méthodes utilisant des noyaux élastiques et notamment s'appuyant sur les principes de la distance *Dynamic Time Warping* [Thureau 2012][Cuturi 2007][Marteau 2009].

## 1.4 Organisation du document

La première contribution de ce document portera sur le problème de la reconstruction de séries temporelles de LAI par assimilation de données. Elle sera présentée dans le **chapitre 2**. Ce chapitre correspond à un article en révision dans la revue *IEEE Transactions on Geosciences and Remote Sensing* [Gong sionb]. On trouvera en premier lieu une description des principales variables biophysiques et de la manière dont on peut les estimer par imagerie. Ensuite, le modèle GreenLab sera présenté ainsi que son adaptation aux images satellites à moyenne résolution spatiale. Les techniques mathématiques liées à l'assimilation de données seront alors introduites, en se focalisant notamment sur les approches stochastiques qui seront plus particulièrement utilisées [Doucet 2001], notamment pour gérer la non-linéarité liée à la fois au modèle GreenLab et au modèle PROSAIL. L'approche résultante sera testée et validée sur des données synthétiques et réelles.

Dans le **chapitre 3**, la théorie liée à la méthode de démélangeage SAGA sera présentée. Ce chapitre correspond à l'article *SAGA: Sparse And Geometry-Aware non-negative matrix factorization through non-linear local embedding* publié dans la revue *Machine Learning* [Courty 2014]. L'approche SAGA est une méthode supervisée qui vise à décomposer un jeu de données  $X$  (un ensemble de séries temporelles dans notre cas) par le produit de deux matrices :  $X = FG^T$  (plus précisément, on recherche  $F$  et  $G$  minimisant  $\|X - FG^T\|$ ). Certaines contraintes sont appliquées sur  $F$  et  $G$ , notamment le fait que les colonnes de  $F$  soient composées d'éléments de  $X$  (la base de représentation est alors issue du jeu de données, ce qui constituera une aide pour la recherche d'éléments purs) et les éléments de  $G$  sont positifs et somment à 1 (ils contiennent l'abondance de chaque élément pur). Tout cela est écrit dans un espace RKHS (mentionné plus haut) afin d'améliorer la représentativité des données, cette réécriture s'effectuant par des fonctions noyaux [Courty 2014]. Les principes sont validés sur des jeux de données synthétiques et réels.

Dans le **chapitre 4**, l'adaptation de la méthode SAGA au problème de démélangeage de séries temporelles est présentée. Ce chapitre correspond à l'article *Time series unmixing with kernel space non-negative matrix factorization* soumis à la revue *IEEE Transactions on Geosciences and Remote Sensing* [Gong siona]. Dans ce chapitre, nous explorons des noyaux adaptés aux séries temporelles [Cuturi 2011][Marteau 2009] pour mener à bien cette tâche. Les limites des méthodes conventionnelles de démélangeage hyperspectral pour des données temporelles sont discutées, ainsi que les défis à relever sur cet aspect. Une procédure complète en quatre étapes est proposée et les performances sont discutées.

Enfin, le **chapitre 5** présente une synthèse et des perspectives associées à ces travaux.

Chaque chapitre est introduit par un résumé illustré de deux pages en français.

Au cours de mes travaux de thèse, trois articles de revues ont été publiés, quatre conférences internationales et deux articles sont en révision/soumission. Ces publications sont listées dans le dernier chapitre.



## Part I

# Reconstruction de séries temporelles de LAI à partir d'images MODIS





# Biophysical variables reconstruction

## Résumé du chapitre

Dans ce chapitre, on s'intéresse à la reconstruction des séries temporelles cohérentes de LAI à partir d'images MODIS "brutes" qui contiennent un certain nombre d'artefacts, comme illustré sur la figure 2.1. Pour cela, nous suggérons

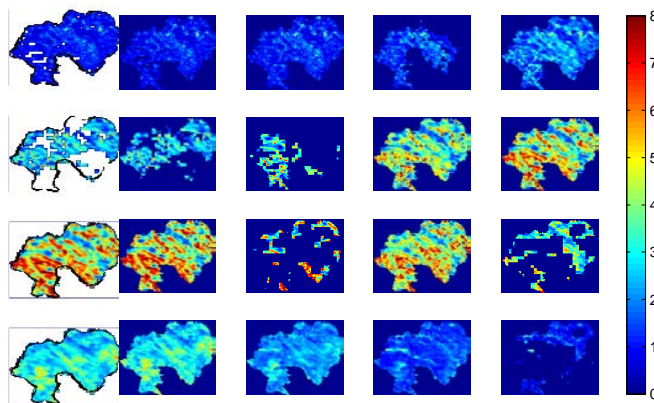


Figure 2.1: LAI calculé sur une série d'images brutes MODIS sur le site de **Chizé, France**. Les zones en blanc (à l'intérieur de la région d'intérêt) correspondent à des données manquantes.

d'exploiter des techniques d'assimilation de données appliquées sur des mesures de LAI, issues de l'inversion du modèle PROSAIL. Le modèle dynamique sur lequel nous menons l'assimilation est le modèle GreenLab, qui est un modèle fonctionnel/structurel de croissance de plantes. Nous nous intéressons à la partie fonctionnelle (qui décrit notamment l'évolution et la répartition de la biomasse) sur quatre types de cultures: maïs, blé, colza et tournesol (principales cultures sur ce site). Les résultats montrent, sur des données synthétiques et réelles, qu'il est possible de recouvrer des valeurs cohérentes de LAI, mais aussi d'estimer la répartition de la biomasse au sein d'une culture (partie dans le fruit, la végétation ou dans les branches/tiges). Ce chapitre est structuré de la manière suivante :

- La **section 2.1** présente une introduction au contexte ainsi que les problématiques;

- La **section 2.2** présente le modèle GreenLab et la manière dont on l'exploite pour des images MODIS;
- La **section 2.3** établit une présentation des techniques d'assimilation de données, et notamment des techniques particulières qui sont utilisées ici. Leur adaptation au modèle GreenLab précédemment introduit est également proposée ;
- La **section 2.4** présente des résultats expérimentaux;
- Une **annexe** détaille la manière dont on estime le LAI à partir d'images.

Ce chapitre correspond à un article en révision dans la revue *IEEE Transactions on Geosciences and Remote Sensing*.

---

---

## Abstract

The main objective of this chapter is concerned with the recovery of Leaf Area Index (LAI) time series in intense agriculture areas from medium resolution remote sensing data (such as MODIS or future SENTINEL). Although their moderate resolution limits an analysis at a parcel level, their high temporal rate still enables to monitor land use through the temporal evolution of key biophysical parameters as LAI. However in practice, a frame-by-frame estimation is unsatisfactory since the quality of each single data is subjected to undesirable effects due to atmosphere disturbance, lighting conditions, shooting angle, etc. These effects lead to a lack of temporal consistency of resulting time series.

In this chapter, we propose an approach to recover time consistent series of LAI from noisy and sometimes corrupted sequential remote-sensing data. It combines the prior information of a plant growth model, namely GreenLab, and low quality remote sensing observations using stochastic data assimilation techniques. Our experiments demonstrate the approach is robust on various challenging situations and enables to extract additional information about observed fields.

## Keywords

Leaf Area Index, GreenLab model, data assimilation, noisy/missing data

## 2.1 Introduction

Land surface on earth is mainly covered by green plants including forests, grasslands or agricultural fields [Townshend 1991]. These vegetation areas are of great importance for global ecological systems and food supply for human beings. Therefore in the past several decades, many satellites have been launched to observe and monitor these important areas [Richards 2013, Trotter 1991]. Since remote sensing data came into being, a large number of important tools have been developed to extract some crucial information on crops and plants from satellite images. This step is indeed a prerequisite in many agricultural applications such as monitoring the plant physiology and/or ecology at different scales [Gao 1996, Huete 1988, Liang 2005]. One of the most popular vegetation index is the NDVI (Normalized Difference Vegetation Index), based on the spectral reflectance property of green leaves (chlorophyll) at different wavelength, NDVI is a normalized measurement of the difference between the canopy reflectance in the red and near-infrared wavebands. While NDVI is a sensitive indicator of canopy structure and chemical content (plant biomass, leaf area index, chlorophyll content in sparse canopies and foliar nitrogen content), it is still a limited index for total canopy biomass in high-density vegetation areas [Gamon 1995]. Some other more sophisticated indexes have been proposed, such as fCover (vegetation cover fraction), fAPAR (fraction of photo-synthetically active radiation absorbed by the canopy) or LAI (Leaf Area Index) which is largely used in practice [Baret 2007]. Readers can find in [Carlson 1997, Gamon 1995] more details about these biophysical variables. In practice, the estimation of these biophysical variables from remote sensing is a delicate task. Several models, based on light interception by plant canopies and vegetation reflectance

in terms of biophysical characteristics, have been designed and inverted to retrieve biophysical information from optical satellite images. Among popular models in this area, the SAIL (Scattering by Arbitrary Inclined Leaves) canopy bidirectional reflectance model [Verhoef 1984] and the PROSPECT one (based on leaf optical properties) are widely used [Schaepman-Strub 2006]. For example in [Weiss 2000], the authors estimate LAI and fCover by inverting SAIL radiative transfer model. Despite reliable results, the associated variables still suffer from a lack of consistency. In [Combal 2003], results over simulated reflectance data sets were improved by incorporating some prior information such as ancillary data (leaf water content) and canopy type. In [Huemmrich 2001], Huemmrich has combined the SAIL model with geometric models to simulate horizontally discontinuous canopies. Even though the results were globally satisfactory, strong improvements have been obtained when one mixes these two models. The resulting one, PROSAIL, has allowed to describe both the spectral and directional variation of canopy reflectance based on single static images [Jacquemoud 2009, Lecerf 2008, González-Sanpedro 2008]. All these methods have been designed for single images and only few of them [Roerink 2000, Jonsson 2002] exploit the time series observation to derive the plant growth evolution information : in [Roerink 2000], Roerink *et al* have developed a Fourier analysis method to reconstruct cloud free NDVI composites, while in [Jonsson 2002], the authors have proposed a technique based on non-linear least squares fits of asymmetric Gaussian model functions to derive information about the seasonal vegetation development from cloud contaminated observations. Despite the fact that they provide interesting information, these methods only make use of consistency in frequency or temporal domain but no prior knowledge about plant or vegetation growth is taken into consideration. In this study we suggest to exploit the periodicity of satellites to improve the quality of estimated biophysical variables. Though we face here the problem of LAI estimation (LAI being one of the most used parameter), our methodology can be applied to any biophysical variable provided that a physical model of its temporal evolution exists. Given a sequence of  $T$  observations  $\bar{O} = \{O_{t_k} | k = 1, \dots, T\}$  at times  $\{t_k | k = 1, \dots, T\}$ , we aim at reconstructing a sequence of  $M$  LAI values  $\{S_{t_k} | k = 1, \dots, M\}$  at times  $\{t_k | k = i, \dots, M\}$  with  $M > T$ , meaning that we reconstruct series of LAI even at locations where no observations are available. To this end, instead of relying on classic interpolation, we propose to achieve the reconstruction by a physical plant growth model.

Plant growth models have initially been developed for agronomic purposes, harvest prediction or optimum crop management [Brisson 2003] and rely on *physical* properties: they describe the flux of external and internal resources through the element “plant” in order to expect its yield. Meanwhile, descriptive models as plant *architectural* ones have been introduced for computer graphics purposes. Associated modeling concepts and software for constructing 3D plants date back to the invention of L systems [Lindenmayer 1968], which is a parallel rewriting system and a type of formal grammar to model plant morphology.

In practice, plants are complex systems that involve metabolic physical mechanisms depending on the topology of the plant. Therefore, *structural-functional* plant models have been developed to perform efficient and realistic dynamic simulations of plant morphogenesis (see for example [Jallas 2000, Perttunen 1996]). The GreenLab plant growth model [YAN 2004] belongs to this family and has been designed to provide dynamic representations of the

morphogenesis and architecture of plants. It consists in a recursive process that produces in each step:

1. a total biomass yield (consumption/production) ;
2. a new plant architecture based on the repartition of the biomass on each organ of the plant (internodes, leaves, fruits, etc).

Obviously, the GreenLab model requires a number of plant-dependent parameters [Zhan 2003] that can be classified into two categories: those which can be estimated from direct (likely destructive) plant measurements (geometry, weight and number of fruits, internodes, leaves, petals, etc) and hidden ones deeply implied in the recursive simulation process (external effects as the photosynthesis efficiency or the biomass demand of various organs, etc). A manual collection of adequate observations and measurements required by GreenLab to fix the parameters would be time-consuming and work-intensive. This indeed requires cultivating a plantation of given plant species, to dig out a small amount of individual plant units at a regular time interval and to perform extensive and destructive measurements.

Retrieving all these parameters from remote sensing data is also unlikely to be satisfied since too many phenomena are involved. Fortunately, most vegetation studies from remote sensing require only certain aspects of plant growth (for example the internal architecture of each individual plant is not required). Therefore a simplified Greenlab model with significantly reduced number of parameters can be adequate for remote sensing. From such a model, the use of data assimilation methods enable to infer these parameters. This is the scope of this study.

The overall paper is organized as follows: the next section introduces the main principles of GreenLab plant growth model and the way we adapt it to MODIS time series. Then, in section 2.3, the proposed data assimilation technique that combines GreenLab and MODIS data is presented. Finally, section 2.4 show some experiments on synthetic and real data.

## 2.2 Introduction to the Greenlab model

GreenLab plant growth model is a powerful and sophisticated tool that simulates plant growth evolution. For each type of plant, this model is calibrated with ad-hoc observations of growth made under controlled environment and where extensive and even destructive measurements of individual plants are made regularly [Guo 2006, Kang 2008]. This enables to derive at first a deterministic description of a plant evolution. In more recent research, some stochasticity has been introduced to bring about plant growth variations [Lopez 2008, Wang 2011]. As mentioned in previous section, two steps are required (biomass yield and biomass repartition in plant organs) that are roughly introduced below.

### 2.2.1 Biomass generation

The biomass production comes from a large number of factors as photosynthesis efficiency of plant species, total leaf area, environmental context, etc. In

GreenLab, the production is done over “growth cycles” through the relation:

$$Q_{t_k} = E \cdot S_p \cdot \frac{1}{r} (1 - \exp(-k \cdot \frac{S_{t_k}}{S_p})) \quad (2.1)$$

where:

- $Q_{t_k}$  is the Biomass generated at the growth cycle between  $t_k$  and  $t_{k+1}$  ;
- $E$  is an empirical environmental factor (sun lighting, temperature, etc);
- $S_p$  is the land occupation of the plant;
- $k$  is photosynthesis efficiency;
- $r$  is an resistance parameters, setting the effect of mutual shading;
- $S_{t_k}$  is the total leaf area at  $t_k$ .

From Eq. (2.1), one observes that biomass generation is largely dependent of the leaf area  $S_{t_k}$  and is also restricted by land occupation  $S_p$ , surrounding environment  $E$  and species-dependent coefficient constant  $r, k$ . Let us now introduce the biomass repartition.

## 2.2.2 Biomass repartition

Once the biomass is produced, it has to be redistributed among different organs of the plant. GreenLab models the sink-demand strategy for biomass partition as follows [YAN 2004]:

$$D_{t_k} = \sum_{o \in \Omega} \sum_{i=1}^n N_o(i) \cdot d_o(i) \quad (2.2)$$

$$d_o(i) = S_o \cdot \phi(i) \quad (2.3)$$

$$\Delta_{q_o(t_k)} = d_o(i) \cdot \frac{Q_{t_{k-1}}}{D_{t_k}} \quad (2.4)$$

with

- $D_{t_k}$  is the total biomass demand at the growth cycle between  $t_k$  and  $t_{k+1}$  ;
- $\Omega$  is the set of various organs;
- $d_o$  is the biomass demand of an organ  $o \in \Omega$ ;
- $S_o$  is the sink-strength of organ  $o \in \Omega$ ;
- $i$  is the growth age for individual organ;
- $N_o(i)$  is the number of organ of type  $o$  at growth age  $i$ ;
- $\phi_o(i)$  is beta law function, it is small at the beginning and end of organ growth  $t_o$ , relatively large during the growth process [Guo 2006], takes the following form:

$$\phi_o(i) = (i - 0.5)^{a_o-1} \cdot (t_o - i + 0.5)^{b_o-1} \quad (2.5)$$

The idea behind this model is that all organs of a plant share a common biomass pool (generated by Eq. (2.1)) and are therefore in competition each others. The association of the biomass to one specific organ is based on its demand  $d_o$  (cf Eq. (2.3)) and depends on its sink strength  $S_o$  and growth age. The strength is plant-dependent: for example concerning fruit trees, the fruit organ tends to have more demand than a tiny root. Therefore  $S_o$  in Eq. (2.3) is bigger for the fruit organ than tiny root one. As for the age component, GreenLab assumes that an organ grows relatively slowly at the beginning and end of his life and accumulate most portion of biomass in the middle part. This is modeled by function  $\phi_o(i)$ . Finally, the role of Eq. (2.4) is to distribute the biomass among organs based on individual demands and the total available biomass.

In agricultural remote sensing applications, we are particularly interested on the total leaf area of the plants. In practice for most agriculture crops, other interactions related to wither (mainly in autumn) occur. This withering can be modeled through the following rule: at  $t_{k+1}$ , the total leaf area transition from  $t_k$  is modeled as:

$$S_{t_{k+1}} = \alpha \cdot (S_{t_k} - \beta_{t_k}) + \Delta_{q_l(t_k)} \quad (2.6)$$

In above equation, the withering of leaves at the growth cycle between  $t_k$  and  $t_{k+1}$  is denoted as  $\beta_{t_k}$ . This parameter is plant-dependent and is conditioned by the age of leaf which in practice is available through the structural sub-model and balanced by a parameter  $\alpha \in [0, 1]$  which stands for the spoilage in natural growth. The term  $\Delta_{q_l(n)}$  is the new growth of total leaf area, calculated from its share of biomass in Eq. (2.4).

The equations (2.1–2.6) are the core of the GreenLab model for biomass generation and partition. Of course in practice, this generation and partition is likely to be influenced by many other natural factors. In our applications which combine GreenLab with physical observation, we will model the contribution of all these factors with Gaussian white noises, as will be shown in the next section. Before that let us now discuss about parameter redundancy in GreenLab.

### 2.2.3 Parameter Redundancy

As shown above, plant growth is a very complex process which involves a large number of parameters that may be unidentifiable from remote sensing observation data. According to Yang *et al* in [Yang 2008], The GreenLab model includes redundant parameters which are intractable in practice when one observes images only. Fortunately, Not all these parameters are required for vegetation monitoring with medium resolution sensors. We then suggest to simplify and adapt the initial GreenLab model to remote sensing. These manipulations are detailed bellow.

#### 2.2.3.1 Focus main organs

we decide to take into consideration only three important organs responsible of the majority of biomass production: leaf, fruit and branch (internode). These three organs are of main ones encountered in most agricultural applications. The remaining ones are modeled with noises.



### 2.2.3.2 Plant structure simplification

in order to describe plant structure evolution, we keep an empirical Look-Up-Table,  $U = \{U_i | i = 1, \dots, N\}$  which store statistical number for different appearing organs. For instance,  $U_4 = \{l = 3, b = 2, f = 0\}$  means that at 4<sup>th</sup> growth cycle, there would be 3 new leaves, 2 new internodes and no new fruits. If organ ageing is not taken into account, at any given time, one can calculate the number of different organs by a simple summation and the growth age of individual organ by differentiating the current time with its appearing time point. These statistical information for plant species could be easily estimated by biological specialist and is the prior knowledge of plant growth. It is important to outline that the topological architecture is not kept here (meaning we do not need to know which fruit grows on which branch). This drastically simplified structure model offers necessary information for GreenLab functional model.

### 2.2.3.3 Spoilage and organ aging

for organs, expansion time  $t_o$  in Eq. (2.5) and life expectance  $t_e$  are key parameters: the organ only grows within its expansion time and after this step, it does not join the biomass repartition in Eq. (2.2) anymore (this models for example the fact that stalks stop to grow after autumn); as for life expectance parameter  $t_e$ , it models how long an organ could stay alive on plant. For instance, at time  $t_k$ , the photosynthesis active number of leaves  $N_s(t_k)$  could be calculated by Eq. (2.7):

$$N_s(t_k) = \sum_{i=t_k-t_e}^{t_k} U_i\{l\}. \quad (2.7)$$

Expansion time and life expectance are also given by biologists. Concerning  $\beta_{t_k}$  in Eq. (2.6), it is computed with following formula based on leaves's appearing time and life expectance  $t_e$ :

$$\beta_{t_k} = \frac{U_{t_k-t_e}\{l\}}{N_s(t_k)} \cdot S_{t_k} \quad (2.8)$$

These three points lead us to a simplified version of GreenLab, which depends now only on leaf mutual shading parameter  $r$ , photosynthesis efficiency  $k$  in Eq. (2.1), and natural leaf foliage  $\alpha$  in Eq. (2.6). In Figure. 2.2, we give some illustrations of the leaf area index evolution depending on these parameters. This new version of the model can now be used to interpret medium resolution remote sensing data. In the following part, we introduce the data assimilation scheme we used to combine images with our simplified GreenLab model.

## 2.3 Data assimilation

Combining different sources of information driven by a physical model, commonly named “data assimilation”, has always been a hot research topic for which a large panel of approaches exist (variational [DIMET 1986, Lions 1971] & stochastic [Bain 2009] in particular). Many problems in earth observation require to estimate a state of a system along time using a sequence of noisy measurements. Data assimilation performs this task by combining different

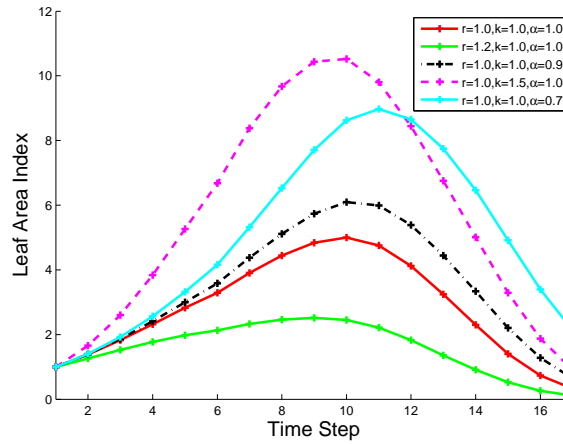


Figure 2.2: Leaf area index evolution with different  $r, k, \alpha$

sources of information, namely observation and prior knowledge as mathematical dynamical models and statistics. It is a widespread mathematical tool in various research areas such as fluid mechanics, atmosphere physics, oceanology or even pattern recognition [Evensen 1994]. Researchers in this area have developed a number of efficient inference tools, such as stochastic methods [Doucet 2000], Kalman filter [Welch 1995] and variational approaches as the well-known 4DVAR algorithm [Courtier 1994].

In all these techniques one can model the problem as follows: the data to estimate, driven by a dynamical model, are represented through a system state discretized into a sequence of hidden states [Kitagawa 1996] and are observed directly or indirectly through an observation system. The goal is to recover the best system state consistent both with observations and prior dynamical model. In general the solving process consists of two sub-steps. First, a state transfer system based on the dynamical model performs a first estimation of hidden states; secondly, with the use of an observation operator, a correction of this estimation (by comparing them with observations) is applied.

In this paper, the leaf area index  $S_{t_k}$  is the hidden variable at time point  $t_k$  and the model is introduced below. Let us now introduce the state transition model.

### 2.3.1 State transition model

The state transition model aims at switching from LAI at time  $t_k$  to  $t_{k+1}$  is written as:

$$S_{t_{k+1}} = M_k[S_{t_k}] + \eta_{t_k} \quad (2.9)$$

where  $\eta_{t_k}$  is the system noise to estimate that models associated uncertainties. The prediction model  $M_k$  is based on Greenlab and is modelled in Eq. (2.6).

### 2.3.2 Observation model

The observation model maps hidden variables to observation data:

$$O_{t_k} = H[S_{t_k}] + \epsilon_{t_k} \quad (2.10)$$

with  $O_{t_k}$  the observation at time  $t_k$ ,  $H$  the associated observation model and  $\epsilon_{t_k}$  associated uncertainties to estimate on the observation system. In our application, the observed data is the intensity value on MODIS images time serie. Because of the complexity of the relationship between the image luminance and the LAI, the definition of a complete observation system  $H$  that directly links the LAI to MODIS data would obviously be tricky. We therefore prefer to adopt a two stage strategy where at first we estimate the leaf area index thanks to the PROSAIL model [Jacquemoud 2009, Lecerf 2008] and then use an identity observation operator in  $H$ . This “pseudo-observation” approach is commonly used when data assimilation systems involve complex relationships between observations and system states [Courtier 1987]. In practice, a neural network is used to inverse PROSAIL model to estimate biophysical variables from MODIS time series (more details are shown in Appendix). In the following we describe the way we estimate the system state and associated uncertainties.

### 2.3.3 Model inference and parameter estimation techniques

Among possible solutions mentioned above to perform model inference, variational and stochastic techniques are the most popular. Variational techniques, also known as variational data assimilation, perform the estimation by minimizing a cost-function issued from a deterministic formalization in a Bayesian framework of the problem. It extracts the best compromise between observations, dynamic model and confidence measurements. Due to a rewriting in a dual space [DIMET 1986], the gradient of the cost-function can efficiently be obtained using a forward-backward integration of the dynamical model and such techniques are well adapted when one deals with large system states. However non-linear dynamic models need to be managed in an incremental framework corresponding to a succession of linearized problems. When one deals with smaller system’s state and non-linear dynamic models, stochastic techniques as the particle filter or the particle smoother, related to Monte Carlo approaches, are often preferred. The main idea consists in manipulating a set of particles more or less connected to the final state to estimate, this latter resulting from linear combination of such particles. Particle filtering, also known as Sequential Monte Carlo [Doucet 2001], is an attractive way to recover the system state for all time steps  $t = \{t_1, \dots, t_k, \dots, t_M\}$ . Starting from:

- An observation sequence  $\bar{O}_{t_k} = \{O_{t_1}, O_{t_2}, \dots, O_{t_k}\}$  available sequentially;
- An initial distribution of the system’s state  $p(S_{t_1})$ ;
- Transition and observation models:  $p(S_{t_k}|S_{t_{k-1}})$  and  $p(O_{t_k}|S_{t_k})$  respectively, related to the stochastic processes  $S$  and  $H$  presented above;
- Common assumption of zero-mean time-independent Gaussian noise for  $\eta, \epsilon$  in (2.9-2.10), leading to  $p(S_{t_k}|S_{t_{k-1}}) \sim \mathcal{N}(M_k(S_{t_{k-1}}), \eta)$  and  $p(O_{t_k}|S_{t_k}) \sim \mathcal{N}(H_k(S_{t_k}), \epsilon)$  with  $\mathcal{N}$  the normal distribution,

it can be shown (see [Doucet 2001] for details) that a sequential estimation of  $(S_{t_k})$  is obtained through the two-fold prediction/correction system:

1. **Prediction step:** given the available observation sequence  $\bar{O}_{t_{k-1}} = \{O_{t_1}, O_{t_2}, \dots, O_{t_{k-1}}\}$ , we predict next hidden state  $S_{t_k}$  via a propagation within the transition system  $p(S_{t_k} | S_{t_{k-1}})$ :

$$p(S_{t_k} | \bar{O}_{t_{k-1}}) = \int p(S_{t_k} | S_{t_{k-1}}) p(S_{t_{k-1}} | \bar{O}_{t_{k-1}}) dS_{t_{k-1}} \quad (2.11)$$

2. **Correction step:** we update this prediction with a new observation  $O_{t_k}$ :

$$p(S_{t_k} | \bar{O}_{t_k}) = \frac{p(O_{t_k} | S_{t_k}) p(S_{t_k} | \bar{O}_{t_{k-1}})}{p(O_{t_k} | \bar{O}_{t_{k-1}})} \quad (2.12)$$

with

$$p(O_{t_k} | \bar{O}_{t_{k-1}}) = \int p(O_{t_k} | S_{t_k}) p(S_{t_k} | \bar{O}_{t_{k-1}}) dS_{t_k} \quad (2.13)$$

The direct computation of the filtering distribution  $p(S_{t_k} | \bar{O}_{t_k})$  is in practice intractable for real systems. The main idea of particle methods is to sample this distribution by a set of  $N$  *particles* weighted by  $w^i(S_{t_k} | t_{k-1})$ ,  $i = \{1 \dots N\}$  [Bonet 1999, Nummiaro 2003] such that:

$$p(S_{t_k} | \bar{O}_{t_k}) = \sum_{i=1}^N w^i(S_{t_k} | t_{k-1}) \delta(S_{t_k}^i) \quad (2.14)$$

where  $\delta(S_{t_k}^i)$  is a dirac function centered in  $S_{t_k}^i$  and the weights  $w$  depend on the relative adequacy of the current observation with all particles:

$$w^i(S_{t_k} | t_{k-1}) = \frac{p(O_{t_k} | S_{t_k}^i)}{\sum_{j=1}^N p(O_{t_k} | S_{t_k}^j)} \quad (2.15)$$

Therefore, once the system at time  $t_{k-1}$  has been obtained, the process consists in generating a prediction of all particles at time  $t_k$  thanks to the transition model and the available observation (prediction step in Eq. (2.11)). These predictions are then corrected by taking into account the new observation  $\bar{O}_{t_k}$  in the correction step. The final estimated distribution is obtained from the set of particles and their associated weights. This is the main principle of the sequential particle filtering.

When the whole sequence is available, i.e. all observations  $\bar{O}_{t_T} = \{O_{t_1}, O_{t_2}, \dots, O_{t_T}\}$  are available for all time  $t = \{t_1, \dots, t_T\}$ , a smoothing version of the previous technique can be applied by taking into account future observations. From the first estimation issued from the previous process, the idea consists in performing a backward exploration in order to correct the weights of the different particles. This is the so-called **forward-backward smoother**. The idea consists in reweighting the particles recursively backward in time, starting from the end time  $t_T$  to the initial one. It can be shown that rewriting the distribution  $p(S_{t_k} | \bar{O}_{t_T})$  in terms of backward transitions  $p(S_{t_k} | S_{t_{k+1}}, \bar{O}_{t_T})$  yields, after

several manipulations, to reweight all the particles with:

$$w^i(S_{t_k|t_T}) = w^i(S_{t_k}) \left[ \sum_{j=1}^N w^j(S_{t_{k+1}|t_T}) \frac{p(S_{t_{k+1}}^j | S_{t_k}^i)}{\sum_{\ell=1}^N w^\ell(S_{t_k}) p(S_{t_{k+1}}^j | S_{t_k}^\ell)} \right]. \quad (2.16)$$

Therefore, the process consists in first performing a sequential filtering and then, to reweight the particles backward in time. More details can be found in [Doucet 2001]. This is the process we suggest to use in this application to recover consistent LAI values from noisy observed ones based on the GeenLab model we derived. and it is described in the following.

