



HAL
open science

Data-driven approaches for ocean remote sensing: from the non-negative decomposition of operators to the reconstruction of satellite-derived sea surface dynamics

Manuel Lopez Radcenco

► **To cite this version:**

Manuel Lopez Radcenco. Data-driven approaches for ocean remote sensing: from the non-negative decomposition of operators to the reconstruction of satellite-derived sea surface dynamics. Signal and Image processing. Ecole nationale supérieure Mines-Télécom Atlantique, 2018. English. NNT : 2018IMTA0107 . tel-02083495

HAL Id: tel-02083495

<https://theses.hal.science/tel-02083495>

Submitted on 29 Mar 2019

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 DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPERIEURE MINES-TELECOM ATLANTIQUE
BRETAGNE PAYS DE LA LOIRE - IMT ATLANTIQUE
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : Signal, Image, Vision

Par

Manuel LOPEZ RADCENCO

Data-driven approaches for ocean remote sensing:

from the non-negative decomposition of operators to the reconstruction of satellite-derived sea surface dynamics

Thèse présentée et soutenue à IMT Atlantique – Campus de Brest, le 12 décembre, 2018
Unité de recherche : Lab-STICC, Technopôle Brest-Iroise, CS 83818, 29238 Brest Cedex 3, France
Thèse N° : 2018IMTA0107

Composition du Jury :

Président :	Florence TUPIN	Professeur, Télécom ParisTech
Rapporteurs :	Abdouramane ATTO Emmanuel COSME	Maître de Conférences, Université Savoie-Mont-Blanc Maître de Conférences, Université Grenoble-Alpes
Examineurs :	Anne CUZOL Ananda PASCUAL	Maître de Conférences, Université de Bretagne-Sud Chercheuse, Institute Méditerranéen d'Etudes Avancés
Directeur de thèse :	Ronan FABLET	Professeur, IMT Atlantique
Co-directeur de thèse :	Abdeldjalil AISSA EL BEY	Professeur, IMT Atlantique
Invités :	Pierre ALLIOT	Maître de conférences, Université de Bretagne Occidentale

Acknowledgements

A Lucía, por estar siempre conmigo y hacerme sentir que nunca estoy solo...

A mi familia y amigos, y particularmente a mi madre, por su apoyo incondicional y su cariño a la distancia...

Un grand remerciement à mes tuteurs, Prof. Ronan Fablet et Prof. Abdeldjalil Aissa-El-Bey, et à mon encadrant, Dr. Pierre Ailliot, pour le support, non seulement scientifique, mais aussi psychologique et moral, lors du déroulement de la thèse. Je remercie aussi les chercheurs de l'IMEDEA qui m'ont accueilli chaleureusement à Majorque, et plus particulièrement à Ananda Pascual et Laura Gomez Navarro. Merci également à Redouane Lguensat pour le template LaTeX. Je remercie Cédric Herzet pour son aide et sa bonne disposition. Finalement, merci aux rapporteurs et examinateurs pour avoir donné de son temps pour lire ce document et participer à ma soutenance de thèse.

Merci aux amis qui m'ont appuyé lors du déroulement de la thèse: Aurélien, Antony, Nicolas A., Steph, Thomas, Nicolas G., Alma, Sabrina, Redouane, Said, et d'autres que j'oublie probablement mais qui ne sont pas, néanmoins, moins importants.

On a personal note, besides acknowledging everyone that made this work possible, I would like to point out the deep feeling of personal and professional growth achieved during these last three years of research work. In particular, many fruitful interactions, not only with researchers directly involved in my daily research activities, but also with multiple researchers from a wide variety of scientific backgrounds (during conferences, symposiums and scientific visits) have helped me understand the need for scientific collaboration and the development of multidisciplinary research groups in order to ensure scientific advancement fully benefits from the interaction between different scientific disciplines. This would have not been possible without the help and support from PhD tutors and advisors, department colleagues and fellow PhD candidates.

Contents

Acknowledgments	I
Résumé long en français	VII
Abstract	XXVII
Acronyms	XXIX
Notation	XXXIII
1 Introduction and thesis summary	1
1.1 Context	1
1.2 Geophysical motivations for data-driven approaches	3
1.3 Contributions	5
1.4 Dissertation outline	7
1.5 Publications and Valorization	7
1.5.1 International Conference papers	7
1.5.2 National Conference papers	8
1.5.3 Journal papers	8
1.5.4 Poster presentations	9
1.5.5 Invited talks	9
1.5.6 Other activities	9
I Non-negative decomposition of local operators	11
2 State of the art and related work	13
2.1 Introduction	13
2.2 Blind source separation	14
2.2.1 Proposed formulations	15
2.2.2 Associated numerical methods	20
2.3 Operator decomposition	31
2.3.1 Orthogonality-based decompositions	32
2.3.2 Dynamic mode decomposition	37

2.3.3	Tikhonov-regularized decompositions	45
2.4	Conclusion	47
3	Proposed models and algorithms	49
3.1	Introduction	49
3.2	General model	50
3.3	Non-negative linear superposition of linear modes	51
3.3.1	Latent class model	52
3.3.2	Parameter estimation	53
3.3.3	Alternating least squares algorithm	54
3.4	Reformulation based on local linear operators	58
3.4.1	Parameter Estimation	59
3.4.2	Dictionary-based decomposition of local linear operators	61
3.4.3	Model training and application	62
3.4.4	Computational complexity analysis	62
3.5	Conclusion	63
4	Algorithm evaluation and benchmarking	65
4.1	Introduction	65
4.2	Synthetic dataset generation	66
4.3	Estimation performance under ideal settings	67
4.4	Estimation performance with noisy mixing coefficients	72
4.5	Latent class initialization evaluation	78
4.6	Conclusion	82
5	Applications	85
5.1	Introduction	85
5.2	Data-driven forecasting of dynamical systems	86
5.3	Multi-tracer synergies for the characterization of upper ocean dynamics	90
5.3.1	Segmentation of upper ocean dynamics from SST/SSS relationships in the Alboran Sea	90
5.3.2	Segmentation of upper ocean dynamics from SST/SSH relationships in the Agulhas region	92
5.4	Conclusion	105
II	Data-driven interpolation of sea level anomaly fields from satellite- derived ocean remote sensing data	107
6	Context and related work	109
6.1	Introduction	109

6.2	Satellite altimetry	111
6.2.1	Along-track nadir altimetry	111
6.2.2	Surface Water Ocean Topography altimetry mission	115
6.3	Geophysical field interpolation	118
6.3.1	Exemplar-based methods	119
6.3.2	Projection-based methods	119
6.3.3	Data assimilation	119
6.3.4	Analog Forecasting	124
6.3.5	AnDA: The analog data assimilation framework	125
6.4	Conclusion	128
7	Locally-adapted convolutional models for the reconstruction of SLA fields from satellite along-track data	129
7.1	Introduction	129
7.2	Model formulation	131
7.2.1	Problem statement	131
7.2.2	Unconstrained model calibration	132
7.2.3	Dictionary-based decompositions	133
7.2.4	Locally-adapted dictionary-based convolutional models	134
7.2.5	Data and Methodology	135
7.2.6	Results	137
7.3	Conclusion	138
8	Data-driven fusion of nadir along-track and SWOT data	141
8.1	Introduction	142
8.2	Benchmarked methods	144
8.3	Case study and data	145
8.3.1	Data	145
8.3.2	Observing System Simulation Experiment	145
8.3.3	Pseudo-observations	146
8.4	Method comparison and benchmarking	146
8.5	AnDA Parameter Sensitivity Analysis	148
8.5.1	Patch size	149
8.5.2	Number of neighbours	149
8.5.3	Assimilation lag	150
8.5.4	Pseudo-observations half-window size	150
8.6	AnDA Noise sensitivity analysis	152
8.6.1	Nadir along-track data	152
8.6.2	SWOT data	153
8.7	Nadir along-track vs. SWOT data	153

8.8	Altimetric data fusion	158
8.9	Additional regressors for analog forecasting	167
8.10	Assimilation of observation gradients	170
8.11	Conclusion	172
III	Conclusion and Perspectives	175
9	Conclusion and Perspectives	177
9.1	Conclusions	177
9.2	Perspectives	179
9.3	Final thoughts	181
	List of Figures	194
	List of Tables	197
	Bibliography	220

Résumé long en français

Résumé

Mots clés : *Téledétection satellitaire, Masses de données, Décomposition d'opérateurs, Non-négativité, Parcimonie, Dynamiques de l'océan, Fouille de données, Apprentissage statistique, Fusion de données multi-capteurs, Reconstruction haute-résolution, Interpolation.*

Au cours des dernières années, la disponibilité toujours croissante de données de télédétection multi-source de l'océan a été un facteur clé pour améliorer notre compréhension des dynamiques de la surface de l'océan, de la circulation océanique et des interactions atmosphère-océan. A cet égard, il existe un potentiel largement sous-exploité dans les ensembles de données actuellement disponibles. Il est donc essentiel de mettre au point des approches efficaces pour mieux exploiter ces ensembles de données afin de résoudre, par des méthodes de fouille de données et d'apprentissage statistique, des problèmes tels que la fusion multi-capteurs, la reconstruction haute résolution ou l'analyse multi-paramètres des processus d'intérêt. En particulier, la décomposition des processus géophysiques en modes pertinents est une question clé pour les problèmes de caractérisation, de prédiction et de reconstruction. Inspirés par les progrès récents en séparation aveugle des sources issus de l'introduction de formulations non-négatives et parcimonieuses, nous visons, dans la première partie de cette thèse, à étendre les modèles de séparation aveugle de sources sous contraintes au problème de la caractérisation et décomposition d'opérateurs ou fonctions de transfert entre variables d'intérêt, en se concentrant sur des formulations non-négatives. Dans ce travail, nous développons des schémas computationnellement efficaces reposant sur des fondations mathématiques solides, y compris la reformulation du problème de décomposition d'opérateurs sous contraintes comme la décomposition par dictionnaire d'opérateurs linéaires estimés localement, pour améliorer ainsi la flexibilité du modèle. Nous illustrons la pertinence des modèles de décomposition proposés dans différentes applications impliquant l'analyse, la segmentation et la prédiction de dynamiques géophysiques. Par la suite, étant donné que la disponibilité toujours croissante d'ensembles de données multi-sources supporte l'exploration des approches pilotées par les données en tant qu'alternative aux formulations classiques basées sur des modèles, nous explorons, dans la deuxième partie de ce travail, des approches basées sur les données récemment introduits pour l'interpolation des champs géophysiques à partir d'observations satellitaires irrégulièrement échantillonnées. De plus, en vue de la future

mission SWOT, la première mission satellitaire à produire des observations d'altimétrie par satellite complètement bidimensionnelles et à large fauchée, nous nous intéressons à évaluer dans quelle mesure les données SWOT, combinées ou non aux données altimétriques conventionnelles, permettraient une meilleure reconstruction des champs altimétriques.

Introduction

Au cours des dernières années, la télédétection satellitaire a produit une grande quantité de données d'observation provenant d'une grande variété de sources. En effet, une vaste gamme de types de capteurs différents (imagerie multi et hyper-spectrale, imagerie SAR, imagerie micro-ondes, etc.) permet l'observation de différents paramètres géophysiques et géochimiques terrestres, océaniques et atmosphériques (hauteur de la surface la mer, température de la surface de la mer, couverture végétale, développement urbain, etc.) à différentes résolutions spatio-temporelles, ce qui représente une énorme quantité de données très peu exploitées.

Une meilleure exploitation de ces ensembles de données nous permettrait d'améliorer la qualité, la puissance et la précision des modèles et des représentations des processus géophysiques et géochimiques, ainsi que de dépasser les limites associées tant aux données actuellement disponibles, comme le fait qu'aucun capteur satellite ne permet de fournir des observations haute-résolution des dynamiques de l'océan à la fois temporellement et spatialement, qu'aux modèles et représentations actuels, trop complexes pour permettre une analyse efficace et une reconstruction haute résolution basée sur des observations, ou trop simples pour appréhender toute la variabilité spatio-temporelle des processus étudiés. Cependant, pour exploiter pleinement le potentiel de ces nouveaux ensembles de données, une compréhension approfondie de la relation entre les différentes variables impliquées est nécessaire.

À cet égard, la mise en œuvre de nouvelles approches pour exploiter pleinement le potentiel des ensembles de données actuellement disponibles apparaît clairement comme un enjeu de recherche majeur. En particulier, les approches d'exploration de données et d'apprentissage automatique pourraient être utilisées dans ce contexte pour une grande variété d'applications, allant de la fusion de données multi-sources et la reconstruction à haute résolution à l'analyse multi-paramétrique des processus géophysiques ou géochimiques, l'analyse spatio-temporelle et la segmentation de la dynamique des systèmes.

Dans ce contexte, cette thèse a deux objectifs majeurs :

1. Explorer les formulations non-négatives et parcimonieuses [69, 155, 200, 233] pour développer de nouvelles méthodes non supervisées pour caractériser les relations entre variables à partir d'un ensemble représentatif d'observations. En effet, étant donné le succès récent des nouvelles formulations [69, 155, 200, 233] pour s'attaquer aux problèmes de séparation

aveugle des sources [201], elles apparaissent comme une alternative intéressante pour résoudre le problème de la caractérisation des opérateurs à partir d'observations. Dans le contexte de la télédétection océanique, la disponibilité toujours croissante des observations motive davantage l'exploration d'approches pilotées par les données comme une alternative puissante aux schémas classiques basés sur des modèles.

2. Exploiter les méthodes et les formulations développées pour la résolution des problèmes inverses en télédétection océanique multi-sources/multi-paramètres, avec un accent particulier sur la reconstruction à haute résolution des champs des dynamiques de la surface de la mer à partir de données satellitaires. En particulier, nous exploiterons des nouveaux développements en altimétrie satellitaire produisant des observations satellitaires bidimensionnelles à large fauchée, dans le cadre de la future mission satellitaire SWOT [59, 74, 78].

Motivations géophysiques pour des approches pilotées par les données

Comme indiqué précédemment, la quantité de données de télédétection océanique multi-sources et de données in situ disponibles a connu une croissance considérable au cours des dernières décennies, ce qui a été un facteur clé pour des possibilités d'analyse plus approfondies menant à une meilleure compréhension des dynamiques de la surface de l'océan, de la circulation océanique et des interactions atmosphère-océan. En particulier, des efforts considérables ont été déployés pour comprendre les relations et les interactions entre les différentes quantités physiques océaniques impliquées [4, 33, 97, 114, 160, 227].

En ce qui concerne la télédétection par satellite, les missions satellitaires actuelles (et futures) comportent généralement des observations partielles et irrégulièrement échantillonnées, ce qui est liée à la conception des orbites considérées, la géométrie des trajectoires des satellites, l'occlusion par des nuages, etc. De plus, les techniques d'interpolation utilisées actuellement ne parviennent pas à reconstruire avec précision les dynamiques de la surface de l'océan à plus petite échelle [34, 68, 73, 74]. A cet égard, l'identification des limites des missions de télédétection satellitaires actuelles a conduit au développement de nombreuses améliorations des technologies de télédétection, telles que la future mission altimétrique SWOT. Il est important de noter que SWOT sera la première mission d'altimétrie satellitaire à exploiter l'interférométrie radar pour produire des observations altimétriques bidimensionnelles à large fauchée, par opposition aux altimètres conventionnels qui ne peuvent produire des mesures altimétriques que sur la trace nadir du satellite [59, 74, 78]. Ces développements motivent des efforts actuels qui se concentrent à la fois sur l'exploration de la synergie entre les différents traceurs dynamiques de l'océan pour améliorer la reconstruction des dynamiques de la surface de l'océan, et sur le développement de méthodes alternatives de reconstruction à haute résolution des dynamiques océaniques à partir d'observations satellitaires à échantillonnage irrégulier.

D'autre part, la hauteur de la surface de la mer (SSH) est, comme indiqué dans [244], une quantité intégrée en profondeur qui contient des informations sur la structure de densité de la colonne d'eau et que capture des structures à méso-échelle de 50 km à quelques centaines de kilomètres, ce qui signifie que les courants de surface peuvent être directement extraits à partir des champs de SSH en exploitant le balance géostrophique. Au niveau des méso-échelles, ces courants interagissent avec les champs de température de surface de la mer (SST) à grande échelle, ce qui implique que les dynamiques de la surface de l'océan forment un système complexe d'interactions qui varient sur une large gamme d'échelles, à la fois dans l'espace et dans le temps [244].

En outre, l'existence d'un mode dynamique de la surface de l'océan, appelé mode SQG (Quasi-Géostrophie de surface), caractérisé par une fonction de transfert linéaire entre la température (SST) et la hauteur (SSH) de la surface de l'océan, a été démontrée dans des nombreux études théoriques et opérationnels [97, 99, 115, 131, 132, 142, 143, 244].

En effet, des travaux récents soulignent que la dynamique de la surface de l'océan peut être caractérisée par des relations linéaires locales SST-SSH qui correspondent précisément à des opérateurs Laplaciens fractionnaires [97, 115, 132, 142, 143, 244] qui peuvent être exprimés dans le domaine de Fourier comme :

$$\mathcal{F}_H(\widehat{SSH}) = -\gamma |\mathbf{k}|^{-2\alpha} \mathcal{F}_T(\widehat{SST}) \quad (1)$$

où \mathbf{k} est le vecteur de longueur d'onde horizontale, \mathcal{F}_T et \mathcal{F}_H sont respectivement des filtres linéaires de SST et SSH et γ est un coefficient de normalisation [99, 244]. La variation du paramètre α , qui contrôle le couplage effectif entre la SST et la SSH, conduit à différents modèles théoriques classiques [244]. Pour $\alpha = 1/2$, en particulière, on recourt au modèle quasi-géostrophique de surface [99, 131, 143, 244].

Sous l'hypothèse SQG, deux conclusions principales peuvent être tirées de l'équation (1) :

1. Les courants de surface, donnés par les gradients orthogonaux de la SSH, peuvent alors être calculés à partir des dérivés spatiales d'une version filtrée de la SST [114, 244].
2. Une seule fonction de transfert linéaire ne suffit pas pour capturer complètement la grande complexité des dynamiques de la surface de l'océan. En effet, la relation SQG (1) suppose des conditions de mélange vertical spécifiques, qui ne sont pas forcément valables partout et à tout moment. [85, 116, 143]. De plus, lorsque la dynamique de la surface de l'océan est réellement déterminée par des dynamiques SQG, le paramètre γ peut varier à la fois dans l'espace et dans le temps [99, 143, 244], par exemple en fonction de la profondeur de la couche de mélange.

Ces résultats ont motivé le développement de nouvelles méthodes statistiques pilotées par les données pour combiner l'information disponible dans les observations de la SSH avec d'autres

données de télédétection, telles que la SST ou les données in situ [4, 36, 244, 268]. Tandis que [85, 116] ont étudié des couplages linéaires SSH-SST en exploitant des représentations dans le domaine de Fourier (domaine fréquentiel), Tandeo et al. [244] se sont appuyés sur le fait que, pour des échelles horizontales entre 50 km et quelques centaines de kilomètres, la turbulence de surface est compatible avec la théorie des turbulences géostrophiques, pour conclure que la dynamique de la surface de l’océan peut être prédite à partir des variations de la densité horizontale en surface, possiblement dominées par des champs de SST. Ils ont exploité cette idée pour explorer les fonctions de transfert linéaires entre la SST et la SSH en introduisant une décomposition multimodale à l’aide d’un modèle de régression à classe latente, de sorte que les relations locales entre la SST et la SSH sont décrites par une fonction de transfert linéaire choisi parmi un nombre prédéfini de fonctions de transfert possibles.

Cependant, les résultats rapportés dans la littérature montrent qu’une seule fonction de transfert linéaire de type SQG ne suffit pas pour représenter avec précision toute les dynamiques de méso-échelle de la surface de l’océan dans une région donnée [244]. De plus, une limite importante des modèles de régression à classe latente est qu’ils ne peuvent tenir compte que d’un ensemble fini de fonctions de transfert linéaires. En revanche, le mode SQG est caractérisé par une seule classe de fonctions de transfert linéaires, à savoir l’opérateur Laplacien fractionnaire $\gamma\Delta^{1/2}$, mais implique un paramètre scalaire positif γ [99, 143, 244], qui est liée à des caractéristiques géophysiques locales (par exemple, la profondeur de la couche de mélange). Par conséquent, les dynamiques de la surface de de l’océan de type SQG, caractérisées par un paramètre γ variant continuellement, peuvent ne pas être bien représentées par des modèles de régression à classe latente.

Contributions

Dans le contexte présenté ci-dessus, la nécessité de formulations plus complexes permettant de répondre aux limites des modèles actuels, ainsi que la réussite des modèles sous contraintes dans des applications de séparation aveugle de sources, nous ont amenés à explorer les développements récents dans ce domaine pour essayer de développer des modèles de décomposition plus efficaces. Formellement, le problème de séparation aveugle de sources implique la décomposition d’un signal ou d’une image donné comme la superposition de différentes composantes :

$$\mathbf{y} = \sum_{k=1}^K \alpha_k \mathbf{s}_k + \boldsymbol{\omega} \quad (2)$$

où $\mathbf{y} \in \mathbb{R}^I$, le coefficient $\alpha_k \in \mathbb{R}$ quantifie la contribution de la composante $\mathbf{s}_k \in \mathbb{R}^I$, qui correspond au signal ou à l’image de référence k et $\boldsymbol{\omega} \in \mathbb{R}^I$ est un processus de bruit blanc Gaussien avec une covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{I \times I}$ qui modélise l’erreur d’estimation. Le principe de la séparation aveugle de sources consiste alors à exploiter des observations des signaux ou des images \mathbf{y} pour identifier et séparer les sources \mathbf{s}_k en quantifiant les coefficients de mélange α_k ,

de sorte que la reconstruction finale se rapproche du signal approximée \mathbf{y} . À cet égard, des contraintes supplémentaires peuvent être considérées pour imposer des restrictions physiques ou structurelles à la décomposition afin d'améliorer l'identifiabilité et l'interprétabilité du modèle. Ces dernières années, les contraintes de non-négativité et de parcimonie se sont révélées particulièrement intéressantes pour produire des décompositions plus pertinentes et plus interprétables.

Dans ce travail, nous visons à étendre les modèles de séparation aveugle des sources sous contraintes au problème de la caractérisation et décomposition d'opérateurs à partir des observations :

$$\mathbf{y}_n = \sum_{k=1}^K \alpha_{nk} f_k(\mathbf{x}_n) + \boldsymbol{\omega}_n \quad (3)$$

où $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$, et nous visons à identifier $\alpha_{nk} \in \mathbb{R}$, les coefficients de mélange qui modélisent la contribution de chaque mode à la reconstruction de \mathbf{y}_n pour un \mathbf{x}_n donné, et $f_k : \mathbb{R}^J \rightarrow \mathbb{R}^I$, une fonction, linéaire ou non-linéaire, associée au mode k . $\boldsymbol{\omega}_n \in \mathbb{R}^I$ est un processus de bruit, généralement un bruit Gaussien additif, représentant l'incertitude du modèle. Sur la base des progrès récents, nous nous concentrerons particulièrement sur des formulations linéaires sous contraintes de non-négativité ou parcimonie des coefficients de mélange α_{nk} . Même si cela dépasse le cadre de ce travail, nous soulignons qu'il est possible également d'envisager des contraintes supplémentaires sur les fonctions de transfert modales f_k .

Il est à noter qu'en étendant les formulations de séparation aveugle de sources au problème de la décomposition d'opérateurs, nous suggérons que les modes dynamiques en jeu ne s'excluent pas les uns les autres, mais se superposent. Ceci est significativement différent de l'hypothèse de base faite par les modèles de régression à classe latente, comme dans [244], qui supposent qu'un seul mode est actif à un emplacement spatio-temporel donné. A notre connaissance, ceci est la première étude à montrer, à partir d'une analyse pilotée par les données, la superposition continue de multiples modes dynamiques associés à différents types de fonction de transfert.

De plus, tandis que des travaux précédents se sont concentrés soit sur des couplages linéaires spatialement homogènes [85, 116], soit sur un ensemble fini de fonctions de transfert linéaires [244], nous allons considérer ici des représentations beaucoup plus riches, qui considèrent toutes les possibilités de mélange parmi un ensemble fini de familles de relations linéaires.

Dans le contexte de l'océanographie physique, cette thèse s'ajoute à un corpus de travaux en cours, à la fois théoriques et pratiques, qui montrent que la dynamique à méso-échelle de la surface l'océan peut être caractérisée par un couplage linéaire local entre la SST et la SSH [97, 115, 132, 142, 143, 244]. Nous soulignons de plus que des approches multimodales sont nécessaires, car une seule fonction de transfert linéaire ne suffit pas pour capturer complètement la variabilité spatio-temporelle, complexe et non-stationnaire, des dynamiques méso-échelle de la

surface de l’océan. Des travaux précédents ont plutôt exploré des fonctions de transfert moyennes régionales [85, 116] ou des modèles de mélanges à classes latentes [244]. En ce qui concerne la télédétection par satellite, dans le cadre de cette thèse, nous poursuivons des efforts pour améliorer les méthodes d’interpolation actuelles [57] ou développer des approches alternatives pour la reconstruction à haute résolution des dynamiques de la surface de l’océan à partir des données satellitaires échantillonnées irrégulièrement [67, 162, 257]. Plus précisément, nous nous concentrons sur l’exploitation des observations à large fauchée pour l’interpolation à haute résolution des champs d’anomalie de la hauteur de la surface de l’océan (SLA), dans la perspective de la future mission altimétrique SWOT.

Partie I : Décomposition non-négative d’opérateurs

Modèles et algorithmes

Comme nous l’avons indiqué précédemment, inspirés par le succès récent des approches de séparation de sources sous contraintes de non-négativité et parcimonie, nous visons à étendre les modèles de séparation sous contraintes à la caractérisation d’opérateurs à partir de données. Dans ce travail, nous traitons ces questions et développons des schémas mathématiques solides et efficaces. Nos principales contributions sont :

- Une formulation basée sur la méthode des moindres carrés dans l’espace d’observations, sous contrainte de non-négativité, ainsi que le développement de différents algorithmes d’estimation associées.
- Une reformulation du problème considérée comme un problème d’apprentissage de dictionnaires, pour gagner en flexibilité de modélisation, y compris la possibilité d’envisager d’autres contraintes, tels que la parcimonie.
- L’évaluation expérimentale des schémas numériques proposés, qui soulignent la pertinence du cadre d’apprentissage de dictionnaires.

Généralement, nous nous concentrerons sur les cas où des contraintes sont imposées soit sur les coefficients de décomposition, soit sur les modes de décomposition eux-mêmes. En particulier, nous introduisons un nouveau modèle de décomposition non-négative pour les opérateurs linéaires et étudions différents algorithmes d’estimation des paramètres, y compris une reformulation du modèle qui exploite les techniques d’apprentissage de dictionnaires. Plus particulièrement, cette reformulation nous permettra d’obtenir une plus grande souplesse et la possibilité de modifier les contraintes du modèle de façon simple.

Modèle de superposition non-négative de modes linéaires

Prenons un ensemble de données d’observation à plusieurs variables $\{\mathbf{x}, \mathbf{y}\}_n$, où $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$ désigne la n -ème paire d’observations. Les variables \mathbf{x}_n et \mathbf{y}_n peuvent, par exemple, faire référence

à des vecteurs, des patches d'images pour différentes modalités ou des états successifs d'un système dynamique, en fonction du cas d'étude considéré. Nous nous concentrons sur le modèle (3) en supposant que la relation potentiellement non-linéaire entre \mathbf{x}_n et \mathbf{y}_n , donnée par la réponse fonctionnelle $f_k(\mathbf{x}_n)$, peut être approximée localement, avec une précision raisonnable, par un opérateur linéaire. Nous considérons la décomposition de l'opérateur linéaire $f_k(\mathbf{x}_n)$ reliant les variables \mathbf{x}_n et \mathbf{y}_n sous des contraintes de non-négativité. Cela se traduit par le modèle suivant :

$$\begin{aligned} \mathbf{y}_n &= \sum_{k=1}^K \alpha_{nk} \beta_k \mathbf{x}_n + \boldsymbol{\omega}_n \\ \text{s.c. } &\begin{cases} \alpha_{nk} \geq 0, & \forall k \in \llbracket 1, K \rrbracket, \forall n \in \llbracket 1, N \rrbracket \\ \|\beta_k\|_F = 1, & \forall k \in \llbracket 1, K \rrbracket \end{cases} \end{aligned} \quad (4)$$

où $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$, $\alpha_{nk} \in \mathbb{R}^+$ sont des coefficients de mélange non-négatifs quantifiant la contribution du mode linéaire k à la reconstruction de \mathbf{y}_n pour un \mathbf{x}_n donné, $\beta_k \in \mathbb{R}^{I \times J}$ est une matrice de régression représentant le mode k , $\|\cdot\|_F$ est la norme de Frobenius et $\boldsymbol{\omega}_n \in \mathbb{R}^I$ est un processus de bruit Gaussien centré avec matrice de covariance Σ , représentant à la fois les incertitudes du modèle et les erreurs d'observation. N et K indiquent respectivement le nombre total d'observations et de modes, alors que $k \in \llbracket 1, K \rrbracket$ et $n \in \llbracket 1, N \rrbracket$ indiquent respectivement le mode et observation actuelles.

Une contrainte de non-négativité a été imposée sur les coefficients de mélange α_{nk} , inspiré par le succès des décompositions non-négatives pour des applications dans lesquelles une superposition positive de différentes parties se produit naturellement [155]. La contrainte de non-négativité permet de distinguer la forme des relations linéaires entre les variables \mathbf{x}_n et \mathbf{y}_n de la magnitude de ces relations. De plus, une contrainte de normalisation sur les matrices de régression modales β_k a été ajoutée pour éliminer les indéterminations d'échelle et améliorer l'identifiabilité du problème.

Le modèle (4) nous permet de résoudre des problèmes de décomposition impliquant des contributions linéaires mixtes et de généraliser les problèmes de mélange linéaires impliquant des régressions linéaires et les modèles de régression à classes latentes [47, 244]. Par rapport aux modèles de l'état de l'art, tels que la régression classique et les modèles de régression à classes latentes [47, 244], les principales caractéristiques de notre formulation proposée sont doubles : premièrement, elle tient compte des amplitudes possiblement variables des relations linéaires ; deuxièmement, elle évalue explicitement l'importance relative des différentes relations linéaires. Cela présente un grand intérêt pour diverses applications telles que les tests d'hypothèses de régression, l'identification de fonction de transfert, etc.

Algorithmes

Nous énonçons l'estimation des paramètres du modèle (4) à partir d'un ensemble d'observations $\{\mathbf{x}, \mathbf{y}\}_n$ comme la solution d'un problème d'optimisation sous contraintes non-linéaire et non-convexe :

$$\forall n, \begin{cases} [\hat{\alpha}_{nk}, \hat{\beta}_k] = \arg \min_{\alpha_{nk}, \beta_k} \sum_{m=1}^N W_m^n \left\| \mathbf{y}_m - \sum_{k=1}^K \alpha_n \beta_k \mathbf{x}_m \right\|_{\Sigma}^2 \\ \alpha_{nk} \geq 0, & \forall n \in \llbracket 1, N \rrbracket, \forall k \in \llbracket 1, K \rrbracket \\ \|\beta_k\|_F = 1, & \forall k \in \llbracket 1, K \rrbracket \end{cases} \quad (5)$$

où $\|\cdot\|_{\Sigma}$ est une norme pondérée selon la covariance Σ . Nous supposons que, selon les facteurs de pondération W_m^n , plusieurs paires d'observation $(\mathbf{x}_m, \mathbf{y}_m)$ peuvent partager des coefficients de mélange relativement similaires $\{\alpha_{nk}\}$. A cette égard, plus W_m^n est grand, plus les coefficients de mélange $\{\alpha_{nk}\}$ et $\{\alpha_{mk}\}$ sont similaires. La matrice de pondération \mathbf{W} peut représenter à la fois des a priori de régularité espace-temps, de sorte que les paires d'observations proches dans l'espace et/ou le temps doivent partager des décompositions d'opérateurs similaires, ainsi que des a priori de similarité dans l'espace des variables, de sorte que les paires d'observations similaires partagent des décompositions similaires. Cela semble raisonnable pour des nombreuses applications où les paramètres devraient être corrélés et varier de façon continue dans le domaine spatio-temporel considéré. Le paramétrage de la matrice de pondération \mathbf{W} devrait alors dépendre de l'application et peut être lié à des idées similaires utilisées dans la modélisation basée sur la covariance [16] et les schémas non-locaux [25, 207]. Concernant les problèmes d'identifiabilité, si le nombre de modes K vérifie $K > I$, l'estimation des coefficients de mélange α_{nk} devient impossible à partir d'une seule paire d'observations $(\mathbf{x}_n, \mathbf{y}_n)$. A cette égard, la matrice de pondération \mathbf{W} fournit également un moyen d'aborder l'estimation des paramètres de mélange dans de telles situations.

Algorithme des moindres carrés alternés

Approche par minimisation directe Compte tenu de la nature non-linéaire et non-convexe du problème de minimisation sous contraintes (5), l'estimation conjointe des paramètres de modèle α_{nk} et β_k n'est pas simple. Heureusement, ce problème de minimisation conjointement non-convexe devient convexe lorsque l'estimation est effectuée pour un ensemble de paramètres uniquement, tout en considérant l'autre ensemble de paramètres comme fixe. Naturellement, cela suggère une approche de minimisation alternée, qui conduit aux mises à jour suivantes des ensembles de paramètres du modèle α_{nk} et β_k , itérées jusqu'à la convergence :

- **β -step** : Minimisation sur β_k sous contraintes de normalisation avec α_{nk} fixe.
- **α -step** : Minimisation sur α_{nk} sous contraintes de non-négativité avec β_k fixe.

Approche par descente de gradient L'inconvénient de la simplicité de l'approche de minimisation directe est qu'elle est sujette aux problèmes numériques. Comme indiqué dans la littérature sur la séparation aveugles des sources [10], les projections alternées sur les espaces de solutions non-contraintes et contraintes peuvent induire un comportement divergent ou numériquement instable. Pour traiter un tel problème, l'approche de minimisation directe peut être adoucie en considérant une descente de gradient. Ceci est ensuite combiné avec une projection sur l'espace des solutions non-négatives, ce qui revient à une méthode de gradient proximal [37].

Algorithme d'apprentissage de dictionnaires

Le problème de décomposition considéré peut être reformulé comme un problème d'apprentissage de dictionnaires. Dans (4), l'opérateur linéaire $\sum_{k=1}^K \alpha_{nk} \beta_k$ peut être considéré comme la décomposition de l'opérateur linéaire local reliant les variables \mathbf{y} et \mathbf{x} pour l'index n . Cet opérateur linéaire local peut être estimé comme suit selon un critère des moindres carrés pondérés utilisant la matrice de pondération \mathbf{W} :

$$\Theta_n = \left(\sum_{m=1}^N W_n^m \mathbf{y}_m \mathbf{x}_m^T \right) \left(\sum_{m=1}^N W_n^m \mathbf{x}_m \mathbf{x}_m^T \right)^{-1} \quad (6)$$

où W_m^n sont des coefficients de pondération qui rendent compte des contributions relatives des paires d'observations $(\mathbf{x}_m, \mathbf{y}_m)$ à l'estimation de l'opérateur linéaire Θ_n associé à la paire d'observations $(\mathbf{x}_n, \mathbf{y}_n)$. Cette estimation des moindres carrés revient à résoudre indépendamment le critère des moindres carrés pour chaque index n dans (4).

A partir des opérateurs linéaires locaux $\{\Theta_n\}_n$, le problème (4) est lié à la décomposition non-négative de chaque opérateur linéaire Θ_n :

$$\begin{aligned} \Theta_n &= \sum_{k=1}^K \alpha_{nk} \beta_k + \Upsilon_n \\ \text{s.c.} \quad &\begin{cases} \alpha_{nk} \geq 0, & \forall k \in \llbracket 1, K \rrbracket, \forall n \in \llbracket 1, N \rrbracket \\ \|\beta_k\|_F = 1, & \forall k \in \llbracket 1, K \rrbracket \end{cases} \end{aligned} \quad (7)$$

ou la nouvelle matrice d'erreur Υ_n est une matrice Gaussienne.

De plus, ce problème de décomposition sous contraintes peut être approximée comme un problème de décomposition basée sur l'apprentissage de dictionnaires :

$$\begin{cases} [\hat{\mathbf{A}}, \hat{\mathbf{B}}] = \underset{\mathbf{A}, \mathbf{B}}{\operatorname{argmin}} & \|\Phi - \mathbf{B}\mathbf{A}\|_F^2 \\ \mathbf{A}_{kn} \geq 0, & \forall k \in \llbracket 1, K \rrbracket, \forall n \in \llbracket 1, N \rrbracket \\ \|\mathbf{B}_{:k}\|_2 = 1, & \forall k \in \llbracket 1, K \rrbracket \end{cases} \quad (8)$$

où la matrice $\Phi \in \mathbb{R}^{IJ \times N}$ est obtenue en concaténant des opérateurs vectorisés $\theta_n = \text{vec}(\Theta_n)$ (i.e. $\Phi = [\theta_1 | \dots | \theta_N]$), les colonnes de la matrice $\mathbf{A} \in \mathbb{R}^{K \times N}$ contiennent les coefficients de mélange α_{nk} quantifiant la contribution de chaque mode k pour la reconstruction de l'opérateur linéaire local vectorisé θ_n et les colonnes de $\mathbf{B} \in \mathbb{R}^{IJ \times K}$ (notées $[\mathbf{B}]_{:,k}$) contiennent des versions vectorisées des matrices de régression linéaire modales β_k .

L'estimation des paramètres du modèle (8) revient, sous cette nouvelle formulation, à un problème classique d'apprentissage de dictionnaires couplé à une contrainte de non-négativité. L'apprentissage de dictionnaires est un problème classique en traitement du signal, pour lequel de nombreuses méthodes, exploitant différentes contraintes, ont été proposées [3, 10, 13, 155, 156]. Ici, puisque nous considérons une contrainte de non-négativité, nous résolvons la minimisation (8) en utilisant une méthode de gradient proximal [37] pour tenir compte de la non-négativité de la matrice de coefficients de mélange \mathbf{A} . Cela implique l'itération des deux étapes suivantes jusqu'à la convergence :

- L'estimation par des moindres carrés de la matrice de dictionnaire. \mathbf{B} sous des contraintes de normalisation.
- L'estimation de la matrice de coefficients de mélange \mathbf{A} à l'aide d'une méthode de gradient proximal [37] pour appliquer la non-négativité.

Alternativement, on peut choisir une technique d'apprentissage de dictionnaires différente pour appliquer une contrainte différente (par exemple, KSVD [3] pour la parcimonie). Cela donne à la formulation basée sur l'apprentissage de dictionnaires une flexibilité et une adaptabilité accrues, étant donné que des contraintes de modèle alternatives peuvent ainsi être introduites de manière transparente dans le modèle (4).

Comparaison des algorithmes proposés

Nous avons évalué les performances des algorithmes présentés pour traiter le modèle de décomposition général (4) dans des cas idéaux et non-idéaux. Nous considérons les algorithmes introduits précédemment, à savoir l'algorithme des moindres carrés alternés et la décomposition d'opérateurs linéaires locaux par apprentissage de dictionnaires.

Nous étudions et comparons les algorithmes proposés en termes d'identification de paramètres et de performances de reconstruction des observations dans divers environnements expérimentaux, notamment des environnements idéaux sans bruit, des cas impliquant un nombre variable de modes de décomposition et des configurations considérant des observations et des paramètres bruitées, afin de mieux comprendre la robustesse et les limites des algorithmes proposés. Nous discutons ensuite de l'apport principal de notre approche par rapport aux méthodes de l'état de l'art pour l'analyse et la décomposition d'opérateurs.

Nous rapportons des expériences numériques pour évaluer les modèles et les algorithmes proposés. Nous exploitons des données synthétiques pour effectuer une analyse quantitative de la performance d'estimation et une analyse de sensibilité par rapport à des paramètres clés et des hypothèses de modélisation.

Globalement, la décomposition par dictionnaire des opérateurs linéaires locaux semble offrir les meilleures performances en termes d'identification du modèle, de stabilité et de complexité de calcul dans des conditions favorables. Alternativement, dans des configurations non-idéales, des algorithmes moins stables, tels que l'algorithme des moindres carrés alternés, peuvent néanmoins s'avérer utiles pour l'identification du modèle et la reconstruction des observations.

À cet égard, les résultats rapportés suggèrent la nécessité de considérer des contraintes et/ou des a priori de régularisation supplémentaires pour résoudre les problèmes d'identifiabilité du modèle (4) dans des configurations non-idéales. En effet, même si nos expériences suggèrent que le modèle et les algorithmes proposés offrent de bonnes performances de reconstruction dans la plupart des configurations considérées, ce qui les rend adaptés à la plupart des problèmes de reconstruction/prédiction, l'identification du modèle semble être considérablement sensible aux configurations non-idéales, où l'hypothèse de partage des paramètres est assouplie ou lorsque le nombre ou la sélection d'observations auxiliaires pour l'estimation des paramètres induit des erreurs. Les résultats suggèrent également que l'identifiabilité du modèle peut être améliorée en introduisant des approches d'estimation robustes pour les opérateurs linéaires locaux et/ou des contraintes de modèle supplémentaires.

Cependant, les modèles et algorithmes proposés ont été utilisés avec succès dans des applications de reconstruction/prédiction et de segmentation [165, 166, 169]. Ces applications soulignent la pertinence de la décomposition non-négative proposée par rapport aux méthodes basées sur l'orthogonalité ou les régressions à classe latente, qui sont considérées dans la plupart des travaux précédents [85, 90, 116, 244].

Applications

Dans ce travail, nous présentons aussi l'application des algorithmes proposés à plusieurs problèmes impliquant à la fois la segmentation et la reconstruction/prédiction. Nous décrivons les avantages des formulations proposées par rapport à des modèles classiques pour une série de problèmes inverses en océanographie physique et en géosciences. En particulier, nous nous concentrons sur la caractérisation des dynamiques de la surface de l'océan à partir de différents traceurs satellitaires océaniques (SST, SSH, SSS), ainsi que sur des problèmes de reconstruction/prédiction des systèmes dynamiques.

D’abord, nous illustrons l’utilité du modèle de décomposition linéaire non-négative proposé pour la prédiction par analogues [175] des systèmes dynamiques, avec un cas d’étude considérant le système dynamique de Lorenz ’96. Les résultats obtenus indiquent qu’intégrer le modèle de décomposition non-négative d’opérateurs locaux dans un schéma de prédiction par analogues améliore l’identifiabilité du modèle pour des petits ensembles de données d’apprentissage et rends la méthode plus robuste au bruit.

Nous illustrons une deuxième application pour la caractérisation des dynamiques de la surface de l’océan à partir de la synergie présentée par différents champs géophysiques, à savoir la température de la surface de la mer (SST) et la salinité de la surface de la mer (SSS). Nous analysons les relations entre la SST et la SSS dans la mer d’Alboran, une région qui présente de fortes patterns saisonniers associées à l’apport d’eau froide de l’Atlantique à travers le détroit de Gibraltar. Nos résultats suggèrent que la décomposition proposée est capable de capturer ces patterns saisonniers et de séparer avec précision deux relations SST-SSS opposées, démontrant ainsi la pertinence du schéma proposé pour la séparation des processus physiques à partir de la synergie entre différents traceurs.

Dans une troisième application, nous illustrons la pertinence des modèles proposés pour étudier et décomposer les dynamiques de la surface de l’océan à partir des synergies entre données multi-capteurs, à savoir la température de la surface de la mer (SST) et la hauteur de la surface de la mer (SSH), dans la région d’Agulhas. Nous rappelons que, comme indiqué précédemment, d’un point de vue théorique [131, 143], les champs de SST et SSH devraient être liés entre eux par des fonctions de transfert linéaires qui correspondent exactement à des opérateurs Laplaciens fractionnés [97, 115, 132, 142, 143, 244]. Notre modèle surpasse de manière significative le modèle de régression à classes latentes développé dans [244] de plus d’un ordre de grandeur en termes d’EQM de prédiction de la SSH. De plus, nous montrons que notre modèle récupère un premier mode dynamique fortement corrélé à la dynamique SQG, c’est-à-dire à une fonction de transfert Laplacien fractionnaire, ainsi qu’un deuxième mode dynamique qui agit comme une correction locale par rapport au premier mode. A cet égard, notre modèle fournit un moyen simple d’évaluer localement dans quelle mesure la dynamique SQG s’applique. Enfin, ces découvertes corroborent réellement l’existence de modes dynamiques superposés, plutôt que de modes mutuellement exclusifs, comme supposé dans [244].

Les applications rapportées illustrent clairement la pertinence du modèle proposé en tant qu’alternative intéressante aux formulations classiques faisant appel à des a priori d’orthogonalité ou à des formulations plus simples, telles que la régression linéaire classique ou les modèles à classes latentes.

Partie II : Interpolation de champs d'altimétrie à partir de données de télédétection satellitaire

Dans une deuxième partie de ce travail, nous nous éloignons des problèmes inverses au sens général, pour nous orienter vers un type spécifique de problème inverse, l'interpolation à haute résolution des cartes d'anomalie de la hauteur de la surface de l'océan (SLA) à partir d'observations satellitaires. Dans ce contexte, nous suivons initialement [66] et explorons une reformulation de notre modèle de décomposition non-négative comme un problème de super-résolution d'images. Par la suite, nous suivons le travail présenté dans [67, 161, 162] et abordons le problème du point de vue de l'assimilation de données. Nous explorons aussi des stratégies pour mieux exploiter les observations satellitaires altimétriques et leur synergie, ainsi que des stratégies potentielles pour inclure des traceurs océaniques supplémentaires (température de la surface de la mer, courants géostrophiques, etc.).

Aujourd'hui, la plupart de sources de données de télédétection utilisent différentes stratégies d'échantillonnage à plusieurs résolutions spatio-temporelles, et impliquent généralement un échantillonnage irrégulier de la surface de l'océan ainsi qu'une grande proportion de données manquantes, en raison de l'occlusion par des nuages, les caractéristiques des orbites des satellites, etc. Dans ce contexte, l'élaboration de produits d'altimétrie dans une grille régulière et sans données manquantes est d'un intérêt majeur pour la recherche et l'industrie.

Comme nous l'avons déjà mentionné, le problème d'interpoler un champ dans une grille régulière à partir de données échantillonnées de manière irrégulière appartient à la famille des problèmes inverses, qui ont été étudiés de manière approfondie dans la littérature [63, 177, 195, 207]. En sciences de la mer, les méthodes de pointe reposent sur l'interpolation optimale (OI) [20, 57, 152], une approche basée modèle qui consiste à modéliser la structure de covariance spatio-temporelle du champ à interpoler, généralement sous l'hypothèse que cette structure de covariance reste constante dans l'espace et dans le temps. Une telle hypothèse n'est pas toujours vérifiée, ce qui conduit à un manque de précision dans la représentation des dynamiques des structures à plus petite échelle. En effet, l'hypothèse Gaussienne inhérente aux techniques OI provoque le lissage des champs reconstruits, et il a été vérifié qu'il n'est pas possible de récupérer les structures à petite échelle (entre 10 et 100 km) à l'aide d'OI [34, 68, 73].

D'une autre part, les techniques d'assimilation de données [61] sont considérées comme un outil basé modèle particulièrement puissant pour l'interpolation de champs géophysiques à haute résolution, à condition que le modèle physique choisi soit suffisamment précis et n'implique pas un coût computationnel prohibitif. De plus, la plupart des méthodes alternatives qui ne reposent pas sur l'OI ou l'assimilation de données restent principalement basées sur des modèles, malgré la disponibilité croissante de jeux de données volumineux, issus à la fois de la télédétection

satellitaire ou d'observations in situ, ainsi que de simulations numériques ou de réanalyses, qui soutient l'exploration des approches pilotées par les données comme une alternative puissante et efficace [67, 207].

Dans ce contexte, nous suivons des développements récents [65–67, 162, 207] qui, compte tenu de la grande variété de données d'observation, de simulation et de ré-analyse à haute résolution disponibles, explorent des stratégies pilotées par les données comme une alternative puissante et efficace aux approches basées sur des modèles pour l'interpolation de champs géophysiques à haute résolution. Les approches pilotées par les données visent, en général, à exploiter les ensembles de données disponibles pour émuler le modèle physique derrière le processus d'intérêt et mieux capturer les variabilités spatio-temporelles qui ne peuvent pas être complètement expliqués par des modèles simplifiés, purement numériques.

D'un point de vue opérationnel, parmi les avancements récentes en télédétection, la mission satellite SWOT (Surface Water and Ocean Topography), issue d'une collaboration entre la NASA, le Centre Français d'Études Spatiales (CNES) et les agences spatiales britannique et canadienne, vise à fournir des cartes altimétriques à haute résolution pour l'hydrologie et l'océanographie. Plus précisément, la mission SWOT sera la première mission à exploiter des altimètres d'interférométrie radar en bande Ka produisant, pour la première fois, des observations satellitaires bidimensionnelles à large fauchée [59, 74, 78]. De plus, la très haute résolution spatiale de la mission (~ 1 km) devrait nous permettre de mieux capturer et de mieux comprendre les processus à méso-échelle et subméso-échelle [59, 74], ce qui constitue l'un des plus grands défis dans la recherche océanographique actuelle.

Par la suite, nous tenterons d'aborder le problème de l'interpolation de champs géophysiques selon différentes perspectives méthodologiques pilotées par les données. Suivant les formulations précédemment introduites [66, 169], nous formulerons d'abord le problème comme un problème de super-résolution d'images exploitant des modèles de convolution, pour étendre ainsi les approches de décomposition d'opérateurs présentées précédemment à ce nouveau problème. Ensuite, nous suivrons [67, 161, 162] pour porter notre attention vers des approches d'assimilation pilotées par les données. En préparation pour le lancement de la mission SWOT en 2021, nous explorons également le potentiel d'exploiter les observations SWOT pour la reconstruction de champs altimétriques haute-résolution dans le contexte des méthodes pilotées par les données.

Modèles convolutionnels localement adaptés pour la super-résolution des champs d'altimétrie

Dans un premier temps, nous nous concentrons sur l'application des modèles et méthodes introduits dans la première partie de cette thèse au problème de l'interpolation pilotée par les données des champs de SLA à partir d'observations satellitaires échantillonnées de manière irré-

gulière. À cette fin, nous suivons [63, 66, 169], et reformulons le problème en tant que problème de super-résolution d'images pour exploiter les modèles de décomposition d'opérateurs proposés pour la décomposition d'opérateurs convolutionnels de super-résolution.

La super-résolution est un problème classique en traitement de l'image qui a été largement étudié dans la littérature [82, 231]. Les avancements récents en apprentissage automatique et l'avènement d'outils de calcul de plus en plus puissants rendent les modèles de super-résolution particulièrement intéressants pour des problèmes complexes, tels que ceux liés aux géosciences, à l'océanographie et à la météorologie. À cet égard, des techniques de super-résolution ont déjà été appliquées à des problèmes d'amélioration des images de télédétection [272].

Contrairement aux applications classiques de super-résolution, des nombreuses applications en télédétection satellitaire impliquent non seulement des images à basse résolution, mais également des informations complémentaires à haute résolution échantillonnées de manière irrégulière. À cet égard, la disponibilité de telles données partielles à haute résolution motive le développement de modèles de super-résolution adaptés localement, dans le but de prendre en compte la variabilité spatio-temporelle des processus d'intérêt. À cette fin, la décomposition est exploitée comme un moyen d'adapter localement des opérateur convolutionnels de super-résolution à des domaines spatiaux plus petits, augmentant ainsi la localisation du modèle et les performances de reconstruction. Nous explorons également d'autres contraintes, notamment l'orthogonalité et la parcimonie, pour illustrer les avantages de la formulation non-négative dans ce contexte particulier.

Nous traitons donc un problème de super-résolution d'image à partir d'informations haute résolution échantillonnées de manière irrégulière. Suivant des approches récentes de super-résolution [2, 252, 253], nous considérons des modèles convolutionnels adaptés localement. Nos contributions méthodologiques sont doubles :

- Les modèles convolutionnels proposés combinent à la fois des images basse résolution et une source d'information secondaire à haute résolution, permettant ainsi la fusion de différentes sources de données hétérogènes dans un cadre de super-résolution simple et unifié.
- Nous explorons les représentations basées sur des dictionnaires pour des opérateurs de convolution avec différents types de contraintes, à savoir des contraintes d'orthogonalité, de non-négativité et de parcimonie [12, 273]. Ces représentations sous contraintes basées sur des dictionnaires sont particulièrement pertinentes pour générer des modèles de super-résolution adaptés localement et calibrés à partir d'un faible nombre de données d'entraînement à haute résolution.

Nous appliquons la méthode proposée à des données multi-source de télédétection océanique. Nous abordons la reconstruction d'images haute résolution de la SLA à partir de données

d'altimétrie satellitaire échantillonnées irrégulièrement, d'images haute résolution de la SST, et d'images basse résolution de la SLA. De cette façon, la super-résolution multimodale des champs altimétriques est faite en prenant en compte une source d'information haute résolution supplémentaire. A cette égard, la fusion de données pose un problème dans un grand nombre d'applications de télédétection, où plusieurs capteurs associés à différentes stratégies d'échantillonnage peuvent contribuer à la reconstruction d'une image haute résolution, ce qui rend la formulation proposée particulièrement pertinente pour des applications réelles.

D'un point de vue méthodologique, nous complétons aussi les modèles précédents de super-résolution par convolution [2,253] par l'évaluation de différentes décompositions sous contraintes basées sur des dictionnaires, et par l'utilisation d'une source d'information haute résolution complémentaire. À cet égard, les décompositions basées sur les dictionnaires sont considérées comme un moyen pour mieux prendre en compte les variabilités spatio-temporelles par le biais des paramétrisations du modèle adaptées localement.

Nous considérons un cas d'étude en Méditerranée Occidentale, où nous rapportons des simulations numériques qui exploitent des expériences de simulation d'un système d'observation (OSSE) pour évaluer les performances de la formulation proposée pour l'interpolation des champs de SLA, en considérant plusieurs dictionnaires sous différentes contraintes. Nos expériences démontrent la pertinence des modèles de super-résolution proposés, en particulier sous des contraintes de non-négativité, pour obtenir une meilleure adaptation locale et produire une meilleure reconstruction des détails à haute résolution, par rapport aux approches classiques telles que l'OI.

Interpolation pilotée par les données de champs d'altimétrie à partir de données de télédétection satellitaire

Malgré les résultats prometteurs obtenus avec les modèles convolutionnels, le succès récent des modèles alternatifs pilotés par les données [8, 66, 67, 162, 169, 213] nous a motivé à explorer d'autres méthodes pilotées par les données pour l'interpolation de champs altimétriques à partir de données satellitaires. En effet, la disponibilité toujours croissante des jeux de données de télédétection, in situ et de simulation motive véritablement l'exploration des approches pilotées par les données en tant qu'une alternative puissante aux schémas classiques basés sur des modèles. Ces schémas pilotés par les données [8, 66, 67, 162, 169, 213] permettent d'avoir un cadre générique et efficace pour l'exploration de la synergie entre les données multi-source, cars ils ne nécessitent pas d'une vérification préalable des a priori géophysiques spécifiques. De plus, ils surpassent considérablement les approches classiques telles que l'OI. Parmi les schémas évalués, l'assimilation analogue (AnDA) [67, 161, 162] apparaît comme particulièrement pertinente pour

modéliser explicitement les dépendances temporelles des dynamiques de la surface de l’océan.

En particulier, nous nous concentrons sur l’exploitation des observations synthétiques SWOT, en préparation du lancement de la mission SWOT en 2021 [59,74]. À cet égard, nous explorons le potentiel des observations SWOT, combinées ou non avec des données nadir (au long de la trace satellitaire), pour surmonter les limites des produits altimétriques actuellement disponibles pour la reconstruction de champs d’altimétrie à haute résolution. Il est important de noter qu’une attention particulière est portée à la caractérisation de la contribution des observations SWOT pour la reconstruction des dynamiques de la surface de l’océan à plus petite échelle.

Nous considérons un cas d’étude en Méditerranée Occidentale, une région présentant des processus à méso-échelle avec des échelles horizontales caractéristiques inférieures à 100 km [6,181,203,204,225]. Nous effectuons des expériences de simulation d’un système d’observation (OSSE) en considérant des stratégies d’échantillonnage spatio-temporel réelles appliquées à des données synthétiques de vérité terrain à haute résolution, afin d’évaluer les performances de différents schémas d’interpolation pilotés par les données pour la reconstruction de champs de SLA à partir de données satellitaires multi-sources. Il est important de noter que nous considérons différents stratégies d’échantillonnage spatio-temporel, notamment celles liées à des sources altimétriques conventionnelles produisant des observations uniquement au long des traces satellitaires [204], ainsi que celles liées à la future mission SWOT produisant des observations bidimensionnelles à large fauchée, et explorons l’analyse conjointe des deux sources de données altimétriques. En particulier, la pertinence d’AnDA en tant que méthode privilégiée pour évaluer l’impact des différentes stratégies d’échantillonnage spatio-temporel considérées est démontrée. De plus, nous explorons ensuite l’utilisation de sources de données alternatives, telles que les champs de SLA à basse résolution (résolus par OI) ou les champs de gradient de SLA (résolus numériquement), comme sources d’informations complémentaires pour améliorer les performances de reconstruction d’AnDA.

Enfin, les OSSE considérées impliquent des niveaux de bruit non corrélés réalistes à la fois pour les données nadir et pour les données SWOT. Toutefois, pour les données SWOT, des sources de bruit corrélées sont également attendues [59]. Étant donné qu’AnDA s’est révélée extrêmement sensible à ces sources de bruit corrélées, nos résultats devraient être considérés comme une limite supérieure de l’amélioration attendue de la mission SWOT pour la reconstruction de champs altimétriques. Il est important de noter que ces sources de bruit doivent être correctement prises en compte dans le modèle d’interpolation considéré [171]. À cet égard, nous introduisons l’assimilation des gradients des observations SWOT (résolus numériquement) comme stratégie possible pour traiter les sources de bruit corrélé contaminant les données SWOT.

Globalement, nos résultats confirment la pertinence d'AnDA par rapport aux approches de l'état de l'art (OI [57,204], MS-VE-DINEOF [161,213] et NN-LLOD [169]) pour une reconstruction améliorée des structures à méso-échelle des champs de SLA, pour des échelles horizontales allant de ~ 20 km à ~ 100 km. En effet, nous rapportons une amélioration de 32% (12%) en termes d'EQM (corrélation) pour la reconstruction des champs de SLA et de 29% (27%) en termes d'EQM (corrélation) pour la reconstruction des champs de gradient de SLA par AnDA à partir des données nadir (par rapport à OI). Nos expériences suggèrent également une amélioration supplémentaire des performances de reconstruction si nous considérons l'assimilation AnDA conjointe des observations SWOT et nadir : 42% (12%) en termes d'EQM (corrélation) pour la reconstruction des champs de SLA et 34% (29%) en termes d'EQM (corrélation) pour la reconstruction des champs de gradient de SLA (par rapport à OI).

Ces résultats mettent en évidence le potentiel des observations SWOT pour fournir plus d'informations que les observations altimétriques nadir conventionnelles. Globalement, nos expériences confirment réellement le potentiel des observations SWOT pour mieux contraindre AnDA et améliorer la reconstruction des dynamiques de méso-échelle des champs altimétriques. Nous pouvons tirer trois conclusions principales :

- L'utilisation conjointe des observations nadir et des observations SWOT dans AnDA apparaît naturellement comme la meilleure stratégie, avec un gain relatif d'interpolation d'environ 14,71% (resp. 42,36%) en termes d'EQM par rapport à AnDA (resp. OI) appliqué uniquement aux données nadir.
- Contrairement à ce qui se passe pour les données nadir, l'accumulation de données SWOT sur plusieurs jours à une incidence négative sur les performances d'interpolation.
- L'échantillonnage temporel irrégulier des observations SWOT pour une région donnée, impliquant généralement une période de revisite de 10 jours pour les latitudes moyennes, entraîne une variabilité temporelle des performances d'interpolation relativement grande.

Conclusion

Au long de cette thèse, nous avons cherché à explorer des approches pilotées par les données pour traiter différents problèmes inverses dans le contexte de l'océanographie physique et des géosciences, avec un accent particulier sur la reconstruction des dynamiques de la surface de l'océan à partir d'observations satellitaires. À cet égard, nous avons exploré l'exploitation de méthodologies pilotées par les données pour la résolution des problèmes de segmentation et de reconstruction dans des contextes scientifiques multiples. Il est important de noter que ces travaux contribuent à des efforts conjoints actuels des communautés des géosciences et du traitement du signal pour exploiter la disponibilité toujours croissante de données de télédétection, in situ et de modélisation/réanalyse afin d'améliorer notre compréhension des dynamiques des

océans et des interactions océan-atmosphère. À cet égard, cette thèse est, compte tenu de son sujet, à l'interface entre les sciences des données et les géosciences. En tant que tel, nous avons essayé de combler en partie le fossé entre le traitement du signal et les problèmes en géosciences, afin d'améliorer les modèles utilisés actuellement pour la segmentation et la reconstruction des dynamiques de la surface de l'océan.

Inspirés par le succès récent des formulations de séparation aveugle de sources sous contraintes de non-négativité et parcimonie pour des applications de traitement du signal et des images, dans la première partie de ce travail nous avons étendu les modèles de séparation aveugle de sources sous contraintes au problème de la caractérisation et décomposition d'opérateurs à partir d'observations. Nous avons développé des modèles pertinents et proposé des algorithmes efficaces reposant sur des fondations mathématiques solides pour l'estimation des paramètres du modèle. Nous avons également exploré plusieurs applications pour la segmentation des dynamiques de la surface de l'océan à partir de données de télédétection satellitaire et pour la reconstruction/prédiction de systèmes dynamiques.

Dans une deuxième partie de cette thèse, nous nous sommes concentrés sur un problème particulier, l'interpolation de champs géophysiques à haute résolution à partir de données échantillonnées de manière irrégulière, avec un intérêt particulier pour l'interpolation des champs d'anomalie de la hauteur de la surface de l'océan (SLA) à partir d'observations altimétriques satellitaires. En préparation du lancement de la mission SWOT en 2021, nous avons aussi exploré l'effet de différentes stratégies d'échantillonnage, plus particulièrement celles relatives à l'altimétrie nadir classique et à l'interférométrie radar à large fauchée.

Pour conclure, nous voudrions souligner le potentiel des modèles de décomposition d'opérateurs à partir d'observations en tant qu'outil d'analyse des couplages de différents traceurs océaniques. Nous espérons en particulier que les modèles développés permettront de mieux comprendre les interactions entre les différentes variables océaniques, les processus régissant les dynamiques des océans, et le rôle de l'océan au sein du système terrestre mondial.

Nous espérons également que nos découvertes contribueront à des nouveaux développements en termes de segmentation des dynamiques de la surface de l'océan, d'interpolation à haute résolution des champs géophysiques à partir d'observations haute résolution multi-sources, et d'autres applications connexes, telles que la reconstruction à haute résolution des courants de surface de la mer, entre autres.

Abstract

Keywords: *Satellite remote sensing, Massive datasets, Operator decomposition, Non-negativity, Sparsity, Ocean dynamics, Data mining, Statistical learning, Multi-sensor data fusion, High-resolution reconstruction, Interpolation.*

In the last few decades, the ever-growing availability of multi-source ocean remote sensing data has been a key factor for improving our understanding of upper ocean dynamics, ocean circulation and atmospheric-ocean interactions. In this regard, a largely under-exploited potential exists within available datasets. Developing efficient approaches to better exploit these datasets is of major importance to address, through data mining and statistical learning approaches, problems such as multi-sensor data fusion, high-resolution reconstruction or multi-parameter analysis of processes of interest. Particularly, the decomposition of geophysical processes into relevant modes is a key issue for characterization, forecasting and reconstruction problems. Inspired by recent advances in blind source separation brought forth with the introduction of non-negative and sparse formulations, we aim, in the first part of this thesis dissertation, at extending constrained blind source separation models to the problem of observation-based characterization and decomposition of linear operators or transfer functions between variables of interest, with an emphasis on a non-negative setting. In this work, we develop mathematically-sound and computationally-efficient schemes, including the reformulation of the constrained operator decomposition problem as the constrained dictionary-based decomposition of locally estimated linear operators, thus enhancing model flexibility. We illustrate the relevance of the proposed decomposition models in different applications involving the analysis, segmentation and forecasting of geophysical dynamics. Subsequently, given that the ever-increasing availability of multi-source datasets supports the exploration of data-driven alternatives to classical model-driven formulations, we explore, in the second part of this work, recently introduced data-driven models for the interpolation of geophysical fields from irregularly-sampled satellite-derived observations. Importantly, with a view towards the future SWOT mission, the first satellite mission to produce complete two-dimensional wide-swath satellite altimetry observations, we focus on assessing the extent to which SWOT data, combined or not with conventional nadir along-track data, may lead to an improved reconstruction of altimetry fields.

Part of this work has been presented at, submitted for review to and/or published in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on Image Processing (ICIP), IEEE International Geoscience and Remote Sensing Symposium (IGARSS) and IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. Submissions available at [166, 169–171]. This work was supported by ANR (grant ANR-13-MONU-0014), Labex CominLabs project SEACS and OSTST project MANATEE.

Acronyms

For the sake of legibility, the meaning of abbreviations and acronyms defined here is also recalled at its first appearance in the text of a chapter.

ADT	Absolute Dynamic Topography
ALS	Alternating Least Squares
ANR	Anchored Neighborhood Regression
AT	Along-track
A+	Adjusted Anchored Neighborhood Regression
CCA	Canonical Correlation Analysis
CNN	Convolutional Neural Networks
DINEOF	Data Interpolating Empirical Orthogonal Functions
VE-DINEOF	Variable EOFs DINEOF
MS-VE-DINEOF	Multi-scale Variable EOFs DINEOF
DMD	Dynamic Mode Decomposition
EDMD	Extended Dynamic Mode Decomposition
EM	Expectation-Maximization
EnKFS	Ensemble Kalman Filter and Smoother
EnKS	Ensemble Kalman Smoother
EOF	Empirical Orthogonal Functions
FSLE	Finite Size Liapunov Exponents
FVE	Fraction of Variance Explained
FVU	Fraction of Variance Unexplained
ICA	Independent Component Analysis
JOR	Jointly Optimized Regressors
KLT	Karhunen–Loève Transform
KNN	K Nearest Neighbours
LLE	Local Linear Embedding
LLOD	Local Linear Operator Decomposition
MADT	Maps of Absolute Dynamic Topography

MAP	Maximum A Posteriori
MDT	Mean Dynamic Topography
MFS	Mediterranean Ocean Forecasting System
MICE	Moment-based Iterative Conditional Estimation
MIMO	Multiple Input Multiple Output
MMSE	Minimum Mean Square Error
MP	Matching Pursuit
MSE	Mean Squared Error
nMSE	Normalized Mean Squared Error
rMSE	Relative Mean Squared Error
MSS	Mean Sea Surface
NNLS	Non-negative Least Squares
NMF	Non-negative Matrix Factorization
NN	Non-negative
NN-LLOD	Non-negative Local Linear Operator Decomposition
ODE	Ordinary Differential Equation
OI	Optimal Interpolation
OMP	Orthogonal Matching Pursuit
OSS	Observing System Simulation Experiment
OST	Ocean Surface Topography
PCA	Principal Component Analysis
PDF	Probability Density Function
POD	Proper Orthogonal Decomposition
PSF	Point Spread Function
QG	Quasi-Geostrophy
RMSE	Root Mean Squared Error
rRMSE	Relative Root Mean Squared Error
nRMSE	Normalized Root Mean Squared Error
ROMS	Regional Ocean Modeling System
SAR	Synthetic Aperture Radar
SLA	Sea Level Anomaly
SNR	Signal-to-Noise Ratio
SQG	Surface Quasi-Geostrophy
SSH	Sea Surface Height
SSS	Sea Surface Salinity
SST	Sea Surface Temperature

SVD	Singular Value Decomposition
SWH	Significant Wave Height
SWOT	Surface Water Ocean Topography
WMOP	Western Mediterranean Operational Forecasting System

Notation

We define hereafter the notation to be used throughout the present thesis dissertation.

Linear algebra

x	Scalar
\mathbf{x}	Vector
\mathbf{X}	Matrix
\mathcal{X}	Set

Sets

\mathbb{C}	Set of complex numbers
\mathbb{R}	Set of real numbers
\mathbb{N}	Set of natural numbers
\mathbb{R}^+	Set of positive real numbers
\mathbb{R}^N	Set of real-valued vectors of size N
$\mathbb{R}^{N \times M}$	Set of real-valued matrices of size $N \times M$
$\{\cdot\}_n$	Set of elements enumerated according to n
$\#\mathcal{X}$	Cardinality of set \mathcal{X} , <i>i.e.</i> , the number of elements in set \mathcal{X}

Symbols and operators

\mathbf{I}_n	Identity matrix of size $n \times n$
$\mathbb{1}_{\mathcal{X}}$	Indicator function with value $0 \forall x \in \mathcal{X}$ and $+\infty$ otherwise
\mathbf{X}^T	Transposition operator applied to matrix \mathbf{X}
\mathbf{C}^H	Conjugate transposition operator applied to complex-valued matrix \mathbf{C}
\odot	Hadamard (element-wise) product
$[\mathbf{X}]_{kl}$	Element from the k -th line and l -th column of matrix \mathbf{X}
$[\mathbf{X}]_k$	k -th line of matrix \mathbf{X}

$[\mathbf{X}]_{:l}$	l -th column of matrix \mathbf{X}
$[\mathbf{X}]_{k,m:n}$	Vector built from the m -th to n -th elements of the k -th line of matrix \mathbf{X}
$[\mathbf{X}]_{k:l,n}$	Vector built from the k -th to l -th elements of the n -th column of matrix \mathbf{X}
$[\mathbf{X}]_{k:l,:}$	Matrix built from the k -th to l -th lines of matrix \mathbf{X}
$[\mathbf{X}]_{:,m:n}$	Matrix built from the m -th to n -th columns of matrix \mathbf{X}
$[\mathbf{X}]_{k:l,m:n}$	Matrix built from the k -th to l -th lines and the m -th to n -th columns of matrix \mathbf{X}
$[\mathbf{x}]_k$	k -th element of vector \mathbf{x}
$\ \mathbf{X}\ _F$	Frobenius norm, <i>i.e.</i> , $\sqrt{\text{tr}\mathbf{X}^T\mathbf{X}}$
$\ \mathbf{x}\ _2$	Euclidean (ℓ_2) norm, <i>i.e.</i> , $\sqrt{\sum_i x_i^2}$
$\ \mathbf{x}\ _1$	ℓ_1 -norm, <i>i.e.</i> , $\sum_i x_i $
$\ \mathbf{x}\ _0$	ℓ_0 -norm, <i>i.e.</i> , number of non-zero elements of \mathbf{x}
$\ \mathbf{X}\ _*$	Nuclear norm, <i>i.e.</i> , $\sum_r \lambda_r(\mathbf{X})$, where $\lambda_r(\mathbf{X})$ are the singular values of \mathbf{X}
$\text{tr}(\cdot)$	Trace of a square matrix
$\text{vec}(\cdot)$	Vectorization operator
$\nabla(\cdot)$	Nabla operator (Gradient)
δ_{ij}	Kronecker delta with value 1 if $i = j$ and 0 otherwise

Probability, statistics and random variables

X	Random variable
$E(X)$	Expectation/Expected value of X
$E(X Y)$	Conditional expectation of X with respect to Y
$\text{var}(X)$	Variance of X , <i>i.e.</i> , $E(X^2) - E(X)^2$
$\text{var}(\mathbf{x})$	Covariance matrix of vector \mathbf{x} , <i>i.e.</i> , $E(\mathbf{x}\mathbf{x}^T) - E(\mathbf{x})^2$
$\text{cov}(X, Y)$	Covariance of X and Y , <i>i.e.</i> , $E((Y - E(X))(Y - E(Y)))$
$\text{cov}(\mathbf{x}, \mathbf{y})$	Cross-covariance matrix of vectors \mathbf{x} and \mathbf{y} , <i>i.e.</i> , $E((\mathbf{x} - E(\mathbf{x}))(\mathbf{y} - E(\mathbf{y}))^T)$
$\mathcal{N}(\mu, \sigma^2)$	Univariate Gaussian distribution with mean μ and variance σ^2
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\mathcal{U}_{\mathcal{X}}$	Uniform distribution over set \mathcal{X}
$X \sim \mathcal{L}$	Random variable X with probability distribution \mathcal{L}

Introduction and thesis summary

"The man who moves a mountain begins by carrying away small stones."

Confucius

1.1	Context	1
1.2	Geophysical motivations for data-driven approaches	3
1.3	Contributions	5
1.4	Dissertation outline	7
1.5	Publications and Valorization	7
1.5.1	International Conference papers	7
1.5.2	National Conference papers	8
1.5.3	Journal papers	8
1.5.4	Poster presentations	9
1.5.5	Invited talks	9
1.5.6	Other activities	9

1.1 Context

In the last few years, satellite remote sensing has produced great amounts of observation data from a wide variety of sources. A vast array of different sensor types (multi and hyper-spectral imaging, SAR imaging, microwave imaging, etc) allows for the observation of different terrestrial, oceanic and atmospheric geophysical and geobiochemical parameters (sea surface height, sea surface temperature, vegetation cover, urban development, etc) at different spatio-temporal resolutions, which amounts to a huge quantity of greatly underexploited data.

Successfully exploiting these datasets would allow us to improve the quality, power and accuracy of geophysical and geobiochemical process models and representations and tackle shortcomings associated both with currently available datasets, like the fact that no satellite sensor

is capable of producing high-resolution data both in space and time, and with current models and representations, which are either too complex to allow for efficient observation-based high-resolution analysis and reconstruction or too simple to fully capture the spatio-temporal variability of the considered processes. To fully exploit the potential within these new datasets, however, a thorough understanding of the relationship between the involved variables is needed.

In this regard, the development of new approaches to fully exploit the potential within current datasets appears clearly as a major research issue. Particularly, data mining and machine learning approaches could be used in this context for a wide variety of applications, ranging from multi-source data fusion and high-resolution reconstruction to multi-parameter analysis of geophysical or geobiochemical processes and spatio-temporal analysis and segmentation of system dynamics.

In this context, this thesis has two major objectives:

1. Explore non-negative and sparse formulations [69, 155, 200, 233] to develop new unsupervised methods to characterize variable relationships from a representative set of observations. Indeed, given the recent success of novel formulations [69, 155, 200, 233] to tackle blind source separation issues [201], they appear as an appealing alternative to tackle the problem of the observation-based characterization of operators. In the context of ocean remote sensing, the ever increasing availability of observations further motivates the exploration of data-driven approaches as an powerful alternative to classic model-driven schemes.
2. Exploit the developed methods and formulations for the resolution of inverse problems in multi-source/multi-parameter ocean remote sensing, with a special focus on the high-resolution reconstruction of sea surface dynamics fields from satellite-derived data. Particularly, we will exploit new developments in satellite altimetry producing wide-swath satellite observations in the context of the future SWOT satellite mission [59, 74, 78].

Additionally, given the vast amounts of available data, special attention must be put into the scalability of the developed solutions, thus ensuring a smooth transition towards the processing of massive datasets of high-dimensional variables.

This project is part of the Cominlabs-Lebesgue-Mer SEACS (Stochastic modEl dAta Coupled representationS for the analysis, simulation and reconstruction of upper ocean dynamics) project, that gathers researchers in applied mathematics, ICTs and oceanography and aims at proposing new stochastic models and representations for ocean dynamics. One of its main research axis, which focuses on developing data mining and machine learning approaches, provides the main setting for this thesis.

1.2 Geophysical motivations for data-driven approaches

As previously stated, the amount of multi-source ocean remote sensing and in situ data available has experienced considerable growth in the last few decades, which has been a key factor for more thorough analysis possibilities leading to a better understanding of upper ocean dynamics, ocean circulation and atmosphere-ocean interactions. In particular, a considerable amount of effort has been put into understanding the relationships and interactions between different oceanic physical quantities [4, 33, 97, 114, 160, 227].

Regarding satellite remote sensing, current (and future) satellite missions usually involve partial, irregularly-sampled observations due to orbit design, satellite track geometry, cloud occlusion, etc. Moreover, current state-of-the-art interpolation techniques fail to accurately reconstruct smaller scale upper ocean dynamics features [34, 68, 73, 74]. In this respect, limitations identified within the current generation of remote sensing satellite missions have led to the development of new, improved remote sensing technologies, such as the future SWOT altimetry mission. Importantly, SWOT will be the first mission to exploit radar interferometry to produce two-dimensional wide-swath altimetry observations, as opposed to current generation altimeters that can only produce altimetry measurements along a narrow nadir track [59, 74, 78]. Such developments further motivate current efforts focusing on both the exploration of the synergy between different ocean tracers to enhance the reconstruction of upper ocean dynamics and the development of alternative high-resolution ocean dynamics reconstruction methods for irregularly-sampled satellite observations.

On the other hand, sea surface height (SSH) is, as stated in [244], a depth-integrated quantity that contains information on the density structure of the water column, and captures mesoscale structures from 50 km up to a few hundred kilometers, which means that surface currents can be directly retrieved from SSH fields using the geostrophy balance. At the mesoscale level, such currents further stir large-scale sea surface temperature (SST) fields, which shows that upper ocean dynamics form a complex system of interactions that vary over a wide range of scales, both in space and time [244].

Additionally, the existence of a sea surface dynamical mode, referred to as the SQG mode (Surface Quasi-Geostrophy), characterized by a linear transfer function between SST and SSH has been exhibited both from theoretical and observation-driven studies [97, 99, 115, 131, 132, 142, 143, 244].

Indeed, recent work points out that upper ocean dynamics may be characterized by local SSH-SST linear relationships that correspond precisely to fractional Laplacian operators [97,

115, 132, 142, 143, 244], which can be expressed in the Fourier domain as:

$$\mathcal{F}_H(\widehat{\text{SSH}}) = -\gamma |\mathbf{k}|^{-2\alpha} \mathcal{F}_T(\widehat{\text{SST}}) \quad (1.1)$$

where \mathbf{k} is the horizontal wavelength vector, \mathcal{F}_T and \mathcal{F}_H are linear filters of SST and SSH respectively and γ is a normalization coefficient [99, 244]. Varying parameter α , which controls the effective coupling between SST and SSH, leads to different classical theoretical models [244]. For $\alpha = 1/2$, one resorts to the surface quasi-geostrophic model [99, 131, 143, 244].

Under the SQG hypothesis, two main conclusions can be drawn from Equation (1.1):

1. Surface currents, given as the orthogonal gradient of the SSH, can then be derived as spatial derivatives of a filtered version of the SST [114, 244];
2. A single linear transfer function may not be enough to capture the high complexity of upper ocean dynamics. The SQG relationship (1.1) assumes specific vertical mixing conditions, which may not be valid anywhere or at any time [85, 116, 143]. Furthermore, when upper ocean dynamics are truly driven by SQG dynamics, parameter γ may vary both in space and time [99, 143, 244], for instance in relation to the mixed layer depth.

These findings motivated the development of new statistical observation-driven methods to combine available SSH information with other remote sensing data, such as microwave SST, or *in situ* data [4, 36, 244, 268]. Whereas [85, 116] investigated Fourier-based representations of linear SSH-SST couplings, Tandeo et al. [244] relied on the fact that, for horizontal scales between 50 km and a few hundred kilometers, the upper ocean turbulence is compatible with the geostrophy turbulence theory to conclude that upper ocean dynamics may be predicted from surface density horizontal variations possibly dominated by SST fields. They exploited this idea to explore linear transfer functions between SSH and SST by introducing a multi-modal decomposition, using a latent class regression model, such that local SSH-SST relationships are described by one among a predefined number of possible linear transfer functions.

However, results reported in the literature show that a single SQG-like linear transfer function does not suffice to capture the whole mesoscale upper ocean dynamics of a particular region [244]. Moreover, an important limitation of latent class regression models is that they can only account for a finite set of linear transfer functions. By contrast, the SQG mode is characterized by a single class of linear transfer functions, namely the fractional Laplacian operator $\gamma\Delta^{1/2}$, but involves a free positive scalar parameter γ [99, 143, 244], which relates to local geophysical features (e.g., mixed layer depth). As a result, SQG-like upper ocean dynamics, characterized by continuously-varying γ parameters, may not be well represented by latent class regression models.

Such considerations support the introduction of a linear mixture model where the relationship between SST and SSH is modeled as a non-negative linear combination of K linear regressions:

$$\mathbf{y}_n = \sum_{k=1}^K \alpha_{nk} \beta_k \mathbf{x}_n + \boldsymbol{\omega}_n \quad (1.2)$$

where $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$ relate, respectively, to the SST and SSH fields, $\alpha_{nk} \in \mathbb{R}$ are mixing coefficients that model the contribution of each linear mode to the reconstruction of \mathbf{y}_n given \mathbf{x}_n , $\beta_k \in \mathbb{R}^{I \times J}$ is the regression matrix associated with mode k , and $\boldsymbol{\omega}_n \in \mathbb{R}^I$ is a noise process, usually a Gaussian noise. Model (1.2) can then be interpreted as a characterization of the relationship between \mathbf{x}_n and \mathbf{y}_n as the superposition of K linear relations, where observation-wise non-negative adjustable amplitude parameters α_{nk} are used to take into account local variations in the strength of the relationships, in a manner similar to coefficient γ in Equation (1.1).

Furthermore, Model (1.2) bears a close resemblance to constrained blind source separation formulations. This naturally suggest the extension of constrained blind source separation models to the problem of operator decomposition. Importantly, non-negative and sparse blind source separation issues, which have proven to be of particular interest to obtain more relevant and/or interpretable decompositions, constitute an important inspiration source for our work, and will then considerably influence the models and algorithms proposed in this thesis dissertation.

1.3 Contributions

In the context presented above, the need for more complex formulations that can tackle the shortcomings of current models, and the success of constrained models in blind source separation applications, led us to seek inspiration in recent developments in blind source separation to try and develop more efficient decomposition models. In this work, we aim at extending constrained blind source separation models to the problem of observation-based characterization and decomposition of operators:

$$\mathbf{y}_n = \sum_{k=1}^K \alpha_{nk} f_k(\mathbf{x}_n) + \boldsymbol{\omega}_n \quad (1.3)$$

where $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$, and we aim at identifying $\alpha_{nk} \in \mathbb{R}$, the mixing coefficients that model the contribution of each mode to the reconstruction of \mathbf{y}_n given \mathbf{x}_n , and $f_k : \mathbb{R}^J \rightarrow \mathbb{R}^I$, a linear or non-linear function associated with mode k . $\boldsymbol{\omega}_n \in \mathbb{R}^I$ is a noise process, usually considered to be additive Gaussian noise, representing model uncertainty. Following recent advancements, we will particularly focus on linear formulations considering additional non-negativity or sparsity constraints on mixing coefficients α_{nk} . Even though it falls beyond the scope of this work, we point out that one may also consider additional constraints on modal transfer functions f_k .

It should be noted that, by extending blind source separation formulations to the problem of operator decomposition, we suggest that the dynamical modes in play do not exclude each other but are rather superimposed. This is significantly different from the basis assumption made by latent class regression models, as in [244], which assume that only one mode is active at any space-time location. To our knowledge, this is the first study to exhibit, from an observation-driven analysis, the continuous superimposition of multiple dynamical modes associated with different types transfer functions.

Moreover, whereas previous works focused either on spatially homogeneous linear couplings [85, 116] or on a finite set of linear transfer functions [244], we will consider here much richer representations, which account for all mixing possibilities between a finite set of families of linear relationships.

In the context of physical oceanography, this thesis adds to an ongoing body of work, both theoretical and practical, that shows that mesoscale upper ocean dynamics may be characterized by a local linear coupling between SST and SSH [97, 115, 132, 142, 143, 244]. We further stress that multi-modal approaches are necessary, as a single linear transfer function does not suffice to capture the complex non-stationary space-time variability of mesoscale upper ocean dynamics. Previous work rather explored regional mean transfer functions [85, 116] or latent class mixture models [244]. As far as satellite remote sensing is considered, in this thesis work we further extend current efforts to either improve current interpolation methods [57] or develop alternative approaches for the high-resolution reconstruction of upper ocean dynamics from irregularly-sampled satellite data [67, 162, 257]. Specifically, we concentrate on the exploitation of wide-swath observations for the high-resolution interpolation of sea level anomaly (SLA) fields, with a view towards the future SWOT altimetry mission. Importantly, the exploration of the potential of the SWOT mission led to a fruitful scientific collaboration with the Mediterranean Institute for Advanced Studies (IMEDEA) in Mallorca, Spain, in the context of OSTST project MANATEE.

Finally, we would like to point out that even though our original motivation was to overcome the drawbacks of classical models for physical oceanography applications, the proposed model and algorithms seem to have much greater capabilities. In fact, they constitute a tool for analyzing relationships between variables, i.e., for analyzing and synthesizing variable correlation. Such a tool would allow for the analysis, segmentation and reconstruction of the relationship between any two interest variables, thus allowing for deeper insight into the nature of said relationship and leading to a more thorough understanding of the interactions between variables and the processes and phenomena behind them.

Hopefully, our findings will open new research avenues in terms of characterization of upper ocean dynamics, high-resolution in-painting and reconstruction of satellite-derived geophysical fields and multi-source high-resolution reconstruction of sea surface currents, amongst other possible applications.

1.4 Dissertation outline

This thesis dissertation is organized in three parts, as follows.

In Part I we focus on the extension of constrained blind source separation models to the problem of the data-driven characterization and decomposition of operators. Chapter 2 presents the state-of-the-art methods for constrained blind separation, introduces the problem of operator decomposition and reviews some of the most relevant applications involving operator decomposition. In Chapter 3, we present our proposed models and algorithms for model characterization, and we evaluate their performance in Chapter 4. Some relevant examples of application of the proposed models and algorithms for both reconstruction and segmentation issues in Geosciences are presented in Chapter 5.

In Part II, we focus on the high-resolution reconstruction of sea level anomaly (SLA) fields from partial, irregularly-sampled satellite observations, with a particular interest in the exploration of the potential of SWOT observations for the reconstruction of upper ocean dynamics. In Chapter 6 we introduce the context of the problem of interest, including a brief review of satellite remote sensing altimetry, as well as the state-of-the-art algorithms for the high-resolution reconstruction of sea surface dynamics and related work. In Chapter 7, we present an application of the non-negative decomposition models and associated algorithms introduced in Part I to the problem of interest. The potential of SWOT for the high-resolution reconstruction of SLA fields in the context of a fully data-driven data assimilation framework is explored in Chapter 8.

In Part III, we conclude this thesis dissertation by presenting our main conclusions and future work perspectives in Chapter 9.

1.5 Publications and Valorization

1.5.1 International Conference papers

- G. Maze, H. Mercier, R. Fablet, P. Lenca, M. Lopez-Radcenco, P. Tandeo, C. Le Goff, and Ch. Feucher. Ocean heat content structure revealed by un-supervised classification of hydrographic profiles. In *OSD 2016: Ocean Sciences Meeting*, New Orleans, United States, February 2016

- M. Lopez-Radcenco, A. Aissa-El-Bey, P. Ailliot, R. Fablet, and P. Tandeo. Non-negative decomposition of linear relationships: application to multi-source ocean remote sensing data. In *ICASSP 2016 : 41st IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4179–4183, 2016
- M. Lopez-Radcenco, A. Aissa-El-Bey, P. Ailliot, and R. Fablet. Non-negative decomposition of geophysical dynamics. In *ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, April 2017
- R. Fablet, M. Lopez-Radcenco, J. Verron, B. Mourre, B. Chapron, and A. Pascual. Learning multi-tracer convolutional models for the reconstruction of high-resolution SSH fields. In *IGARSS 2017: 2017 IEEE International Geoscience and Remote Sensing Symposium*, Fort Worth, Texas, USA, July 2017
- M. Lopez-Radcenco, R. Fablet, A. Aissa-El-Bey, and P. Ailliot. Locally-adapted convolution-based super-resolution of irregularly-sampled ocean remote sensing data. In *ICIP 2017 : IEEE International Conference on Image Processing*, Beijing, China, September 2017
- M. Lopez-Radcenco, A. Pascual, L. Gomez-Navarro, A. Aissa-El-Bey, and R. Fablet. Analog data assimilation for along-track nadir and SWOT altimetry data in the Western Mediterranean Sea. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Valencia, Spain, July 2018

1.5.2 National Conference papers

- M. Lopez-Radcenco, A. Aissa-El-Bey, P. Tandeo, and R. Fablet. Décomposition Non-négative de Dynamiques Géophysiques. In *GRETSI 2017: XXVIème colloque du GRETSI*, Juan-Les-Pins, France, September 2017
- M. Lopez-Radcenco, A. Pascual, L. Gomez-Navarro, A. Aissa-El-Bey, and R. Fablet. Assimilation par Analogues de Données Altimétriques Nadir et SWOT dans la Mer Méditerranée Occidentale. In *Conférence Française de Photogrammétrie et de Télédétection (CFPT)*, Marne-la-Vallée, France, June 2018

1.5.3 Journal papers

- G. Maze, H. Mercier, R. Fablet, P. Tandeo, M. Lopez-Radcenco, P. Lenca, Ch. Feucher, and C. Le Goff. Coherent heat patterns revealed by unsupervised classification of Argo temperature profiles in the North Atlantic Ocean. *Progress in Oceanography*, 151:275 – 292, 2017
- M. Lopez-Radcenco, R. Fablet, and A. Aissa-El-Bey. Non-negative observation-based decomposition of operators. *IEEE Transactions on Signal Processing*, Submitted

- M. Lopez-Radcenco, A. Pascual, L. Gomez-Navarro, A. Aissa-El-Bey, B. Chapron, and R. Fablet. Analog Data Assimilation of Along-track Nadir and Wide-swath SWOT Altimetry Observations in the Western Mediterranean Sea. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Forthcoming

1.5.4 Poster presentations

- M. Lopez-Radcenco, A. Aissa-El-Bey, P. Ailliot, P. Tandeo, and R. Fablet. Non-negative decomposition of sea surface dynamics from multi-source ocean remote sensing data. *A connected ocean conference (ACO), Sea Tech Week*, Brest, France, October 2016
- M. Lopez-Radcenco, A. Aissa-El-Bey, P. Ailliot, and R. Fablet. Non-negative Decomposition of Geophysical Dynamics. *GODAE OceanView Summer School. New frontiers in operational oceanography*, Port de Pollensa, Balearic Islands, Spain, October 2017
- M. Lopez-Radcenco, A. Pascual, L. Gomez-Navarro, A. Aissa-El-Bey, and R. Fablet. Analog data assimilation of along-track nadir and SWOT altimetry observations in the Western Mediterranean Sea. *European Geosciences Union (EGU) General Assembly 2018*, Vienna, Austria, April 2018

1.5.5 Invited talks

- M. Lopez-Radcenco, A. Aissa-El-Bey, P. Ailliot, and R. Fablet. Non-negative and Sparse Decomposition of Geophysical Dynamics. *Data Science & Environment Workshop (DSE2017)*, Brest, France, July 2017

1.5.6 Other activities

- Reviewer for *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Part I

Non-negative decomposition of local operators

State of the art and related work

"But arts and sciences should be like mines, where the noise of new works and further advances is heard on every side."

Sir Francis Bacon

2.1	Introduction	13
2.2	Blind source separation	14
2.2.1	Proposed formulations	15
2.2.2	Associated numerical methods	20
2.3	Operator decomposition	31
2.3.1	Orthogonality-based decompositions	32
2.3.2	Dynamic mode decomposition	37
2.3.3	Tikhonov-regularized decompositions	45
2.4	Conclusion	47

2.1 Introduction

In order to provide the methodological and theoretical framework for this thesis dissertation, this chapter briefly introduces state-of-the-art methods and models for the problem of blind source separation, including recently introduced formulations involving sparsity and non-negativity and methodological developments to efficiently characterize the proposed models. Formally, the blind source separation problem involves decomposing a given signal or image as the superposition of K different components:

$$\mathbf{y} = \sum_{k=1}^K \alpha_k \mathbf{s}_k + \boldsymbol{\omega} \quad (2.1)$$

where $\mathbf{y} \in \mathbb{R}^I$, coefficient $\alpha_k \in \mathbb{R}$ quantifies the contribution of component $\mathbf{s}_k \in \mathbb{R}^I$, which corresponds to the k -th reference signal or image and $\boldsymbol{\omega} \in \mathbb{R}^I$ is a white Gaussian noise process with covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{I \times I}$ that models the estimation residual. The principle behind blind

source separation is then to exploit observations of signal or image \mathbf{y} to identify and separate sources \mathbf{s}_k by quantifying mixing coefficients α_k , so that the final reconstruction correctly approximates approximate signal \mathbf{y} . In this respect, additional constraints may be considered to impose physically or structurally-motivated restrictions on the decomposition to enhance model identifiability and interpretability. In recent years, non-negative and sparsity constraints have proven to be of particular interest to produce relevant and interpretable decompositions. In this respect, we aim at extending these approaches to the problem of operator decomposition:

$$\mathbf{y}_n = \sum_{k=1}^K \alpha_{nk} f_k(\mathbf{x}_n) + \boldsymbol{\omega}_n \quad (2.2)$$

where $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$, and we aim at identifying $\alpha_{nk} \in \mathbb{R}$, the mixing coefficients that model the contribution of each mode to the reconstruction of \mathbf{y}_n given \mathbf{x}_n , and $f_k : \mathbb{R}^J \rightarrow \mathbb{R}^I$, a linear or non-linear function associated with mode k . $\boldsymbol{\omega}_n \in \mathbb{R}^I$ is a noise process, usually considered to be additive Gaussian noise, representing model uncertainty. This problem has been considerably less studied despite its close similarity to the problem of blind source separation.

In this chapter, we establish the theoretical basis of constrained blind source separation models and operator decomposition formulations. In Section 2.2, we review the problem of blind source separation and explore different constraints proposed in the literature, including orthogonality, sparsity and non-negativity (Section 2.2.1), as well as the associated methods and algorithms developed for model characterization (Section 2.2.2). Subsequently, we present current models used to address the problem of operator decomposition in Section 2.3, discussing the most classically used techniques and their theoretical bases, applications, and limitations. We also explore some recently introduced alternative models for operator decomposition. Building on these results, we will focus, in the following chapters, on the identified shortcomings of current models and address the extension of the blind source separation approaches presented here to the problem of observation-based characterization and decomposition of operators, with a particular focus on applications to physical oceanography.

2.2 Blind source separation

The separation and identification of contributions associated with different types of sources or processes from multiple observations (without any a priori knowledge of the processes behind the generation of such observations), formally known as blind source separation, is a general problem in signal and image processing [40, 157, 201]. It provides the methodological base for the models and algorithms developed in this thesis dissertation. Broadly speaking, blind source separation comes to decomposing a signal or image as the sum of K components:

$$\mathbf{y} = \sum_{k=1}^K \alpha_k \mathbf{s}_k + \boldsymbol{\omega} \quad (2.3)$$

where $\mathbf{y} \in \mathbb{R}^I$, coefficient $\alpha_k \in \mathbb{R}$ quantifies the contribution of component $\mathbf{s}_k \in \mathbb{R}^I$, which corresponds to the k -th reference signal or image and $\boldsymbol{\omega} \in \mathbb{R}^I$ is a white Gaussian noise process with covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{I \times I}$ that models the estimation residual. The objective is to identify the sources \mathbf{s}_k and mixing coefficients α_k that better approximate signal \mathbf{y} .

Methodologically, the most general method for blind source separation is Independent Component Analysis (ICA) [39, 40, 111–113, 158]. Under general hypothesis of non-Gaussianity and statistical independence of sources \mathbf{s}_k , ICA comes to the separation of sources \mathbf{s}_k by optimizing a given independence criterion. The choice of this criterion is problem-dependent and closely relates to the choice of constraints in the case of constrained blind source separation. Particularly, classical ICA implementations consider independence criteria such as the minimization of mutual information or the maximization of non-Gaussianity, among others. Principal Component Analysis (PCA), on the other hand, is an alternative, widely used source separation method, which we will further develop in the following sections. In particular, PCA aims at identifying a set of orthogonal sources \mathbf{s}_k that best explain the variability of the considered data, in the sense that the variance explained by each source \mathbf{s}_k is maximal. In this respect, one may consider PCA as a particular case of ICA, under the choice of an appropriate independence criterion.

Importantly, the identifiability of model (2.3) will depend on the dimension of observations I , the number of observations N and the number of sources considered K , so that the problem may become under-determined if the number of sources to detect is too big or if the number of observations considered is too small. In this respect, depending on the specific characteristics of the dataset or application considered, model (2.3) may become intractable, which justifies the introduction of additional constraints in order to improve model identifiability. In particular, significant advances in blind source separation have been reported in the last decade with the introduction of concepts such as non-negativity and sparsity [69, 155, 200, 233].

2.2.1 Proposed formulations

2.2.1.1 Orthogonal decompositions

Orthogonality is probably the most widely used constraint in most scientific disciplines, given its convenient mathematical properties, methodological simplicity and ease of interpretation. Formally, the orthogonally constrained blind source separation problem involves constraining model (2.3) as follows:

$$\mathbf{y} = \sum_{k=1}^K \alpha_k \mathbf{s}_k + \boldsymbol{\omega}, \quad \text{subject to } \langle \mathbf{s}_p, \mathbf{s}_q \rangle = 0, \forall p \neq q, p, q \in \llbracket 1, K \rrbracket \quad (2.4)$$

where $\langle \cdot, \cdot \rangle$ is the internal product in the vector space of signals \mathbf{s}_k , i.e., the vector dot product. This model then translates into a constrained least squares minimization problem:

$$\begin{cases} [\hat{\alpha}_k, \hat{\mathbf{s}}_k] = \arg \min_{\alpha_k, \mathbf{s}_k} \left\| \mathbf{y} - \sum_{k=1}^K \alpha_k \mathbf{s}_k \right\|_{\Sigma}^2 \\ \langle \mathbf{s}_p, \mathbf{s}_q \rangle = 0, & \forall p \neq q, p, q \in \llbracket 1, K \rrbracket \end{cases} \quad (2.5)$$

where $\|\cdot\|_{\Sigma}$ is the covariance-weighted norm.

In oceanography and environmental sciences in particular, orthogonality-based decomposition approaches [90, 206] have been classically used for the analysis and decomposition of ocean and atmosphere dynamics. The flagship method exploiting orthogonality for dimensionality reduction, analysis and segmentation is known as Empirical Orthogonal Functions (EOF) [77, 140, 174, 197, 198] within the oceanographic and meteorologic communities, while it is known as Principal Component Analysis (PCA) [104–106, 206] within the signal processing community.

2.2.1.2 Tikhonov regularization

Tikhonov regularization is among the most widely used methods for constraining ill-posed problems in signal and image processing. Even though similar formulations were proposed independently in a variety of scientific contexts [70, 102, 103, 209, 247], the work of Andrey Tikhonov [247–251] is usually given credit for popularizing it.

Tikhonov regularization involves a weighted ℓ_2 -norm constraint for coefficients vectors $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_K]^T$:

$$\mathbf{y} = \sum_{k=1}^K \alpha_k \mathbf{s}_k + \boldsymbol{\omega}, \quad \text{subject to } \|\mathbf{\Gamma} \boldsymbol{\alpha}\|_2^2 \leq \epsilon \quad (2.6)$$

where $\epsilon \in \mathbb{R}^+$ is a user-set tolerance parameter, and $\mathbf{\Gamma}$ is a weighting matrix, the Tikhonov matrix, determining the type of constraint one wishes to impose. In particular, the Tikhonov matrix is usually chosen as a multiple of the identity matrix $\mathbf{\Gamma} = c\mathbf{I}$, so that the constraint comes to a classic ℓ_2 -norm minimization penalizing solutions with bigger norms [196]. Alternatively, one may consider other operators for $\mathbf{\Gamma}$ to enforce alternative priors. Model (2.6) then translates into the following constrained minimization problem:

$$\begin{cases} [\hat{\alpha}_k, \hat{\mathbf{s}}_k] = \arg \min_{\alpha_k, \mathbf{s}_k} \left\| \mathbf{y} - \sum_{k=1}^K \alpha_k \mathbf{s}_k \right\|_{\Sigma}^2 \\ \|\mathbf{\Gamma} \boldsymbol{\alpha}\|_2^2 \leq \epsilon \end{cases} \quad (2.7)$$

or, equivalently:

$$\begin{cases} [\hat{\alpha}_k, \hat{\mathbf{s}}_k] = \arg \min_{\alpha_k, \mathbf{s}_k} \|\mathbf{\Gamma}\boldsymbol{\alpha}\|_2^2 \\ \left\| \mathbf{y} - \sum_{k=1}^K \alpha_k \mathbf{s}_k \right\|_{\boldsymbol{\Sigma}}^2 \leq \epsilon \end{cases} \quad (2.8)$$

where $\|\cdot\|_{\boldsymbol{\Sigma}}$ is the covariance-weighted norm.

Intuitively, Tikhonov regularization can be interpreted in a Bayesian sense, with matrix $\mathbf{\Gamma}$ relating to the a priori uncertainties of mixing coefficients α_k . Under a Gaussianity hypothesis, the Tikhonov solution comes to the Maximum A Posteriori (MAP) estimate of vector $\boldsymbol{\alpha}$, provided that $\mathbf{\Gamma}$ is chosen accordingly [260]. More generally, for homoscedastic problems with uncorrelated errors, the Tikhonov solution relates to the minimal unbiased estimator [7].

2.2.1.3 Non-negativity

In the context of the blind source separation problem formulated in Equation (2.3), non-negativity constraints have been exploited extensively to develop part-based representations with multiple applications, ranging from learning parts of faces or semantic features of text [155] to audio blind source separation of convolutive mixtures [69]. They were first introduced by drawing inspiration from situations in nature where a process or phenomena can be explained by the naturally occurring addition of multiple contributing factors, and where an inherent non-negativity exists within the problem solution (e.g. physical measurements, pixel intensities, frequency counts, etc). Under such conditions, they allow us to impose structural/physical constraints on models in order to avoid physically impossible or absurd results.

Formally, the non-negatively constrained blind source separation problem involves constraining model (2.3) as follows:

$$\mathbf{y} = \sum_{k=1}^K \alpha_k \mathbf{s}_k + \boldsymbol{\omega}, \quad \text{subject to } \alpha_k \geq 0, \forall k \in \llbracket 1, K \rrbracket \quad (2.9)$$

which translates into a constrained least squares minimization problem:

$$\begin{cases} [\hat{\alpha}_k, \hat{\mathbf{s}}_k] = \arg \min_{\alpha_k, \mathbf{s}_k} \left\| \mathbf{y} - \sum_{k=1}^K \alpha_k \mathbf{s}_k \right\|_{\boldsymbol{\Sigma}}^2 \\ \alpha_k \geq 0, \quad \forall k \in \llbracket 1, K \rrbracket \end{cases} \quad (2.10)$$

where $\|\cdot\|_{\boldsymbol{\Sigma}}$ is the covariance-weighted norm.

In practice, non-negativity can be enforced in several manners, most notably by means of optimization schemes such as active set algorithms [145], iterative update rules that maintain

non-negativity [155], and constrained alternating least squares [275] exploiting proximal operators [37], amongst others. Classical non-negatively constrained problems include Non-negative Least Squares (NNLS) [145] and Non-negative Matrix Factorization (NMF) [69, 155, 200].

Non-negative matrix factorization The problem of non-negative matrix factorization (NMF) [11, 107, 155] involves approximating a non-negative matrix \mathbf{Y} as the product of two lower rank matrices (\mathbf{W}, \mathbf{H}) : $\mathbf{Y} \approx \mathbf{WH}$, where $\mathbf{Y} \in \mathbb{R}^{I \times N}$, $\mathbf{W} \in \mathbb{R}^{I \times K}$, $\mathbf{H} \in \mathbb{R}^{K \times N}$, with $K \leq \min\{I, N\}$.

Provided that matrix \mathbf{Y} contains relevant data whose physical origin or structure can be effectively modeled as the non-negative superposition of multiple physical processes, such decompositions may allow to effectively separate physically or structurally mixed data sources. Multiple applications of NMF have been proposed, including text mining [155] and audio source separation [69], amongst others [11].

In the case of blind source separation, one may reformulate a constrained version of Equation (2.3) as follows:

$$\mathbf{Y} = \mathbf{SA} + \mathbf{W}, \quad \text{subject to: } \begin{cases} [\mathbf{S}]_{ik} \geq 0, & \forall i \in \llbracket 1, I \rrbracket, \forall k \in \llbracket 1, K \rrbracket \\ [\mathbf{A}]_{kn} \geq 0, & \forall k \in \llbracket 1, K \rrbracket, \forall n \in \llbracket 1, N \rrbracket \end{cases} \quad (2.11)$$

which induces the following constrained minimization problem:

$$\begin{cases} [\hat{\mathbf{S}}, \hat{\mathbf{A}}] = \underset{\mathbf{S}, \mathbf{A}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{SA}\|_F^2 \\ [\mathbf{S}]_{ik} \geq 0, & \forall i \in \llbracket 1, I \rrbracket, \forall k \in \llbracket 1, K \rrbracket \\ [\mathbf{A}]_{kn} \geq 0, & \forall k \in \llbracket 1, K \rrbracket, \forall n \in \llbracket 1, N \rrbracket \end{cases} \quad (2.12)$$

where the columns of matrix $\mathbf{Y} \in \mathbb{R}^{I \times N}$ contain N observations \mathbf{y}_n to be decomposed into K sources, represented as columns of matrix $\mathbf{S} \in \mathbb{R}^{I \times K}$, so that each line of matrix $\mathbf{A} \in \mathbb{R}^{K \times N}$ contain N coefficients α_{nk} relating to one of the resolved sources. A key difference with model (2.9), however, lies in the fact that in the non-negative matrix factorization formulation sources are also considered to be non-negative ($\mathbf{s}_k \geq 0$), which may be adequate (or even more suitable) for multiple physical problems.

Among relevant extensions of classic NMF formulations we may cite those involving additional constraints or hypotheses, such as sparsity [81, 107], low-rank approximations [81, 277] and matrix symmetry [136].

2.2.1.4 Sparsity

Sparsity constraints were introduced for dimensionality reduction. Their purpose is to allow for the development of simpler representations of high-dimensional data [246]. The main interest for this is the development of simpler models and representations that are easier to understand. Effectively, constraining solutions to be as sparse as possible will nullify all but the strongest parts or components of the solution, thus allowing for these to be given greater relative importance in the final reconstruction (provided that an adequate number of non-zero components is parametrized, of course).

Formally, sparsity constraints impose restrictions on the number of non-zero coefficients α_k :

$$\mathbf{y} = \sum_{k=1}^K \alpha_k \mathbf{s}_k + \boldsymbol{\omega}, \quad \text{subject to } \#\{\alpha_k \neq 0\} \leq M \quad (2.13)$$

In practice, this should be achieved by restricting the ℓ_0 -norm of coefficients vectors $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_K]^T$: (i.e. the number of non-zero elements of vector $\boldsymbol{\alpha}$),

$$\mathbf{y} = \sum_{k=1}^K \alpha_k \mathbf{s}_k + \boldsymbol{\omega}, \quad \text{subject to } \|\boldsymbol{\alpha}\|_0 \leq M \quad (2.14)$$

which translates into the following constrained minimization problem:

$$\begin{cases} [\hat{\alpha}_k, \hat{\mathbf{s}}_k] = \arg \min_{\alpha_k, \mathbf{s}_k} \left\| \mathbf{y} - \sum_{k=1}^K \alpha_k \mathbf{s}_k \right\|_{\boldsymbol{\Sigma}}^2 \\ \|\boldsymbol{\alpha}\|_0 \leq M \end{cases} \quad (2.15)$$

or, equivalently:

$$\begin{cases} [\hat{\alpha}_k, \hat{\mathbf{s}}_k] = \arg \min_{\alpha_k, \mathbf{s}_k} \|\boldsymbol{\alpha}\|_0 \\ \left\| \mathbf{y} - \sum_{k=1}^K \alpha_k \mathbf{s}_k \right\|_{\boldsymbol{\Sigma}}^2 \leq \epsilon \end{cases} \quad (2.16)$$

where $\|\cdot\|_{\boldsymbol{\Sigma}}$ is the covariance-weighted norm, and $\epsilon \in \mathbb{R}^+$ and $M \in \mathbb{Z}$ are user-set tolerance parameters.

However, since this minimization is an NP-hard, usually intractable problem, the constraint is generally relaxed and the ℓ_1 -norm is considered instead:

$$\begin{cases} [\hat{\alpha}_k, \hat{\mathbf{s}}_k] = \arg \min_{\alpha_k, \mathbf{s}_k} \left\| \mathbf{y} - \sum_{k=1}^K \alpha_k \mathbf{s}_k \right\|_{\boldsymbol{\Sigma}}^2 \\ \|\boldsymbol{\alpha}\|_1 \leq M \end{cases} \quad (2.17)$$

or, equivalently:

$$\begin{cases} [\hat{\alpha}_k, \hat{\mathbf{s}}_k] = \arg \min_{\alpha_k, \mathbf{s}_k} \|\boldsymbol{\alpha}\|_1 \\ \left\| \mathbf{y} - \sum_{k=1}^K \alpha_k \mathbf{s}_k \right\|_{\boldsymbol{\Sigma}}^2 \leq \epsilon \end{cases} \quad (2.18)$$

where $\|\cdot\|_{\boldsymbol{\Sigma}}$ is the covariance-weighted norm, and $\epsilon \in \mathbb{R}^+$ and $M \in \mathbb{Z}$ are user-set tolerance parameters.

Several sparse representation algorithms exist to compute (or approximate) the solution of this minimization problem, including (but not restricted to) KSVD [3] and proximal splitting methods [37]. Other relevant sparse representation approaches include dictionary learning schemes exploiting joint sparsity and non-negativity constraints, such as K-WEB [13].

2.2.2 Associated numerical methods

2.2.2.1 Methods for orthogonality formulations

Principal Component Analysis To provide some historical context and background, orthogonal decomposition techniques have their origin in social sciences, going back to the works of Pearson [206]. The seminal work by Hotelling [104–106] formally introduced Principal Component Analysis (PCA), probably the most widespread orthogonal decomposition technique. In the geosciences domain, EOFs were introduced in the 1940s [77, 140, 174, 197, 198]. PCA/EOFs has also been used in several other domains. To cite a few, it is known as the Hotelling transform [104, 104, 106] in multivariate quality control and also as proper orthogonal decomposition (POD) [178, 179] in mechanical engineering. Moreover, it also relates closely to techniques such as the discrete Karhunen–Loève transform (KLT) [164] in signal processing, and factor analysis [95]. Even though PCA is mainly an exploratory analysis technique, it can be used for both dimensionality reduction [92] and filtering [197] in a straightforward manner.

Generally speaking, PCA aims at decomposing a set of observations of possibly correlated variables as a linear combination of an ensemble of uncorrelated variables (or principal components). To achieve this, principal components are chosen to ensure that each one of them accounts for the largest possible variability while being orthogonal to each other. In this respect, PCA closely relates to ICA, and one may think of PCA as a particular implementation of ICA involving an independence criterion that enforces independence up to the second-order moments, thus involving the maximization of the variance explained by each component under the assumption that components are orthogonal [39]. Conversely, ICA may be viewed as an extension of PCA, with the particularity that PCA imposes independence up to the second-order moments only (given the maximization of variance of orthogonal components), whereas ICA may impose

independence up to higher orders, depending on the considered criterion [39].

Formally, given a dataset of observations $\mathbf{y}_n \in \mathbb{R}^I, n = 1, \dots, N$ (which we can assume to be centered without loss of generality), PCA will resort to a decomposition of the covariance structure underlying the data in order to discover the directions of maximal variance. It is important to note that, for geophysical applications, observations \mathbf{y}_n may contain values of a geophysical field at I distinct locations s_i at time t_n , in which case the covariance matrix elements $[\mathbf{C}]_{ij}$ contains the covariances between full time series at location pairs (s_i, s_j) . Alternatively, observations \mathbf{y}_n may reflect a geophysical field value at I distinct times t_i at a fixed location s_n ; covariance matrix elements $[\mathbf{C}]_{ij}$ will then be the covariances between full spatial maps of \mathbf{y} at different times (t_i, t_j) .

More specifically, given the data matrix $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]^T \in \mathbb{R}^{N \times I}$, PCA first computes the empirical covariance matrix as:

$$\mathbf{C} = \frac{1}{N} \mathbf{Y}^T \mathbf{Y} \quad (2.19)$$

PCA aims at finding the orthogonal basis of vectors $\mathbf{u}_k, k = 1, \dots, K$ (the principal components) that better explain the variability of the dataset, in the sense that each vector will have the maximum possible variance. This yields the following optimization problem:

$$\begin{cases} \mathbf{u}_k = \arg \max_{\mathbf{u}} \mathbf{u}^T \mathbf{C} \mathbf{u} \\ \mathbf{u}^T \mathbf{u} = 1 \end{cases} \quad (2.20)$$

which can be easily identified as an eigendecomposition problem

$$\mathbf{C} \mathbf{u} = \lambda^2 \mathbf{u} \quad (2.21)$$

The vector basis we are looking for is then the ensemble of the K eigenvectors associated with the K largest eigenvalues λ_k^2 of covariance matrix \mathbf{C} .

$$\lambda_k^2 = \mathbf{u}_k^T \mathbf{C} \mathbf{u}_k = \frac{1}{N} \|\mathbf{Y} \mathbf{u}_k\|^2 \quad (2.22)$$

which is proportional to the variance accounted for by the k -th principal component \mathbf{v}_k , i.e the variance of the projection of observations onto each one of the found directions \mathbf{u}_k : $\mathbf{v}_k = \mathbf{Y} \mathbf{u}_k$.

In practice, computing the covariance matrix \mathbf{C} can be avoided by resorting to the SVD decomposition of matrix \mathbf{Y}

$$\mathbf{Y} = \mathbf{V} \mathbf{\Lambda} \mathbf{U}^T \quad (2.23)$$

where $\mathbf{V} \in \mathbb{R}^{N \times R}$ and $\mathbf{U} \in \mathbb{R}^{I \times R}$ are unitary matrices (i.e., $\mathbf{V}^T \mathbf{V} = \mathbf{U}^T \mathbf{U} = \mathbf{I}_R$), R is the rank of data matrix \mathbf{Y} and the diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{R \times R}$ contains the singular values $\lambda \in \mathbb{R}^+$ of \mathbf{Y} .

One may then express covariance matrix \mathbf{S} as

$$\mathbf{C} = \frac{1}{N} \mathbf{U} \mathbf{\Lambda} \mathbf{\Lambda}^T \mathbf{V}^T \mathbf{V} \mathbf{U}^T = \frac{1}{N} \mathbf{U} \mathbf{\Lambda} \mathbf{\Lambda}^T \mathbf{U}^T = \frac{1}{N} \mathbf{U} \mathbf{\Lambda} \mathbf{\Lambda}^T \mathbf{U}^T = \frac{1}{N} \mathbf{U} \mathbf{\Lambda}^2 \mathbf{U}^T \quad (2.24)$$

Then, columns of matrix \mathbf{U} correspond to principal components and the term $1/N$ can be integrated into matrix $\mathbf{\Lambda}$ to yield eigenvalues λ_k^2 .

We may then sort the eigenvalues in decreasing order and sort the columns of \mathbf{U} accordingly, so that the order of the eigenvalues and components corresponds to the level of explained variability.

Going back to Equation (2.23), one may reconstruct the whole data matrix using the estimated principal components and eigenvalues:

$$\mathbf{Y} = \mathbf{V} \mathbf{\Lambda} \mathbf{U}^T = \sum_{k=1}^R \lambda_k \mathbf{v}_k \mathbf{u}_k^T \quad (2.25)$$

where $\mathbf{v}_k = [\mathbf{V}]_{:k}$ is the k -th column of matrix \mathbf{V} . It is then easy to see that matrix $\mathbf{V} \mathbf{\Lambda}$ (vectors $\lambda_k \mathbf{v}_k$) contains the coefficients that allow the reconstruction of the data from the principal components. This quantity is then referred to as the PCA scores or expansion coefficients. In this regard, PCA coefficients $\lambda_k \mathbf{v}_k$ characterize the decomposition of observations \mathbf{y}_n as a linear combination of R principal components \mathbf{u}_k .

It is possible, alternatively, to reconstruct observation \mathbf{y}_n independently by computing:

$$\mathbf{y}_n = [\mathbf{V}]_{n:} \mathbf{\Lambda} \mathbf{U}^T = \sum_{k=1}^R \lambda_k [\mathbf{V}]_{nk} \mathbf{u}_k^T \quad (2.26)$$

where $[\mathbf{V}]_{n:}$ is the n -th line of matrix \mathbf{V} and $[\mathbf{V}]_{nk}$ is the k -th element of $[\mathbf{V}]_{n:}$.

Additionally, we may also use PCA as a dimensionality reduction technique by keeping only the M first principal components and exploiting the smaller-dimensional representation yielded by PCA scores $\lambda_k \mathbf{v}_k$. Indeed, PCA will yield an orthogonal base of principal components \mathbf{u}_k on which one may project the original data, thus obtaining a new M -dimensional representation in the space of PCA scores $\lambda_k \mathbf{v}_k$. For high values of I , choosing M appropriately may help tackle high dimensional problems. Usually, M is chosen so as to ensure that a certain proportion of the total variability of the dataset, given by $var_{explained}(M) \in [0, 1]$, is kept. Moreover,

$var_{explained}(M)$ can be easily computed with the aid of eigenvalues λ_k^2 .

$$var_{explained}(M) = \frac{\sum_{p=1}^M \lambda_p^2}{\sum_{q=1}^R \lambda_q^2} \quad (2.27)$$

Similarly, it is also possible to reconstruct the original data from a reduced number of principal components using Equations (2.25) and (2.26) and obtain a smoothed version of the original observations, thus performing a PCA-based filtering.

Extensions of PCA Several extensions to classical PCA have been developed to tackle some of its identified shortcomings. Even though the orthogonality constraint gives PCA rather useful properties, it also imposes the non-correlation of identified principal components, whereas real physical phenomena tend to be correlated/non-orthogonal. This means that PCA modes tend to be difficult to interpret, as physically coherent modes/dynamics may be spread across or mixed within the identified principal components. Some of the extensions to tackle this shortcoming include rotated EOFs [95, 117, 118, 126, 130, 221], which involves rotating part of the identified principal components, thus breaking orthogonality, in such a way so as to minimize a given simplicity criterion (cost function) that penalizes model complexity, thus inducing a more interpretable decomposition. In the same spirit, Simplified Component Technique-Lasso EOFs (SCoTLASS EOFs) [91, 121, 122] tries to simultaneously impose orthogonality and model simplicity by adding a Lasso-type sparsity-promoting ℓ_1 -norm constraint to identified principal components. However, this is a non-convex optimization problem whose resolution involves advanced numerical methods that rely on the integration of an appropriately defined ordinary differential equation (ODE), with a high associated computational cost. Nonetheless, some reformulations have been proposed to achieve simpler solutions for SCoTLASS [42, 265]. A related work dealing with sparse variants of PCA involves different techniques, including exploiting random effects models for principal components and/or incorporating shrinkage of eigenvalues [154]. Finally, a number of authors have focused on sparse formulations of PCA for high-dimensional cases where the number of observations is smaller than the dimension of observations ($N \ll I$) by modeling irrelevant variables as uncorrelated noise [15, 119, 159].

A second shortcoming of PCA relates to its sensitivity to outliers within the analyzed dataset. Early attempts to deal with this limitation gave birth to the first variants of what is now referred to as Robust PCA, and focused on exploiting robust estimation of covariance and correlation matrices in the context of PCA [88, 108–110, 226]. The ever increasing availability of information and the advent of big data has recently sparked a renewed interest in Robust PCA techniques in the context of machine learning and data mining [45, 120]. A particularly interesting approach [267] formulates Robust PCA as a decomposition of data matrix $\mathbf{Y} = \mathbf{L} + \mathbf{S}$

as the sum of a low rank matrix \mathbf{L} and a sparse matrix \mathbf{S} by minimizing a two-term cost function:

$$\left[\hat{\mathbf{L}}, \hat{\mathbf{S}}\right] = \underset{\mathbf{L}, \mathbf{S}}{\operatorname{argmin}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \quad (2.28)$$

where $\|\mathbf{L}\|_*$ is the nuclear norm of \mathbf{L} , given by the sum of its singular values and λ is a parameter that controls the relative importance trade-off between \mathbf{L} and \mathbf{S} . The main idea behind this formulation lies in the fact that, for many applications, low-rank representations are associated with general, mean solutions or patterns, such as a faces in facial recognition or background images in video surveillance, while sparse representations (also referred to as the "noise" components) relate to variations or disturbances to these mean states, e.g. variations in faces due to illumination changes or moving objects in video sequences [120]. This problem, also known as principal component pursuit, has been the object of several studies. While some authors focus on more theoretical aspects, such as model identification, parameter selection, algorithmic computational complexity and extensions to cases considering missing data [30], others propose more complex formulations, particularly for the structure of the "noise" component matrix \mathbf{S} [18, 276].

Another shortcoming of classic PCA in the domain of geophysics lies on its inability to maximize variance on both the temporal and spatial domains simultaneously [90]. Indeed, depending on how the data matrix \mathbf{Y} is built and structured, only either spatial or temporal covariance is considered. Extended EOF [22, 23, 71, 214, 263] is an approach that deals with this limitation by considering augmented observations that include both spatial and temporal information. Considering spatio-temporal information at decomposition allows this approach to not only discover static spatial or temporal modes, but also to put forward and extract propagating structures and periodic signals buried within the original data. An alternative approach to obtain similar insight, known as Complex EOF [21, 94, 137, 215, 235], relies on the decomposition of complex fields built from both the original considered field and spatially or temporally lagged (displaced) versions of the original field. Such representation will then intrinsically contain spatio-temporal information that will be discoverable via a PCA decomposition performed over the complex valued field.