### 2.3.4 Application to the recovery of LAI using GreenLab

Following equations (2.1-2.6), the set of parameters  $\{\Theta = r, k, \alpha, \eta, \epsilon\}$  has to be fixed in order to describe properly the LAI evolution of a plant. Parameters  $r, k$  are related to leaf mutual shading parameter and photosynthesis efficiency in Eq. (2.1),  $\alpha$  to natural leaf foliage in Eq. (2.6),  $\eta$  to the system noise in Eq. (2.9) and  $\epsilon$  to the observation variance in Eq. (2.10). We assume that the mean state of particles  $\{S_{t_k}^r | i = 1, 2, \dots, M\}$  at a given time  $t_k$  is a good estimation of M-length hidden state sequence:

$$S_{t_k}^r = \int S_{t_k} \cdot p(S_{t_k} | \bar{O}_{t_k}) dS_{t_k} = \sum_{i=1}^N S_{t_k} w^i(S_{t_k}) \quad (2.17)$$

where  $w^i(S_{t_k})$  is defined by Eq. (2.15) for particle filter, or by Eq. (2.16) for particle smoother. From above relation we can estimate the parameters  $\{r, k, \alpha\}$  of GreenLab model by minimizing the relation with gradient descent algorithm:

$$\{r, k, \alpha\} = \arg \min_{r, k, \alpha} \sum_{i=1}^M \|S_{t_k}^r - S_{t_k}^e\|^2 \quad (2.18)$$

with  $S_{t_k}^e$  the LAI evolution of GreenLab model, depending on  $\{r, k, \alpha\}$  (see illustration in Fig. 2.2) and  $S_{t_k}^r$  is computed by iteratively through Eq. (2.6) and Eq. (2.14-2.17).

As for the system noise  $\eta$  in Eq. (2.9), its distribution is assumed to be time independent zero-mean Gaussian. Its variance  $\sigma_\eta$  can then directly be computed with the relation:

$$\sigma_\eta = \frac{1}{M} \sum_{i=1}^M \|S_{t_k}^r - S_{t_k}^e\|^2. \quad (2.19)$$

The same holds for the estimation of  $\epsilon$ : noisy estimation  $S_{t_k}^o$  of Leaf Area Index being acquired from each image  $O_{t_k}$  by an inversion of PROSAIL model (as shown in Appendix), the uncertainty  $\sigma_\epsilon$  related to the observation system can be extracted through:

$$\sigma_\epsilon = \frac{1}{T} \sum_{i=1}^T \|S_{t_k}^r - S_{t_k}^o\|^2. \quad (2.20)$$

Concerning the initialization of the system, we estimate noisy LAI sequence  $\{S_{t_k}^o | k = 1, \dots, T\}$  from input MODIS images. The sequence  $\{S_{t_k}^r | k = 1, \dots, M\}$  is approximated with a polynomial regression from  $\{S_{t_k}^o\}$ . All these steps result in the following iterative algorithm:

**Incremental technique for LAI estimation and parameter inference**

- Initializations:
  - From MODIS images  $\{O_{t_k}\}$ , compute noisy estimations  $\{S_{t_k}^o\}$  of LAI as shown in Appendix;
  - Perform a polynomial regression on  $\{S_{t_k}^o\}$  in order to initialize  $\{S_{t_k}^r | k = 1, \dots, M\}$ ;
- 1. Estimate hidden parameters  $\{r, k, \alpha\}$  with Eq. (2.18)
- 2. Compute associated uncertainties  $\{\eta, \epsilon\}$  with Eq. (2.19-2.20)
- 3. Perform LAI inference with Eq. (2.14–2.17) through particle methods described in section 2.3.3
- 4. **Loop** to step (1) until the estimation of LAI converges
  - Once converged, the reconstructed Leaf Area Index sequence  $\{S_{t_k}^r | k = 1, \dots, M\}$  and associated hidden parameters  $\Theta = \{r, k, \alpha, \eta, \epsilon\}$  are available.

From this algorithm, we observe that the model and observation uncertainties  $\eta$  and  $\epsilon$  are automatically adapted during the process. A lowerbound as 20 percents of the summation of  $\eta, \epsilon$  is enforced to prevent the over-confidence on observation or model. Let us now turn to some experiments on real and simulated data.

## 2.4 Experimental Results

This part aims at validating the proposed LAI reconstruction technique. Our approach has been tested both on mastered and real data sets and in order to analyze the benefits of our method under various situations. Before entering into details, let us discuss about some general aspects of the experimental setup.

### 2.4.1 Experimental settings

The following experimental conditions have been applied:

- In each experiment, the number of particles has been set to 200;
- The efficiency of the reconstruction process presented in this paper has been compared with polynomial regression and wavelet filter. The order of the polynomial is chosen to be 5 via cross-validation. As for the wavelet procedure, we have removed the high frequency band coefficients and used the Daubechies wavelets with 5-level. These parameters have also been set by cross-validation;
- We also have indicated the results with fitted GreenLab model, GreenLab parameters  $(r, k, \alpha)$  are calibrated with the proposed procedure in Section 2.3.4. Then we could iteratively predict the LAI values without the

correction steps, as demonstrated in Fig. 2.2. this could be considered as a model-based regression;

Let us now present the results.

### 2.4.2 Data with ground truth

To evaluate our technique on mastered conditions, we have used available data in our lab. They are issued from real MODIS observations (a sequence of 17 images) on which LAI estimations, issued from the PROSAIL model, have been properly reconstructed and validated (see [Lecerf 2008, Lecerf 2005]). In addition, associated classifications of large agricultural parcels (we mean by large the fact that they recover more than one MODIS pixel) were available, providing us confident LAI time series and associated land cover. In practice three types of agricultural crops were used: colza, maize and wheat. Some LAI profiles are visible in Fig. 2.3. These data are considered as ground truth in our experiments.

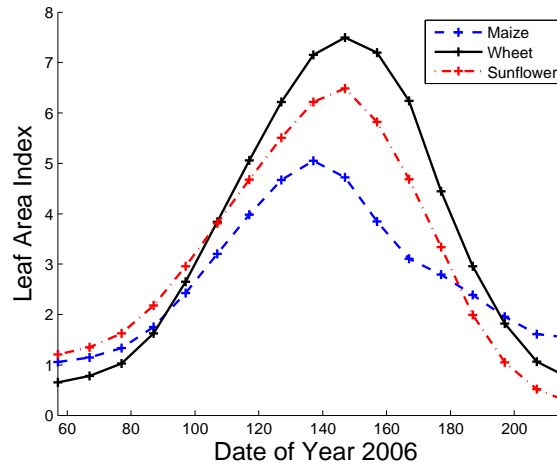


Figure 2.3: A demonstration of uncorrupted LAI time series for three sorts of agriculture crops in Brittany, France

In order to model the various artifacts, that are prone to occur in remote sensing images, several processes have been applied on these data:

- *Sequence of noisy observations.* Data are equally subsampled (one image every 10 days) from ground truth data and a Gaussian white noise has been introduced as shown in Fig. 2.4.
- *Sequence of sparse observations.* Here data at only 6 observations have been kept. The motivation is to evaluate the performance of our algorithm under a short length of remote sensing observation time series (this appear for example in very cloudy regions where few data can be exploited), as shown in Fig. 2.5.

- *Sequence of missing observations.* Like previously, very few observations were used (7 in practice) but here we simulate the absence of data during a long series. Therefore missing data are consecutive in time as shown in Fig. 2.6.

Experimental results are presented below.

#### 2.4.2.1 Result on sequence of noisy observations

This kind of data represents a great challenge in remote sensing analysis since all observations are in general corrupted by noise. Many methods rely on signal processing filters (spatially or temporally) [Rabiner 1975][Burrus 1998] to remove such noise. However this requires additional parameters and our method has the great advantage to get rid of this problem. In the inference process, the noises on the plant growth evolution as well as on observations  $\eta, \epsilon$  are indeed estimated along the procedure. This is to our opinion a great advantage compared to other denoising processes.

In order to make a quantitative evaluation, various levels of Gaussian zero-mean noise, with Signal-to-noise ratio(SNR) ranging from 2 to 10, were tested . The table 2.1 represents the root-mean-square error (RMSE) between reconstructed LAI (with particle filter and smoother and other reconstruction techniques introduced in section 2.4.1) and groundtruth data. It is interesting to observe in this table that the introduction of GreenLab plant growth model in the reconstruction process systematically reduces the error residual and outperforms all other methods. Particle smoother almost always gives a better performance than particle filter, which is quite rational as when estimating the LAI at any time point, particle filter only exploits information of up-to-date observations, while particle smoother would take into account all observations in time series. In order to have a visual inspection of the reconstructions, some illustrations are visible in Fig. 2.4 and confirm that our outputs are of reliable quality. Even if other curves are sometimes consistent with ground truth, particle filter and/or smoother represent the best compromise since strong errors are likely to appear with other techniques (as the polynomial one very sensible to little variations). Let us now turn to sequences with sparse observations.

Table 2.1: The Root-Mean-Square Error (RMSE) of Different Methods on Sequences of various SNR level.

RMSE	SNR=2	SNR=4	SNR=6	SNR=8	SNR=10
Observations	1.3222	0.8712	0.7747	0.6707	0.6313
Polynomial	0.7971	0.6120	0.5760	0.5351	0.5136
Wavelet filter	0.8018	0.7040	0.6795	0.6481	0.6365
Fitted model	<b>0.7063</b>	0.5749	0.5259	0.5189	0.4586
Particle filter	0.7755	0.6633	0.5973	0.5452	0.5403
Particle smoother	0.7072	<b>0.5660</b>	<b>0.5014</b>	<b>0.4837</b>	<b>0.4341</b>



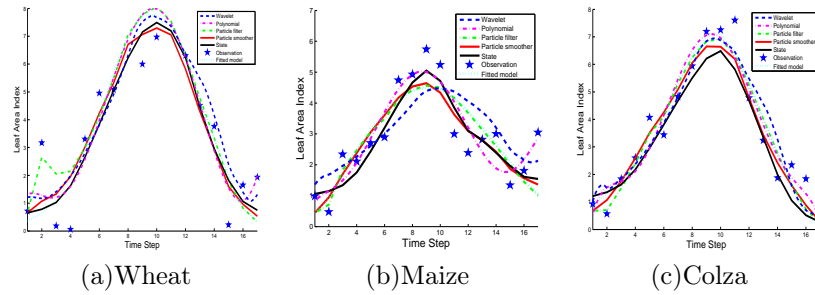


Figure 2.4: **LAI reconstruction with various techniques on the sequence of noisy observations (with SNR=6)**. Dark line are uncorrupted MODIS observations (ground truth state), blue dots are the input noisy observations, red line is the result of reconstruction with particle smoother, cyan line from particle filter, green line from fitted model, pink line from polynomial regression, blue line from wavelet filter. Even in this situation where the noise is large, the best fitted curves are issued from techniques based on the GreenLab model

#### 2.4.2.2 Sequence with sparse observations

As mentioned above, because of satellite orbit, weather condition, etc. it appears that observations are often very sparse in space or time. Therefore a robust remote sensing technique should be able to handle this common problem. In our dataset, we have tried to simulate this situation by removing a large number of data. In practice, we have randomly taken only six observations to reconstruct the whole time series. These observations are the blue dotted points in Fig. 2.5. As one can observe, they are more or less issued from a uniform sample along the time period. Quantitative results are depicted in table 2.2. and also prove, in this situation, the performance and the benefit of relying on GreenLab growth model compared with other techniques. Reconstructed curves are also visible in figure 2.5 and demonstrate the great ability of our technique to recover proper series. Let us now present the last experiment of this section: missing data.

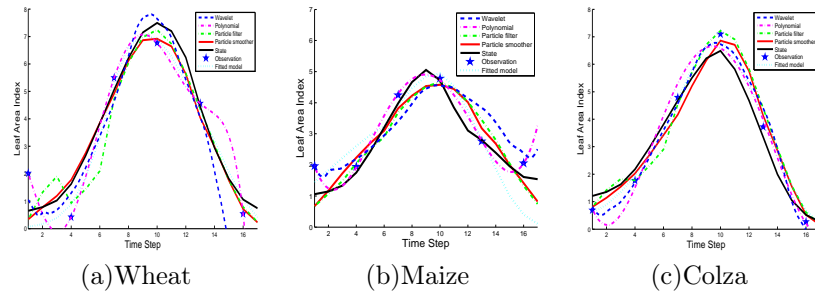


Figure 2.5: **LAI reconstruction with various techniques on the sequence of sparse observations (with SNR=6)**. Dark line are uncorrupted MODIS observations (ground truth state), blue dots are the input noisy observations, red line is the result of reconstruction with particle smoother, cyan line from particle filter, green line from fitted model, pink line from polynomial regression, blue line from wavelet filter.

Table 2.2: The Root-Mean-Square Error(RMSE) of Different Methods on Sequence with Sparse Observations.

RMSE	SNR=2	SNR=4	SNR=6	SNR=8	SNR=10
Observations	1.4425	1.2365	0.9455	0.7818	0.6887
Polynomial	1.4943	1.2852	0.1447	0.8807	0.7867
Wavelet filter	1.1932	1.1424	1.0390	0.7390	0.7229
Fitted model	1.0346	1.0425	0.8574	0.6767	0.7126
Particle filter	1.1762	1.1233	0.8523	0.7931	0.7549
Particle smoother	<b>1.0198</b>	<b>0.9622</b>	<b>0.9451</b>	<b>0.5923</b>	<b>0.5919</b>

### 2.4.2.3 Sequence of missing observations

To model the fact that sometimes there are no observations during a long period of time (unavailability of the satellite due to technical problems), we have removed long period of consecutive data. Some illustrations are visible in Fig. 2.6 where one can observe that no observations during the growing period are available, which is a very delicate problem. In spite of these difficulties, the proposed framework is able to provide very good reconstruction of LAI time series, as shown in Tab. 2.3 and in Fig. 2.6. Obviously here, only the dynamical model prior is able to estimate consistent values during the critical period. This is to our opinion a very interesting behavior of our estimators.

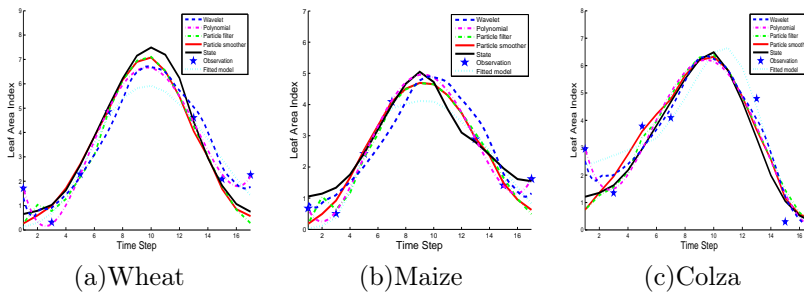


Figure 2.6: **LAI reconstruction with various techniques on the sequence of missing observations (with SNR=6)** . Dark line are uncorrupted MODIS observations (ground truth state), blue dots are the input noisy observations, red line is the result of reconstruction with particle smoother, cyan line from particle filter, green line from fitted model, pink line from polynomial regression, blue line from wavelet filter.

### 2.4.2.4 Overview of these experiments

In this three experiments, various reconstruction methods have been tested. In all situations, the use of GreenLab as prior plant growth model has enabled to improve the reconstruction of LAI over long time periods. This is to our opinion a very interesting behavior likely to have many potential applications. It should

Table 2.3: The Root-Mean-Square Error(RMSE) of Different Methods on Sequence of Missing Observations.

RMSE	SNR=2	SNR=4	SNR=6	SNR=8	SNR=10
Observations	1.2934	1.0424	0.9218	0.7011	0.7118
Polynomial	1.2606	1.1733	1.0649	1.1239	0.8877
Wavelet filter	1.3656	1.0531	0.8184	0.7741	0.7840
Fitted model	<b>1.2036</b>	1.0689	0.8079	0.7072	0.6894
Particle filter	1.2617	0.9596	0.8526	0.7232	0.7044
Particle smoother	1.2065	<b>1.0253</b>	<b>0.8055</b>	<b>0.7050</b>	<b>0.6892</b>

also be outlined that our technique not only reconstructs the LAI but is also able to extract deviations with respect to observations and dynamical model (parameters  $\{\eta, \epsilon\}$ ) and internal parameters of plants  $\{r, k, \alpha\}$  related to the photosynthesis efficiency, resistance of the plant and spoilage in natural growth (cf section 2.2). The estimation of these parameters is a very interesting point that may have interesting applications in the future. This will be illustrated in the next section.

This first series of experiments has quantitatively demonstrated the efficiency of our technique. Let us now present experiments on real MODIS data.

### 2.4.3 Application on large scale study site

In order to visually evaluate the efficiency of our technique, we have applied our methods on a large scale study site in Chizé, France. This study site is abundant with various sorts of vegetation, such as wood, grassland, maize, colza, wheat. In practice 16 images from March to October 2006 were available. On these images, the PROSAIL model has been applied to derive LAI time series. An illustration of available data is visible in the top of Fig. 2.9 where red stands for high values of LAI (indicating dense plant canopy) and blue for low LAI. However at some location, observation data are corrupted, mainly because of cloud coverage, on certain dates. Therefore LAI variables are missing in top of Fig. 2.9 are noisy and inconsistent in time.

In the bottom of figure 2.9, the results from particle smoother under our proposed framework are depicted. Obviously, when comparing (a) with (b), it appears clearly that our method is more in accordance with the expected ground truth (of course on this study site this latter is not available). To go into more details in the evaluation, we have therefore plotted in Fig. 2.7 some curves randomly taken from the images, where the reconstructed data seem to be consistent with plant growth knowledge. Once again this experiment demonstrate the efficiency of our proposed model.

We would like to point out here that such approach not only recover consistent LAI time series from GreenLab model and available satellite sequential observations, but also our method provides additional information of the state of plants, in particular the total biomass production and its distribution among different organs. To illustrate this, Fig. 2.8 shows the estimated biomass production and its partition for different organs at each continuous growth cycles

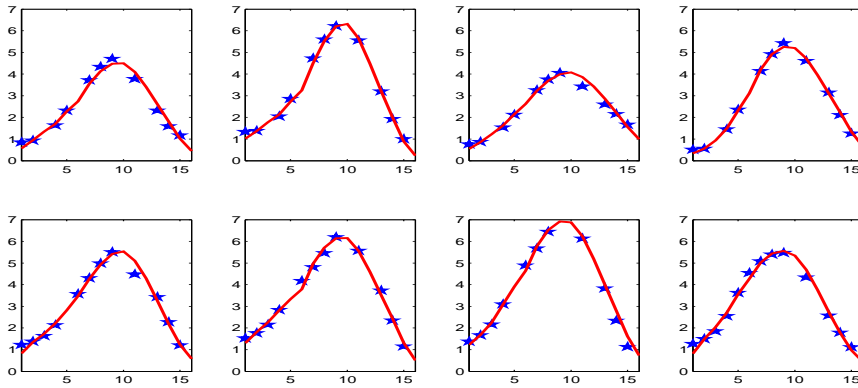


Figure 2.7: Several examples of LAI time series reconstruction with particle smoother at a study site in Chizé, France

of a unit wheat field issued from the real sequence from Fig. 2.4(a). This information is of prime importance for agriculture monitoring and is also provided by our technique. This is to our opinion a key point.

## 2.5 Conclusion

In this paper, we have proposed a technique that estimates leaf area index time series using data assimilation. The dynamical model is based on the GreenLab plant growth model where we have modified it to be applied in a remote sensing context. LAI observations are issued from the inversion of the PROSAIL model. As shown in our experiments, this technique is efficient in many various situations as very sparse observations or missing/noisy data. Quantitatively, it outperforms other reconstruction strategies that do not rely on a physical prior knowledge.

Another key point of our process is its ability to estimate internal parameters of the plant unobservable from images. This is to our opinion a very good property that may have other applications in the future.

## Appendix

### LAI estimation procedures

In practice satellite data are MODIS level 1B images time series acquired from March 2000 to April 2008. Images are firstly atmospherically corrected using the 6S model (Second Simulation of a Satellite Signal in the Solar Spectrum vector code) [Vermote 1997]. In a second step time series are screened out to remove the influence of cloud. This is implemented by a constant threshold on blue band (B0) to acquire a mask of cloud covered area [Lisens 2000].

Biophysical variables as LAI can be derived from satellite observations. Among the different techniques available, we are based in this study on inversions of the SAIL+PROSPECT radiative transfer model [Baret 2007, Jacquemoud 2009,

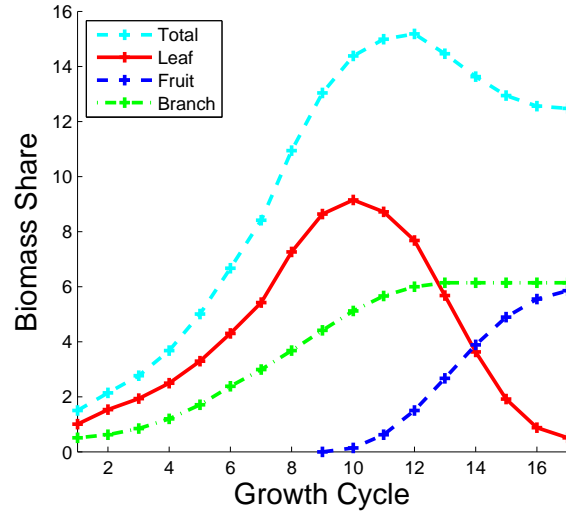


Figure 2.8: **Estimated biomass production and its partition during continuous growth cycles of unit wheat field.**

[Lecerf 2008]. The SAIL model deals with light scattering by leaf layers with application to canopy reflectance model and PROSPECT is about leaf optical properties spectra. A neural network is used to estimate biophysical variable, LAI, from each image. In order to calibrate the neural network, top of canopy reflectance for near nadir viewing angle are simulated by using PROSAIL model with priori defined distribution of soil reflectance, canopy characteristics and sun position, as detailed [Baret 2007]. The simulated dataset is used as the training dataset for neural network to inverse the PROSAIL model. It has been proved in [Lecerf 2008, Weiss 2007] that this approach performs efficiently for low, medium, high and very high-resolution data and is therefore adapted to the variability of satellite images available.

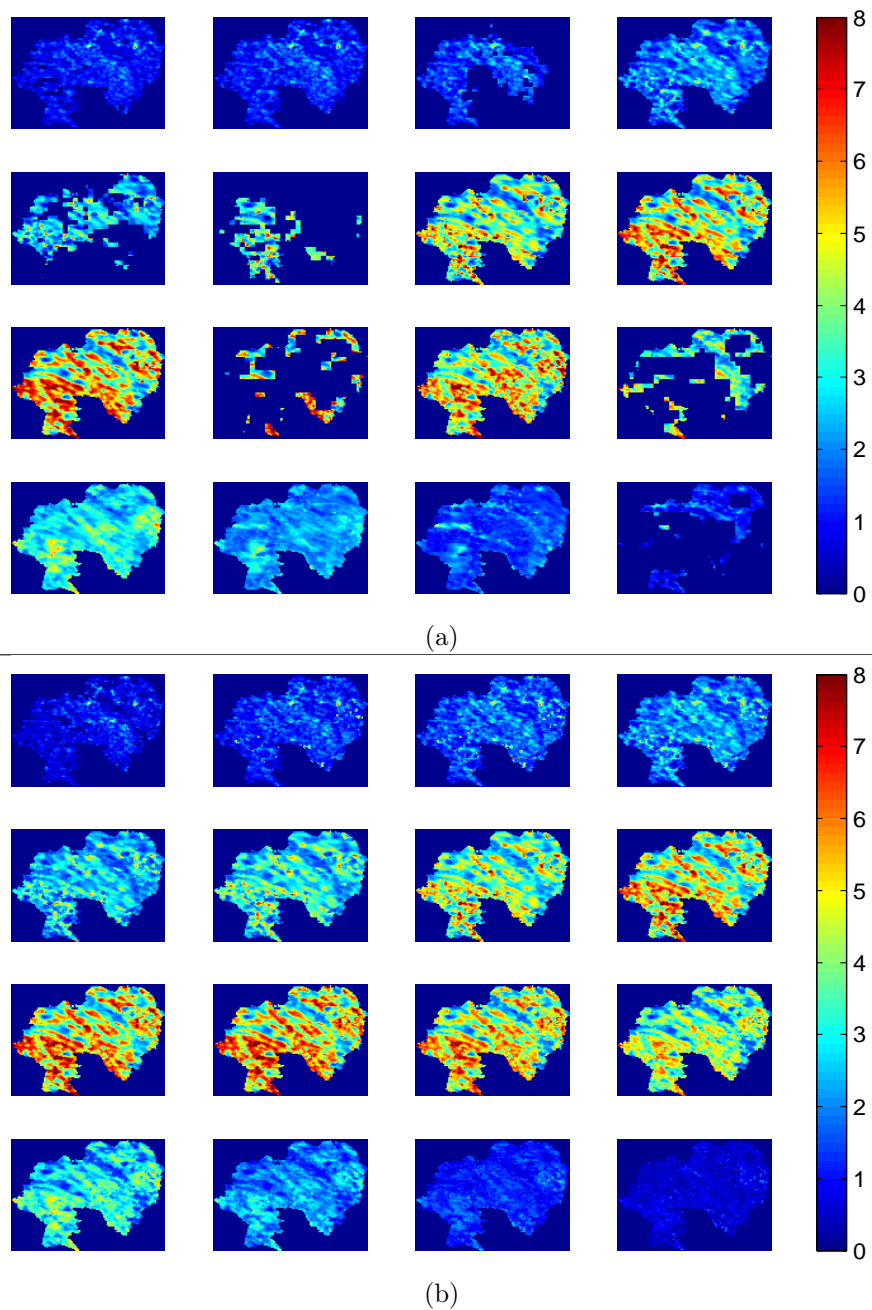


Figure 2.9: **Result of particle smoother at a study site in Chizé, France.** Top : input LAI values computed from MODIS data ; Bottom : reconstructed data



## Part II

# Démélangeage non-linéaire : aspects théoriques





# Non-negative Matrix Factorization and application to unmixing

---



---

## Résumé du chapitre

Dans ce chapitre, on s'intéresse au démixage lorsque les données ne sont pas linéairement composées. Pour cela, nous exploitons des techniques de factorisation non-négative de matrices. L'ensemble des données dont on dispose est représenté par une matrice  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]_{(\mathbf{x}_i \in \mathbb{R}^p)}$  de dimension  $p \times n$  qui contient les  $n$  échantillons de dimension  $p$  ( $\in \mathbb{R}^p$ ). La factorisation non-négative de matrice s'appuie sur le même principe que l'Analyse en Composante Principale (ACP) dans le sens où on cherche une autre forme de représentation des données, de préférence dans un espace de dimension réduit. Cependant, la base de décomposition est estimée de manière différente.

Pour rappel, une ACP va chercher à transformer l'information pour ne représenter les  $n$  variables que sur  $p'$  dimensions ( $p' < p$ ), au moyen d'une relation du type :

$$\mathbf{X} = V_{p'} \mathbf{X}'^T$$

où  $V_{p'}$  est la matrice de changement de base, de dimension  $p \times p'$  et composée des  $p'$  vecteurs propres associés aux plus grandes valeurs propres de la matrice de variance-covariance de  $\mathbf{X}$ . La matrice  $\mathbf{X}'$  est de dimension  $n \times p'$  et représente les coordonnées des  $n$  échantillons dans la base  $V_{p'}$ .

Dans le cas de la factorisation non négative, l'idée de la factorisation de matrices consiste à représenter  $\mathbf{X}$  par le produit suivant:

$$\mathbf{X} = \mathbf{F}\mathbf{G}^T$$

où  $\mathbf{F}$  et  $\mathbf{G}$  sont deux matrices telles que les colonnes de  $\mathbf{F} \in \mathbb{R}^{p \times \ell}$  soient composées d'éléments de  $\mathbf{X}$  et que  $\mathbf{G} \in \mathbb{R}^{n \times \ell}$  soit positive ou nulle, chaque colonne sommant à l'unité. Ainsi, les nouvelles coordonnées sont exprimées comme une combinaison linéaire de certains éléments de l'ensemble initial. S'ils sont judicieusement choisis, ces éléments peuvent être des *endmembers* et la matrice  $\mathbf{G}$  contient ainsi l'*abondance* de chaque entrée.

En pratique, pour gérer la non-linéarité des données, nous effectuons ces opérations dans un espace transformé des données (engendré par le RKHS : Reproducing Kernel Hilbert Space, qui autorise une meilleure séparabilité des données) et les end-members sont choisis comme étant ceux qui maximisent le

volume engendré par les données dans le RKHS, comme illustré sur la figure 3.1(a). Comme nous le verrons, nous choisissons un ensemble d’endmembers “sur-complet” (*overcomplete*), c’est-à-dire qui comprend plus de données que d’éléments purs, pour traiter efficacement la diversité des valeurs associées à une classe.

Les valeurs d’abondance sont quant à elles obtenues en minimisant  $\|\mathbf{X} - \mathbf{F}\mathbf{G}^T\|^2$  sous la contrainte que les éléments de  $\mathbf{G}$  sont positifs ou nuls et que les colonnes somment à 1. Ainsi, comme nous le verrons, chaque élément de  $\mathbf{X}$  dans le RKHS appartient à un simplexe. Ceci est illustré en figure 3.1(b).

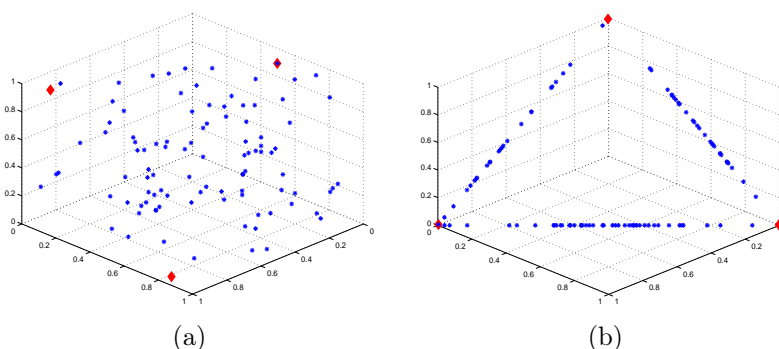


Figure 3.1: **Illustration des approches qui seront présentées.** (a) extraction des endmembers (en rouge) par maximisation du volume et (b) estimation des abondances.

Ce chapitre est organisé comme suit:

- La **section 3.1** introduit le contexte de la recherche en factorisation non-négative;
- La **section 3.2** présente l’approche de maximisation du volume du simplexe (SiVM) et notamment dans le RKHS;
- La **section 3.3** se focalise sur l’estimation de l’abondance ;
- La **section 3.4** résume la procédure complète pour le démixage;
- La **section 3.5** présente des résultats expérimentaux.

Ce chapitre correspond à un article publié dans la revue *Machine Learning*. Le travail a en grande partie été réalisé par Prof. Nicolas Courty avec qui j’ai beaucoup interagi pour la construction de la méthode (je suis deuxième auteur du papier). Ensuite, comme nous le verrons dans le chapitre 4, je me suis focalisé sur l’application de la méthode développée ici au cas spécifique des séries temporelles.

## Abstract

This chapter presents a new non-negative matrix factorization (NMF) technique which (1) allows the decomposition of the original data on multiple latent factors accounting for the geometrical structure of the manifold embedding the data; (2) provides an optimal representation with a controllable level of sparsity; (3) has an overall linear complexity allowing handling in tractable time large and high dimensional datasets. It operates by coding the data with respect to local neighbors with non-linear weights. This locality is obtained as a consequence of the simultaneous sparsity and convexity constraints. Our method is demonstrated over several experiments, including a feature extraction and classification task, where it achieves better performances than the state-of-the-art factorization methods, with a shorter computational time.