Limitations of PCA PCA(EOF) and similar approaches rely on a strong assumption of mode orthogonality and on the hypothesis that an orthogonal decomposition yields the best representation, which may not always be the case. As clearly demonstrated in [189], it should be noted that there are no guarantees that an orthogonal decomposition will yield individual dynamical modes or modes that relate to individual kinematic degrees of freedom. Moreover, EOF derived modes will probably not be statistically independent, and will be strongly influenced by the modal non-locality needed to ensure that variance is maximized globally.

2.2.2.2 Methods for Tikhonov regularization

The greatest advantage of Tikhonov regularization is that it allows for the direct derivation of an analytic expression for parameter estimators, given that the ℓ_2 -norm exploited is a continuous and derivable function for all real valued vectors $\boldsymbol{\alpha} \in \mathbb{R}^K$. Importantly, this means that no indirect methods need to be developed to compute the Tikhonov-regularized solution for the general blind source separation problem (2.6). Indeed, in the case of blind source separation, one may reformulate Equation (2.6) as follows:

$$\mathbf{Y} = \mathbf{S}\mathbf{A} + \mathbf{W}, \quad \text{subject to: } \|\mathbf{A}\|_F^2 < \epsilon \quad (2.29)$$

which induces the following constrained minimization problem:

$$\left[\hat{\mathbf{S}}, \hat{\mathbf{A}} \right] = \underset{\mathbf{S}, \mathbf{A}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{S}\mathbf{A}\|_F^2 + \|\mathbf{A}\|_F^2 \quad (2.30)$$

where the columns of matrix $\mathbf{Y} \in \mathbb{R}^{I \times N}$ contain N observations \mathbf{y}_n to be decomposed into K sources, represented as columns of matrix $\mathbf{S} \in \mathbb{R}^{I \times K}$, so that each line of matrix $\mathbf{A} \in \mathbb{R}^{K \times N}$ contain N coefficients α_{nk} relating to one of the resolved sources.

The computation of the solution for minimization problem (2.30) yields the following analytic expression:

$$\mathbf{A} = \left[\mathbf{S}^T \boldsymbol{\Sigma}^{-1} \mathbf{S} + \boldsymbol{\Gamma}^T \boldsymbol{\Gamma} \right]^{-1} \mathbf{S}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y} \quad (2.31)$$

which simplifies to

$$\mathbf{A} = \left[\mathbf{S}^T \boldsymbol{\Sigma}^{-1} \mathbf{S} + \mathbf{I}_K \right]^{-1} \mathbf{S}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y} \quad (2.32)$$

when $\boldsymbol{\Gamma} = \mathbf{I}_K$.

2.2.2.3 Methods for non-negative formulations

Active-set algorithms Active sets algorithms are based on the assumption that only some of the n inequality constraints involved in minimization (2.10) are active (i.e. the associated coefficient α_k is zero). The main idea is that constrained minimization problem (2.10) boils down to an unconstrained minimization problem on the passive set only (i.e. on coefficients α_k which are not null), assuming that the active set is known. Active set coefficients α_k are then simply set to zero to obtain the full solution of the constrained problem.

The simplest version of an active set method will then exploit an alternating least squares scheme to approximate the solution. In this respect, an initial feasible solution is found and all n constraints are considered to be in the passive set. At each iteration, the unconstrained minimization problem on the passive set is solved, variables in the active set are identified and

removed from the passive set, and the process is repeated. The algorithm stop is given by a convergence condition, usually on some measure of the optimality of the solution.

Methods for non-negative matrix factorization Numerically, algorithms for non-negative matrix factorization rely mostly on alternating least squares and multiplicative update rules (introduced in [155]). Given that minimization problem (2.12) involves a matrix product, it is non-convex if a joint minimization over both \mathbf{A} and \mathbf{S} is attempted. However, the sub-problem of optimizing \mathbf{A} (resp. \mathbf{S}) while considering \mathbf{S} (resp. \mathbf{A}) as fixed is indeed convex. This strongly motivates the application of ALS-type algorithms. At each iteration, one may then exploit gradient-splitting algorithms using proximal operators or similar optimization techniques. Alternatively, Lee and Seung introduced in [155] a set of multiplicative update rules that ensure that the cost function introduced in constrained minimization problem (2.12) decreases at each iteration (even though such update rules provide no assurance of convergence towards a global minimum).

2.2.2.4 Methods for sparse formulations

KSVD KSVD [3] is a sparse dictionary learning algorithm designed to iteratively solve problem (2.15) by alternating optimizations over sparse mixing coefficients α_k and signals \mathbf{s}_k . In this respect, KSVD closely relates to alternating least squares approaches. Specifically, the optimization of reference signals \mathbf{s}_k is performed by rewriting problem (2.15) in matricial form:

$$\mathbf{Y} = \mathbf{S}\mathbf{A} + \mathbf{W}, \quad \text{subject to: } \|\mathbf{A}_{:,n}\|_0 \leq M, \forall n \in \llbracket 1, N \rrbracket \quad (2.33)$$

This new formulation induces the following constrained minimization problem:

$$\begin{cases} \left[\hat{\mathbf{S}}, \hat{\mathbf{A}} \right] = \underset{\mathbf{S}, \mathbf{A}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{S}\mathbf{A}\|_F^2 \\ \|\mathbf{A}_{:,n}\|_0 \leq M, \forall n \in \llbracket 1, N \rrbracket \end{cases} \quad (2.34)$$

where the columns of matrix $\mathbf{Y} \in \mathbb{R}^{I \times N}$ contain N observations \mathbf{y}_n to be decomposed into K sources, represented as columns of matrix $\mathbf{S} \in \mathbb{R}^{I \times K}$, so that each line of matrix $\mathbf{A} \in \mathbb{R}^{K \times N}$ contain N coefficients α_{nk} relating to one of the resolved sources, and $\mathbf{A}_{:,n}$ denotes the n -th column of matrix \mathbf{A} .

KSVD updates the dictionary of reference signals \mathbf{S} iteratively, by sequentially updating each reference signal \mathbf{s}_k (column $[\mathbf{S}]_{:,k}$ in matrix \mathbf{S}). To this end, the cost function of minimization problem (2.34) can be rewritten as

$$\|\mathbf{Y} - \mathbf{S}\mathbf{A}\|_F^2 = \left\| \mathbf{Y} - \sum_{j=1}^K \mathbf{s}_j [\mathbf{A}]_{j:} \right\|_F^2 = \left\| \left(\mathbf{Y} - \sum_{j \neq k}^K \mathbf{s}_j [\mathbf{A}]_{j:} \right) - \mathbf{s}_k [\mathbf{A}]_{k:}^T \right\|_F^2 = \|\mathbf{E}_k - \mathbf{s}_k [\mathbf{A}]_{k:}\|_F^2 \quad (2.35)$$

so that the product $\mathbf{S}\mathbf{A}$ is decomposed as the sum of K rank-1 matrices. \mathbf{E}_k is then the error matrix when all but the k -th reference signal are considered. The objective now is to find the reference signal \mathbf{s}_k so that $\mathbf{s}_k[\mathbf{A}]_{k:}$ is the rank-1 matrix that best approximates \mathbf{E}_k , while respecting the sparsity constraint imposed on matrix \mathbf{A} and enforced in Equation (2.35) by vector $[\mathbf{A}]_{k:}$.

To ensure that sparsity is respected, KSVD identifies which observations \mathbf{y}_i use signal \mathbf{s}_k , given by non-zero elements of $[\mathbf{A}]_{k:}$. It then defines index vector $\boldsymbol{\omega}_k$ containing the indexes of the actually-exploited observations

$$\boldsymbol{\omega}_k = \{i | 1 \leq i \leq K, [\mathbf{A}]_{ki} \neq 0\} \quad (2.36)$$

and then uses this index vector to define matrix $\boldsymbol{\Omega}_k \in \mathbb{R}^{N \times |\boldsymbol{\omega}|}$ containing ones at positions $(\boldsymbol{\omega}_k(i), i)$ and zeros elsewhere.

Matrix $\boldsymbol{\Omega}_k$ is then used to constraint Equation (2.35) so that only observations that use signals \mathbf{s}_k given by $[\mathbf{A}]_{k:}$ are considered. If we compute $\mathbf{Y}_k^R = \mathbf{Y}\boldsymbol{\Omega}_k \in \mathbb{R}^{N \times |\boldsymbol{\omega}|}$ we obtain a matrix that only contains observations that use signal \mathbf{s}_k . Similarly, the product $[\mathbf{A}]_{k:}^R = [\mathbf{A}]_{k:}\boldsymbol{\Omega}_k$ will discard all zero entries of $[\mathbf{A}]_{k:}$. It can then be deduced that $\mathbf{E}_k^R = \mathbf{E}_k\boldsymbol{\Omega}_k \in \mathbb{R}^{N \times |\boldsymbol{\omega}|}$ will contain only error columns involving signal \mathbf{s}_k . If we apply this matrix product and go back to Equation (2.35), we obtain:

$$\|\mathbf{E}_k - \mathbf{s}_k[\mathbf{A}]_{k:}\|_F^2 = \|\mathbf{E}_k^R - \mathbf{s}_k[\mathbf{A}]_{k:}^R\|_F^2 \quad (2.37)$$

where sparsity compliance has been assured by the product with matrix $\boldsymbol{\Omega}_k$, and we may now jointly optimize over \mathbf{s}_k and $[\mathbf{A}]_{k:}^R$ without contradicting the imposed sparsity.

The optimization over \mathbf{s}_k and $[\mathbf{A}]_{k:}^R$ is achieved by using SVD [83] to decompose matrix $\mathbf{E}_k^R = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}$ and setting $\mathbf{s}_k = [\mathbf{U}]_{:1}$ and $[\mathbf{A}]_{k:}^R = [\boldsymbol{\Delta}]_{11}[\mathbf{V}]_{:1}$. Note that such decomposition ensures that $[\mathbf{A}]_{k:}^R$ has at most the same number of non-zero elements as $[\mathbf{A}]_{k:}$, if not less, thus ensuring sparsity is respected.

The optimization of coefficients α_k , on the other hand, is performed by means of orthogonal matching pursuit.

Orthogonal matching pursuit Matching pursuit (MP) [183] is a sparse coding algorithm that relies on sequentially finding the best sparse approximation for signal \mathbf{y} given reference signals \mathbf{s}_k . To achieve this, at each step the reference signal \mathbf{s}_k^* most correlated to \mathbf{y} is identified. An approximation of \mathbf{y} using only \mathbf{s}_k^* is then computed by estimating the associated coefficient α_k^* . This approximation is then subtracted from signal \mathbf{y} . In this way, the selection of the next most correlated signal at the next iteration is computed based on residual $\mathbf{y} - \alpha_k^*\mathbf{s}_k^*$ from the previous iteration. The process is repeated until a maximum number of M signals are selected

(or, alternatively, until the residual error is smaller than a threshold ϵ_r).

It should be noted that MP relies on the hypothesis that the dictionary of signals \mathbf{s}_k is over-complete, which means that $K > I$, where $\mathbf{y} \in \mathbb{R}^I$. Moreover, MP computes an approximate solution of minimization problem (2.15). However, it can be shown that the algorithm will converge monotonically for any signal \mathbf{y} spanned by reference signals \mathbf{s}_k [3, 43, 183, 205].

Orthogonal matching pursuit (OMP) [43, 205] is a natural extension of MP, which consists in updating all coefficients of already extracted signals \mathbf{s}_k at each iteration, rather than only updating the coefficient of the currently selected signal \mathbf{s}_k^* . This boils down to an orthogonal projection of signal \mathbf{y} onto the space spanned by the previously selected signals \mathbf{s}_k , hence the name of the extension.

2.2.2.5 Methods for multiple constraints

Alternating Least Squares For problems where multiple parameters must be optimized, the principle behind alternating least squares (ALS) [10, 58, 96, 238, 274] lies on sequentially computing the optimal value of each parameter while the remaining parameters remain fixed. In this way, the alternating least squares approach will optimize \mathbf{s}_k while considering fixed coefficients α_k and subsequently optimize coefficients α_k while considering fixed sources \mathbf{s}_k . This process is repeated until convergence.

This general scheme for multi-parameter optimization is a particular case of non-linear Gauss-Seidel optimization schemes, which implies that its convergence is not always ensured. However, local convergence of ALS-type algorithms can be ensured under appropriate conditions [10, 58], usually verified in practice. We refer the interested reader to [10, 58] for a more detailed discussion of the mathematical criteria influencing convergence for ALS algorithms.

At each step, the optimization of the considered parameter relies on an adequate minimization scheme, such as gradient descent, direct minimization, gradient-splitting methods exploiting proximal operators, etc., depending on the considered parameter and on any desired constraint.

Proximal gradient methods One of the main challenges of using non-negativity/sparsity constraints lies on the non-differentiability of the constraints that prevents the direct exploitation of classical optimization methods, such as gradient descent. In this respect, proximal gradient methods are a family of optimization methods developed to deal with possibly non-differentiable cost functions.

Formally, we consider the following multivariate minimization problem:

$$\min_{\mathbf{x}} F(\mathbf{x}) + G(\mathbf{x}) \quad (2.38)$$

where $\mathbf{x} \in \mathbb{R}^N$, $F(\mathbf{x}) : \mathbb{R}^N \rightarrow]-\inf, +\inf] \in \mathcal{C}^1$, i.e., $F(\mathbf{x})$ is differentiable and both $F(\mathbf{x})$ and $\frac{\partial F(\mathbf{x})}{\partial \mathbf{x}}$, its derivative with respect to \mathbf{x} , are continuous, and $G(\mathbf{x}) : \mathbb{R}^N \rightarrow]-\inf, +\inf]$ is a lower-semicontinuous, not necessarily differentiable function. In the context of blind source separation, $F(\mathbf{x})$ is a data fit function measuring the distance between observed data and model reconstruction, while $G(\mathbf{x})$ is the desired constraint function.

The proximal operator of function $G(\mathbf{x})$ is defined as

$$\text{prox}_G(\mathbf{u}) = \operatorname{argmin}_{\mathbf{x}} G(\mathbf{x}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_2^2 \quad (2.39)$$

so that the proximal operator yields the closest point to \mathbf{u} that minimizes function $G(\mathbf{x})$. We notice that if function $G(\mathbf{x})$ is the indicator function $\mathbb{1}_{\mathcal{S}}$ of some convex set $\mathcal{S} \subset \mathbb{R}^N$, then the associated proximal operator comes to the projection of \mathbf{u} onto \mathcal{S} . The proximal operator is, then, an extension of the notion of projection.

Proximal operators have a number of properties that make them specially suitable for iterative minimization algorithms [37], such as verifying the following inclusion:

$$\mathbf{p} = \text{prox}_G(\mathbf{u}) \iff \mathbf{u} - \mathbf{p} \in \partial G(\mathbf{p}), \quad \forall \mathbf{u} \in \mathbb{R}^N, \forall \mathbf{p} \in \mathbb{R}^N \quad (2.40)$$

where $\partial G(\mathbf{p})$ is the subdifferential of function G , a set-valued function defined as

$$\partial G : \mathbb{R}^N \rightarrow 2^{\mathbb{R}^N} : \mathbf{x} \rightarrow \left\{ \mathbf{u} \in \mathbb{R}^N \mid \forall \mathbf{y} \in \mathbb{R}^N, (\mathbf{y} - \mathbf{x})^T \mathbf{u} + G(\mathbf{x}) \leq G(\mathbf{y}) \right\} \quad (2.41)$$

which can be thought of as the set of gradients (slopes) of the affine minorants of G at \mathbf{x} . This means that if \mathbf{p} is given by the proximal operator of function $G(\mathbf{x})$ at point \mathbf{x} , then the difference vector $\mathbf{x} - \mathbf{p}$ belongs to the set of the subgradients of $G(\mathbf{x})$.

Another interesting property of proximal operators is that they are firmly non-expansive:

$$\|\text{prox}_G(\mathbf{x}) - \text{prox}_G(\mathbf{y})\|^2 + \|(\mathbf{x} - \text{prox}_G(\mathbf{x})) - (\mathbf{y} - \text{prox}_G(\mathbf{y}))\|^2 \leq \|\mathbf{x} - \mathbf{y}\|^2, \quad \forall \mathbf{x} \in \mathbb{R}^N, \forall \mathbf{y} \in \mathbb{R}^N \quad (2.42)$$

which implies that proximal operators have a non-empty set of fixed points that corresponds to the set of the minima of G [37].

From these properties, it is shown in [38] that minimization problem (2.38) has at least one solution, and that its solutions verify the following fixed point equation:

$$\mathbf{x} = \text{prox}_{\gamma G}(\mathbf{x} - \gamma \nabla F(\mathbf{x})), \forall \gamma \in \mathbb{R}^+ \quad (2.43)$$

which suggest the following iterative approach:

$$\mathbf{x}_{n+1} = \text{prox}_{\gamma G}(\mathbf{x}_n - \gamma \nabla F(\mathbf{x}_n)), \forall \gamma \in \mathbb{R}^+ \quad (2.44)$$

which comes to applying a classic gradient descent, followed by the proximal operator associated to G .

Proximal gradient methods for constrained blind source separation For the problems of non-negative and sparse blind source separation introduced in Equations (2.9) and (2.14), for example, one may design an alternating minimization approach in which two steps are iterated until convergence.

A first step comes to optimizing reference signals \mathbf{s}_k with fixed coefficients α_k .

A second step comes to the optimization of constrained coefficients α_k with fixed reference signals \mathbf{s}_k . To achieve this, coefficients are first estimated without constraints using a classic gradient descent:

$$\hat{\alpha}_k^{i+1} = \hat{\alpha}_k^i + \frac{\mathbf{s}_k^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{y} - \sum_{p=1}^K \hat{\alpha}_p^i \mathbf{s}_p \right)}{\mathbf{s}_k^T \boldsymbol{\Sigma}^{-1} \mathbf{s}_k}, \forall k \in \llbracket 1, K \rrbracket \quad (2.45)$$

The desired constraint $G(\alpha_k)$ is then imposed by means of the appropriate proximal operator

$$\hat{\alpha}_k^{i+1} = \text{prox}_G(\hat{\alpha}_k^{i+1}), \forall k \in \llbracket 1, K \rrbracket \quad (2.46)$$

In this respect, non-negativity can be enforced by setting

$$G(\alpha_k) = \mathbb{1}_{\mathbb{R}^+}(\alpha_k) = \begin{cases} 0, & \alpha_k \geq 0 \\ +\text{inf}, & \alpha_k < 0 \end{cases}, \forall k \in \llbracket 1, K \rrbracket \quad (2.47)$$

whose associated proximal operator is

$$\text{prox}_{\mathbb{1}_{\mathbb{R}^+}}(\alpha_k) = \begin{cases} \alpha_k, & \alpha_k \geq 0 \\ 0, & \alpha_k < 0 \end{cases}, \forall k \in \llbracket 1, K \rrbracket \quad (2.48)$$

Sparsity, on the other hand, involves a weighted ℓ_1 -norm penalty:

$$G(\boldsymbol{\alpha}) = \gamma \|\boldsymbol{\alpha}\|_1 \quad (2.49)$$

where weight γ controls the compromise between reconstruction similarity and the ℓ_1 -norm regularization.

The associated proximal operator for this function is the soft-thresholding operator:

$$\left[\text{prox}_{\gamma \|\boldsymbol{\alpha}\|_1}(\boldsymbol{\alpha}) \right]_k = \text{soft}_{\gamma}(\alpha_k) = \begin{cases} \alpha_k - \gamma, & \alpha_k > \gamma \\ 0, & |\alpha_k| \leq \gamma \\ \alpha_k + \gamma, & \alpha_k < -\gamma \end{cases} \quad (2.50)$$

Alternatively, sparsity constraints may also involve a weighted ℓ_0 -norm penalty:

$$G(\boldsymbol{\alpha}) = \gamma \|\boldsymbol{\alpha}\|_0 \quad (2.51)$$

where weight γ controls the compromise between reconstruction similarity and the ℓ_0 -norm regularization.

The associated proximal operator for this function is the hard-thresholding operator:

$$\left[\text{prox}_{\gamma \|\boldsymbol{\alpha}\|_0}(\boldsymbol{\alpha}) \right]_k = \text{hard}_{\sqrt{2\gamma}}(\alpha_k) = \begin{cases} \alpha_k, & |\alpha_k| > \sqrt{2\gamma} \\ 0, & |\alpha_k| \leq \sqrt{2\gamma} \end{cases} \quad (2.52)$$

2.3 Operator decomposition

Even though extensive literature exists on the problem of blind source separation, the very similar problem of observation-based characterization and decomposition of operators (relationships between variables of interest) has not been studied as extensively. As stated in our introductory chapter, we aim at extending constrained blind source separation models to the problem of the observation-based characterization and decomposition of operators:

$$\mathbf{y}_n = \sum_{k=1}^K \alpha_{nk} f_k(\mathbf{x}_n) + \boldsymbol{\omega}_n \quad (2.53)$$

where $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$, and we aim at identifying $\alpha_{nk} \in \mathbb{R}$, the mixing coefficients that model the contribution of each mode to the reconstruction of \mathbf{y}_n given \mathbf{x}_n , and $f_k : \mathbb{R}^J \rightarrow \mathbb{R}^I$, a linear or non-linear function associated with mode k . $\boldsymbol{\omega}_n \in \mathbb{R}^I$ is a noise process, usually considered to be additive Gaussian noise, representing model uncertainty. Similarly to the problem of blind source separation, one may also consider additional constraints on either mixing coefficients α_{nk}

or modal transfer functions f_k to better constraint the problem to the specific application or dataset considered.

In this section, we provide insights into some of the most relevant applications where operator decomposition has been studied, in domains ranging from fluid dynamics and dynamical system analysis to image super-resolution.

2.3.1 Orthogonality-based decompositions

2.3.1.1 Joint EOFs

To study the relationship between multiple variables, one may wish to apply PCA simultaneously to more than one field, in order to discover relevant information about the coupling between them. For vectorial fields, such as sea surface winds or geostrophic velocities, one may again resort to the PCA decomposition of a complex-valued field built from the quantities of interest. In more general settings, however, either an Extended EOFs analysis of the relationship between the two fields of interest or, alternatively, an SVD decomposition of their cross-covariance matrix, may be more viable alternatives.

The most straightforward way of performing a Joint EOF analysis involves building extended observations combining the considered fields, in a manner similar to Extended EOFs, so that both field observations $\mathbf{x}_n \in \mathbb{R}^J$ and $\mathbf{y}_n \in \mathbb{R}^I$ are represented in the extended observation $\mathbf{z}_n \in \mathbb{R}^{I+J}$: $\mathbf{z}_n = [\mathbf{x}_n, \mathbf{y}_n]$. In this regard, the PCA decomposition of extended observations \mathbf{z}_n will produce jointly estimated PCA components $\mathbf{u}_k = [\mathbf{u}_k^x, \mathbf{u}_k^y]$. Principal components \mathbf{u}_k^x and \mathbf{u}_k^y then relate, respectively, to the decomposition of \mathbf{x}_n and \mathbf{y}_n , whereas PCA scores $\lambda_k \mathbf{v}_k$ are shared between the two representations, as they have been jointly estimated so that each extended PCA component \mathbf{u}_k individually explains a maximal fraction of the variability of the extended dataset \mathbf{Z} built from extended observations \mathbf{z}_n . Interestingly, if the fields under study are uncorrelated, this approach will produce the same components as the independent PCA decomposition of each separate field, sorted according to their variance.

The main advantage of such an approach is that the covariance matrix of extended vector \mathbf{z}_n will consider both the covariance structures of \mathbf{x}_n and \mathbf{y}_n independently, as well as their cross covariance structure. However, this approach may fail to account for statistical differences in the fields and requires the relative weighting and normalization of the extended observations to take differences in magnitude, considered units and sampling into account, and to avoid any bias towards the field with the highest variability.

Alternatively, one may directly apply an SVD decomposition to the cross-covariance matrix of observations \mathbf{x}_n and \mathbf{y}_n

$$\mathbf{C} = \frac{1}{N} \mathbf{X}^T \mathbf{Y} = \mathbf{V} \mathbf{\Lambda} \mathbf{U}^T = \sum_{r=1}^R \lambda_r \mathbf{v}_r \mathbf{u}_r^T \quad (2.54)$$

where data matrices $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T \in \mathbb{R}^{N \times J}$ and $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]^T \in \mathbb{R}^{N \times I}$ are build from observations from both fields, $\mathbf{V} \in \mathbb{R}^{J \times R}$ and $\mathbf{U} \in \mathbb{R}^{I \times R}$ are unitary matrices (i.e., $\mathbf{V}^T \mathbf{V} = \mathbf{U}^T \mathbf{U} = \mathbf{I}_R$), R is the rank of cross-covariance matrix \mathbf{C} and the diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{R \times R}$ contains the singular values $\lambda \in \mathbb{R}^+$ of \mathbf{C} .

Similarly to classical PCA, resorting to the SVD decomposition of matrix \mathbf{C} will yield an orthogonal decomposition of the cross-covariance into the R highest covariability modes of the two fields. It can indeed be shown that the SVD decomposition will lead to pairs of principal components for fields \mathbf{x}_n and \mathbf{y}_n given, respectively, by singular vectors \mathbf{u}_k and \mathbf{v}_k , so that each pair of principal components explains as much as possible of the mean-squared cross-covariance between the considered fields [19]. This is due to the fact that SVD will select singular vectors pairs so that the covariance between the projections of datasets \mathbf{X} and \mathbf{Y} onto, respectively, the left and right singular vectors \mathbf{v}_k and \mathbf{u}_k is maximal for each pair. In this respect, the maximal covariance explained by each pair of singular vectors is given by their corresponding singular value, and the PCA expansion coefficients $\mathbf{L}_x \in \mathbb{R}^{J \times R}$ and $\mathbf{L}_y \in \mathbb{R}^{I \times R}$ are then obtained by projecting matrices \mathbf{X} and \mathbf{Y} onto orthonormal basis \mathbf{V} and \mathbf{U} , respectively [19]:

$$\mathbf{L}_x = \mathbf{XV} \quad (2.55)$$

$$\mathbf{L}_y = \mathbf{YU} \quad (2.56)$$

One can then sort the singular values in decreasing order and sort the columns of matrices \mathbf{U} , \mathbf{V} , $\mathbf{\Lambda}$, \mathbf{L}_x and \mathbf{L}_y accordingly, so that the order of the singular values and joint PCA components corresponds to the level of covariability explained.

In this context, assuming a linear operator $\mathbf{A} \in \mathbb{R}^{J \times I}$ can accurately approximate the relationship between datasets \mathbf{X} and \mathbf{Y} , then an expression for operator \mathbf{A} can be derived:

$$\mathbf{Y} = \mathbf{XA} = \mathbf{X} \left[N \left(\mathbf{V} \mathbf{D}_x^{-1} \mathbf{\Lambda} \mathbf{U}^T \right) \right] \quad (2.57)$$

where $\mathbf{D}_x = \mathbf{L}_x^T \mathbf{L}_x = \mathbf{V}^T \mathbf{X}^T \mathbf{XV}$ is the covariance matrix of \mathbf{L}_x . Importantly, matrix \mathbf{D}_x is diagonal by construction, as \mathbf{L}_x corresponds to the projection of observations \mathbf{X} onto the space spanned by left singular vectors \mathbf{v}_k (which, by definition, diagonalizes the covariance matrix of \mathbf{x}_n).

Alternatively, an equivalent expression for observations \mathbf{x}_n and \mathbf{y}_n can be obtained:

$$\mathbf{y}_n = \mathbf{A}^T \mathbf{x}_n = N \left(\mathbf{U} \mathbf{\Lambda} \mathbf{D}_x^{-1} \mathbf{V}^T \right) \mathbf{x}_n \quad (2.58)$$

2.3.1.2 Canonical correlation analysis

Canonical correlation analysis (CCA) [19,93,106,236,261] is a related approach for the analysis of the covariability between two variables $\mathbf{x}_n \in \mathbb{R}^J$ and $\mathbf{y}_n \in \mathbb{R}^I$. The principle behind this method lies in identifying maximally correlated linear transformations of the two variables of interest [19,93,106,236,261]. In this regard, CCA will iteratively look for the linear transformations $\mathbf{v} \in \mathbb{R}^J$ and $\mathbf{u} \in \mathbb{R}^I$ that, when applied to observations \mathbf{x}_n and \mathbf{y}_n , produce a pair of transformed variables with maximal correlation:

$$[\mathbf{v}, \mathbf{u}] = \underset{\mathbf{v}, \mathbf{u}}{\operatorname{argmax}} \operatorname{corr}(\mathbf{v}^T \mathbf{x}_n, \mathbf{u}^T \mathbf{y}_n) \quad (2.59)$$

subject to the constraint that each new pair of identified canonical variables $(\mathbf{v}^T \mathbf{x}_n, \mathbf{u}^T \mathbf{y}_n)$ must be uncorrelated (orthogonal) to the canonical variables previously identified. An additional constraint imposing that the variance of canonical variables is unitary, i.e., $\operatorname{var}(\mathbf{v}^T \mathbf{x}_n) = \operatorname{var}(\mathbf{u}^T \mathbf{y}_n) = 1$, is needed to ensure the unicity of the decomposition. Intuitively, we may view CCA as a projection of variables \mathbf{x}_n and \mathbf{y}_n onto lower-dimensional subspaces where they are maximally correlated [19,93,106,236,261].

Methodologically, this can be expressed as the following constrained minimization problem

$$\begin{aligned} [\mathbf{V}, \mathbf{U}] &= \underset{\mathbf{V}, \mathbf{U}}{\operatorname{argmax}} \operatorname{tr} \left(\mathbf{V}^T \mathbf{X}^T \mathbf{Y} \mathbf{U} \right) \\ \text{Subject to } &\begin{cases} \mathbf{V}^T \mathbf{X}^T \mathbf{X} \mathbf{V} = \mathbf{I}_L \\ \mathbf{U}^T \mathbf{Y}^T \mathbf{Y} \mathbf{U} = \mathbf{I}_L \end{cases} \end{aligned} \quad (2.60)$$

where data matrices $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T \in \mathbb{R}^{N \times J}$ and $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]^T \in \mathbb{R}^{N \times I}$ are build from observations from both fields, $\mathbf{V} \in \mathbb{R}^{J \times L}$ and $\mathbf{U} \in \mathbb{R}^{I \times L}$ are matrices whose k -th columns correspond to the k -th linear transformation pair $(\mathbf{v}_k, \mathbf{u}_k)$ yielding the k -th canonical variable pair $(\mathbf{v}_k^T \mathbf{x}_n, \mathbf{u}_k^T \mathbf{y}_n)$, and L is the number of pairs of canonical variables considered.

In this regard, it can be proven [19] that the solution to this minimization problem is given by

$$\mathbf{V} = \mathbf{C}_x^{-1/2} \hat{\mathbf{V}} \quad (2.61)$$

$$\mathbf{U} = \mathbf{C}_y^{-1/2} \hat{\mathbf{U}} \quad (2.62)$$

where $\mathbf{C}_x = \frac{1}{N} \mathbf{X}^T \mathbf{X}$ and $\mathbf{C}_y = \frac{1}{N} \mathbf{Y}^T \mathbf{Y}$ are, respectively, the covariance matrices of \mathbf{x}_n and \mathbf{y}_n , and $\hat{\mathbf{V}}$ and $\hat{\mathbf{U}}$ are given by the first L components of the SVD decomposition:

$$\mathbf{C} = \mathbf{C}_x^{-1/2} \mathbf{C}_{xy} \mathbf{C}_y^{-1/2} = \hat{\mathbf{V}} \mathbf{\Lambda} \hat{\mathbf{U}}^T \quad (2.63)$$

where $\mathbf{C}_{xy} = \frac{1}{N} \mathbf{X}^T \mathbf{Y}$ is the cross-covariance matrix between observations \mathbf{x}_n and \mathbf{y}_n , $\hat{\mathbf{V}} \in \mathbb{R}^{J \times R}$ and $\hat{\mathbf{U}} \in \mathbb{R}^{I \times R}$ are unitary matrices (i.e., $\hat{\mathbf{V}}^T \hat{\mathbf{V}} = \hat{\mathbf{U}}^T \hat{\mathbf{U}} = \mathbf{I}_R$), R is the rank of matrix \mathbf{C} and the diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{R \times R}$ contains the singular values $\lambda \in \mathbb{R}^+$ of \mathbf{C} [19].

Compared to the SVD approach presented before, we may regard CCA as an SVD decomposition of the cross-covariance matrix between two datasets previously normalized by their respective covariance matrices. In this regard, as in the SVD decomposition presented in the previous section, a linear operator relating variables \mathbf{x}_n and \mathbf{y}_n may be approximated from the CCA decomposition. Importantly, when compared to the SVD decomposition of the cross-covariance matrix presented before, each pair of identified canonical variables will be more strongly correlated between them, but explain a lower fraction of the covariability between the two fields of interest [19].

A far as practical applications are concerned, CCA is useful to gain insight into how different components of variables \mathbf{x}_n and \mathbf{y}_n relate to each other, the amount of variability shared among different components within the two datasets and to model relationships between the two variables [254]. Indeed, CCA has been successfully used in a variety of applications, including classification, multivariate regression and dimensionality reduction [93, 236, 259].

2.3.1.3 Singular value decomposition of linear operators

Let's consider a linear system given by

$$\mathbf{y}_n = \mathbf{A} \mathbf{x}_n \quad (2.64)$$

where $\mathbf{y}_n \in \mathbb{R}^I$, $\mathbf{x}_n \in \mathbb{R}^J$ are observations of signals or images, while $\mathbf{A} \in \mathbb{R}^{I \times J}$ is a linear operator governing the relationship between the observed variables.

For the sake of simplicity, we assume that either operator \mathbf{A} is known or, alternatively, that we have a dataset of N observation pairs $\{\mathbf{x}, \mathbf{y}\}_n$ large enough that operator \mathbf{A} may be accurately estimated in a mean-squared-error sense: $\mathbf{A} = \mathbf{Y} \mathbf{X}^T [\mathbf{X} \mathbf{X}^T]^{-1}$, where columns of matrices $\mathbf{X} \in \mathbb{R}^{J \times N}$ and $\mathbf{Y} \in \mathbb{R}^{I \times N}$ correspond, respectively, to observations \mathbf{x}_n and \mathbf{y}_n .

One may then use an SVD decomposition to decompose linear operator \mathbf{A} :

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{U}^T = \sum_{r=1}^R \lambda_r \mathbf{v}_r \mathbf{u}_r^T = \sum_{r=1}^R \lambda_r \mathbf{A}_r \quad (2.65)$$

where $\mathbf{V} \in \mathbb{R}^{I \times R}$ and $\mathbf{U} \in \mathbb{R}^{J \times R}$ are orthogonal matrices (i.e., $\mathbf{V}^T \mathbf{V} = \mathbf{U}^T \mathbf{U} = \mathbf{I}_R$), R is the rank of operator \mathbf{A} and the diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{R \times R}$ contains the singular values $\lambda_r \in \mathbb{R}^+$ of matrix \mathbf{A} . λ_r is then the r -th singular value with associated left and right singular vectors \mathbf{v}_r and \mathbf{u}_r . As such, matrix \mathbf{A} is then decomposed as the sum of R rank-1 matrices $\mathbf{A}_r = \mathbf{v}_r \mathbf{u}_r^T$.

Moreover, columns of matrices \mathbf{U} , \mathbf{V} and $\mathbf{\Lambda}$ can be sorted according to their respective singular values λ_r in decreasing order. Singular values λ_r characterize then the decomposition of operator \mathbf{A} as a combination of rank-1 matrices $\mathbf{A}_r = \mathbf{v}_r \mathbf{u}_r^T$. In this respect, matrix \mathbf{A} can be seen as the weighted, ordered sum of R separable matrices \mathbf{A}_r , in the sense that each matrix \mathbf{A}_r can be written as the product of two vectors \mathbf{v}_r and \mathbf{u}_r . This may be useful for applications where separable models naturally arise, such as the separation of image filters into horizontal and vertical components.

Finally, it can be proven [55, 234] that reconstructing \mathbf{A} using only the largest K singular values and associated singular vectors yields the best rank- K approximation of operator \mathbf{A} , which may be useful for applications where low-rank operator or matrix approximations are required.

2.3.1.4 Eigendecomposition of linear dynamical operators

More specifically, we may consider a linear dynamical system given by

$$\mathbf{x}_n = \mathbf{A} \mathbf{x}_{n-1} \quad (2.66)$$

where $\mathbf{x}_n \in \mathbb{R}^J$ is the system state vector at time index n , while $\mathbf{A} \in \mathbb{R}^{J \times J}$ is a dynamical operator governing the system evolution.

As in the previous section we assume, for the sake of simplicity, that either operator \mathbf{A} is known or, alternatively, that we have a dataset of N observations $\{\mathbf{x}\}_n$ large enough that operator \mathbf{A} may be accurately estimated in a mean-squared-error sense: $\mathbf{A} = \mathbf{X}_n \mathbf{X}_{n-1}^T \left[\mathbf{X}_{n-1} \mathbf{X}_{n-1}^T \right]^{-1}$, where columns of matrices $\mathbf{X}_n \in \mathbb{R}^{J \times (N-1)}$ and $\mathbf{X}_{n-1} \in \mathbb{R}^{J \times (N-1)}$ correspond, respectively, to observations \mathbf{x}_n and \mathbf{x}_{n-1} .

Provided linear operator \mathbf{A} is diagonalizable, we may perform a spectral decomposition of the dynamics governing the evolution of state variable \mathbf{x}_n by means of an eigendecomposition

of operator \mathbf{A} , which yields:

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1} = \sum_{j=1}^J \lambda_j [\mathbf{U}]_{:j} [\mathbf{U}^{-1}]_j \quad (2.67)$$

where $\mathbf{U} \in \mathbb{C}^{J \times J}$ is a square matrix whose columns $[\mathbf{U}]_{:j} = \mathbf{u}_j$ are the eigenvectors of matrix \mathbf{A} with associated eigenvalues $\lambda_j \in \mathbb{C}$, which are contained in the diagonal matrix $\mathbf{\Lambda} \in \mathbb{C}^{J \times J}$, and eigenvalues and eigenvectors verify:

$$\mathbf{A}\mathbf{u}_j = \lambda_j \mathbf{u}_j \quad (2.68)$$

It should be noted that matrix \mathbf{U} is not necessarily unitary ($\mathbf{U}^{-1} = \mathbf{U}^H$) for the general case of a diagonalizable operator \mathbf{A} , so that the eigendecomposition of an arbitrary linear operator will not necessarily yield an orthogonal decomposition. For the decomposition to be orthogonal, \mathbf{A} must be a normal matrix ($\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A}$), so that the eigendecomposition yields a unitary matrix \mathbf{U} :

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H = \sum_{j=1}^J \lambda_j \mathbf{u}_j \mathbf{u}_j^H \quad (2.69)$$

In particular, if the considered dynamical system is self-adjoint, i.e., if operator matrix \mathbf{A} is symmetrical ($\mathbf{A} = \mathbf{A}^T$), then all its eigenvalues $\lambda_j \in \mathbb{R}$ are real and \mathbf{U} is a real orthogonal matrix ($\mathbf{U}^{-1} = \mathbf{U}^T$):

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T = \sum_{j=1}^J \lambda_j \mathbf{u}_j \mathbf{u}_j^T \quad (2.70)$$

Matrix \mathbf{A} is then decomposed as the sum of J matrices $\mathbf{U}_j = \mathbf{u}_j \mathbf{u}_j^T$, so that eigenvalues λ_j then characterize the decomposition of dynamical system matrix \mathbf{A} as a combination of J matrices $\mathbf{U}_j = \mathbf{u}_j \mathbf{u}_j^T$.

2.3.1.5 Limitations of orthogonality-based decompositions

Similarly to classical PCA (EOF) and similar approaches, the orthogonality-based variability analysis and operator decomposition formulations presented here rely on a strong assumption of mode orthogonality, and on the hypothesis that an orthogonal decomposition yields the best representation, which may not always be the case. Once again we refer to the fact that there are no guarantees that an orthogonal decomposition will yield individual dynamical modes or modes that relate to individual kinematic degrees of freedom [189]. Moreover, for PCA-based approaches, derived modes will probably not be statistically independent, and will be strongly influenced by the modal non-locality needed to ensure that variance is maximized globally [189].

2.3.2 Dynamic mode decomposition

Dynamical system analysis and decomposition is a key issue in multiple and varied disciplines such as fluid dynamics [229], medical care [17], neuroscience [24], epidemiology [216], robotic control [9], image and video processing [139], finance [184] and power system analysis [237], to cite a

few. Among a wide variety of approaches, Dynamic Mode Decomposition (DMD) [138,228,264] is one of the most widely used techniques to decompose operators representing dynamical systems into physically relevant modes. Importantly, as illustrated by the previously cited examples, applications of DMD span a vast number of different scientific disciplines. In particular, DMD relies on SVD decomposition [83] and exploits a finite-dimensional approximation of the Koopman operator [134,138], an infinite-dimensional linear representation of a finite-dimensional non-linear dynamical system.

2.3.2.1 The Koopman Operator

Let's consider a classical discrete-time, possibly non-linear dynamical system given by

$$\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n) \quad (2.71)$$

where $\mathbf{x}_n \in \mathbb{R}^J$ is a state vector, lying on a smooth manifold $\mathcal{M} \subset \mathbb{R}^J$, that accurately represents system dynamics.

The Koopman operator [26,134,138,187,188] is defined as an infinite-dimensional linear operator \mathcal{K} acting on the space of **all** possible observable functions of \mathbf{x}_n , $g(\mathbf{x}) \in \mathcal{G}$, $g(\mathbf{x}) : \mathcal{M} \rightarrow \mathbb{C}$, that verifies:

$$\mathcal{K}g(\mathbf{x}_n) = g(\mathbf{f}(\mathbf{x}_n)) = g(\mathbf{x}_{n+1}) \quad (2.72)$$

so that a linearization of the non-linear dynamics is obtained by a projection of the dynamical system from the space of system states \mathbf{x} to the space of observable functions $g(\mathbf{x})$, at the cost of considering now an infinite-dimensional problem. Indeed, given that the definition of the Koopman operator considers the infinite-dimensional space of **all** possible observable functions $g(\mathbf{x})$, the existence of a linear operator that verifies Equation (2.72) can be guaranteed.

Moreover, from Equation (2.72) it is possible to see that this linear representation of a dynamical system allows one to evolve the system (represented in the space spanned by observables $g(\mathbf{x})$) simply by a multiplication with the Koopman operator.

As previously stated, the Koopman operator is defined on the space of **all** scalar measuring functions $g(\mathbf{x})$. In this sense, moving from a classical state space representation to a Koopman representation in the space of observables $g(\mathbf{x})$ involves a shift from nonlinear but finite-dimensional dynamics to linear but infinite-dimensional dynamics. On one hand, linear representations offer multiple advantages and increased simplicity. On the other hand, the infinite dimensionality of the new representation tends to be difficult to handle. In practice, however, a sufficiently large but finite approximation of the infinite dimensional Koopman operator usually provides adequate results.

If one considers the eigendecomposition of the Koopman operator

$$\mathcal{K}\phi_k = \lambda_k\phi_k \quad (2.73)$$

where $\lambda_k \in \mathbb{C}$ are the Koopman eigenvalues, then $\phi_k : \mathcal{M} \rightarrow \mathbb{C}$, the Koopman eigenfunctions, define a set of coordinates for the observables g on which we can represent and advance the system dynamics by means of a linear operator. Indeed, following spectral decomposition principles, any vector of observables \mathbf{g} can be written as:

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) \\ g_2(\mathbf{x}) \\ \vdots \\ g_I(\mathbf{x}) \end{bmatrix} = \sum_{k=1}^{+\infty} \phi_k(\mathbf{x}) \begin{bmatrix} v_k^1 \\ v_k^2 \\ \vdots \\ v_k^I \end{bmatrix} = \sum_{k=1}^{+\infty} \phi_k(\mathbf{x}) \mathbf{v}_k \quad (2.74)$$

where v_k^i , the i -th element of vector \mathbf{v}_k , is the k -th coefficient of observable function g_i when projected onto the basis of eigenfunctions $\{\phi_k\}$, so that each function g_i is expressed as a weighted sum of an infinite number of eigenfunctions ϕ_k . \mathbf{v}_k is thus the k -th Koopman mode of observable vector \mathbf{g} , associated with the k -th Koopman eigenfunction ϕ_k .

The main advantage of this representation is that it allows for the evolution of the system in the space of observables simply by means of a product involving the Koopman eigenvalues:

$$\mathbf{g}(\mathbf{x}_{n+1}) = \mathcal{K}\mathbf{g}(\mathbf{x}_n) = \mathcal{K} \sum_{k=1}^{+\infty} \phi_k(\mathbf{x}_n) \mathbf{v}_k = \sum_{k=1}^{+\infty} \mathcal{K}\phi_k(\mathbf{x}_n) \mathbf{v}_k = \sum_{k=1}^{+\infty} \lambda_k \phi_k(\mathbf{x}_n) \mathbf{v}_k \quad (2.75)$$

Applying the same principle recursively from $g(\mathbf{x}_0)$, we obtain:

$$\mathbf{g}(\mathbf{x}_1) = \mathcal{K}\mathbf{g}(\mathbf{x}_0) = \mathcal{K} \sum_{k=1}^{+\infty} \phi_k(\mathbf{x}_0) \mathbf{v}_k = \sum_{k=1}^{+\infty} \mathcal{K}\phi_k(\mathbf{x}_0) \mathbf{v}_k = \sum_{k=1}^{+\infty} \lambda_k \phi_k(\mathbf{x}_0) \mathbf{v}_k \quad (2.76)$$

$$\mathbf{g}(\mathbf{x}_2) = \mathcal{K}\mathbf{g}(\mathbf{x}_1) = \mathcal{K} \sum_{k=1}^{+\infty} \lambda_k \phi_k(\mathbf{x}_0) \mathbf{v}_k = \sum_{k=1}^{+\infty} \lambda_k \mathcal{K}\phi_k(\mathbf{x}_0) \mathbf{v}_k = \sum_{k=1}^{+\infty} \lambda_k^2 \phi_k(\mathbf{x}_0) \mathbf{v}_k \quad (2.77)$$

\vdots

$$\mathbf{g}(\mathbf{x}_{n+1}) = \mathcal{K}\mathbf{g}(\mathbf{x}_n) = \mathcal{K} \sum_{k=1}^{+\infty} \lambda_k^n \phi_k(\mathbf{x}_0) \mathbf{v}_k = \sum_{k=1}^{+\infty} \lambda_k^n \mathcal{K}\phi_k(\mathbf{x}_0) \mathbf{v}_k = \sum_{k=1}^{+\infty} \lambda_k^{n+1} \phi_k(\mathbf{x}_0) \mathbf{v}_k \quad (2.78)$$

so that the final Koopman decomposition is given by:

$$\mathbf{g}(\mathbf{x}_{n+1}) = \sum_{k=1}^{+\infty} \lambda_k^{n+1} \phi_k(\mathbf{x}_0) \mathbf{v}_k = \sum_{k=1}^{+\infty} \lambda_k^{n+1} \boldsymbol{\omega}_k \quad (2.79)$$

where we have absorbed scalar values $\phi_k(\mathbf{x}_0)$ into Koopman mode $\boldsymbol{\omega}_k = \phi_k(\mathbf{x}_0) \mathbf{v}_k$.

In this regard, the Koopman decomposition describes observable functions $\mathbf{g}(\mathbf{x})$ as a sum of fixed modes \mathbf{v}_k whose time variability and dynamical evolution is encoded by Koopman eigenvalues λ_k , which provides additional information relating to spatial structures and their temporal evolution [135, 239]. Importantly, for any given Koopman dynamical mode, its frequency and rate of decay are given, respectively, by the phase and module of the associated Koopman eigenvalue λ_k [239]. In general, Koopman modes (and thus Koopman eigenfunctions) then relate to the spatial nature of the problem of interest, while temporal information is encoded in Koopman eigenvalues. In this respect, Koopman analysis then usually focuses on studying the spatial distribution of Koopman modes (eigenfunctions) and their temporal evolution, which is governed by Koopman eigenvalues λ_k [135, 239, 241].

Interestingly, in the original formulation of his theory [134], Koopman worked with measure preserving Hamiltonian flows, so that the Koopman operator is then unitary and eigenfunctions ϕ_k are orthonormal in the space of observables $g(\mathbf{x})$ [138], which yields:

$$\mathbf{v}_k = \begin{bmatrix} v_k^1 \\ v_k^2 \\ \vdots \\ v_k^I \end{bmatrix} = \begin{bmatrix} \langle \phi_k, g_1 \rangle \\ \langle \phi_k, g_2 \rangle \\ \vdots \\ \langle \phi_k, g_I \rangle \end{bmatrix} \quad (2.80)$$

The orthogonality of Koopman modes (eigenfunctions), however, can only be ensured under certain conditions, as we will discuss later.

In this context, all non-linear dynamics are captured by the Koopman operator, while the Koopman eigenfunctions produce a basis of non-linear functions in which the dynamical system in question becomes linear. To deal with this infinite-dimensional representation, one can truncate the infinite series of Koopman eigenfunctions to an appropriate finite number of terms by considering only the first K eigenfunctions. In this respect, this truncation also involves restricting observables g to a Koopman-invariant subspace $\mathbf{G} \subset \mathcal{G}$ [138, 228, 239, 241], so that the restriction of \mathcal{K} to this subspace induces a finite-dimensional linear operator \mathbf{K} that allows us to evolve the system dynamics within the restricted subspace.

$$\mathbf{g}(\mathbf{x}_{n+1}) = \mathbf{K}\mathbf{g}(\mathbf{x}_n) = \mathbf{K} \sum_{k=1}^K \phi_k(\mathbf{x}_n) \mathbf{v}_k = \sum_{k=1}^K \mathbf{K} \phi_k(\mathbf{x}_n) \mathbf{v}_k = \sum_{k=1}^K \lambda_k \phi_k(\mathbf{x}_n) \mathbf{v}_k = \dots = \sum_{k=1}^K \lambda_k^{n+1} \omega_k \quad (2.81)$$

Finding the Koopman eigenfunctions and associated Koopman eigenvalues of a dynamical system, however, is one of the most challenging parts of Koopman spectral analysis. One of the most popular approaches developed to solve this problem, dynamic mode decomposition (DMD), relies on exploiting a set of measures of system dynamics observables to estimate the Koopman modes and the associated Koopman eigenvalues.

2.3.2.2 The dynamic mode decomposition algorithm

Dynamic mode decomposition (DMD) [138, 223, 228, 256, 264] is a data-driven approach that allows for the approximation of Koopman modes and eigenvalues directly from a set of measures of observables of a dynamical system¹.

Let's consider a set of observables g_1, \dots, g_I , with $g_i \in \mathcal{M}, \forall i \in \llbracket 1, I \rrbracket$ that span, approximately, a Koopman invariant subspace and denote then the observable vector $\mathbf{g} = [g_1, \dots, g_I]^T$.

Assuming we have an ensemble of measures $\mathbf{g}(\mathbf{x}_1), \dots, \mathbf{g}(\mathbf{x}_n)$ of our dynamical system, we build the following data matrices:

$$\mathbf{Y}_{n-1} = [\mathbf{g}(\mathbf{x}_1), \dots, \mathbf{g}(\mathbf{x}_{n-1})] \quad (2.82)$$

$$\mathbf{Y}_n = [\mathbf{g}(\mathbf{x}_2), \dots, \mathbf{g}(\mathbf{x}_n)] \quad (2.83)$$

Intuitively, one may infer that $\mathbf{Y}_n = \mathbf{A}_Y \mathbf{Y}_{n-1}$, with \mathbf{A}_Y a finite dimensional approximation of the complete infinite-dimensional Koopman operator \mathcal{K} . This finite dimensional approximation corresponds to the restriction of \mathcal{K} to the subspace of observables g_1, \dots, g_I , and relates to the finite-dimensional nature of our ensemble of observations.

It is straightforward, then, that \mathbf{A}_Y may be computed as

$$\mathbf{A}_Y = \mathbf{Y}_n \mathbf{Y}_{n-1}^\dagger \quad (2.84)$$

where $(\cdot)^\dagger$ denotes the Moore-Penrose pseudo-inverse.

Formally, the Koopman mode decomposition theorem [138, 228] states that, given an eigenfunction $\phi_k \in \text{span}\{g_i\}$ of \mathcal{K} with associated eigenvalue λ_k that verifies:

$$\phi_k(\mathbf{x}) = \sum_{i=1}^I z_i g_i(\mathbf{x}) = \mathbf{z}^T \mathbf{g}(\mathbf{x}) \quad (2.85)$$

for some $\mathbf{z} = [z_1, \dots, z_I]^T \in \mathbb{C}^I$, if $\mathbf{z} \in R(\mathbf{Y}_{n-1})$ where $R(\mathbf{Y}_{n-1})$ is the range of \mathbf{Y}_{n-1} , i.e., the subspace spanned by the columns of \mathbf{Y}_{n-1} , then \mathbf{z}^H is a left eigenvector of \mathbf{A}_Y with eigenvalue λ_k : $\mathbf{z}^H \mathbf{A}_Y = \lambda_k \mathbf{z}^H$.

¹Depending on the author and the scientific community involved, the exact definition of DMD may not be identical. Certain authors follow the historical developments of the methodology and refer to DMD when the decomposition is applied directly to state vector \mathbf{x} , i.e., when no observable functions are considered, and use the term Extended DMD (EDMD) to refer to the technique considering a set of observables $g(\mathbf{x})$. More recent developments use the name DMD to refer to the more general approach involving observable functions. One may indeed consider the identity observable function $g(\mathbf{x}) = \mathbf{x}$, so that both approaches are related. In this chapter, we follow the latter convention and refer to the more general approach as DMD.

Assuming, without loss of generality, that the left and right eigenvectors of \mathbf{A}_Y , \mathbf{z}_k^H and \mathbf{v}_k , are normalized, i.e., $\mathbf{z}_k^H \mathbf{v}_{k'} = \delta_{kk'}$, $\forall k, k' \in \llbracket 1, K \rrbracket$, it can be shown that observable functions $\mathbf{g}(\mathbf{x})$ verify [138, 228, 239]:

$$\mathbf{g}(\mathbf{x}) = \sum_{k=1}^K \phi_k(\mathbf{x}) \mathbf{v}_k \quad (2.86)$$

which implies that, if the set of available observables is sufficiently large to verify $\phi_k \in \text{span}\{g_i\}$ and rich enough that $\mathbf{z} \in R(\mathbf{Y}_{n-1})$, then the Koopman decomposition of observable functions \mathbf{g} can be accurately approximated from the DMD eigenvalues and associated modes of matrix \mathbf{A}_Y [138]. We may then compute the eigendecomposition of matrix \mathbf{A}_Y : $\mathbf{A}_Y \boldsymbol{\omega}_k = \lambda_k \boldsymbol{\omega}_k$ in order to obtain the Koopman modes and eigenvalues.

Alternatively, rather than directly computing \mathbf{A}_Y and its eigendecomposition directly, DMD resorts to the SVD decomposition of \mathbf{Y}_{n-1} :

$$\mathbf{Y}_{n-1} = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (2.87)$$

so that

$$\mathbf{Y}_n = \mathbf{A}_Y \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (2.88)$$

$$\mathbf{U}^T \mathbf{Y}_n \mathbf{V} \mathbf{S}^{-1} = \mathbf{U}^T \mathbf{A}_Y \mathbf{U} = \mathbf{A}'_Y \quad (2.89)$$

One can then compute matrix \mathbf{A}'_Y and its eigendecomposition $\mathbf{A}'_Y \boldsymbol{\omega}'_k = \lambda_k \boldsymbol{\omega}'_k$ to obtain its eigenvalues λ_k and eigenvectors $\boldsymbol{\omega}'_k$. Given that \mathbf{A}_Y and \mathbf{A}'_Y are related by a coordinate change associated with the orthonormal basis given by columns of matrix \mathbf{U} , then it is straightforward that \mathbf{A}_Y and \mathbf{A}'_Y share the same eigenvalues λ_k . Additionally, Koopman modes $\boldsymbol{\omega}_k$ can be computed as $\boldsymbol{\omega}_k = \mathbf{U} \boldsymbol{\omega}'_k$, which implies:

$$\mathbf{A}_Y \boldsymbol{\omega}_k = \mathbf{A}_Y \mathbf{U} \boldsymbol{\omega}'_k = \lambda_k \boldsymbol{\omega}_k \quad (2.90)$$

and given that

$$\mathbf{Y}_n = \mathbf{A}_Y \mathbf{U} \mathbf{S} \mathbf{V}^T \implies \mathbf{A}_Y = \mathbf{Y}_n \mathbf{V} \mathbf{S}^{-1} \mathbf{U}^T \quad (2.91)$$

we obtain

$$\boldsymbol{\omega}_k = \frac{1}{\lambda_k} \mathbf{Y}_n \mathbf{V} \mathbf{S}^{-1} \boldsymbol{\omega}'_k \quad (2.92)$$

which gives us another expression for the computation of Koopman modes.

If one wishes to compute the Koopman eigenfunctions $\phi(\mathbf{x}_n)$ rather than the Koopman modes, these can be approximated from the left eigenvectors \mathbf{z}_k^H of matrix \mathbf{A}_Y as $\phi_k(\mathbf{x}_n) = \mathbf{z}_k^H \mathbf{g}(\mathbf{x}_n)$.

2.3.2.3 Orthogonality of the Koopman decomposition

Generally speaking, DMD results in a data-driven non-orthogonal decomposition onto a set of modes that attempts to capture the behaviour of a dynamical system from a set of observations [135]. In this respect, exploiting non-orthogonal decompositions may enhance DMD's capabilities to better represent the behaviour of the dynamical system under study, despite the increased computational cost associated with non-orthogonal decompositions [135]. DMD can be regarded, then, as a data-driven alternative to more classical model decomposition techniques, such as PCA or POD [135]. In this regard, to yield a decomposition onto orthogonal Koopman modes, DMD would require the dynamical system in question to be normal (or self-adjoint), so that the linear operator \mathbf{K} , the restriction of infinite-dimensional Koopman operator \mathcal{K} to the space of observable functions \mathbf{G} , is a normal (or symmetrical) matrix with a unitary (or orthogonal) eigendecomposition. Under this conditions, the DMD decomposition closely relates to the eigendecomposition of normal and self-adjoint operators presented previously in Section 2.3.1.4.

2.3.2.4 Limitations of dynamic mode decomposition

In general, DMD and related approaches rely on strong assumptions, such as the time invariance of the Koopman operator, the finite-dimensional approximation of the infinite-dimensional Koopman operator and the correct selection and availability of observable functions that span a Koopman-invariant subspace, all of which are necessary to ensure the feasibility of DMD. For a wide range of applications, however, these hypotheses may not always be valid, which may prevent DMD (or related techniques) to be applied or to provide significant or interpretable results.

2.3.2.5 Beyond DMD

Given the non-trivial limitations of DMD, several extensions have been proposed for the DMD algorithm. To tackle the problem of the Koopman-invariance of the subspace spanned by observable functions $g(\mathbf{x})$, multiple approaches have been developed, including the use of non-linear basis functions [264], reproducing kernels [129] or delay coordinates [87, 271]. Particularly, some studies propose to learn Koopman invariant subspaces, which comes down to learning observable functions $g(\mathbf{x})$ directly from data. While [240] exploits neural networks to this end, another study [163] uses dictionary-learning techniques to approximate observable functions $g(\mathbf{x})$. Interestingly, a combination of both techniques is proposed in [264], where DMD is coupled with a dictionary learned using neural networks.

Other studies focus on characterizing DMD in the context of noisy observables $g(\mathbf{x})$ [52, 202] and developing suitable approaches to treat observation noise explicitly [44, 100] or indirectly [239, 242]. In particular, [239] formulates the DMD algorithm in the context of a Bayesian

framework, thus allowing for the characterization of observation noise. Moreover, given that the Bayesian formulation allows for the automatic tuning of hyper-parameters, the authors propose to estimate the number of considered modes by exploiting a sparsity-promoting prior. They further benefit from the probabilistic formulation by using Monte Carlo techniques to compute parameter and modeling uncertainties. Alternatively, [242] deals not only with observation noise but also with process noise in the context of stochastic dynamical systems by exploiting an orthogonal projection onto past system dynamics.