The non-negative matrix factorization technique, described in this chapter, is performed in reproducing kernel Hilbert space (RKHS), and the various definitions of kernel function offer a flexibility to measure the similarity between data samples. In Chapter 4, we will exploit this kernel-space NMF method and elastic kernels to perform pixel unmixing based on time series data.

## Keywords

Non-negative matrix factorization, manifold hull, simplex, kernel

## 3.1 Introduction

### 3.1.1 Context

Non-negative matrix factorization (or NMF for short) has long been studied and used as a powerful data analysis tool providing a basis for numerous processing, such as dimensionality reduction, clustering, denoising, unmixing, etc. This article is concerned with a variant of NMF, where one tries to decompose a given matrix  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]_{(\mathbf{x}_i \in \mathbb{R}^p)}$  formed by a set of  $n$  samples described with  $p$  variables ( $\in \mathbb{R}^p$ ) as a product of two matrices  $\mathbf{F} \in \mathbb{R}^{p \times \ell}$  and  $\mathbf{G} \in \mathbb{R}^{n \times \ell}$  such that  $\|\mathbf{X} - \mathbf{F}\mathbf{G}^\top\|^2$  is minimized, under the constraints that the elements of  $\mathbf{G}$  are positive or nil, and that its rows sum to one. In other words,  $\mathbf{G}$  serves as a convex surrogate description of  $\mathbf{X}$  in the reduced embedding formed by  $\ell$  prototypes (the columns of  $\mathbf{F}$ ).

This convexity property on  $\mathbf{G}$  is desirable in a lot of applications for its relations with the physics of the underlying observations: in this acceptance, the datapoints are usually observations of a mixing process where the components of the mixture are not known. Hyperspectral images are a good example, as each pixel describes a spectrum that can be defined as a combination of pure materials spectra (trees, concrete, water, etc.); the combination occurring because of the captor spatial resolution and different scattering effects [Esser 2012, Bioucas-Dias 2012]. Other examples of applications are found in archetypal analysis [Mørup 2012], or biology [Kersting 2012].

### 3.1.2 Notations

Bold capital letters, such as  $\mathbf{X}$  or  $\mathbf{M}$  refer to matrices. The transposition operator is denoted  $\cdot^\top$  (so that  $\mathbf{M}^\top$  refers to the transpose of  $\mathbf{M}$ ). The set of  $m \times n$  real matrices (respectively real matrices with nonnegative entries) is denoted  $\mathbb{R}^{m \times n}$  (respectively  $\mathbb{R}_+^{m \times n}$ ).  $\mathbf{I}_n$  is the  $n \times n$  identity matrix, and  $\|\cdot\|$  refers to the Frobenius norm. For any matrix  $\mathbf{M}$ ,  $\mathbf{M}_{\bullet i}$  (respectively  $\mathbf{M}_{i\bullet}$ ) corresponds to the  $i$ th column (respectively  $i$ th row) of  $\mathbf{M}$ . However, in the particular case of  $\mathbf{X}$ , each column  $\mathbf{X}_{\bullet i}$  corresponds to the datum  $\mathbf{x}_i$ , so that this latter more intuitive notation is preferred.  $\Delta^n(\mathbf{M})$  is a  $(n-1)$ -simplicial polytope formed by  $n$  columns of  $\mathbf{M}$ . The unit  $(n-1)$ -simplex  $\Delta^n(\mathbf{I}_n)$  is simply referred to as  $\Delta^n$ .

### 3.1.3 Related work

Despite appealing properties, NMF presents a number of difficulties: First, while real data are often embedded on complex manifolds, the seminal convex NMF formulation prohibits a large number of latent factors that would explain the data in a geometry preserving way; Second, the level of sparsity in the final embedding (the matrix  $\mathbf{G}$ ) is rarely controllable, and far from data analyst expectations; finally, the computational time is prohibitive when it comes to handling very large size matrices. Even if to date, various state-of-the-art methods already tackle one of or two of these issues, no method addresses these three issues all together.

The convexity constraint we consider on  $\mathbf{G}$  differs from that of what is classically referred to as Convex NMF [Ding 2010], i.e. NMF where the columns of  $\mathbf{F}$  are convex combinations of columns of  $\mathbf{X}$ . However, our problem is also not unheard of in the literature, as, beyond its easiness of interpretation, it provides a very appealing computational framework: if  $\mathbf{G}$  encodes convex combinations, the  $\ell$  columns of  $\mathbf{F}$  are expected to characterize a hull of  $\mathbf{X}$  in  $\mathbb{R}^p$ , as data points lying within are perfectly reconstructed. Based on this idea, it is possible to separate the NMF problem into two steps, namely the computation of  $\mathbf{F}$  (in order to determine, or to best approximate the hull of the dataset) and of  $\mathbf{G}$  (which corresponds to the projection of the dataset onto the region bound by the hull). From a computational point of view, this is really interesting, as it is now possible to decompose the NMF (which amounts to finding two matrices with close relationship) into two simple problems, each focused on a particular matrix: first one computes  $\mathbf{F}$  regardless  $\mathbf{G}$ , and second, one computes  $\mathbf{G}$  by projection of the columns of  $\mathbf{X}$  onto  $\mathbf{F}$ . With such an approach, it is possible to expect a lower complexity than that of elder methods based on singular value decomposition (with a  $o(n^3)$  complexity), or than that of demanding procedures that iteratively alternates between the minimization of the two matrices, such as in [Ding 2010].

Initially,  $\mathbf{F}$  was first related to the *convex hull of  $\mathbf{X}$*  [Cutler 1994]: Notably, numerous works in the remote sensing and hyperspectral imaging community (see [Bioucas-Dias 2012] as well as its references) have pointed out that  $\mathbf{F}$  should be chosen so that  $\Delta^\ell(\mathbf{F})$  best encompasses  $\mathbf{X}$ , while other investigations focused on how to find this simplicial convex hull [Wang 2010, Thureau 2010, Civril 2009]. More recently, a series of works [Arora 2012, Kumar 2013, Gillis 2013, Recht 2012] focused on the *conical hull of  $\mathbf{X}$* : Here, the idea is to find  $\mathbf{F}$  so that

it spans a cone encompassing the dataset. The reason of the recent focus on the conical hull is that it is the geometric translation of an important assumption in NMF, the *separability assumptions* proposed by Donoho in [Donoho 2003]. This separability assumption reads that (1) the residue  $\mathbf{X} - \mathbf{F}\mathbf{G}^\top$  is nil, (2) columns of  $\mathbf{F}$  are collinear to columns of  $\mathbf{X}$ . This second condition regarding the collinearity of the columns of  $\mathbf{F}$  and  $\mathbf{X}$  is interesting beyond the separability assumption: whatever the type of hull defined by  $\mathbf{F}$ , it is possible to assume that the components of the mixture belong to the data, and that any datum is a convex combination of a restricted number of particular selected datapoints (those forming  $\mathbf{F}$ ). This approach drastically reduces the complexity of finding  $\mathbf{F}$  as this latter is simply defined by a *column subset selection* (CSS) of  $\mathbf{X}$ .

Fig. 3.2 illustrates a dataset lying on a non-linear manifold resulting from some hidden factors that could be of interest, as well as its conical and convex hulls. It clearly appears that the geometries of these hulls are not adapted to that of the dataset. So far, there has been little interest in trying to characterize the boundary of the manifold dataset in spite of its non-convexity; in the sequel, we shall address this boundary with the shorter and imaged name *manifold hull*, that is illustrated on Fig. 3.2(d). Naturally, precisely characterizing such a manifold hull would require to increase  $\ell$ , the number of datapoints involved in the CSS; which stresses to a larger extent the need for an adapted control of the sparsity level that is already sought for in numerous applications: Each point should be described as a convex combination of a restricted number  $\lambda$  of prototype points among the  $\ell$  of the CSS.

Whatever the type of hull (convex or conical), separating  $\mathbf{F}$  and  $\mathbf{G}$  computations has drastically reduced the overall NMF complexity of the state-of-the-art methods: [Thureau 2010, Thureau 2012] compute  $\mathbf{F}$  in linear time and few works [Gillis 2013, Recht 2012] even reach overall linear-complex NMF. However, among them, none allows characterizing the manifold hull, and none allows controlling the solution sparsity. Even if to date, more computationally efficient methods than adding the classical  $L^1$  penalty [Hoyer 2004, Esser 2012] have been developed, such as [Kim 2007, Gillis 2012], none reaches a linear complexity.

Even if the characterization of the manifold hull has never been addressed so far, it is well-known that the kernel trick is an efficient way to provide a manifold preserving description of a dataset. In fact, several pre-existing works applied the kernel trick to NMF [Zhang 2006, Buciu 2008, Cai 2011]. However, the convexity constraint on  $\mathbf{G}$  (and consequently the notion of manifold hull) does not appear. Moreover, neither sparsity nor any linear complexity is achieved.

### 3.1.4 Proposal

In this article, we present SAGA (*Sparse And Geometry-Aware*), a new NMF method avoiding the aforementioned limits. It operates in a Reproducing Kernel Hilbert Space (RKHS) [Schölkopf 2002], where  $\mathbf{F}$  is defined by a CSS. Then, according to an expected sparsity level,  $\mathbf{G}$ , the best dataset projection onto the simplex formed by  $\mathbf{F}$  is computed. The advantages of our method are:

*First*, kernelization is interesting to several extents: *i)* it makes the algorithm compliant with dataset where only relationships among objects are available; *ii)* its regularization property improves robustness to noise; *iii)* it allows using more latent factors than the dimensions of the input data, providing an insightful

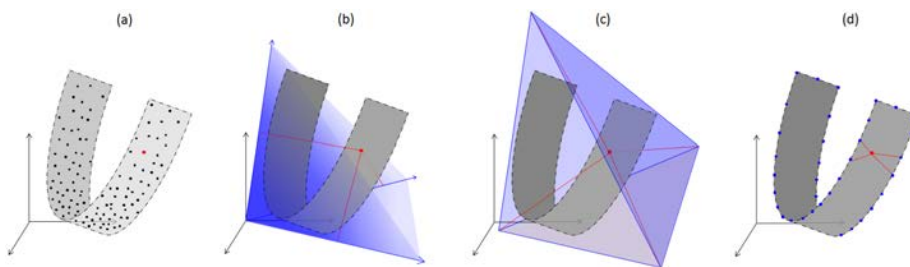


Figure 3.2: (a) A dataset  $\mathbf{X}$  embedded in a U-shaped manifold; (b) the separability assumption assumes the dataset is encompassed in a cone spanned by the columns of  $\mathbf{F}$ ; a datum is described in a column of  $\mathbf{G}$  through its non-negative coordinates in the cone; (c) Convex hull: the dataset is embedded in a simplex, and each datum is described with barycentric coordinates; (d) Manifold hull: one uses a great enough number of reference points to precisely characterize the manifold, while forcing the sparsity to achieve some local coding. Note that, in order to make these imaged representations clearer, we did not represent  $\mathbf{F}$  as a CSS of  $\mathbf{X}$ .

tool to consider the geometric structure of the data manifold. As both the number of latent factors and the sparsity level are increasing, it appears that the locality of the support of  $\mathbf{G}$  is increasing, turning the factorization problem in a *non-linear local embedding* of the original data. As shown in some recent works [Yu 2009, Guillemot 2012], this kind of embedding is powerful to describe the non-linear structure of the data manifold, and as such serve as a very good feature extraction framework.

*Second*, the CSS is defined thanks to a manifold subsampling method [Shroff 2011], which reaches a linear complexity with respect to  $n$ , the size of the dataset.

*Third*, the computation of  $\mathbf{G}$  corresponds to a sparse RKHS simplex projection, which is a non-linear optimization problem. Based on recent advances on sparse projected gradient methods, the projection is solved with an algorithm of linear complexity (with respect to  $n$ ), while naturally embedding sparsity mechanism control.

*Forth* and finally, since both computations of  $\mathbf{F}$  and  $\mathbf{G}$  are linear, the overall complexity of SAGA is linear. This makes this algorithm perfectly suitable for very big data.

To the best of our knowledge, no state-of-the-art method simultaneously compiles all these advantages.

### 3.1.5 Contributions

SAGA has been developed on the top of several state-of-the-art algorithms, thanks to several technical extensions that are listed here. The computation of matrix  $\mathbf{F}$  is largely inspired by the simplex volume maximization (SiVM) approach of [Thureau 2012]. However, the method itself has been extended in several manners:

1. It is slightly generalized in order to operate in the Hilbert space reproducing the Gaussian kernel, rather than in  $\mathbb{R}^P$ .

2. Most importantly, while SiVM proposes an approximate optimization (for several simplifications are made to reach a linear complexity), we propose another solution to the simplex volume maximization which performs an exact optimization.

In spite of these two extensions, our method remains of linear complexity, so that finally, the computation of  $\mathbf{F}$  with SiVM is both a particular case and an approximation of the one produced by SAGA. Then, the computation of matrix  $\mathbf{G}$  is not inspired by any other NMF technique, but is based on a recent sparse projected gradient strategy [Kyriilidis 2013]:

3. This latter is adapted to the projection over the CSS in the RKHS, and its linear complexity is kept.
4. We provide with theoretical bounds on the convergence of the projector (linked to the kernel bandwidth, the minimum pairwise distance between the CSS and the sparsity level).

### 3.1.6 Outline

We solve our NMF problem by separating the computations of the CSS matrix and of the projection matrix. This is why, Sections 3.2 and 3.3 focus on the computations of  $\mathbf{F}$  and  $\mathbf{G}$  respectively. More specifically, the structure of Sections 3.2 is the following: First a small introductory paragraph recalls the basics of the kernel trick, as well as why it is interesting to operate in a RKHS to fulfill our objectives. Then, Section 3.2.1 investigates the consequences of working in such RKHS; They lead us to a particular strategy, which is to characterize the manifold hull of the dataset in  $\mathbb{R}^p$  via a simplicial convex hull in the RKHS; and to define this latter with a manifold sampling strategy. At this point, computational constraints direct us toward SiVM-like procedures rather than toward more resource-demanding ones. Section 3.2.2 jointly presents SiVM, such as defined in the literature [Thureau 2012], as well as the kernel generalization we propose, while Section 3.2.3 describes our modifications to reach exact maximal simplex volume. In a similar way, Section 3.3 is divided into two parts: the first one (Section 3.3.1) explains how projecting the image in the RKHS of any datum  $\mathbf{x}_i$  onto the image of the CSS in the RKHS, along with sparsity constraints; Section 3.3.2 provides with a proof that the projector defined in Section 3.3.1 converges. As it is established in the literature [Garg 2009] that the Restricted Isometry Property (RIP, [Candes 2008]) implies the convergence, we prove that our projector respect the RIP for some particular tuning of its parameters. At this stage, most of the required mathematics is established. Then, in Section 3.4, one summarizes the entire procedure made of the concatenation of the two matrices computation through an easy to implement algorithm, completed by some theoretical assessments of the linear complexity of the algorithm. Finally, Section 3.5 is devoted to experimental validations. In the first series of experiments, one focuses on toy examples, in order to illustrate the behavior of the manifold hull. Then, follow several experiments on simulated datasets, in order to compare the computational efficiency and the computational precision of SAGA with respect to the state-of-the-art. Finally, we consider real datasets through challenging image classification tasks.



## 3.2 Geometry aware CSS procedure

Despite living in  $\mathbb{R}^p$ ,  $\mathbf{X}$  may span a nonlinear manifold of intrinsic dimensionality lower than  $p$ . A major problem is thus to extend to this nonlinear manifold the classical statistics that are used to work in a vector space. To do so, an essential element is to replace the Euclidean distances by geodesic distances. Depending on the manifold, the associated metric may be difficult to formally define. However, it is possible to characterize it through the time-scale of the well-studied heat diffusion process, the kernel formulation of which is well approximated by the Gaussian kernel [Lafferty 2005, Lafon 2006]: The geometry of the manifold in which  $\mathbf{X}$  lies is captured by  $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)$  with variance  $\sigma^2$ . Let us call  $\phi(\cdot)$  the implicit feature map from  $\mathbb{R}^p$  onto  $\mathcal{H}$ , the RKHS associated to  $k(\cdot, \cdot)$ . We use the shorthand notation  $\Phi = \phi(\mathbf{X}) = [\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_n)]_{(\phi(\mathbf{x}_i) \in \mathcal{H})}$ . Then, following the notations of [Ding 2010], the SAGA solution amounts to finding the indicator matrix  $\mathbf{W}^{(\ell)} \in \{0, 1\}^{n \times \ell}$  defining the CSS<sup>1</sup> and the projection matrix  $\mathbf{G}$  where  $\phi(\mathbf{F}) = \phi(\mathbf{X}\mathbf{W}^{(\ell)}) = \Phi\mathbf{W}^{(\ell)}$ , such that  $\|\Phi - \Phi\mathbf{W}^{(\ell)}\mathbf{G}^\top\|^2$  is minimized under convexity constraints.

### 3.2.1 CSS as a manifold sampling procedure

Let us start by a basic remark,

*Remark 1.* In the input space, we must have  $\ell \leq n$  and  $\ell \leq p$ . In the Gaussian RKHS, one still has  $\ell \leq n$ , however,  $\ell > p$  becomes possible, for each sample spans its own dimension.

which leads to the following property:

*Property 1* (Non-Separability). Separability assumption does not hold for NMF in the Gaussian RKHS.

*Proof.* According to Remark 1, any datum not included in  $\mathbf{X}\mathbf{W}$  cannot be expressed as a linear combination of elements of  $\phi(\mathbf{X}\mathbf{W})$ .  $\square$

Thus, one should not consider conical hull in the RKHS. However, Remark 1 leads to:

**Corollary 1.** *In the Gaussian RKHS, it is possible to use more than  $p$  points in the CSS. The latter forms a non-simplicial polytope in  $\mathbb{R}^p$  while their image in the Gaussian RKHS is a simplex.*

In other words, the manifold hull of  $\mathbf{X}$  can be characterized through the simplicial convex hull  $\Delta^\ell(\phi(\mathbf{F}))$ . Then, it follows that:

**Corollary 2.** *No sample lies in  $\Delta^\ell(\phi(\mathbf{F}))$  and all the samples will be projected on hyperfaces of  $\Delta^\ell(\phi(\mathbf{F}))$ , leading to approximate reconstructions. Yet, such an approximation comes with the appealing property of sparsity, discussed later in the article.*

At this point, finding  $\Delta^\ell(\phi(\mathbf{F}))$  amounts to finding  $\mathbf{W}$ , which turns out to subsample the boundary of the data manifold. Most of the methods from the

<sup>1</sup>We write  $\mathbf{W}$  instead of  $\mathbf{W}^{(\ell)}$  when  $\ell$  does not matter, or is implicit regarding the context.

literature address it with the objective of maximizing the representation of the dataset, while here, we are interested in its boundary, which makes the sampling completely different: For instance, a kernel  $k$ -means sampling, although very efficient to subsample a given manifold [Lafon 2006], leads to Convex NMF of [Ding 2010], the aim of which is completely different of ours.

However, our problem is not completely unheard of: In [Shroff 2011], the authors consider maximizing the diversity of the selected samples using a Karcher variance<sup>2</sup> maximization criterion. Alternative formulation exists, where one seeks for the maximum volume parallelepiped in the data matrix [Çivril 2009]. Interestingly enough, whatever the interpretation (diversity or volume criterion), the corresponding computation can reduce to recursive QR decompositions of the data matrix [Gillis 2013, Shroff 2011]. However, in the RKHS, since no explicit coordinates are available, those methods cannot be transposed. The recent proposal of [Courty 2013] regarding a kernel rank revealing Cholesky technique is also of interest, unfortunately, it fails to scale up to big data, because it implies a computationally demanding decomposition at each selection step.

Finally, among all the methods available in the literature, if one discards those (1) which do not sample the boundary, (2) which cannot be conducted in a RKHS, (3) which do not have a linear complexity, we are aware of a single remaining method: the Simplex Volume Maximization (SiVM) [Thureau 2012].

### 3.2.2 Original Simplex Volume Maximization

We begin with a review of the original formulation of Thureau and co-workers [Thureau 2012] and its direct transposition to the kernel framework. SiVM tries to maximize  $\text{Vol}(\Delta^\ell(\mathbf{F}))$ , the volume of the simplex spanned by  $\Delta^\ell(\mathbf{F})$ , which reads:

$$\text{Vol}(\Delta^\ell(\mathbf{F})) = \sqrt{\frac{-1^\ell \cdot \text{cmd}(\mathbf{F})}{2^{\ell-1}(\ell-1)!}}, \text{ with } \text{cmd}(\mathbf{F}) = \det \begin{pmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 1 & 0 & d_{1,2}^2 & d_{1,3}^2 & \dots & d_{1,\ell}^2 \\ 1 & d_{2,1}^2 & 0 & d_{2,3}^2 & \dots & d_{2,\ell}^2 \\ 1 & d_{3,1}^2 & d_{3,2}^2 & 0 & \dots & d_{3,\ell}^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & d_{\ell,1}^2 & d_{\ell,2}^2 & d_{\ell,3}^2 & \dots & 0 \end{pmatrix}. \quad (3.1)$$

$\text{cmd}(\mathbf{F})$  is the Cayley-Menger determinant of  $\mathbf{F}$  (i.e. the determinant of the matrix accounting for the pairwise distances in  $\mathbf{F}$ ) and  $d_{i,j}^2$  is the square Euclidean distance between elements  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . This obviously transposes to the RKHS: The volume of  $\Delta^\ell(\Phi\mathbf{W})$  can be expressed by replacing the Cayley-Menger determinant by that of a matrix accounting for pairwise distances in  $\mathcal{H}$ :

$$\text{Vol}(\Delta^\ell(\Phi\mathbf{W})) = \sqrt{\frac{-1^\ell}{2^{\ell-1}(\ell-1)!} \det(\mathbf{A})} \quad (3.2)$$

with  $\mathbf{A}$  similar to the matrix of Eq. (3.1) except that  $\forall i, j \leq \ell$ ,  $d_{i,j}^2$  is replaced by:

$$\mathbf{A}_{(i+1)(j+1)} = \|\phi(\mathbf{x}_i) - \phi(\mathbf{x}_j)\|^2 = k(\mathbf{x}_i, \mathbf{x}_i) + k(\mathbf{x}_j, \mathbf{x}_j) - 2k(\mathbf{x}_i, \mathbf{x}_j)$$

The comparison of the volume of the simplices spanned by all the possible CSS is computationally prohibitive. Thus, an approximate search in linear time

<sup>2</sup>The Karcher variance is a variance accounting for Riemannian distances over the manifold, rather than Euclidean ones.

of the best simplex is proposed in [Thureau 2012]. This is possible thanks to two tricks: The first one is a strong result (Theorem 1 of [Thureau 2012]), which states that if one adds an element to the simplex, it can only make the reconstruction error smaller or equal. Thus, a simple greedy algorithm can be used to compute  $\mathbf{W}$ : Starting with the best 1-simplex (*i.e.*  $\mathbf{W}$  has 2 columns), one adds the best third sample (according to the maximization of Eq. (3.1)), then the fourth, and so on until  $\ell$  samples are selected. Practically, this procedure can be transposed to the RKHS: At iteration  $p - 1$  one selects the element  $\phi(\mathbf{x}_i)$  of  $\Phi$  such that

$$i = \arg \max_q \text{Vol}(\Delta^p(\Phi \mathbf{W}) \cup \phi(\mathbf{x}_q)). \quad (3.3)$$

However, this procedure still requires the computation of several Cayley-Menger determinants (each of them being computationally intensive).

At this point shows up the second trick: If one makes the simplifying assumption that the distances between the elements of the CSS defined in the previous iteration are constant and noted  $a$ , and if  $\alpha_{j,q} = d_{j,q}^2/2$ , then, Eq. 3.3 amounts to finding  $\phi(\mathbf{x}_i)$  such that

$$i = \arg \max_q \left[ \sum_{k=1}^p \alpha_{kq} \cdot \left( a^2 + 2 \sum_{j=k+1}^p \alpha_{jq} \right) - (p-1) \cdot \sum_{k=1}^p \alpha_{kq}^2 \right], \quad (3.4)$$

where  $d_{j,q}$  refers to the distance between a point of the CSS  $\mathbf{x}_j$  and a point out of the CSS  $\mathbf{x}_q$  which is considered for adjunction to the CSS (thus, in the RKHS, one has  $\alpha_{j,q} = 1 - k(\mathbf{x}_j, \mathbf{x}_q)$ ). Finally, the computation of Eq. 3.4 is sped up by considering that  $d_{i,q}^2 \approx d_{i,q}$  and  $a^2 \approx a$ , leading to

$$i = \arg \max_q \left[ \sum_{k=1}^p d_{kq} \cdot \left( a + \sum_{j=k+1}^p d_{jq} \right) - \frac{p-1}{2} \cdot \sum_{k=1}^p d_{kq}^2 \right]. \quad (3.5)$$

*Remark 2.* Naturally, the magnitude to which the constant distance assumption is violated strongly depends on the dataset. As a consequence, the instantiation of Eq. 3.3 into Eq. 3.5 may lead to some approximations as already noted in [Gillis 2013], where Gillis and Vavasis showed that SiVM can underperform on ill-conditioned datasets.

Despite this remark, the general principle advocated in Eq. 3.3 remains unalterably valid. This is why, we rely on it to propose an alternative to SiVM, which provides exact volume computation, with a similar linear complexity.

### 3.2.3 Exact simplex volume maximization

Let us consider a simplex  $\Delta^p$  of dimensionality  $p - 1$  spanned by a subset  $\mathbf{XW}^{(p)} = \{\mathbf{x}_1, \dots, \mathbf{x}_p\}$  of  $p$  points of  $\mathbf{X}$ . If we add a  $(p + 1)$ th point  $\mathbf{x}_i \in \mathbf{X} \setminus \mathbf{XW}^{(p)}$  to the simplex, the new volume is given by:

$$\text{Vol}(\Delta^{p+1}) = \frac{\text{Vol}(\Delta^p) \times \text{dist}(\mathbf{x}_i, \mathbf{XW}^{(p)})}{p} \quad (3.6)$$

where  $\text{dist}(\mathbf{x}_i, \mathbf{XW}^{(p)})$  is the distance between  $\mathbf{x}_i$  and its projection onto the subspace spanned by  $\mathbf{XW}^{(p)}$ . According to [Courty 2011], in  $\mathcal{H}$ , this distance reads:

$$\text{dist}(\phi(\mathbf{x}_i), \Phi \mathbf{W}^{(p)}) = 1 - (\mathbf{k}_{\mathbf{x}_i}^\top \cdot \mathbf{K}_p^{-1} \cdot \mathbf{k}_{\mathbf{x}_i}) \quad (3.7)$$

where  $\mathbf{K}_p^{-1}$  is the inverse of the kernel matrix of the elements of the CSS, *i.e.*  $\mathbf{K}_p^{-1} = (\mathbf{W}^{(p)\top} \Phi^\top \Phi \mathbf{W}^{(p)})^{-1}$ , and where  $\mathbf{k}_{\mathbf{x}_i}$  is a vector of length  $p$  such that  $\mathbf{k}_{\mathbf{x}_i} = [k(\mathbf{x}_j, \mathbf{x}_i)]_{\mathbf{x}_j \in \mathbf{XW}^{(p)}}$ . Then, it is possible to use a greedy procedure similar to that of SiVM, where Eq. 3.3 translates into:

$$\begin{aligned} i &= \arg \max_q \frac{\text{Vol}(\Delta^p(\Phi \mathbf{W}^{(p)})) \times \text{dist}(\phi(\mathbf{x}_q), \Phi \mathbf{W}^{(p)})}{p} \\ &= \arg \min_q \left[ \mathbf{k}_{\mathbf{x}_q}^\top \cdot \mathbf{K}_p^{-1} \cdot \mathbf{k}_{\mathbf{x}_q} \right]. \end{aligned} \quad (3.8)$$

*Remark 3.* This procedure tends to add datum that most changes the geometry of the manifold spanned by the CSS. As such, it acts as the spectral sampling procedure proposed in [Öztireli 2010] for sampling a 3D mesh in a computer graphics context.

Even if  $\mathbf{K}_p^{-1}$  is computed a single time at each iteration (it does not depend on  $\mathbf{x}_q$ ), a matrix inversion remains a resource demanding operation. Moreover, if  $\ell$  elements are to be selected, then  $\ell$  inversions of matrices of increasing size are expected. Fortunately, it is possible to bypass this inversion, by iteratively constructing  $\mathbf{K}_p^{-1}$  on the basis of the Schur complement [Boyd 2004]. Once  $i$ , the index of the best  $(p+1)$ th point to add to the CSS is defined (Eq. 3.8), one computes

$$\begin{aligned} \mathbf{K}_{p+1}^{-1} &= \begin{bmatrix} \mathbf{K}_p & \mathbf{k}_{\mathbf{x}_i} \\ \mathbf{k}_{\mathbf{x}_i}^\top & 1 \end{bmatrix}^{-1} = \mathcal{K} \cdot \begin{bmatrix} (1 - \mathbf{k}_{\mathbf{x}_i}^\top \cdot \mathbf{K}_p^{-1} \cdot \mathbf{k}_{\mathbf{x}_i})^{-1} & \mathbf{0}_p^\top \\ \mathbf{0}_p & \mathbf{K}_p^{-1} \end{bmatrix} \cdot \mathcal{K}^\top \quad (3.9) \\ \text{with } \mathcal{K} &= \begin{bmatrix} -\mathbf{K}_p^{-1} \cdot \mathbf{k}_{\mathbf{x}_i} & \mathbf{I}_p \\ 1 & \mathbf{0}_p^\top \end{bmatrix} \end{aligned}$$

where  $\mathbf{I}_p$  is the identity matrix of size  $p$  and  $\mathbf{0}_p$  is a vector of  $p$  zeros. The computation works as long as  $\mathbf{k}_{\mathbf{x}_i}^\top \cdot \mathbf{K}_p^{-1} \cdot \mathbf{k}_{\mathbf{x}_i}$  differs from 1, which is always true in the Gaussian RKHS as long as the data points are separated.

### 3.3 Sparse projections onto the RKHS simplex

In this section, we focus on the computation of  $\mathbf{G}$ . We give the formulation of our sparse RKHS simplex projector, and then we discuss its convergence. We notably show some analytical bounds required for the convergence of the method.

#### 3.3.1 Projection on the RKHS simplex

We search for the projection of any point  $\phi(\mathbf{x}_i)$  onto the simplex  $\Delta^\ell(\Phi \mathbf{W})$ , *i.e.* the point of the simplex which minimizes the Euclidean distance to  $\phi(\mathbf{x}_i)$ . It amounts to solving the  $n$  independent problems of computing the rows of  $\mathbf{G}$ :

$$\mathbf{G}_{i\bullet} = \arg \min_{\mathbf{G}_{i\bullet}} \|\phi(\mathbf{x}_i) - \Phi \mathbf{W} \mathbf{G}_{i\bullet}^\top\|^2 \text{ s. t. } \sum_j \mathbf{G}_{ij} = 1, \mathbf{G}_{ij} \geq 0, \forall j \quad (3.10)$$

The constraint  $\sum_j \mathbf{G}_{ij} = 1$ ,  $\mathbf{G}_{ij} \geq 0, \forall j$  is equivalent to have  $\mathbf{G}_{i\bullet}$  in the unit standard simplex  $\Delta^\ell$ , ( $\mathbf{G}$  encodes the barycentric coordinates of  $\mathbf{X}$  in  $\Delta^\ell(\Phi\mathbf{W})$ ). At this point, it is possible to force the sparsity of the projection to  $\lambda < \ell$ , without any extra computational cost: We only need to replace the previous constraint by  $\mathbf{G}_{i\bullet} \in \Delta^\lambda$ . Thus, Eq. (3.10) reads:

$$\mathbf{G}_{i\bullet} = \arg \min_{\mathbf{G}_{i\bullet}} \|\phi(\mathbf{x}_i) - \Phi\mathbf{W}\mathbf{G}_{i\bullet}^\top\|^2 \text{ s. t. } \mathbf{G}_{i\bullet} \in \Delta^\lambda \quad (3.11)$$

Instead of considering quadratic programming, such as in [Thureau 2012] or [Kumar 2013], we follow some recent work on the projection on the unit standard simplex [Kyrillidis 2013], and we propose to use a simple projected gradient descent algorithm to solve Eq. (3.11), which amounts to iterating through different possible solutions of

$$\mathbf{G}_{i\bullet}^{t+1} = \mathcal{P}_\lambda \left( \mathbf{G}_{i\bullet}^t - \varepsilon_t \nabla (\|\phi(\mathbf{x}_i) - \Phi\mathbf{W}\mathbf{G}_{i\bullet}^t\|^2) \right) \quad (3.12)$$

$t$  being the iteration index,  $\varepsilon_t$  a (possibly varying) step size,  $\nabla(\cdot)$  the gradient operator and  $\mathcal{P}_\lambda(\cdot)$  the projector onto  $\Delta^\lambda$ . This kind of projected gradient descent method has recently shown its computational efficiency and is also endowed with theoretical convergence guarantees [Garg 2009]. The gradient reads (we omit the iteration index  $t$  for clarity):

$$\begin{aligned} \nabla(\|\phi(\mathbf{x}_i) - \Phi\mathbf{W}\mathbf{G}_{i\bullet}^\top\|^2) &= \nabla((\phi(\mathbf{x}_i) - \Phi\mathbf{W}\mathbf{G}_{i\bullet}^\top)^\top (\phi(\mathbf{x}_i) - \Phi\mathbf{W}\mathbf{G}_{i\bullet}^\top)) \\ &= \nabla(\mathbf{G}_{i\bullet}^\top \mathbf{K}_\ell \mathbf{G}_{i\bullet}^\top - 2\mathbf{k}_{\mathbf{x}_i} \mathbf{G}_{i\bullet}^\top + k(\mathbf{x}_i, \mathbf{x}_i)), \\ &= 2(\mathbf{G}_{i\bullet}^\top \mathbf{K}_\ell - \mathbf{k}_{\mathbf{x}_i}). \end{aligned} \quad (3.13)$$

As for  $\mathcal{P}_\lambda(\cdot)$ , we rely on the Greedy Selector and Simplex Projector (GSSP) algorithm of [Kyrillidis 2013] which can be summarized as a two-step procedure: firstly the coordinates of the vector are sorted by magnitude, and then the  $\lambda$  greatest values are projected on the unit simplex  $\Delta^\lambda$  (while the other vector entries are set to zero).

*Remark 4.* Since the GSSP procedure projects  $\mathbf{G}_{i\bullet}$  on the subspace spanned by the  $\ell$  columns of  $\mathbf{F}$ , the coordinates of  $\mathbf{G}_{i\bullet}$  embeds the projection over each of the  $\ell$  selected elements in the feature space, so that the sparsity support is chosen in the closest elements in the feature space.

The Gaussian kernel is monotonically decreasing according to the neighboring distance; This implies that for each datum, the sparsity support is made of the closest CSS elements in the input space. It follows that:

*Property 2* (Non-linear local coding). The sparse RKHS simplex projector describes any element  $\mathbf{x}_i$  of  $\mathbf{X}$  with  $\mathbf{G}_{i\bullet}$ , which interprets as its non-linear (because of the kernel non-linearity) barycentric coordinates according to  $\lambda$  prototype points. These prototype points are found in the closest elements of the CSS, thus providing a non-linear local coding for  $\mathbf{x}_i$ .

### 3.3.2 Convergence of the projector

Finally, we establish the convergence of that projector, which ensures [Kyrillidis 2013] that the final vector  $\mathbf{G}_{i\bullet}$  is the best  $\lambda$ -sparse solution. To do so, we rely on [Garg 2009], which states that for Eq. 3.11 to be minimized via the projected gradient approach,  $\Phi\mathbf{W}$  has to satisfy  $\lambda$ -restricted isometry property (or  $\lambda$ -RIP for short), with  $\delta_\lambda \in [0, 1[$ . This latter reads:

**Definition 1.** The linear operator  $\Phi\mathbf{W}$  respects the  $\lambda$ -restricted isometry property [Candes 2008] with constant  $\delta_\lambda$  if

$$(1 - \delta_\lambda)\|\mathbf{x}\|_2^2 \leq \|\Phi\mathbf{W}\mathbf{x}\|_2^2 \leq (1 + \delta_\lambda)\|\mathbf{x}\|_2^2 \quad (3.14)$$

where  $\|\cdot\|_2$  refers to the  $L^2$  norm, and for every  $\lambda$ -sparse vector  $\mathbf{x}$ .

Equivalently, the constant  $\delta_\lambda$  can also be defined as:

$$\delta_\lambda := \max_{\substack{\mathcal{L} \subseteq \{1, \dots, \ell\}, \\ |\mathcal{L}| = \lambda}} \|(\Phi\mathbf{W})_{\mathcal{L}}^\top (\Phi\mathbf{W})_{\mathcal{L}} - \mathbf{I}_\lambda\|_2 \quad (3.15)$$

where  $(\Phi\mathbf{W})_{\mathcal{L}}$  denotes a subset matrix of  $\Phi\mathbf{W}$ , with  $\lambda$  columns corresponding to a subset  $\mathcal{L}$  of cardinality  $\lambda$  picked up among the  $\ell$  indices of the CSS. Thus,  $(\Phi\mathbf{W})_{\mathcal{L}}^\top (\Phi\mathbf{W})_{\mathcal{L}}$  is simply the related Gram matrix, noted  $\mathbf{K}_\lambda$ .  $\delta_\lambda$  is defined according to the subset providing a maximum among all the possible combinations of those columns.

As a matter of fact, such convergence holds for particular values of  $\lambda$  and  $\sigma$ , such as stated by the following proposition:

**Proposition 1.** If  $\lambda > 2$ , and if  $\sigma < \frac{d_{min}}{\sqrt{2 \ln(\lambda-1)}}$ , then,  $\Phi\mathbf{W}$  satisfies the  $\lambda$ -RIP with constant  $\delta_\lambda \in [0, 1[$ . If  $\lambda \leq 2$ , there is no particular bound to  $\sigma$ .

*Proof.* Let us first note that since we are working in the Gaussian RKHS, the columns of  $\Phi\mathbf{W}$  have unit norms. The diagonal entries of  $\mathbf{K}_\lambda$  are 1, and  $\forall i, j \in \mathcal{L}$  the off-diagonal element  $(i, j)$  is  $k(\mathbf{x}_i, \mathbf{x}_j)$ .

Let  $\nu_i$  be an eigenvalue of a matrix  $\mathbf{A}$ . By the Gershgorin circle theorem, we know that:

$$|\nu_i - \mathbf{A}(i, i)| < \sum_{j \leq \lambda, i \neq j} |\mathbf{A}(i, j)| \quad (3.16)$$

Thus, for  $\mathbf{A} = \mathbf{K}_\lambda - \mathbf{I}_\lambda$ , we obtain  $|\nu_i| < \sum_{j \leq \lambda, i \neq j} k(\mathbf{x}_i, \mathbf{x}_j)$ . Let  $\mu$  be the greatest dot product between the elements of  $(\Phi\mathbf{W})_{\mathcal{L}}$ , i.e.  $\mu := \max k(\mathbf{x}_i, \mathbf{x}_j)$ ,  $\forall i, j \in \mathcal{L}, i \neq j$ . We can write:

$$\sum_{j \leq \lambda, i \neq j} k(\mathbf{x}_i, \mathbf{x}_j) \leq (\lambda - 1)\mu \quad (3.17)$$

and we have a bound for every eigenvalue  $\nu_i$  and every subset  $\mathcal{L}$  of cardinality  $\lambda$ . Thus,  $\Phi\mathbf{W}$  follows the  $\lambda$ -RIP with  $\delta_\lambda = (\lambda - 1)\mu$  [Bandeira 2012]. Let  $d_{min}^2$  be the minimum squared distance between two distinct elements of  $\mathbf{X}\mathbf{W}$ , i.e.  $d_{min}^2 = \min_{i, j, i \neq j} \|\mathbf{x}_i - \mathbf{x}_j\|^2$ . As  $\mu = \exp(-d_{min}^2/2\sigma^2)$ , one has:

$$\delta_\lambda < 1 \Leftrightarrow (\lambda - 1) \exp\left(\frac{-d_{min}^2}{2\sigma^2}\right) < 1 \Leftrightarrow \sigma < \frac{d_{min}}{\sqrt{2 \ln(\lambda - 1)}}, \lambda > 2 \quad (3.18)$$

□

This allows deriving conditions on  $\sigma$  for different convergence or approximation guarantees, such as  $\delta_{2\lambda} < 1/3$  for convergence [Garg 2009] (Theorem 2.1). This point is discussed in [Kyrillidis 2013] (Section 2). We note that similar developments using the Gershgorin circle theorem have been used for the deterministic creation of projection matrices in the domain of compressed sensing [Bandeira 2012] (Section 2.1).

### 3.4 Complete SAGA procedure

We give in Alg. 1 the complete SAGA matrix factorization procedure<sup>3</sup>. We then discuss its computational complexity.

#### 3.4.1 Algorithm

In addition to the data matrix  $\mathbf{X}$ , the user needs to tune the following parameters:  $\sigma$  the Gaussian kernel bandwidth,  $\lambda$  the expected sparsity level, and  $\ell$  the number of prototypes in the CSS (with  $\lambda \leq \ell$ ). Additional parameters for the gradient descent can be tuned to optimize the computation load, yet it is not mandatory. At first, one computes the CSS, then the matrix of projections. Regarding the CSS, the procedure is initialized (Lines 1 to 4) as in [Thureau 2012]: First, one randomly selects a datum. Then, one finds the most distant datum to that first one. Finally, the most distant datum to this second datum is selected, and is considered as the first element of the CSS (see Lines 2 and 3 in Alg. 1). After initialization, the iterative increment of the simplex dimensionality is implemented in the loop from Line 5 to 8. Regarding the projection, each datum is processed separately thanks to the loop from Line 9 to 16. Within this loop, another loop deals with the gradient descent up to a stopping criterion (loop from Line 11 to 16).

#### 3.4.2 Computational complexity

*Property 3* (Computational complexity for  $\mathbf{W}$  – noted as **First step** in Alg. 1). The selection of the CSS defining  $\mathbf{W}$  based on the exact and incremental simplex volume maximization has complexity of  $o(n\ell^3)$ , i.e. it has a linear complexity with respect to  $n$ .

*Proof.* The entire CSS construction is based on Eq. 3.8 and 3.9. At each step of the selection, the procedure amounts to a linear scanning procedure where the volume increment is computed for each element in the dataset (size  $n$ ). For one element  $\mathbf{x}_i$ , with  $i \in [1, n]$ , this requires to compute the associated bilinear form  $\mathbf{k}_{\mathbf{x}_q}^\top \cdot \mathbf{K}_{i-1}^{-1} \cdot \mathbf{k}_{\mathbf{x}_q}$  vector. with an asymptotical computational complexity of  $o(p^2)$ . As  $p$  varies 1 to  $\ell$ , one ends up with a cubical complexity term regarding  $\ell$ .  $\square$

We note that thanks to Eq. 3.9, almost all the values (except the one corresponding to the kernel evaluation with the last chosen element) have already been computed in the previous iteration and do not need to be computed again. This makes the overall CSS computation that efficient. However, for exceptionally large data matrices, it is possible to improve it with the randomized approximate search of [Thureau 2010], that can be directly adapted. Finally, let us remark that, in spite of being linear in terms of  $n$ , the number of elements in the dataset, the procedure is not linear with respect to  $\ell$ . However, as  $\ell$  is classically several order of magnitude smaller than  $n$ , so that it is seldom important. This is why, in a similar way, SiVM is also not linear with respect to  $\ell$ .

The complexity of the computation of the projection matrix  $\mathbf{G}$  is by construction linear with respect to  $n$  as in the algorithm, it is decomposed into a

<sup>3</sup>The MATLAB source code is available on <http://people.irisa.fr/Nicolas.Courty/SAGA>

---

**Algorithm 1:** The SAGA Matrix Factorization algorithm

---

**input** :  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]_{(\mathbf{x}_i \in \mathbb{R}^p)}$  the data matrix  
 $\sigma$  the Gaussian kernel bandwidth  
 $\lambda$  the expected sparsity level  
 $\ell$  the number of prototypes  
 $E = \{\epsilon, (\epsilon_t)_{(t \in \mathbb{N})}\}$  additional parameters for the gradient descent  
**output:**  $\mathbf{W}$  the indicator matrix  
 $\mathbf{G}$  the reduced sparse convex embedding of  $\mathbf{X}$

**// First step:** Column Subset Selection as a manifold subsampling  
1  $\mathbf{W}^{(i=1)} \leftarrow 0$  //  $W^{(i)} \in \{0, 1\}^{n \times i}$   
2  $t \leftarrow \arg \min_q [k(\mathbf{x}_q, \mathbf{x}_{\text{rand}[1, n]})]$   
3  $e \leftarrow \arg \min_q [k(\mathbf{x}_q, \mathbf{x}_t)]$  //  $\mathbf{x}_e$  is the first element  
4  $\mathbf{W}_{et}^{(i=1)} \leftarrow 1$   
5 **for**  $i \leftarrow 2$  **to**  $\ell$  **do**  
6  $e \leftarrow \arg \min_q [\mathbf{k}_{\mathbf{x}_q}^\top \cdot \mathbf{K}_{i-1}^{-1} \cdot \mathbf{k}_{\mathbf{x}_q}]$   
7  $\mathbf{W}_{\bullet i}^{(i)} \leftarrow 0$ ;  $\mathbf{W}_{ei}^{(i)} \leftarrow 1$   
8 compute  $\mathbf{K}_i^{-1}$  from  $\mathbf{K}_{i-1}^{-1}$  using Eq. 3.9

**// Second step:** Sparse Projection over the defined simplex  
9 **for**  $\mathbf{x}_i \in \mathbf{X}$  **do**  
10  $\mathbf{G}_{i\bullet}^{(k=0)} \leftarrow [1/\ell, \dots, 1/\ell]$   
11 **repeat**  
12  $\mathbf{G}_{i\bullet}^{(k+1)} \leftarrow \mathbf{G}_{i\bullet}^{(k)} - \epsilon_t (\mathbf{G}_{i\bullet}^{(k)\top} \mathbf{K}_\ell - \mathbf{k}_{\mathbf{x}_i})$   
13 find the indices of the  $\lambda$  greatest values of  $\mathbf{G}_{i\bullet}^{(k+1)}$   
14 project the corresponding elements of  $\mathbf{G}_{i\bullet}^{(k+1)}$  onto  $\Delta^\lambda$   
[Maculan 1989]  
15 set the other elements to 0  
16 **until**  $\|\mathbf{G}_{i\bullet}^{(k+1)} - \mathbf{G}_{i\bullet}^{(k)}\|^2 < \epsilon$ ;

---



succession of  $n$  independent projections. However, the complexity of the projection with respect to  $\ell$  is linear:

*Property 4* (Computational complexity for  $\mathbf{G}$  – noted as **Second step** in Alg. 1). Each projection has a linear complexity regarding  $\lambda$ .

*Proof.* Each projection is computed through a gradient descent, each iteration of which has a complexity dominated by that of the GSSP (Eq. 3.12). The latter requires getting the  $\lambda$  greatest values of  $\mathbf{G}_{i\bullet}^t$ . To do so, the GSSP classically relies on sorting the elements of  $\mathbf{G}_{i\bullet}^t$  (a vector of size  $\ell$ ), with a  $o(\ell \log(\ell))$  complexity. However, it is possible to be more efficient by achieving this task in  $o(\lambda)$  thanks to the median-finding algorithm [Maculan 1989]. Finally, if  $q$  iterations are needed in the descent, then the total complexity of one projection is  $o(q\lambda)$ .  $\square$

Let us note here that  $q$  typically depends on the choice of the magnitude of the gradient step  $\varepsilon_t$ , which can be efficiently set following the results of [Garg 2009]. In practice, only a few tens of iterations are necessary and if one has  $n \gg q\lambda$ , the influence of the number of iterations is immaterial.

Finally, as both the definition of the CSS and the projection have a linear complexity with respect to the dataset size, the overall SAGA procedure also has. This allows factorizing very large matrices in tractable time.

## 3.5 Experiments and results

In this Section, we first observe the behavior of SAGA on some toy datasets. Our goal is to verify the behavior of our algorithm with respect to the theoretical properties given in the previous section, with a special focus on the nature of the subsampling occurring the RKHS. Then, the volume maximization strategy is discussed, as well as its impact on the reconstruction errors over a toy and a real dataset. We finally discuss the potential use of SAGA in a feature extraction context for classification purpose. The performances of our method are then compared to a selection of state-of-the-art methods performing NMF with characteristics shared by our method (sparsity, kernels, convexity, etc.).

### 3.5.1 Experiments on toy datasets

The SAGA paradigm is first illustrated on toy datasets (Fig. 3.3, 3.4 and 3.5). In the first examples we consider simulated datasets (ring and S shaped respectively) made of 600 points in  $\mathbb{R}^2$ : In the input space, the points of the CSS form a non-simplicial polytope which intuitively fits with the manifold hull idea (both inner and outer contours are displayed for the ring). If one projects points onto the corresponding CSS ( $\ell = 30$  for the ring, and  $\ell = 50$  for the S), the magnitude of the reconstruction error fits with the non-linear geometry of the simplex in the input space (Fig. 3.3). The sparsity and the locality of the reconstruction are displayed on Fig. 3.4: It appears that for each point, the number of non-null components is smaller than or equal to  $\lambda = 3$  (ring) or  $\lambda = 5$  (S). Moreover, these non-nil components are all located in the very close neighborhood of the projected point, as expected for a non-linear local embedding. Figure 3.5(a) shows the evolution of the reconstruction error and the locality

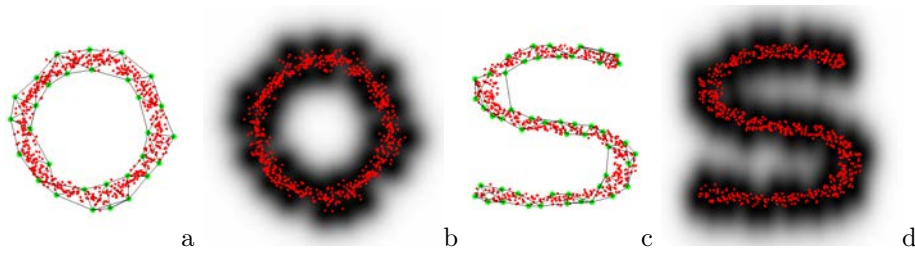


Figure 3.3: A ring and an S-shape datasets (a-c) It is possible to have  $\ell > p$  elements in the CSS forming a non-simplicial polytope (here illustrated by a 4-NN graph) which approximates well the contour of the shape. (b-d) Each pixel of the image is projected in the RKHS onto the 29-simplex (b) or 49-simplex (d). Each pixel is colored according to its reconstruction error (black for low values, white for high ones).

of the samples used for the reconstruction. This last term is measured as the radius of the minimum volume enclosing ball (computed as a smallest enclosing ball problem [Gärtner 1999]). As expected, this mean radius is decreasing as more samples are taken from the dataset.

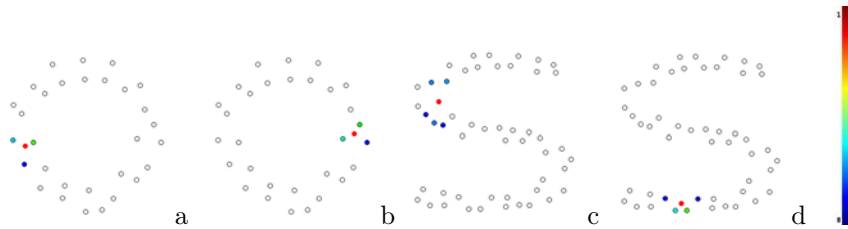


Figure 3.4: (a-b) The 29-simplex on the ring shape dataset: The red point is projected in the RKHS as a 3-sparse vector. Each vertex color accounts for the corresponding convex coordinates (from blue= 0 to red= 1) (c-d) The S dataset with 5-sparse vectors (in (d), only 4 projections are non-nil).

Next, the reconstruction error of SAGA is evaluated with respect to the chosen sparsity level and compared to the original version of SiVM [Thureau 2010], yet in the Gaussian RKHS, *i.e.* a simplex volume maximization of Eq. 3.5, followed by a projection based on quadratic programming. We draw 2,000 points in  $\mathbb{R}^{30}$  according to a Gaussian distribution with covariance matrix  $\sigma = 0.5\mathbf{I}$ . We measure for SiVM and SAGA the normalized reconstruction error  $\|\Phi - \Phi\mathbf{W}\mathbf{G}^\top\|^2/\|\Phi\|^2$ . We remark here that this formula is correct for SiVM if it is implemented through a kernel form with the linear kernel. Results are displayed in Figure 3.5(b). When  $\ell \leq p = 30$ , SAGA (with or without sparsity constraint) as well as SiVM directly operates as a dimensionality reduction method: As the SiVM curve is below that of SAGA  $\lambda = \ell$ , SiVM appears as more reliable, which makes sense, as the Gaussian distribution provides a rather convex dataset. However, even on such a dataset, if the reconstruction error is considered along with sparsity, it appears that, whatever the value  $\lambda^*$  chosen

for parameter  $\lambda$ , it is always possible to find a value  $\ell^*$  for  $\ell$ , such that the reconstruction error with SAGA tuned with  $(\ell = \ell^*, \lambda = \lambda^*)$  is smaller than with SiVM tuned with  $(\ell = \lambda^*)$ . This illustrates well that it is possible to reduce the reconstruction error while constraining the solution sparsity. If one considers the case where  $\ell \geq p = 30$ , SiVM fails in producing reconstruction error which decreases when  $\ell$  increases: From Eq. (3.1), the addition of a  $p + 1$  vertex leads to a simplex of null volume, which is impossible to maximize, and turning SiVM into a random projection method. Thus, the reconstruction error becomes greater than with SAGA, the latter enhancing the reconstruction quality, despite strong sparsity constraints.

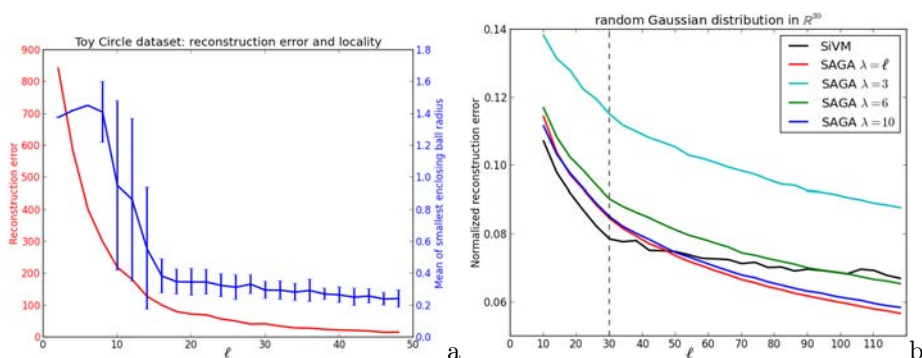


Figure 3.5: (a) Reconstruction error (in red) and mean radius of smallest enclosing ball (blue) for the ring dataset. (b) Normalized reconstruction error.

### 3.5.2 Comparison on subsampling strategies

The quality of the overall matrix factorization procedure relies on two elements: the ability of the CSS to correctly define the manifold hull, and the precision of the projection. A particular focus is given here on the first one, as various sampling strategies are compared and discussed.

The goal of our first comparison is to confirm that, among the methods based on volume maximization, ours is the most accurate. To this end, we compare three volume maximization methods (operating in the RKHS, to fit the objectives of this work): The first one is very method from [Thureau 2012], reported through Eq. 3.5, including the constant distance assumption as well as the replacement of squared distances by simple distances. As this point is only supported by computational considerations, we also consider the method summarized by Eq. 3.4, where the constant distance assumption still holds, but where the squared distances are kept. These two methods are referred to as SiVM-approx-(3.4) and SiVM-approx-(3.5). Naturally, the third one is that of SAGA, based on Eq. 3.8 and 3.9. As a reference, we consider the exact volume computation based on the Cayley-Menger determinant (CMD), and we compare the ratios of the volumes derived by the three methods over the reference one. Due to the computation cost of CMD, we have restrict the size of the CSS to  $\ell = 8$ .

Three types of datasets are used for this comparison. First datasets are composed of 2,000 points in  $\mathbb{R}^{30}$  according to a uniform distribution. The second type of datasets is used to test the robustness of the methods to ill-conditioned data; so, 2,000 elements in  $\mathbb{R}^{50}$  are generated and their singular values are transformed such as described in [Gillis 2013]. Finally, our last dataset is obtained through a random selection of 160 images from the real dataset COIL-20 (Columbia University Image Library). COIL-20 contains  $128 \times 128$  gray images of 20 objects at different view angles, for a total number of sample of 1,440 [Nene 1996]. We use 30 datasets of each type and we compute the mean ratio (and variance) of the volumes, as described above (see Table 3.1).

Table 3.1: Mean ratio and variance (in %, over 30 runs) of maximum simplex volumes found by both approximations from Eq. 3.4 and Eq. 3.5, as well as SAGA over the exact CMD approach. For each dataset, the most accurate result is in bold font.

	Uniform	Ill-conditioned	COIL-20
SiVM-approx-(3.4)	98.93 (0.07)	59.18 (1.6)	97.37 (0.12)
SiVM-approx-(3.5)	96.03 (0.21)	55.85 (1.5)	96.28 (0.13)
SAGA	<b>100.13</b> ( $2 \cdot 10^{-3}$ )	<b>100.04</b> ( $6 \cdot 10^{-3}$ )	<b>100.00</b> (0.00)

Some conclusions can be drawn from Table 3.1. First, the lower ratios of SiVM-approx-(3.4) and SiVM-approx-(3.5) for ill-conditioned datasets confirm the conclusion of [Gillis 2013] regarding the constant distance approximation, while SAGA is not bothered. Second, it is possible to notice the slight decrement of the performances due to the supplemental approximation in SiVM-approx-(3.5) where the squared distances are not considered anymore for computational reasons. Also, for all datasets, SAGA finds the most similar simplex volumes to the reference ones, as the ratios are the closest to 1. Oddly enough, the ratios are sometimes even slightly greater than 1, due to the numerical imprecisions of the determinant computations which may induce different choices for the simplex vertices. In the meantime, while volume differences between SAGA and CMD are low, their respective computational performances are quite different. As an example, with COIL dataset, each CMD computation requires  $\approx 10$  sec. whereas SAGA takes  $\approx 0.01$  seconds in an unoptimized implementation. As a conclusion, SAGA provides a volume as large as what of CMD approach, yet in much less time. However, the computational accuracy of SAGA is fully investigated in the next subsection.

Now that it is established that SAGA provides a better hull for the manifold than classical methods based on a kernelization of the simplex volume maximization, let us compare it to other manifold sampling strategies. As it is not meaningful to use the simplex volume as a criterion, we consider the normalized reconstruction errors, such as with toy datasets. We can also consider larger CSS values, ranging from 10 to more than 100, as the computation of the CMD is not an issue anymore. In the comparison, we keep SiVM-approx-(3.5), as it corresponds to a kernelized version of the original paper [Thureau 2012]. We also consider random sampling (the NMF reducing to a random projection algorithm), kernel  $k$ -means and kernel rank-revealing Cholesky of [Courty 2013],

as a surrogate for the rank-revealing QR method of [Shroff 2011] which cannot be kernelized.

Two datasets are used, namely the UCI Wine dataset [Bache 2013] composed of 178 instances with 13 attributes and a simulated dataset which consist of 2,000 points in  $\mathbb{R}^{50}$  according to a Gaussian distribution. Reconstruction error curves are presented in Fig 3.6 and demonstrate the best overall performances of SAGA with increasing  $\ell$  values. However this general trend differs according to the considered dataset and the  $\ell$  values. For example if SAGA clearly outperforms SiVM with the Wine dataset, their performances are very close with the Gaussian one. With lowest  $\ell$  values, kernel  $k$ -means and kernel rank-revealing Cholesky get better results on the Wine dataset, however, the reconstruction error remains high whatever the strategy.

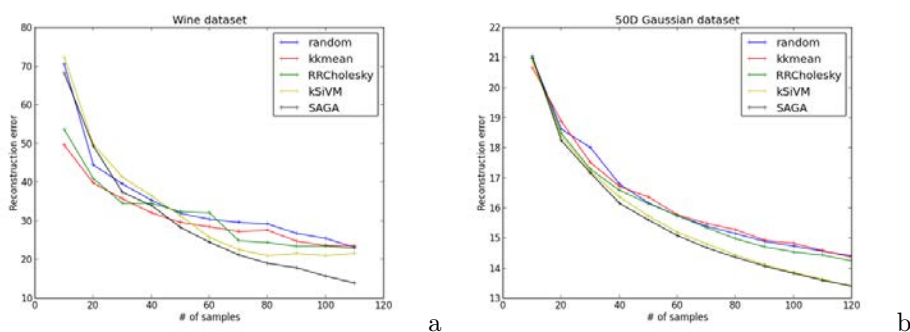


Figure 3.6: Comparison of the reconstruction error according to the size of the CSS: (a) Wine dataset; (b) simulated dataset (50 dimensional Gaussian).

### 3.5.3 Computational complexity

First, we consider the complexity of the CSS computation. A random dataset in  $\mathbb{R}^{30}$  of size  $n$ , with  $n$  ranging in [1000–10000] is considered, with  $\ell = 10$ , for both linear and Gaussian kernel. Each time the experiment is repeated 20 times to stabilize the measure, and the computation times are reported in Fig. 3.7(a). As expected, we observe a near linear trend with both different versions of kernel. The computational differences occurring between the two kernels are mostly due to the evaluation of the kernel: the linear kernel results in a simple dot product operation, whereas evaluating the Gaussian kernel involves computations of transcendental operators.

If we now turn to the complexity of the projection, it is useless to consider it with respect to  $n$  as the projection is dealt datum by datum. However, it is interesting to consider the complexity regarding  $\ell$ . To do so, we use the same experimental setting, yet,  $n$  is fixed and  $\ell$  varies from 1 to 120. Fig. 3.7(b) clearly highlights the outperformance of our projector compared to the quadratic programming approach used in SiVM.