Non-negative and sparse DMD To improve the interpretability and representation power of DMD, a few studies have investigated extensions involving the low-rank estimation of system dynamics [35, 48, 98, 123, 269] and the exploitation of additional constraints, such as joint sparsity and non-negativity [241]. Closely related to our subject of interest, in [241], most notably, Takeishi et al. propose a constrained version of DMD, so that dynamic modes are estimated under joint sparsity and non-negativity constraints. To achieve this, the authors propose to reformulate DMD as a block-multiconvex optimization problem. By exploiting a polar-coordinate expression for Koopman eigenmodes $\mathbf{m}_k = \phi_k(\mathbf{x}_0)\mathbf{v}_k$ and eigenvalues λ_k :

$$\mathbf{m}_k^i = q_k^i e^{j\theta_k^i}, \quad q_k^i, \theta_k^i \in \mathbb{R} \quad (2.93)$$

$$\lambda_k = r_k e^{j\psi_k}, \quad r_k, \psi_k \in \mathbb{R} \quad (2.94)$$

where j is the imaginary unit and $\mathbf{m}_k^i = [\mathbf{m}_k]_i$ is the i -th element of eigenmode \mathbf{m}_k . We can now define matrices $[\mathbf{Q}]_{ki} = q_k^i$, $[\Theta]_{ki} = e^{(j\theta_k^i)}$, $[\mathbf{R}]_{kt} = r_k^{t-1}$, $[\Psi]_{kt} = e^{(j(t-1)\psi_k)}$, with $k = 1, \dots, K$, $i = 1, \dots, I$ and $t = 1, \dots, n+1$, where K is the maximum number of considered modes.

With this new notation, the DMD problem can be rewritten as the following minimization problem:

$$[\hat{q}_k^i, \hat{\theta}_k^i, \hat{r}_k, \hat{\psi}_k] = \arg \min_{q_k^i, \theta_k^i, r_k, \psi_k} \|\mathbf{Y} - (\mathbf{Q} \odot \Theta)(\mathbf{R} \odot \Psi)\|_F^2 \quad (2.95)$$

where $\mathbf{Y} = [\mathbf{g}(\mathbf{x}_0), \dots, \mathbf{g}(\mathbf{x}_n)]$ is the full data matrix and \odot denotes the elementwise product.

Once DMD is reformulated as a minimization problem, the authors use a block coordinate descent, which comes down to a particular implementation of gradient descent for block-multiconvex functions [270]. The purpose of this reformulation is to allow for the introduction of additional constraints on variable q , θ , r and ψ for problem (2.95):

$$[\hat{q}_k^i, \hat{\theta}_k^i, \hat{r}_k, \hat{\psi}_k] = \arg \min_{q_k^i, \theta_k^i, r_k, \psi_k} \|\mathbf{Y} - (\mathbf{Q} \odot \Theta)(\mathbf{R} \odot \Psi)\|_F^2 + c(q, \theta, r, \psi) \quad (2.96)$$

The authors propose, in particular, to impose non-negativity and an ℓ_1 -norm regularization on dynamic modes, i.e., fixing $q \geq 0$, $\theta = 0$ and adding a regularization term proportional to $|q|$:

$$c(q, \theta, r, \psi) = c(q) = \gamma \sum_{k=1}^K \sum_{i=1}^I |q_k^i| + \mathbb{1}_{q \geq 0}(q) \quad (2.97)$$

with γ being a regularization parameter and $\mathbb{1}_{q \geq 0}(q)$ being an indicator function whose value is 0 if $q \geq 0$ and $+\infty$ otherwise. After the introduction of the additional constraints, the authors rely on a proximal Newton-type method coupled with a block coordinate descent to solve optimization problem (2.96).

As far as applications are concerned, Takeishi et al. justify the choice of a jointly non-negative and sparse formulation as a means to better adapt the DMD algorithm to the problem of separating background and foreground in video sequences, as both constraints appear as naturally relevant for this particular task.

Otherwise, operator decomposition formulations similar to DMD, but involving alternative constraints, have been introduced in other scientific domains. Most notably, Tikhonov-regularized formulations have been successfully exploited within the image processing community to tackle the problem of image super-resolution, as detailed in the following section.

2.3.3 Tikhonov-regularized decompositions

Within the image processing community, recent super-resolution methods exploit linear patch-based operators and their decomposition to tackle memory constrained super-resolution applications [2].

Image super-resolution or upscaling refers to the problem of creating a coherent and visually pleasant high-resolution version of a low-resolution image, without loss of sharpness or detail. It closely relates to the problem of deblurring, with the exception that in super-resolution the low-resolution initial image is considered to be sharp at its original size [253]. Since super-resolution involves the interpolation of new values at sub-pixel locations, it is inherently an ill-posed problem, as each pixel in the low-resolution image must be mapped into multiple pixels [2, 41, 253].

Among the variety of approaches proposed to tackle this problem, exemplar-based methods use a patch-based (or feature-based) approach to learn a function between low-resolution and high-resolution spaces [252], based on hypothesis that patches lie in local manifolds that are in correspondence, i.e., similar patches in the low-resolution space are mapped to similar patches in the high-resolution space [253]. The simplest method exploiting this idea consists in approximating each low-resolution patch as a linear combination of its most similar patches and using the estimated coefficients to obtain a reconstruction from the corresponding patches in the high-

resolution space. This technique relates to approaches such as neighbour embedding [101], local linear embedding [222] and non-local means [25, 207]. Here, and in more elaborate approaches that learn a non-linear function to go from a low-resolution space to a high-resolution space, we require a training dataset of low-resolution/high-resolution patch pairs.

Usually, a single global function to be applied to every patch will most probably not be able to adequately super-resolve an image, given the complex, non-linear nature of the problem [41]. In this respect, some methods propose to locally approximate the non-linear function relating the low-resolution and high-resolution spaces by multiple local linear functions. In particular, Jointly Optimized Regressors (JOR) [41] separates patches pairs in the training dataset into clusters using K-means [180] and computes a single linear regression (called a regressor) per cluster. The problem of super-resolution for a new image boils down to assigning each patch of the image to the cluster that yields the lower super-resolution error and applying the associated regressor.

Building on this principle, Anchored Neighborhood Regression (ANR) [252] performs the initial clustering by computing a low-resolution sparse dictionary using KSVD [3]. Each element of the dictionary will then have an associated regressor, so that super-resolution comes to finding the dictionary element most correlated to the low-resolution patch to super-resolve. In ANR, regressors for each dictionary element are computed from its K nearest elements in the dictionary. Adjusted Anchored Neighborhood Regression (A+) [253] is an extension of ANR that improves performance by looking for the K nearest neighbours for the computation of each regressor in the whole dataset of low-resolution training patches (rather than only in the low-resolution dictionary elements).

More specifically, from a set of low-resolution patches $\mathcal{X} \in \mathbb{R}^{d_L}$ and corresponding high-resolution patches $\mathcal{Y} \in \mathbb{R}^{d_H}$, A+ uses KSVD to compute low-resolution dictionary $\mathbf{D}^L = [\mathbf{d}_1^L, \dots, \mathbf{d}_N^L] \in \mathbb{R}^{d_L \times N}$ and the corresponding high-resolution dictionary $\mathbf{D}^H = [\mathbf{d}_1^H, \dots, \mathbf{d}_N^H] \in \mathbb{R}^{d_H \times N}$. For each element of the low resolution dictionary \mathbf{d}_i^L , A+ builds a regressor \mathbf{W}_i using ridge regression

$$\mathbf{W}_i = \underset{\mathbf{W}}{\operatorname{argmin}} \|\mathbf{W}\mathbf{X}_i - \mathbf{Y}_i\|_F^2 + \lambda \|\mathbf{W}\|_F^2 \quad (2.98)$$

where matrix $\mathbf{X}_i = [\mathbf{x}_1^i, \dots, \mathbf{x}_K^i] \in \mathbb{R}^{d_L \times K}$ contains the K low-resolution patches closest to \mathbf{d}_i^L in the training dataset and matrix $\mathbf{Y}_i = [\mathbf{y}_1^i, \dots, \mathbf{y}_K^i] \in \mathbb{R}^{d_H \times K}$ contains the corresponding high-resolution patches. To super-resolve a new image, for each patch \mathbf{x} , we find its closest dictionary element \mathbf{d}_x^L and apply the associated regressor \mathbf{W}_x :

$$\hat{\mathbf{y}} = \mathbf{W}_x \mathbf{x} \quad (2.99)$$

One of the main disadvantages of the above mentioned approaches lies on the high memory requirements needed to stock the estimated regressors. Indeed, standard A+ implementations consider $N = 65,536$ regressors. Given that the memory requirement of A+ is $\mathcal{O}(Nd_Hd_L + Nd_L)$, this means, for example, a memory footprint of 1.2GB for a x4 upscaling ($d_L = 31, d_H = 144$, single precision) [2]. To deal with this limitation, Regressor Basis Learning (RB) [2] builds on A+ and decomposes regressors onto a set of R basis regressors $\{\tilde{\mathbf{W}}^j\}_{j=1,\dots,R}$ and representation coefficients $\{\alpha_i^j\}_{j=1,\dots,R}$:

$$\mathbf{W}_i = \sum_{j=1}^R \alpha_i^j \tilde{\mathbf{W}}^j \quad (2.100)$$

so that only basis regressors and coefficients are stored, which diminishes the memory complexity at the expense of some extra computational complexity, given that at run-time one must now reconstruct regressor \mathbf{W}_i before applying it.

One of the main limitations of this approach, however, is that it relies on Tikhonov regularization (also known as ridge regression), i.e., an ℓ_2 -norm penalization, for the decomposition of regressors \mathbf{W}_i :

$$[\tilde{\mathbf{W}}^j, \alpha_i^j] = \arg \min_{\tilde{\mathbf{W}}^{j'}, \alpha_i^{j'}} \sum_{i=1}^N \left\| \left(\sum_{j=1}^R \alpha_i^j \tilde{\mathbf{W}}^j \right) \mathbf{X}_i - \mathbf{Y}_i \right\|_F^2 + \lambda \left\| \tilde{\mathbf{W}} \right\|_F^2 \quad (2.101)$$

where $\tilde{\mathbf{W}} = [\tilde{\mathbf{W}}^1, \dots, \tilde{\mathbf{W}}^R]$.

However, more strongly constrained decompositions might be more suitable. In particular, considering a sparsity constraint on coefficients α_i^j may help reduce memory requirements even further.

2.4 Conclusion

In this chapter, we introduced the problem of blind source separation and associated models and algorithms, which we aim at extending to the problem of operator decomposition. Furthermore, we also introduced some of the problems and applications on which operator decomposition methods have been developed and applied, so as to provide some context relating to the general problem of operator decomposition and the wide variety of disciplines to which it relates.

Broadly speaking, blind source separation formulations provide efficient algorithms exploiting alternative constraints to classical orthogonal decomposition issues, most notably involving non-negativity and sparsity. These novel formulations have recently proved to be a powerful alternative to orthogonality-based models such as PCA, allowing for the development of more relevant and/or more interpretable decompositions. As far as applications are considered, the DMD algorithm exploits, in the context of dynamical system analysis, spectral decomposition principles to

Table 2.1 – Synthesis of constrained blind source separation algorithms. *Comparison criteria include enforced constraints, convergence properties, flexibility and code availability.*

Algorithm	Orthogonal	Sparse	Non-negative	Convergence	Flexibility	Code
PCA	✓	×	×	+	–	+
KSVD	×	✓	×	–	–	+
ALS	×	✓	✓	–	+	–
Proximal operators	✓	✓	✓	+	+	–

decompose the Koopman operator, an infinite-dimensional linearization of a spatio-temporally invariant dynamical system. Given its non-trivial shortcomings, a few extensions of DMD have been proposed to tackle its limitations, and related operator decomposition approaches exploiting other constraints have also been proposed in other scientific domains. Nonetheless, the problem of constrained operator decomposition remains, in general, considerably unexplored. This strongly supports the need for alternatively constrained operator decomposition models. In this respect, this thesis work aims at extending the constrained blind source separation formulations explored in this chapter to the problem of data-driven operator decomposition. To guide our efforts throughout this thesis dissertation, we present a synthesis of the explored constrained blind source separation algorithms in Table 2.1, where comparison criteria include enforced constraints, convergence properties, flexibility and code availability. These results will help us better conceive the operator decomposition models and algorithms we will introduce in the following chapters.

In the next chapter, we will thus draw inspiration from the applications and formulations introduced here, and from their limitations, in order to develop novel models by extending classical blind source separation issues to the problem of observation-based operator decomposition. Importantly, the methodological basis introduced here establishes the core foundations on which our proposed models will rely.

Proposed models and algorithms

"Essentially, all models are wrong, but some are useful."

George E. P. Box

3.1	Introduction	49
3.2	General model	50
3.3	Non-negative linear superposition of linear modes	51
3.3.1	Latent class model	52
3.3.2	Parameter estimation	53
3.3.3	Alternating least squares algorithm	54
3.4	Reformulation based on local linear operators	58
3.4.1	Parameter Estimation	59
3.4.2	Dictionary-based decomposition of local linear operators	61
3.4.3	Model training and application	62
3.4.4	Computational complexity analysis	62
3.5	Conclusion	63

3.1 Introduction

As stated in the previous chapter, inspired by the recent success of non-negative and sparse blind source separation, we aim at extending constrained blind source separation models to the data-driven characterization of operators. Overall, given the need for more complex formulations that can tackle the shortcomings of current models briefly identified and explored in previous sections, recent advances in blind source separation applications using sparse and non-negative constraints make them particularly appealing to address the observation-based characterization and decomposition of operators. This chapter addresses these issues and develops mathematically-sound and computationally-efficient schemes. Our main contributions are three-fold:

- A least-square formulation in the observation space under non-negativity constraints associated with different estimation algorithms;
- A reformulation of the considered non-negative issue as a dictionary learning problem to gain modeling flexibility, including the ability to consider alternative priors, such as sparsity;
- The experimental evaluation, in the next chapter, of the proposed numerical schemes, which point out the relevance of the dictionary learning framework.

Typically, we will focus on cases where constraints are imposed on either decomposition coefficients and/or on decomposition modes themselves. In particular, we introduce a novel non-negative decomposition model for linear operators and investigate different parameter estimation algorithms, including a reformulation of the model that exploits dictionary learning techniques. Most notably, this reformulation will allow us to gain increased flexibility and the ability to change model constraints in a simple way.

3.2 General model

In order to extend constrained blind source separation models to the problem of observation-based decomposition of operators, we start by formalizing the problem of operator decomposition.

Formally, the general and unconstrained observation-based decomposition of operators amounts to considering operators which relate variables of interest \mathbf{x} and \mathbf{y} , and state a general decomposition according to K modes as the superposition of K responses to input variable \mathbf{x} . For a dataset $\{\mathbf{x}_n, \mathbf{y}_n\}$, this may be given by:

$$\mathbf{y}_n = \sum_{k=1}^K \alpha_{nk} f_k(\mathbf{x}_n) + \boldsymbol{\omega}_n \quad (3.1)$$

where $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$, $\alpha_{nk} \in \mathbb{R}$ are mixing coefficients that model the contribution of each mode to the reconstruction of \mathbf{y}_n given \mathbf{x}_n , $f_k : \mathbb{R}^J \rightarrow \mathbb{R}^I$ is a linear or non-linear function associated with mode k , and $\boldsymbol{\omega}_n \in \mathbb{R}^I$ is a noise process, usually a Gaussian noise.

For the sake of simplicity, and given the envisioned applications, we will focus on the linear version of problem (3.1). The general, unconstrained linear problem can be expressed as the linear decomposition of the operator relating two observable variables \mathbf{x}_n and \mathbf{y}_n into K linear modes:

$$\mathbf{y}_n = \sum_{k=1}^K \alpha_{nk} \boldsymbol{\beta}_k \mathbf{x}_n + \boldsymbol{\omega}_n \quad (3.2)$$

where $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$, $\alpha_{nk} \in \mathbb{R}$ are mixing coefficients that model the contribution of each linear mode to the reconstruction of \mathbf{y}_n given \mathbf{x}_n , $\beta_k \in \mathbb{R}^{I \times J}$ is the regression matrix associated with mode k , and $\omega_n \in \mathbb{R}^I$ is a noise process, usually a Gaussian noise. Model (3.2) can be interpreted as a characterization of the relationship between \mathbf{x}_n and \mathbf{y}_n as the superposition of K linear relations.

3.3 Non-negative linear superposition of linear modes

Let us consider a multivariate observation dataset $\{\mathbf{x}, \mathbf{y}\}_n$, where $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$ denote the n^{th} observation pair. Variables \mathbf{x}_n and \mathbf{y}_n may, for instance, refer to feature vectors, image patches for different modalities or successive states of a dynamical system, depending on the targeted case-study. We focus on model (3.2) under the assumption that the potentially non-linear relationship between \mathbf{x}_n and \mathbf{y}_n , given by functional response $f_k(\mathbf{x}_n)$, can be locally approximated, with reasonable accuracy, by a linear operator. The idea of exploiting a local linear approximation of non-linear operators directly relates to classical approaches such as local linear embedding [222] and non-local means [25, 207]. We consider the decomposition of the approximated linear operator relating variables \mathbf{x}_n and \mathbf{y}_n under non-negativity constraints. As stated in [166], this translates into the following model for the relationship between variables \mathbf{x}_n and \mathbf{y}_n :

$$\begin{aligned} \mathbf{y}_n &= \sum_{k=1}^K \alpha_{nk} \beta_k \mathbf{x}_n + \omega_n \\ \text{Subject to } &\begin{cases} \alpha_{nk} \geq 0, & \forall k \in \llbracket 1, K \rrbracket, \forall n \in \llbracket 1, N \rrbracket \\ \|\beta_k\|_F = 1, & \forall k \in \llbracket 1, K \rrbracket \end{cases} \end{aligned} \quad (3.3)$$

where $\mathbf{x}_n \in \mathbb{R}^J$, $\mathbf{y}_n \in \mathbb{R}^I$, $\alpha_{nk} \in \mathbb{R}^+$ are non-negative mixing coefficients quantifying the contribution of linear mode k to the reconstruction of \mathbf{y}_n for a given \mathbf{x}_n , $\beta_k \in \mathbb{R}^{I \times J}$ is a regression matrix representing mode k , $\|\cdot\|_F$ is the Frobenius norm and $\omega_n \in \mathbb{R}^I$ is a centered Gaussian noise process with covariance matrix Σ , representing both model uncertainties and observation errors. N and K denote, respectively, the total number of observations and modes, while $k \in \llbracket 1, K \rrbracket$ and $n \in \llbracket 1, N \rrbracket$ indicate, respectively, the current mode and observation.

A non-negativity constraint has been imposed on mixing coefficients α_{nk} , drawing inspiration from the success of non-negative decompositions in applications where naturally occurring positive superposition of parts exists [155]. The non-negativity constraint provides the mean for distinguishing the form of the linear relationships between variables \mathbf{x}_n and \mathbf{y}_n from the magnitude of these relationships. Additionally, a normalization constraint on modal regression matrices β_k has been added to eliminate scaling indeterminacies and improve the identifiability

of the problem.

Model (3.3) allows us to address decomposition problems involving mixed linear contributions and generalizes linear mixture problems involving linear regressions and latent class regression models [47, 244]. The later might be viewed as a simplified version of our model (3.3), where mixing coefficients do not depend on index n and only one mode is actually active for each sample pair. Besides, a classical linear regression resorts to model (3.3) with $K = 1$ and $\alpha_{nk} = \alpha^* \quad \forall k \in \llbracket 1, K \rrbracket, \forall n \in \llbracket 1, N \rrbracket$.

Compared to state-of-the-art models like classical regression and latent class regression models [47, 244], the key features of our proposed formulation are two-fold: first, it accounts for possibly varying magnitudes of the linear relationships; second, it explicitly evaluates the relative importance of different linear relationships. This is of wide interest for a variety applications such as regression hypothesis testing, transfer function identification, regime-switching dynamics, etc.

Finally, it may also be noted that any non-linear decomposition model (3.2) may be restated as a linear decomposition model (3.3) according to the vector of regression variables $(f_0(\mathbf{x}_n), \dots, f_K(\mathbf{x}_n))$. In the subsequent, we assume that candidate non-linear functional responses $(f_K(\mathbf{x}_n))$ are given a priori and we address the estimation of mixing coefficients (α_{nk}) and regression matrices (β_k) .

3.3.1 Latent class model

Under the assumption that for each pair of variables \mathbf{x}_n and \mathbf{y}_n there is a strongly dominant mode, that is to say that for any sample pair $(\mathbf{x}_n, \mathbf{y}_n)$ only one of the mixing coefficients α_{nk} is non-null, we can introduce a hidden variable z_n that indicates the mode which operates for each sample pair $(\mathbf{x}_n, \mathbf{y}_n)$, and derive a simplified latent class formulation for model (3.3) :

$$\begin{cases} P(\mathbf{y}_n | z_n = k) \sim \mathcal{N}(0, \boldsymbol{\Sigma}_k, r(\mathbf{x}_n, \mathbf{y}_n)) \\ r(\mathbf{x}_n, \mathbf{y}_n) = \mathbf{y}_n - \alpha_{nk}^* \beta_k \mathbf{x}_n \\ \alpha_{nk}^* = \operatorname{argmin}_{\alpha_{nk}} \|\mathbf{y}_n - \alpha_{nk} \beta_k \mathbf{x}_n\|_{\boldsymbol{\Sigma}_k}^2 \end{cases} \quad (3.4)$$

where $\mathcal{N}(0, \boldsymbol{\Sigma}_k, \cdot)$ is a centered Gaussian distribution with covariance matrix $\boldsymbol{\Sigma}_k$ and $\|\cdot\|_{\boldsymbol{\Sigma}_k}$ is the mode dependent covariance-weighted distance.

It may be noted that this latent class model still generalizes latent class regression models as used in [244], since non-null coefficients α_{nk} may still vary with respect to sample index n . Given its lower computational complexity, this latent class model is of particular interest in the proposed multi-stage calibration process for model (3.3), and may be used as an initialization.

3.3.2 Parameter estimation

We state the estimation of model parameters for model (3.3) from a set of observations $\{\mathbf{x}, \mathbf{y}\}_n$ as the resolution of the following non-linear, non-convex constrained optimization problem:

$$\forall n, \begin{cases} [\hat{\alpha}_{nk}, \hat{\beta}_k] = \arg \min_{\alpha_{nk}, \beta_k} \sum_{m=1}^N W_m^n \left\| \mathbf{y}_m - \sum_{k=1}^K \alpha_{nk} \beta_k \mathbf{x}_m \right\|_{\Sigma}^2 \\ \alpha_{nk} \geq 0, & \forall n \in \llbracket 1, N \rrbracket, \forall k \in \llbracket 1, K \rrbracket \\ \|\beta_k\|_F = 1, & \forall k \in \llbracket 1, K \rrbracket \end{cases} \quad (3.5)$$

where $\|\cdot\|_{\Sigma}$ is a weighted norm according to covariance Σ . We assume that, according to weighing factors W_m^n , multiple observation pairs $(\mathbf{x}_m, \mathbf{y}_m)$ may share relatively similar mixing coefficients $\{\alpha_{nk}\}$. The greater W_m^n , the more similar the expected mixing coefficients $\{\alpha_{nk}\}$ and $\{\alpha_{mk}\}$. Weighing matrix \mathbf{W} may encode both space-time smoothness priors, such that observation pairs close in space and/or time are expected to share similar operator decompositions, as well as observation-space similarity priors, for instance that observation pairs with similar regression variables may share similar decompositions. This seems reasonable for many applications where parameters are expected to correlate and vary smoothly in the considered spatio-temporal space. The parameterization of weighing matrix \mathbf{W} is expected to be application-dependent and may be related to similar ideas used in covariance-based modeling [16] and non-local schemes [25, 207]. Regarding identifiability issues, if the number of modes K verifies $K > I$ (where I is the dimension of observation vector \mathbf{y}_n), the estimation of mixing coefficients α_{nk} becomes intractable from a single observation pair $(\mathbf{x}_n, \mathbf{y}_n)$ ¹. As such, weighing matrix also provides a means to address the estimation of mixing parameters in such situations.

3.3.2.1 Single observation solution

An interesting particular case of model (3.3), studied in [166], arises when $K \leq I$. In this case, model parameters may be estimated from only observation pair $(\mathbf{x}_n, \mathbf{y}_n)$ for each index n , with relates to the parameterization of weighing matrix \mathbf{W} as

$$W_m^n = \begin{cases} 1, & m = n \\ 0, & m \neq n \end{cases} \quad (3.6)$$

¹For a fixed set of linear modes β_k , the estimation of mixing coefficients α_{nk} requires solving a linear system involving K unknowns and I equations

and translates to the following constrained minimization problem:

$$\forall n, \begin{cases} [\hat{\alpha}_{nk}, \hat{\beta}_k] = \arg \min_{\alpha_{nk}, \beta_k} \left\| \mathbf{y}_n - \sum_{k=1}^K \alpha_{nk} \beta_k \mathbf{x}_n \right\|_{\Sigma}^2 \\ \alpha_{nk} \geq 0, & \forall n \in \llbracket 1, N \rrbracket, \forall k \in \llbracket 1, K \rrbracket \\ \|\beta_k\|_F = 1, & \forall k \in \llbracket 1, K \rrbracket \end{cases} \quad (3.7)$$

Nonetheless, it should be noted that the condition $K \leq I$ does not singlehandedly guarantee that Equation (3.7) will have a solution. Indeed, for pathological cases where the system's Gramian matrix is not invertible more observations need to be considered to compute a solution (as in Equation (3.5)). In this respect, a compromise exists between the number of observations considered, which will increase model robustness and numerical stability, and the locality of the model, which increases as less observations are considered.

3.3.3 Alternating least squares algorithm

3.3.3.1 Direct minimization approach

Given the non-linear, non-convex nature of constrained minimization problem (3.5), the joint estimation of model parameters α_{nk} and β_k is not straightforward. Conveniently, this jointly non-convex minimization problem becomes convex when estimation is performed for one set of parameters only while considering the other set of parameters to be fixed. Naturally, this suggests an alternating minimization approach, which leads to the following updates of model parameter sets α_{nk} and β_k being iterated until convergence:

β -step: Minimization over β_k with fixed α_{nk} and externally forced normalization constraints²

$$\hat{\beta}_k^{i+1} = \hat{\beta}_k^i + \left[\sum_{n=1}^N \hat{\alpha}_{nk}^i \left(\mathbf{y}_n - \sum_{p=1}^K \hat{\alpha}_{np}^i \hat{\beta}_p^i \mathbf{x}_n \right) \mathbf{x}_n^T \right] \left[\sum_{n=1}^N \left(\hat{\alpha}_{nk}^i \right)^2 \mathbf{x}_n \mathbf{x}_n^T \right]^{-1} \quad (3.8)$$

$$\hat{\beta}_k^{i+1} = \frac{\hat{\beta}_k^{i+1}}{\|\hat{\beta}_k^{i+1}\|_F}, \quad \forall k \in \llbracket 1, K \rrbracket \quad (3.9)$$

α -step: Minimization over α_{nk} with fixed β_k and externally forced non-negativity constraints

$$\hat{\alpha}_{nk}^{i+1} = \hat{\alpha}_{nk}^i + \frac{\sum_{m=1}^N W_m^n \left[\mathbf{x}_m^T \left(\hat{\beta}_k^i \right)^T \Sigma^{-1} \left(\mathbf{y}_m - \sum_{p=1}^K \hat{\alpha}_{np}^i \hat{\beta}_p^i \mathbf{x}_m \right) \right]}{\sum_{m=1}^N W_m^n \left[\mathbf{x}_m^T \left(\hat{\beta}_k^i \right)^T \Sigma^{-1} \hat{\beta}_k^i \mathbf{x}_m \right]} \quad (3.10)$$

²Since linear modes β_k are shared by all observation pairs, one set of regression matrices is estimated using all observation pairs in the training dataset. In this respect, all observation pairs are weighted equally for the estimation of modal regression matrices, under the assumption that they contribute uniformly to the estimation of the globally shared linear modes. Hence, to correctly fit the model, all global linear modes should be adequately sampled so as to be represented equally within the training dataset.

$$\hat{\alpha}_{nk}^{i+1} = \max \left\{ 0, \hat{\alpha}_{nk}^{i+1} \right\} \quad (3.11)$$

For the case where a single observation pair suffices to estimated model parameters (Equation (3.7)), the alternating minimization approach presented here is still valid, but requires Equation (3.10) to be modified in order to consider the current observation pair $(\mathbf{x}_n, \mathbf{y}_n)$ only:

$$\hat{\alpha}_{nk}^{i+1} = \hat{\alpha}_{nk}^i + \frac{\mathbf{x}_n^T (\hat{\boldsymbol{\beta}}_k^i)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{y}_n - \sum_{p=1}^K \hat{\alpha}_{np}^i \hat{\boldsymbol{\beta}}_p^i \mathbf{x}_n \right)}{\mathbf{x}_n^T (\hat{\boldsymbol{\beta}}_k^i)^T \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{\beta}}_k^i \mathbf{x}_n} \quad (3.12)$$

3.3.3.2 Gradient-splitting approach

The downside to the simplicity of the alternating minimization approach is that it is prone to numerical issues. As acknowledged in the blind source separation literature [10], the alternating projections on the unconstrained and constrained solution spaces may induce divergent or numerically unstable behaviour. To handle such a problem, the direct minimization introduced in Equation (3.10) may be softened by considering a gradient descent:

$$\hat{\alpha}_{nk}^{i+1} = \hat{\alpha}_{nk}^i + 2\delta \left[\sum_{m=1}^N W_m^n \mathbf{x}_m^T (\hat{\boldsymbol{\beta}}_k^i)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{y}_m - \sum_{p=1}^K \hat{\alpha}_{mp}^i \hat{\boldsymbol{\beta}}_p^i \mathbf{x}_m \right) \right] \quad (3.13)$$

where δ is the used defined gradient descent step.

For the single-observation case (Equation (3.7)), the α -step reduces to:

$$\hat{\alpha}_{nk}^{i+1} = \hat{\alpha}_{nk}^i + 2\delta \left[\mathbf{x}_n^T (\hat{\boldsymbol{\beta}}_k^i)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{y}_n - \sum_{p=1}^K \hat{\alpha}_{np}^i \hat{\boldsymbol{\beta}}_p^i \mathbf{x}_n \right) \right] \quad (3.14)$$

This is then combined with a projection onto the constrained non-negative solution space (Equation (3.11)), which comes down to a gradient based proximal splitting method [37].

Even though less necessary (since the renormalization constraint imposed in Equation (3.9) comes down to a simple rescaling), the same gradient based reformulation can be used for the estimation of modal linear regression matrices $\boldsymbol{\beta}_k$:

$$\hat{\boldsymbol{\beta}}_k^{i+1} = \hat{\boldsymbol{\beta}}_k^i + 2\delta \left[\sum_{n=1}^N \hat{\alpha}_{nk}^i \boldsymbol{\Sigma}^{-1} \left(\mathbf{y}_n - \sum_{p=1}^K \hat{\alpha}_{np}^i \hat{\boldsymbol{\beta}}_p^i \mathbf{x}_n \right) \mathbf{x}_n^T \right] \quad (3.15)$$

3.3.3.3 Latent class model calibration

For latent class model (3.4), we derive a classical iterative Expectation-Maximization (EM) procedure [14, 46, 232] to infer model parameters $\boldsymbol{\beta}_k$ and α_{nk} according to a maximum likelihood criterion. At iteration i , the E-step comes to compute the posterior of hidden variables z_n given

current model parameters β_k^i and α_{nk} :

$$\tau_{nk}^{i+1} = P(z_n = k | \mathbf{x}_n, \mathbf{y}_n) = \frac{\pi_k^i \mathcal{N}(0, \Sigma_k, \mathbf{y}_n - \hat{\alpha}_{nk}^i \hat{\beta}_k^i \mathbf{x}_n)}{\sum_{l=1}^K \pi_l^i \mathcal{N}(0, \Sigma_k, \mathbf{y}_n - \hat{\alpha}_{nl}^i \hat{\beta}_l^i \mathbf{x}_n)} \quad (3.16)$$

The M-step then updates model parameters as follows:

$$\pi_k^{i+1} = P(z = k) = \frac{1}{N} \sum_{n=1}^N \tau_{nk}^{i+1} \quad (3.17)$$

$$\hat{\beta}_k^{i+1} = \left[\frac{1}{N} \sum_{n=1}^N \tau_{nk}^{i+1} \mathbf{y}_n \hat{\alpha}_{nk}^i \mathbf{x}_n^T \right] \left[\frac{1}{N} \sum_{n=1}^N \tau_{nk}^{i+1} \hat{\alpha}_{nk}^i \mathbf{x}_n \hat{\alpha}_{nk}^i \mathbf{x}_n^T \right]^{-1} \quad (3.18)$$

$$\hat{\alpha}_{nk}^{i+1} = \frac{\mathbf{y}_n^T (\Sigma_k^i)^{-1} \hat{\beta}_k^{i+1} \mathbf{x}_n}{\mathbf{x}_n^T (\hat{\beta}_k^{i+1})^T (\Sigma_k^i)^{-1} \hat{\beta}_k^{i+1} \mathbf{x}_n} \quad (3.19)$$

$$\Sigma_k^{i+1} = \frac{\sum_{n=1}^N \tau_{nk}^{i+1} (\mathbf{y}_n - \hat{\alpha}_{nk}^{i+1} \hat{\beta}_k^{i+1} \mathbf{x}_n) (\mathbf{y}_n - \hat{\alpha}_{nk}^{i+1} \hat{\beta}_k^{i+1} \mathbf{x}_n)^T}{\sum_{n=1}^N \tau_{nk}^{i+1}} \quad (3.20)$$

Non-negativity constraints over α_{nk} and normalization constraints over β_k are forced after each EM iteration (Equations (3.9) and (3.11)).

Experimentally, it has been noted that an ill-conditioned or nearly singular covariance matrix may result in numerical instabilities in the update of scalars α_{nk} (Equation (3.19)). Hence, we may consider a variant of the algorithm with a spherical covariance matrix, $\Sigma_k^i = \sigma^2 \mathbf{I}_I, \forall k \in [1, K]$. Similar approaches have proven to be effective in the context of Feasible Generalized Least Squares (FGLS), where considering an spherical covariance structure has proven to improve algorithm robustness [220, 266]. We will refer to this variant of the EM algorithm as the pseudo-EM algorithm.

3.3.3.4 Moment-based estimation of regression parameters

Regarding initialization issues, we also derive a moment-based estimation of regression matrices β_k . Noting that all amplitude information is contained in coefficients α_{nk} , a direct application of the orthogonality principle [190] yields that regression matrix β_k that minimizes the root mean square estimation error is the matrix that causes vectors $\beta_k \mathbf{x}_n$ and \mathbf{y}_n to be collinear for all sample pairs $(\mathbf{x}_n, \mathbf{y}_n)$ assigned to mode k . Therefore, assuming that we are provided with some estimates or initial values of posteriors τ_{nk} (Equation (3.16)), we can derive the following moment-based estimate for regression matrix β_k :

$$\hat{\beta}_k = \left[\frac{1}{N} \sum_{n=1}^N \tau_{nk} \mathbf{y}_n \mathbf{x}_n^T \right] \left[\frac{1}{N} \sum_{n=1}^N \tau_{nk} \mathbf{x}_n \mathbf{x}_n^T \right]^{-1} \quad (3.21)$$

We then define a moment-based iterative conditional estimation (MICE) algorithm, following [5,210–212]. It consists in replacing the update of regression matrices β_k in the M-step of the EM procedures (Equation (3.18)) by the above moment-based update (Equation (3.21)). As detailed below, this moment-based estimation is considered as initialization to the EM iterations, since it does not require any knowledge regarding initial values for mixing coefficients α_{nk} .

3.3.3.5 Full implementation of the ALS algorithm

We detail in Algorithm 1 the overall procedure for the calibration of model (3.3). We adopt a greedy approach to improve convergence properties. As initialization, we perform an initial K class clustering of the data using a simple k-means algorithm [128,180,192], compute initial binary posteriors τ_{nk} , and exploit the MICE update of regression matrices β_k and mixing coefficients α_{nk} . We then iterate the EM (or pseudo-EM) procedure until convergence for latent class model (3.4) and follow with iterations of the ALS procedure (exploiting either a direct minimization scheme or a gradient descent approach) until convergence for general model (3.3).

Algorithm 1 Non-negative Alternating Least Squares. *Taken from [166]. © 2016 IEEE.*

- 1: Input: $\{\mathbf{x}_n\}, \{\mathbf{y}_n\}$
 - 2: Initialization:
 - $\tau_{nk}^0 \leftarrow \text{k-means}(\mathbf{x}_n, \mathbf{y}_n)$
 - $[\hat{\alpha}_{nk}^1, \hat{\beta}_k^1, \tau_{nk}^1, \pi_k^1, \Sigma_k^1] \leftarrow \text{MICE}(\mathbf{x}_n, \mathbf{y}_n, \tau_{nk}^0)$
 - 3: **repeat**
 - $[\hat{\alpha}_{nk}^{i+1}, \hat{\beta}_k^{i+1}, \tau_{nk}^{i+1}, \pi_k^{i+1}, \Sigma_k^{i+1}] \leftarrow \text{EM}(\mathbf{x}_n, \mathbf{y}_n, \hat{\alpha}_{nk}^i, \hat{\beta}_k^i, \tau_{nk}^i, \pi_k^i, \Sigma_k^i)$
 - or**
 - $[\hat{\alpha}_{nk}^{i+1}, \hat{\beta}_k^{i+1}, \tau_{nk}^{i+1}, \pi_k^{i+1}, \Sigma_k^{i+1}] \leftarrow \text{pseudo-EM}(\mathbf{x}_n, \mathbf{y}_n, \hat{\alpha}_{nk}^i, \hat{\beta}_k^i, \tau_{nk}^i, \pi_k^i, \Sigma_k^i)$
 - 4: **until** convergence
 - 5: **repeat**
 - $[\hat{\alpha}_{nk}^{i+1}, \hat{\beta}_k^{i+1}] \leftarrow \text{ALS}(\mathbf{x}_n, \mathbf{y}_n, \hat{\alpha}_{nk}^i, \hat{\beta}_k^i)$
 - 6: **until** convergence
-

3.3.3.6 Estimator bias and variance

It is straightforward to demonstrate that the parameter estimators $\hat{\alpha}_{nk}$ and $\hat{\beta}_k$ (Equations (3.9), (3.10) and (3.12)) are unbiased (i.e. $\mathbb{E}(\hat{\alpha}_{nk}) = \alpha_{nk}$ and $\mathbb{E}(\hat{\beta}_k) = \beta_k$) and have variances given by:

$$\text{var}(\hat{\beta}_k) = \left[\sum_{n=1}^N \alpha_{nk}^2 \left\| \left(\sum_{n=1}^N \alpha_{nk}^2 \mathbf{x}_n \mathbf{x}_n^T \right)^{-1} \mathbf{x}_n \right\|_F^2 \right] \Sigma \quad (3.22)$$

$$\text{var}(\hat{\alpha}_{nk}) = \frac{\sum_{m=1}^N (W_m^n)^2 \|\boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}_k \mathbf{x}_m\|_F^2}{\left(\sum_{m=1}^N W_m^n \|\boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}_k \mathbf{x}_m\|_F^2 \right)^2} \quad (3.23)$$

For the single observation case, we obtain:

$$\text{var}(\hat{\alpha}_{nk}) = \frac{1}{\|\boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}_k \mathbf{x}_n\|_F^2} \quad (3.24)$$

In this context, Equation (3.22) gives the covariance between lines $[\hat{\boldsymbol{\beta}}_k]_{i \cdot}$ of the estimator of the k^{th} -mode linear regression matrix. Alternatively, one can compute the element-wise second order moment of estimator $\hat{\boldsymbol{\beta}}_k$:

$$\begin{aligned} \text{var}([\hat{\boldsymbol{\beta}}_k]_{ij}) &= \mathbb{E} \left\{ \left([\hat{\boldsymbol{\beta}}_k]_{ij} - \overline{[\hat{\boldsymbol{\beta}}_k]_{ij}} \right) \left([\hat{\boldsymbol{\beta}}_k]_{i^* j^*} - \overline{[\hat{\boldsymbol{\beta}}_k]_{i^* j^*}} \right) \right\} \\ &= \sum_{n=1}^N \alpha_{nk}^2 \mathbf{x}_n^T \left[\left(\sum_{n=1}^N \alpha_{nk}^2 \mathbf{x}_n \mathbf{x}_n^T \right)^{-1} \right]_{:j} \left[\left(\sum_{n=1}^N \alpha_{nk}^2 \mathbf{x}_n \mathbf{x}_n^T \right)^{-1} \right]_{j^*} \mathbf{x}_n [\boldsymbol{\Sigma}]_{ii^*} \end{aligned} \quad (3.25)$$

where $\overline{[\hat{\boldsymbol{\beta}}_k]_{ij}}$ is the first order moment of estimator $\hat{\boldsymbol{\beta}}_k$, given by:

$$\overline{[\hat{\boldsymbol{\beta}}_k]_{ij}} = [\boldsymbol{\beta}]_{i \cdot} \left[\sum_{n=1}^N \alpha_{nk}^2 \mathbf{x}_n \mathbf{x}_n^T \right] \left[\left(\sum_{n=1}^N \alpha_{nk}^2 \mathbf{x}_n \mathbf{x}_n^T \right)^{-1} \right]_{:j} = [\boldsymbol{\beta}]_{ij} \quad (3.26)$$

3.4 Reformulation based on local linear operators

As detailed below, the considered decomposition issue may be restated as a dictionary learning problem. In (3.3), linear operator $\sum_{k=1}^K \alpha_{nk} \boldsymbol{\beta}_k$ can be regarded as a decomposition of the local linear operator relating variables \mathbf{y} and \mathbf{x} for index n . This local linear operator may be estimated as follows according to a weighted least-square criterion using weighing matrix \mathbf{W} :

$$\boldsymbol{\Theta}_n = \left(\sum_{m=1}^N W_m^n \mathbf{y}_m \mathbf{x}_m^T \right) \left(\sum_{m=1}^N W_m^n \mathbf{x}_m \mathbf{x}_m^T \right)^{-1} \quad (3.27)$$

where again W_m^n are weighting coefficients that account for the relative contributions of observation pairs $(\mathbf{x}_m, \mathbf{y}_m)$ to the estimation of the linear operator $\boldsymbol{\Theta}_n$ relating observation pair $(\mathbf{x}_n, \mathbf{y}_n)$. This least-square estimate comes to solve independently the least-square criterion for each index n in (3.3). Here, as in model (3.3), there is also a compromise between model robustness and computational stability and model locality, ultimately determined by the number of auxiliary observations considered for the estimation of local linear operators $\boldsymbol{\Theta}_n$.

Given local models $\{\Theta_n\}_n$, problem (3.3) relates to the non-negative decomposition of linear operators Θ_n . It can be shown that model (3.3) (which yields optimization problem (3.5)) can be reformulated as:

$$\Theta_n = \sum_{k=1}^K \alpha_{nk} \beta_k + \overbrace{\left(\sum_{m=1}^N W_m^n \omega_m \mathbf{x}_m \mathbf{x}_m^T \right) \left(\sum_{m=1}^N W_m^n \mathbf{x}_m \mathbf{x}_m^T \right)^{-1}}^{\Upsilon_n} \quad (3.28)$$

Subject to $\begin{cases} \alpha_{nk} \geq 0, & \forall k \in \llbracket 1, K \rrbracket, \forall n \in \llbracket 1, N \rrbracket \\ \|\beta_k\|_F = 1, & \forall k \in \llbracket 1, K \rrbracket \end{cases}$

The reformulation introduced by the estimation of local linear operators Θ_n induces an error matrix $\Upsilon_n \in \mathbb{R}^{I \times J}$ that depends directly on observations \mathbf{x}_m and weights W_m^n . However, given the Gaussian nature of the original error term ω_m , the new error matrix Υ_n , being a linear combination of Gaussian terms, is a Gaussian matrix.

Computation of the first and second order moments of the new error matrix elements $[\Upsilon_n]_{ij}$ introduced in Equation (3.28) gives:

$$\mathbb{E}\{[\Upsilon_n]_{ij}\} = 0 \quad (3.29)$$

$$\begin{aligned} [\Psi_n]_{(ij)(i^*j^*)} &= \mathbb{E}\{[\Upsilon_n]_{ij}[\Upsilon_n]_{i^*j^*}\} \\ &= \sum_{m=1}^N (W_m^n)^2 \mathbf{x}_m^T \left[\left(\sum_{m=1}^N W_m^n \mathbf{x}_m \mathbf{x}_m^T \right)^{-1} \right]_{:j} \left[\left(\sum_{m=1}^N W_m^n \mathbf{x}_m \mathbf{x}_m^T \right)^{-1} \right]_{j^*} \mathbf{x}_m [\Sigma]_{ii^*} \end{aligned} \quad (3.30)$$

where sub-indexes $[A]_{:j}$ and $[A]_{l:}$ denote, respectively, the j -th column and the l -th line of matrix A . This leads to the conclusion that Υ_n (and thus Θ_n) is heteroscedastic, i.e., its elements present a non-constant variance Ψ_n that depends on the observations \mathbf{x}_n and weights W_m^n used to estimate the considered linear operator Θ_n . In this respect, adequately choosing the linear regression weights W_m^n should allow us to better manage the heteroscedastic nature of model (3.28).

3.4.1 Parameter Estimation

Given local models $\{\Theta_n\}_n$, parameter estimation for model (3.28) then translates to the following constrained optimization problem:

$$\begin{cases} [\hat{\alpha}_{nk}, \hat{\beta}_k] = \arg \min_{\alpha_{nk}, \beta_k} \sum_{n=1}^N \left(\left\| \Theta_n - \sum_{k=1}^K \alpha_{nk} \beta_k \right\|_{\Psi_n}^2 \right) \\ \alpha_{nk} \geq 0, & \forall n \in \llbracket 1, N \rrbracket, \forall k \in \llbracket 1, K \rrbracket \\ \|\beta_k\|_2 = 1, & \forall k \in \llbracket 1, K \rrbracket \end{cases} \quad (3.31)$$

3.4.1.1 Direct minimization

The unconstrained direct minimization of the cost function introduced in (3.31) yields the following estimators for model parameters:

$$\hat{\alpha}_{nk} = \frac{\text{vec}(\boldsymbol{\beta}_k)^T (\boldsymbol{\Psi}_n)^{-1} \left(\text{vec}(\boldsymbol{\Theta}_n) - \sum_{\substack{p=1 \\ p \neq k}}^K \alpha_{np} \text{vec}(\boldsymbol{\beta}_p) \right)}{\text{vec}(\boldsymbol{\beta}_k)^T (\boldsymbol{\Psi}_n)^{-1} \text{vec}(\boldsymbol{\beta}_k)} \quad (3.32)$$

$$\hat{\boldsymbol{\beta}}_k = \frac{\sum_{n=1}^N \alpha_{nk} \left(\boldsymbol{\Theta}_n - \sum_{\substack{p=1 \\ p \neq k}}^K \alpha_{np} \boldsymbol{\beta}_p \right)}{\sum_{n=1}^N \alpha_{nk}^2} \quad (3.33)$$

where $\text{vec}(\cdot)$ denotes the matrix vectorization operator and $\boldsymbol{\Psi}_n$ is the covariance matrix of vectorized error matrix $\mathbf{v}_n = \text{vec}(\boldsymbol{\Upsilon}_n)$. It is important to notice that vectorized error matrix \mathbf{v}_n remains heteroscedastic, with its observation-dependent covariance matrix $\boldsymbol{\Psi}_n$ containing the second order moments of error matrix $\boldsymbol{\Upsilon}_n$.

3.4.1.2 Estimator bias and variance

It can be proven that estimators (3.32) and (3.33) are unbiased and present variances given by:

$$\text{var}(\hat{\alpha}_{nk}) = \frac{1}{\text{vec}(\boldsymbol{\beta}_k)^T \boldsymbol{\Psi}_n^{-1} \text{vec}(\boldsymbol{\beta}_k)} \quad (3.34)$$

$$\text{var}(\hat{\boldsymbol{\beta}}_k) = \frac{\sum_{n=1}^N \alpha_{nk}^2 \boldsymbol{\Psi}_n}{\left(\sum_{n=1}^N \alpha_{nk}^2 \right)^2} \quad (3.35)$$

3.4.1.3 ALS-based formulation

Similarly to the formulation introduced for (3.7), an ALS-base formulation that couples these estimators (either via direct minimization or a gradient-based proximal-splitting method) with projections onto the corresponding constrained sub-spaces could be used to numerically resolve constrained minimization problem (3.31). However, the minimization of this least-square criterion using ALS or gradient-splitting schemes would require the computation of an error covariance matrix $\boldsymbol{\Psi}_n$ for each local linear operator $\boldsymbol{\Theta}_n$, which implies a considerable increase in the computational complexity. Therefore, we propose an alternative method based on the dictionary-based decomposition of vectorized versions of local linear operators $\boldsymbol{\Theta}_n$ where, for the sake of simplicity, we also consider a simplified homoscedastic covariance structure.

3.4.2 Dictionary-based decomposition of local linear operators

The constrained minimization problem presented in Equation (3.31) can be reformulated as a blind dictionary learning based decomposition. To do so, we consider the set $\{\Theta\}_n$ of all N local linear operators, to which we apply the vectorization operator in order to rewrite Equation (3.31) as:

$$\begin{cases} [\hat{\mathbf{A}}, \hat{\mathbf{B}}] = \underset{\mathbf{A}, \mathbf{B}}{\operatorname{argmin}} & \|\Phi - \mathbf{B}\mathbf{A}\|_F^2 \\ \mathbf{A}_{kn} \geq 0, & \forall k \in \llbracket 1, K \rrbracket, \forall n \in \llbracket 1, N \rrbracket \\ \|\llbracket \mathbf{B} \rrbracket_{:,k}\|_2 = 1, & \forall k \in \llbracket 1, K \rrbracket \end{cases} \quad (3.36)$$

where matrix $\Phi \in \mathbb{R}^{IJ \times N}$ is obtained by concatenating vectorized operators $\theta_n = \operatorname{vec}(\Theta_n)$ (i.e. $\Phi = [\theta_1 | \dots | \theta_N]$), columns of matrix $\mathbf{A} \in \mathbb{R}^{K \times N}$ contain mixing coefficients α_{nk} quantifying the contribution of each mode k for the reconstruction of vectorized local linear operator θ_n and columns of $\mathbf{B} \in \mathbb{R}^{IJ \times K}$ (noted as $\llbracket \mathbf{B} \rrbracket_{:,k}$) contain vectorized versions of modal linear regression matrices β_k , i.e., $\llbracket \mathbf{B} \rrbracket_{:,k} = \operatorname{vec}(\beta_k)$.

The estimation of model parameters for model (3.36) resorts, under this new formulation, to a classical dictionary learning problem coupled with a non-negativity constraint. Dictionary learning is a classical problem in signal processing, for which numerous methods, exploiting different constraints, have been proposed [3, 10, 13, 155, 156]. Moreover, the dictionary learning based reformulation has some considerable advantages when compared to other proposed approaches [166]. Effectively, this formulation is more flexible and adaptable than previously introduced models, since model constraints can be changed seamlessly simply by changing the blind dictionary learning approach used to solve constrained minimization (3.36) (e.g. NMF [155], KSVD [3], etc). Here, since we consider a non-negative constraint, we solve minimization (3.36) using a proximal splitting method [37] to account for the non-negativity of mixing coefficients matrix \mathbf{A} . It involves the iteration of the following two steps until convergence:

1. The least-squares estimation of dictionary matrix \mathbf{B} under normalization constraints

$$\|\llbracket \mathbf{B} \rrbracket_{:,k}\|_2 = 1, \forall k:$$

$$\mathbf{B}^{i+1} = \Phi (\mathbf{A}^i)^T \left(\mathbf{A}^i (\mathbf{A}^i)^T \right)^{-1} \quad (3.37)$$

$$\llbracket \mathbf{B}^{i+1} \rrbracket_{:,k} = \frac{\llbracket \mathbf{B}^{i+1} \rrbracket_{:,k}}{\|\llbracket \mathbf{B}^{i+1} \rrbracket_{:,k}\|_2} \quad \forall k \in \llbracket 1, K \rrbracket \quad (3.38)$$

2. The estimation of the mixing coefficients matrix \mathbf{A} using a gradient descent based proximal splitting method [37] to enforce non-negativity:

$$\mathbf{A}^{i+1} = \mathbf{A}^i - 2\lambda (\mathbf{B}^i)^T (\Phi - \mathbf{B}^i \mathbf{A}^i) \quad (3.39)$$

$$\left[\mathbf{A}^{i+1}\right]_{kn} = \max\left\{0, \left[\mathbf{A}^{i+1}\right]_{kn}\right\}, \forall k \in \llbracket 1, K \rrbracket, \forall n \in \llbracket 1, N \rrbracket \quad (3.40)$$

Alternatively, one may choose a different dictionary-learning technique to enforce a different constraint (e.g. KSVD [3] for sparsity). This gives the dictionary-based formulation increased flexibility and adaptability, since alternative model constraints can thus be introduced seamlessly into model (3.3).

3.4.3 Model training and application

We may distinguish two different situations in terms of model parameter estimation for this dictionary-based formulation:

- **Model training:** Regression matrices $\hat{\beta}_k$ (matrix $\hat{\mathbf{B}}$ in formulation (3.36)) and mixing coefficients $\hat{\alpha}_{nk}$ (matrix $\hat{\mathbf{A}}$ in formulation (3.36)) are jointly estimated for a set of local linear operators Θ_n obtained from a training dataset $\{\mathbf{x}, \mathbf{y}\}_n$. Estimated regression matrices $\hat{\beta}_k$ will be considered as the dictionary of regression modes β_k when the model is applied to new observations and are thus stored for future use.
- **Model application:** Given a trained dictionary of operators $\{\hat{\beta}\}_k$ (matrix $\hat{\mathbf{B}}$ in formulation (3.36)), mixing coefficients $\hat{\alpha}_{nk}$ (matrix $\hat{\mathbf{A}}$ in formulation (3.36)) are estimated for a new observation dataset $\{\mathbf{x}^*, \mathbf{y}^*\}_n$. Two approaches may be considered. Similarly to the training step, linear operators $\{\Theta^*\}_n$ can be estimated for the new dataset, and mixing coefficients can be computed by projecting these operators onto the previously trained dictionary (using a non-negativity constraint). Alternatively, mixing coefficients can be estimated directly from observations using a least-squares criterion derived from model (3.3) without the prior estimation of linear operators $\{\Theta^*\}_n$. Both approaches can be implemented using proximal operators, as in the model training step, or classical non-negative least-squares solvers [145]. It should be noted that the estimation of mixing coefficients α_{nk}^* for new observations $(\mathbf{x}_n^*, \mathbf{y}_n^*)$ may exploit only data from the training dataset, which, in the context of dynamical system prediction, provides the algorithm with actual prediction capabilities (since no knowledge of $\mathbf{y}_n^* = \mathbf{x}_{n+1}^*$ is needed for the estimation of mixing coefficients α_{nk}).

3.4.4 Computational complexity analysis

Table 3.1 presents a summary of complexity of the different algorithms, namely the alternating least squares exploiting a direct minimization (Equations (3.8), (3.9), (3.10) and (3.11)), the alternating least squares exploiting a gradient descent (Equations (3.8), (3.9), (3.13) and (3.11)) and the dictionary-based local linear operator decomposition (Equations (3.37), (3.38), (3.39) and (3.40)), expressed in number of operations. Subsequently, we will refer to these algorithms as ALS-direct, ALS-gradient and LLOD, respectively. From these results, it is clear that differences

Table 3.1 – Computational cost of the different steps of the proposed algorithms, in number of operations.

	ALS-direct	ALS-gradient	LLOD
Θ_n estimation	-	-	$NM[2J^2 + 2IJ]$ $+N[J^3 + IJ^2]$
α/\mathbf{A} -step	$2NK[MKIJ + 1]$ $+NK[I^2 + IJ + 2I + 1]$ $+MNK[I^2 + 2IJ + J + 1]$	$2NK[MKIJ + 1]$ $+NK[I^2 + IJ + 2I + 1]$ $+1$	$NK[2IJ + 2]$ $+NIJ + 1$
$\alpha/\mathbf{A} \geq 0$	NK	NK	NK
β/\mathbf{B} -step	$NK[IJ + I]$ $+N[2J^2 + 2IJ + I + 1]$ $+J^3 + IJ^2 + IJ$	$NK[IJ + I]$ $+N[2J^2 + 2IJ + I + 1]$ $+J^3 + IJ^2 + IJ$	$NK^2 + K^2IJ$ $+KIJ + K^3$
β/\mathbf{B} normalization	$K[3IJ + 1]$	$K[3IJ + 1]$	$K[3IJ + 1]$

in computational complexity arise from the different strategies used to approximate the unconstrained solution, as the cost of implementing model constraints is identical for all algorithms. ALS-gradient is more computationally demanding than ALS-direct, which seems in agreement with the more gradual manner in which the solution is approximated. In this respect, the added computational cost comes as a downside of having a more regular, smoother approach. As far as LLOD is concerned, complexity is shifted from the optimization stage to the estimation of local linear operators. Globally, however, LLOD involves a lower computational complexity than both variants of the ALS algorithm.

3.5 Conclusion

In this chapter, we addressed the extension of constrained blind source separation models to the observation-based decomposition of operators. We formally introduced a non-negative additive mixing model for operators, including a dictionary-based reformulation, and derived associated estimation algorithms. The dictionary-based formulation led to a greater modeling flexibility and possible straightforward extensions considering other constraints, including, for example, sparsity-based priors. Importantly, the proposed models and algorithms have been used for multiple applications successfully presented in a number of national and international publications [62, 165–167, 169], and are also the main subject of a journal article [168].

In the next chapter, we focus on performing numerical experiments to evaluate and compare the model identification and observation reconstruction performance of the proposed algorithms under both ideal and non-ideal settings.

Algorithm evaluation and benchmarking

"True genius resides in the capacity for evaluation of uncertain, hazardous, and conflicting information."

Sir Winston Churchill

4.1	Introduction	65
4.2	Synthetic dataset generation	66
4.3	Estimation performance under ideal settings	67
4.4	Estimation performance with noisy mixing coefficients	72
4.5	Latent class initialization evaluation	78
4.6	Conclusion	82

4.1 Introduction

In this chapter we evaluate the performance of the algorithms introduced in the previous chapter to address the general decomposition model (3.3) under ideal and non-ideal settings. We consider the three algorithms introduced in the previous chapter, namely ALS-direct, ALS-gradient and LLOD.

We study and compare the proposed algorithms in terms of parameter identification and observation reconstruction performance in a variety of experimental settings, including ideal noiseless settings, cases involving a variable number of decomposition modes and configurations considering noisy observations and parameters, in order to gain insight into the robustness and limitations of the proposed algorithms. We further discuss the main contribution of our approach compared with state-of-the-art methods for the analysis and decomposition of operators.

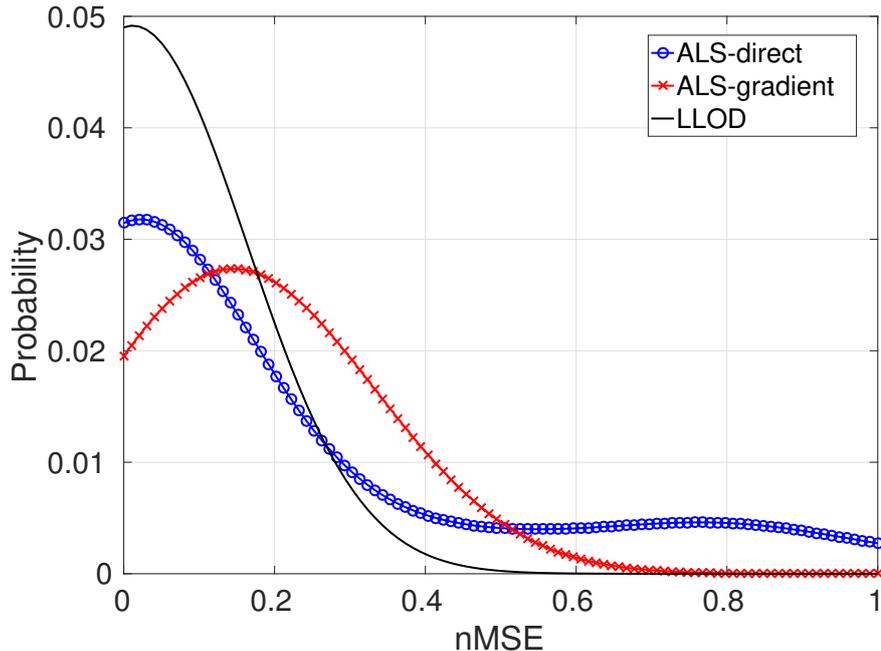


Figure 4.1 – Probability density function (PDF) for the normalized mean squared estimation error (nMSE) for mixing coefficients α_{nk} . Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD). All presented probability distributions were computed using a Gaussian kernel.

We report numerical experiments to evaluate the proposed models and algorithms. We exploit synthetic data to perform a quantitative analysis of the estimation performance and a sensitivity analysis w.r.t. key parameters and modeling hypotheses.