### 3.5.4 Application to classification on real datasets

In order to propose a more application-oriented test of SAGA, we propose to use the result of various NMF methods as feature extraction processes in a

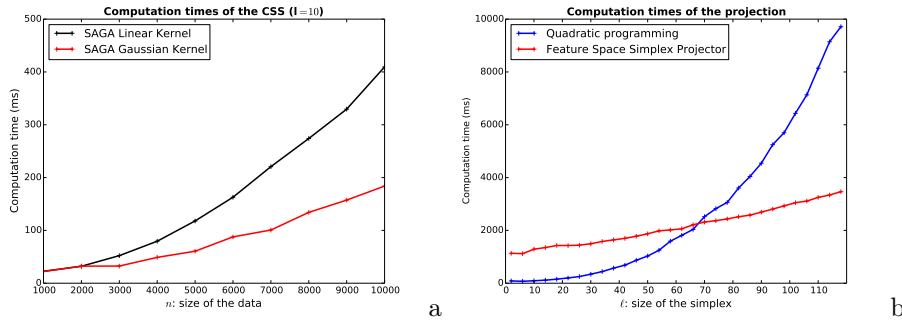


Figure 3.7: Comparison of the computational times for the CSS (a) and the projection (b) for SiVM and SAGA.

classification problem. Experiments are conducted on four publicly available datasets. The previously described COIL-20 dataset, the CMU PIE dataset, the MNIST dataset and the CIFAR-10 dataset: The CMU PIE dataset contains the 32 gray scale facial images of 68 people, any of each having 42 images at different illumination and facial expression conditions [Sim 2002]. The MNIST dataset [LeCun 1998] contains the  $28 \times 28$  gray scale images of handwritten numbers (from 0 to 9). Its total sample number is 70,000. Finally, CIFAR-10 is made of 60,000  $32 \times 32$  color images in 10 classes, with 6,000 images per class.

Each dataset is partitioned into two sets: one for training (one tenth of the dataset), and the other for testing. On the training dataset  $\mathbf{X}_t$ , a factorization is conducted and leads to derive  $\mathbf{F}_t$  and  $\mathbf{G}_t$ .  $\mathbf{F}_t$  plays the role of a visual coding dictionary, and is used to reconstruct the test set. We note here that  $\mathbf{F}_t$  could have been constructed with respect to the entire dataset, in a unsupervised learning manner, but it was not the case.  $\mathbf{G}_t$  is used to train a SVM classifier (with Gaussian kernel), with parameters optimized by standard cross-validation. The testing set  $\mathbf{X}_s$  is then projected over  $\mathbf{F}_t$  which allows deriving  $\mathbf{G}_s$ , used for testing with the SVM classifier. This process is repeated 10 times for each value of  $\ell \in \{10, 20, 30, 40, 50\}$ , and for each  $\ell/\lambda$  ratio  $\in \{1, 1.5, 2\}$ , in order to stabilize the performances. In this setting, we have compared the result of SAGA to other state-of-the-art algorithms: Sparse NMF [Kim 2007] (with  $\ell/\lambda = 2$ ), Kernel NMF [Li 2012], Convex NMF [Ding 2010] and Kernel Convex NMF [Ding 2010] (or KC NMF for short). We have used the same kernel variance for all kernel-based techniques, on the basis of a standard rule of thumb [Von Luxburg 2007]. The mean results over the 10 repetitions are displayed in Fig. 3.8. It shows that SAGA produces the most accurate results on three datasets out of four. In the case of CIFAR-10 dataset, one notices first the very low performances of all the methods with respect to the state-of-the-art works focusing on classification performances. The reason is that the size of the CSS remains rather low in our experimental setting: the point of this comparison is not to exhibit the highest possible accuracies, but rather to provide a sound experimental setting across various datasets of heterogeneous difficulty. However, this does not explain why SAGA does not compete with Sparse NMF and Kernel NMF on this dataset. A possible explanation stems from the complexity of the manifold hull which cannot be described efficiently by so few elements. In this particular case, SAGA

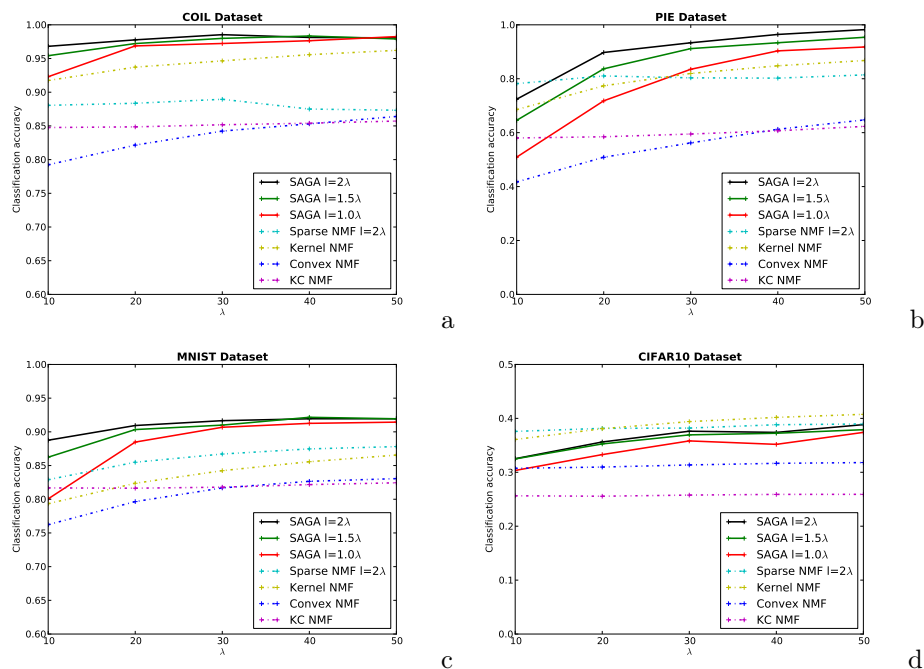


Figure 3.8: Feature extraction evaluation on (a) COIL-20, (b) PIE, (c) MNIST and (d) CIFAR-10 datasets at different reduced dimension.

would perform slightly worse than other state-of-the-art methods. Interestingly enough, whatever the dataset, when  $\lambda$  stays the same, the SAGA classification accuracy improves with the increasing number  $\ell$  of simplex vertices. However,  $\ell$  exerts little influence when  $\lambda$  is already large. As discussed in [Yu 2009], we can relate the optimal value of  $\ell$  to the intrinsic dimensionality of the manifold where the data live.

Table 3.2 summarizes the results of Fig. 3.8 by averaging the performances  $\forall \ell \in \{10, 20, 30, 40, 50\}$ , (for Sparse NMF and SAGA, with  $\ell/\lambda = 2$ ): The variances which are not displayed on Fig. 3.8 for clarity sakes are given here. SAGA variance is sometimes important due to the strong increment of the performances when  $\ell$  increases. To allow for more complete comparisons, we have also added a kernel version of the original SiVM with our projection method, noted Sparse Kernel SiVM (SK SiVM for short), which basically amounts to using SAGA, yet with Eq. 3.5 instead of Eq. 3.8 and 3.9; it appears that it is always less accurate than SAGA.

Finally, on the majority of the datasets, the superiority of SAGA is established. Also, it is the fastest of all, as can be seen in Fig. 3.9, where the factorization performances are reported for SAGA and the considered state-of-the-art NMF methods for the whole PIE dataset when varying the size of the CSS. This illustrates the computational benefits of our approach.

Table 3.2: Mean accuracy and variance NMF-based classification. For each dataset, the most accurate method is in bold font. This property is assessed thanks to a paired Student’s T-test between the SAGA and the best other method (apart from SK SiVM which can be seen as a particular case of SAGA)

	COIL-20	PIE	MNIST	CIFAR-10
Sparse NMF	88.04 (0.66)	80.24 (1.25)	86.08 (1.99)	38.34 (0.56)
Kernel NMF	94.38 (1.75)	79.89 (7.25)	83.61 (2.86)	<b>38.90 (1.87)</b>
Convex NMF	83.46 (2.84)	54.95 (9.07)	80.66 (2.80)	31.31 (0.45)
KC NMF	85.18 (0.39)	59.80 (1.75)	81.94 (0.35)	25.76 (0.16)
SK SiVM	96.79 (1.34)	83.73 (9.32)	89.71 (3.78)	35.50 (2.20)
SAGA	<b>97.23 (1.58)</b>	<b>84.44 (13.62)</b>	<b>89.92 (3.18)</b>	35.59 (2.44)
$t$ -statistics	3.2327	1.8872	2.3257	X
Confidence	$\geq 99\%$	$\geq 95\%$	$\geq 97.5\%$	X

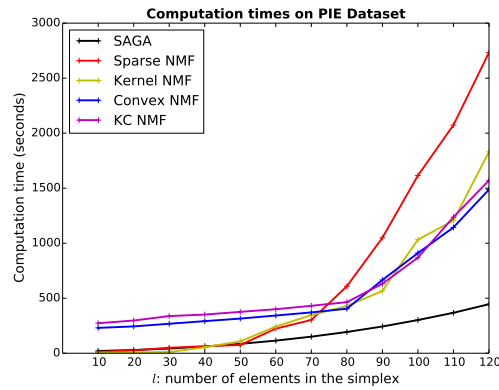


Figure 3.9: Computational performances between SAGA and the other considered NMF methods on the whole PIE dataset.

## 3.6 Conclusion

SAGA (Sparse and Geometry Aware) is a new matrix factorization algorithm which has the following properties: (1) it operates in the Gaussian RKHS, which accounts for potential nonlinearity in the dataset geometry; (2) it provides sparse and convex projections onto a reduced embedding spanned by selected typical samples, which facilitates the human interpretation, and leads to a non-linear local representation of the data; (3) it has a complexity linear with the number of data entries, which allows dealing with big data. SAGA relies on both a manifold sampling strategy and a data projection. This latter has been proved to converge under some conditions regarding the Gaussian kernel variance. Finally, SAGA has been tested on toy, simulated and real datasets. The following conclusions can be drawn from the experiments: we observed in accordance with the theory that SAGA encodes the data as convex combinations of neighbor sam-



ples; the proposed volume maximization heuristic leads to better subsampling of the original data with respect to the volume of the simplex formed by the CSS; and the performances of its feature extraction have proved to outperform the selected state-of-the-art other NMF methods on classification tasks.

## Part III

# Démélangeage de séries temporelles



# Time series unmixing

## Résumé du chapitre

Dans cette partie, nous appliquons la technique vue au chapitre précédent au cas du démélangeage de séries temporelles de variables biophysiques associées à chaque pixel d'une séquence d'images MODIS. Nous cherchons plus précisément à estimer les fractions associées à chaque classe au sein d'une série. L'apport

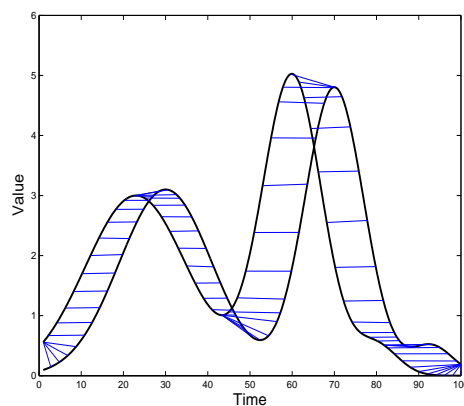


Figure 4.1: **Principe du "Dynamic Time Warping" (DTW) pour l'alignement de deux séries temporelles.** Les deux courbes présentent un profil similaire associé à une même classe. Malgré cela, la distance euclidienne entre ces séries est importante. Le principe de DTW est de chercher le meilleur alignement entre les deux courbes (droites bleues) et de calculer l'erreur sur les courbes ré-alignées, mettant ainsi en évidence le fait que les deux courbes sont similaires.

méthodologique principal est l'emploi de noyaux adaptés aux séries temporelles pour effectuer la transformation des données (le démélangeage étant effectué au moyen de la technique proposée dans le chapitre précédent). Ces noyaux s'appuient sur des *distances élastiques* (comme le *Dynamic Time Warping* illustré en Fig 4.1) qui permettent de comparer de manière plus efficace des *profils temporels* par rapport à la distance euclidienne qui ne prend pas en compte l'aspect corrélé des valeurs en chaque point d'une courbe. Ce chapitre est organisé comme suit :

- La **section 4.1** rappelle les enjeux du démélangeage de séries temporelles ;

- La **section 4.2** établit un état de l'art dans ce domaine;
- La **section 4.3** présente la méthodologie, qui combine les aspects théoriques du chapitre précédent avec des noyaux temporels;
- La **section 4.4** présente des résultats expérimentaux sur des données MODIS.

Soulignons ici que cette approche visant à démêler directement des séries temporelles n'a, selon notre connaissance, pas encore été exploitée. Ce chapitre étudie donc les possibilités d'une telle démarche et les résultats préliminaires sont encourageants. Les travaux présentés ici correspondent à un article soumis dans la revue *IEEE Transactions on Geosciences and Remote Sensing*.

---

## Abstract

The main objective of this chapter is to explore the possibility of pixel unmixing related to remote sensing time series.

In remote sensing research, mixed pixels are the result of the combination of the distinct materials that occupy a single pixel due to the spatial resolution of the sensor. The distinct materials associated with the surface are called endmembers, and the fractions in which they appear in a pixel are called abundances. The pixel unmixing problem involves estimating all or some of: the number of endmembers, their signatures and their abundances in each pixel. Traditionally, pixel unmixing is implemented on hyperspectral data, based on the assumption that the spectral signatures of endmembers are time invariant. Of course this time invariance is not valid for time series. To deal with this specificity, similarly than other state-of-art approach and previous chapter, we rely on non-negative matrix factorization (NMF) framework in feature space. The kernel-based approach allows us to make use of *elastic* kernels when measuring the similarity between two time series. Based on kernel space simplex volume maximisation (SiVM) method, we extract an over-complete set of endmembers and corresponding abundances. The preliminary experimental results, obtained on real dataset, illustrate the effectiveness and the generalization capacities of the proposed approach.

## Keywords

Time series, pixel unmixing, kernel, endmember, over-complete endmember set.

## 4.1 Introduction

At the moment, the large archive produced by satellite remote-sensing missions is becoming an increasing valuable source of information [Jensen 1996][Sabins Jr 1978]. This leads to a better interpretation of land-cover and land-use by analyzing the accumulated satellite data. The quality of remote sensing data is dependent of its spatial, spectral, radiometric and temporal resolutions [Lillesand 2004] and a given image is a tradeoff among these four resolutions. For instance, due to the limit of its orbit and imaging equipment, the high temporal and radiometric resolution satellite image would likely be of low spatial resolution [Richards 1999].

The orbit of satellites for high spatial resolution has to be close to the earth surface, making it subjected to the influence of the atmosphere. The associated operational life span is relatively short compared with other sort of satellites and their revisit frequency is low. On the drawback, for median or low spatial resolution satellite image, the temporal rate is better but a single pixel generally corresponds to a large square ground area [Harsanyi 1994], possibly embedding multiple distinct materials. As a consequence, researchers have concentrated their efforts to the pixel unmixing problem, the objectives of which are to estimate all or some of the number of endmembers, their signatures and their abundances in each pixel.

For pixel unmixing, a popular motivation is to exploit the information of hyperspectral image [Camps-Valls 2005][Chan 2009]. Hyperspectral cameras (HSCs) measure electromagnetic energy scattered in their instantaneous field

view in hundreds or thousands of spectral channels with higher spectral resolution than multispectral cameras [Bioucas-Dias 2012][Keshava 2002]. The unmixing approaches for HSCs, generally called *hyperspectral unmixing* (HU) methods, assume that the spectral signature of each endmember is persistent and uniform over time. Many models searching for robust, stable, tractable and accurate hyperspectral unmixing algorithms have been developed the past decades [Bioucas-Dias 2012]. For example the approach in [Smith 1985] is applicable to remote sensing data of planetary surfaces for quantitative determinations of mineral abundances. In a similar way Adams et al propose a spectral mixture model for the analysis of rock and soil types [Adams 1986] while in [Swayze 1992], the hyperspectral unmixing technique is used to obtain mineral maps from AVIRIS data of Cuprite mining district.

However, although the spectral signatures of minerals, soils or water areas may be time invariant, this assumption is not always valid for most of agricultural land-covers such as forest, grassland or agricultural field [Goenaga 2013]. Broadly speaking, the state of ecological systems would likely differ in various seasons [Walthall 1992], resulting in rhythm patterns of remote sensing sequential observations in a given area. A way to deal with this is to exploit remote sensing time series which could potentially contain much more information than single images [Sakamoto 2005]. Many satellites such as MODIS [Justice 1998] or Landsat [Vogelmann 2001] have been working on orbit for long, yielding today remote sensing time series data which are a valuable source of observation on which the unmixing problem can be applied.

The attempt to exploit time series also brings about new challenges. First, time series could be of various length: the observation dates and frequency of two time series in two distinct areas may be not identical due to cloud coverage and satellite availability. Second, the climate fluctuation of two different years would result in semantically *similar* time series but associated with *large* Euclidean distance between each others. These issues are also encountered in other research fields as for example speech recognition where two audio clips of different length could convey exact sentence [Myers 1980]. Many efforts have been devoted to handle such special properties of time series the past three decades. In remote sensing, several works have also been designed to deal with time series (related to luminance, spectral data, biophysical parameters, ...). For example in [Petitjean 2012] Petitjean et al. use specific measurements (and in particular the Dynamic Time Warping discussed later) to classify various remote sensing time series while in [Petitjean 2011] they propose specific strategies for their averaging and clustering. A similar approach with various comparison criteria has been proposed in [Viovy 2000, Dusseux 2014b] where in this latter article, the application is related to the identification of grassland management practices.

If one intends to exploit temporal sequences, it is essential to measure the similarity between two time series. Despite other notable attempts [Jebara 2004], the most popular standard to compare time series remains the Dynamic Time Warping (DTW) distance and its associated elastic kernels. The DTW framework has been extensively studied since it was first proposed by Sakoe et al [Sakoe 1978] and used since in a large number of application papers. Even though time series data is not new in remote sensing, only a few recent papers attempt to apply it to pixel unmixing. In [Hemissi 2013], Hemissi et al propose to summarize temporal sequence of spectral signatures as a 3-D mesh and measure the similarity of the temporal spectral signature sequence by comparing the

shapes of 3-D meshes for mixed pixel unmixing. The paper in [Hall 1987] combines the several spectral signatures in distinct seasons to construct an extended signature which is used for pixel unmixing. However none of these methods have fully explored the time shift property of time series under DTW framework for pixel unmixing.

Among the state-of-the-art pixel unmixing approaches, efficient and recent ones are implemented under the non-negative matrix factorization (NMF) framework [Pauca 2006][Zhu 2014]. Non-negative matrix factorization [Courty 2014] (or NMF for short) try to decompose a given matrix  $\mathbf{X}$  as a product of two matrices  $\mathbf{M}$  and  $\mathbf{A}$ , such that  $\|\mathbf{X} - \mathbf{MA}\|$  is minimized, under the constraints that the elements of  $\mathbf{A}$  are positive or nil. Many interesting properties can arise (as for instance unmixing) with such a constraint. Among existing techniques, Li et al have transposed NMF onto a reproducing kernel Hilbert space (RKHS), resulting in an efficient *kernel NMF* approach [Li 2012]. In their application, each column of  $\mathbf{M}$  is a combination of all columns in  $\mathbf{X}$ , and corresponding columns in  $\mathbf{A}$  are weight coefficients. Recently NMF framework has been used for pixel unmixing in hyperspectral data [Bioucas-Dias 2012]. However, this technique is pixel-based and does not specifically take into account the fact that we are dealing with time series.

Motivated by this kernel NMF technique of [Li 2012] and elastic kernels devoted to deal with time series (cf for example in [Dusseux 2014b]), we propose here to perform the time series unmixing in a RKHS space using elastic kernels.

Before going into details, let us first formalize the time series unmixing problem. Let  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$  be a matrix of  $T \times n$  corresponding to a dataset of  $n$  remote sensing time series  $\mathbf{x}_i$  of length  $T$ , which are observed in image pixels<sup>1</sup>. We denote  $\mathbf{H} = [\mathbf{s}_1, \dots, \mathbf{s}_c]$  an ideal  $T \times c$  matrix composed of the  $c$  semantic classes (such as forest, grassland, urban area, ...) corresponding to all pure elements of dataset  $\mathbf{X}$ . The unsupervised unmixing problem can be described with Eq. (4.1) as:

$$\mathbf{X} = \mathbf{MA} : \quad \text{such that } \sum_j \mathbf{A}_{ji} = 1, \forall i \quad \text{and } \mathbf{A}_{ji} \geq 0, \forall i, j \quad (4.1)$$

where  $\mathbf{M}$  is a  $T \times \ell$  matrix composed of the  $\ell$  endmember signatures and  $\mathbf{A}$  is a  $\ell \times n$  matrix called abundance map. Ideally, we wish that one signature could perfectly represent one semantic class ( $\ell = c$  and  $\mathbf{M} = \mathbf{H}$ ), so that the  $i_{th}$  column of matrix  $\mathbf{A}$  (noted  $\mathbf{A}_{\bullet,i}$ ) would be the fractions of time series  $\mathbf{x}_i$  corresponding to pure elements. However, this is quite unlikely in real situation, as the intra-class variability may be important, yielding most of the time multiple endmembers to represent one class ( $\ell > c$ ). As will be discussed in the next section, the relation between endmember set  $\mathbf{M}$  and class set  $\mathbf{H}$  can be estimated from other sources of information. Once estimated, from the abundance  $\mathbf{A}$  one can compute the fraction map of dataset  $\mathbf{X}$  corresponding to the semantic classes  $\mathbf{H}$ .

The overall paper is organized as follows: the next section introduces the state-of-the-art pixel unmixing techniques, which has inspired our time series unmixing research. Then, in section 4.3, our proposed methodology based on non-negative matrix factorization and elastic kernels is presented. Finally, section 4.4 shows some experiments on real datasets.

<sup>1</sup>For clarity, we assume that length  $T$  is the same for all time series but in practice they can be of various lengths.



## 4.2 Related work

The pixel unmixing process is generally composed of two major steps: *endmember extraction* and *abundance estimation*.

- Endmembers  $\mathbf{M}$  are the signatures representing the unique materials in a given image such as mineral, water body, soil, ... In general they have to be extracted from the dataset  $\mathbf{X}$ ;
- Abundances  $\mathbf{A}$  are the percentage of each endmember within a given pixel. Once endmember signatures are extracted, this abundance map can be estimated with various strategies depending on the application (see for example the review paper in [Bioucas-Dias 2012]).

The linear mixing model assumes that the mixed pixel is a linear combination of endmembers based on corresponding abundances. New approaches [Broadwater 2007][Kwon 2006] generalize the linear mixing model by introducing nonlinearities through kernel functions. We will now give a brief introduction for both sorts of methods.

### 4.2.1 Endmember extraction introduction

In general we assume that for large datasets, there exists at least few *pure inputs* that contain only the corresponding material [Nascimento 2005][Winter 1999]. Under the linear model, the mixed data should locate within the simplex constructed by the pure endmember pixels, as shown in [Courty 2014]. The endmembers may then be inferred by selecting some pixels which comprise a maximum volume (MV) simplex to the data. This rather simple and yet powerful idea underlies several geometrical based endmember extraction methods.

The vertex component analysis (VCA) algorithm [Nascimento 2005] enables to solve this fitting problem by iteratively projecting data onto a direction orthogonal to the subspace spanned by the already determined endmembers. The signature corresponding to the extreme of the projection is selected as the new endmember. The popular NFINDR method [Winter 1999] assumes that the volume defined by a simplex of purest pixels is larger than any other volume defined by any other combination of pixels. Therefore this algorithm searches for the set of pixels defining the largest volume by inflating a simplex inside the data. The simplex identification via variable splitting and augmented Lagrangian (SISAL) [Bioucas-Dias 2009] allows the positivity constraint on  $\mathbf{A}$  to be violated, so that some very noisy data could possibly lay outside the simplex (this is a robust version of the maximum volume technique). For above methods, the number of endmember is restricted by the signature dimension.

Among the endmember extraction methods under non-linear mixing model, we are particularly interested on kernel-based methods, as they offers a degree of freedom by choosing various kernel functions (see [Bishop 2006][Bioucas-Dias 2012] for more details). In [Thureau 2012], Thureau et al propose an approach for determining an over-complete endmember set by using only pairwise distances between two signatures. Over-complete endmembers are very interesting since they allow more flexibility when dealing with classes embedding a large variability. As we will explain later, the technique in [Thureau 2012] could be easily transposed to kernel-based method. The over-complete endmember set would

comprise a manifold hull, which better describes the property of dataset  $\mathbf{X}$ . Let us now discuss about abundance estimation.

### 4.2.2 Abundance estimation

Once endmembers  $\mathbf{M}$  for dataset  $\mathbf{X}$  are extracted, the next step consists in estimating the abundance map  $\mathbf{A}$ . This can naturally be obtained by minimizing the Least Square Error (LSE). However, some constraints can be included in the abundance map  $\mathbf{A}$ , as its elements should be non-negative, and the sum of each column in  $\mathbf{A}$  should be one (each column stands for the percentage of each endmember). Another useful constraint is to impose sparsity in matrix  $\mathbf{A}$  since a mixed data is in general made of a small subset of endmembers. For linear mixing model, the abundance estimation problem can then be formulated as Eq. (4.2):

$$\min_{\mathbf{A}_{\bullet,i}} \frac{1}{2} \|\mathbf{M}\mathbf{A}_{\bullet,i} - \mathbf{x}_i\|_F^2 + \lambda \|\mathbf{A}_{\bullet,i}\|_1 \text{ s.t } \mathbf{A} \geq 0 \quad \sum \mathbf{A}_{\bullet,i} = 1 \quad (4.2)$$

This relation, commonly named Constrained Sparse Regression (CSR) (as we apply a sparse convex constraint on  $\mathbf{A}$ ) embeds the Frobenius norm  $F$  defined for a matrix  $M$  composed of elements  $(m_{ij})$  as  $\|M\|_F = \sqrt{\sum_i \sum_j |m_{ij}|^2}$ . The applicability of sparse regression to pixel unmixing is studied in [Iordache 2011].

Broadwater et al [Broadwater 2007] propose to generalize the Eq. (4.2) with kernel techniques. This approach is called kernel fully constrained least square (KFCLS) algorithm. Despite efficient properties, it suffers from a too large computational coast. A more efficient approach is proposed by combining RKHS space gradient and simplex unit projection [Courty 2014]. This will be detailed in the next section.

## 4.3 Methodology for time series unmixing

In this section we describe into details the general methodology for time series unmixing. It is based on simplex volume maximization (for end member extraction) and projection onto it (for abundance estimation). All these steps are realized in a feature space embedded by a specific kernel devoted to time series, introduced in the next paragraph.

### 4.3.1 Kernel time series

There exists a great amount of computer disciplines which are faced with the problem of measuring the similarity between two time series (as for instance financial and stock data analysis, astronomy, medicine, audio processing, ...). Let  $\mathbf{x} = (x_1, \dots, x_m)$  and  $\mathbf{y} = (y_1, \dots, y_k)$  be two time series from a dataset  $\mathbf{X}$  ( $\mathbf{x}, \mathbf{y} \in \mathbf{X}$ ). As shown earlier, a kernel function can be used to describe the dot product in a given feature space. In a context of time series analysis, popular kernels can be classified into three categories:

1. Nonelastic kernels that do not support any time shifting. The usual Gaussian kernel in equation 4.3 stands in this category;

2. Elastic kernels that tolerate time shifting but are not positive definite, such as the Dynamic Time Warping (DTW) based kernel described later;
3. Elastic kernels that tolerate time shifting and are positive definite (p.d.), such as global alignment kernel described in the end of this paragraph.

The kernels in first category generally compare two time series point-by-point, assuming therefore that all series have the same length. In addition, it is based on the idea that all *instantaneous* differences are significant: this means that any shift between two time series will generate errors, preventing from identifying two similar patterns delayed in time. These kernels can be regarded as standard kernels for feature vectors and the most famous one is the Gaussian kernel, as defined in Eq. (4.3):

$$k_{Gauss}(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{1}{\gamma} \|\mathbf{x} - \mathbf{y}\|^2\right). \quad (4.3)$$

The second category of kernels are often similar with the gaussian one but the  $L_2$  distance of relation (4.3) is replaced by an *elastic criteria*, for example based on Dynamic Time Warping (DTW) [Noma 2002][Bahlmann 2002] or Edit Distances [Cortes 2004]. The main idea behind DTW is to assign low values for similar temporal profiles but with different amplitudes or delay on time. It is based on the extraction of a minimal path to switch from a series  $\mathbf{x}$  to another  $\mathbf{y}$  of different lengths.

In practice to evaluate the DTW between series  $\mathbf{x} = [x_1, x_2, \dots, x_m]^T$  and  $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$ , we compute a  $m \times n$  penalty matrix  $P$  such that  $p(i, j)$  represents the cost to switch from  $x_i$  to  $y_j$ . It is generally computed as:

$$p(i, j) = |x_i - y_j|. \quad (4.4)$$

A warping path  $W = w_1, \dots, w_K, K \in [\max(m, n), m + n - 1]$  is the set of coordinates in matrix  $P$  to switch from  $p(1, 1)$  to  $p(m, n)$ . It should respect the following properties:

- $w_1 = (1, 1)$  and  $w_K = (m, n)$  (start and end points);
- $w_{i+1}$  is adjacent to  $w_i$  for all  $i \in [1, K - 1]$  (continuity constraint);
- $(w_{i+1} - w_i)(w_i - w_{i-1}) > 0$  for all  $i \in [2, K - 1]$  (monotonicity constraint).

The DTW distance  $D_{dtw}(\mathbf{x}, \mathbf{y})$  is the warping path that has the minimum cumulated penalties:

$$D_{dtw}(\mathbf{x}, \mathbf{y}) = \min \frac{\sum_{k=1}^K p(w_k)}{K}. \quad (4.5)$$

Efficient implementations to find this optimal path from  $P$  can be obtained using dynamic programming. In practice a distance matrix  $D$  of size  $m \times n$  can be constructed with  $D(1, 1) = p(1, 1)$  and:

$$D(i, j) = \min \begin{cases} D(i-1, j) + p(i-1, j) \\ D(i-1, j-1) + p(i-1, j-1) \\ D(i, j-1) + p(i, j-1) \end{cases}. \quad (4.6)$$

Any value  $D(i, j)$  gives the DTW distance between the  $i$  first elements of  $\mathbf{x}$  and the  $j$  first elements of  $\mathbf{y}$  (and therefore the DTW distance is  $D_{dtw}(\mathbf{x}, \mathbf{y}) = D(m, n)$ ).