4.2 Synthetic dataset generation

We consider synthetic data $(\mathbf{x}_n, \mathbf{y}_n)$ so that we are provided with ground-truth data. We proceed as follows. Mixing coefficients α_{nk} are simulated by means of a clustering-based approach so that they involve state-dependent variabilities. Elements of linear regression matrices β_k are sampled from a normal distribution $\mathcal{N}(0, 1)$, and regression matrices are subsequently normalized. A cluster-based strategy is used to generate observation pairs, so that N_c cluster centroids \mathbf{x}_c are sampled from a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \sigma_c^2 \mathbf{I})$ and N_x clustered observations are sampled for each cluster from a multivariate normal distribution $\mathcal{N}(\mathbf{x}_c, \sigma_x^2 \mathbf{I})$ centered around each cluster centroid. Mixing coefficients α_{nk} are emulated by sampling the same mixing coefficient for all observation pairs in a given cluster from a uniform distribution $\mathcal{U}_{[0, G_\alpha]}$. Corresponding observations \mathbf{y}_n are then generated by applying model (3.3) to \mathbf{x}_n .

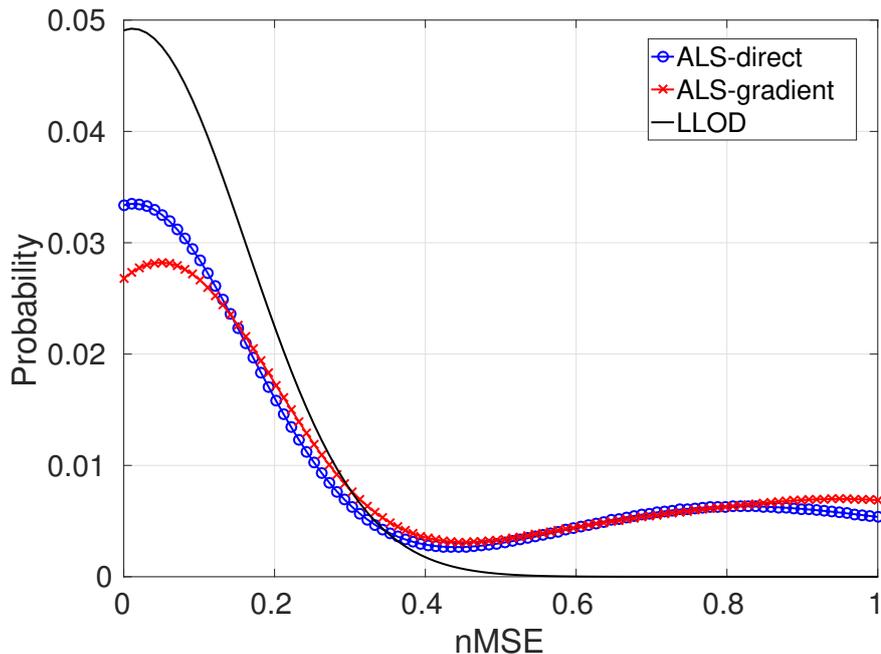


Figure 4.2 – Probability density function (PDF) for the normalized mean squared estimation error (nMSE) for linear modes β_k . Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD). For ALS, two different initialization schemes for α_{nk} are presented: a random initialization and an initialization based on the binary clustering of observations \mathbf{x}_n into K classes. All presented probability distributions were computed using a Gaussian kernel.

4.3 Estimation performance under ideal settings

We first evaluate estimation performance under ideal noise-free conditions, i.e., when no observation noise is present, which means that noise process ω_n in Equation (3.3) represents modeling error only. Moreover, we consider that all observations pairs within the same cluster share exactly the same operator decomposition, in the sense that no parameter noise in either mixing coefficients α_{nk} or modal regression matrices β_k is considered. The minimal inter-cluster distance d_{min} verifies $d_{min} > 6\sigma_x$, which ensures a nearest neighbour search will only select points within the same cluster, such that they truly share the same mixing coefficients. All considered algorithms (ALS-direct, ALS-gradient, LLOD) were applied to a dataset generated considering $I = 30$, $J = 2$, $K = 2$, $N_c = 100$, $N_x = 300$, $\sigma_c^2 = 1$ and $\sigma_x^2 = d_{min}^2/360$. $M = 100$ nearest neighbours are used to estimate model parameters for each observation pair $(\mathbf{x}_n, \mathbf{y}_n)$, with uniform weighting $W_m^n = 1/M, \forall n, m$. The experience is repeated 100 times and results are averaged over all runs to ensure statistical significance.

We consider two different initialization schemes for mixing coefficients α_{nk} . The random initialization involves a random sampling from a uniform distribution $\mathcal{U}_{[0, G_\alpha]}$. The clustering-based initialization relies on an initial clustering of observations \mathbf{x}_n into K clusters and a subse-

Table 4.1 – Error statistics (after convergence) for the different algorithms considered, computed over 100 runs of the algorithms with randomly generated data.

Parameter	Setting	Initialization	Mean	Median	Variance
α_{nk}	ALS-direct	Clustered	7.5371×10^{-1}	1.1134×10^{-2}	2.4298×10^1
		Random	1.5546×10^0	1.9618×10^{-2}	9.2158×10^1
	ALS-gradient	Clustered	1.5744×10^{-1}	1.2882×10^{-1}	9.5170×10^{-3}
		Random	1.5194×10^{-1}	1.2397×10^{-1}	8.8158×10^{-3}
LLOD		1.1983×10^{-2}	9.2061×10^{-3}	1.1242×10^{-4}	
β_k	ALS-direct	Clustered	2.7868×10^{-1}	1.4808×10^{-2}	1.7997×10^{-1}
		Random	2.8436×10^{-1}	9.7638×10^{-3}	2.0602×10^{-1}
	ALS-gradient	Clustered	3.2184×10^{-1}	6.4106×10^{-2}	1.7588×10^{-1}
		Random	2.8090×10^{-1}	5.9336×10^{-2}	1.3902×10^{-1}
LLOD		1.1750×10^{-2}	8.7974×10^{-3}	9.9564×10^{-5}	

quent binary assignment of mixing coefficients α_{nk} according to the corresponding cluster [166]. Given that both initializations yield similar results, we focus, unless explicitly specified, on the clustering-based initialization, which does not require an initial guess of the maximum amplitude G_α of the mixing coefficients. For completeness, both initialization schemes' error statistics are presented in Table 4.1.

Figure 4.1 presents the probability density function (PDF) of the normalized mean squared estimation error (nMSE) for mixing coefficients α_{nk} , defined as $nMSE(\alpha_{nk}, \hat{\alpha}_{nk}) = 1/K \cdot \sum_{k=1}^K \left[\sum_{n=1}^N (\alpha_{nk} - \hat{\alpha}_{nk})^2 / \sum_{n=1}^N (\alpha_{nk})^2 \right]$, with α_{nk} being the real mixing coefficients and $\hat{\alpha}_{nk}$ being the estimated mixing coefficients. Figure 4.2 presents similar results for linear modes β_k . All PDFs were computed from the 100 simulation runs using a non-parametric Gaussian kernel based estimation. The dictionary-based LLOD algorithm yields a better reconstruction performance for both α_{nk} and β_k , with an error PDF presenting higher values around zero and a rapidly decaying tail for higher error levels. By contrast, the two ALS schemes depict similar patterns for the estimation of regression matrices β_k , with a secondary mode of the PDF centered around high nMSE values. These patterns indicate that the ALS algorithms do not converge for a significant fraction of cases. For 11% (resp. 12%) of the simulations, the nMSE is greater than 0.5 for the ALS-direct (resp. ALS-gradient) scheme, whereas it remains at 0% for the LLOD algorithm. For mixing coefficients α_{nk} , ALS-gradient presents a wider, non-zero-centered mode, which reflects a lower parameter identification performance. As far as the ALS-direct is considered, even though it depicts higher probability levels around zero, its PDF presents, nonetheless, a slowly decaying tail, which reflects a higher instability, with high error values (greater than 0.5) for a significant percentage (approximately 20%) of the simulations. Of the considered algorithms, only the LLOD approach displays consistent and stable performance for the identification of both mixing coefficients α_{nk} and linear modes β_k .

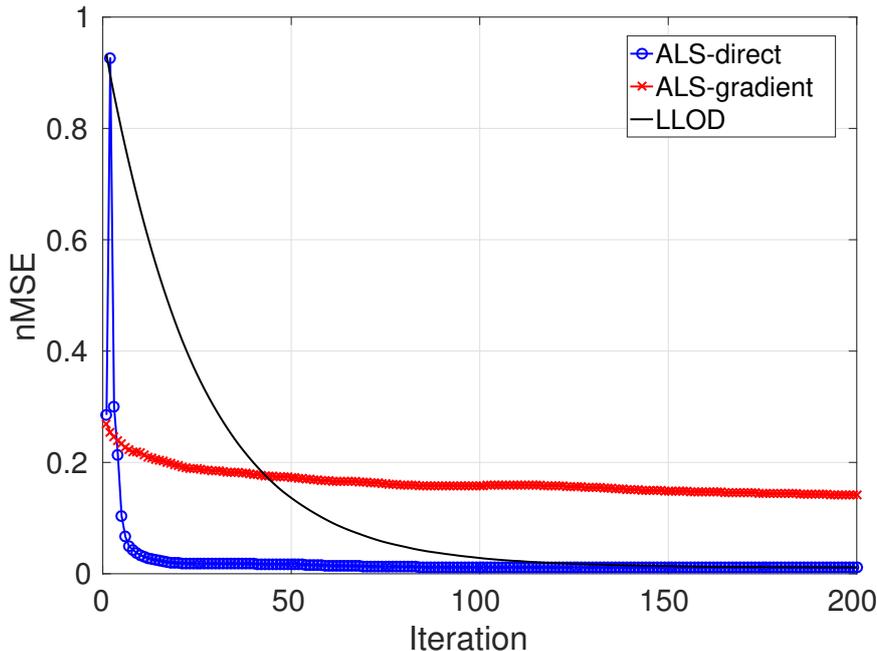


Figure 4.3 – Normalized mean squared estimation error (nMSE) median evolution for mixing coefficients α_{nk} . Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).

Estimation statistics reported in Table 4.1 further support these conclusions. For both the mean, median and variance of the estimation error after convergence (500 iterations), LLOD outperforms the both ALS algorithms by at least one order of magnitude. Most notably, the instability of ALS-direct can be observed in the high variance levels and considerable difference between mean and median values of parameters estimation errors.

Regarding convergence properties, we report in Figures 4.3 and 4.4 the median nMSE (at convergence) as a function of the iteration number for mixing coefficients α_{nk} and linear modes β_k . The LLOD approach presents a much slower and smoother convergence than the two ALS schemes for mixing coefficients α_{nk} , while also converging to a lower nMSE value. Conversely, for linear modes β_k , convergence is significantly slower for the ALS-gradient algorithm, while both the ALS-direct scheme and the LLOD approach present fast convergence towards low nMSE values. Overall, the ALS-direct scheme depicts a fast convergence (about 10 iterations) for both parameters, but we may underline that the convergence towards the actual parameters is not guaranteed as shown above. Regarding the LLOD approach, convergence is reached in about 10 iterations for linear modes β_k and 100 iterations for mixing coefficients α_{nk} .

A complementary experiment addresses the evaluation of estimation performance with respect to the number of classes K . We vary the number of classes $K = 2, \dots, 10$ and generate observations using the same procedure as previously. The experience is repeated 100 times for

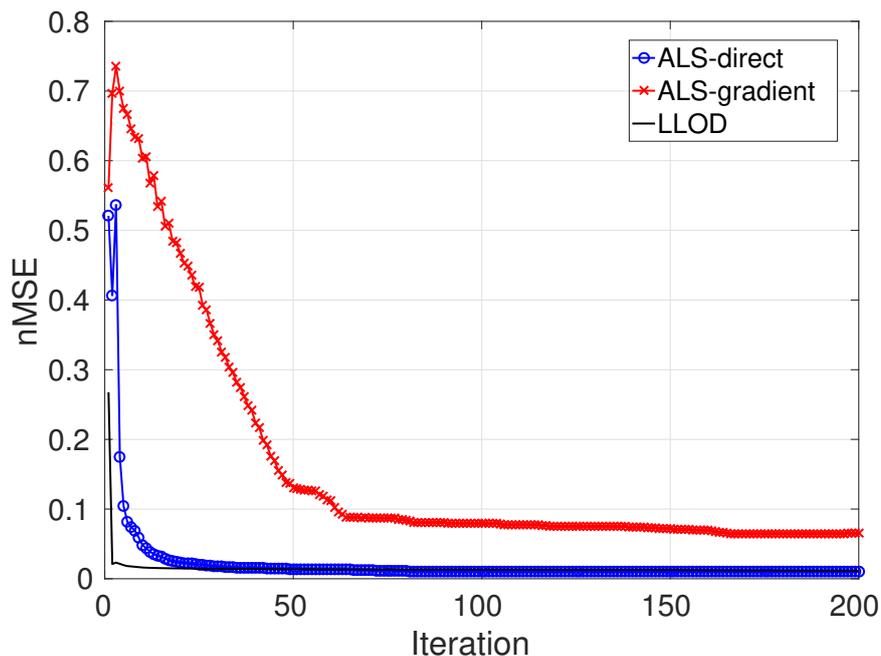


Figure 4.4 – Normalized mean squared estimation error (nMSE) median evolution for linear modes β_k . Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).

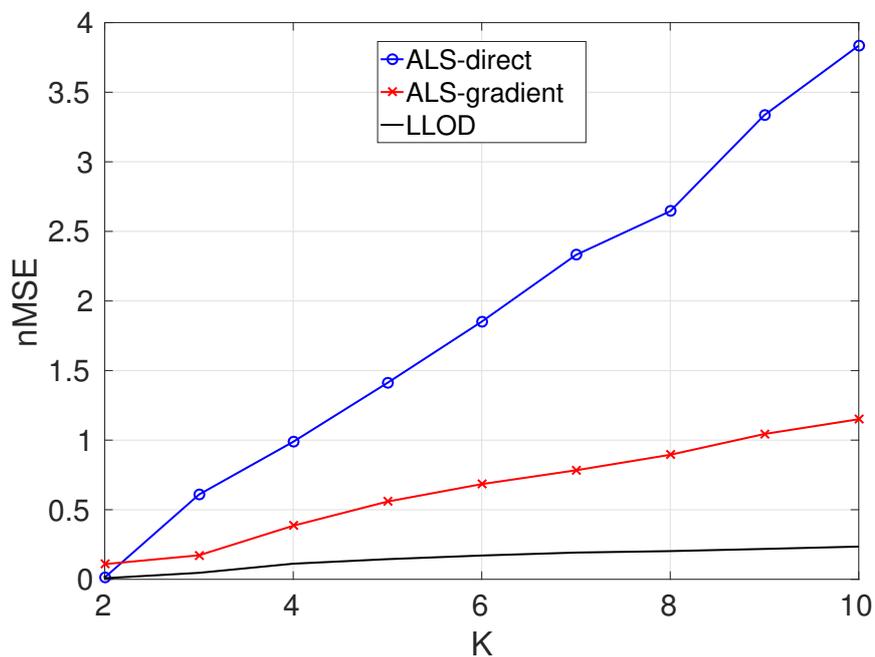


Figure 4.5 – Normalized mean squared estimation error (nMSE) final median value (at convergence) for mixing coefficients α_{nk} as a function of the number of classes K considered. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).

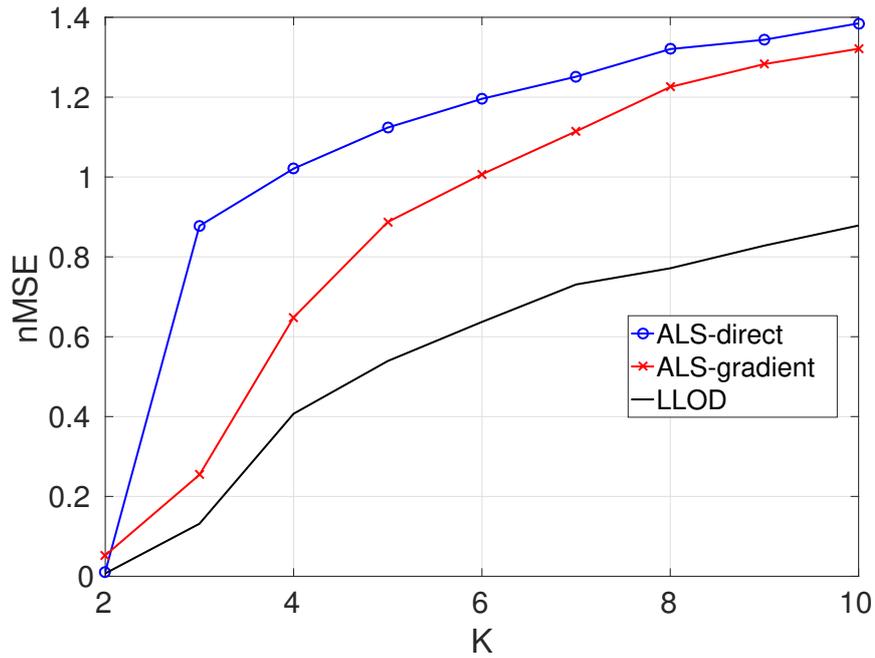


Figure 4.6 – Normalized mean squared estimation error (nMSE) final median value (at convergence) for linear modes β_k as a function of the number of classes K considered. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).

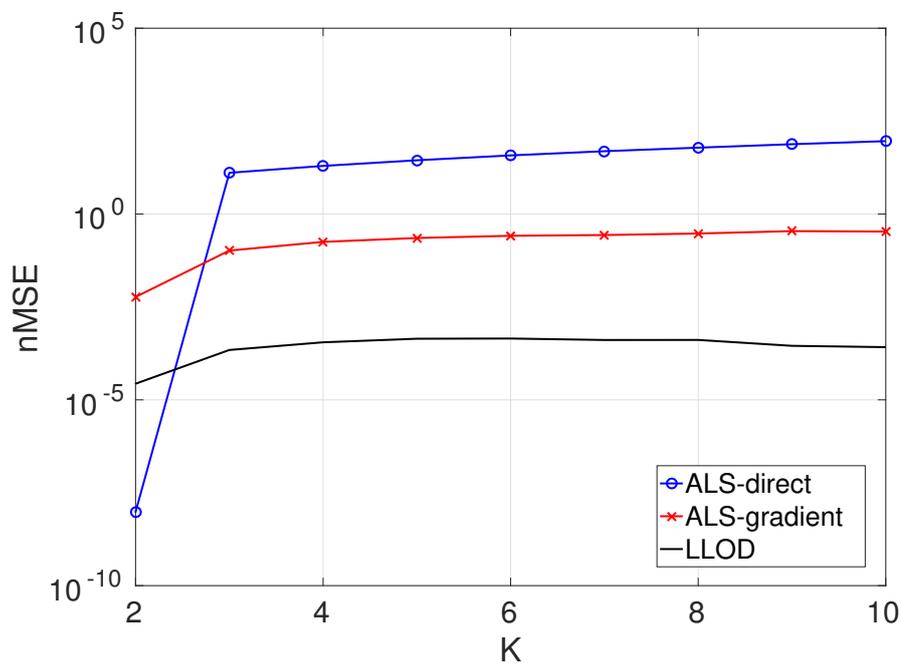


Figure 4.7 – Normalized mean squared \mathbf{y}_n reconstruction error (nMSE) final median value (at convergence) as a function of the number of classes K considered. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).

each number of classes K and results are averaged over all runs for each value of K . Figures 4.5 and 4.6 present the median nMSE (at convergence) for mixing coefficients α_{nk} and linear modes β_k as a function of the number of classes K , for the different algorithms considered. We also depict the median nMSE (at convergence) for the reconstruction of variables $\{\mathbf{y}_n\}$ in Figure 4.7. Obtained results show that the LLOD outperforms both variants of the ALS for the recovery of both mixing coefficients α_{nk} and linear modes β_k . As expected, ALS-direct is the least performant algorithm, which can be explained by a greater numerical instability and a higher estimation variance. Specifically, results show that the high performance degradation for $K > 3$ for ALS-direct is related to the existence of rapid oscillations between multiple local minima at each iteration, most probably due to the instabilities brought about by the alternating projections onto the constrained and unconstrained solution spaces. Overall, parameter recovery performance is degraded as K increases, so that we report good parameter recovery performance only for a low number of classes ($K < 4$). Reconstruction performance, on the other hand, is weakly affected by the number of classes K , with low nMSE values for LLOD and rather poor nMSE levels both ALS variants (and particularly ALS-direct). These results relate to the identifiability of the model. This identifiability becomes weaker as the number of classes K increases, since so does the number of parameters to be estimated (given by $K(N + IJ)$), while the quantity of available information to estimate these parameters remains constant (since N , I and J , the number and dimensions of observations \mathbf{x}_n and \mathbf{y}_n , remain unchanged).

4.4 Estimation performance with noisy mixing coefficients

We further evaluate the robustness of the proposed algorithms in the case of noisy mixing coefficients, that is to say that for a given observation index n in Equation (3.5) not all auxiliary observations pairs with index m and non-zero coefficients W_m^n may share exactly the same mixing coefficients α_{nk} . The considered experiment proceeds as follows. A random Gaussian noise is added to the initially cluster-specific mixing coefficients α_{nk} in order to obtain observation-specific coefficients, which will no longer be shared by observations in the same cluster. To prevent the existence of negative mixing coefficients due to the addition of Gaussian noise, the initial cluster-specific mixing coefficients are now sampled from a uniform distribution $\mathcal{U}_{[100G_\alpha, 101G_\alpha]}$.

This simulation setting implies that the $M = 100$ nearest-neighbors of sample n involve varying mixing coefficients, such that model (3.3) does not hold exactly and is only an approximation. In this respect, parameter similarity for close observations will now depend on the noise variance and, thus, on the signal-to-noise ratio (SNR) between the generated mixing coefficients α_{nk} and the added noise. As noise variance increases (SNR decreases), the relationship between observation similarity (in terms of distance and of belonging to a given cluster) and parameter similarity becomes weaker. The minimal inter-cluster distance d_{min} verifies $d_{min} > 6\sigma_x$, which ensures a nearest neighbour search will only select points within the same cluster.

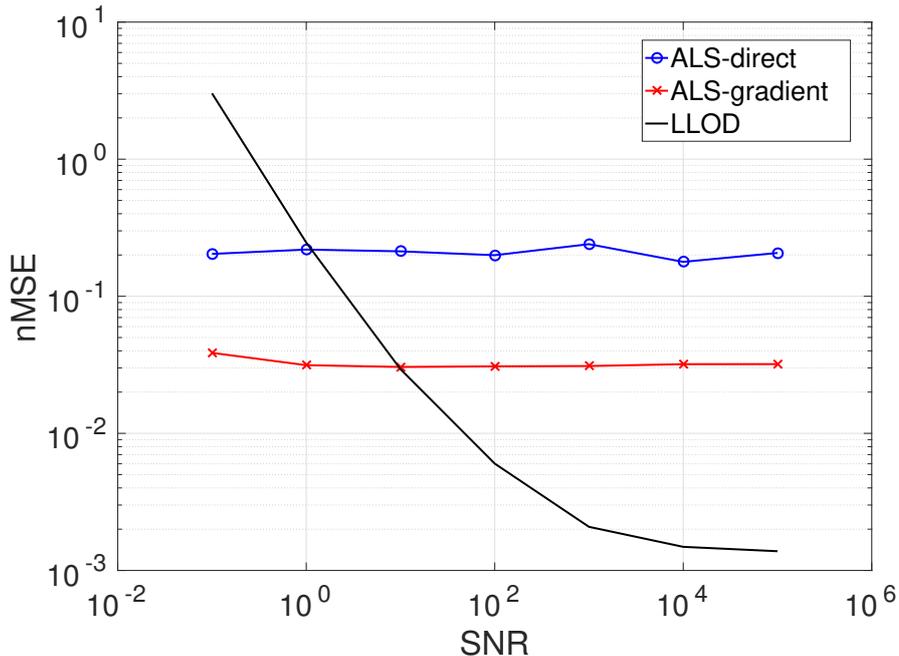


Figure 4.8 – Normalized mean squared estimation error (nMSE) final median value (at convergence) for mixing coefficients α_{nk} as a function of mixing coefficient signal-to-noise ratio (SNR) when Gaussian noise is added to cluster-specific mixing coefficients. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).

All considered algorithms (ALS-direct, ALS-gradient, LLOD) were applied to a dataset generated considering $I = 30$, $J = 2$, $K = 2$, $N_c = 100$, $N_x = 300$, $\sigma_c^2 = 1$ and $\sigma_x^2 = d_{min}^2/360$. Again, no observation noise is present, so that noise process ω_n in Equation (3.3) represents modeling error only. $M = 100$ nearest neighbours are used to estimate model parameters for each observation pair $(\mathbf{x}_n, \mathbf{y}_n)$, with uniform weighting $W_m^n = 1/M, \forall n, m$. The experience is repeated 100 times and results are averaged over all runs, to ensure statistical significance. Moreover, these 100 simulation runs are repeated considering varying SNR levels: $SNR = \{10^{-1}, 10^0, \dots, 10^5\}$.

Figures 4.8 and 4.9 present the median nMSE (at convergence) for mixing coefficients α_{nk} and linear modes β_k as a function of the SNR of mixing coefficients. We also depict the median nMSE (at convergence) for the reconstruction of variables $\{\mathbf{y}_n\}$ in Figure 4.10. The LLOD approach is clearly the most sensitive to noisy mixing coefficients. Linear modes β_k are highly affected even for low noise levels. For SNR values below 10^4 , the approach converges to linear modes significantly different from the ground-truth ones. By contrast, the retrieval of mixing coefficients α_{nk} as well as reconstruction performance for variables $\{\mathbf{y}_n\}$ from observations $\{\mathbf{x}_n\}$ seems consistent for SNR levels greater than 10^1 . As far as ALS algorithms are concerned, their performance is weakly affected by noisy mixing coefficients α_{nk} as illustrated by Figures 4.8 and 4.9. They however lead to poor estimation performance for the identification of linear modes β_k even for SNR values greater than 10^4 . Overall, these experiments suggest identifiability issues

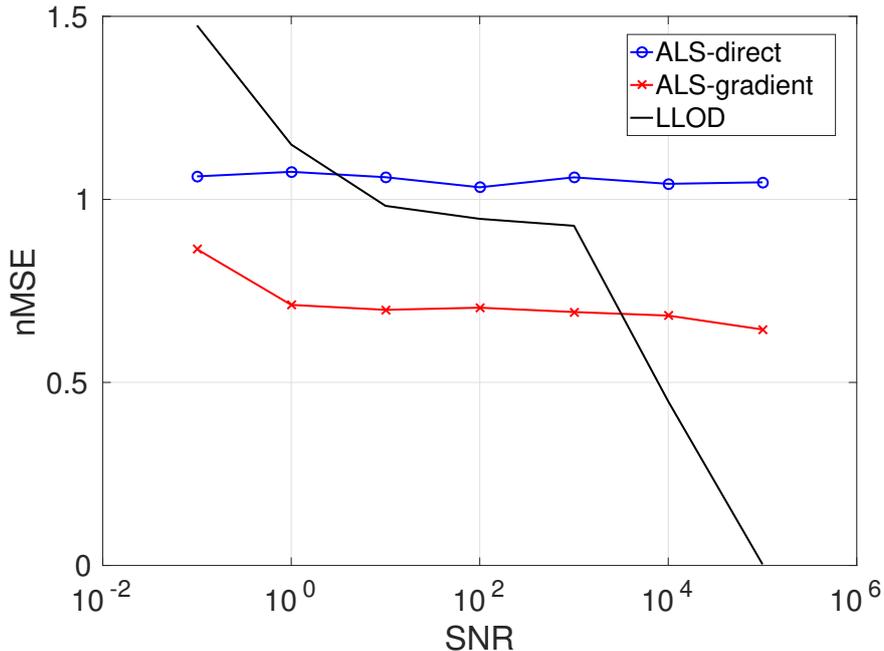


Figure 4.9 – Normalized mean squared estimation error (nMSE) final median value (at convergence) for linear modes β_k as a function of mixing coefficient signal-to-noise ratio (SNR) when Gaussian noise is added to cluster-specific mixing coefficients. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).

for model (3.3) for noisy mixing coefficients even at high SNR values. It seems that there may exist a set of estimated linear regression matrices $\hat{\beta}_k$, different from the true modal regression matrices β_k , that lead to low reconstruction errors (typically nMSE values below 0.01). This implies that the proposed algorithms will be suitable for reconstruction applications, but will also suffer from non-unique solutions for the identification of regression modes $\{\beta_k\}$. Nonetheless, it is worth noting that the non-uniqueness of the solution will not necessarily prevent the algorithms to be considered for identification/segmentation applications using mixing coefficients $\{\alpha_{nk}\}$ as illustrated in [165, 166]. From a computational point of view, one may investigate additional constraints or priors onto mixing coefficients α_{nk} and/or linear modes β_k to overcome such identifiability issues.

We further evaluate the extent to which we may account for other noise configurations, especially when neighbours in the observation space may not share similar mixing patterns. To study such situations, we simulate possibly overlapping clusters. As such, neighboring observation pairs $(\mathbf{x}_m, \mathbf{y}_m)$ and $(\mathbf{x}_n, \mathbf{y}_n)$, which are associated with non-null weighing coefficients W_m^n , may belong to different clusters and have, hence, different mean mixing coefficients α_{mk} and α_{nk} . Numerically, we proceed as follows to simulate such datasets. Initial cluster centroids are sampled from a multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \sigma_c^2)$. To ensure initial cluster separation, an additional acceptance/rejection sampling strategy is used to reject all cluster centroids

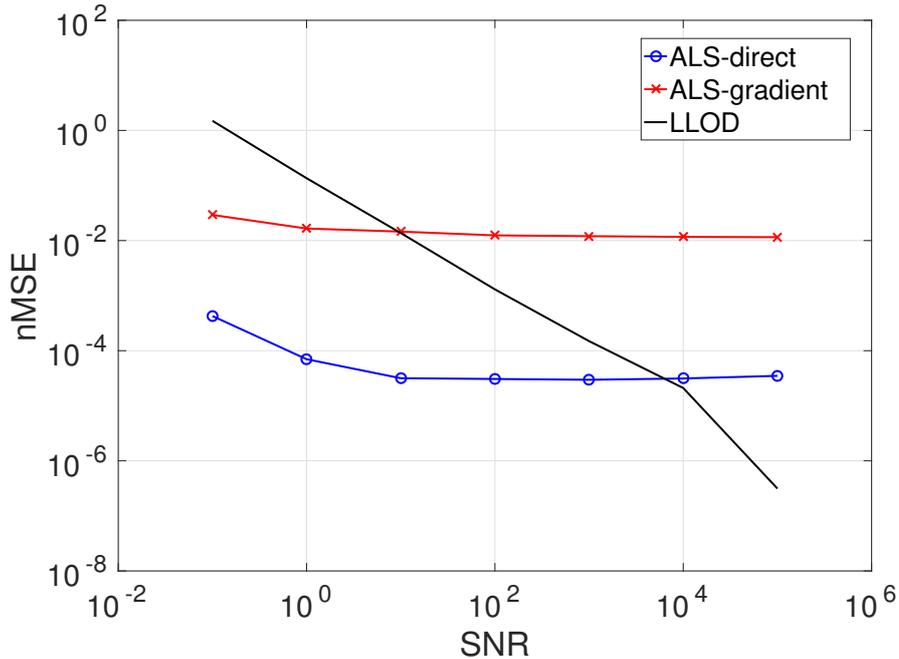


Figure 4.10 – Normalized mean squared \mathbf{y}_n reconstruction error (nMSE) final median value (at convergence) as a function of mixing coefficient signal-to-noise ratio (SNR) when Gaussian noise is added to cluster-specific mixing coefficients. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).

that are too close to other centroids, according to a minimal distance d_{min} . Given the Gaussian nature of the centroid sampling distribution, the distance between cluster centroids will follow a Rayleigh distribution with scale parameter σ_c . Taking this into account, the minimal distance is chosen as $d_{min} = \sigma_c/e$, which ensures a relatively uniform spatial distribution of cluster centroids. For each cluster, we sample observation data $\{\mathbf{x}_n\}$ from Gaussian distributions $\mathcal{N}(\mathbf{x}_c, \sigma_x^2)$ with a standard deviation σ_x ranging from $\frac{1}{100}d_{min}$ to $\frac{30}{100}d_{min}$. For a standard deviation of $\frac{1}{100}d_{min}$, the simulation leads to non-overlapping clusters, whereas overlapping starts to occur from standard deviation values of $\frac{1}{6}d_{min}$ and above. We then evaluate estimation performance as a function of parameter $\lambda = 6\frac{\sigma_x}{d_{min}}$. Figures 4.11 and 4.12 present the median nMSE (at convergence) for mixing coefficients α_{nk} and linear modes β_k as a function of parameter $\lambda = 6\frac{\sigma_x}{d_{min}}$. Obtained results indicate, most notably, that ALS-gradient seems to be unable to correctly recover mixing coefficients α_{nk} as soon as clusters are close enough so that observations from neighbouring clusters start to intervene in the estimation of model parameters, which occurs at around $\sigma_x/d_{min} = 1/6$, i.e. when $\lambda = 6\sigma_x/d_{min} = 1$. ALS-direct, on the other hand, seems more robust to cluster overlapping, with a slightly increasing nMSE as clusters merge. Moreover, LLOD seems to perform worst in the intermediate variance ranges, where parameters for observations near the cluster edge will be computed using wrongly selected neighbours from nearby clusters, while parameters for observations closer to the centroid will be estimated correctly from observations selected from the same cluster. Such behaviour can be also observed

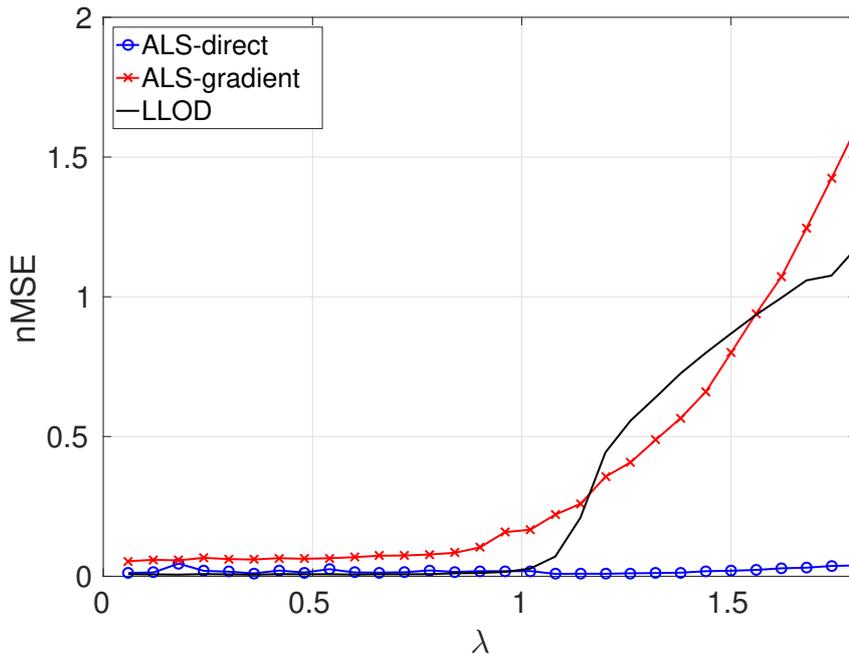


Figure 4.11 – Normalized mean squared estimation error (nMSE) final median value (at convergence) for mixing coefficients α_{nk} as a function of the ratio between cluster standard deviation σ_x and minimal distance d_{min} (parameter $\lambda = 6 \frac{\sigma_x}{d_{min}}$), for initially non-overlapping clusters. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).

for the estimation of linear modes β_k , whereas ALS-based algorithms seems to remain relatively robust to cluster overlap for the recovery of regression matrices β_k .

Taking all previous considerations into account, it seems clear that ALS-direct should be used when cluster overlap may exist or when doubts may arise over how many auxiliary observations should be used and whether the chosen number of auxiliary observations may lead to the incorrect selection of nearest neighbours from nearby clusters. LLOD, on the other hand, should be used for cases where cluster overlap and the correct selection of the number of neighbours M are not an issue, since in such cases it will allow for a better model identification performance, both in terms of mixing coefficients α_{nk} and linear modes β_k .

We specifically investigate robust estimation schemes to improve the performance of LLOD w.r.t. such overlap patterns. When cluster overlapping occurs, local linear operators Θ_n for points near the clusters' edge are computed using observations from both the current and neighbouring clusters. When compared to local linear operators computed for observations closer to the cluster centroid (which are estimated using only observations from the current cluster), the later local linear operators tend to involve considerably larger values and will thus dominate the dictionary-based decomposition (Equation (3.36)). To tackle this problem, two different strategies are explored. The first strategy comes to compute the mean m_{Φ} and standard de-

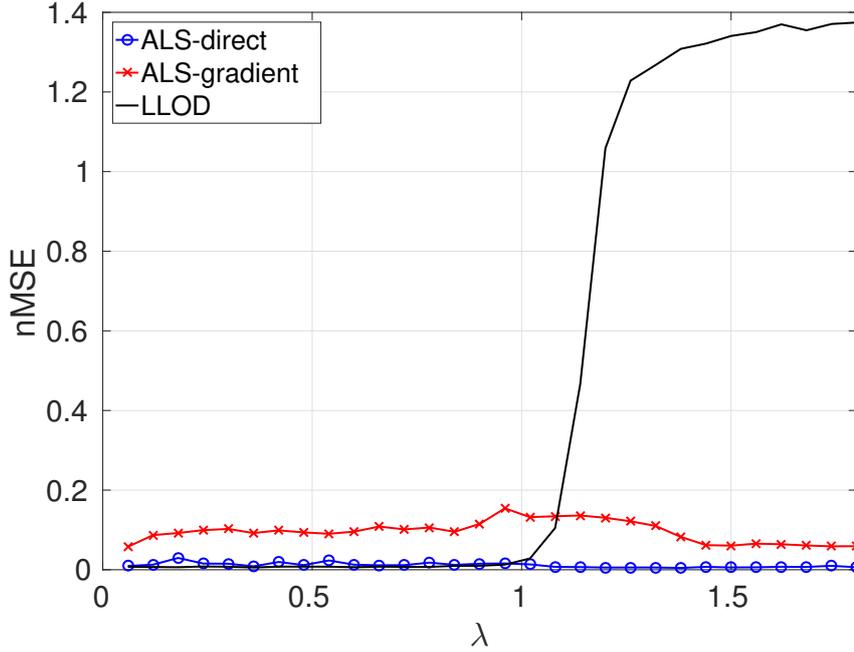


Figure 4.12 – Normalized mean squared estimation error (nMSE) final median value (at convergence) for linear modes β_k as a function of the ratio between cluster standard deviation σ_x and minimal distance d_{min} (parameter $\lambda = 6 \frac{\sigma_x}{d_{min}}$), for initially non-overlapping clusters. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).

viation σ_{Φ} of mean values of estimated local linear operators and filter all observations whose associated local linear operator Θ_n mean value deviates from m_{Φ} , with cutoff values $m_{\Phi} \pm f_c$, where $f_c = n\sigma_{\Phi}$, for $n = \{1, 2, 3\}$. The second strategy involves the robust estimation of local linear operators Θ_n using an iterative re-weighted least squares approach (considering $i = 25$ iterations) [29]. Figures 4.13 and 4.14 present the median nMSE (at convergence) for mixing coefficients α_{nk} and linear modes β_k as a function of parameter $\lambda = 6 \frac{\sigma_x}{d_{min}}$, for the original LLOD and for the two robust variants considered. For the sake of simplicity, only the most performant filtering strategy, namely that considering $f_c = \sigma_{\Phi}$, is depicted. Reported results suggest that both approaches increase the robustness of LLOD, with best results obtained with the filtering scheme with the lowest cutoff value, closely followed by the robust regression approach, which has the additional advantage of not discarding any observations. These approaches consistently improve the working range of LLOD, which we define as the range of values for λ in which $nMSE < 0.1$. The working range of the original LLOD is $\lambda \in [0, \sim 1.08]$, which corresponds to a maximum overlap (in term of percentage of overlapping points between two clusters) of 0.55%, while the working range of the robust LLOD variants is $\lambda \in [0, \sim 1.26]$, which corresponds to a maximum overlap of 1.73%.

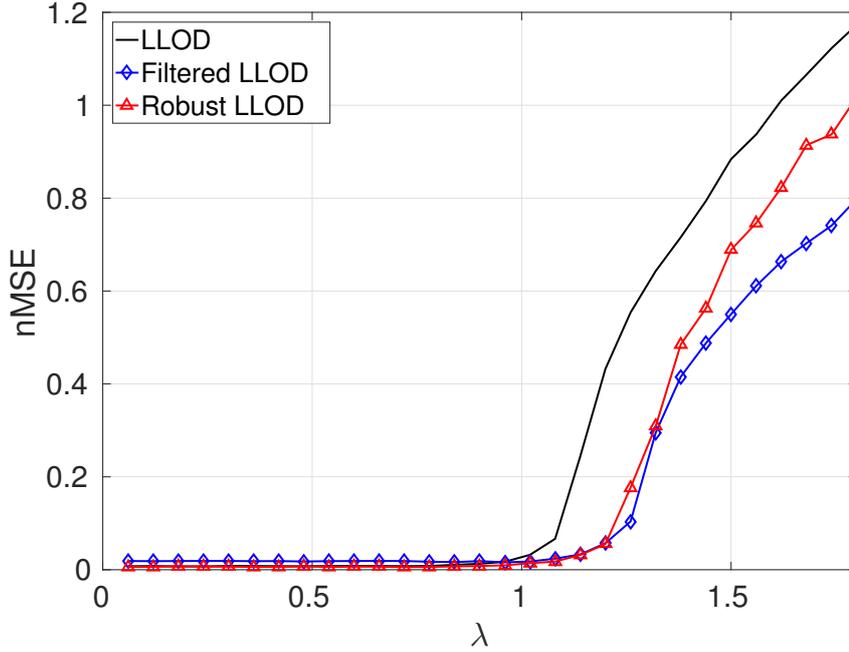


Figure 4.13 – Normalized mean squared estimation error (nMSE) final median value (at convergence) for mixing coefficients α_{nk} as a function of the ratio between cluster standard deviation σ_x and minimal distance d_{min} (parameter $\lambda = 6 \frac{\sigma_x}{d_{min}}$), for initially non-overlapping clusters. Results presented for the original dictionary-based local linear operator decomposition (LLOD) and for two robust variants, namely Filtered LLOD, a filtering of local linear operators with mean value deviation higher than $f_c = \sigma_{\Phi}$ (where σ_{Φ}^2 is the local linear operator mean value variance), and Robust LLOD, which involves the iterative re-weighted least squares estimation of local linear operators Θ_n .

4.5 Latent class initialization evaluation

Following our initial developments, presented in [166], we evaluate the proposed latent class model initialization procedures using synthetic datasets and considering the direct minimization version of ALS (ALS-direct), as it is the simplest parameter characterization solution. We consider the single observation solution formulation (3.7), with an additional focus on comparing both initialization variants of the ALS algorithm (EM and pseudo-EM) for latent class model (3.4). We simulate data according to model (3.3), considering $I = 3$, $J = 2$, $N = 100000$, $K = 2$ and randomly generated mixing coefficients α_{nk}^* and regression matrices β_k^* , with $G_\alpha = 100$. For this experiment, we consider a single cluster of observations $\mathbf{x}_n \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$, with elements of $\boldsymbol{\mu}$ being drawn from a uniform distribution $\mathcal{U}_{[-G_\mu, G_\mu]}$ with $G_\mu = 2$, and σ^2 being drawn from a uniform distribution $\mathcal{U}_{[0, G_\sigma]}$ with $G_\sigma = 20$. For the generation of observations \mathbf{y}_n , $\boldsymbol{\omega}_n$ is sampled from a multivariate centered normal distribution with a diagonal covariance matrix $\boldsymbol{\Sigma}_d$. Each element of the diagonal of covariance matrix $\boldsymbol{\Sigma}_d$ was sampled from a uniform distribution $\mathcal{U}_{[0, G_{\Sigma_d}^2]}$, with $G_{\Sigma_d} = 1$. Given that we focus on the single observation formulation, we generate a different α_{nk} for each observation pair $(\mathbf{x}_n, \mathbf{y}_n)$, so that no parametrizations are shared between any two different observation pairs.

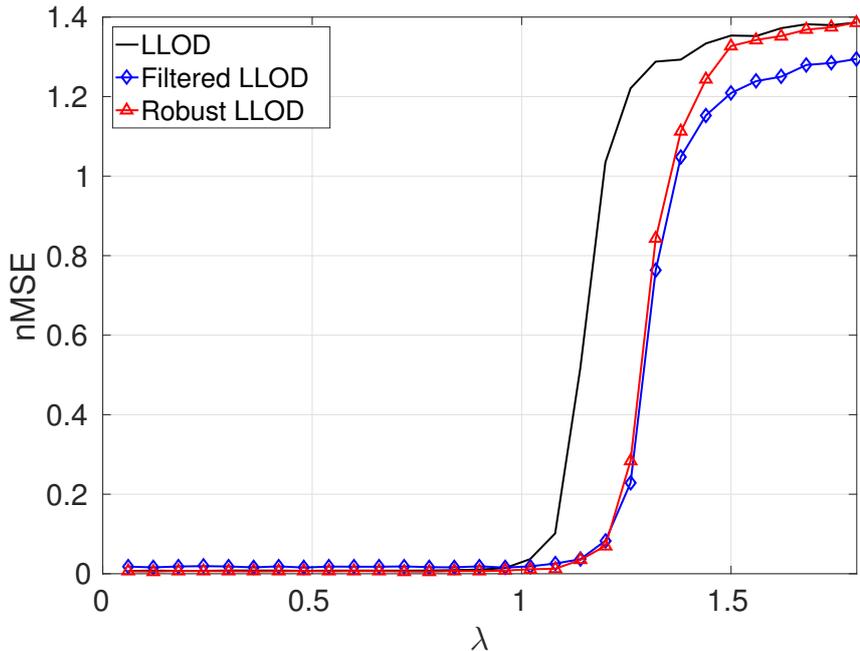


Figure 4.14 – Normalized mean squared estimation error (nMSE) final median value (at convergence) for linear modes β_k as a function of the ratio between cluster standard deviation σ_x and minimal distance d_{min} (parameter $\lambda = 6 \frac{\sigma_x}{d_{min}}$), for initially non-overlapping clusters. Results presented for the original dictionary-based local linear operator decomposition (LLOD) and for two robust variants, namely Filtered LLOD, a filtering of local linear operators with mean value deviation higher than $f_c = \sigma_{\Phi}$ (where σ_{Φ}^2 is the local linear operator mean value variance), and Robust LLOD, which involves the iterative re-weighted least squares estimation of local linear operators Θ_n .

Both versions of the ALS-direct algorithm are then applied to the generated dataset to estimate its mixing coefficients α_{nk} and regression matrices β_k . With a view to characterizing calibration performance, we analyze the evolution of the normalized mean squared error (nMSE) between our parameter estimates $\hat{\beta}_k$ and $\hat{\alpha}_{nk}$ and the real parameter values β_k^* and α_{nk}^* ($e_{\alpha_{nk}}$ and e_{β_k}) for both variants of the algorithm. The obtained results are presented in Figure 4.15. We also compute the final log-likelihood $\mathcal{L}(\alpha_{nk}, \beta_k) = \sum_{n=1}^N \log(p(\mathbf{y}_n | \mathbf{x}_n, \alpha_{nk}, \beta_k))$ for latent class initialization model (3.4), and the final relative mean squared error (rMSE, presented as a percentage) for the reconstruction of variables $\{\mathbf{y}_n\}$. The results are summarized in Table 4.2.

To further assess the quality of the reconstruction, histograms of the relative mean squared reconstruction error (rMSE) for the EM and pseudo-EM variants of the considered algorithm are presented in figure 4.16, with a logarithmic scale for the number of occurrences and line plots instead of bars to allow for an easier comparison of the two considered variants.

From these results, we can conclude that the proposed algorithm accurately estimates model parameters with reasonable error statistics (Table 4.2). The EM variant of the algorithm outperforms the pseudo-EM variant, since it reaches lower error values and higher log-likelihood

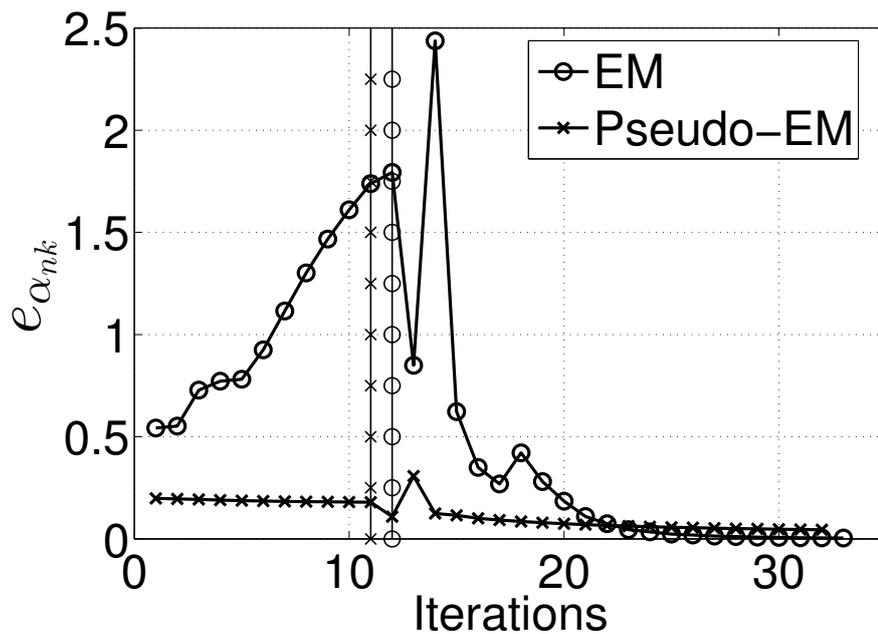
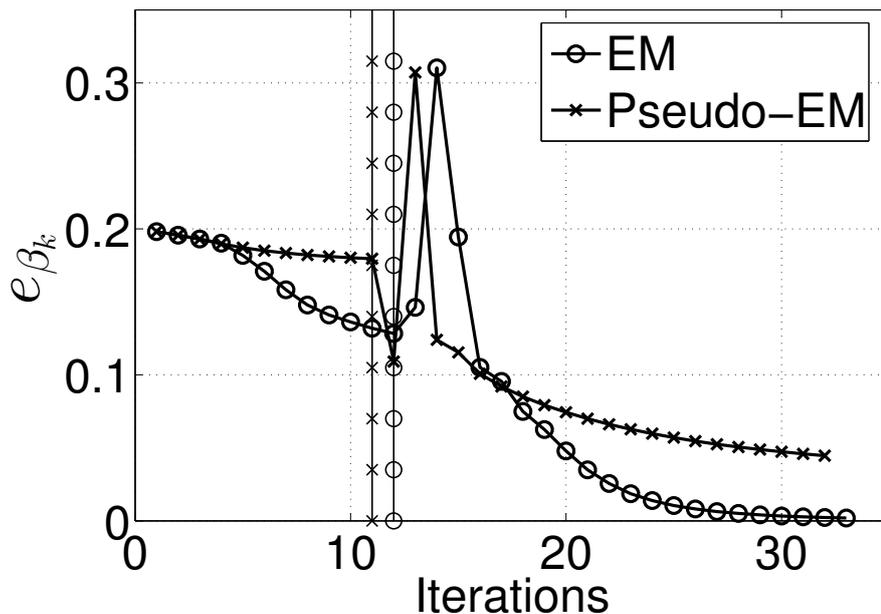
(a) α_{nk} normalized mean squared error (nMSE)(b) β_k normalized mean squared error (nMSE)

Figure 4.15 – Normalized mean squared parameter estimation errors (nMSE) for the single observation version of the ALS-direct algorithm (Algorithm 1) applied to a single-cluster ground-truth synthetic dataset without parameter sharing between observations. Results presented for both the EM and pseudo-EM initialization variants of the considered algorithm. Reported initialization variants are based on model (3.4), a latent class simplification of the proposed decomposition model (3.3). Vertical lines indicate the iteration number at which the change between the EM (pseudo-EM) procedure for the latent class model (3.4) and the ALS procedure for general model (3.3) occurs. Taken from [166]. © 2016 IEEE.

Table 4.2 – Estimation features for the single observation version of the ALS-direct algorithm (Algorithm 1) applied to a single-cluster ground-truth synthetic dataset without parameter sharing between observations. Reported results include mixing coefficients α_{nk} normalized mean squared estimation error ($e_{\alpha_{nk}}$), modal matrices β_k normalized mean squared estimation error (e_{β_k}), final log-likelihood ($\mathcal{L}(\alpha_{nk}, \beta_k)$) for the latent class initialization using model (3.4), and final relative mean squared error (rMSE, presented as a percentage) for the reconstruction of variables $\{\mathbf{y}_n\}$. Results presented for both the EM and pseudo-EM initialization variants of the considered algorithm. Reported initialization variants are based on model (3.4), a latent class simplification of the proposed decomposition model (3.3). Taken from [166]. © 2016 IEEE.

	EM variant	pseudo-EM variant
$e_{\alpha_{nk}}$	0.19 %	8.02 %
e_{β_k}	0.40 %	4.47 %
rMSE	0.07 %	0.48 %
$\mathcal{L}(\alpha_{nk}, \beta_k)$	-8.8305×10^5	-1.4104×10^6

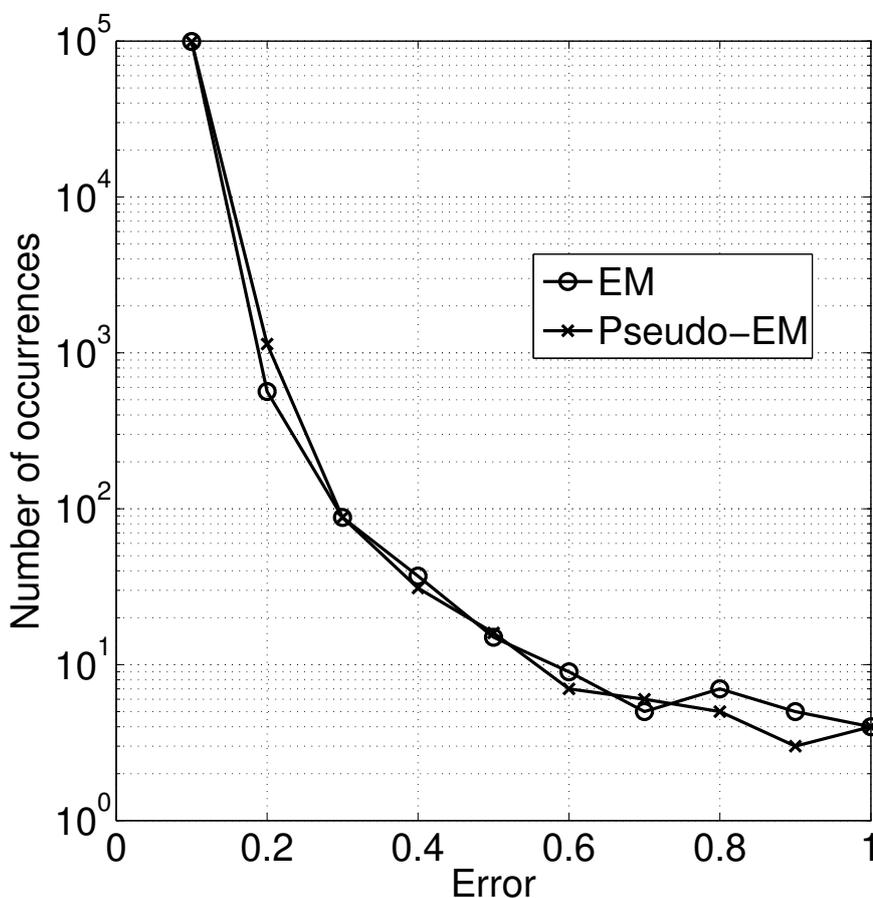


Figure 4.16 – Relative mean squared reconstruction error (rMSE) histograms for the single observation version of the ALS-direct algorithm (Algorithm 1) applied to a single-cluster ground-truth synthetic dataset without parameter sharing between observations. Results presented for both the EM and pseudo-EM initialization variants of the considered algorithm. Reported initialization variants are based on model (3.4), a latent class simplification of the proposed decomposition model (3.3).

values (Table 4.2). This result is explained by the fact that the pseudo-EM variant does not take the covariance structure of estimation residuals into account for the estimation of model

parameters. As far as reconstruction of observations \mathbf{y}_n is concerned, the reported relative mean squared error (rMSE) values (Table 4.2) indicate that both variants have good reconstruction performance, with a greater accuracy for the EM variant.

Furthermore, to analyze the convergence properties of our two latent class initialization variants, we study the evolution of the log-likelihood $\mathcal{L}^{(i)}(\hat{\beta}_k, \hat{\alpha}_{nk})$. Convergence is considered to have been reached once the relative change in log-likelihood $\Delta\mathcal{L}^{(i)}$ is smaller than a given threshold $\lambda_{\mathcal{L}} = 10^{-4}$:

$$\Delta\mathcal{L}^{(i)} = \frac{\mathcal{L}(\hat{\beta}_k^i, \hat{\alpha}_{nk}^i) - \mathcal{L}(\hat{\beta}_k^{i-1}, \hat{\alpha}_{nk}^{i-1})}{\mathcal{L}(\hat{\beta}_k^{i-1}, \hat{\alpha}_{nk}^{i-1})} < \lambda_{\mathcal{L}} = 10^{-4} \quad (4.1)$$

The evolution of the log-likelihood $\mathcal{L}^{(i)}$ and the relative change in log-likelihood $\Delta\mathcal{L}^{(i)}$ (equation (4.1)) is presented for both the EM and pseudo-EM variants of our algorithm in Figure 4.17.

From Figure 4.17, it may be concluded that the pseudo-EM variant seems to reach convergence faster in general, which may be partially explained by the fact that, contrary to the EM variant, it does not suffer from numerical instabilities associated with the inversion of a possibly ill-conditioned or nearly singular covariance matrix.

Taking previous results into account, we conclude that the EM variant of the algorithm should be used when possible, but in cases where numerical instability problems may arise, the pseudo-EM variant could be a viable choice. Choosing between these two algorithm variants implies making a compromise between accuracy and numerical stability and will ultimately depend on the nature of each specific problem and dataset.

4.6 Conclusion

In the previous chapter, we formally introduced a non-negative additive mixing model for operators, including a dictionary-based reformulation, and derived associated estimation algorithms. In this chapter, we performed numerical experiments to evaluate the estimation performance of the proposed algorithms. Regarding the valorization of our contributions, results reported in this chapter were presented in an international conference [166] and are, along with the theoretical developments presented in the previous chapter, the subject matter of a journal article [168].

Overall, the dictionary-based decomposition of local linear operators seems to provide the best performance in terms of model identification, stability and computational complexity under favorable settings. Alternatively, under non-ideal settings, less stable algorithms, such as the ALS-direct, may nonetheless prove useful for model identification and observation reconstruc-

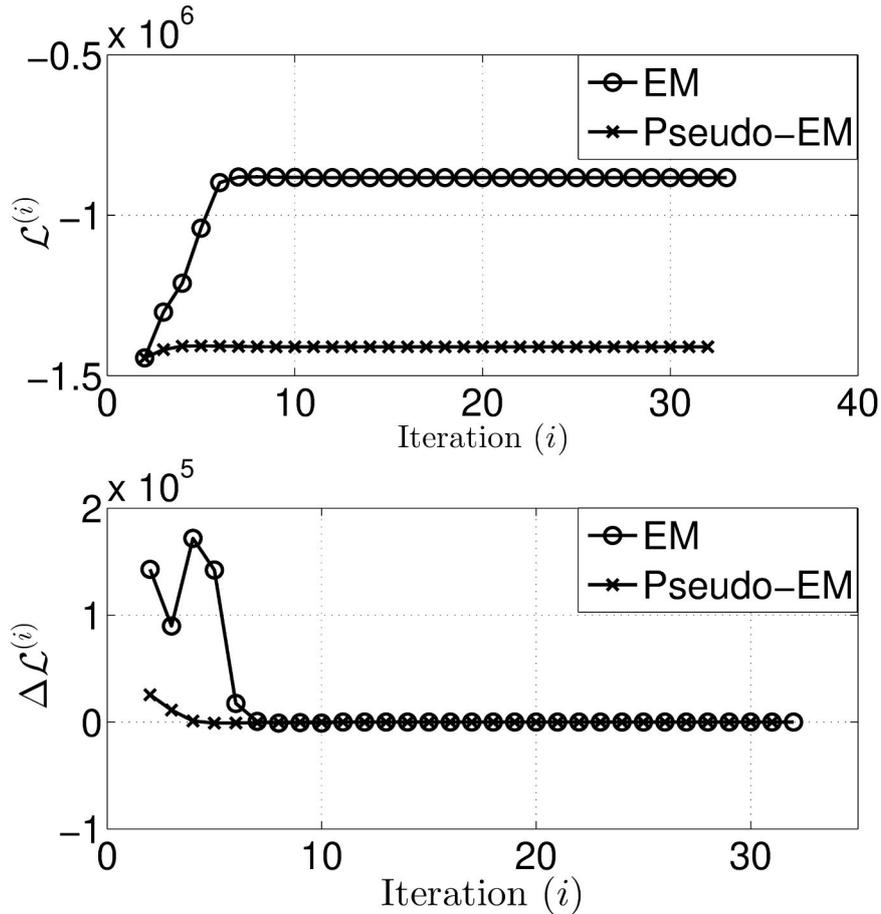


Figure 4.17 – Log-likelihood $\mathcal{L}^{(i)}$ (top) and relative change in log-likelihood $\Delta \mathcal{L}^{(i)}$ (bottom) evolution for the single observation version of the ALS-direct algorithm (Algorithm 1) applied to a single-cluster ground-truth synthetic dataset without parameter sharing between observations. Results presented for both the EM and pseudo-EM initialization variants of the considered algorithm. Reported initialization variants are based on model (3.4), a latent class simplification of the proposed decomposition model (3.3).

tion.