Bellow we give an illustration of the described approach. Suppose we want to compare two time series  $\mathbf{x} = (2, 3, 6, 9, 5, 4, 3)$  of size  $m = 7$  and  $\mathbf{y} = (1, 2, 5, 9, 4, 2)$  of size  $n = 6$ . The penalty  $P$  and distance  $D$  matrices read:

$$P = \begin{array}{c|cccccc|c} & & \mathbf{y} & & & & & \\ & & \mathbf{1} & \mathbf{2} & \mathbf{5} & \mathbf{9} & \mathbf{4} & \mathbf{2} & i= \\ \hline \mathbf{2} & \mathbf{1} & \mathbf{0} & 3 & 7 & 2 & 0 & 1 \\ \mathbf{3} & 2 & \mathbf{1} & 2 & 6 & 1 & 1 & 2 \\ \mathbf{6} & 5 & 4 & \mathbf{1} & 3 & 2 & 4 & 3 \\ \mathbf{9} & 8 & 7 & 4 & \mathbf{0} & 5 & 7 & 4 \\ \mathbf{5} & 4 & 3 & 0 & 4 & \mathbf{1} & 3 & 5 \\ \mathbf{4} & 3 & 2 & 1 & 5 & \mathbf{0} & 2 & 6 \\ \mathbf{3} & 2 & 1 & 2 & 6 & 1 & \mathbf{1} & 7 \\ \hline j = & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} & \end{array} \quad (4.7)$$

$$D = \begin{array}{c|cccccc|c} & & \mathbf{y} & & & & & \\ & & \mathbf{1} & \mathbf{2} & \mathbf{5} & \mathbf{9} & \mathbf{4} & \mathbf{2} & i= \\ \hline \mathbf{2} & \mathbf{1} & \mathbf{1} & 4 & 11 & 13 & 13 & 1 \\ \mathbf{3} & 3 & \mathbf{2} & 3 & 9 & 10 & 11 & 2 \\ \mathbf{6} & 8 & 6 & \mathbf{3} & 6 & 8 & 12 & 3 \\ \mathbf{9} & 16 & 13 & 7 & \mathbf{3} & 8 & 15 & 4 \\ \mathbf{5} & 20 & 16 & 7 & 7 & \mathbf{4} & 7 & 5 \\ \mathbf{4} & 23 & 18 & 8 & 12 & \mathbf{4} & 6 & 6 \\ \mathbf{3} & 25 & 19 & 10 & 14 & 5 & \mathbf{5} & 7 \\ \hline j = & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} & \end{array} \quad (4.8)$$

On this basis, the DTW kernel is defined in Eq. (4.9) :

$$k_{DTW}(\mathbf{x}, \mathbf{y}) = \exp(-DTW(\mathbf{x}, \mathbf{y})) \quad (4.9)$$

For implementation, the optimal alignment is determined via dynamical programming [Sakoe 1978]. The DTW kernel has been exploited by many authors for various applications [Noma 2002][Zhou 2010]. However, as discussed in [Cuturi 2011], the DTW kernel is not positive definite, and this contradicts the mathematical foundations of kernel methods from the theory of reproducing kernel Hilbert spaces [Bishop 2006].

From this difficulty, many efforts have been devoted to construct positive definite kernels suitable for time series similarity measurement with elastic principles. In this third category of kernels one can find the one proposed by Marteau [Marteau 2009] based on time warp edit distance which satisfies the triangle inequality, this property being attained by enforcing stiffness adjustment and as a result could be used to define a p.d. kernel. Another kernel has been proposed by Cuturi et al [Cuturi 2007][Cuturi 2011] who use the soft-minimum of the costs of all the alignments that can map a time series onto another to define a positive definite kernel, namely, Global Alignment kernel in Eq. (4.10):

$$k_{GA}(\mathbf{x}, \mathbf{y}) = \sum_{\pi \in \mathcal{A}(\mathbf{x}, \mathbf{y})} \prod_{i=1}^{|\pi|} \kappa(x_{\pi_1(i)}, y_{\pi_2(i)}) \quad (4.10)$$

where  $\mathcal{A}(\mathbf{x}, \mathbf{y})$  includes all possible alignment  $\pi$  between  $\mathbf{x}, \mathbf{y}$ . As proved in [Cuturi 2011],  $k_{GA}$  is p.d. if  $\kappa/(1 + \kappa)$  is p.d. and  $\kappa$  could be Gaussian kernel.

Both kernels in (4.9) and (4.10) will be tested in our application. Let us now turn to the endmember extraction issue.

### 4.3.2 Endmember extraction

Traditional pixel unmixing approaches assume that there exists a single signature which well represents each semantic class. However, as illustrated in the work of [Hemissi 2013], the signatures of basic classes can have natural variability that needs to be taken into consideration during endmember extraction, abundance estimation and any further analysis. In [Goenaga 2013], the authors have applied *local unmixing* strategy to deal with this problem by assuming that only for a small local area, one endmember for each class is sufficient. For the entire image, they merge the results of all local areas to compute the abundance map of all classes. Because of the efficiency of this technique, we propose in this article to use multiple signatures for each class, which means that we would select several endmembers for one single semantic class. These endmembers would in fact describe the intra-class variability. The set of all endmembers compose an *over-complete* basis set for further unmixing problem. Recent research [Yang 2009][Olshausen 1997] indicates that such over-complete basis set is beneficial when a sparse constraint on abundance map is enforced.

The basic idea of our endmember extraction technique is based on the non-negative matrix factorization method we proposed in [Courty 2014] (minimize  $\|\mathbf{X} - \mathbf{M}\mathbf{A}\|^2$  in kernel space). We construct the columns of endmember matrix  $\mathbf{M}$  directly from the dataset  $\mathbf{X}$ : by this way the endmembers compose a manifold hull which characterizes the boundary of the dataset  $\mathbf{X}$ . To construct this matrix we choose the Simplex Volume Maximization (SiVM) [Thureau 2012], due to the following reasons:

1. SiVM samples the boundary. Based on the idea that a linear combination of pure pixels are inside a simplex composed of endmembers, the maximization of the simplex would result in that the extracted points are more likely composed of a single distinct materials [Bioucas-Dias 2012], i.e. endmembers, as illustrated in Fig. 4.2;
2. SiVM can be easily conducted in a RKHS, thus we could use the kernel described in Section 4.3.1 to deal with the properties of time series. It is also possible to derive a simplex with the vertex number to be much larger than the original dimension of time series [Cristianini 2000];
3. It has a linear complexity, thus it can be applied for a large datasets.

In the following we describe more formally this approach.

Let  $\Delta^\ell(\mathbf{M})$  be a  $(\ell - 1)$ -simplicial polytope formed by  $\ell$  columns of  $\mathbf{M}$ , and  $\phi(\mathbf{M}) = [\phi(m_1), \dots, \phi(m_n)]$  be the feature vectors of endmembers in a RKHS. We begin with a review of the original formulation of Thureau and co-workers [Thureau 2012] and its direct transposition to the kernel framework. SiVM tries to maximize  $\text{Vol}(\Delta^\ell(\mathbf{M}))$ , the volume of the simplex spanned by

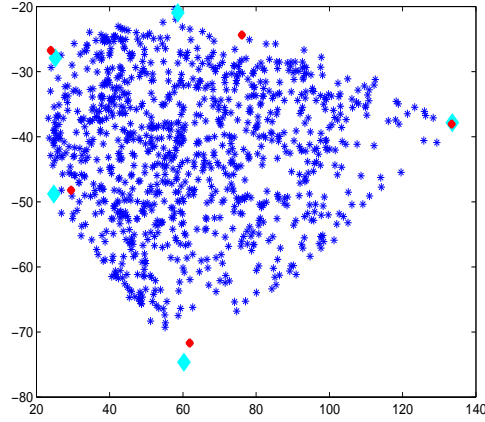


Figure 4.2: **An endmember extraction demo on toy dataset.** The blue points are the observation data of mixed pixels, they contain more than one class substance, the cyan points represent various distinct materials (pure pixel), and the red points are extracted endmember, they are quite close to be pure pixel.

$\Delta^\ell(\mathbf{M})$ , which reads:

$$\text{Vol}(\Delta^\ell(\mathbf{M})) = \sqrt{\frac{-1^\ell \cdot \text{cmd}(\mathbf{M})}{2^{\ell-1}(\ell-1)!}} \quad (4.11)$$

where  $\text{cmd}(\mathbf{M})$  is the Cayley-Menger determinant [Michelucci 2004] of  $\mathbf{M}$  (i.e. the determinant of the matrix accounting for the pairwise distances in  $\mathbf{M}$ ) and  $d_{i,j}^2$  is the square Euclidean distance between elements  $m_i$  and  $m_j$ :

$$\text{cmd}(\mathbf{M}) = \det \begin{pmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 1 & 0 & d_{1,2}^2 & d_{1,3}^2 & \dots & d_{1,\ell}^2 \\ 1 & d_{2,1}^2 & 0 & d_{2,3}^2 & \dots & d_{2,\ell}^2 \\ 1 & d_{3,1}^2 & d_{3,2}^2 & 0 & \dots & d_{3,\ell}^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & d_{\ell,1}^2 & d_{\ell,2}^2 & d_{\ell,3}^2 & \dots & 0 \end{pmatrix}. \quad (4.12)$$

The formulation of  $\text{cmd}(\mathbf{M})$  only involves pairwise distances  $d_{i,j}$ , so that its kernelization appears straightforward: the volume of  $\Delta^\ell(\mathbf{M})$  can be expressed by replacing the Cayley-Menger determinant by that of a matrix accounting for pairwise distances, derived from kernel function:

$$\text{Vol}(\Delta^\ell(\phi(\mathbf{M}))) = \sqrt{\frac{-1^\ell}{2^{\ell-1}(\ell-1)!} \det(\mathbf{P})} \quad (4.13)$$

with  $\mathbf{M}$  similar to the matrix of Eq. (4.12) except that  $\forall i, j \leq \ell$ ,  $d_{i,j}^2$  is replaced by:

$$\begin{aligned} \mathbf{P}_{(i+1)(j+1)} &= \|\phi(m_i) - \phi(m_j)\|^2 \\ &= k(m_i, m_j) + k(m_j, m_j) - 2k(m_i, m_j) \end{aligned} \quad (4.14)$$

The comparison of the volume of the simplices in Eq. (4.13) spanned by all the possible endmember sets is computationally prohibitive (computation complexity  $O(N^\ell)$ ). Thus, the greedy search method for the best simplex, which we propose in [Courty 2014], is used to construct an over-complete endmember set. This method is an improvement of SiVM in [Thureau 2012] and enables then to extract an over-complete set of endmembers representing the variability (in terms of time series) of each pure materials.

Now that we have an over-complete endmember set, its relation with the  $c$  semantic classes has to be determined. Let  $\{\mathbf{M} \leftarrow \mathbf{H}\}$  be a mapping from semantic class  $\mathbf{H}$  to endmember set  $\mathbf{M}$ , and let  $\mathbf{H}_{\bullet i}$  denoting the fraction of  $m_i$  for each class. Because of possible errors in the maximization process for simplex volume maximization, we assume that all endmembers are not necessary pure materials but they also can be composed of mixed elements. In this study from remote sensing images, once extracted, the composition of elements in  $\mathbf{M}$  are set manually using high spatial resolution aerial images in the corresponding area. It can also be set by experts or any additional information, users just need to identify the composition of the few endmembers (compared to the whole dataset) automatically extracted by simplex volume maximization. As  $\ell$  is much smaller than dataset size  $n$ , this strategy could save a great amount of human work. Active learning [Tong 2002] and similar strategies are tools that can also be exploited to reduce labeling workload for training data and will be investigated in a future step. In the following we discuss about abundance estimation.

### 4.3.3 Abundance estimation

Once the endmember set  $\mathbf{M}$  is estimated with the approach described in previous section, one needs to compute the abundance map  $\mathbf{A}$ . As  $\mathbf{M}$  is over-complete, it appears rational to promote sparse solutions of  $\mathbf{A}$ , as done in [Wright 2009]. It also appears natural to rely on the similarity measurement kernels discussed in section 4.3.1 for the estimation of  $\mathbf{A}$ . Given these constraints, we search for the projection of any point  $\phi(\mathbf{x}_i)$  onto the simplex  $\Delta^\ell(\phi(\mathbf{M}))$ , *i.e.* the point of the simplex which minimizes the Euclidean distance in the feature space to  $\phi(\mathbf{x}_i)$ . It can be shown that this amounts to solve the  $n$  independent problems of computing the columns of  $\mathbf{A}$  [Courty 2014]:

$$\begin{aligned} \mathbf{A}_{\bullet i} &= \arg \min_{\mathbf{A}_{\bullet i}} \|\phi(\mathbf{x}_i) - \phi(\mathbf{M})\mathbf{A}_{\bullet i}\|^2 \\ \text{s. t. } &\sum_j \mathbf{A}_{ji} = 1, \mathbf{A}_{ji} \geq 0, \forall i. \end{aligned} \quad (4.15)$$

The constraint  $\sum_j \mathbf{A}_{ji} = 1, \mathbf{A}_{ji} \geq 0, \forall i$  is equivalent to have  $\mathbf{A}_{\bullet i}$  in the unit standard simplex  $\Delta^\ell$ , (in fact  $\mathbf{A}$  encodes the barycentric coordinates of  $\mathbf{X}$  in  $\Delta^\ell(\phi(\mathbf{M}))$ ). At this point, it is possible to force the sparsity of the projection to  $\lambda < \ell$ , without any extra computational cost: we only need to replace the previous constraint by  $\mathbf{G}_{i\bullet} \in \Delta^\lambda$ . Thus, Eq. (4.15) reads:

$$\mathbf{A}_{\bullet i} = \arg \min_{\mathbf{A}_{\bullet i}} \|\phi(\mathbf{x}_i) - \phi(\mathbf{M})\mathbf{A}_{\bullet i}\|^2 \text{ s. t. } \mathbf{G}_{i\bullet} \in \Delta^\lambda \quad (4.16)$$

Instead of considering quadratic programming, such as in [Thureau 2012] or [Kumar 2013], we follow some recent work on the projection on the unit

standard simplex [Kyrillidis 2013], and we propose to use a simple projected gradient descent algorithm to solve Eq. (4.16), which amounts to iterating through different possible solutions of

$$\mathbf{A}_{\bullet i}^{t+1} = \mathcal{P}_\lambda (\mathbf{A}_{\bullet i}^t - \varepsilon_t \nabla (\|\phi(\mathbf{x}_i) - \phi(\mathbf{M})\mathbf{A}_{\bullet i}^t\|^2)) \quad (4.17)$$

where  $t$  is the iteration index,  $\varepsilon_t$  a (possibly varying) step size,  $\nabla(\cdot)$  the gradient operator and  $\mathcal{P}_\lambda(\cdot)$  the projector onto  $\Delta^\lambda$ . This kind of projected gradient descent method has recently shown its computational efficiency and is also endowed with theoretical convergence guarantees [Garg 2009]. As proved in [Courty 2014], the gradient reads (we omit the iteration index  $t$  for clarity):

$$\nabla (\|\phi(\mathbf{x}_i) - \phi(\mathbf{M})\mathbf{A}_{\bullet i}\|^2) = 2(\mathbf{A}_{\bullet i}\mathbf{K}_\ell - \mathbf{k}_{\mathbf{x}_i})$$

where  $K_\ell$  is the Gram Matrix [Bishop 2006] of feature vectors  $\phi(\mathbf{M})$  and  $\mathbf{k}_{\mathbf{x}_i}$  stands the kernel function results between  $\mathbf{x}_i$  and each endmember in  $\mathbf{M}$ .

As for projector  $\mathcal{P}_\lambda(\cdot)$ , we rely on the Greedy Selector and Simplex Projector (GSSP) algorithm of [Kyrillidis 2013] which can be summarized as a two-step procedure: first the coordinates of the vector are sorted by magnitude, and second the  $\lambda$  greatest values are projected on the unit simplex  $\Delta^\lambda$  (while the other vector entries are set to zero). The convergence of the projector is analyzed in [Courty 2014]. The overall procedure is summarized in next section.

#### 4.3.4 Overall procedure

The endmember extraction and abundance estimation are both performed in RKHS space. This enables a better separability of input data and to take advantage of the elastic kernels described in section 4.3.1 to measure the similarity between two time series. After these two steps, we obtain the over-complete endmember set  $\mathbf{M}$ , and the whole dataset  $\mathbf{X}$ 's abundance map  $\mathbf{A}$ . To perform the unmixing, we also need to determine the fraction map of endmember set  $\mathbf{M}$  on semantic classes  $\mathbf{H}$  in the interested area. As mentioned before, the estimation of  $\mathbf{H}$  requires knowledge or information from other sources, such as expert field investigation. In our experiment, it is estimated from human inspection on high resolution aerial image at corresponding endmember location. From all these informations, we could easily merge the intermediate results to obtain the dataset  $\mathbf{X}$ 's fraction map  $\mathbf{F}$  on semantic classes:  $\mathbf{F} = \mathbf{H}\mathbf{A}$ .

In practice, in addition to the data matrix  $\mathbf{X}$ , the user needs to tune the following parameters:  $k$  the choice of suitable kernel,  $\gamma$  the parameters associated with the kernel in Eq. (4.14),  $\ell$  the number of the extracted endmembers and  $\lambda$  the expected sparsity level (with  $\lambda \leq \ell$ ). The parameters could be fixed with cross-validation [Kohavi 1995] and we find in our experiments that these parameters are not very sensitive. All these steps result in the following procedure for time series unmixing:



**Time series unmixing process**

**Inputs:**  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$  the time series dataset  
 $k$  the kernel for similarity measurement  
 $\gamma$  the parameter for kernel  $k$  in Eq. (4.14)  
 $\lambda$  the expected sparse level for the number of endmembers composing each pixels  
 $\ell$  the number of extracted endmembers  
 $\epsilon$  optimization parameters for gradient descent in Eq. (4.17)  
 $\mathbf{H}$  endmembers' fraction map on semantic classes

**Outputs:**  $\mathbf{M}$  the extracted endmember set  
 $\mathbf{A}$ : abundance map of  $\mathbf{X}$  corresponding to  $\mathbf{M}$   
 $\mathbf{F}$ : fraction map of  $\mathbf{X}$  corresponding to pure materials

**step-1:** use SiVM for endmember extraction, obtain an over-complete endmember set  $\mathbf{M}$ .

**step-2:** fix endmember set  $\mathbf{M}$ 's fraction map  $\mathbf{F}$  corresponding on semantic class  $\mathbf{H}$  (by expert investigation in practice).

**step-3:** sparse projections onto the RKHS simplex to obtain  $X$ 's abundance map  $\mathbf{A}$  corresponding to  $\mathbf{M}$ .

**step-4:** analyze the intermediate results to obtain  $X$ 's fraction map  $\mathbf{F} = \mathbf{H}\mathbf{A}$  corresponding to semantic classes  $\mathbf{H}$ .

The manual estimation of semantic class fractions in  $\ell$  endmember locations to fix  $\mathbf{H}$  would take some effort, yet this procedure is a substantial reduction of workload, compared with the fraction estimation of a whole learning data set for endmember extraction in each semantic class. The manual estimation of  $\mathbf{H}$  could be considered as the *labeling* steps for active learning [Ertekin 2007], as we only need to offer groundtruth information on a small number of pixels, selected by endmember extraction algorithm.

## 4.4 Experimental validation

This part aims at validating the proposed time series unmixing framework. Our approach has been tested on real dataset. Endmember extraction technique will be first evaluated and then, a comparison between pixel unmixing results of various approaches will be provided. Before entering into details, let us introduce some general aspects of experimental setup.

### 4.4.1 Experimental settings

The following experimental conditions have been applied:

- In each experiment, we have attempted to evaluate the influence of varying value of  $\ell$  (the number of extracted endmembers) ;
- When using the proposed time series unmixing framework, the two introduced elastic kernels, namely Dynamic Time Warping kernel in Eq. (4.9) and Global Alignment kernel in Eq. (4.10), have been investigated. Their comparison with Gaussian kernel in Eq. (4.3) is also provided to demonstrate the merit of elastic criteria;

- All internal parameters, such as kernel parameter  $\gamma$ , sparse level  $\lambda$ , gradient descent parameter  $\epsilon$  have been determined with cross-validation [Kohavi 1995].

#### 4.4.2 Dataset properties

The dataset is issued from MODIS observation of Belledonne in southeast France and consisted of 562 time series samples manually labeled by experts. The samples are related to 4 types of land covers: Urban, Wood, Forest, Grassland and our pixel dataset has been acquired from the location where exists mainly these 4 classes objects. We have used MODIS satellite observations in 2006, 2008, 2009, as shown in Fig. 4.3. In each year, 36 observations at 10 days time interval [Lecerf 2008] are available. For validation, 4 types of time series are used:

- Time series of Leaf Area Index (LAI) [Tripathi 2014], which is defined as the one-sided green leaf area (for broadleaf canopies) or projected conifer needle surface area to ground area (for coniferous plants) per unit horizontal earth surface area. The satellite-derived LAI value should corresponds to the total green leaves of the all the canopy layers, including the understorey which may represent a very significant contribution, particularly for dense vegetation area, such as forest.
- Time series of Fraction of green vegetation Cover (FCover) [Gutman 1998], which corresponds to the fraction of ground area covered by the green vegetation, the considered area being generally a pixel. Fcover is independent from the illumination direction and sensitive to the vegetation amount.
- Time series of MODIS B1 band data, [Xiong 2006], which is the reflectance data at red band (wavelength 620–670 nm). The spatial resolution is 250m and the solar radiation absorption rate of live leaves is low at this band.
- Time series of MODIS B2 band data, [Xiong 2006], which is the reflectance data at near infrared band (wavelength 841–876 nm). The spatial resolution is 250m, and the solar radiation absorption rate of live leaves is high.

LAI and FCover are biophysical variables, which are widely used by remote sensing community to describe the state of vegetation area. In our experiments, these variables are estimated from MODIS multispectral data by inverting the PROSAIL model [Jacquemoud 2009]. The detailed procedure is presented in [Lecerf 2008][Baret 2007]. The B1 and B2 are original MODIS reflectance observations, their spatial resolution (250m) is highest among all MODIS spectral bands. From MODIS spectral bands, NDVI [Carlson 1997] (Normalized Difference Vegetation Index) can also be computed. However here, most presented results are derived from LAI time series since it has been observed that this biophysical parameter is more stable and precise than NDVI [Dusseux 2014b].

According to our observation, in the time series dataset, the intra-class variability is not negligent (see in Fig. 4.3), which justifies our argument that multiple endmembers should be used to describe a single semantic class. Moreover translation in time can be observed : in 2006/2008, the starting date of MODIS

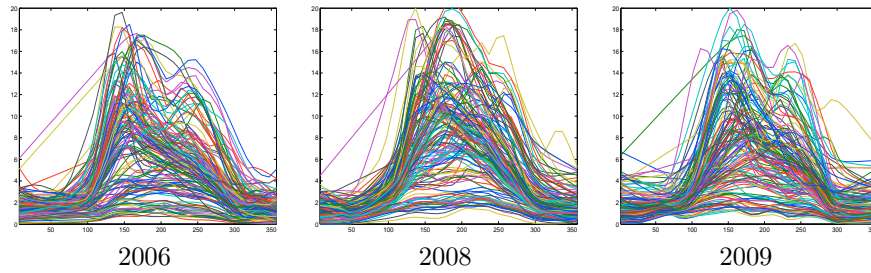


Figure 4.3: **Leaf Area Index evolution of various years in Belledonne, France.** the horizontal axis is date of the year, the vertical axis is Leaf Area Index

observation is on January, 7th while in 2009, this starting date is January, 2nd. In addition, because of different climate in each year, there exists a shift for time series data even in the same location. As will be shown, elastic kernels are able to deal with this time shifting property. Let us now evaluate the endmember extraction procedure.

#### 4.4.3 Endmember extraction validation

As explained before, the goal of our proposed endmember extraction method is to find an over-complete set for representing the various semantic classes of land cover within this ground area. We have tested our proposed approach presented in section 4.3 respectively with Gaussian, DTW, and GA kernels. In Fig. 4.4 (a,b,c) is depicted a visual illustration of our extracted endmembers (shown in red diamond points) with these 3 kernels, when we experiment with LAI time series and the number of endmembers has been set as 22. As the time series length is 36, in order to visually demonstrate the results, all time series have been projected onto 2-D space with Principal Component Analysis (PCA). It can be observed that our extracted endmembers with DTW kernel in (Fig. 4.4b), GA kernel (Fig. 4.4c) could effectively represent the dataset structure. It also appears that the performances of DTW kernel and GA kernel seem better than Gaussian kernel (result in Fig. (4.4a)) since their distribution seems to be more consistent with the initial one. This confirms our previous suggestion that elastic kernels appear better suited for measuring similarity between time series. In order to make a comparison with other existing endmember extraction methods, we have also implemented 3 state-of-the-art approaches: VCA [Nascimento 2005], NFINDR [Winter 1999] and SISAL [Bioucas-Dias 2009]. All the three method would strive to extract a endmember simplex, while other pixels would lay within the simplex as much as possible. The results are illustrated in Fig. (4.4e-4.4f). From a visual comparison, it appears that the presented approach seems to be much more adapted to time series endmember extraction, for similar reasons than previously (in practice SISAL has extracted endmembers far from the original dataset).

In order to make a quantitative comparison, we have manually determined class fraction of extracted endmembers with aerial high resolution images. For SISAL algorithm, the extracted endmembers do not necessarily exist in original dataset (Fig. (4.4f)), as the endmember signatures are calculated by solving a

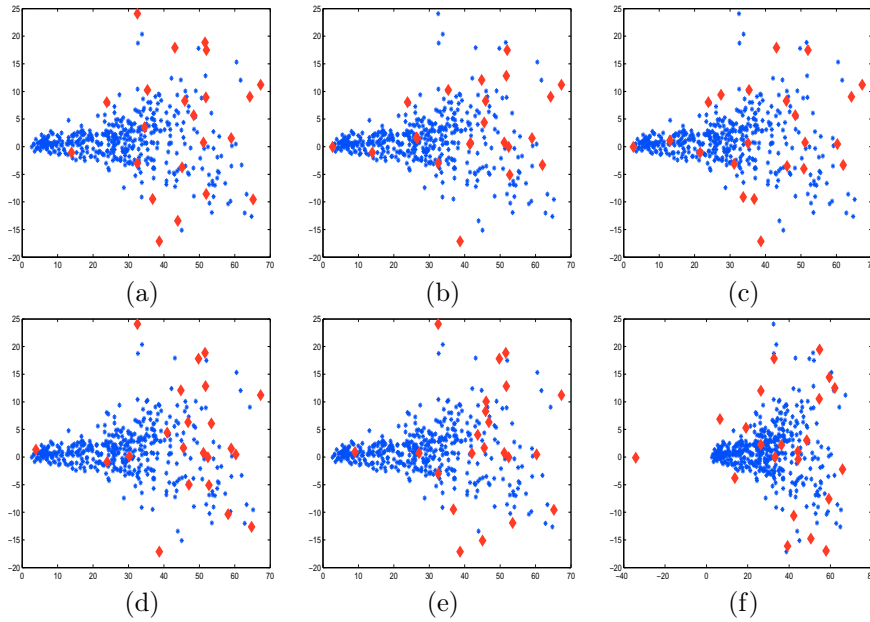


Figure 4.4: **An illustration of endmember extraction results of various methods**, time series are projected to 2D with PCA, and red diamond points are extracted endmembers. (a) SiVM with Gaussian kernel, (b) SiVM with DTW kernel, (c) SiVM with GA kernel, (d) VCA, (e) NFINDR, (f) SISAL.

sequence of convex optimization problems using the method of augmented Lagrange multipliers, therefore the class fraction of nearest time series in dataset is used. A good endmember set should properly describe all semantic classes and it is better that the endmembers are evenly distributed among various classes. In the light of this, we have used two criteria, entropy [Wang 1984] and gini impurity [Fayyad 1992]. The high value of these two criteria would indicate all semantic classes have their share in the extracted endmember set. Table 4.1 presents the quantitative results. From this table, one can observe two phenomena:

1. Our proposed method with various kernels has outperformed the 3 state-of-the-art algorithms for constructing a over-complete endmember set;
2. Elastic kernels (DTW and GA) have demonstrated their merits for time series, compared with Gaussian kernel.

These observations confirm our choices and demonstrate the ability of the technique for endmember extraction related to time series. In the following we evaluate the pixel unmixing approach.

#### 4.4.4 Pixel unmixing result evaluation

For pixel unmixing problem, the ultimate objective is to retrieve all existing classes and their fraction within a pixel area. In contrast, land cover classification only requires the algorithm to assign a single class label to each pixel.

Table 4.1: **Quantitative comparison of endmember extraction** performances for the six tested approaches

Criterion	VCA	NFINDR	SISAL	Gaussian	DTW	GA
Entropy	1.7150	1.6923	1.7044	1.8125	1.8582	<b>1.8656</b>
Gini impurity	0.6786	0.6694	0.6737	0.7005	0.7051	<b>0.7117</b>

Table 4.2: The Salient Component Detect Rate for LAI time series unmixing. Gaussian, DTW, GAK represent results of our proposed methods, with Gaussian, DTW, GAK, kernels and  $\ell$  stands for various endmember numbers.

Method	$\ell = 4$	$\ell = 10$	$\ell = 16$	$\ell = 22$	$\ell = 28$
<i>VCA</i>	17.47%	23.35%	24.95%	26.37%	27.11%
<i>NFINDR</i>	16.60%	27.10%	32.44%	33.86%	33.66%
<i>SISAL</i>	15.16%	24.80%	28.20%	30.56%	31.61%
<i>Gaussian</i>	18.41%	28.02%	32.29%	35.32%	37.02%
<i>DTW</i>	<b>25.81%</b>	<b>37.19%</b>	41.86%	46.44%	47.69%
<i>GA</i>	19.93%	32.92%	<b>42.64%</b>	<b>48.04%</b>	<b>48.50%</b>

At the moment, it remains too challenging to obtain an accurate estimation of fraction coefficients of existing classes within one pixel. In many applications, we would only require the detection of salient classes. The salient class within a pixel are defined as the class whose fraction in this pixel are more than 25%. If our method successfully detects all salient classes, and does not give any wrong salient class for a pixel, we would associate this as a good pixel unmixing result. For this reason, we take *salient component detection rate* as our quantitative comparison criterion between different pixel unmixing methods. Table 4.2 gives the quantitative results of salient component detection rate on LAI time series. Gaussian, DTW, GA represent our proposed method with respective kernels. VCA, NFINDR, SISAL are endmember extraction methods, for them, the supplementary abundance estimation method is chosen as constrained sparse regression (CSR) in Eq. (4.2). The other steps remain identical as our method. From this table, it is easy to show that our method have outperformed all state-of-the-art ones and that elastic kernels perform efficiently. Moreover, Global Alignment kernels perform better as soon as the number of endmembers is more that  $\sim 10$ , as also illustrated in Fig. (4.5). As the value of  $\ell$  increases, the pixel unmixing results become better. This phenomenon also validates the beneficiary of over-complete endmember set ( $\ell = 28$ ) and it is expected that a small endmember ( $\ell = 4$ ) set is unable to describe the intra-class variability.

We have also made an experimental inspection on the performances of various sources of time series. The time series sources generally fall into two categories:

1. Biophysical variables, which describe the state of vegetation area. In this

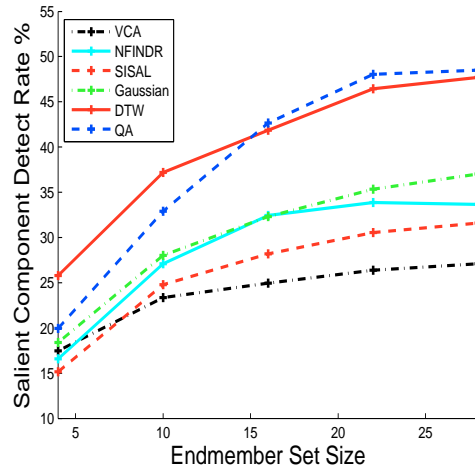


Figure 4.5: Time series evaluation of various methods with different end-member set size.

Table 4.3: The Saliency Component Detect Rate comparison between MODIS B1, B2, FCover, LAI time series ( $\ell = 28$ ).

Source	VCA	NFINDR	SISAL	Gaussian	DTW	GA
B1	19.54%	19.89%	20.70%	30.50%	29.95%	31.91%
B2	23.75%	28.83%	22.14%	33.27%	33.63%	34.70%
FCover	26.19%	31.35%	24.80%	35.41%	40.04%	37.19%
LAI	<b>27.11%</b>	<b>33.66%</b>	<b>31.61%</b>	<b>37.02%</b>	<b>47.69%</b>	<b>48.50%</b>

paper, we take LAI and FCover as examples ;

2. MODIS reflectance band observations, which describe the spectral reflectance properties of the ground. In this paper, we take MODIS B1 and B2 band data as examples.

The pixel unmixing results are shown in Table. 4.3. In summary, biophysical variables appear as better indexes for time series unmixing. This is probably explained by the fact that the original reflectance data is more influenced by illumination and noise and in contrast, biophysical variables are more stable. As FCover saturates in dense vegetation area and LAI incorporates information of understory layer leaves, the result using LAI are naturally better than FCover, which is consistent with similar observations related to the efficiency of LAI biophysical parameter [Carlson 1997].

## 4.5 Conclusion

In this paper, we have studied the possibility of remote sensing time series unmixing. The objective was to estimate the existing classes fraction information

in each pixel from the entire time series, which is to our opinion a new approach in remote sensing. To this end we have proposed a framework based on the exploitation of non-negative matrix factorization and elastic kernels. An over-complete endmember set (multiple endmembers are used to represent a single semantic class) has been extracted on which a sparse convex constraint is enforced for the definition of the abundance matrix. The experimental part has both validated our endmember extraction strategy and compared the performance of various pixel unmixing techniques. The presented preliminary results are quite promising and encouraging for future operational time series unmixing processes.

# Conclusion générale





# Conclusion & Perspectives

---

Dans ce chapitre, nous présentons un résumé des contributions ainsi que des perspectives associées à ce travail.

## 5.1 Résumé des contributions

Les images de télédétection sont précieuses pour le suivi de l'environnement et l'arrivée de plus en plus importante de séries temporelles autorise maintenant l'accès à de l'information de plus en plus précise sur l'usage des sols (état des cultures, modes de gestion, ...). Cependant, analyser et interpréter des séries temporelles d'images satellites n'est pas une tâche aisée. Cette thèse a proposé plusieurs contributions méthodologiques sur cet aspect. Celles-ci sont résumées ci-dessous:

- Proposition et validation d'un cadre d'assimilation de données pour la reconstruction de variables biophysiques en combinant les LAI estimés sur des images MODIS avec le modèle de croissance de plantes GreenLab (chapitre 2).

Les défauts des approches traditionnelles d'estimation de variables biophysiques est qu'elles ne s'appliquent qu'à une seule image. Par conséquent, elles ne sont pas insensibles aux différents artefacts présents (couverture nuageuse par exemple). Bien que des techniques d'interpolation temporelle existent, nous avons proposé une approche originale visant à combiner ces observations bruitées avec un modèle de croissance de plantes, GreenLab, développé en partie au LIAMA à Pékin. Nous avons fait une simplification de ce modèle en ignorant l'aspect structurel des plantes pour ne conserver que les informations statistiques liées à l'évolution de paramètres biophysiques (LAI, biomasse, ...). Cela autorise le modèle GreenLab à être appliqué aux données de télédétection. Le cadre de l'assimilation de données retenu est basé sur des méthodes particulières, qui sont performantes pour gérer des modèles dynamiques complexes et non linéaires, comme c'est le cas du modèle GreenLab. Les résultats, sur des données synthétiques et réelles, ont mis en avant l'intérêt de cette approche, comparativement à d'autres méthodes de l'état de l'art. On peut notamment citer que notre approche permet une meilleure estimation des variables biophysiques sur des données de télédétection bruitées et que par ailleurs, il est possible d'en déduire une information interne sur l'état des plantes (répartition de la biomasse dans ses différents organes).

- Dans un second temps, nous nous sommes intéressés au problème de démélangeage de séries temporelles (chapitres 3 et 4).

A l'heure actuelle, nous ne connaissons pas de techniques établissant un démélangeage sur des séries temporelles, ce problème étant encore difficile et récent. Pour l'affronter, nous avons exploré des techniques avancées de démélangeage, notamment des approches de factorisation non-négative dans un espace de grande dimension engendré par le RKHS (Reproducing Kernel Hilbert Space). Cette technique est présentée dans le chapitre 3 et consiste en l'extraction d'endmembers en recherchant les points qui maximisent un simplexe dans l'espace engendré par le RKHS. Ensuite les autres points sont projetés sur ce simplexe pour estimer les abondances dans chaque mélange. Dans le chapitre 4, l'application de cette technique aux séries d'images satellites a été proposée. Pour cela, nous avons utilisé des noyaux élastiques qui permettent de gérer proprement les spécificités des séries temporelles (décalage notamment), donnant des résultats très encourageants. Etant donnée la variabilité des séries associées à chaque classe, nous avons opté pour l'extraction d'un ensemble d'endmembers "sur-complet" (plus important que le nombre réel). L'affectation du pourcentage d'éléments purs associés à chaque endmember (on n'est pas garanti qu'ils soient tous purs) se fait à l'aide d'une image à très haute résolution spatiale. La projection des séries temporelles sur ces endmembers permet ensuite de retrouver le pourcentage d'éléments purs associés à chaque pixels. Ainsi, seuls quelques points (correspondant aux endmembers) sont à labéliser manuellement, ce qui réduit énormément l'intervention humaine.

## 5.2 Perspectives

Un grand nombre de perspectives sont offertes par ce travail.

### Reconstruction de variables biophysique

Le modèle GreenLab utilisé a été simplifié à l'échelle de MODIS et appliqué à quelques types de cultures, dans le but de réaliser une preuve de concept (c'est la première expérience du genre) plutôt que de fournir un outil opérationnel. Bien évidemment, de nombreuses améliorations restent à faire. Celles-ci sont listées ci-dessous:

- L'extension à un nombre plus conséquent de couverts végétaux est une première étape. La question de la structure topologique des plantes se pose notamment pour des cas plus complexes (forêt tropicale par exemple) pour lesquels l'introduction des termes structurels de GreenLab pourrait être envisagée. L'utilisation de techniques de sélection de modèles serait dans ce cas d'une grande aide [Burnham 2002] ;
- La technique d'assimilation de données proposée a été appliquée sur du LAI car cette variable semblait la plus pertinente. Néanmoins, l'adapter à d'autres variables comme le fCover ou fApar représente un intérêt certain;
- Enfin, il serait judicieux de reproduire les assimilations sur plusieurs années afin d'analyser des grandes tendances ou d'observer les influences climatiques (années de sécheresses) sur les variables biophysiques.

## Démélangeage de séries temporelles

Ici aussi, la technique utilisée (Non-Negative Matrix Factorization associée à une méthode de simplexe et combinée à des noyaux élastiques) est assez novatrice et les résultats préliminaires sont très prometteurs. Cependant, il reste un certain nombre de questions à résoudre :

- Les noyaux élastiques adaptés aux séries temporelles peuvent être affinés, notamment afin d'améliorer leur robustesse aux données bruitées. Par ailleurs la prise en compte du lien spatial dans les fonctions à noyau semble aussi judicieuse, surtout dans la perspective de travailler avec des données SENTINEL-2 dont la résolution spatiale sera grandement améliorée;
- L'utilisation d'un ensemble d'endmembers sur-complet a contribué à améliorer les performances des techniques de démélangeage. Cependant, cela nécessite une intervention (limitée tout de même) humaine afin de quantifier leur fraction réelle correspondant aux classes sémantiques. Estimer automatiquement la taille de cet ensemble d'endmembers est alors un axe de recherche important.

## Quelques mots pour conclure

Cette thèse s'est déroulée dans un contexte d'une collaboration franco-chinoise. Elle s'est inscrite également au carrefour de plusieurs communautés scientifiques : modélisation de plantes, télédétection, apprentissage statistique. Malgré ces différences scientifiques et culturelles, ce travail est allé à son terme. J'aimerais remercier et saluer les personnes que j'ai pu rencontrer, à Pékin comme à Rennes.



# Part IV

## Annexe



# GreenLab plant growth model

---

The objective of the GreenLab model [Qi 2010] is to predict plant growth in a given environment based on physiological knowledge. This model is designed to provide a dynamic representation of the morphogenesis and architecture of a plant on the basis of a small set of recurrent mathematical equations and metamorphic rules. It is executed at time steps corresponding to organogenetic growth cycles (GC). The elementary units in GreenLab are metamers, which would consist of a node to which one or several leaves are attached, a subtending internode, axillary buds at the base of leaves and flower buds if they exist. The set of metamers produced during a growth cycle is defined as the growth unit (GU) and A GU could consist of one or several metamers. The organs evolve on their individual, finite, thermal-time axes, and are therefore characterized by physiological age (PA). The chronological age (CA) for entire plant is kept to control the organogenesis process.

The GreenLab model is a combination of two coupling sub-models, 1) the structural model which decides the emergence of metamers in each growth cycle, therefore the plant architecture is determined. 2) the functional model which describes the biomass production and its partition among organs.

## Structural model

The plant architecture is generated by an dual scale automaton, as illustrated in Fig. 6.1 of [YAN 2004]. The evolution of a specific physiological age metamer is represented by a microstate automaton. The GU produced in a given GC is a set of new metamers and called a macrostate. The macrostate could be implemented repeatedly, resulting in a bearing axis (BA). The structural information about plant topological evolution could be stored in macrostate and microstate automaton (namely dual scale automaton), as shown in Fig. 6.1. The structural information includes the number of appearing organs, and their respective locations on the plant at any given growth cycle. The parameters of dual scale automaton could be measured with field observation data [Guo 2006].

## Functional model

It is assumed that the biomass fabricated by photosynthesis during a growth cycle would be gathered in a common pool [YAN 2004], then the biomass allocation for living organs is determined according to the source-sink relations that settle their competitions for biomass acquisition. The biomass generation is modelled as a leaf size dependent function (Eq. (2.1)). This empirical and intuitive formula could well adapt to the real case through parameter optimisation. For each living organ, it has a demand for biomass, depending on its sink strength, physiological age (PA) and chronological age (CA) [Qi 2010]. Biomass is allocated to organs according to their demand ratio over plant total demand.



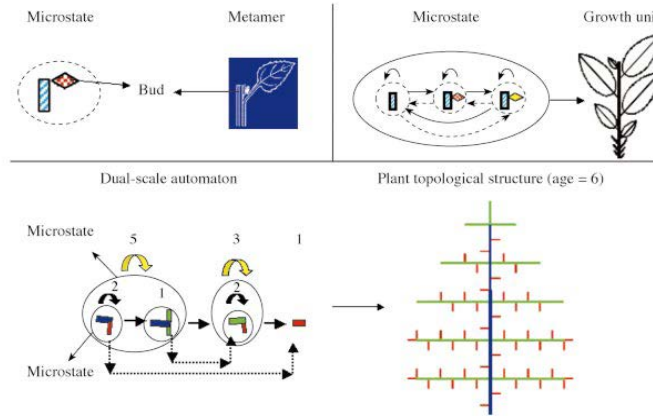


Figure 6.1: Dual scale automaton.

(Eq. (2.4)). The growth of each organ is the accumulation of biomass, similar to Eq. (2.6). Many parameters for functional model have to be estimated with optimisation techniques [Guo 2006].

### GreenLab versions

The number of organs is a key factor, which drives the plant demand for biomass, organ size and the growth evolution of a plant. Since the first proposition of GreenLab in 1998, three versions have been implemented by other researchers [Qi 2010].

1. **Deterministic version** [Zhan 2003][YAN 2004], this is the earliest and simplest version of GreenLab, the number of metamers and active buds at any growth cycle are constant and predefined by the dual scale automaton.
2. **Stochastic version** [Kang 2008], A stochastic variation is introduced to the deterministic GreenLab model, Now the dual scale automaton only determines a potential plant architecture. Whether buds grow or not depends on certain probabilities, which decide the final plant topology.
3. **Mechanistic version** [Letort 2008], The feedback influence between the physiological processes (e.g. biomass fabrication) and the plant development (e.g. metamer production) is modelled. The number of metamers appearing at each growth cycle is dependent on the ratio of plant biomass to plant demand. The plant evolution is more variable, based on this mechanism.

Our implementation of GreenLab model has incorporated the ideas of stochastic and mechanistic versions GreenLab. The probability is introduced through Eq. (2.19). and the correction step (Eq. (2.12)) updates the plant state, thus would influence future biomass production. For large scale remote sensing applications on crop fields, the topological structure for individual plant is not important, therefore a significant simplification of GreenLab is performed to adapt the model to remote sensing data, as done in section 2.2.

The extensive introduction and mathematical equations of GreenLab model could be found in [Qi 2010][YAN 2004].



# Publications

---

## Articles de revues

- [1] Xing Gong and Thomas Corpetti. Adaptive window size estimation in unsupervised change detection. In *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2):991-1003, 2013.
- [2] Nicolas Courty, Xing Gong, Jimmy Vandel, and Thomas Burger. Saga : Sparse and geometry-aware non-negative matrix factorization through non-linear local embedding. In *Machine Learning*, 97(1-2) :205-226, 2014.
- [3] Pauline Dusseux, Xing Gong, Laurence Hubert-Moy, and Thomas Corpetti. Identification of grassland managements from lai time series using dynamic time warping. In *Journal of Applied Remote Sensing*, 8(1) :083559-083559, 2014

## Conférences internationales

- [4] Xing Gong, Thomas, Corpetti. Adaptive patches for change detection. In *IEEE International Conference on Image Processing, ICIP'11*, pages 2789-2792, Brussels, Belgium, September 2011.
- [5] Xing Gong, Thomas, Corpetti. Local patches for change detection in very high resolution remote sensing images. In *IEEE Int. Geoscience and Remote Sensing Symp, IGARSS'11*, pages 237-240, Vancouver, Canada, July 2011
- [6] Xing Gong, Thomas Corpetti, MengZhen Kang, BaoGang Hu and Laurence Hubert-Moy. Reconstruction of leaf area time series using data assimilation on the greenlab plant growth model and remote sensing. In *International Conference on Functional-Structural Plant Models*, pages 223-225, Saariselkä, Finland, June 2013
- [7] Pauline Dusseux, Xing Gong, Thomas Corpetti, Laurence Hubert-Moy, and Samuel Corgne. Contribution of radar images for grassland management identification. In *SPIE Int. Conf. on Remote Sensing for Agriculture, Ecosystems, and Hydrology XIV*, pages 853104-853104, Edinburgh, UK, September 2012

## Articles en soumission/révision

- [8] Xing Gong, Thomas Corpetti, MengZheng Kang, BaoGang Hu, and Laurence Hubert-Moy. Reconstruction of lai time series from medium resolution remote sensing by assimilation of the greenlab plant growth model. In *IEEE Transactions on Geoscience and Remote Sensing*, in revision 2014.

- [9] Xing Gong, Thomas Corpetti, Nicolas Courty, BaoGang Hu and Laurence Hubert-Moy Time series unmixing with kernel space non-negative matrix factorization. In *IEEE Transaction of geoscience and remote sensing*, submitted

# Bibliography

- [Adams 1986] John B Adams, Milton O Smith et Paul E Johnson. *Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site*. Journal of Geophysical Research: Solid Earth (1978–2012), vol. 91, no. B8, pages 8098–8112, 1986.
- [Arora 2012] S. Arora, R. Ge, R. Kannan et A. Moitra. *Computing a nonnegative matrix factorization—provably*. In Proceedings of the 44th symp. on Theory of Computing, pages 145–162. ACM, 2012.
- [Bache 2013] K. Bache et M. Lichman. *UCI Machine Learning Repository*, 2013.
- [Bahlmann 2002] Claus Bahlmann, Bernard Haasdonk et Hans Burkhardt. *Online handwriting recognition with support vector machines—a kernel approach*. In Frontiers in Handwriting Recognition, 2002. Proceedings. Eighth International Workshop on, pages 49–54. IEEE, 2002.
- [Bain 2009] Alan Bain et Dan Crisan. Fundamentals of stochastic filtering, volume 3. Springer, 2009.
- [Bandeira 2012] A. Bandeira, M. Fickus, D. Mixon et P. Wong. *The road to deterministic matrices with the restricted isometry property*. arXiv preprint arXiv:1202.1234, 2012.
- [Baret 2007] Frédéric Baret, Olivier Hagolle, Bernhard Geiger, Patrice Bicheron, Bastien Miras, Mireille Huc, Béatrice Berthelot, Fernando Niño, Marie Weiss, Olivier Samain et al. *LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: Part 1: Principles of the algorithm*. Remote sensing of environment, vol. 110, no. 3, pages 275–286, 2007.
- [Bioucas-Dias 2009] José M Bioucas-Dias. *A variable splitting augmented Lagrangian approach to linear spectral unmixing*. In Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS’09. First Workshop on, pages 1–4. IEEE, 2009.
- [Bioucas-Dias 2012] José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader et Jocelyn Chanussot. *Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches*. Selected Topics in Applied Earth Obs. & Remote Sensing, IEEE J., vol. 5, no. 2, pages 354–379, 2012.
- [Bishop 2006] Christopher M Bishop et al. Pattern recognition and machine learning, volume 1. springer New York, 2006.
- [Bonet 1999] J Bonet et T-SL Lok. *Variational and momentum preservation aspects of smooth particle hydrodynamic formulations*. Computer Methods in applied mechanics and engineering, vol. 180, no. 1, pages 97–115, 1999.

- [Boyd 2004] Stephen Poythress Boyd et Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [Brisson 2003] N Brisson, C Gary, E Justes, R Roche, B Mary, D Ripoché, D Zimmer, J Sierra, P Bertuzzi, P Burgeret *al.* *An overview of the crop model STICS*. *European Journal of agronomy*, vol. 18, no. 3, pages 309–332, 2003.
- [Broadwater 2007] J Broadwater, Rama Chellappa, Amit Banerjee et Philippe Burlina. *Kernel fully constrained least squares abundance estimates*. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pages 4041–4044. IEEE, 2007.
- [Buciu 2008] I. Buciu, N. Nikolaidis et I. Pitas. *Nonnegative matrix factorization in polynomial feature space*. *Neural Networks, IEEE T.*, vol. 19, no. 6, pages 1090–1100, 2008.
- [Burnham 2002] Kenneth P Burnham et David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, 2002.
- [Burrus 1998] C Sidney Burrus, Ramesh A Gopinath, Haitao Guo, Jan E Odegaard et Ivan W Selesnick. *Introduction to wavelets and wavelet transforms: a primer*, volume 23. Prentice hall Upper Saddle River, 1998.
- [Cai 2011] D. Cai, X. He, J. Han et T.S. Huang. *Graph Regularized Nonnegative Matrix Factorization for Data Representation*. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pages 1548–1560, 2011.
- [Camacho 2013] Fernando Camacho, Jesús Cernicharo, Roselyne Lacaze, Frédéric Baret et Marie Weiss. *GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference products*. *Remote Sensing of Environment*, vol. 137, pages 310–329, 2013.
- [Camps-Valls 2005] Gustavo Camps-Valls et Lorenzo Bruzzone. *Kernel-based methods for hyperspectral image classification*. *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 6, pages 1351–1362, 2005.
- [Candes 2008] E. Candes. *The restricted isometry property and its implications for compressed sensing*. *Comptes Rendus Mathématique*, vol. 346, no. 9–10, pages 589–592, Mai 2008.
- [Carlson 1994] Toby N Carlson, Robert R Gillies et Eileen M Perry. *A method to make use of thermal infrared temperature and NDVI measurements to infer surface soil water content and fractional vegetation cover*. *Remote Sensing Reviews*, vol. 9, no. 1-2, pages 161–173, 1994.
- [Carlson 1997] Toby N Carlson et David A Ripley. *On the relation between NDVI, fractional vegetation cover, and leaf area index*. *Remote sensing of Environment*, vol. 62, no. 3, pages 241–252, 1997.
- [Çivril 2009] A. Çivril et M. Magdon-Ismail. *On selecting a maximum volume sub-matrix of a matrix and related problems*. *Theoretical Computer Science*, vol. 410, no. 47–49, pages 4801 – 4811, 2009.

- [Chan 2009] Tsung-Han Chan, Chong-Yung Chi, Yu-Min Huang et Wing-Kin Ma. *A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing*. Signal Processing, IEEE Transactions on, vol. 57, no. 11, pages 4418–4432, 2009.
- [Chen 1997] Jing M Chen, Paul M Rich, Stith T Gower, John M Norman et Steven Plummer. *Leaf area index of boreal forests: Theory, techniques, and measurements*. Journal of Geophysical Research: Atmospheres (1984–2012), vol. 102, no. D24, pages 29429–29443, 1997.
- [Combal 2003] B Combal, F Baret, M Weiss, A Trubuil, D Mace, A Pragnere, R Myneni, Y Knyazikhin et L Wang. *Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem*. Remote sensing of environment, vol. 84, no. 1, pages 1–15, 2003.
- [Cortes 2004] Corinna Cortes, Patrick Haffner et Mehryar Mohri. *Rational kernels: Theory and algorithms*. The Journal of Machine Learning Research, vol. 5, pages 1035–1062, 2004.
- [Courtier 1987] Philippe Courtier et Olivier Talagrand. *Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results*. Quarterly Journal of the Royal Meteorological Society, vol. 113, no. 478, pages 1329–1347, 1987.
- [Courtier 1994] PHILIPPE Courtier, J-N Thépaut et A Hollingsworth. *A strategy for operational implementation of 4D-Var, using an incremental approach*. Quarterly Journal of the Royal Meteorological Society, vol. 120, no. 519, pages 1367–1387, 1994.
- [Courty 2011] N. Courty, T. Burger et L. Johann. *PerTurbo: a new classification algorithm based on the spectrum perturbations of the Laplace-Beltrami operator*. In ECML/PKDD, volume 1, pages 359–374, 2011.
- [Courty 2013] N. Courty et T. Burger. *A kernel view on manifold sub-sampling based on Karcher variance optimization*. In Geometric Science of Information, pages 751–758. Springer, 2013.
- [Courty 2014] Nicolas Courty, Xing Gong, Jimmy Vandael et Thomas Burger. *SAGA: Sparse And Geometry-Aware non-negative matrix factorization through non-linear local embedding*. Machine Learning, vol. 97, no. 1-2, pages 205–226, 2014.
- [Cristianini 2000] Nello Cristianini et John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.
- [Cutler 1994] Adele Cutler et Leo Breiman. *Archetypal analysis*. Technometrics, vol. 36, no. 4, pages 338–347, 1994.
- [Cuturi 2007] Marco Cuturi, J-P Vert, Øystein Birkenes et Tomoko Matsui. *A kernel for time series based on global alignments*. In Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, volume 2, pages II–413. IEEE, 2007.



- [Cuturi 2011] Marco Cuturi. *Fast global alignment kernels*. In Proceedings of the 28th International Conference on Machine Learning (ICML-11), pages 929–936, 2011.
- [DeFries 1994] RS DeFries et JRG Townshend. *NDVI-derived land cover classifications at a global scale*. International Journal of Remote Sensing, vol. 15, no. 17, pages 3567–3586, 1994.
- [DIMET 1986] FRANÇOIS-XAVIER LE DIMET et Olivier Talagrand. *Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects*. Tellus A, vol. 38, no. 2, pages 97–110, 1986.
- [Ding 2010] C. Ding, T. Li et M. Jordan. *Convex and Semi-Nonnegative Matrix Factorizations*. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 32, no. 1, pages 45–55, Janvier 2010.
- [Donoho 2003] David L Donoho et Victoria C Stodden. *When does non-negative matrix factorization give a correct decomposition into parts?* In NIPS, 2003.
- [Doucet 2000] Arnaud Doucet, Simon Godsill et Christophe Andrieu. *On Sequential Monte Carlo Sampling Methods for Bayesian Filtering*. STATISTICS AND COMPUTING, vol. 10, no. 3, pages 197–208, 2000.
- [Doucet 2001] Arnaud Doucet. Sequential monte carlo methods. Wiley Online Library, 2001.
- [Dusseux 2014a] P. Dusseux. *Exploitation de séries temporelles d’images satellites à haute résolution spatiale pour le suivi des prairies en milieu agricole*. Thèse de géographie, Université Rennes 2, 2014.
- [Dusseux 2014b] P. Dusseux, X. Gong, L. Hubert-Moy et T. Corpetti. *Identification of grassland management practices from leaf area index time series*. Journal of Applied Remote Sensing, vol. 8, no. 1, 2014.
- [Ertekin 2007] Seyda Ertekin, Jian Huang, Leon Bottou et Lee Giles. *Learning on the border: active learning in imbalanced data classification*. In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, pages 127–136. ACM, 2007.
- [Esser 2012] E. Esser, M. Moller, S. Osher, G. Sapiro et J. Xin. *A Convex Model for Nonnegative Matrix Factorization and Dimensionality Reduction on Physical Space*. IEEE Trans. on Image Processing, vol. 21, no. 7, pages 3239–3252, 2012.
- [Evensen 1994] Geir Evensen. *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*. Journal of Geophysical Research: Oceans (1978–2012), vol. 99, no. C5, pages 10143–10162, 1994.
- [Fayyad 1992] Usama M Fayyad et Keki B Irani. *The attribute selection problem in decision tree generation*. In AAAI, pages 104–110, 1992.

- [Gamon 1995] John A Gamon, Christopher B Field, Michael L Goulden, Kevin L Griffin, Anne E Hartley, Geeske Joel, Josep Peñuelas et Riccardo Valentini. *Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types*. Ecological Applications, pages 28–41, 1995.
- [Gao 1996] Bo-Cai Gao. *NDWI-a normalized difference water index for remote sensing of vegetation liquid water from space*. Remote sensing of environment, vol. 58, no. 3, pages 257–266, 1996.
- [Garg 2009] R. Garg et R. Khandekar. *Gradient descent with sparsification: an iterative algorithm for sparse recovery with restricted isometry property*. In ICML, pages 337–344, 2009.
- [Gärtner 1999] B. Gärtner. *Fast and Robust Smallest Enclosing Balls*. In Proceedings of the 7th Annual European Symposium on Algorithms, ESA '99, pages 325–338, 1999.
- [Gillis 2012] Nicolas Gillis. *Sparse and unique nonnegative matrix factorization through data preprocessing*. JMLR, vol. 13, pages 3349–3386, 2012.
- [Gillis 2013] N. Gillis et S. Vavasis. *Fast and Robust Recursive Algorithms for Separable Nonnegative Matrix Factorization*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2013.
- [Gobron 2008] Nadine Gobron, Bernard Pinty, Ophélie Aussedat, Malcolm Taberner, Olga Faber, Frédéric Mélin, Thomas Lavergne, Monica Robustelli et Paul Snoeij. *Uncertainty estimates for the FAPAR operational products derived from MERISqX Impact of top-of-atmosphere radiance uncertainties and validation with field data*. Remote Sensing of Environment, vol. 112, no. 4, pages 1871–1883, 2008.
- [Goenaga 2013] Miguel A Goenaga, Maria C Torres-Madronero, Miguel Velez-Reyes, Skip J Van Bloem et Jesus D China. *Unmixing analysis of a time series of Hyperion images over the guánica dry forest in Puerto Rico*. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, vol. 6, no. 2, pages 329–338, 2013.
- [Gong siona] Xing Gong, Thomas Corpetti, Nicolas Courty, BaoGang Hu et Laurence Hubert-Moy. *Time series unmixing with kernel space non-negative matrix factorization*. Transaction of geoscience and remote sensing, in revision.
- [Gong sionb] Xing Gong, Thomas Corpetti, BaoGang Hu, Laurence Hubert-Moy et MengZhen Kang. *Reconstruction of LAI time series from remote sensing using GreenLab plant growth model*. Transaction of geoscience and remote sensing, in revision.
- [González-Sanpedro 2008] MC González-Sanpedro, Thuy Le Toan, J Moreno, L Kergoat et E Rubio. *Seasonal variations of leaf area index of agricultural fields retrieved from Landsat data*. Remote Sensing of Environment, vol. 112, no. 3, pages 810–824, 2008.

- [Gower 1999] Stith T Gower, Chris J Kucharik et John M Norman. *Direct and indirect estimation of leaf area index, fapar, and net primary production of terrestrial ecosystems*. Remote Sensing of Environment, vol. 70, no. 1, pages 29–51, 1999.
- [Guillemot 2012] C. Guillemot et M. Turkan. *Neighbor Embedding with Non-negative Matrix Factorization for image prediction*. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pages 785–788, Kyoto, Japon, 2012.
- [Guo 2006] YAN Guo, Yun-Tao Ma, Zhi-Gang Zhan, Bao-Guo Li, Michael Dingkuhn, Delphine Luquet et Philippe De Reffye. *Parameter optimization and field validation of the functional-structural model GREENLAB for maize*. Annals of Botany, vol. 97, no. 2, pages 217–230, 2006.
- [Gutman 1998] G Gutman et A1 Ignatov. *The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models*. International Journal of Remote Sensing, vol. 19, no. 8, pages 1533–1543, 1998.
- [Hall 1987] Forrest G Hall et Gautam D Badhwar. *Signature-extendable technology: Global space-based crop recognition*. Geoscience and Remote Sensing, IEEE Transactions on, no. 1, pages 93–103, 1987.
- [Harsanyi 1994] Joseph C Harsanyi et Chein-I Chang. *Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach*. Geoscience and Remote Sensing, IEEE Transactions on, vol. 32, no. 4, pages 779–785, 1994.
- [Hemissi 2013] Selim Hemissi, Imed Riadh Farah, Karim Saheb Ettabaa et Basel Solaiman. *Multi-spectro-temporal analysis of hyperspectral imagery based on 3-d spectral modeling and multilinear algebra*. Geoscience and Remote Sensing, IEEE Transactions on, vol. 51, no. 1, pages 199–216, 2013.
- [Hoyer 2004] Patrik O Hoyer. *Non-negative matrix factorization with sparseness constraints*. JMLR, vol. 5, pages 1457–1469, 2004.
- [Huemmrich 2001] KF Huemmrich. *The GeoSail model: a simple addition to the SAIL model to describe discontinuous canopy reflectance*. Remote Sensing of Environment, vol. 75, no. 3, pages 423–431, 2001.
- [Huete 1988] Alfredo R Huete. *A soil-adjusted vegetation index (SAVI)*. Remote sensing of environment, vol. 25, no. 3, pages 295–309, 1988.
- [Iordache 2011] M-D Iordache, José M Bioucas-Dias et Antonio Plaza. *Sparse unmixing of hyperspectral data*. Geoscience and Remote Sensing, IEEE Transactions on, vol. 49, no. 6, pages 2014–2039, 2011.
- [Jacquemoud 1990] S Jacquemoud et F Baret. *PROSPECT: A model of leaf optical properties spectra*. Remote sensing of environment, vol. 34, no. 2, pages 75–91, 1990.