In this respect, reported results suggest the need for additional regularization constraints or priors to tackle identifiability issues for model (3.3) in non-ideal configurations. Indeed, even though our experiments suggest that the proposed model and algorithms have good reconstruction performance in most settings, which makes them suitable for most reconstruction/prediction issues, model identification appears to be considerably sensitive to non-ideal settings, where the parameter sharing hypothesis is relaxed or where the number or selection of auxiliary observations for parameter estimation induces errors. Results also suggest that model identifiability can be improved by introducing robust estimation approaches for local linear operators and/or additional model constraints.

The proposed models and algorithms, however, have been successfully used in both reconstruction/forecasting and segmentation applications [165, 166, 169], which we will explore in the following chapter. These applications stress the relevance of the proposed non-negative decomposition of operators compared with orthogonality-based or latent class settings, which are considered in most previous works [85, 90, 116, 244].

As far as methodological aspects are concerned, reported results suggest that future work should focus on developing strategies for increasing model robustness and algorithm performance, further exploring sparsity and/or other alternative or additional constraints, and identifying and evaluating new possible applications.

Applications

"In theory, there is no difference between theory and practice. In practice, there is."

Richard P. Feynman

5.1	Introduction	85
5.2	Data-driven forecasting of dynamical systems	86
5.3	Multi-tracer synergies for the characterization of upper ocean dynamics	90
5.3.1	Segmentation of upper ocean dynamics from SST/SSS relationships in the Alboran Sea	90
5.3.2	Segmentation of upper ocean dynamics from SST/SSH relationships in the Agulhas region	92
5.4	Conclusion	105

5.1 Introduction

In the previous chapter, we performed an exhaustive performance analysis of the algorithms developed to characterize the proposed non-negative linear decomposition model. Reported results put forward the advantages and limitations of the proposed algorithms, and reveal the need for the introduction of additional constraints and/or robust regression approaches to deal with model identifiability issues under non-ideal conditions. Particularly, proposed approaches seem more suitable for reconstruction applications, as they may involve non-unique solutions in the context of parameter estimation for model characterization and segmentation. As previously reported, however, the proposed model and algorithms have been successfully applied in both segmentation/characterization and reconstruction issues. In this chapter, we present the application of the proposed algorithms to several problems involving both segmentation and reconstruction/prediction of system dynamics. We report the advantages of the proposed formulations with respect to state-of-the-art models for a range of inverse problems in physical oceanography and geosciences.

5.2 Data-driven forecasting of dynamical systems

We further illustrate the interest of the proposed non-negative linear decomposition model for the forecasting of dynamical systems using analog forecasting methods [175], which we will further explore in the second part of this thesis dissertation, most notably in Section 6.3.4. We apply the proposed blind non-negative local linear operator decomposition to Lorenz '96 dynamics, which have been extensively studied in the assimilation and forecasting literature, since they are representative of chaotic geophysical dynamical systems (e.g., the atmosphere).

5.2.0.1 Data and methodology

Essentially, the Lorenz '96 dynamical system is a simple forced dissipative system with quadratic non-linear terms, defined as follows:

$$\frac{\partial \mathbf{s}_i}{\partial t} = (\mathbf{s}_{i+1} - \mathbf{s}_{i-2}) \mathbf{s}_{i-1} - \mathbf{s}_i + F \quad \forall i \in \llbracket 1, L \rrbracket \quad (5.1)$$

where \mathbf{s} is the system's state vector, F is a forcing constant and the boundary conditions $\mathbf{s}_{-1} = \mathbf{s}_{N-1}, \mathbf{s}_0 = \mathbf{s}_N$ and $\mathbf{s}_{N+1} = \mathbf{s}_1$ are assumed to be true. Lorenz introduced this multi-dimensional atmospheric model to study predictability issues in weather forecasting. Despite its simplicity, the system maintains key physical properties of atmospheric systems, such as dissipation, advection, quadratic non-linearity and external forcing. In this simplified model, each variable \mathbf{s}_i is associated to the value of an atmospheric variable (e.g. temperature, humidity, etc) measured along L evenly distributed points around a circle of constant latitude of the earth. Forcing parameter F models external forcing and, for values $F > 8$ the system becomes chaotic. We let the reader refer to [176] for a more detailed description of Lorenz '96 model. An example of a typical Lorenz '96 time series is depicted in Figure 5.3.

Methodologically, we aim at characterizing a linear forecasting operator $\mathbf{A}(\mathbf{s}(t))$ to predict $\mathbf{s}(t + \partial t)$ from $\mathbf{s}(t)$

$$\mathbf{s}(t + \partial t) = \mathbf{A}(\mathbf{s}(t)) \mathbf{s}(t) \quad (5.2)$$

Moreover, the characterization of operator $\mathbf{A}(\mathbf{s}(t))$ involves the dictionary-based non-negative decomposition presented in Section 3.4.2, so that operator $\mathbf{A}(\mathbf{s}(t))$ is decomposed as the non-negative superposition of K linear modes:

$$\mathbf{A}(\mathbf{s}(t)) = \sum_{k=1}^K \alpha_k(\mathbf{s}(t)) \boldsymbol{\beta}_k \quad (5.3)$$

where $\alpha_k(\mathbf{s}(t)) \in \mathbb{R}^+$ are non-negative mixing coefficients quantifying the contribution of linear mode k to the reconstruction of $\mathbf{s}(t + \partial t)$ for a given $\mathbf{s}(t)$. and $\boldsymbol{\beta}_k \in \mathbb{R}^{I \times J}$ is a regression matrix representing mode k .

Following the principle of analog forecasting, forecasting operator $\mathbf{A}(\mathbf{s}(t))$ is to be estimated by means of a data-driven approach exploiting a catalog of simulations of Lorenz' 96 time series, as explained below.

We simulate Lorenz '96 40-dimensional time series with forcing parameter $F = 8$ and time step $\partial t = 0.05$. We build training and test datasets from independent time series corresponding respectively to 2×10^5 and 200 consecutive time steps. Following [161, 175], we consider a locally-linear analog model. It comes to fitting a multivariate linear regression $\Theta_n \in \mathbb{R}^{I \times J}$:

$$\mathbf{y}_n = \Theta_n \mathbf{x}_n \quad (5.4)$$

where, for given time series \mathbf{s} and variable index $l^* \in \llbracket 1, 40 \rrbracket$, variable \mathbf{x}_n is given by a 21-dimensional vector of consecutive variables at time t^*

$$\mathbf{x}_n = \begin{bmatrix} \mathbf{s}(t^*, l^* - 10) \\ \vdots \\ \mathbf{s}(t^*, l) \\ \vdots \\ \mathbf{s}(t^*, l^* + 10) \end{bmatrix} \quad (5.5)$$

and variable \mathbf{y}_n corresponds to the 3-dimensional vector of corresponding consecutive variables at time $t^* + \partial t$

$$\mathbf{y}_n = \begin{bmatrix} \mathbf{s}(t^* + \partial t, l^* - 1) \\ \mathbf{s}(t^* + \partial t, l) \\ \mathbf{s}(t^* + \partial t, l^* + 1) \end{bmatrix} \quad (5.6)$$

Following our proposed local linear operator decomposition model, local linear operators Θ_n are estimated for each observation pair $(\mathbf{y}_n, \mathbf{x}_n)$ under the hypothesis that the nearest observation pairs in the training dataset, referred to as analogs, share the same Θ_n . In agreement with the proposed non-negative local linear operator decomposition model, the estimation of Θ_n resorts, numerically, to a weighted least-square estimate from the dataset formed by the analogs $(\mathbf{y}_m, \mathbf{x}_m)$ of pair $(\mathbf{y}_n, \mathbf{x}_n)$, the weights accounting for relative similarities.

We illustrate here the application of the proposed non-negative decomposition with $K = 4$ modes. Given the dictionary $\{\hat{\beta}\}_k$ learned from the training dataset by means of our proposed local linear operator decomposition algorithm, we state the analog forecasting operators of the test dataset according to model (3.31). It resorts to the estimation of mixing coefficients in model (3.31). More precisely, for the current observation pair $(\mathbf{y}_n^*, \mathbf{x}_n^*)$ in the test dataset, the associated analog forecasting operator involves the non-negative projection of the previously estimated locally-linear operator Θ_n^* , computed for test observation $(\mathbf{y}_n^*, \mathbf{x}_n^*)$ from its M analog pairs $(\mathbf{y}_m^*, \mathbf{x}_m^*)$ in the training dataset, onto the manifold spanned by the estimated modes $\{\hat{\beta}\}_k$.

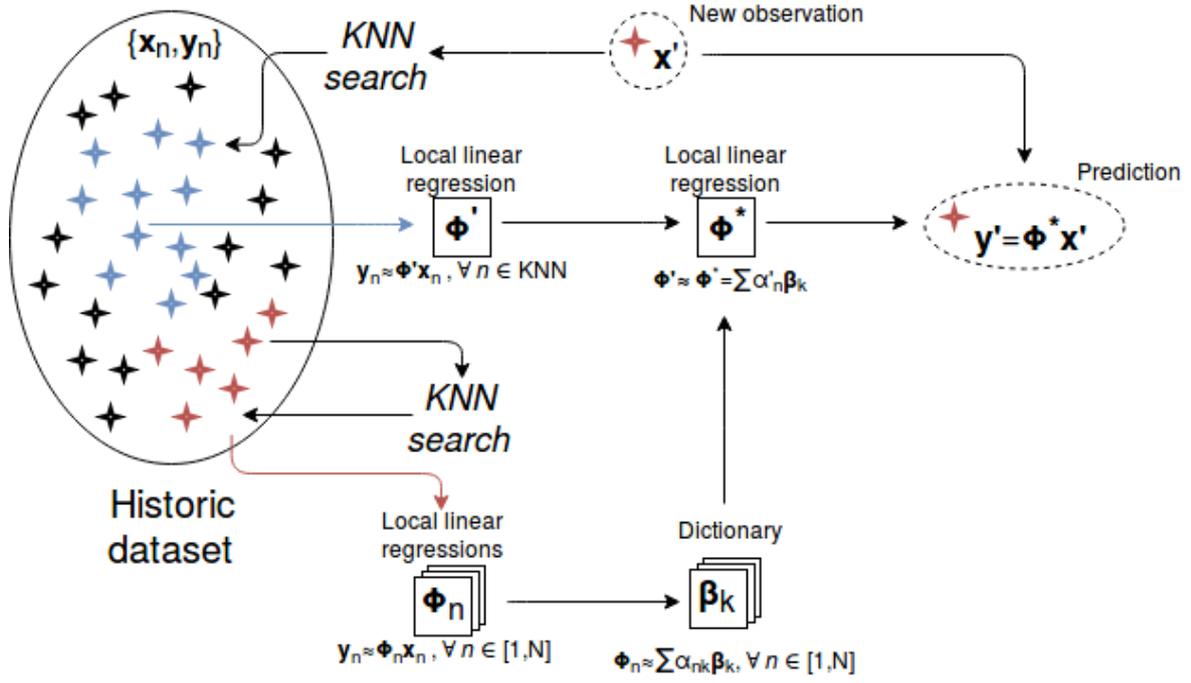


Figure 5.1 – Illustration of the complete dictionary-based local linear analog forecasting operator non-negative decomposition. *Nearest neighbours are searched (in the train dataset) for observations in the train dataset to estimate local linear operators, which are then decomposed to obtain a dictionary of modal matrices β_k . Local linear operators are subsequently estimated for observations in the test dataset (new observations) from nearest neighbours in the train dataset, and then projected onto the manifold spanned by modal matrices β_k under non-negativity constraints. The forecasting for the next time step simply amounts to the application of the projected linear operator to the current state.*

This provides the algorithm with true prediction capabilities, as all local linear operators for observations in the test dataset are estimated from observations in the train dataset only. The forecasting for the next time step simply amounts to the application of the projected linear operator to the current state.

A schema illustrating the complete dictionary-based local linear analog forecasting operator decomposition approach is presented in Figure 5.1

5.2.0.2 Results

Figure 5.2 illustrates the prediction performance for the decomposition of local linear analog forecasting operators, for different parameter settings with a focus on variable index $l^* = 20$ of the Lorenz '96 state. We report the normalized root mean square error (RMSE) of the forecasting as a function of the number of analogs used, for the locally-linear analog forecasting with no decomposition (full line) and the proposed analog forecasting using a non-negative decomposition (dashed line). Three scenarios are simulated: i) a noise-free scenario with a large catalog ($N = 2 \times 10^5$ exemplars) (Figure 5.2a), ii) a noisy scenario with a large catalog and noise variance $\sigma_{noise}^2 = 0.1$ (Figure 5.2b), iii) a noise-free scenario with a small catalog ($N = 2 \times 10^3$ exemplars)

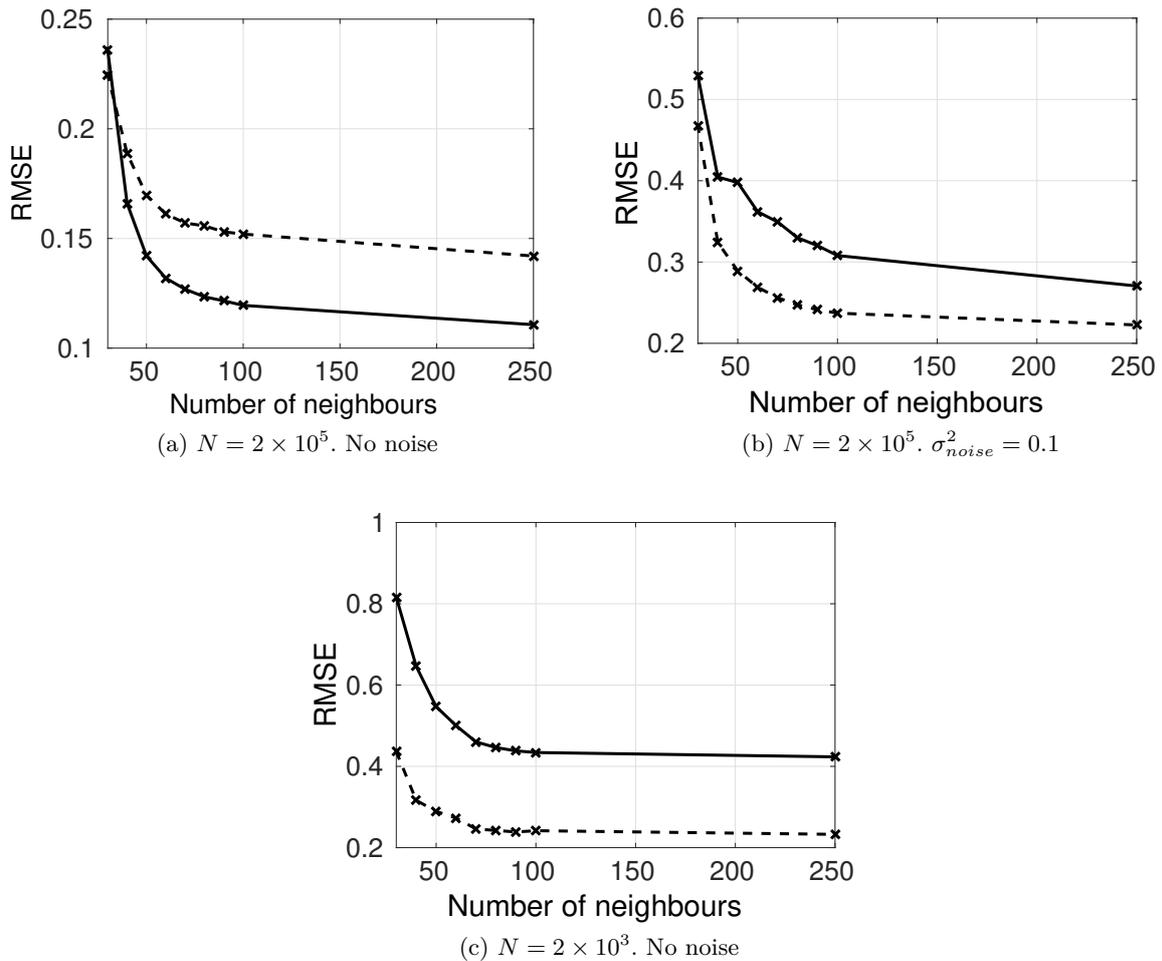


Figure 5.2 – Forecasting performance of local linear analog forecasting operators: we depict the normalized forecasting RMSE vs. the number of analogs used to estimate the locally-linear analog forecasting operator for Lorenz '96 dynamics for: locally-linear operators (*Full line*), the proposed non-negative decomposition of locally-linear operators with $K = 4$ modes (*Dashed line*). We consider three experimental settings: a noise-free scenario with a large catalog of 2×10^5 exemplars, a noisy scenario with a large catalog of 2×10^5 exemplars, a noise-free scenario with a smaller catalog of 2×10^3 exemplars.

(Figure 5.2c). For illustration purposes, and for the sake of completeness, we also depict in Figure 5.3 an example of the real Lorenz '96 time series alongside with the reconstructed Lorenz '96 for the base case involving a large, noise-free catalog and considering $K = 250$ neighbours.

Reported results indicate that the non-negative decomposition of local operators in an analog-based prediction scheme clearly reduces forecasting errors when the analogs are sampled from a noisy or reduced catalog. This is in agreement with the fact that this decomposition can be seen as a projection of the original operator into a lower-dimensional space, which makes the identification of the model feasible for small datasets and improves robustness to noise.

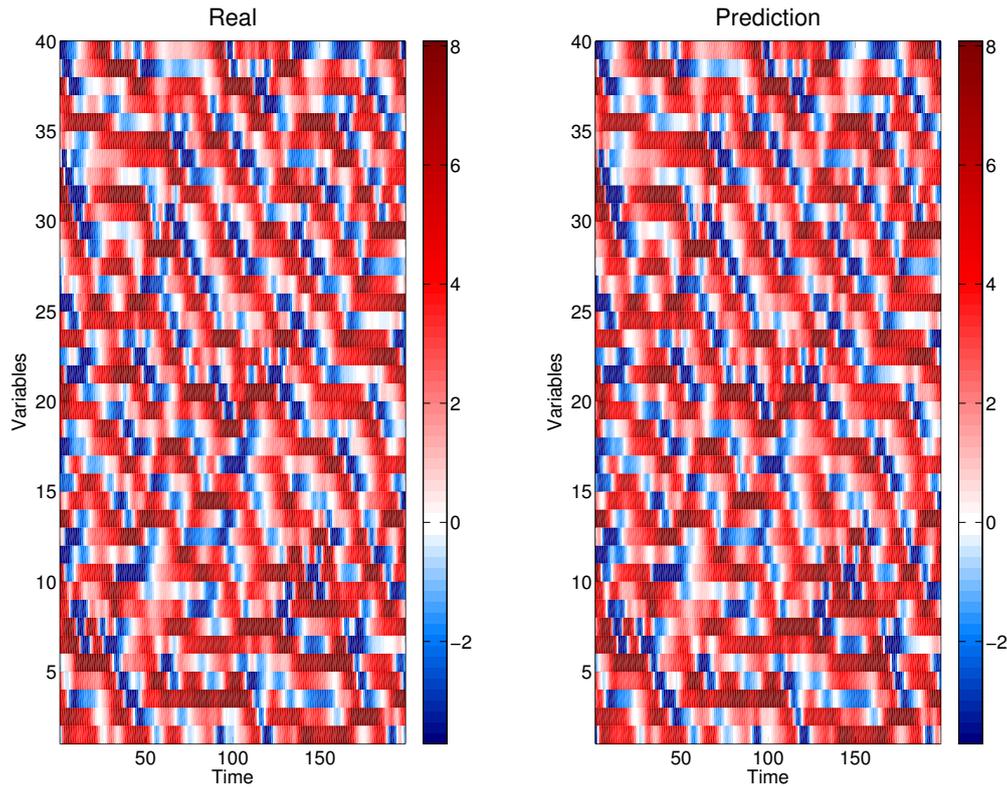


Figure 5.3 – Illustration of the reconstruction of Lorenz '96 dynamics using the proposed dictionary-based local linear analog forecasting operator non-negative decomposition. *Real Lorenz '96 time series alongside with the reconstructed Lorenz '96 for the base case involving a large, noise-free catalog and considering $K = 250$ neighbours.*

5.3 Multi-tracer synergies for the characterization of upper ocean dynamics

5.3.1 Segmentation of upper ocean dynamics from SST/SSS relationships in the Alboran Sea

We illustrate a second application of the proposed non-negative decomposition to the characterization of upper ocean dynamics from the synergy exhibited by different sea surface geophysical fields, namely sea surface temperature (SST) and sea surface salinity (SSS). As illustrated in Figure 5.5a, we analyze the relationships between SST and SSS in the Alboran Sea ($35^{\circ}N - 38^{\circ}N, 0^{\circ}W - 5^{\circ}W$). This region involves strong seasonal patterns associated with the intake of cold Atlantic water through the Gibraltar strait, which strongly affects the SST signature in the Alboran Sea and results in a shift from positive to negative correlations between SST and SSS fields, as can be observed in the time series of the SSS-SST correlation depicted in Figure 5.4. We expect the proposed non-negative decomposition to capture this seasonal patterns.

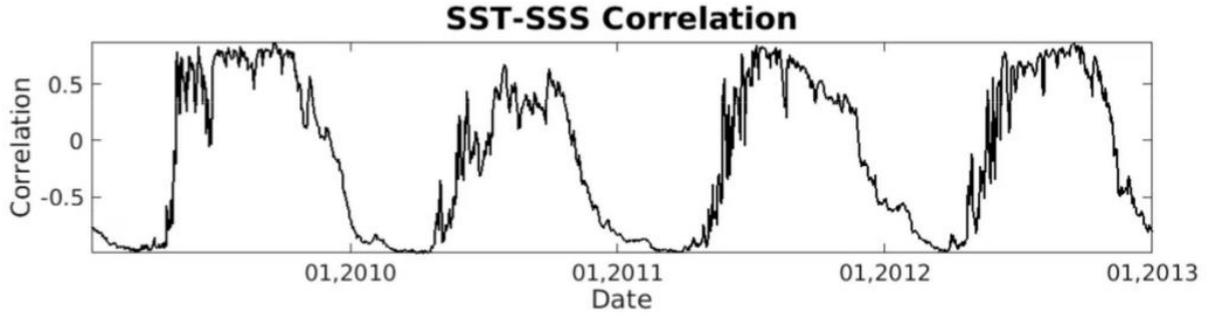


Figure 5.4 – Correlation time series of the considered ground-truth sea surface salinity (SSS) and sea surface temperature (SST) fields in the Alboran Sea.

5.3.1.1 Data and methodology

For the reported experiments, we exploit $1/16^\circ$ operational Regional Ocean Modeling System (ROMS) simulations of the Western Mediterranean Operational Forecasting System (WMOP) numerical model from SOCIB [124, 125] from 2009 to 2012. WMOP is a regional reparametrization of the ROMS general ocean circulation model, nested in the Mediterranean Ocean Forecasting System (MFS), focused on producing high-resolution realistic ocean dynamics simulation in the Mediterranean Sea.

Following our dictionary-based local linear operator decomposition formulation (3.31), we analyze the relationships between daily SST and SSS images (Figure 5.5a) using a convolutional model for 3×3 image patches, which comes to considering SSS pixel values can be approximated daily by a single shared linear regression $\Theta(t_n)$ on values from the corresponding 3×3 SST patches, under the hypothesis that all daily patches share the same linear relationship:

$$SSS(t_n, p) = \Theta(t_n) \text{vec}(\mathcal{P}_{SST}(t_n, p)), \forall p \quad (5.7)$$

where $SSS(t_n, p)$ is the SSS pixel value at location p for day t_n , $\mathcal{P}_{SST}(t_n, p)$ is the corresponding 3×3 SST patch centered at location p for day t_n , $\text{vec}(\cdot)$ is the matrix vectorization operator and $\Theta(t_n)$ is the local linear operator relating daily SSS pixel values to the corresponding SST patch values.

The estimation of local linear regressions $\Theta(t_n)$ follows the proposed local linear operator decomposition model (3.31). Particularly, the choice of auxiliary observation pairs $(\mathbf{x}_m, \mathbf{y}_m)$ in Equation (3.27) relies here on a time domain similarity criterion for neighbour selection, in the sense that all patches extracted from the same daily SST-SSS image pair at day t_n are considered as auxiliary observations and used for the estimation of local linear operators.

Daily estimated linear operators are subsequently decomposed into $K = 2$ modes using the proposed non-negative local linear operator decomposition (3.31):

$$\Theta(t_n) = \sum_{k=1}^K \alpha_k(t_n) \beta_k \quad (5.8)$$

to obtain modal regression matrices β_k and non-negative mixing coefficients $\alpha_k(t_n)$. Given the learned dictionary, mixing coefficients are then re-estimated directly from SST-SSS observations.

Importantly, it should be noted that, given the time domain locality criterion considered, the proposed decomposition comes to the analysis and characterization of the temporal variability of the SST-SSS relationship.

5.3.1.2 Results

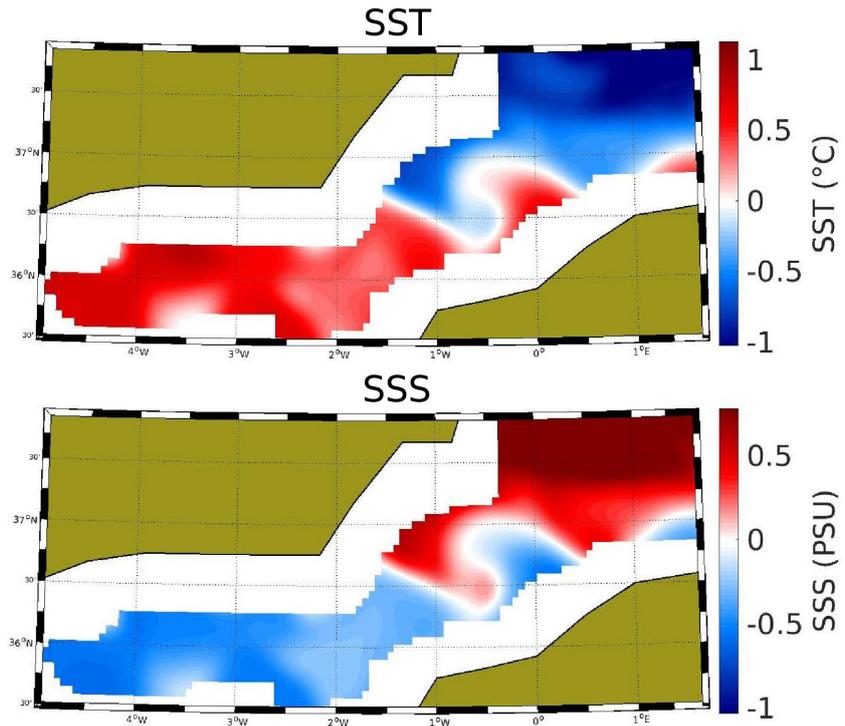
Figure 5.5b presents the SSS fields predicted by each mode. The first mode clearly captures an inversion of the SST, while the second mode captures a sign-coherent SST-SSS relationship. This is further illustrated in Figure 5.6 by the SST-SSS correlation probability density functions computed independently, via a Gaussian-kernel estimation, for dates when either one of the estimated modes are dominant. These results suggest that the proposed decomposition is capable of accurately separating the two opposite SST-SSS relationships, and demonstrate thus the relevance of the proposed scheme for the separation of physical processes from multi-tracer synergies.

5.3.2 Segmentation of upper ocean dynamics from SST/SSH relationships in the Agulhas region

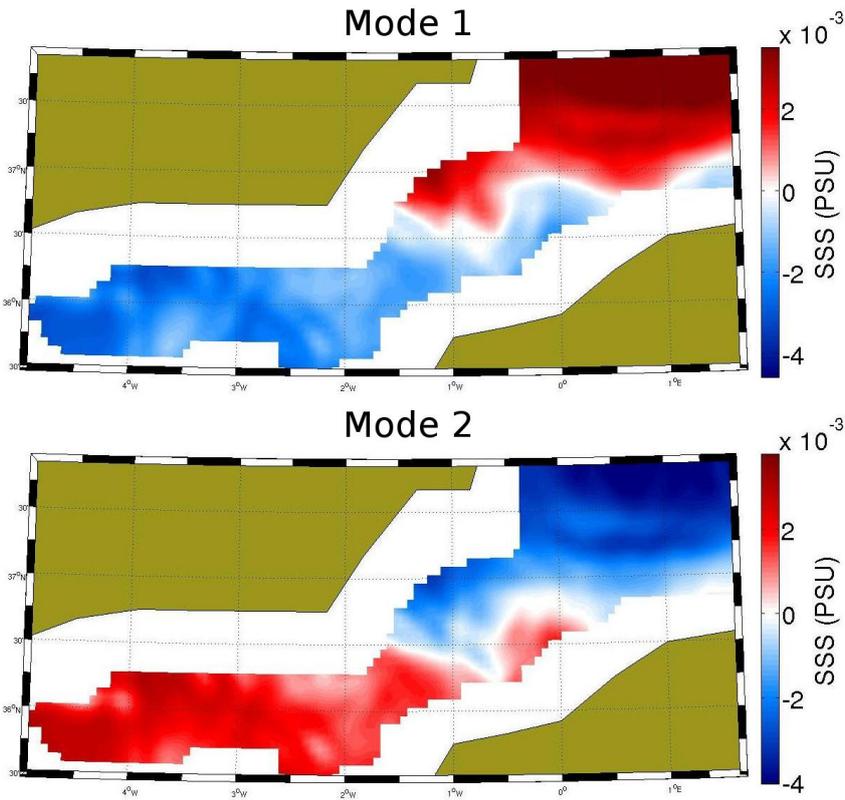
The main objective of this section is to illustrate the relevance of the proposed models to investigate and decompose upper ocean dynamics from multi-sensor data synergies, namely sea surface temperature (SST) and sea surface height (SSH). We recall that, as previously stated, from a theoretical point of view [131, 143], SST and SSH images are expected to be related through linear transfer functions that correspond precisely to fractional Laplacian operators [97, 115, 132, 142, 143, 244] and can be expressed in the Fourier domain as:

$$\mathcal{F}_H(\widehat{\text{SSH}}) = -\gamma |\mathbf{k}|^{-2\alpha} \mathcal{F}_T(\widehat{\text{SST}}) \quad (5.9)$$

where \mathbf{k} is the horizontal wavelength vector, \mathcal{F}_T and \mathcal{F}_H are linear filters of SST and SSH respectively and γ is a normalization coefficient [99, 244]. Varying parameter α , which controls the effective coupling between SST and SSH, leads to different classical theoretical models [244]. For $\alpha = 1/2$, for example, one resorts to the surface quasi-geostrophic model [99, 131, 143, 244]. For this application, we consider the direct minimization ALS method introduced in Section 3.3.3.1.



(a) Simulated fields



(b) Modal SSS predictions

Figure 5.5 – Non-negative decomposition of SST-SSS relationships in the Alboran Sea: (a) Ground-truth SST and SSS fields on March 22nd, 2011, (b) SST-derived predictions of the SSS fields for each mode of the considered dictionary-based local linear operator decomposition considering $K = 2$ modes.

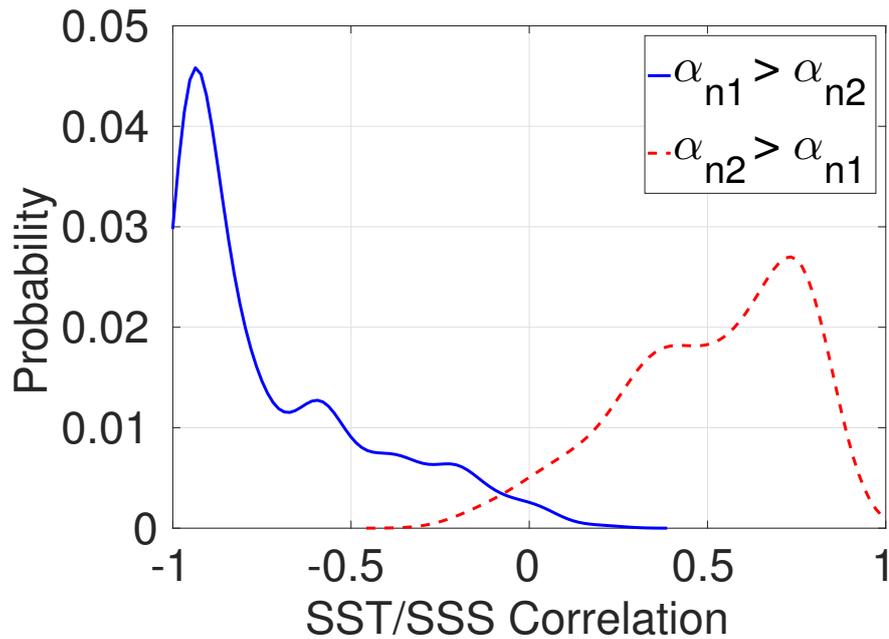
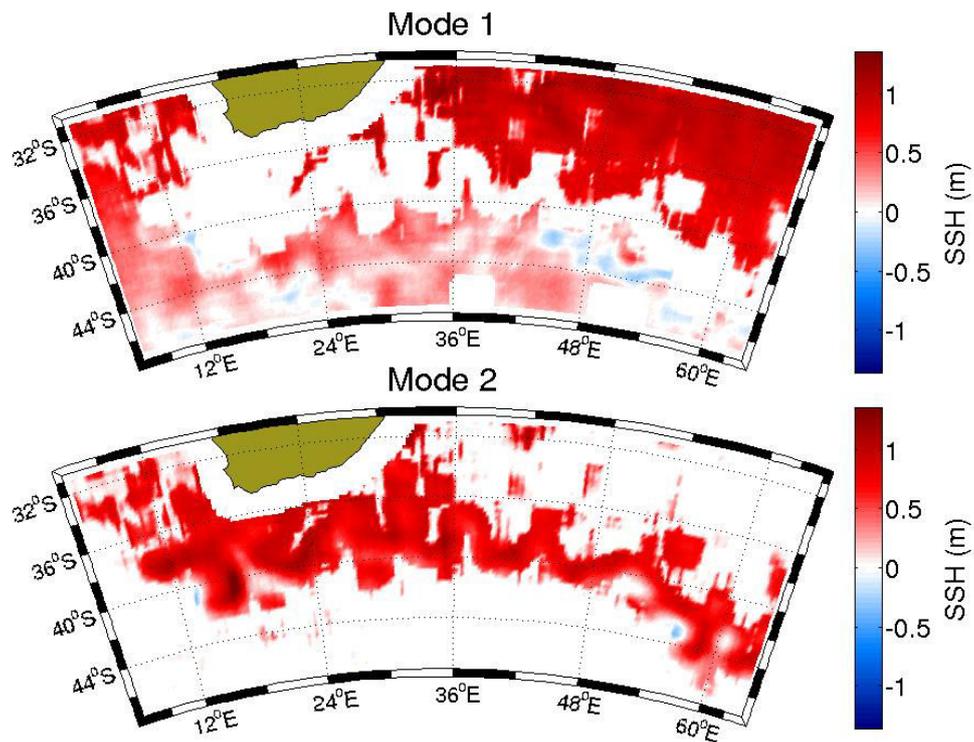


Figure 5.6 – Distribution of SST-SSS correlation when mode 1 (resp. mode 2) of the two-class dictionary-based local linear operator decomposition dominates, i.e. $\alpha_{n1} > \alpha_{n2}$ (resp. $\alpha_{n2} > \alpha_{n1}$).

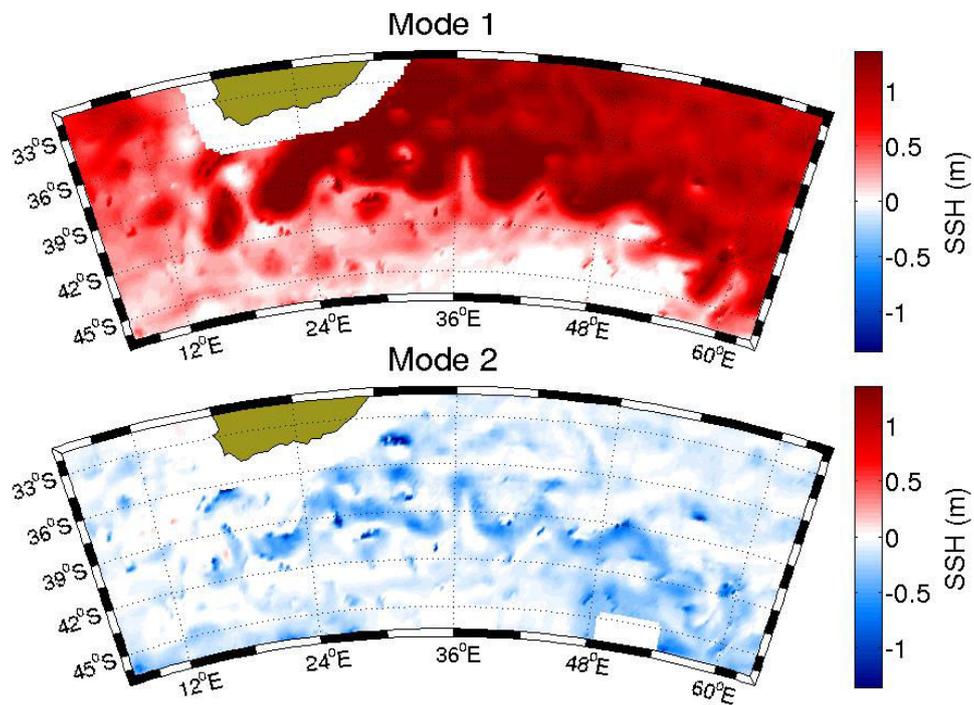
5.3.2.1 Remote Sensing Data

Following [244], we exploit the same SST-SSH dataset considered therein, which the authors used to perform the latent class segmentation of upper ocean dynamics. As SSH data, we use the daily delayed time Maps of Absolute Dynamic Topography (MADT) produced by Collecte Localisation Satellites (CLS) (available online at <http://www.avis.oceanobs.com/>). This information combines the signal of several altimeters onto a $1/3^\circ$ Mercator projection grid. We consider data from 2004 since four altimeters were available (Jason-1, Envisat or ERS-2, Topex/Poseidon and GFO). As SST data, we use optimally interpolated microwave SSTs provided by Remote Sensing System (RSS) (available online at <http://www.ssmi.com/>). It combines the signal of three microwave radiometers (Tropical rainfall measuring missions Microwave Image (TMI), Advanced Microwave Scanning Radiometer Earth observing system (AMSR-E) and WindSAT) which are robust to the presence of clouds. The spatial resolution is $1/4^\circ \times 1/4^\circ$ and the temporal resolution is the same as the MADT data, i.e. daily. We bilinearly interpolate the MADT data onto the SST grid.

We focus on the Agulhas region, a dynamical ocean region between longitudes $5^\circ E$ to $65^\circ E$ and latitudes $30^\circ S$ to $48^\circ S$. Given the joint series of satellite observations, we extract SST patches (noted as \mathbf{y}) and the associated concentric SSH patches (noted as \mathbf{x}). Overall, the processed dataset is composed of approximately 5×10^6 pairs of vectors \mathbf{y} and \mathbf{x} . To infer the parameters of the considered model, i.e. mixing coefficients α_{nk} and regression matrices β_k for each dynamical mode in model (3.3), we first build a training dataset as a random sample of



(a) Latent class regression model



(b) Proposed model

Figure 5.7 – Ocean surface dynamics segmentation. Predicted SSH per mode for the 1st January, 2004. Reported results correspond to both considered models, namely our proposed non-negative decomposition model and a latent class regression model introduced in [244], considering $K = 2$ classes. Taken from [166]. © 2016 IEEE.

$N = 10^5$ elements taken over the whole year 2004 (for a given day, we use about 2% of the data to fit the model). In a second step, we apply the inferred model to the entire processed dataset (which consists of approximately 1.3×10^4 patch pairs for each day of the year 2004) by considering the trained regression matrices β_k and re-estimating mixing coefficients α_{nk} on a daily basis to extract the spatio-temporal characteristics of the different dynamical modes.

5.3.2.2 Methodology

Following our proposed approach, we consider a generalized linear mixture model where the relationship between SST and SSH is a non-negative linear combination of K linear regressions. Given the dimensionality of the problem and involved variables, we consider here the simplified, single observation formulation of model (3.3), which relates to constrained optimization problem (3.7). In this respect, a unique mixing coefficient α_{nk} is affected to each observation pair $(\mathbf{x}_n, \mathbf{y}_n)$, with no two pairs sharing the same parametrization. The observation-wise adjustable amplitude parameters $\alpha_k \in \mathbb{R}^+$ are here used to take into account local variations in the strength of the relationships, in a manner similar to coefficient γ in Equation (5.9). In the real domain, rewriting linear transfer functions as matricial operators and formulating a patch-based combination of linear regressions, we resort to:

$$\mathbf{y}(s_i, t_i) = \sum_k \alpha_k(s_i, t_i) \beta_k \mathbf{x}(s_i, t_i) \quad (5.10)$$

where $\mathbf{y}(s_i, t_i) \in \mathbb{R}^{I \times 1}$ and $\mathbf{x}(s_i, t_i) \in \mathbb{R}^{J \times 1}$ are vectorized versions of local patches centered at location s_i at time t_i , as shown in Figure 5.8. The matricial operator $\beta_k \in \mathbb{R}^{I \times J}$ characterizes the k -th additive component, which relates SSH and SST fields through a linear filter. It is given by I vectorized versions of spatial convolution matrices, where \sqrt{I} is the considered SSH patch width. We typically set $\sqrt{I} = 3$ for our experiments, in order to be able to capture two-dimensional information. Similarly to [244], the SST patch width \sqrt{J} is set according to the Rossby radius of the study region, i.e., the mean size of mesoscale structures. For the Agulhas region, we considered $\sqrt{J} = 9$ which, at the spatial resolution of the dataset used, corresponds to a Rossby radius of 200 km. Coefficients α_k may be regarded as mixing coefficients for the different dynamical modes in play at a given space-time location. With respect to notations in Equation (5.9), linear filter β_k accounts for both the transfer function in (5.9) as well as linear filters \mathcal{F}_H and \mathcal{F}_T .

Similarly to [244], a key aspect here is the choice of the number of classes K , with several possible criteria for finding an optimal value being proposed in the literature. In our case, empirical results show that for the Agulhas region as few as $K = 2$ classes suffice to accurately characterize upper ocean dynamics, with a higher number of classes only resulting in a random partitioning of a class into several similar sub-classes. Additional modes do not bring significant improvement in terms of mean square error statistics either.

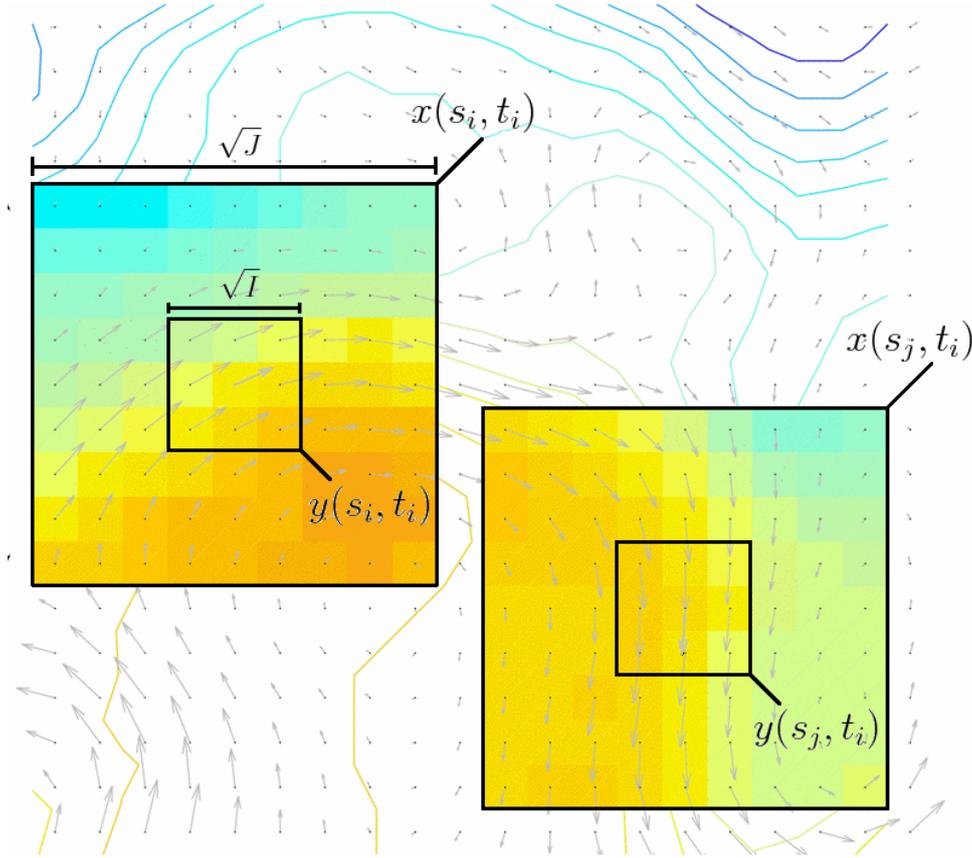


Figure 5.8 – Sketch of the considered patch-based representation of the SST field (in degrees represented in false colors), denoted by \mathbf{x} , and the SSH field (in meters represented by contour lines), denoted by \mathbf{y} , at locations s_i and s_j and time t_i . Adapted from [244]. © 2014 IEEE.

We calibrate model parameters using the direct minimization ALS algorithm applied to the training dataset. The retrieved regression matrices β_k are then used as the true regression matrices for the test dataset, and the ALS algorithm is used with fixed β_k to estimate mixing coefficients α_{nk} for the test dataset. For comparison purposes, we also apply a classical regression model, a latent class regression model with two classes [244], and a one-mode version of model (3.3) (*i.e.*, $K = 1$) to the same datasets.

Spatio-temporal Analysis of Dynamical Modes and SSH Prediction We exploit the estimated parameters of model (3.3) to perform the spatio-temporal analysis of upper ocean dynamics. For each day of the year 2004, each term of the non-negative decomposition can be computed separately to obtain a modal SSH prediction $\hat{\mathbf{y}}_{nk}$:

$$\hat{\mathbf{y}}_{nk} = \alpha_{nk} \beta_k \mathbf{x}_n \quad (5.11)$$

Besides, the computation of the complete non-negative decomposition allows us to perform a global estimation of a SSH patch from the corresponding SST patch:

$$\hat{\mathbf{y}}_n = \sum_{k=1}^K \alpha_{nk} \beta_k \mathbf{x}_n \quad (5.12)$$

Using these tools, we analyze the spatio-temporal variabilities, accuracy and evolution of both modal and global predictions and prediction errors to evaluate the relevance of the proposed model for the spatio-temporal analysis and characterization of upper ocean dynamics.

Model Learning with SQG Mode Forcing We introduce an alternative version of model (5.10) that comes to constrain the first mode of the non-linear decomposition to involve a SQG transfer function [99, 131, 143]. It amounts to setting the linear regression matrix β_k for the first mode to a matrix associated to a fractional Laplacian operator with $\alpha = 1/2$. For this model, parameter estimation proceeds similarly to the ALS approach for generic model (3.3), except that the regression matrix for the SQG mode is never updated. We use, as numerical approximation of the fractional Laplacian operator, the regression matrix that best approximates the Fourier-based transfer function (5.9) in the real domain for the considered patch sizes \sqrt{I} and \sqrt{J} .

5.3.2.3 Results

Model comparison and benchmarking We illustrate the intrinsic difference between the latent class regression model ($K = 2$) and the proposed model ($K = 2$) through the contribution of each mode to the reconstruction of the SSH field (Figure 5.7). Whereas the latent class regression model performs some binary spatial segmentation, the proposed model decomposes the SSH field according to two additive components: the first one providing a coarse prediction of the SSH field comprising the mean north-south gradient, the second one consisting of a negative anomaly along the Agulhas current (the region depicting the greater current values).

These results clearly demonstrate the relevance of the non-negative decomposition. It allows us to distinguish the form of the retrieved linear relationships between the SST and SSH fields, *i.e.* the normalized regression matrices β_k , from the local magnitude of these linear relationships, *i.e.* the mixing coefficients α_{nk} . This is a key feature, as the local magnitude of the retrieved linear relationships is expected to vary in space and time, especially in relation to the depth of the mixing layer. Our model also provides the means to test for the relative relevance of one-mode against two-mode models. The two-mode model leads to a significant relative improvement by a factor of 20 (Table 5.1)

From a quantitative point view, we report in Table 5.1 the relative mean square error (rMSE, given as a percentage) for the reconstruction of the SSH field for each considered model. With

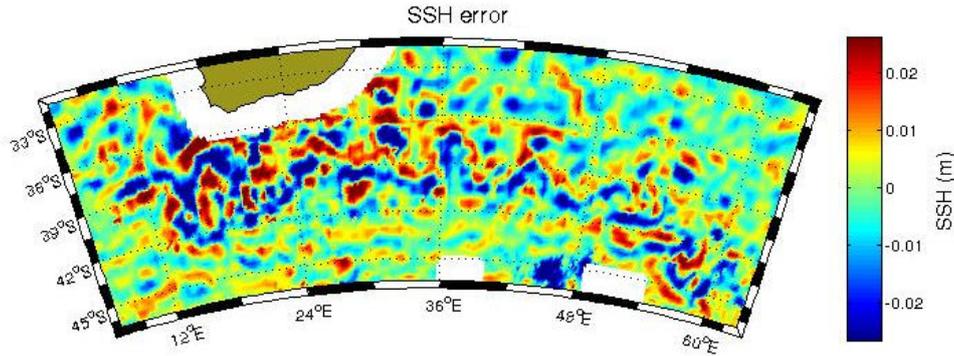


Figure 5.9 – SSH reconstruction error for the 1st January, 2004. Reported results correspond to our proposed non-negative decomposition model considering $K = 2$ classes. The Jet colormap was explicitly chosen to enhance contrast between positive and negative reconstruction error values.

Table 5.1 – Relative mean square error (rMSE, given as a percentage) for the different models considered. [Reported results correspond to a simple global linear regression; the two-class latent regression model introduced in [244] and our proposed non-negative decomposition model considering $K = 1$ and $K = 2$ classes. Taken from [166]. © 2016 IEEE.]

	rMSE
Linear regression	28.69 %
Two-class latent regression model [244]	27.22 %
Proposed model with $K = 1$	1.67 %
Proposed model with $K = 2$	0.08 %

respect to [244], the proposed model reduces rRMSE values by a factor of approximately 16 (resp. 340) for the reconstruction considering $K = 1$ (resp. $k = 2$) modes.

We further illustrate the reconstruction capabilities of the proposed algorithm by depicting, in Figure 5.9, the reconstruction error field $\mathbf{e}_n = \mathbf{y}_n - \hat{\mathbf{y}}_n$ for the 1st January, 2004. As far as reconstruction performance is concerned, the proposed formulation allows for the accurate reconstruction of the SSH field from the observed SST field and the retrieved modal regression matrices. Indeed, the reconstruction error for the SSH field is relatively small compared to the ground-truth SSH field. Moreover, higher error values are mostly found in or near the Agulhas current path. This is to be expected, given that this is the most dynamic part of the region and involves, thus, the higher local variance. It should also be noted that even though the algorithm is capable of retrieving and reconstructing most of the Agulhas current dynamics, it is still unable to completely and accurately reconstruct some smaller scale structures.

SQG Forcing vs. No Forcing We compare SSH prediction statistics for two-mode model (3.3) with and without SQG forcing for the first mode. We compute the spatio-temporal normalized root-mean squared error (nRMSE) of the global SSH prediction by normalizing SSH reconstruction RMSE by the standard deviation of the real SSH. The same analysis is performed

Table 5.2 – Normalized root mean squared error (nRMSE) values for global SSH and SSH gradient predictions for a two-mode model with and without external SQG forcing, and for a single-mode model without SQG forcing. *nRMSE values were calculated by normalizing the spatio-temporal global prediction RMSE for SSH and SSH gradient fields by the spatio-temporal standard deviation of the real SSH and SSH gradient fields.*

	SSH	SSH gradient
No forcing ($K = 2$)	0.0297	0.1302
SQG forcing ($K = 2$)	0.0370	0.1635
No forcing ($K = 1$)	0.1312	0.3164

for real and predicted SSH gradients. Additionally, we report RMSE statistics for a single mode model ($K = 1$) without SQG forcing. Results are summarized in Table 5.2.

We report greater nRMSE values for model (3.3) with SQG forcing compared to model (3.3) without SQG forcing. Interestingly, the most energetic mode of this second model is highly correlated to a SQG mode ($R^2 > 0.95$, $p < 0.001$). These results suggest that SST-SSH observation data truly embeds SQG-like dynamics as a dominant dynamical mode. We interpret the improved SSH prediction issued from model (3.3) without SQG forcing as a result of its better ability to account for the transfer functions applied during SST and SSH pre-processing steps (e.g., point spread function (PSF) of the instrument, space-time interpolation, etc.). As expected, model (3.3) considering 2 modes without SQG forcing also clearly outperforms the one-mode version of model (3.3).

From these results, we select model (3.3) without SQG forcing as our reference model to further investigate the spatio-temporal variabilities of upper ocean dynamics from SST-SSH relationships.

Spatio-temporal Distribution and Seasonality We first analyze the temporal variability of the considered multi-modal SST-SSH decomposition. We report in Figure 5.10 the time series of daily fraction of variance explained (FVE), computed as $1 - FVU$, with FVU being the fraction of variance unexplained, i.e., the ratio between the reconstruction error variance and the ground-truth field variance, for SSH and SSH gradient predictions for model (3.3). Except for one specific day, which may involve specific numerical issues, FVE values show a low variability for both cases, with typical values greater than 99.5% for SSH fields and greater than 98% for SSH gradient fields. The analysis of the FVE for the first mode of the decomposition, i.e. the SQG-like mode (Figure 5.10, top), indicates that this mode captures the overall meridional structure of the SSH field and accounts for 80%-85% of the overall variability. Interestingly, it exhibits a weak seasonal pattern with maximum FVE values in winter, around July and August. This is in agreement with previous studies [85, 115, 143, 244], as wintertime involves a lower upper stratification and is expected to depict stronger SQG-like SST-SSH relationships.

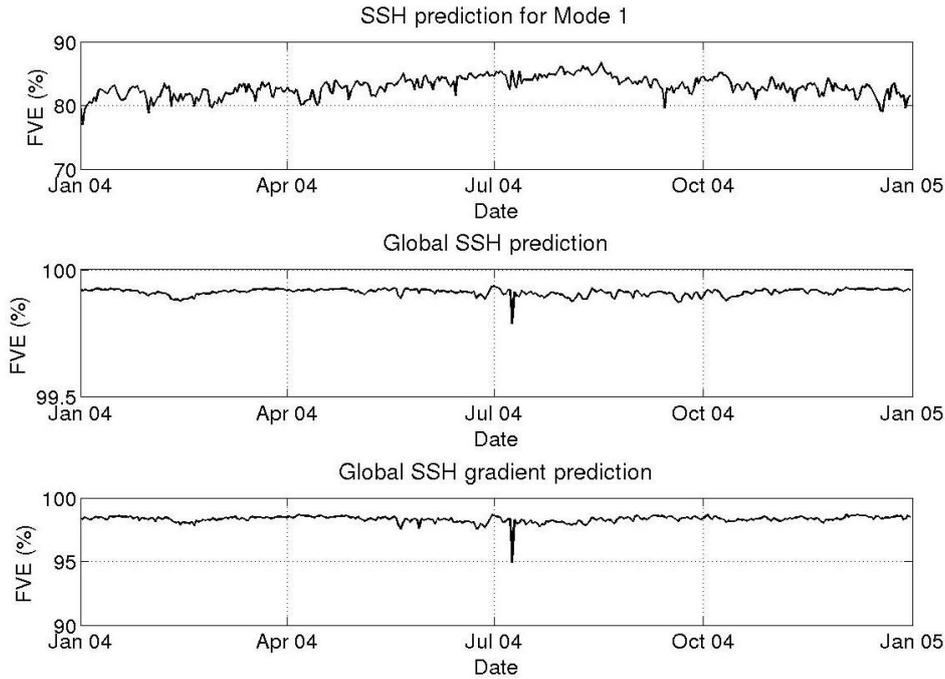


Figure 5.10 – Time series of the daily FVE (Fraction of Variance Explained) for model (3.3): top, SSH prediction for the first mode (SQG-like mode); middle, global SSH prediction; bottom, global SSH gradient prediction. *FVE* is computed as $1 - FVU$, with *FVU* being the fraction of variance unexplained, *i.e.*, the ratio between the reconstruction error variance and the ground-truth field variance.

To characterize the spatial distribution of the retrieved modes, we present in Figure 5.11 the mean annual SSH prediction for each mode. The first mode clearly captures a local mean SSH level, which is consistent with the FVE levels previously reported. The second mode is mainly active in the strong current area and acts as a negative compensation around the mean value captured by the first mode. The second mode can be regarded as a local correction to the first mode so as to more accurately reconstruct the global SSH field at each spatio-temporal location. The spatial mapping of standard deviation of the SSH prediction error from the first mode in Figure 5.12 further stresses that the prediction errors to be compensated by the second mode are mainly located within the frontal region associated with the strongest currents. This is in agreement with Figure 5.11 (bottom) which clearly highlights a stronger contribution of the second mode within this area. Effectively, the statistical analysis of the relationship between the first mode SSH prediction error standard deviation and the second mode SSH prediction reveals a high correlation ($R^2 > 0.90$, $p < 0.001$).

Analysis of mixing parameters We further investigate the spatio-temporal distribution and physical interpretation of model parameters, particularly of mixing coefficients α_{nk} . We compute the mean annual mapping of the mixing coefficients for each mode for the year 2004 as well as the time series of their daily spatial mean. Results are presented in figures 5.13 and 5.14 respectively.

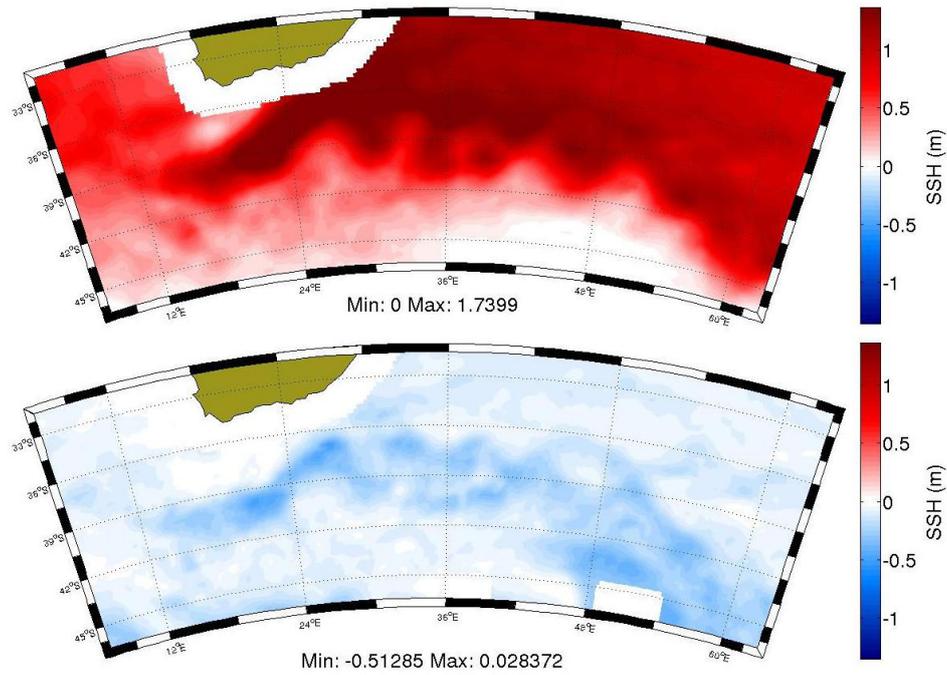


Figure 5.11 – Mean annual SSH field prediction for each dynamical mode of model (3.3): top, first mode (SQG-like mode); bottom, second mode.