- [Jacquemoud 2009] Stéphane Jacquemoud, Wout Verhoef, Frédéric Baret, Cédric Bacour, Pablo J Zarco-Tejada, Gregory P Asner, Christophe François et Susan L Ustin. *PROSPECT+ SAIL models: A review of use for vegetation characterization*. Remote Sensing of Environment, vol. 113, pages S56–S66, 2009.
- [Jallas 2000] Eric Jallas, Pierre Martin, Ronaldo Sequeira, Sammy Turner, Michel Cretenet et Edward Gérardaux. *Virtual COTONS®, the first-born of the next generation of simulation model*. In Virtual Worlds, pages 235–244. Springer, 2000.
- [Jebara 2004] Tony Jebara, Risi Kondor et Andrew Howard. *Probability product kernels*. The Journal of Machine Learning Research, vol. 5, pages 819–844, 2004.
- [Jensen 1996] John R Jensen *et al.* Introductory digital image processing: a remote sensing perspective. Numeéro Ed. 2. Prentice-Hall Inc., 1996.
- [Jonsson 2002] Per Jonsson et Lars Eklundh. *Seasonality extraction by function fitting to time-series of satellite sensor data*. Geoscience and Remote Sensing, IEEE Transactions on, vol. 40, no. 8, pages 1824–1832, 2002.
- [Justice 1998] Christopher O Justice, Eric Vermote, John RG Townshend, Ruth Defries, David P Roy, Dorothy K Hall, Vincent V Salomonson, Jeffrey L Privette, George Riggs, Alan Strahler *et al.* *The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research*. Geoscience and Remote Sensing, IEEE Transactions on, vol. 36, no. 4, pages 1228–1249, 1998.
- [Kang 2008] Meng-Zhen Kang, Paul-Henry Cournède, Philippe De Reffye, Daniel Auclair et Bao-Gang Hu. *Analytical study of a stochastic plant growth model: application to the GreenLab model*. Mathematics and Computers in Simulation, vol. 78, no. 1, pages 57–75, 2008.
- [Kersting 2012] K. Kersting, Z. Xu, M. Wahabzada, C. Bauckhage, C. Thureau, C. Roemer, A. Ballvora, U. Rascher, J. Leon et L. Pluemer. *Pre-Symptomatic Prediction of Plant Drought Stress Using Dirichlet-Aggregation Regression on Hyperspectral Images*. AAAI Computational Sustainability and AI Track, 2012.
- [Keshava 2002] Nirmal Keshava et John F Mustard. *Spectral unmixing*. Signal Processing Magazine, IEEE, vol. 19, no. 1, pages 44–57, 2002.
- [Kim 2007] Hyunsoo Kim et Haesun Park. *Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis*. Bioinformatics, vol. 23, no. 12, pages 1495–1502, 2007.
- [Kitagawa 1996] Genshiro Kitagawa. *Monte Carlo filter and smoother for non-Gaussian nonlinear state space models*. Journal of computational and graphical statistics, vol. 5, no. 1, pages 1–25, 1996.

- [Knyazikhin 1998] Yu Knyazikhin, JV Martonchik, RB Myneni, DJ Diner et SW Running. *Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data*. Journal of Geophysical Research: Atmospheres (1984–2012), vol. 103, no. D24, pages 32257–32275, 1998.
- [Kohavi 1995] Ron Kohavi et al. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. In IJCAI, volume 14, pages 1137–1145, 1995.
- [Kumar 2013] Abhishek Kumar, Vikas Sindhwani et Prabhanjan Kambadur. *Fast Conical Hull Algorithms for Near-separable Non-negative Matrix Factorization*. In ICML, pages 231–239, 2013.
- [Kwon 2006] Heesung Kwon et Nasser M Nasrabadi. *Kernel matched subspace detectors for hyperspectral target detection*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 28, no. 2, pages 178–194, 2006.
- [Kyrillidis 2013] A. Kyrillidis, S. Becker et V. Cevher. *Sparse projections onto the simplex*. JMLR W&CP: Proceedings of The 30th International Conference on Machine Learning (ICML 2013), vol. 28, no. 2, pages 235–243, 2013.
- [Lafferty 2005] John D Lafferty et Guy Lebanon. *Diffusion kernels on statistical manifolds*. 2005.
- [Lafon 2006] S. Lafon et A.B. Lee. *Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization*. PAMI, vol. 28, no. 9, pages 1393–1403, 2006.
- [Lecerf 2005] R. Lecerf, T. Corpetti, L. Hubert-Moy et V. Dubreuil. *Monitoring land use and land cover changes in oceanic and fragmented landscapes with reconstructed MODIS time series*. In Third International Workshop on the Analysis of Multi-temporal Remote Sensing Images, Multitemp, pages 195–199, Biloxi, Mississippi USA, May 2005.
- [Lecerf 2008] Rémi Lecerf, Laurence Hubert-Moy, Thomas Corpetti, Frédéric Baret, Bassam Abdel Latif et Hervé Nicolas. *Estimating biophysical variables at 250 m with reconstructed EOS/MODIS time series to monitor fragmented landscapes*. In Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, volume 2, pages II–954. IEEE, 2008.
- [LeCun 1998] Y. LeCun, L. Bottou, Y. Bengio et P. Haffner. *Gradient-based learning applied to document recognition*. Proceedings of the IEEE, vol. 86, no. 11, pages 2278–2324, 1998.
- [Letort 2008] Veronique Letort. *Adaptation du modèle de croissance GreenLab aux plantes à architecture complexe et analyse multi-échelle des relations source-puits pour l'identification paramétrique*. PhD thesis, Ecole Centrale Paris, 2008.

- [LI 2003] Xiao-qin LI, Dan-feng SUN et Feng-rong ZHANG. *Landscape Pattern Analysis on Change in the Fraction of Green Vegetation Based on Remotely Sensed Data in Beijing Mountainous Area [J]*. Journal of Mountain Research, vol. 3, 2003.
- [Li 2012] Yifeng Li et Alioune Ngom. *A new kernel non-negative matrix factorization and its application in microarray data analysis*. In Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2012 IEEE Symposium on, pages 371–378. IEEE, 2012.
- [Liang 2005] Shunlin Liang. Quantitative remote sensing of land surfaces, volume 30. John Wiley & Sons, 2005.
- [Lillesand 2004] Thomas M Lillesand, Ralph W Kiefer, Jonathan W Chipman et al. Remote sensing and image interpretation. Numéro Ed. 5. John Wiley & Sons Ltd, 2004.
- [Lindenmayer 1968] Aristid Lindenmayer. *Mathematical models for cellular interactions in development I. Filaments with one-sided inputs*. Journal of theoretical biology, vol. 18, no. 3, pages 280–299, 1968.
- [Lions 1971] Jacques Louis Lions et Sanjoy K Mitter. Optimal control of systems governed by partial differential equations, volume 1200. Springer Berlin, 1971.
- [Lisens 2000] G Lisens, P Kempencers, F Fierens et J Van Rensbergen. *Development of cloud, snow, and shadow masking algorithms for VEGETATION imagery*. In Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International, volume 2, pages 834–836. IEEE, 2000.
- [Lopez 2008] Gerardo Lopez, Romeo R Favreau, Colin Smith, Evelyne Costes, Przemyslaw Prusinkiewicz et Theodore M DeJong. *Integrating simulation of architectural development and source-sink behaviour of peach trees by incorporating Markov chains and physiological organ function submodels into L-PEACH*. Functional Plant Biology, vol. 35, no. 10, pages 761–771, 2008.
- [Maculan 1989] N. Maculan et G. de Paula. *A linear-time median-finding algorithm for projecting a vector on the simplex of  $R^n$* . Operations research letters, vol. 8, no. 4, pages 219–222, 1989.
- [Marteau 2009] P-F Marteau. *Time warp edit distance with stiffness adjustment for time series matching*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 31, no. 2, pages 306–318, 2009.
- [Michelucci 2004] Dominique Michelucci et Sebti Foufou. *Using Cayley-Menger determinants for geometric constraint solving*. In Proceedings of the ninth ACM symposium on Solid modeling and applications, pages 285–290. Eurographics Association, 2004.
- [Mørup 2012] Morten Mørup et Lars Kai Hansen. *Archetypal analysis for machine learning and data mining*. Neurocomputing, vol. 80, pages 54–63, 2012.

- [Myers 1980] Cory Myers, Lawrence Rabiner et Aaron E Rosenberg. *Performance tradeoffs in dynamic time warping algorithms for isolated word recognition*. Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 28, no. 6, pages 623–635, 1980.
- [Myneni 1997] Ranga B Myneni, Ramakrishna Ramakrishna, R Nemani et Steven W Running. *Estimation of global leaf area index and absorbed PAR using radiative transfer models*. Geoscience and Remote Sensing, IEEE Transactions on, vol. 35, no. 6, pages 1380–1393, 1997.
- [Nascimento 2005] José MP Nascimento et José M Bioucas Dias. *Vertex component analysis: A fast algorithm to unmix hyperspectral data*. Geoscience and Remote Sensing, IEEE Transactions on, vol. 43, no. 4, pages 898–910, 2005.
- [Nene 1996] S. Nene, S. Nayar et H. Murase. *Columbia object image library (COIL-20)*. Dept. Comput. Sci., Columbia Univ., New York, vol. 62, 1996.
- [Noma 2002] Hiroshi Shimodaira Ken-ichi Noma. *Dynamic time-alignment kernel in support vector machine*. Advances in neural information processing systems, vol. 14, page 921, 2002.
- [Nummiaro 2003] Katja Nummiaro, Esther Koller-Meier et Luc Van Gool. *An adaptive color-based particle filter*. Image and Vision Computing, vol. 21, no. 1, pages 99–110, 2003.
- [Olshausen 1997] Bruno A Olshausen et David J Field. *Sparse coding with an overcomplete basis set: A strategy employed by V1?* Vision research, vol. 37, no. 23, pages 3311–3325, 1997.
- [Öztireli 2010] C. Öztireli, M. Alexa et M. Gross. *Spectral Sampling of Manifolds*. In SIGGRAPH ASIA, Dec. 2010.
- [Pauca 2006] V Paul Pauca, Jon Piper et Robert J Plemmons. *Nonnegative matrix factorization for spectral data analysis*. Linear algebra and its applications, vol. 416, no. 1, pages 29–47, 2006.
- [Perttunen 1996] Jari Perttunen, R SIEV ÄNEN, Eero Nikinmaa, H Salminen, H Saarenmaa et al. *LIGNUM: a tree model based on simple structural units*. Annals of Botany, vol. 77, no. 1, pages 87–98, 1996.
- [Petitjean 2011] François Petitjean, Alain Ketterlin et Pierre Gançarski. *A global averaging method for dynamic time warping, with applications to clustering*. Pattern Recognition, vol. 44, no. 3, pages 678–693, 2011.
- [Petitjean 2012] François Petitjean, Jordi Inglada et Pierre Gançarski. *Satellite image time series analysis under time warping*. Geoscience and Remote Sensing, IEEE Transactions on, vol. 50, no. 8, pages 3081–3095, 2012.
- [Pettorelli 2005] Nathalie Pettorelli, Jon Olav Vik, Atle Mysterud, Jean-Michel Gaillard, Compton J Tucker et Nils Chr Stenseth. *Using the satellite-derived NDVI to assess ecological responses to environmental change*. Trends in ecology & evolution, vol. 20, no. 9, pages 503–510, 2005.

- [Pierce 1988] Lars L Pierce et Steven W Running. *Rapid estimation of coniferous forest leaf area index using a portable integrating radiometer*. Ecology, pages 1762–1767, 1988.
- [Price 1995] John C Price et Walter C Bausch. *Leaf area index estimation from visible and near-infrared reflectance data*. Remote Sensing of Environment, vol. 52, no. 1, pages 55–65, 1995.
- [Qi 2010] Rui Qi. *Optimization and optimal control of plant growth: application of GreenLab model for decision aid in agriculture*. PhD thesis, Ecole Centrale Paris, 2010.
- [Rabiner 1975] Lawrence R Rabiner et Bernard Gold. *Theory and application of digital signal processing*. Englewood Cliffs, NJ, Prentice-Hall, Inc., 1975. 777 p., vol. 1, 1975.
- [Recht 2012] Ben Recht, Christopher Re, Joel Tropp et Victor Bittorf. *Factoring nonnegative matrices with linear programs*. In NIPS, pages 1223–1231, 2012.
- [Richards 1999] John A Richards et JA Richards. Remote sensing digital image analysis, volume 3. Springer, 1999.
- [Richards 2013] John A Richards. Remote sensing digital image analysis: an introduction. Springer, 2013.
- [Ridao 1998] Eduardo Ridao, Jose R Conde et M Inés Miñiguez. *Estimating fAPAR from nine vegetation indices for irrigated and nonirrigated faba bean and semileafless pea canopies*. Remote Sensing of Environment, vol. 66, no. 1, pages 87–100, 1998.
- [Roerink 2000] GJ Roerink, M Menenti et W Verhoef. *Reconstructing cloudfree NDVI composites using Fourier analysis of time series*. International Journal of Remote Sensing, vol. 21, no. 9, pages 1911–1917, 2000.
- [Sabins Jr 1978] Floyd F Sabins Jr et al. *Remote sensing*. Remote sensing., 1978.
- [Sakamoto 2005] Toshihiro Sakamoto, Masayuki Yokozawa, Hitoshi Toritani, Michio Shibayama, Naoki Ishitsuka et Hiroyuki Ohno. *A crop phenology detection method using time-series MODIS data*. Remote sensing of environment, vol. 96, no. 3, pages 366–374, 2005.
- [Sakoe 1978] Hiroaki Sakoe et Seibi Chiba. *Dynamic programming algorithm optimization for spoken word recognition*. Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 26, no. 1, pages 43–49, 1978.
- [Schaepman-Strub 2006] G Schaepman-Strub, ME Schaepman, TH Painter, S Dangel et JV Martonchik. *Reflectance quantities in optical remote sensing-Definitions and case studies*. Remote sensing of environment, vol. 103, no. 1, pages 27–42, 2006.
- [Schölkopf 2002] B. Schölkopf et A. Smola. Learning with kernels: Support vector machines, regularization, optimization, and beyond. the MIT Press, 2002.



- [Shroff 2011] N. Shroff, P. Turaga et R. Chellappa. *Manifold Precis: An Annealing Technique for Diverse Sampling of Manifolds*. In NIPS, pages 154–162, 2011.
- [Sim 2002] T. Sim, S. Baker et M. Bsat. *The CMU pose, illumination, and expression (PIE) database*. In Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on, pages 46–51. IEEE, 2002.
- [Smith 1985] Milton O Smith, Paul E Johnson et John B Adams. *Quantitative determination of mineral types and abundances from reflectance spectra using principal components analysis*. Journal of Geophysical Research: Solid Earth (1978–2012), vol. 90, no. S02, pages C797–C804, 1985.
- [Swayze 1992] Gregg Swayze, Roger N Clark, Fred Kruse, Steve Sutley et Andrea Gallagher. *Ground-truthing AVIRIS mineral mapping at Cuprite, Nevada*. 1992.
- [Thureau 2010] C. Thureau, K. Kersting et C. Bauckhage. *Yes we can: simplex volume maximization for descriptive web-scale matrix factorization*. In CIKM, pages 1785–1788, 2010.
- [Thureau 2012] C. Thureau, K. Kersting, M. Wahabzada et C. Bauckhage. *Descriptive matrix factorization for sustainability Adopting the principle of opposites*. Data Mining & Knowledge Disc., vol. 24, no. 2, pages 325–354, 2012.
- [Tong 2002] Simon Tong et Daphne Koller. *Support vector machine active learning with applications to text classification*. The Journal of Machine Learning Research, vol. 2, pages 45–66, 2002.
- [Townshend 1991] John Townshend, Christopher Justice, Wei Li, Charlotte Gurney et Jim McManus. *Global land cover classification by remote sensing: present capabilities and future possibilities*. Remote Sensing of Environment, vol. 35, no. 2, pages 243–255, 1991.
- [Tripathi 2014] Poonam Tripathi, NR Patel, SPS Kushwaha et VK Dadhwal. *Upscaling of leaf area index in Terai forest plantations using fine-and moderate-resolution satellite data*. International Journal of Remote Sensing, vol. 35, no. 22, pages 7749–7762, 2014.
- [Trotter 1991] Craig M Trotter. *Remotely-sensed data as an information source for geographical information systems in natural resource management a review*. International Journal of Geographical Information System, vol. 5, no. 2, pages 225–239, 1991.
- [Tucker 2005] Compton J Tucker, Jorge E Pinzon, Molly E Brown, Daniel A Slayback, Edwin W Pak, Robert Mahoney, Eric F Vermote et Nazmi El Saleous. *An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data*. International Journal of Remote Sensing, vol. 26, no. 20, pages 4485–4498, 2005.

- [Turner 1999] David P Turner, Warren B Cohen, Robert E Kennedy, Karin S Fassnacht et John M Briggs. *Relationships between leaf area index and Landsat TM spectral vegetation indices across three temperate zone sites*. Remote sensing of environment, vol. 70, no. 1, pages 52–68, 1999.
- [Verhoef 1984] W Verhoef. *Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model*. Remote sensing of environment, vol. 16, no. 2, pages 125–141, 1984.
- [Vermote 1997] Eric F Vermote, Didier Tanré, Jean L Deuze, Maurice Herman et J-J Morcette. *Second simulation of the satellite signal in the solar spectrum, 6S: An overview*. Geoscience and Remote Sensing, IEEE Transactions on, vol. 35, no. 3, pages 675–686, 1997.
- [Viovy 2000] N Viovy. *Automatic classification of time series (ACTS): a new clustering method for remote sensing time series*. International Journal of Remote Sensing, vol. 21, no. 6-7, pages 1537–1560, 2000.
- [Vogelmann 2001] James E Vogelmann, Stephen M Howard, Limin Yang, Charles R Larson, Bruce K Wylie et Nick Van Driel. *Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources*. Photogrammetric Engineering and Remote Sensing, vol. 67, no. 6, 2001.
- [Von Luxburg 2007] Ulrike Von Luxburg. *A tutorial on spectral clustering*. Statistics and computing, vol. 17, no. 4, pages 395–416, 2007.
- [Walthall 1992] Charles L Walthall et Elizabeth M Middleton. *Assessing spatial and seasonal variations in grasslands with spectral reflectances from a helicopter platform*. Journal of Geophysical Research: Atmospheres (1984–2012), vol. 97, no. D17, pages 18905–18912, 1992.
- [Wang 1984] Qing Ren Wang et Ching Y Suen. *Analysis and design of a decision tree based on entropy reduction and its application to large character set recognition*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, no. 4, pages 406–417, 1984.
- [Wang 2010] Fa-Yu Wang, Chong-Yung Chi, Tsung-Han Chan et Yue Wang. *Nonnegative least-correlated component analysis for separation of dependent sources by volume maximization*. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 32, no. 5, pages 875–888, 2010.
- [Wang 2011] Feng Wang, Mengzhen Kang, Qi Lu, Véronique Letort, Hui Han, Yan Guo, Philippe de Reffye et Baoguo Li. *A stochastic model of tree architecture and biomass partitioning: application to Mongolian Scots pines*. Annals of Botany, vol. 107, no. 5, pages 781–792, 2011.
- [Weiss 2000] Marie Weiss, Frédéric Baret, Ranga B Myneni, Agnès Pragnère, Youri Knyazikhin et al. *Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data*. Agronomie, vol. 20, no. 1, pages 3–22, 2000.

- [Weiss 2007] Marie Weiss, Frédéric Baret, Sébastien Garrigues et Roselyne Lacaze. *LAI and fAPAR CYCLOPES global products derived from VEG-ETATION. Part 2: validation and comparison with MODIS collection 4 products*. Remote sensing of Environment, vol. 110, no. 3, pages 317–331, 2007.
- [Welch 1995] Greg Welch et Gary Bishop. *An introduction to the Kalman filter*, 1995.
- [Winter 1999] Michael E Winter. *N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data*. In SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, pages 266–275. International Society for Optics and Photonics, 1999.
- [Wright 2009] John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry et Yi Ma. *Robust face recognition via sparse representation*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 31, no. 2, pages 210–227, 2009.
- [Xiong 2006] Xiaoxiong Xiong et William Barnes. *An overview of MODIS radiometric calibration and characterization*. Advances in Atmospheric Sciences, vol. 23, no. 1, pages 69–79, 2006.
- [YAN 2004] HONG-PING YAN, Meng-Zhen Kang, Philippe De Reffye et Michael Dingkuhn. *A dynamic, architectural plant model simulating resource-dependent growth*. Annals of Botany, vol. 93, no. 5, pages 591–602, 2004.
- [Yang 2008] Shuang-Hong Yang, Bao-Gang Hu et Paul-Henry Cournède. *Structural identifiability of generalized constraint neural network models for nonlinear regression*. Neurocomputing, vol. 72, no. 1, pages 392–400, 2008.
- [Yang 2009] Jianchao Yang, Kai Yu, Yihong Gong et Thomas Huang. *Linear spatial pyramid matching using sparse coding for image classification*. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 1794–1801. IEEE, 2009.
- [Yu 2009] K. Yu, T. Zhang et Y. Gong. *Nonlinear learning using local coordinate coding*. In NIPS, pages 2223–2231, 2009.
- [Zhan 2003] Zhi-Gang Zhan, Philippe De Reffye, François Houllier et Bao-Gang Hu. *Fitting a functional-structural growth model with plant architectural data*. In International Symposium on Plant Growth Modeling, Simulation, Visualization and their Applications-PMA'03, pages 108–117, 2003.
- [Zhang 2006] Z. Zhang Da.and Zhou et S. Chen. *Non-negative matrix factorization on kernels*. In PRICAI 2006: Trends in Artificial Intelligence, pages 404–412. Springer, 2006.
- [Zhou 2002] Xiaodong Zhou, Qijiang Zhu, Shihao Tang, Xue Chen et Menxin Wu. *Interception of PAR and relationship between FPAR and LAI in summer maize canopy*. In Geoscience and Remote Sensing Symposium,

2002. IGARSS'02. 2002 IEEE International, volume 6, pages 3252–3254. IEEE, 2002.
- [Zhou 2010] Feng Zhou, Fernando De la Torre et Jeffrey F Cohn. *Unsupervised discovery of facial events*. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pages 2574–2581. IEEE, 2010.
- [Zhu 2014] Feiyun Zhu, Ying Wang, Shiming Xiang, Bin Fan et Chunhong Pan. *Structured Sparse Method for Hyperspectral Unmixing*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 88, pages 101–118, 2014.



# List of Tables

1.1	Caractéristiques en termes de résolution, nombre de bandes spectrales et fauchée des principaux satellites d’observation de la terre	3
2.1	The Root-Mean-Square Error (RMSE) of Different Methods on Sequences of various SNR level.	27
2.2	The Root-Mean-Square Error(RMSE) of Different Methods on Sequence with Sparse Observations.	29
2.3	The Root-Mean-Square Error(RMSE) of Different Methods on Sequence of Missing Observations.	30
3.1	Mean ratio and variance (in %, over 30 runs) of maximum simplex volumes found by both approximations from Eq. 3.4 and Eq. 3.5, as well as SAGA over the exact CMD approach. For each dataset, the most accurate result is in bold font.	55
3.2	Mean accuracy and variance NMF-based classification. For each dataset, the most accurate method is in bold font. This property is assessed thanks to a paired Student’s T-test between the SAGA and the best other method (apart from SK SiVM which can be seen as a particular case of SAGA)	59
4.1	<b>Quantitative comparison of endmember extraction</b> performances for the six tested approaches	80
4.2	The Salient Component Detect Rate for LAI time series unmixing. Gaussian, DTW, GAK represent results of our proposed methods, with Gaussian, DTW, GAK, kernels and $\ell$ stands for various endmember numbers.	80
4.3	The Salient Component Detect Rate comparison between MODIS B1, B2, FCover, LAI time series ( $\ell = 28$ ).	81



# List of Figures

1.1	<b>Illustration de données brutes MODIS.</b> En raison des mauvaises conditions d'acquisition, des nuages, aérosols, ..., les séries temporelles résultantes sont bruitées. . . . .	4
1.2	<b>Deux séries temporelles représentant un même comportement dynamique mais décalées dans le temps.</b> Ici, une comparaison point-par-point des valeurs de chaque série n'est pas optimale. . . . .	4
2.1	<b>LAI calculé sur une série d'images brutes MODIS sur le site de Chizé, France.</b> Les zones en blanc (à l'intérieur de la région d'intérêt) correspondent à des données manquantes. . . . .	13
2.2	<b>Leaf area index evolution with different <math>r, k, \alpha</math></b> . . . . .	21
2.3	<b>A demonstration of uncorrupted LAI time series for three sorts of agriculture crops in Brittany, France</b> . . . . .	26
2.4	<b>LAI reconstruction with various techniques on the sequence of noisy observations (with SNR=6).</b> Dark line are uncorrupted MODIS observations (ground truth state), blue dots are the input noisy observations, red line is the result of reconstruction with particle smoother, cyan line from particle filter, green line from fitted model, pink line from polynomial regression, blue line from wavelet filter. Even in this situation where the noise is large, the best fitted curves are issued from techniques based on the GreenLab model . . .	28
2.5	<b>LAI reconstruction with various techniques on the sequence of sparse observations (with SNR=6).</b> Dark line are uncorrupted MODIS observations (ground truth state), blue dots are the input noisy observations, red line is the result of reconstruction with particle smoother, cyan line from particle filter, green line from fitted model, pink line from polynomial regression, blue line from wavelet filter. . .	28
2.6	<b>LAI reconstruction with various techniques on the sequence of missing observations (with SNR=6)</b> . Dark line are uncorrupted MODIS observations (ground truth state), blue dots are the input noisy observations, red line is the result of reconstruction with particle smoother, cyan line from particle filter, green line from fitted model, pink line from polynomial regression, blue line from wavelet filter.	29
2.7	<b>Several examples of LAI time series reconstruction with particle smoother at a study site in Chizé, France</b> . . . . .	31
2.8	<b>Estimated biomass production and its partition during continuous growth cycles of unit wheat field.</b> . . . . .	32
2.9	<b>Result of particle smoother at a study site in Chizé, France.</b> Top : input LAI values computed from MODIS data ; Bottom : reconstructed data . . . . .	33
3.1	<b>Illustration des approches qui seront présentées.</b> (a) extraction des endmembers (en rouge) par maximisation du volume et (b) estimation des abondances. . . . .	38



3.2 (a) A dataset  $\mathbf{X}$  embedded in a U-shaped manifold; (b) the separability assumption assumes the dataset is encompassed in a cone spanned by the columns of  $\mathbf{F}$ ; a datum is described in a column of  $\mathbf{G}$  through its non-negative coordinates in the cone; (c) Convex hull: the dataset is embedded in a simplex, and each datum is described with barycentric coordinates; (d) Manifold hull: one uses a great enough number of reference points to precisely characterize the manifold, while forcing the sparsity to achieve some local coding. Note that, in order to make these imaged representations clearer, we did not represent  $\mathbf{F}$  as a CSS of  $\mathbf{X}$ . . . . . 42

3.3 A ring and an S-shape datasets (a-c) It is possible to have  $\ell > p$  elements in the CSS forming a non-simplicial polytope (here illustrated by a 4-NN graph) which approximates well the contour of the shape. (b-d) Each pixel of the image is projected in the RKHS onto the 29-simplex (b) or 49-simplex (d). Each pixel is colored according to its reconstruction error (black for low values, white for high ones). . . . . 53

3.4 (a-b) The 29-simplex on the ring shape dataset: The red point is projected in the RKHS as a 3-sparse vector. Each vertex color accounts for the corresponding convex coordinates (from blue= 0 to red= 1) (c-d) The S dataset with 5-sparse vectors (in (d), only 4 projections are non-nil). . . . . 53

3.5 (a) Reconstruction error (in red) and mean radius of smallest enclosing ball (blue) for the ring dataset. (b) Normalized reconstruction error. . . . . 54

3.6 Comparison of the reconstruction error according to the size of the CSS: (a) Wine dataset; (b) simulated dataset (50 dimensional Gaussian). . . . . 56

3.7 Comparison of the computational times for the CSS (a) and the projection (b) for SiVM and SAGA. . . . . 57

3.8 Feature extraction evaluation on (a) COIL-20, (b) PIE, (c) MNIST and (d) CIFAR-10 datasets at different reduced dimension. . . . 58

3.9 Computational performances between SAGA and the other considered NMF methods on the whole PIE dataset. . . . . 59

4.1 **Principe du “Dynamic Time Warping” (DTW) pour l’alignement de deux séries temporelles.** Les deux courbes présentent un profil similaire associé à une même classe. Malgré cela, la distance euclidienne entre ces séries est importante. Le principe de DTW est de chercher le meilleur alignement entre les deux courbes (droites bleues) et de calculer l’erreur sur les courbes ré-alignées, mettant ainsi en évidence le fait que les deux courbes sont similaires. . . . . 63

4.2 **An endmember extraction demo on toy dataset.** The blue points are the observation data of mixed pixels, they contain more than one class substance, the cyan points represent various distinct materials (pure pixel), and the red points are extracted endmember, they are quite close to be pure pixel. . . . . 73

4.3 **Leaf Area Index evolution of various years in Belledonne, France.** the horizontal axis is date of the year, the vertical axis is Leaf Area Index . . . . . 78

4.4	<b>An illustration of endmember extraction results of various methods</b> , time series are projected to 2D with PCA, and red diamond points are extracted endmembers. (a) SiVM with Gaussian kernel, (b) SiVM with DTW kernel, (c) SiVM with GA kernel, (d) VCA, (e) NFINDR, (f) SISAL. . . . .	79
4.5	<b>Time series evaluation of various methods with different end-member set size.</b> . . . . .	81
6.1	Dual scale automaton. . . . .	92





## Abstract

---

This PhD dissertation is concerned with time series analysis for medium spatial resolution (MSR) remote sensing images. The main advantage of MSR data is their high temporal rate which allows to monitor land use. However, two main problems arise with such data. First, because of cloud coverage and bad acquisition conditions, the resulting time series are often corrupted and not directly exploitable. Secondly, pixels in medium spatial resolution images are often “mixed” in the sense that the spectral response is a combination of the response of “pure” elements.

These two problems are addressed in this PhD. First, we propose a data assimilation technique able to recover consistent time series of Leaf Area Index from corrupted MODIS sequences. To this end, a plant growth model, namely GreenLab, is used as a dynamical constraint. Second, we propose a new and efficient unmixing technique for time series. It is in particular based on the use of “elastic” kernels able to properly compare time series shifted in time or of various lengths.

Experimental results are shown both on synthetic and real data and demonstrate the efficiency of the proposed methodologies.

**Keywords:** Time series, Medium Spatial Resolution, data assimilation, pixel unmixing, elastic kernel.

## Résumé

---

Cette thèse s’intéresse à l’analyse de séries temporelles d’images satellites à moyenne résolution spatiale. L’intérêt principal de telles données est leur haute répétitivité qui autorise des analyses de l’usage des sols. Cependant, deux problèmes principaux subsistent avec de telles données. En premier lieu, en raison de la couverture nuageuse, des mauvaises conditions d’acquisition, ..., ces données sont souvent très bruitées. Deuxièmement, les pixels associés à la moyenne résolution spatiale sont souvent “mixtes” dans la mesure où leur réponse spectrale est une combinaison de la réponse de plusieurs éléments “purs”.

Ces deux problèmes sont abordés dans cette thèse. Premièrement, nous proposons une technique d’assimilation de données capable de recouvrer des séries temporelles cohérentes de LAI (Leaf Area Index) à partir de séquences d’images MODIS bruitées. Pour cela, le modèle de croissance de plantes GreenLab est utilisé. En second lieu, nous proposons une technique originale de démixage, qui s’appuie notamment sur des noyaux “élastiques” capables de gérer les spécificités des séries temporelles (séries de taille différentes, décalées dans le temps, ...)

Les résultats expérimentaux, sur des données synthétiques et réelles, montrent de bonnes performances des méthodologies proposées.

**Mots-clés :** Séries temporelles, moyenne résolution spatiale, assimilation de données, démixage, noyaux élastiques.