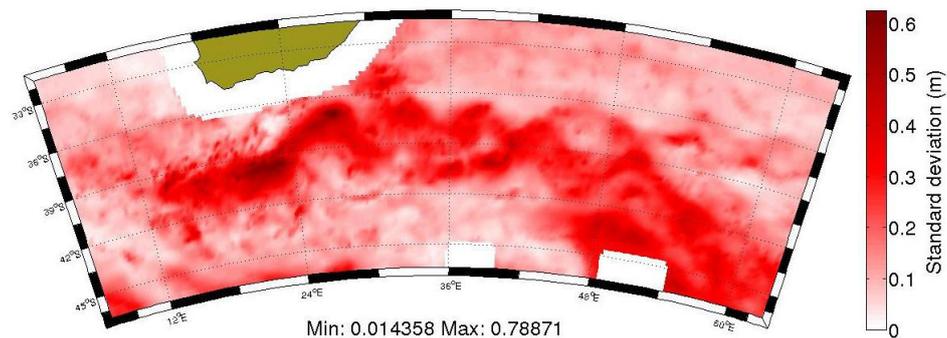


Figure 5.12 – Mean annual SSH prediction error standard deviation for the first mode (SQG-like mode) of model (3.3).

The mean annual spatial distribution of mixing coefficients bears a close resemblance to mean annual SSH field predictions (Figure (5.11)), which indicates that mixing coefficients are strongly related to modal SSH predictions ($R^2 > 0.35$, $p < 0.001$), and to real SSH values for the first mode ($R^2 > 0.95$, $p < 0.001$). As far as the daily evolution of the spatial mean of mixing coefficients is concerned, we can see that both modes are strongly correlated ($R^2 > 0.80$, $p < 0.001$), with a common trend shared by both modes. A clear seasonality can also be observed, with higher values for the mixing coefficients in winter, between the months of July and October, and lower values for the mixing coefficients in summer, between the months of January and April, as well as clear transition periods in between. Mean second mode mixing coefficients may involve two plateaus in winter and summer, whereas mean first mode coefficients depict a

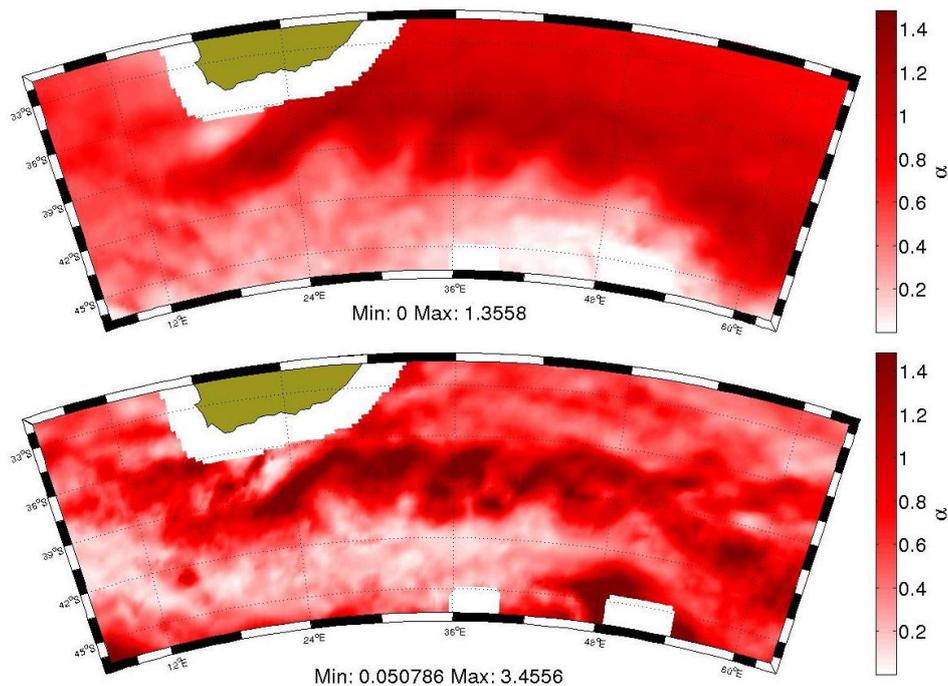


Figure 5.13 – Mean annual mapping of mixing coefficients α for each dynamical mode of model (3.3): top, mode 1 (SQG-like mode); bottom, mode 2.

more continuous pattern.

To further investigate the interpretation of mixing coefficients α_k and better understand their role in the proposed model, we analyze the time series of the mixing coefficients within four $1/2^\circ \times 1/2^\circ$ zones. As detailed in Figure 5.15a, we select two zones within the frontal area, referred to as active zones, and two zones away from the frontal area, referred to as passive zones.

In figure 5.15b, we report the time series of the daily mean mixing coefficients for each zone and for the two dynamical modes of model (3.3). For all zones, first mode mixing coefficients depict a much smaller temporal variability than second mode mixing coefficients. This is in agreement with the interpretation of the first mode as a mean proxy of the SSH and the second mode as a local correction. In passive zones, second mode mixing coefficients involve more periods with very low values, which indicates a weaker local contribution of the second mode in passive zones with respect to active zones. This might suggest some relationship to the local forcing of upper ocean dynamics by ocean surface winds [147]. Regarding active zones, both modes seem to significantly contribute throughout the year. It may also be noted that zones A1 and P2 depict a clear correlation between the mixing coefficients of the two modes (respectively, $R^2 > 0.74$, $p < 0.001$; $R^2 > 0.34$, $p < 0.001$). By contrast, zone P1 shows a much weaker correlation and zone A2 even presents anti-correlation patterns. This observation suggests that the analysis of local space-time patterns of mixing coefficients may provide additional means

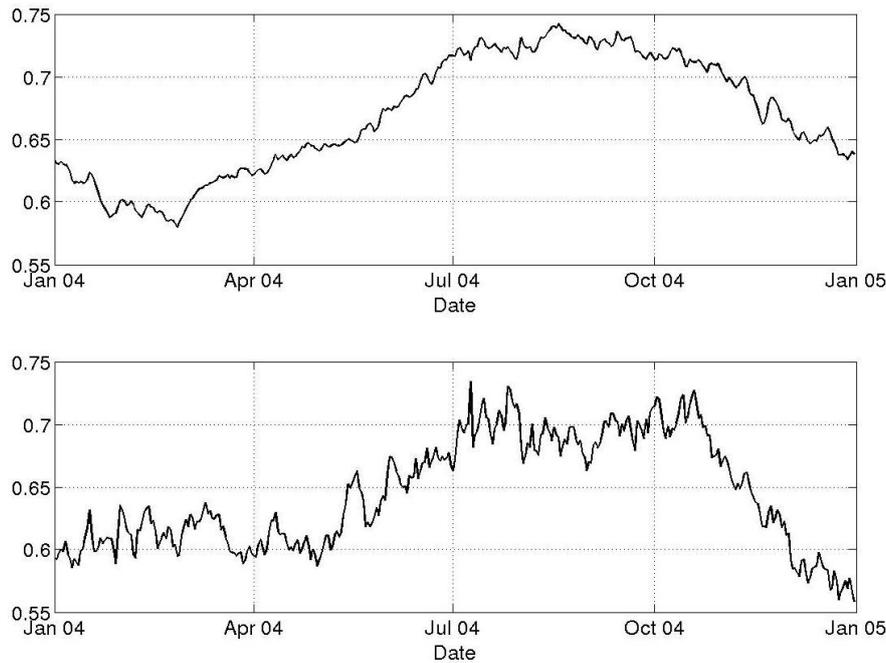
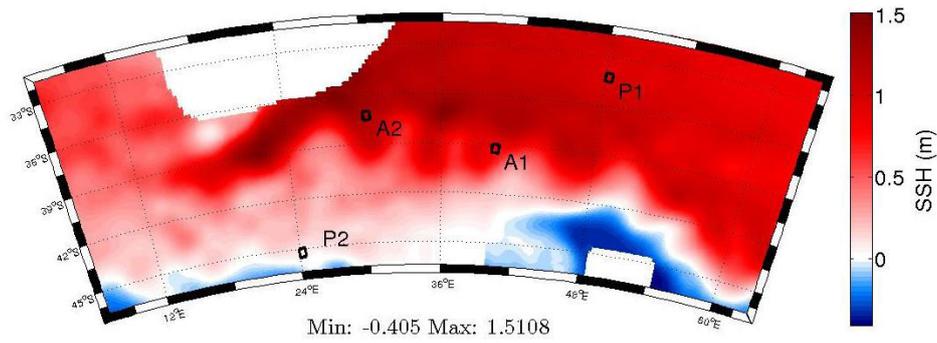


Figure 5.14 – Time series of the daily mean of mixing coefficients α of model (3.3): top, mode 1 (SQG-like mode); bottom, mode 2.

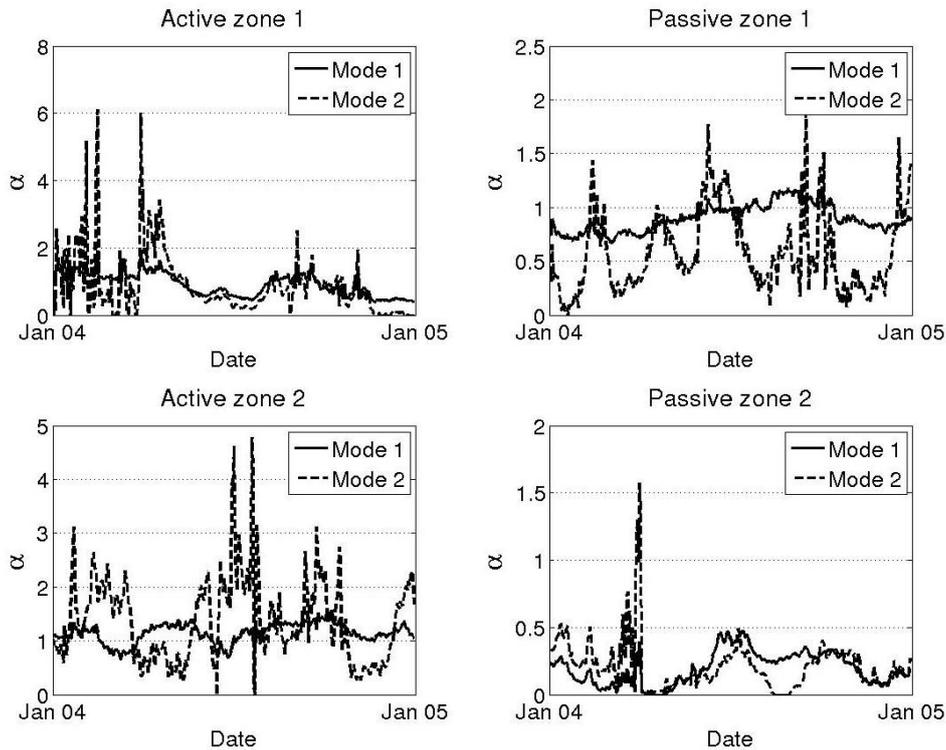
for the characterization of local upper ocean dynamics, for instance related to vertical mixing conditions.

5.3.2.4 Conclusion

Applied to the 2004 daily $1/4^\circ \times 1/4^\circ$ satellite SSH and SST data within an active ocean region off South Africa, our model significantly outperforms the latent class regression model developed in [244] by more than one order of magnitude in terms of SSH prediction (respectively, 0.08% vs 27.22% for the SSH reconstruction rRMSE). Interestingly, we show that our model retrieves a first dynamical mode that is strongly correlated to SQG dynamics, i.e. to a fractional Laplacian transfer function. This dynamical mode accounts for the mean SSH field and captures most of the spatio-temporal variability of the region (above 80%). By contrast, the application of a latent class regression model, as in [244], splits this mode into different ones, which can be interpreted as reparametrization of SQG-like dynamics with different mean mixing coefficient values. As such, our model provides a simple mean to locally evaluate the extent to which SQG dynamics apply. This is achieved at the resolution of the considered field, typically $1/4^\circ \times 1/4^\circ$. By contrast, Fourier-based analyses, as in [85, 116], typically consider regional scales. Moreover, the second dynamical mode acts as a local correction to the first mode. It shows more activity in the frontal area, where the strongest sea surface currents are observed. These findings truly support the existence of superimposed dynamical modes, rather than mutually exclusive ones as assumed in [244]. This seems more in agreement with the expected continuous shift between ideal SQG-like dynamics and non-SQG dynamics, especially with respect to the space-time vari-



(a) Considered zones for the analysis of mixing coefficients in model (3.3). The location of each zone is reported on top of the real SSH annual mean.



(b) Time series of the daily mean of mixing coefficients α for each zone.

Figure 5.15 – Spatio-temporal variability of mixing coefficients α of model (3.3) with respect to four contrasted zones.

ability of the upper ocean stratification and mixed layer depth [85, 99, 115, 116, 131, 143, 244]. Complementary analyses should be undertaken to fully assess the geophysical interpretability of the proposed decomposition.

5.4 Conclusion

We presented, throughout this chapter, a variety of applications of the novel non-negative linear decomposition model to a number of problems in physical oceanography and geosciences. Par-

ticularly, we focus on the characterization of upper ocean dynamics from satellite-derived ocean tracers (SST,SSH,SSS) and on reconstruction/forecasting issues for the Lorenz '96 model.

For each one of the proposed applications, we compared our proposed model and algorithms to state-of-the-art formulations and presented their advantages with respect to more classical models. The reported applications clearly illustrate the relevance of the proposed model as an interesting alternative to more classical formulations involving orthogonality priors or simpler formulations, such as classical linear regression and latent class models. As far as the valorization of our contributions is considered, it is worth mentioning that the applications detailed here have been successfully presented in a number of national and international conferences [165–167].

With this chapter, we conclude the first part of this thesis dissertation. In the next part, we distance ourselves from inverse problems in a general sense, a shift our focus towards a specific type of inverse problem, namely the interpolation of high-resolution sea level anomaly (SLA) maps from satellite-derived observations. In this context, we initially follow [66] and explore a reformulation of our non-negative decomposition model as a super-resolution image processing issue. Subsequently, we develop on the work introduced in [67, 161, 162] and tackle the problem from a data assimilation point of view. We explore strategies to better exploit satellite-derived observations and their synergy and also evaluate potential strategies to include additional oceanic tracers (sea surface temperature, geostrophic currents, etc). The connection and potential integration between the considered data assimilation methodology and previously introduced non-negative/sparse operator decomposition approaches is also briefly explored.

Part II

Data-driven interpolation of sea level anomaly fields from satellite-derived ocean remote sensing data

Context and related work

"For me context is the key – from that comes the understanding of everything."

Kenneth Noland

6.1	Introduction	109
6.2	Satellite altimetry	111
6.2.1	Along-track nadir altimetry	111
6.2.2	Surface Water Ocean Topography altimetry mission	115
6.3	Geophysical field interpolation	118
6.3.1	Exemplar-based methods	119
6.3.2	Projection-based methods	119
6.3.3	Data assimilation	119
6.3.4	Analog Forecasting	124
6.3.5	AnDA: The analog data assimilation framework	125
6.4	Conclusion	128

6.1 Introduction

In the previous chapters, we focused on the extension of constrained blind source separation formulations to the problem of data-driven operator decomposition, with a view towards tackling inverse problems in a general sense. In the second part of this thesis work, we direct our attention towards one particular inverse problem, namely the interpolation of high-resolution geophysical fields (and particularly sea level anomaly fields) from satellite-derived data. In this chapter, we present the context and motivation behind the idea of exploiting data-driven methods for the problem of interpolating geophysical fields from satellite data, including a brief description of satellite altimetry and associated observation datasets, and a review of classical state-of-the-art geophysical field interpolation models, their hypothesis and limitations.

Thanks to recent advancements in remote sensing, in situ measurements and numerical models, an ever increasing amount of data, coming from a wide range of sources, is gathered daily. However, most of these data sources use different sampling strategies at multiple spatio-temporal resolutions, and usually involve irregular sampling patterns and missing data due to cloud occlusion, satellite orbit and track geometry, etc. In this context, the processing of such datasets to produce gridded gap-free products is of major interest for both research and industry.

As already mentioned, the problem of producing a regularly-gridded field from irregularly-sampled data belongs to the family of inverse problems, which have been extensively studied [63, 177, 195, 207]. In ocean sciences, state-of-the-art methods rely on Optimal Interpolation (OI) [20, 57, 152], a model-driven approach that involves modeling the spatio-temporal covariance structure of the field to be interpolated, usually under the hypothesis that this covariance structure remains constant in space and time. Such a hypothesis is not always verified, which leads to a lack of accuracy in the representation of the dynamics of smaller scale structures. Indeed, the Gaussian assumption inherent to OI techniques causes smaller scale structures to be smoothed, and it has been verified that it is not possible to recover fine scale structures (between 10-100 km) using OI [34, 68, 73]. Such limitation also relates to the inherent track separation in conventional altimeter tracks, governed by the orbit design and sampling strategies of currently space-borne altimetry missions.

On the other hand, data assimilation techniques [61] have been regarded as a particularly powerful model-driven tool for the interpolation of high-resolution geophysical fields, provided that the chosen physical model is sufficiently accurate and does not bear a prohibitive computational cost. Nowadays, most alternative methods that do not rely on OI or data assimilation remain mostly model-driven, in spite of the growing availability of massive datasets, issued from both real remote sensing or in situ observations as well as numerical simulations or reanalysis issues, that supports the exploration of data-driven approaches as a powerful and efficient alternative [67, 207].

Here, we follow recent developments [65–67, 162, 207] that, given the great variety of high-resolution observation, simulation and reanalysis datasets available, explore data-driven strategies as a powerful, computationally efficient alternative to model-based approaches for the interpolation of high-resolution geophysical fields. Data-driven approaches aim, in general, at exploiting available datasets, which can be considered to accurately depict the spatio-temporal variability of the fields of interest, to emulate the physical model behind the process of interest and better capture the spatio-temporal variabilities that may not be accounted for in simplified, purely numerical models.

From an operational point of view, among recent advancements in remote sensing, the Surface Water and Ocean Topography (SWOT) satellite mission, a joint effort between the US National Aeronautics and Space Administration (NASA), the French Centre Nationale d’Etudes Spatiales (CNES) and the UK and Canada Space Agencies, aims at providing high resolution altimetry maps for both hydrology and oceanography. Specifically, the SWOT mission will be the first mission to exploit Ka-band radar interferometry altimeters producing, for the first time, complete two-dimensional wide-swath satellite tracks [59, 74, 78], as seen in Figure 6.3. Moreover, the mission’s unparalleled high spatial resolution (1 km) should allow us to better capture mesoscale and sub-mesoscale processes [59, 74], which constitutes one of the great challenges within current oceanography research.

In the following chapters, we will attempt to tackle the problem of the data-driven interpolation of geophysical fields from different methodological perspectives. Following previously introduced formulations [66, 169], we will first formulate the problem as a super-resolution image processing issue exploiting convolutional models and extend the dictionary-based local linear operator decomposition presented in previous chapters to this new problem. Subsequently, we follow [67, 161, 162] and shift our focus towards data-driven data assimilation formulations. In preparation for the SWOT mission launch in 2021, we also explore, in subsequent chapters, the potential of exploiting SWOT-like observations in the context of data-driven methods for the reconstruction of high-resolution altimetry fields.

6.2 Satellite altimetry

In this section, we present a brief overview of satellite altimetry, including an overview of the satellite remote sensing technologies and methodology behind the data products to be used for the interpolation of regularly-gridded, gap free SLA fields, with particular focus on their spatio-temporal characteristics, noise properties and the fundamental differences between them.

6.2.1 Along-track nadir altimetry

In the last few decades, multiple ocean altimetry missions have been implemented and subsequently launched into space, starting from the first generation altimeters such as Seasat (1978) and Geosat (1985) to currently airborne and operative missions such as Jason 2-3, Sentinel 3A-3B, Cryosat or Saral/Altika. The ocean surface topography (OST) measurements provided by these altimeters have permitted considerable advancements in our understanding of ocean dynamics and ocean-atmosphere interactions [72, 193, 194].

Conventional altimetry missions rely on either traditional radar altimeters or more complex synthetic aperture radar (SAR) instruments. Both technologies work by emitting radar pulses towards the ocean surface and measuring the return time Δt to estimate the ocean surface to-

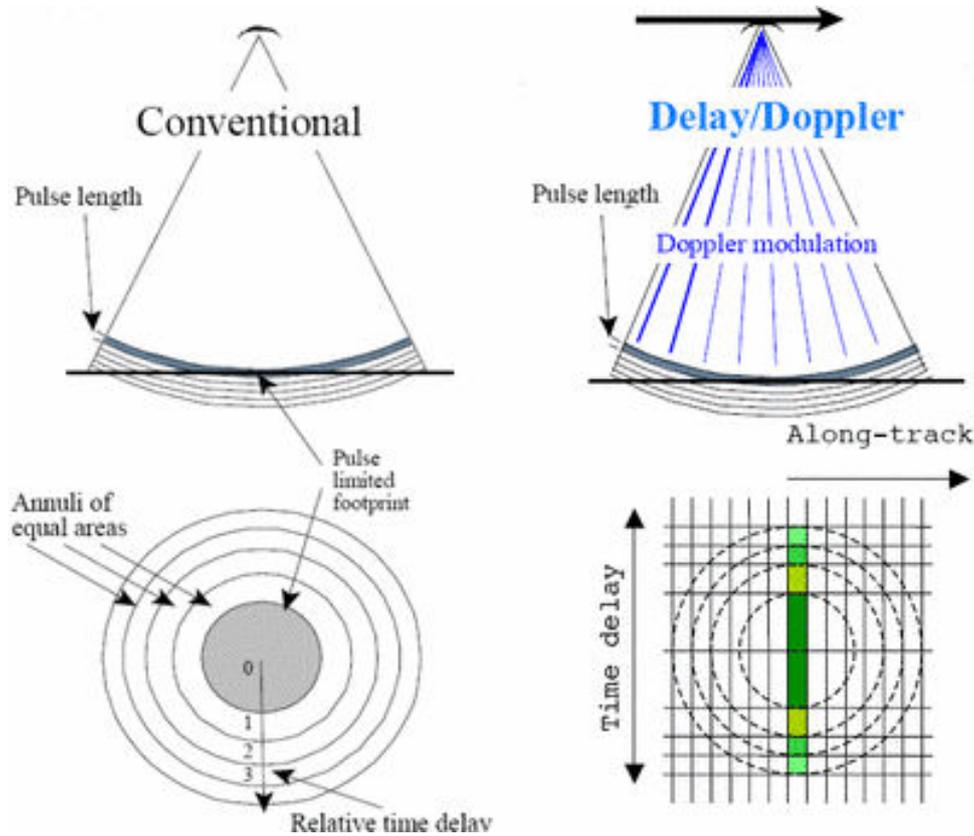


Figure 6.1 – Comparison of conventional pulse-limited radar altimetry (left) and synthetic aperture radar (SAR) altimetry (right). Credits: R.K. Raney, Johns Hopkins University Applied Physics Laboratory¹. Used with permission.

pography, with the difference that SAR altimetry performs a batch processing of coherent groups of multiple pulses transmitted at successive times. In SAR altimetry, successive pulses are transmitted as the satellite moves, and by processing the return time, amplitude, phase and Doppler frequency a two-dimensional view with higher along-track resolution can be achieved [219]. Figure 6.1¹ presents an illustration of conventional and SAR altimetry. As opposed to conventional altimetry, which only takes measurements directly in the nadir direction, SAR senses each along-track spatial location as long as it is illuminated by the satellite beam and exploits the satellite movement to simulate a larger antenna (hence the name synthetic aperture antenna). In this way, the full Doppler bandwidth (the frequency range of signals produced as a result of the ocean surface falling in and out of sight of the radar as the satellite moves) is exploited, so that the power reflected from the ocean surface is more efficiently used, thanks to SAR instruments not being pulse-limited like traditional single-pulse radar altimeters.

More precisely, satellite altimetry measures the distance between the satellite and the ocean surface, also known as the altimetric range, by measuring the return time Δt of an emitted radar pulse. From the measured altimetric range, multiple ocean tracers can be computed, each one

¹ <http://www.altimetry.info/radar-altimetry-tutorial/how-altimetry-works/delay-doppler-or-sar-altimetry/>

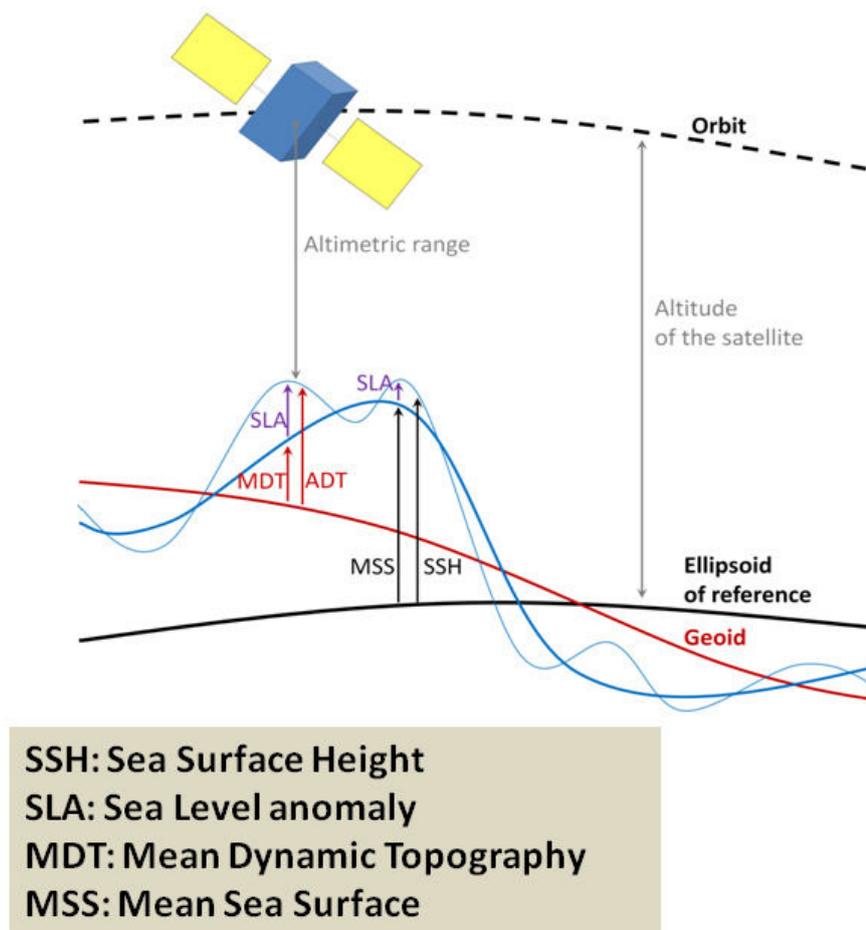


Figure 6.2 – Different quantities derived from altimetric measurements with reference to the Earth's Geoid and the Ellipsoid. Credits: CLS. Source: duacs.cls.fr². Used with permission.

comporting a particular interest for the study of different oceanic processes and phenomena. Figure 6.2² presents a schema explaining the most relevant tracers that can be obtained from the satellite-measured distance and other relevant reference levels, such as the geoid (a surface of equal gravitational potential, perpendicular to the gravity field, and relating to the mean sea level and the equilibrium position of the ocean) and the reference ellipsoid (the best fit of the earth true shape). In future chapters, in particular, we will focus on Sea Level Anomaly (SLA), which can be seen as either variations of the Sea Surface Height (SSH) around the Mean Sea Surface (MSS) (with reference to the Ellipsoid), or variations of the Absolute Dynamic Topography (ADT) around the Mean Dynamic Topography (MDT) (with reference to the Geoid).

The main limitation of conventional altimetry is that it can only produce measures along the nadir of the satellite track, with a radar altimeter footprint width of around 2-10 km, as observed in Figure 6.3. This implies that, when compared to other remote sensing data sources (e.g. microwave sea surface temperature (SST)), conventional altimeter observations may involve

² <https://duacs.cls.fr/faq/different-sea-surface-heights-used-in-altimetry/>

a very scarce sampling of the ocean surface with higher rates of missing data, bigger gaps and increased sensibility to orbit characteristics such as track separation and revisit time. Indeed, current generation altimeters have spatial gaps of the order of 100 km between tracks. In particular, it should be noted that a compromise exists between track separation and revisit time, with smaller track separation involving longer revisit times and vice-versa. In this respect, no satellite mission is currently capable of providing high-resolution observations in both space and time. As far as instrument noise and observation errors are considered, the instrument characteristics and along-track nature of conventional altimetry induces observation error processes that can be modeled as centered, white, Gaussian noise processes. In this respect, the covariance of observation noise for current generation altimetry missions lies in the range $\sigma_n^2 \in [3, 5] \text{ cm}^2$ [204].

Preliminary studies on the exploitation of nadir along-track satellite-derived data have proven that at least two altimeters are needed to accurately resolve the main spatio-temporal scales of global ocean processes [133], with further studies validating and illustrating this point [50, 51]. In [191], Morrow and Le Traon show that at least three altimeters are needed to capture mesoscale signals and processes. Subsequent studies exploiting both numerical models and real satellite data concluded that using up to four altimeters at the same time helps improve the accuracy and reconstruction power of state-of-the-art interpolation techniques exploiting these datasets [75, 148–151, 204]. However, even though experiments combining multiple altimeters allow for the reconstruction of scales that would be unattainable using a single altimeter, no combination of current airborne altimeters is capable of completely resolving the smaller mesoscale (below 50-70 km) [68, 73] or the sub-mesoscale processes (below 10 km) [34], which are nonetheless crucial to increasing our understanding of a great number of oceanic processes [74]. In this respect, multiple studies show the importance of sub-mesoscale processes in ocean circulation [31, 74, 144, 258]. In particular, whereas mesoscale processes relate mainly to horizontal transport, sub-mesoscale processes seem to be associated with stronger vertical velocities and vorticity magnitudes, and would thus play a particularly important role for the transfer of ocean properties from the surface and upper ocean to the deep ocean [144]. Moreover, recent studies show that high-resolution measurements of OST can be used to characterize vertical ocean currents in the upper ocean [53, 132].

In this context, the need for the development of novel altimetry techniques that improve on current altimeters to enhance their observation capabilities in order to better observe and resolve OST at smaller scales appears clearly. The Surface Water Ocean Topography (SWOT) mission, which we discuss in the next section, is one of such new developments, exploiting a novel wide-swath altimetry principle that will allow for unprecedented horizontal resolution for the observation of OST.

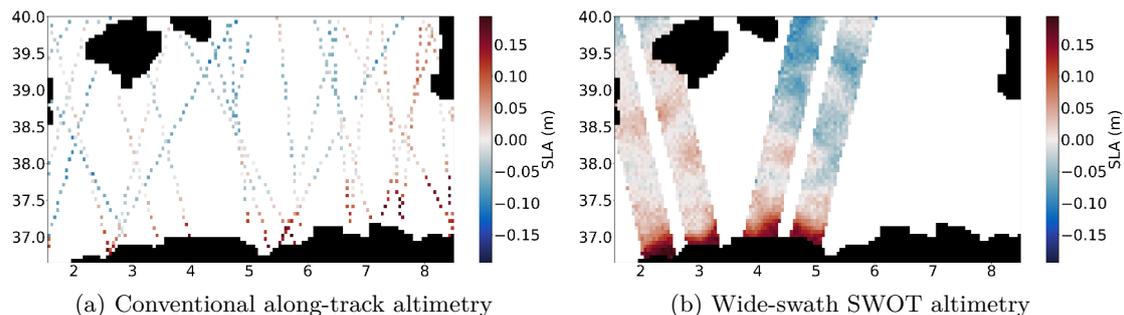


Figure 6.3 – Comparison of classic nadir along-track observations (6.3a) and off-nadir wide-swath SWOT observations (6.3b). Adapted from [170]. © 2019 IEEE.

6.2.2 Surface Water Ocean Topography altimetry mission

The Surface Water Ocean Topography (SWOT) mission is a joint effort between the US National Aeronautics and Space Administration (NASA), the French Centre Nationale d’Etudes Spatiales (CNES) and the UK and Canada Space Agencies, expected to go airborne in 2021. It aims at simultaneously responding to scientific questions from both hydrology and oceanography by providing high resolution altimetry maps for land water masses, rivers, coastal regions and the global ocean. In preparation for the mission launch in 2021, we also explore, in subsequent chapters, the potential of exploiting SWOT-like observations in the context of data-driven methods for the reconstruction of high-resolution altimetry fields.

6.2.2.1 Wide-swath radar interferometry altimetry

Specifically, the SWOT mission will be the first mission to exploit a novel wide-swath Ka-band radar interferometry altimeter producing, for the first time, complete two-dimensional off-nadir wide-swath satellite tracks [59, 74, 78] as seen in Figure 6.3.

A comparison of along-track nadir-looking altimeters and off-nadir wide-swath interferometry altimeters is presented in Figure 6.5³. As observed, the off-nadir track allows for a wider footprint of the instrument, while a higher resolution can be achieved by exploiting interferometry between two off-nadir radiometers.

The main instrument on the SWOT satellite is a Ka-band radar interferometer (KaRIn), based on two SAR radiometry antennas mounted on opposite ends of a 10 m boom, as illustrated in Figure 6.4 [76]. The principle used for the interferometry measurement involves illuminating each swath from a single antenna and measuring the back-scatter for both antennas. The difference in phase between the received signals can then be used to estimate the OST on the two-dimensional swath. Moreover, in order to have simultaneous measurements for both

³ <https://www.avisio.altimetry.fr/en/techniques/altimetry/future-evolutions/interferometers.html>

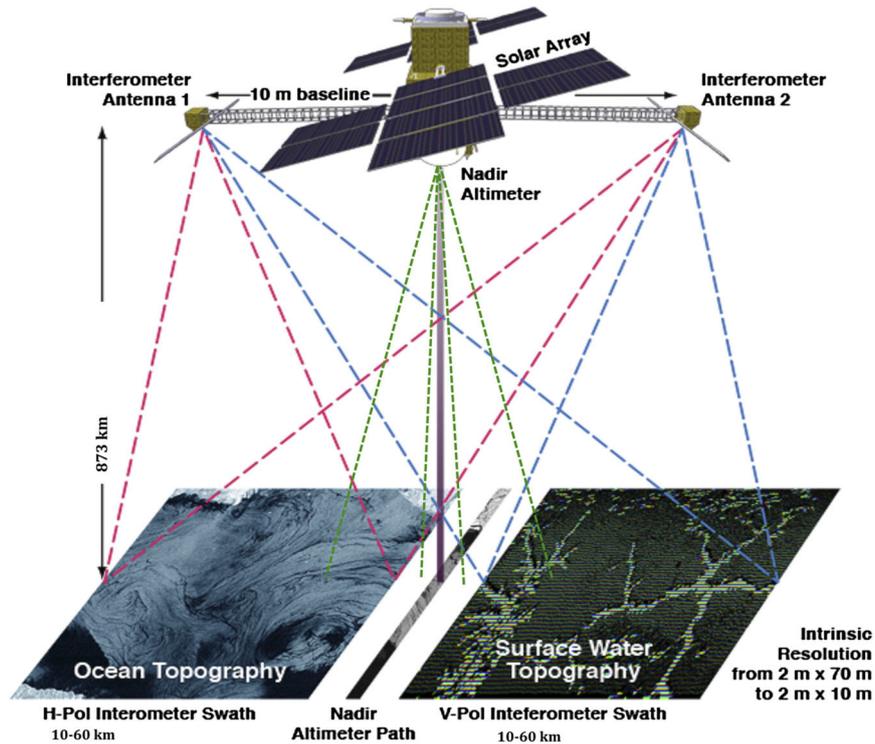


Figure 6.4 – SWOT satellite scheme. A 10 m boom separates two SAR antennas. Each antenna receives the backscatter from both swaths, which are illuminated by a single antenna with two different polarizations (to have two separate simultaneous measurements). Interferometry is used to estimate the OST for both swaths from the received backscatter on both antennas. Taken from [76]. © American Meteorological Society. Used with permission.

swaths, each one of them is illuminated with a different polarization (horizontal or vertical). The satellite payload also includes a classic SAR nadir-looking altimeter to obtain simultaneous nadir along-track OST measurements as well.

The altimetric measurements provided by SWOT will involve two swaths of 50 km, spanning from 10 km to 60 km in the across-track direction (measured from nadir). This amounts to a total observation zone spanning 120 km in the cross-track direction, with two 10 km gaps between the nadir along-track measurement and the interferometer swaths. After post-processing, SWOT measurements are expected to have an along-track and across-track resolution of 1 km and a total noise standard deviation of around $\sigma_{SWOT} = 1$ cm (considering a root-sum-squared (RSS) error analysis criteria) [54, 59]. Moreover, the proposed orbit for SWOT will involve a gap between swaths of around 130 km, with a total orbital cycle of around 23 days and a repetitivity subcycle with a mean revisit period of approximately 10 days. Given the constant track separation and the varying (spatial) revisit time (which decreases as we move away from the Equator), the proposed orbit will involve higher track density for higher latitudes (since earth surface also decreases with latitude). In this respect, the mission's unparalleled high spatial resolution should allow us to better capture mesoscale and sub-mesoscale processes [59, 74], which constitutes one of the great challenges within current oceanography research.

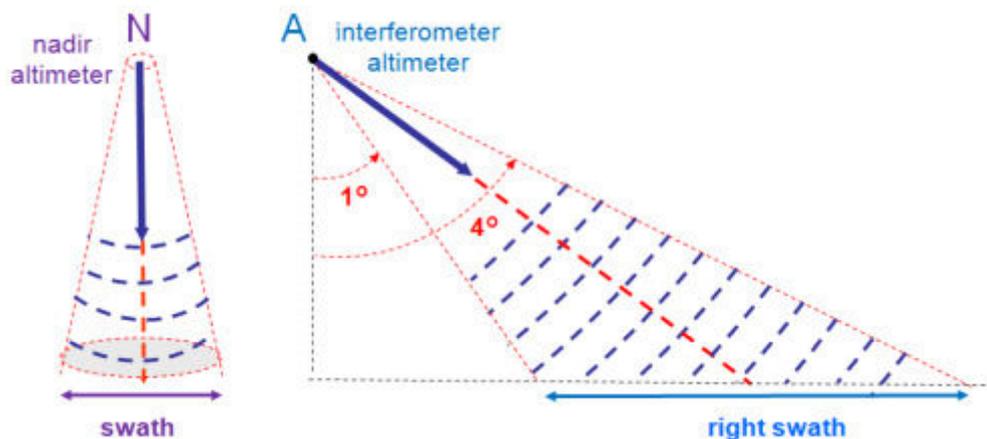


Figure 6.5 – Comparison of along-track nadir-looking altimeters (left) and off-nadir wide-swath interferometry altimeters (right). *The off-nadir track allows for a wider footprint of the instrument, while a higher resolution can be achieved by exploiting interferometry between two off-nadir radiometers.* Credits: AVISO³. Used with permission.

6.2.2.2 Noise sources in SWOT

One of the main limitations concerning the future SWOT mission involves the numerous sources of error associated with radar interferometry measurements.

Indeed, given the instrument characteristics and the geometric nature of the wide-swath radar interferometry measurements, SWOT will be subjected to different error sources, with two main type of errors being identified. Below we specify the most relevant error sources. We refer the interested reader to [59] for an in-depth discussion of all the error sources identified for SWOT and to [78] for details concerning their numerical simulation. Additionally, most error sources specified below are also taken into account in realistic numerical simulations of SWOT observations [78].

Instrumental errors:

- **KaRIn instrument noise:** the observation error associated with the KaRIn interferometer radar instruments is a centered, uncorrelated white Gaussian noise, whose variance depends on the ocean significant wave height (SWH) and on cell (pixel) surface, and hence on the distance to nadir (since pixel size increases as we move away from nadir).
- **Roll errors:** It corresponds to errors introduced due to uncertainties in the satellite roll estimation (roll knowledge/gyro error) or due to uncertainties in roll corrections performed during the stabilization of the satellite (roll control error). Given the geometrical nature of this error, it is a spatially correlated random process in the across-track direction.
- **Phase errors:** It involves uncertainties due to changes in the relative phase between the two back-scatter signal pairs in the interferometric pair [59]. It may be introduced by radar

electronic systems, mechanical/thermal antenna systems and/or multi-path in the back-scatter signal or external signals. Statistically, this noise presents a uniform distribution over each swath and is decorrelated between the two swaths.

- **Baseline dilation errors:** relates to uncertainties in the baseline (distance between the two SAR antennas) due to physical changes in the boom on which the antennas are mounted (thermal or mechanical dilation/contraction). Given its nature, it is a correlated error in the cross-track direction.
- **Timing errors:** Associated with uncertainties in the measurement of return times for the interferometry measurements. Given that the look angle variation is small across the swath, a timing error will involve a constant height bias across the swath [59]. The timing error is thus correlated and constant in the across-track direction.

Geophysical errors:

- **Wet troposphere errors:** Involves uncertainties due to delay in the radar signal introduced by water vapor in the atmosphere. It is the major geophysical source of error. Statistically, it is characterized by a uniform phase distribution.
- **Dry troposphere errors:** Involves uncertainties due to delay in the radar signal introduced by 'dry' gases in the troposphere. *This error source is not simulated in [78].*
- **Ionosphere errors:** Involves uncertainties due to delay in the radar signal introduced by electromagnetic dispersion in the ionosphere. *This error source is not simulated in [78].*
- **Sea state bias:** relates to bias introduced by local changes in the electromagnetic reflectivity of the ocean surface due to wave crests and troughs (since reflectivity is higher for wave troughs). *This error source is not simulated in [78].*

Among these sources of errors, KaRIn instrument noise, roll and phase errors and wet troposphere errors are the strongest and most relevant. It should also be noted that many of these error sources are correlated, which introduces additional difficulties and limitations for the exploitation of SWOT observations for the interpolation of high-resolution SLA fields.

6.3 Geophysical field interpolation

We dedicate this section to providing a brief introduction of the most widely used approaches for the interpolation of geophysical fields in the context of oceanography and geosciences applications.

6.3.1 Exemplar-based methods

Initially developed for image processing issues, patch-based and exemplar-based models rely on the representation of images using a dictionary of representative exemplars extracted from the considered dataset. Among these approaches, non-local means and non-local priors [25, 207] have recently known some success in remote sensing applications [63, 177]. Importantly, integrating these approaches into data assimilation formulations led to the development of analog data assimilation [89, 161, 243], which we formulate later within this Section as a data assimilation issue.

6.3.2 Projection-based methods

Projection-based approaches have recently been applied successfully to the interpolation of geophysical fields. They involve the data-driven computation of a set of basis functions providing a low-dimensional representation of the considered dataset, and the interpolation of missing values by exploiting a projection onto the estimated basis functions. Among these approaches, orthogonal projection approaches such as Data Interpolating Empirical Orthogonal Functions (DINEOF) [8] and Variable EOFs DINEOF (VE-DINEOF) [213] have been particularly successful in remote sensing interpolation issues. However, given that orthogonal decompositions may not be fully geophysically-consistent, alternative constraints for projections, such as non-negativity, have also been explored in the context of geophysical field interpolation [169].

6.3.2.1 Multi-scale VE-DINEOF

VE-DINEOF [213] is a state-of-the-art data-driven approach for the interpolation of irregularly-sampled geophysical fields that relies on an EOF [90, 206] decomposition. The principle behind this approach is simple, and involves an iterative procedure in which, at each iteration, missing data values are reconstructed by exploiting an EOF decomposition of the field of interest (with missing values being set to zero for the computation of EOFs). While the number of EOFs considered for the interpolation of missing values is increased sequentially at each iteration in classical DINEOF, VE-DINEOF relies on an improved formulation in which the number of EOFs used for missing data interpolation is optimized at each iteration using cross-validation. Furthermore, in [162] we introduced Multi-scale Variable EOFs DINEOF (MS-VE-DINEOF), a multi-scale, patch-based version of VE-DINEOF. This new formulation was developed to ensure a fair comparison between VE-DINEOF and AnDA in the context of the interpolation of geophysical fields.

6.3.3 Data assimilation

Data assimilation refers, in a general way, to methods aimed at combining the equations governing the behaviour of a dynamical system (which ultimately determine a numerical model)

with observations conveying information related to such dynamical system, which can be either direct observations of the system state variables, or a function of such state variables. The objective is to improve reconstruction/forecasting performance of the numerical model by exploiting the additional information introduced by observations. Formally, a state-space formulation is used [61]:

$$\begin{cases} \mathbf{x}(t) &= \mathcal{M}(\mathbf{x}(t - \delta t)) + \boldsymbol{\epsilon}(t) \\ \mathbf{y}(t) &= \mathcal{H}(\mathbf{x}(t), \Omega(t)) + \boldsymbol{\eta}(t) \end{cases} \quad (6.1)$$

where t is a discrete time index, \mathbf{x} is the hidden state sequence to be reconstructed and \mathbf{y} is the observed data sequence. \mathcal{M} is a dynamical model relating the current state $\mathbf{x}(t)$ to the previous state $\mathbf{x}(t - \delta t)$. \mathcal{H} is an observation operator, where $\Omega(t)$ is a mask accounting for missing data at time t and $\boldsymbol{\epsilon}(t)$ and $\boldsymbol{\eta}(t)$ are random noise processes accounting for modeling and observation uncertainties.

In classical data assimilation, model \mathcal{M} is applied at each time step to produce an initial forecast \mathbf{x}^f . This forecast is then corrected using observation $\mathbf{y}(t)$ to produce an improved forecast, usually referred to as the analysis \mathbf{x}^a . The assimilation of observation $\mathbf{y}(t)$ to produce analysis \mathbf{x}^a can be achieved by means of either stochastic approaches (such as Kalman [127] or particle filters [86]) aiming at producing the maximum a posteriori estimate (i.e. the most probable state) given the current state and observation sequence, or variational approaches (such as 3DVAR [173]) aiming at optimizing a cost function (that penalizes observation error), by means of gradient-descent approaches, to produce an improved prediction \mathbf{x}^a . For cases considering both a linear model $\mathcal{M} = \mathbf{M}$ and a linear observation operator $\mathcal{H} = \mathbf{H}$ subject to centered, Gaussian noises $\boldsymbol{\eta}(t)$ and $\boldsymbol{\epsilon}(t)$, both stochastic and variational data assimilation produce equivalent results, boiling down to an optimal estimation of the analysis $\mathbf{x}^a(t)$ in a minimum-mean-square-error (MMSE) sense. Comparatively, however, stochastic data assimilation approaches have the advantage that, since they are formulated within a probabilistic Bayesian framework, they allow for the characterization of uncertainty, thus not only producing an improved prediction, but also characterizing its probability distribution and allowing for the computation of prediction and model uncertainty.

Data assimilation has been classically used in geosciences for the interpolation of geophysical fields from irregularly-sampled satellite-derived observations, for example sea level anomaly (SLA) fields [153, 204]. Optimal Interpolation (OI), which relies on linearity and Gaussianity hypothesis, is certainly the most widely used framework.

In the remainder of this section, we introduce the most relevant data assimilation methods, namely the classic Kalman Filter, the Ensemble Kalman Filter, introduced as an extension of the classic Kalman Filter for high-dimensional, non-linear systems, and Optimal Interpolation, the most popular state-of-the-art method for the interpolation of geophysical fields.

6.3.3.1 Kalman Filter

The Kalman filter [127] is probably the most widely used approach for data assimilation, despite the fact that it involves considerably restrictive hypotheses, including linearity of both the model operator $\mathcal{M} = \mathbf{M}$ and the observation operator $\mathcal{H} = \mathbf{H}$, and Gaussianity of both the model error $\boldsymbol{\epsilon}(t)$ and the observation error $\boldsymbol{\eta}(t)$, with known covariance matrices \mathbf{B} and \mathbf{R} respectively.

Under these hypotheses, the Kalman filter provides recursive expressions for the characterization of the probability distribution $P(\mathbf{x}(t)|\mathbf{Y}_{1:t})$ of the state $\mathbf{x}(t)$ given the series of past noisy observations $\mathbf{Y}_{1:t} = [\mathbf{y}(1), \dots, \mathbf{y}(t)]$, assuming all model parameters are known. In the conditions given by the Kalman filter hypotheses, $P(\mathbf{x}(t)|\mathbf{Y}_{1:t})$ is a Gaussian distribution, so that the Kalman filter characterizes such distribution by giving expressions for its mean $\hat{\mathbf{x}}_{t|t}$ (i.e. the mean state $\mathbf{x}(t)$ at time t given all previous observations $\mathbf{Y}_{1:t}$) and its associated covariance matrix $\mathbf{P}_{t|t}$.

Methodologically, the Kalman filter resorts to the computation of the MMSE estimation of state $\mathbf{x}(t)$ which, in the context of stochastic data assimilation, is given by the mean state $\hat{\mathbf{x}}_{t|t}$ being determined by the conditional expectation of $\mathbf{x}_t = \mathbf{x}(t)$ given all previous observations $\mathbf{Y}_{1:t}$:

$$\hat{\mathbf{x}}_{t|t} = E[\mathbf{x}_t | \mathbf{Y}_{1:t}] \quad (6.2)$$

Developing on this expression and exploiting the Gaussian assumptions of the Kalman formulation, the Kalman recursion can be derived. We refer the reader to [127] for derivation details. The Kalman recursion involves two steps that are repeated until convergence.

- **Prediction step:** Involves applying model \mathbf{M} to obtain the estimation of forecast $\hat{\mathbf{x}}_{t|t-1}$ given the previous analysis $\hat{\mathbf{x}}_{t-1|t-1}$ and all past observations up to time $t-1$, and the computation of the associated covariance matrix $\mathbf{P}_{t|t-1}$.

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{M}\hat{\mathbf{x}}_{t-1|t-1} \quad (6.3)$$

$$\mathbf{P}_{t|t-1} = \mathbf{M}\mathbf{P}_{t-1|t-1}\mathbf{M}^T + \mathbf{B} \quad (6.4)$$

- **Update step:** Involves exploiting the Gaussian assumptions to compute the Kalman gain \mathbf{K}_t and assimilate current observation \mathbf{y}_t to obtain analysis $\hat{\mathbf{x}}_{t|t}$ and its associated covariance matrix $\mathbf{P}_{t|t}$.

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}^T \left(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T + \mathbf{R} \right)^{-1} \quad (6.5)$$

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t \left(\mathbf{y}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1} \right) \quad (6.6)$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{P}_{t|t-1} \quad (6.7)$$

It should be noted that similar recursions can be defined by considering all past and future observations $\mathbf{Y}_{1:T} = [\mathbf{y}(1), \dots, \mathbf{y}(T)]$, thus allowing for the processing of entire datasets in order to exploit all information within an observation time series (past and future) to improve state dynamics reconstruction. Such schemes are known as Kalman smoothers, and usually involve a forward pass in which past observations are used to correct current state predictions, as in a standard Kalman filter, and a backward pass in which time is inverted and the procedure is repeated backwards, so that future observations are exploited to correct past estimations of the system state.

The inherent limitations of the Kalman filter associated with the strongly restrictive Gaussian, linear hypotheses motivated the development of numerous extensions more suitable to address more complex situations. For the proposed approaches we will exploit in this thesis work, we rely on one of such extensions, namely the Ensemble Kalman Filter [27, 61].

6.3.3.2 Ensemble Kalman Filter

As previously stated, the interpolation of geophysical fields may involve non-linear dynamical models \mathcal{M} . Additionally, given the high-dimensional nature of most geophysical datasets, the use of Kalman filters is usually not possible, given the prohibitive computational cost associated with high-dimensionality, particularly for matrix inversions and storage. To tackle these shortcomings, several extensions of the Kalman filter have been proposed. In geosciences, the Ensemble Kalman filter [27, 60, 61] is particularly appealing given its simplicity and straightforward implementation.

The basic principle behind the Ensemble Kalman filter relies on the Gaussianity hypothesis, particularly for the filtering posterior distribution $P(\mathbf{x}(t)|\mathbf{Y}_{1:t})$. Indeed, the Ensemble Kalman filter relies on Monte Carlo simulations to estimate the mean and covariance of the filtering posterior distribution, as opposed to computing it directly as in a classic Kalman filter. Similarly to a Kalman filter, the Ensemble Kalman filter starts by considering an initialization relating to the initial prior distribution of the system state vector, and then proceeds by sampling an ensemble of Gaussian vectors $\mathbf{x}_i^f(0) \sim \mathcal{N}(\mathbf{x}(0), \mathbf{B}) \forall i = 1, \dots, N$ (called members). It then iterates prediction and update steps exploiting Monte Carlo simulations based on this ensemble:

- **Prediction step:** Involves applying model \mathcal{M} to ensemble members to generate a new ensemble $\mathbf{x}_i^f(t)$ and estimating the forecast $\mathbf{x}^f(t)$ and its covariance matrix $\mathbf{P}^f(t)$ from the new ensemble.

$$\mathbf{x}^f(t) = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i^f(t) \quad (6.8)$$

$$\mathbf{P}^f(t) = \frac{1}{N-1} \sum_{i=1}^N [\mathbf{x}_i^f(t) - \mathbf{x}^f(t)] [\mathbf{x}_i^f(t) - \mathbf{x}^f(t)]^T \quad (6.9)$$

- **Update step:** Involves computing the Kalman gain $\mathbf{K}^a(t)$, generating an ensemble of samples $\mathbf{y}_i^f(t) \sim \mathcal{N}(\mathbf{H}\mathbf{x}^f(t), \mathbf{R}) \forall i = 1, \dots, N$, and exploiting these samples and observations $\mathbf{y}(t)$ to update ensemble members and characterize the filtering posterior distribution $P(\mathbf{x}(t)|\mathbf{Y}_{1:t})$ from the updated ensemble.

$$\mathbf{K}^a(t) = \mathbf{P}^f(t)\mathbf{H}^T [\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}]^{-1} \quad (6.10)$$

$$\mathbf{x}_i^a(t) = \mathbf{x}_i^f(t) + \mathbf{K}^a(t) (\mathbf{y}(t) - \mathbf{y}_i^f(t)) \quad (6.11)$$

$$\mathbf{x}^a(t) = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i^a(t) \quad (6.12)$$

$$\mathbf{P}^a(t) = \frac{1}{N-1} \sum_{i=1}^N [\mathbf{x}_i^a(t) - \mathbf{x}^a(t)] [\mathbf{x}_i^a(t) - \mathbf{x}^a(t)]^T \quad (6.13)$$

As with the Kalman filter, an Ensemble Kalman smoother can be derived by starting at $t = T$ and running a backward pass in time, with the particularity that, for each time t , the update step involves the computation of a Kalman smoother gain $\mathbf{K}^s(t) = \mathbf{P}^a(t)\mathcal{M}^T [\mathbf{P}^f(t+1)]^{-1}$, so that $\mathbf{P}^a(t)\mathcal{M}$ needs to be estimated as the sample covariance of ensemble members for cases where operator \mathcal{M} is not linear [208, 245].

Compared to a classic Kalman filter, it has been proven that for linear dynamical systems \mathbf{M} and observation operators \mathbf{H} with Gaussian uncertainties $\boldsymbol{\epsilon}(t)$ and $\boldsymbol{\eta}(t)$, the Ensemble Kalman Filter converges to the same solution as the Kalman filter. This is, however, not true for cases involving non-linear systems or non-Gaussian uncertainties [146].

6.3.3.3 Optimal Interpolation

Optimal Interpolation (OI), also known as Objective Analysis, is among the most widely used data assimilation techniques in oceanography and meteorology. It was first introduced by Gandin in [141] for meteorological applications and latter applied in the context of physical oceanography in [20]. Numerous currently available ocean datasets, e.g. CMEMS altimetry data products, are produced by exploiting optimal interpolation.

Methodologically, OI relies on a Bayesian formulation to derive the Maximum A Posteriori (MAP) estimate of an irregularly sampled field \mathbf{x} , under the hypothesis that it follows a Gaussian distribution with a constant covariance matrix and that the observation operator $\mathcal{H} = \mathbf{H}$ is linear. In this respect, the length of the considered covariance structure will then represent a trade-off between model locality, relating to fine scale reconstruction, and observation availability, relating to the size of the gaps to be filled. It may be noted that the Kalman smoother provides a sequential solution of the OI problem.

Indeed, under the assumption that model error $\boldsymbol{\epsilon}$ and observation error $\boldsymbol{\eta}$ are uncorrelated, and that their respective covariance matrices \mathbf{B} and \mathbf{R} are known, OI aims at finding the best linear unbiased estimation at each time step t :

$$\mathbf{x}(t) = \mathbf{x}(t - \delta t) + \mathbf{K}(\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t - \delta t)) \quad (6.14)$$

where the objective is to estimate the optimal weight matrix \mathbf{K} . In the context of stochastic data assimilation, and given the Gaussian and linear hypothesis, the optimal weighting matrix \mathbf{K} is given by:

$$\mathbf{K} = E \left[(\mathbf{x}(t) - \mathbf{x}(t - \delta t))(\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t - \delta t))^T \right] E \left[(\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t - \delta t))(\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t - \delta t))^T \right]^{-1} \quad (6.15)$$

$$= E \left[(\boldsymbol{\epsilon}(t)(\boldsymbol{\eta}(t) - \mathbf{H}\boldsymbol{\epsilon}(t))^T) \right] E \left[(\boldsymbol{\eta}(t) - \mathbf{H}\boldsymbol{\epsilon}(t))(\boldsymbol{\eta}(t) - \mathbf{H}\boldsymbol{\epsilon}(t))^T \right]^{-1} \quad (6.16)$$

$$= E \left[\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}(t)^T \right] \mathbf{H}^T \left(E \left[\boldsymbol{\eta}(t)\boldsymbol{\eta}(t)^T \right] + \mathbf{H}E \left[\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}(t)^T \right] \mathbf{H}^T \right)^{-1} \quad (6.17)$$

$$= \mathbf{B}\mathbf{H}^T \left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T \right)^{-1} \quad (6.18)$$

which is in agreement with the result obtained with a Kalman filter in the context of a Gaussian, linear problem.

Even though OI was originally formulated in the context of stochastic data assimilation, it should be noted that, since OI assumes the Gaussian and linear hypotheses to hold true, the same result can be obtained by means of a variational data assimilation techniques minimizing the following cost function:

$$C(\mathbf{x}) = [\mathbf{x}(t) - \mathbf{x}(t - \delta t)]\mathbf{B}^{-1}[\mathbf{x}(t) - \mathbf{x}(t - \delta t)] + [\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t - \delta t)]\mathbf{R}^{-1}[\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t - \delta t)] \quad (6.19)$$

Finally, given that the Gaussian hypothesis involves spatio-temporally invariant mean covariance structures that tend to smooth out fine scale details, it should be expected that OI will not be able to fully exploit high-resolution information available within observation datasets associated with both current and/or future altimetry missions, such as SWOT. In this respect, considerable efforts are being made to improve OI [57] or find alternative approaches [67,162,257]. Escudier et al. [57], for example, propose to improve OI by considering an additional bathymetry constraint. Following these efforts, we will focus in the following chapters on completely data-driven approaches that exploit available observations to emulate dynamical model \mathcal{M} , thus avoiding the use of restrictive model hypothesis, such as spatio-temporally invariant covariance structures.

6.3.4 Analog Forecasting

As previously stated, in the following chapters we will focus on data-driven approaches for the interpolation of geophysical fields. The approaches we will consider for this purpose rely on the

principle of analog forecasting, which we will briefly introduce here.

Analog forecasting is a data-driven forecasting technique first introduced by Lorenz in the context of atmospheric forecasting [175]. The principle behind analog forecasting relies on exploiting an ensemble of past observations of system dynamics \mathbf{x} to obtain a forecast. In this sense, the implementation of analog forecasting is simple and straightforward, and consists in finding similar situations (or analogs) to current state $\mathbf{x}(t)$ within the ensemble of past system dynamics, retrieve the future state associated with the analogs (called the successors) and exploit this information to produce a forecast. Once the analogs and successors have been retrieved, models of varying complexity can be formulated on the base of these analogs and successors.

The two main components involved in an analog forecasting scheme, independently of the complexity of the analog model considered, are the catalog of past observations (or simulations) of state \mathbf{x} and a distance measure. The catalog must be large enough to be representative of the complete system dynamics, so that sufficiently similar analogs can be found, specially given that the probability of finding a perfect analog is small [175]. The choice of the distance measure is also important, with the selection of analogs, and hence the results, being considerably affected by the choice of the distance.

Methodologically, the principle of analog forecasting bears close resemblance to classical approaches such as K nearest neighbours (KNN), non-local means [25, 207] and local linear embedding (LLE) [222]. Similarly to these related techniques, analog forecasting strongly depends on the correct selection of analogs, and is thus considerably sensitive to the curse of dimensionality. In this respect, the exploitation of analog forecasting schemes when high-dimensional data is considered usually requires the previous application of a dimensionality reduction technique, such as PCA [90, 206].

6.3.5 AnDA: The analog data assimilation framework

The analog data assimilation (AnDA) [161] was introduced as a data-driven alternative to classical data assimilation, and should allow us to tackle the fact that classical data assimilation strongly depends on the accuracy and computational efficiency of numerical model \mathcal{M} . AnDA is based on the hypothesis that dynamical model \mathcal{M} can be emulated by exploiting a catalog of realistic observations/simulations of the dynamical system.

In this way, at each time step, dynamical model \mathcal{M} is replaced by a data-driven numerical approximation, thus replacing the model-based forecasting step with a data-driven alternative.

Formally, this boils down to the following state-space formulation:

$$\begin{cases} \mathbf{x}(t) &= \mathcal{F}(\mathbf{x}(t - \delta t)) + \boldsymbol{\epsilon}(t) \\ \mathbf{y}(t) &= \mathcal{H}(\mathbf{x}(t), \Omega(t)) + \boldsymbol{\eta}(t) \end{cases} \quad (6.20)$$

where t is a discrete time index, \mathbf{x} is the hidden state sequence to be reconstructed and \mathbf{y} is the observed data sequence. \mathcal{H} is an observation operator, where $\Omega(t)$ is a mask accounting for missing data at time t and $\boldsymbol{\epsilon}(t)$ and $\boldsymbol{\eta}(t)$ are random noise processes accounting for modeling and observation uncertainties. Compared to classical data assimilation model (6.1), this new formulation introduces \mathcal{F} , an analog forecasting operator that exploits analog forecasting techniques to emulate dynamical model \mathcal{M} in formulation (6.1). Specifically, for current system state \mathbf{x} , and assuming a catalog of past state dynamics large enough to accurately depict the underlying processes is available, operator \mathcal{F} is built from the K most similar states to \mathbf{x} within the catalog, referred to as analogs. Given that the future state of these analogs, known as successors, is known and exists within the catalog, this information can be used to adjust a forecasting model that can be subsequently applied to the current state \mathbf{x} to obtain forecast \mathbf{x}^f . In this respect, the procedure relies on the similarity between the current state \mathbf{x} and simulated states stored in the catalog.

In the context of stochastic data assimilation, the analog forecasting operator \mathcal{F} involves the data-driven characterization of the transition probability between analogs and successors $p(\mathbf{x}(t)|\mathbf{x}(t - \delta t))$, so that the analog forecast \mathbf{x}^f is obtained by sampling from the estimated distribution. Assuming a Gaussian prior $\mathbf{x}(t)|\mathbf{x}(t - \delta t) \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$, as in [161], the mean $\boldsymbol{\mu}_t$ and covariance matrix $\boldsymbol{\Sigma}_t$ are estimated from the K nearest neighbours of the current system state within the catalog. Multiple strategies can be used for the estimation of these parameters, and we refer the reader to [67, 161, 162] for an in-depth discussion. Here, we shall consider a locally-linear model that involves fitting a weighted linear regression between the K nearest analogs and their successors. The fitted linear regression is then applied to the current state to obtain mean $\boldsymbol{\mu}_t$. Covariance matrix $\boldsymbol{\Sigma}_t$ then comes to the empirical covariance of the residuals of the fitted linear regression. In this formulation, linear regression weights are introduced to account for the distance between the current state and its analog within the catalog, so that closer analogs will have a stronger contribution to the estimation of the forecasting linear regression. To this end, following [162], a Gaussian kernel, normalized by the median distance between analogs and the current system state, is used.

Once this step is completed, analog data assimilation proceeds similarly to classical data assimilation and exploits a stochastic scheme to combine forecast \mathbf{x}^f with observation \mathbf{y} and produce the analysis \mathbf{x}^a . In our current implementation, we consider an Ensemble Kalman filter/smoothing (EnKFS) to assimilate partial observations sampled at a δt time step.

The application of the proposed framework to the reconstruction of high-resolution geophysical fields from irregularly-sampled data requires, however, the implementation of a number of strategies to deal with both the two-dimensional geometry as well as the high-dimensional nature of the data, as discussed in the following section.

6.3.5.1 Multi-scale patch-based analog data assimilation for high-dimensional geophysical fields

6.3.5.2 Multi-scale decomposition

Given that we focus on resolving smaller scales not resolved by current methods, and that OI already provides an efficient way of estimating large scale dynamics, AnDA follows classical scale space decomposition principles [182] and exploits a multi-scale decomposition:

$$\mathbf{x} = \bar{\mathbf{x}} + d\mathbf{x} + \zeta \quad (6.21)$$

so that larger scales $\bar{\mathbf{x}}$ are resolved by means of an OI scheme, while the reconstruction of the anomaly field $d\mathbf{x}$ relies on the proposed analog data assimilation, with unresolved scales being represented by ζ .

6.3.5.3 Patch-based state space formulation

Following recent advancements in image processing [63,207] and data assimilation [67] issues, the two-dimensional nature of the considered data is tackled by extracting and vectorizing $W_p \times W_p$ overlapping patches of anomaly field $d\mathbf{x} = \mathbf{x} - \bar{\mathbf{x}}$. We consider overlapping patches, so that the analog data assimilation is thus performed independently at the patch level, with the final anomaly field reconstruction being obtained by means of a simple overlapping patch averaging scheme. We denote a patch centered at position s at time t as $\mathcal{P}(s, t)$.

6.3.5.4 EOF-based representation of patches

Depending on the chosen patch size W_p , the vectorized anomaly patches $\mathbf{p}(s, t) = \text{vec}(\mathcal{P}(s, t))$ may probably involve a high number of dimensions. This is particularly problematic for the proposed framework, given that the analog forecasting is highly sensitive to the curse of dimensionality, as it relies on a K nearest neighbour search, with the relevance of the neighbour selection being highly degraded as dimension increases. To deal with this drawback, an EOF-based decomposition [90] of vectorized anomaly patches $\mathbf{p}(s, t)$ is used:

$$\mathbf{p}(s, t) = \sum_{k=1}^{N_{EOF}} \alpha_k(s, t) B_k \quad (6.22)$$

where B_k are the EOF basis functions learned from the ensemble of all patches $\mathbf{p}(s, \cdot)$ at location s , and $\alpha_k(s, t)$ are the coefficients of the decomposition of patch $\mathbf{p}(s, t)$, at location s and

time t , onto these EOF basis functions. At each location, only the N_{EOF} strongest EOF basis functions are retained, so that analog data assimilation is subsequently performed on coefficients $\alpha_k(s, t)$ representing each patch decomposition onto its locally computed EOF basis functions. In this way, analog forecasting (including the nearest neighbour search) and data assimilation are performed on the projection of patches $\mathbf{p}(s, t)$ onto a lower-dimensional space of their EOF-based representation, given by coefficients $\alpha_k(s, t)$, rather than on the high-dimensional space of patches $\mathbf{p}(s, t)$.

It should be noted that even though the original AnDA framework exploits PCA-based decompositions for the dimensionality reduction of vectorized patches, it is possible to consider alternative decomposition models exploiting different constraints, such as sparsity or non-negativity. In this respect, depending on the dataset considered, alternative constraints may indeed allow for a more performant decomposition involving fewer dimensions or a more interpretable decomposition giving coefficients $\alpha_k(s, t)$ a physical spatio-temporal signification. To this end, one may indeed exploit blind source separation models and algorithms similar to those introduced in the first part of this thesis dissertation.

6.4 Conclusion

In this chapter, we presented the context and set the methodological bases for the problem of the interpolation of high-dimensional geophysical fields, and more particularly of SLA fields, from irregularly sampled satellite-derived data.

We briefly explained the state-of-the-art methods used to tackle this problem, their basic working principles and limitations, and illustrated the need for more data-driven approaches in this era of big data pouring from space. We also introduced the two main type of altimetry data we will consider and explained their characteristics and fundamental differences. Importantly, we will focus, on this second part of this thesis dissertation, on wide-swath altimeter data, with a view towards the future SWOT mission to be launched in 2021.

In the next chapters, we will build on the concepts presented here to tackle the data-driven interpolation of SLA fields from satellite-derived altimetry data. In particular, we will exploit and compare approaches based on both non-negative decomposition models for convolutional operators, as introduced in [66, 169] and presented in the first part of this work, and on developments recently introduced in [67, 161, 162] for the data-driven analog data assimilation of irregularly-sampled observations and its application to the interpolation of high-dimensional geophysical fields.

Locally-adapted convolutional models for the reconstruction of SLA fields from satellite along-track data

"Think global, act local."

Sir Patrick Geddes

7.1	Introduction	129
7.2	Model formulation	131
7.2.1	Problem statement	131
7.2.2	Unconstrained model calibration	132
7.2.3	Dictionary-based decompositions	133
7.2.4	Locally-adapted dictionary-based convolutional models	134
7.2.5	Data and Methodology	135
7.2.6	Results	137
7.3	Conclusion	138

7.1 Introduction

In this chapter, we focus on the application of the models and methods introduced in the first part of this thesis to the problem of the data-driven interpolation of SLA fields from irregularly-sampled satellite-derived observations.

To this end, we follow [63,66,169] and reformulate the problem as an image super-resolution issue and exploit the proposed decomposition models for the decomposition of super-resolution

convolutional models. The decomposition is exploited as a means to locally-adapt the super-resolution operator to smaller spatial domains, thus increasing model locality and reconstruction performance. We also explore alternative constraints, namely orthogonality and sparsity, to illustrate the advantages of the non-negative formulation in this particular context.

As illustrated in our analysis of the state-of-the-art methods relating to this work, image super-resolution or upscaling appears as a classical problem in image processing, and has been extensively studied in the literature [82, 231]. Recent advancements in machine learning and the advent of ever more powerful computational tools make super-resolution models particularly appealing for complex problems, such as those arising from geosciences, oceanography and meteorology. In this respect, super-resolution techniques have also been applied to remote sensing image enhancement problems [272].

Contrary to the classical image super-resolution setting, numerous satellite remote sensing applications do not only involve low-resolution images but also irregularly-sampled high-resolution information. As illustrated before, the later may be due to specific sampling patterns, such as along-track narrow-swath satellite data, as well as to partial occlusions caused by weather conditions [64, 80]. In this respect, the availability of such partial high-resolution data supports locally-adapted super-resolution models, rather than models fully trained offline, with a view to accounting for the space-time variabilities of the monitored processes.

In the following sections, we address image super-resolution issues from irregularly-sampled high-resolution information. Following state-of-the-art super-resolution models [2, 252, 253], we consider locally-adapted convolution-based models. Our methodological contributions are two-fold:

1. The proposed convolution-based models combine both a low-resolution image and a secondary image source, thus allowing for the fusion of heterogeneous data sources within a simple, unified super-resolution framework.
2. We explore dictionary-based representations of the convolutional operators with different types of constraints, namely orthogonality, non-negativity and sparsity constraints [12, 273]. Such dictionary-based representations and constraints are particularly appealing to resort to locally-adapted super-resolution models calibrated from a low number of high-resolution training data.

As case study, we apply the proposed framework to multi-source ocean remote sensing data, namely the reconstruction of high-resolution SLA (Sea Level Anomaly) images from satellite-derived along-track altimeter data, a high-resolution SST (Sea Surface Temperature) image and a low-resolution SLA image. We report numerical experiments, which demonstrate the relevance of the proposed super-resolution models, especially under non-negativity constraints, compared with classical approaches such as OI.

7.2 Model formulation

7.2.1 Problem statement

In classical image super-resolution problems, we aim at reconstructing a series of high-resolution images $\{\mathbf{Y}(t)\}_t$ at different times $\{t_1, \dots, t_T\}$ from the corresponding series of low-resolution images $\{\mathbf{Y}_{LR}(t)\}_t$. Additionally, in the considered application setting, we are also provided with:

- A complementary source of high-resolution images $\{\mathbf{X}(t)\}_t$, which may depict some local or global correlation with $\{\mathbf{Y}(t)\}_t$;
- An irregularly-sampled dataset of high-resolution point-wise observations $\{\tilde{t}(k), \tilde{s}(k), \tilde{\mathbf{Y}}(k)\}_k$, with $\tilde{t}(k)$, $\tilde{s}(k)$ and $\tilde{\mathbf{Y}}(k)$ respectively the time, location and value of the k^{th} high-resolution observation.

In the context of the interpolation of SLA fields from satellite-data, the irregularly-sampled dataset corresponds to along-track observations of the SLA fields, as observed in Figure 7.1, which presents an example of the considered along-track sampling patterns.

Following [63, 66, 169], the reconstruction of high-resolution image $\mathbf{Y}(t)$ given low-resolution image $\mathbf{Y}_{LR}(t)$ is stated according to the following convolution-based model:

$$\mathbf{Y}(t) = \mathbf{Y}_{LR}(t) + \mathbf{H}_Y * \mathbf{Y}_{LR}(t) + \mathbf{H}_X * \mathbf{X}(t) + \mathbf{N}(t) \quad (7.1)$$

where $\mathbf{N}(t)$ is a space-time noise process. \mathbf{H}_Y (resp. \mathbf{H}_X) is the two-dimensional impulse response of the Y_{LR} (resp. X) component of the proposed convolutional model. \mathbf{H}_Y and \mathbf{H}_X are characterized by $(2W_p + 1) \times (2W_p + 1)$ discrete representations onto the considered high-resolution grid. Importantly, \mathbf{H}_Y and \mathbf{H}_X are space-and-time-varying operators and capture the space-time variabilities of $(\mathbf{Y}, \mathbf{Y}_{LR})$ and (\mathbf{Y}, \mathbf{X}) relationships. This model can be regarded as a patch-based super-resolution approach where high-resolution image \mathbf{Y} at a given location is computed as a linear combination of $(2 * W_p + 1) \times (2 * W_p + 1)$ patches of images \mathbf{X} and \mathbf{Y}_{LR} centered at the same location. Parametrization $\mathbf{H}_X = 0$ clearly relates to regression-based super-resolution models, such as JOR [41], ANR [252] and A+ [253].

From a physical point of view, considering linear couplings between different ocean dynamical tracers relates to recent studies that demonstrate the existence of the SQG mode, characterized by a linear transfer function between SST and SSH [97, 99, 115, 131, 132, 142, 143, 244]. Indeed, as indicated in Chapter 1, recent work points out that upper ocean dynamics may indeed be characterized by local SSH-SST linear relationships that correspond precisely to fractional Laplacian operators [97, 115, 132, 142, 143, 244]. In this respect, operators \mathbf{H}_Y and \mathbf{H}_X may involve numerical approximations of theoretical fractional Laplacian operators. Moreover, the proposed linear

coupling between low and high-resolution sources of information directly relates to inter-scale interactions observed in upper-ocean dynamics, with turbulent flow allowing for the transport of energy between multiple spatio-temporal scales.

7.2.2 Unconstrained model calibration

The calibration of model (7.1) amounts to the estimation of the $(2W_p + 1) \times (2W_p + 1)$ matrix representations of operators \mathbf{H}_Y and \mathbf{H}_X at any space-time location. The availability of the irregularly-sampled dataset $\{\tilde{t}(k), \tilde{s}(k), \tilde{Y}(k)\}_k$ provides the means for this locally-adapted calibration. It may be noted that, in classical image super-resolution issue, such models are trained offline or involve nearest-neighbor techniques using a training dataset of joint low-resolution and high-resolution image patches [252, 253], which closely relates to analog forecasting methods [175], non-local means [25, 207] and local linear embedding [222].

Following these approaches, we proceed as follows. For a given space-time location (t_0, s_0) , we regard all data such that $\tilde{t}(k) \in [t_0 - D_t, t_0 + D_t]$ and $\|\tilde{s}(k) - s_0\|_2^2 \leq D_s$ as observations for model (7.1) at location (t_0, s_0) . Parameters D_t and D_s state respectively the spatio-temporal extent of the considered neighborhood around location (t_0, s_0) . This principle of accumulating observations on a spatio-temporal neighbourhood around the location and time of interest relates to the idea of considering fixed spatial and temporal correlation scales in the context of stochastic data assimilation, as is done, for example, in OI. Moreover, this idea is a key concept further explored, in the next chapter, for the construction of pseudo-observations in the context of a completely data-driven data assimilation framework. Given the irregular sampling of the high-resolution dataset, no guarantees exist that sampling locations $\tilde{s}(k)$ will lie within the considered $\mathbf{X}/\mathbf{Y}_{LR}$ grid, and thus $(2W_p + 1) \times (2W_p + 1)$ high-resolution \mathbf{X} patches and low-resolution \mathbf{Y}_{LR} patches might need to be interpolated around spatio-temporal locations $(\tilde{s}(k), \tilde{t}(k))$ to build a coherent dataset for model calibration.

Irregularly-sampled observations $\tilde{Y}(k)$ and \mathbf{X} and \mathbf{Y}_{LR} patches interpolated around sample locations are then exploited to fit local impulse responses \mathbf{H}_X and \mathbf{H}_Y using a least squares approach, which comes to minimize the mean square reconstruction error $\mathcal{E}(\mathbf{H}_X, \mathbf{H}_Y)$ for the high-resolution detail $\delta\mathbf{Y} = \mathbf{Y} - \mathbf{Y}_{LR}$ at irregularly-sampled dataset positions $(\tilde{s}(k), \tilde{t}(k))$:

$$\mathcal{E}(\mathbf{H}_X, \mathbf{H}_Y) = \sum_k \left\| \delta\tilde{Y}(k) - \widehat{\delta\tilde{Y}}(k) \right\|^2 \quad (7.2)$$

where

$$\widehat{\delta\tilde{Y}}(k) = \mathbf{H}_Y * \mathbf{Y}_{LR}(\tilde{t}(k), \tilde{s}(k)) + \mathbf{H}_X * \mathbf{X}(\tilde{t}(k), \tilde{s}(k)) \quad (7.3)$$

Assuming the number of available observations is sufficiently large, minimization (7.2) resorts to a least-square estimation of operators \mathbf{H}_Y and \mathbf{H}_X .

7.2.3 Dictionary-based decompositions

A critical aspect of the above least-square minimization is the number of available training data points and the underlying balance between locally-adapted and robust parametrizations. With a view to improving estimation robustness as well model interpretability, we explore dictionary-based decomposition approaches. They resort to the following decomposition of operators $\mathbf{H}_{\mathbf{X}}$ and $\mathbf{H}_{\mathbf{Y}}$:

$$\mathbf{H}_{\{\mathbf{X}, \mathbf{Y}\}} = \sum_{k=1}^K \alpha_k \mathbf{D}_k^{\{\mathbf{X}, \mathbf{Y}\}} \quad (7.4)$$

where $\mathbf{D}_k^{\mathbf{Y}}$ (resp. $\mathbf{D}_k^{\mathbf{X}}$) is the k^{th} component of the dictionary of operators for operator $\mathbf{H}_{\mathbf{Y}}$ (resp. $\mathbf{H}_{\mathbf{X}}$) and α_k is the k^{th} scalar coefficient that states the decomposition of operator $\mathbf{H}_{\mathbf{Y}}$ (resp. $\mathbf{H}_{\mathbf{X}}$) onto dictionary element $\mathbf{D}_k^{\mathbf{Y}}$ (resp. $\mathbf{D}_k^{\mathbf{X}}$). It should be noted that a joint dictionary-based representation is considered, so that decomposition coefficients α_k are shared by the two convolutional operators $\mathbf{H}_{\mathbf{Y}}$ and $\mathbf{H}_{\mathbf{X}}$. In this respect, the proposed dictionary-based convolutional operator decomposition closely relates to the operator decomposition models and algorithms presented in the first part of this thesis dissertation, and most notably to the dictionary-based decomposition of local linear operators (LLOD), as convolution operators $\mathbf{H}_{\{\mathbf{X}, \mathbf{Y}\}}$ can be regarded as locally estimated linear operators Θ_n (as defined in Equation (3.27)) to be decomposed onto a set of representative dictionary elements $\mathbf{D}_k^{\{\mathbf{X}, \mathbf{Y}\}}$ under adequately selected constraints. In particular, the selection of the nearest neighbours for the computation of local linear operators Θ_n (Equation (3.27)) in this context responds to a spatio-temporal locality criteria defined by spatio-temporal extent parameters D_t and D_s , which delimit the neighbourhood within which nearest neighbours will be selected.

Following classical dictionary-based settings [199], we explore the application of our previously introduced decomposition model (3.28) for convolution operators. We investigate three different types of constraints for dictionary elements $\{\mathbf{D}_k^{\{\mathbf{X}, \mathbf{Y}\}}\}$ and decomposition coefficients $\{\alpha_k\}$, namely orthogonality, sparsity and non-negativity constraints. Following local linear operator decomposition model (3.28), the calibration of these dictionary-based settings first involves the estimation of dictionary elements $\{\mathbf{D}_k^{\{\mathbf{X}, \mathbf{Y}\}}\}$ using training data. We here assume we are provided with a previously computed representative dataset of unconstrained estimates of operators $\mathbf{H}_{\mathbf{Y}}$ and $\mathbf{H}_{\mathbf{X}}$ from Equation (7.2), denoted by $\{\mathbf{H}_{\mathbf{Y}}^n, \mathbf{H}_{\mathbf{X}}^n\}_n$. More precisely, the considered dictionary-based decompositions are as follows:

- **Orthogonality constraint:** under this constraint, dictionary elements $\{\mathbf{D}_k^{\{\mathbf{X}, \mathbf{Y}\}}\}$ form an orthonormal basis with no other constraints onto coefficients $\{\alpha_k\}$. This decomposition relates to the application of principal component analysis (PCA) [206] to dataset $\{\mathbf{H}_{\mathbf{Y}}^n, \mathbf{H}_{\mathbf{X}}^n\}_n$. Given the trained dictionaries, the estimation of decomposition coefficients $\{\alpha_k\}$ comes to the projection of the unconstrained operator estimates onto dictionary elements $\{\mathbf{D}_k^{\{\mathbf{X}, \mathbf{Y}\}}\}$.

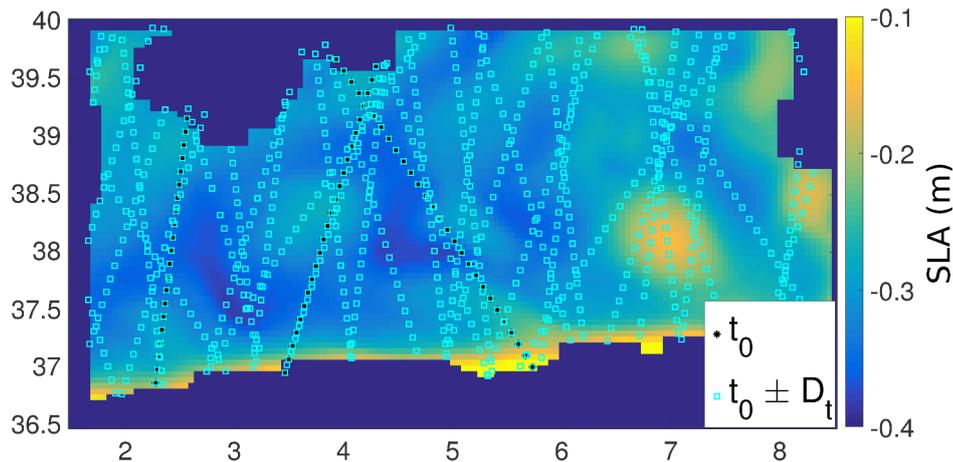


Figure 7.1 – Illustration of the irregular sampling of high-resolution observations associated with ocean remote sensing data: sea level anomaly image with the sampled along-track positions by satellite altimeters (cyan squares) in a ± 10 -day time window around April 20th, 2012. Adapted from [169]. © 2017 IEEE.

- **Sparsity constraint:** the sparse dictionary-based decomposition [3] resorts to complementing Mean Squared Error (MSE) criterion (7.2) with the ℓ_1 -norm of coefficients $\{\alpha_k\}$. We apply a KSVD scheme to dataset $\{\mathbf{H}_Y^n, \mathbf{H}_X^n\}_n$ to train dictionary elements $\{\mathbf{D}_k^{\{\mathbf{X}, \mathbf{Y}\}}\}$. Given the trained dictionaries, we proceed similarly to KSVD and use orthogonal matching pursuit [205] for the sparse estimation of decomposition coefficients $\{\alpha_k\}$ for any new unconstrained operator estimate.
- **Non-negativity constraint:** the non-negative dictionary-based decomposition constrains coefficients $\{\alpha_k\}$ to be non-negative. Given dataset $\{\mathbf{H}_Y^n, \mathbf{H}_X^n\}_n$, the training of dictionary elements $\{\mathbf{D}_k^{\{\mathbf{X}, \mathbf{Y}\}}\}$ resorts to the minimization of reconstruction error (7.2) under non-negativity constraints for the decomposition coefficients. This comes to the application of the LLOD algorithm as it was originally defined in Chapter 3. Following this formulation, we exploit the iterative proximal operator-based algorithm [37] defined by Equations (3.37), (3.38), (3.39) and (3.40). In this respect, given the trained dictionary, the estimation of decomposition coefficients $\{\alpha_k\}$ comes to a least-square estimation under non-negativity constraints.

7.2.4 Locally-adapted dictionary-based convolutional models

The application of the proposed dictionary-based decompositions to the super-resolution of irregularly-sampled high-resolution images involves the following steps. For a given dictionary-based decomposition, we first train the associated dictionaries $\{\mathbf{D}_k^{\mathbf{X}}, \mathbf{D}_k^{\mathbf{Y}}\}$. Considering the entire image time series, we proceed to the unconstrained estimation of operators \mathbf{H}_X and \mathbf{H}_Y from Equation (7.2) for a variety of spatio-temporal neighborhoods with given parameters D_s^{Tr} and D_t^{Tr} . Parameters D_s^{Tr} and D_t^{Tr} are set such that the number of high-resolution observations is high enough to solve for least-square criterion (7.2). We typically sample around 1500 neighbor-

Table 7.1 – Normalized root mean square reconstruction error (nRMSE) for daily high-resolution SLA images $\{Y(t)\}_t$, for a global convolutional model and for locally-adapted decompositions of a global convolutional model using principal component analysis (PCA) [206], KSVD [3] and non-negative decomposition (NN), considering $K = 2$, $K = 5$ and $K = 10$ classes. The nRMSE value for daily low-resolution SLA images $\{Y_{LR}(t)\}_t$ is given as reference (noted as SLA_{LR}). Best results for each number of classes K considered are presented in bold. Results that outperform a global convolutional model are underlined.

	$K = 2$	$K = 5$	$K = 10$
PCA	0.1807	<u>0.1734</u>	<u>0.1680</u>
KSVD	0.2228	0.2228	0.2228
NN	0.1807	<u>0.1734</u>	<u>0.1666</u>
Global model			0.1755
SSH_{LR}			0.2228

hoods to build a representative dataset of operators \mathbf{H}_X and \mathbf{H}_Y .

Given the trained dictionaries, we proceed to the super-resolution of an image at a given date t^* as follows. For any given spatial location s^* , we first estimate the associated decomposition coefficients $\{\alpha_k\}$ from the subset of high-resolution observations in a spatio-temporal neighborhood of space-time location (t^*, s^*) with parameters D_s^{SR} and D_t^{SR} . The later parameters typically define smaller spatio-temporal neighborhoods than training neighborhoods with parameters D_s^{Tr} and D_t^{Tr} . As such, estimated coefficients $\{\alpha_k\}$ come to the projection of more local convolutional operators onto the subspace spanned by the estimated dictionaries, thus yielding a more locally-adapted model (7.1) This calibrated model is then applied to the reconstruction of image \mathcal{Y} in a neighborhood of location (t^*, s^*) . In this way, the dictionary decomposition provides the means to reduce the dimensionality of the problem by projecting local convolutional operators $\{\mathbf{H}_Y^n, \mathbf{H}_X^n\}_n$ onto the subspace spanned by dictionary elements $\{\mathbf{D}_k^{\{X,Y\}}\}$, which allows for the accurate estimation of more local convolutional operators, since less observations are necessary for model calibration in the lower-dimensional subspace. To reduce the computational time, we perform the calibration of locally-adapted models for a regular subsampling of the image grid, typically $D_s^{SR}/2$, and use a spatial averaging of overlapping local reconstructions to obtain a single high-resolution reconstruction of image \mathcal{Y} .

7.2.4.1 Experiments

7.2.5 Data and Methodology

As case study, we consider an application to ocean remote sensing data, more particularly to the reconstruction of sea-surface height (SLA) image time series from along-track altimeter data. As previously explained, satellite altimeters are narrow-swath sensors such that high-resolution altimeter data is only acquired along the satellite track path [218], resulting in an particularly scarce and irregular sampling of the ocean surface as illustrated in Figure 7.1. Interestingly, numerous studies have pointed out the potential contribution of high-resolution sea surface

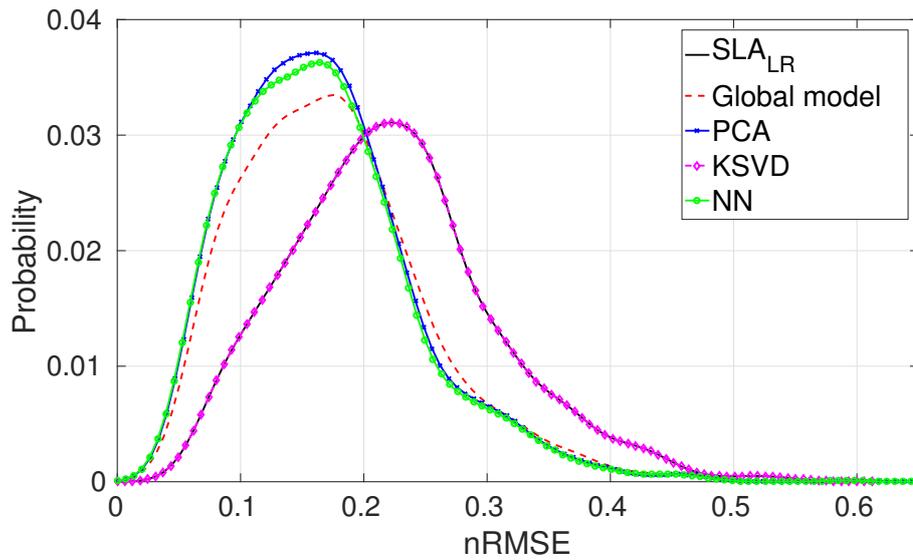


Figure 7.2 – Probability distribution for the normalized root mean square reconstruction error (nRMSE) for daily high-resolution SLA images $\{Y(t)\}_t$, for a global convolutional model and for locally-adapted decompositions of a global convolutional model using principal component analysis (PCA) [206], KSVD [3] and non-negative decomposition (NN) and considering $K = 10$ classes. The probability distribution of the nRMSE for daily low-resolution SLA images $\{Y_{LR}(t)\}_t$ is given as reference (noted as SLA_{LR}).

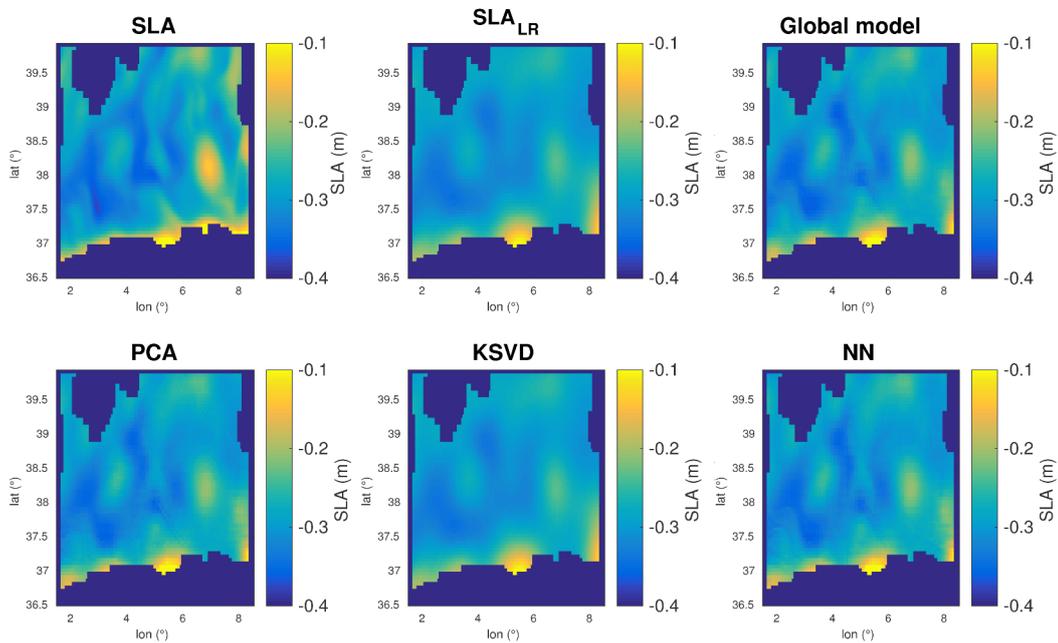


Figure 7.3 – High-resolution SSH image Y reconstruction, April 20th, 2012: first row, from left to right, real high-resolution SSH image Y , low-resolution SSH image Y_{LR} (noted as SSH_{LR}), reconstruction of high-resolution SSH image Y using global convolutional model (7.1); second row, reconstruction of high-resolution SSH image Y using a 10-class locally-adapted decomposition (7.4) of global convolutional model (7.1) using, from left to right, principal component analysis (PCA) [206], KSVD [3] and non-negative decomposition (NN). Adapted from [169]. © 2017 IEEE.

temperature (SST) images to the reconstruction of SSH images (and hence of SLA images as well), as they share common geometrical patterns associated with the underlying upper ocean dynamics [131, 143]. In addition, optimally-interpolated products [218] provide a low-resolution reconstruction of the SLA image. Overall, the reconstruction of high-resolution SLA image time series resorts to a super-resolution issue from irregularly-sampled high-resolution information as stated in Section 7.2. It may be stressed that this case study involves a scaling factor of about 10 between the low-resolution and high-resolution data, which makes it particularly challenging compared with classical image super-resolution issues.

In our experiments, we exploit a ground-truth dataset issued from an observing system simulation experiment (OSSE) for a case study region in the Western Mediterranean Sea ($36.5^\circ N$ to $40^\circ N$, $1.5^\circ E$ to $8.5^\circ E$). A high-resolution numerical simulation of the Western Mediterranean Operational Forecasting System (WMOP) model [124] is used to generate daily high-resolution SLA and SST images from 2009 to 2013 for a $1/20^\circ$ grid. The along-track dataset is simulated by sampling the SLA images at real along-track positions issued from multiple altimetry missions in 2014 and 2015 (see Figure 7.1). Given the simulated along-track dataset, optimally-interpolated SLA fields [218], referred to as low-resolution SLA images Y_{LR} , are computed for a $1/8^\circ$ grid resolution. The calibration of the proposed convolutional operators is performed by considering $W_p = 1$, which corresponds to 3×3 convolutional masks. We use the following parameter setting for spatio-temporal neighborhoods: $t_0 \pm D_t$ -day time windows with $D_t = 10$, and $D_s \times D_s$ spatial neighborhoods with $D_s^{Tr} = 7^\circ$ for the training step and $D_s = 2^\circ$ for the locally-adapted calibration steps.

7.2.6 Results

In Table 7.1, we report the average normalized root mean square reconstruction error (nRMSE) for daily high-resolution SLA images $\{Y(t)\}_t$, for a global convolutional model and for locally-adapted convolutional models, using principal component analysis (PCA) [206], KSVD [3] and non-negative dictionary-based decomposition (NN) and considering $K = 2$, $K = 5$ and $K = 10$ elements in the dictionaries. The reconstruction nRMSE for daily low-resolution SLA images $\{Y_{LR}(t)\}_t$ (noted as SLA_{LR}) is given as reference.

From Table 7.1, locally-adapted convolutional models clearly outperform global models for $K \geq 5$ (with the exception of the KSVD-based decomposition), which can be explained by the improved local adaptation to local spatio-temporal variabilities through locally-adapted decomposition coefficients. In this respect, the non-negative decomposition outperforms alternative approaches, with a **maximum relative gain** (with respect to optimally-interpolated low-resolution SSH images $\{Y_{LR}(t)\}_t$, at $K = 10$) of **25.22% for NN, 24.60% for PCA and 21.23% for a global convolutional model**.

These results are further illustrated by the reconstruction of high-resolution SLA image Y for sample date April 20th, 2012 presented in Figure 7.3 and by the probability distributions of daily reconstruction normalized root mean square error for high-resolution SLA images $\{Y(t)\}_t$, computed for the global convolutional model and for each one of the considered locally-adapted models with $K = 10$, presented in Figure 7.2. Visually, the proposed super-resolution models clearly improve the reconstruction of finer-scale details compared to the low-resolution image. The model using non-negativity constraints seems to involve slightly sharper gradients compared with the unconstrained model. The PCA-based model appears visually less relevant, while the KSVD-based model seems unable to exploit the high-resolution information sources to enhance the low-resolution altimetry field. We hypothesize that the low performance of KSVD may relate to the low number of dictionary elements considered, as KSVD relies on the over-completeness of the dictionary to produce adequate sparse representations. In this way, KSVD may require a much larger number of dictionary elements to achieve good reconstruction performance, which would go in detriment of the dimensionality reduction introduced by the dictionary-based decomposition that ultimately enables the local readjustment of the convolutional model.

7.3 Conclusion

In this chapter, we explored the application of operator decomposition models introduced in Chapter 3 to the interpolation high-resolution of SLA fields from irregularly-sampled along-track satellite data.

Additionally, we addressed the multi-modal super-resolution of irregularly-sampled high-resolution images by considering an additional high-resolution source of information, namely the SST field. Data fusion arises as an issue in a number of remote sensing applications, where several sensors associated with different regular and irregular sampling patterns may contribute to the reconstruction of a given high-resolution image, which makes the proposed formulation particularly relevant for real-world applications.

As a case study, we considered the reconstruction of high-resolution SLA images in the Western Mediterranean Sea, and exploited the WMOP model alongside with an observing system simulation experiment (OSSE) to evaluate the performance of the proposed dictionary-based formulation under different constraints, namely orthogonality, sparsity and non-negativity. From a methodological point of view, we complement previous convolution-based super-resolution models [2, 253] with the evaluation of different constrained dictionary-based decompositions and the use of a complementary high-resolution image source.

In this respect, dictionary-based decompositions are regarded as a means to better account for spatio-temporal variabilities through more locally-adapted model calibrations. Moreover, our numerical experiments support the selection of non-negativity constraints to achieve a better

local adaptation. They demonstrate the relevance of the proposed approach to achieve a better reconstruction of higher-resolution details, compared with the optimally-interpolated fields. In this respect, reported methods have led to relevant results, including the specific application demonstrated in this chapter, which have been presented in different national and international conferences [62, 167, 169].

As far as possible methodological improvements are concerned, future work should include non-local extensions of the proposed model to combine spatio-temporal and similarity-based neighborhoods as considered in regression-based super-resolution models [2, 253]. Non-linear dictionary-based decomposition seems particularly appealing to combine non-linear mapping, for instance convolutional neural networks (CNN) based models [49], and locally-adapted models. As far as ocean remote sensing applications are considered, applying the proposed models to different sampling patterns, for instance along-track narrow-swath satellite data vs. wide-swath satellite data, appears to be of interest, the later possibly enabling the modeling of higher-order geometrical details.

Despite the relevance of the formulation developed in this chapter for the interpolation of high-resolution SLA fields from satellite-derived observations, in the next chapter we shift our focus away from non-negative operator decomposition issues to explore interpolation issues exploiting the Analog Data Assimilation, a promising alternative method introduced in [67, 161, 162] for the interpolation of high-resolution geophysical fields from irregularly-sampled data. Importantly, the Analog Data Assimilation is formulated as a stochastic data assimilation framework, which allows it to explicitly model intrinsic temporal relationships existing within the exploited datasets. Subsequently, we will explore such formulations and compare them to the high-resolution interpolation methods introduced in this chapter. In particular, we will also focus on exploiting SWOT synthetic observations, in preparation for the mission launch in 2021.

Data-driven fusion of nadir along-track and SWOT data

"The hardships that I encountered in the past will help me succeed in the future."

Philip Emeagwali

8.1	Introduction	142
8.2	Benchmarked methods	144
8.3	Case study and data	145
8.3.1	Data	145
8.3.2	Observing System Simulation Experiment	145
8.3.3	Pseudo-observations	146
8.4	Method comparison and benchmarking	146
8.5	AnDA Parameter Sensitivity Analysis	148
8.5.1	Patch size	149
8.5.2	Number of neighbours	149
8.5.3	Assimilation lag	150
8.5.4	Pseudo-observations half-window size	150
8.6	AnDA Noise sensitivity analysis	152
8.6.1	Nadir along-track data	152
8.6.2	SWOT data	153
8.7	Nadir along-track vs. SWOT data	153
8.8	Altimetric data fusion	158
8.9	Additional regressors for analog forecasting	167
8.10	Assimilation of observation gradients	170
8.11	Conclusion	172

8.1 Introduction

In previous chapters, we explored the use of operator decomposition models as a data-driven approach for the interpolation of high-resolution SLA fields from irregularly-sampled satellite observations. Despite the promising results obtained, recent success of alternative data-driven assimilation models, introduced in [67, 161, 162], motivated us to shift our efforts and explore alternative data-driven methods for the interpolation of SLA fields from satellite-derived data, and particularly from SWOT observations.

In this respect, with a view towards the SWOT mission launch in 2021 [59, 74], we focus here on exploiting recent developments in data-driven interpolation schemes to assess the extent to which SWOT data, combined or not with nadir along-track data, may lead to an improved reconstruction of altimetry fields. We use an observing system simulation experiment (OSSE) framework and consider a case-study region in the Western Mediterranean Sea, which is known to involve mesoscale processes with characteristic horizontal scales below 100 km [6, 181, 203, 204, 225].

As exposed in Chapter 6, current and past generation satellite altimetry missions have played a substantial role in improving our understanding of sea surface dynamics, despite only being able to provide measurements along the nadir satellite track, with a small radar altimeter footprint width of around 2-10 km, as observed in Figure 8.2. Even though the fusion of multiple altimeters allows for the reconstruction of scales that would be unattainable using a single altimeter, no combination of currently airborne altimeters is capable of completely resolving the smaller mesoscale (below 50-70 km) [68, 73] or the sub-mesoscale processes (below 10 km) [34], which are nonetheless crucial to increasing our understanding of a great number of oceanic processes [74]. Overall, it is generally considered that current satellite-derived altimetry fields hardly resolve horizontal scales below 100 km. In this respect, the SWOT mission's unparalleled high spatial resolution (2 km) should allow us to better capture mesoscale and sub-mesoscale processes [59, 74], which constitutes one of the great challenges within current oceanography research. The ability to exploit SWOT data within interpolation schemes to improve the reconstruction of smaller scale processes in altimetry fields remains, however, an open question [32, 79, 84, 224, 257].

As previously stated, given the heterogeneity of currently available remote sensing altimetry observations, the processing and fusion of datasets involving different sampling strategies at multiple spatio-temporal resolutions, irregular sampling patterns and missing data to produce gridded gap-free products is of major interest. The problem of interpolating irregularly-sampled data onto a regular grid belongs to the family of inverse problems, which have been extensively studied [63, 177, 195, 207]. In general, geophysical field interpolation issues can be classified into either model-driven or data-driven approaches. Among model-driven approaches, optimal interpolation and data assimilation issues are among the state-of-the-art methods for the interpolation

of geophysical fields from irregularly-sampled data. In this respect, given the current limitations of these approaches, considerable efforts are being made to improve OI [57] or find alternative approaches [257]. Particularly, Ubelmann et al. [257] developed Dynamical Interpolation, an alternative approach that exploits the Quasi-Geostrophy (QG) physical model by imposing a vorticity conservation constraint, and applied it to synthetic SWOT observations. Interestingly, from a physical point of view, Dynamic Interpolation comes to considering locally-varying spatio-temporal covariance structures. As far as data assimilation is considered, preliminary studies involving the assimilation of SWOT data have been performed. In particular, Carrier et al. [32] explore the assimilation of synthetic SWOT observations within a variational data assimilation framework, with promising results for the long-term assimilation of SWOT data. Besides model-driven approaches, the growing availability of large-scale datasets, issued from real remote sensing, in situ observations, numerical simulations and reanalysis, has motivated the development of data-driven approaches as a powerful and efficient alternative to model-driven schemes [67, 207]. Initially developed for image processing issues, patch-based and exemplar-based models have recently known some success in remote sensing applications [63, 177], while integrating these approaches into data assimilation formulations led to the development of analog data assimilation [89, 161, 243]. Alternatively, projection-based approaches, exploiting either classical orthogonal decompositions [8, 213] or alternative constraints like non-negativity [169], are also among state-of-the-art methods for the interpolation of geophysical fields. In this chapter, we focus on such data-driven approaches [8, 66, 67, 162, 169, 213] for the processing of SWOT data, as they provide a generic and computationally-efficient framework to explore the synergy between multi-source satellite-derived data and do not require specific geophysical priors to be verified. In particular, we concentrate here on evaluating the potential of SWOT observations to improve the reconstruction of smaller scale features in satellite-derived altimetry fields.

Regarding methodological considerations, the Analog Data Assimilation (AnDA), a completely data-driven data assimilation approach that exploits available observations to emulate dynamical model \mathcal{M} (thus avoiding the use of restrictive model hypothesis, such as spatio-temporally invariant covariance structures), was introduced in [161] following recent efforts that explore data-driven methods as an alternative to classical model-driven issues for the interpolation of geophysical fields [65, 66, 207]. Moreover, in [67], the authors further demonstrated the relevance of the data-driven AnDA framework for the interpolation of high-dimensional geophysical fields, namely sea surface temperature (SST) fields largely obscured by clouds and irregularly-sampled satellite-derived sea level anomaly (SLA) fields. Extending on this work, we develop in this chapter an application of various data-driven geophysical field interpolation approaches, an particularly of AnDA, to the reconstruction of SLA fields in the Western Mediterranean Sea from multi-source satellite altimetry data. Importantly, we explore different spatial sampling patterns and consider both nadir along-track altimetry sources [204] as well as wide-swath altimetry, with a view towards the upcoming SWOT mission [59, 74]. As previously

explained, we specifically focus on evaluating the potential of SWOT observations to enhance the reconstruction of small-scale features in satellite-derived altimetry fields.

In the remainder of this chapter, we first present a brief overview of the considered data-driven models. Subsequently, the considered case-study, data and observing system simulation experiment (OSSE) developed to evaluate the relevance of SWOT data to improve the reconstruction of altimetry fields are presented. We then report experimental results to assess the performance of the different data-driven approaches considered. Particularly, the relevance of AnDA as the preferred method to assess the impact of the considered space-time sampling of the sea surface is demonstrated. To optimize parameter settings for the proposed data-driven methodology, we perform a preliminary analysis of the sensitivity of the AnDA framework to its most relevant internal parameters and to different noise sources. Subsequently, our experiments exploit the proposed OSSE, which considers real spatio-temporal sampling patterns and high-resolution synthetic ground-truth datasets, to benchmark the proposed AnDA scheme under different experimental settings. This allows us to demonstrate the relevance of the proposed approach for the high-resolution reconstruction of SLA fields from irregularly-sampled data sources involving real satellite-derived spatio-temporal sampling patterns, including nadir along-track altimetry and pseudo-SWOT patterns. Importantly, particular attention is paid towards characterizing the contribution of SWOT observations to the reconstruction of fine-scale upper ocean dynamic features in satellite-derived altimetry fields. The potential for the fusion of multiple data sources, including nadir along-track and SWOT observations as well as additional complementary datasets, to improve SLA field reconstruction performance within the proposed methodology is also clearly demonstrated. Finally, the assimilation of numerically-resolved observation gradients is explored as a possible alternative to deal with correlated noise sources in SWOT observations.

Overall, reported results support the relevance of the proposed AnDA scheme for an improved reconstruction of mesoscale structures for horizontal scales ranging from ~ 20 km to ~ 100 km, with considerable gains with respect to Optimal Interpolation (OI) [57]. Furthermore, we demonstrate both a considerable gain, in terms of reconstruction error, for the reconstruction of smaller scales structures not correctly resolved by current methods, as well as a potential additional gain derived from the joint assimilation of both altimetry data sources considered. Particularly, our results suggest that the data-driven joint analysis of nadir along-track altimetry data and SWOT data leads to the best reconstruction performance.

8.2 Benchmarked methods

To evaluate the relevance of SWOT observations to improve the data-driven reconstruction of high-resolution SLA fields from satellite-derived observations, we consider three different data-driven approaches. Specifically, we evaluate projection-based approaches considering classical or-

thogonality constraints and non-negative constraints by considering MS-VE-DINEOF [161, 213] (Section 6.3.2.1) and the data-driven locally-adapted convolution-based non-negative decomposition approach presented in the previous chapter (NN-LLOD) [169]. Additionally, we consider the Analog Data Assimilation framework, introduced in [161] and presented in Section 6.3.5. For comparison purposes, we also consider Optimal Interpolation [20, 57, 141] (Section 6.3.3.3), the state-of-the-art model-driven approach for the interpolation of geophysical fields.

8.3 Case study and data

We illustrate the relevance of the proposed approaches by implementing an Observing System Simulation Experiment (OSSE), considering a case-study region in the Western Mediterranean Sea ($36.5^{\circ}N$ to $40^{\circ}N$, $1.5^{\circ}E$ to $8.5^{\circ}E$). The chosen region of interest for our experiments, the Western Mediterranean Sea, is a highly dynamic region characterized by relatively small Rossby radii, with smaller structures that make the reconstruction of mesoscale and sub-mesoscale sea surface dynamics from satellite data appear as particularly challenging. In this respect, the Mediterranean Sea can be regarded as a small-scale ocean laboratory, as a wide variety of global mesoscale and sub-mesoscale ocean processes are also present, at smaller scale, within both the basin and sub-basin scales [6, 181, 225, 255]. As such, developing efficient strategies to correctly capture and represent smaller scales within the Mediterranean Sea is of the utmost importance for the exploitation of the scientific potential that exists within the region.

8.3.1 Data

A synthetic ground-truth dataset of daily high-resolution SLA fields from 2010 to 2013 is generated using the Western Mediterranean Operational Forecasting System (WMOP) numerical model from SOCIB [124]. WMOP is particularly relevant for scientific studies in the Mediterranean, given it is a regional re-parametrization of the Regional Ocean Modeling System (ROMS) model nested in the larger scale Mediterranean Ocean Forecasting System (MFS) [124, 255], with a view to producing high-resolution realistic ocean dynamics simulations in the Western Mediterranean Sea. The original grid resolution of the model varies between 1.8 km and 2.2 km ($\sim 1/50^{\circ}$). We refer interested readers to [124] for an in-depth discussion of implementation details and validation of the WMOP model. In our case-study experiments, high-resolution SLA fields produced by WMOP are down-sampled to a $1/20^{\circ}$ resolution.

8.3.2 Observing System Simulation Experiment

The generated ground-truth SLA fields dataset is split into a training dataset comprising years 2010-2012, from which a catalog of analog/successor pairs is created, and a test dataset built from ground-truth SLA fields from 2013, which serves as ground-truth for the generation of

synthetic observations and the evaluation of reconstruction performance.

Synthetic observations emulating satellite altimetry are generated using the proposed OSSE, which exploits spatio-temporal locations derived from real satellite tracks from a four altimeter configuration (comprising Jason-2, Cryosat-2, Saral-Altika and Hy-2A satellite missions) in 2014. The year 2014 was chosen as reference because it provides the greatest number of spatio-temporal locations from real airborne altimetry missions. Acquisition noise is simulated by means of a centered additive Gaussian white noise of variance σ_n^2 .

SWOT-like observations, on the other hand, are generated from the ground-truth high-resolution SLA fields using the SWOT simulator [78], provided by the SWOT science team. Synthetic observations obtained with the SWOT simulator are produced by linearly interpolating the ground-truth fields onto a synthetically generated SWOT-like 2 km resolution grid computed using expected orbit characteristics of the SWOT mission. Additionally, the SWOT simulator emulates most noise sources expected to influence the mission once launched. According to the SWOT mission error budget [59], this includes both correlated noise sources, related to the novel radar interferometry based altimetry technique, its sampling spatio-temporal geometry and atmospheric perturbations (e.g. roll, phase, timing, baseline dilation and wet troposphere errors), and uncorrelated noise related to sensor acquisition error (e.g. KaRIn instrument noise).

Both types of synthetic observations are linearly interpolated onto a regular $1/20^\circ$ grid to match the high-resolution ground-truth SLA fields.

8.3.3 Pseudo-observations

To further investigate alternative sampling strategies, two different synthetic observation generation methods are explored. The first strategy comes to simply considering observations at times $t_0, t_0 + \delta t, \dots, t_0 + n\delta t, \dots$ etc directly. Alternatively, the second strategy relies on building pseudo-observations by accumulating observations on a time window $t_0 \pm D$ centered at current day t_0 . Additionally, for both strategies, we may solely consider either nadir along-track or SWOT observations or, alternatively, we may also combine these two types of observations, thus allowing for the seamless fusion of both data sources within the different data-driven interpolation schemes considered. An illustration of this principle and a visual comparison of these two strategies is presented in Figure 8.2.

8.4 Method comparison and benchmarking

We first perform a preliminary comparative analysis of the considered methods by evaluating their performance when exploiting nadir along-track altimetry observations. For all the reported experiments, performance is measured by means of the root mean squared error (RMSE, in m)

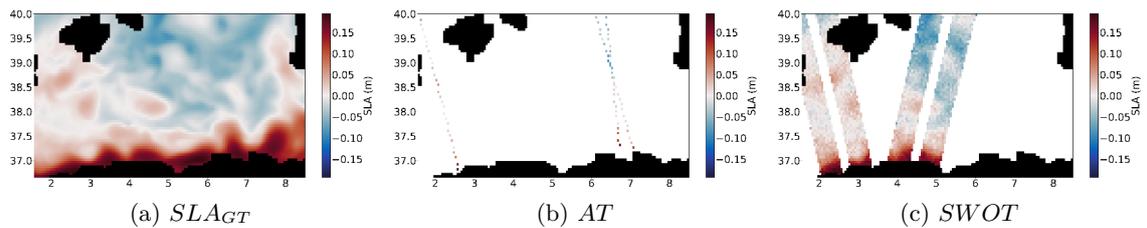


Figure 8.1 – Comparison of nadir along-track observations (8.1b) and SWOT observations (8.1c) generated from ground-truth high-resolution SLA fields (8.1a) using real satellite tracks spatio-temporal locations and the SWOT simulator, respectively. Adapted from [170]. © 2019 IEEE.

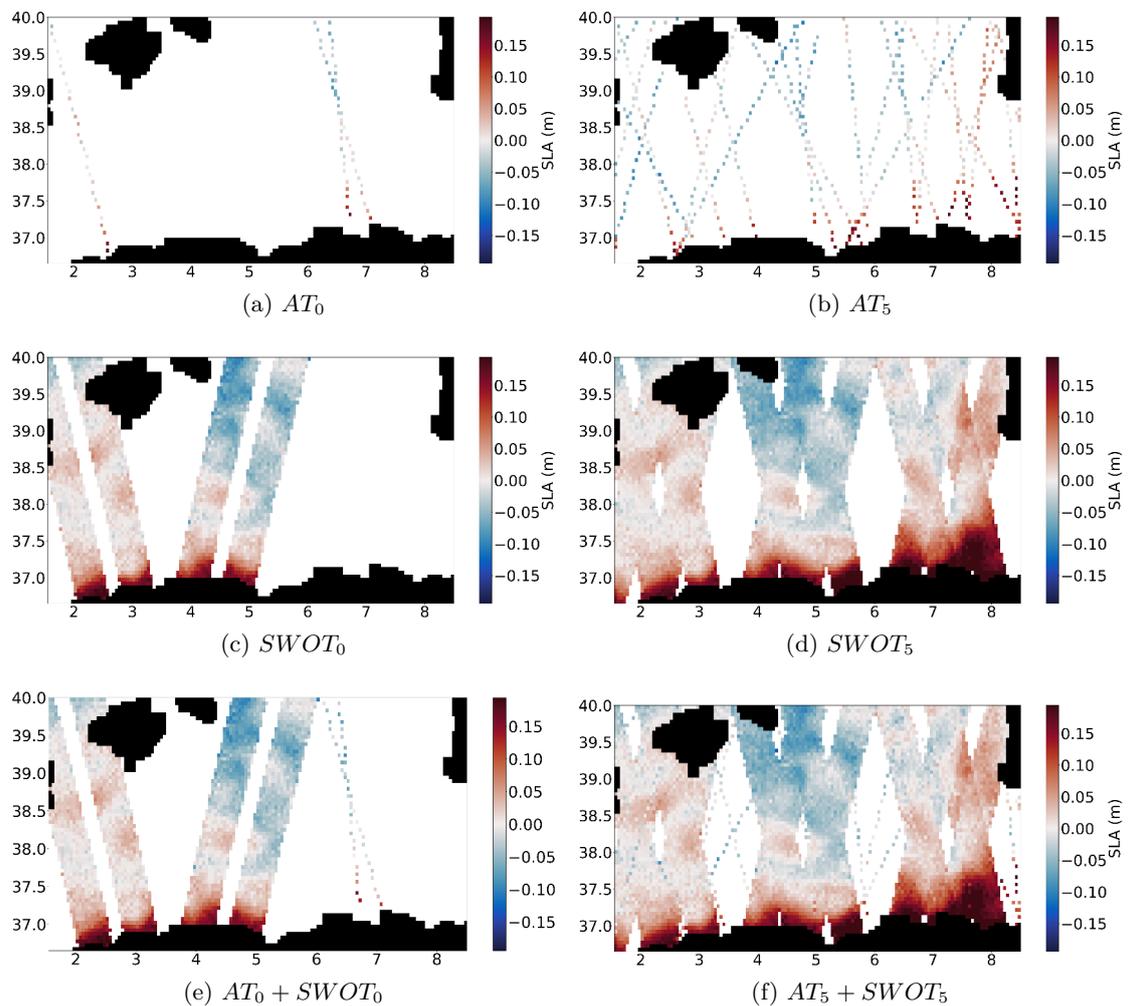


Figure 8.2 – Pseudo-observations generated from the synthetic ground-truth SLA fields by considering either observations at current day t_0 or observations accumulated over a time window $t_0 \pm D$, with $D = 5$ days, for both nadir along-track altimetry (noted as AT_D with $D = \{0, 5\}$) and pseudo-SWOT observations (noted as $SWOT_D$ with $D = \{0, 5\}$) (8.2a) Daily nadir along-track observations. (8.2b) Nadir along-track observations accumulated on a window $t_0 \pm D$ with $D = 5$ days. (8.2c) Daily SWOT observations. (8.2d) SWOT observations accumulated on a window $t_0 \pm D$ with $D = 5$ days. (8.2e) Fusion of daily nadir along-track and SWOT observations. (8.2f) Fusion of nadir along-track and SWOT observations accumulated on a window $t_0 \pm D$ with $D = 5$ days. Adapted from [170]. © 2019 IEEE.

Table 8.1 – Root mean squared error (Correlation) for SLA and SLA gradient (∇ SLA) reconstruction from nadir along-track observations for the different algorithms considered, namely OI [57], NN-LLOD [169], MS-VE-DINEOF [161, 213] and AnDA [161, 162]. Best result in **bold**.

Setting	SLA	∇ SLA
OI	0.02927 (0.8451)	0.006655 (0.6052)
MS-VE-DINEOF	0.02820 (0.9059)	0.006416 (0.6339)
NN-LLOD	0.02115 (0.6965)	0.004506 (0.5893)
AnDA	0.01978 (0.9457)	0.004699 (0.7660)

and correlation coefficient, for both the predicted SLA field and the gradient of the predicted SLA field (∇ SLA). The default parameter settings considered for AnDA and MS-VE-DINEOF are: patch size $W_p = 30$ pixels (~ 150 km, 1 pixel ≈ 5 km), $K = 25$ neighbours, assimilation lag $\delta t = 1$ days, $N_{EOF} = 9$ EOF components, pseudo-observation half-window size $D = 5$ days. The OI SLA reconstruction used a Gaussian covariance parametrization with a temporal correlation scale of 10 days and a spatial correlation scale of 100 km. The NN-LLOD approach exploited an operator basis issued from altimetry data accumulated on a 21-day window. The local fitting of the convolutional operators was achieved for $2^\circ \times 2^\circ$ overlapping regions.

Table 8.1 presents SLA and ∇ SLA RMSE and correlation results obtained when considering solely nadir along-track data for the different methods considered. From Table 8.1, AnDA clearly outperforms the other proposed approaches with a relative SLA RMSE gain of 32.42% (resp., 27.74% and 3.66%) w.r.t. OI (resp., NN-LLOD and MS-VE-DINEOF). This is interpreted as a consequence of the greater ability of AnDA to account for temporal dependencies, as it relies on a data assimilation formulation. The other proposed approaches, on the other hand, only take temporal dependencies into account implicitly within their formulations. From these results, we subsequently focus on exploring the use of SWOT observations within the AnDA framework.

8.5 AnDA Parameter Sensitivity Analysis

We report the sensitivity of the proposed AnDA framework to its most relevant internal parameters, namely the impact of the patch size W_p , the number of neighbours K , the assimilation lag δt and the pseudo-observation half-window size D . For the sensitivity analysis, only daily nadir along-track data is considered, so that the impact of considering SWOT observations can be subsequently explored under ideal conditions involving optimized parameter settings. For all the reported experiments, performance is measured by means of the root mean squared error (RMSE, in m) and correlation coefficient. Performance is evaluated on both the predicted SLA field and the gradient of the predicted SLA field ∇ SLA. Tests are performed by varying the parameter of interest under the following base parameter settings: $W_p = 35$ pixels (~ 175 km, 1 pixel ≈ 5 km), $K = 100$ neighbours, $\sigma_n^2 = 0$, $\delta t = 1$ days, $D = 0$ days, $N_{EOF} = 9$ EOF components.

8.5.1 Patch size

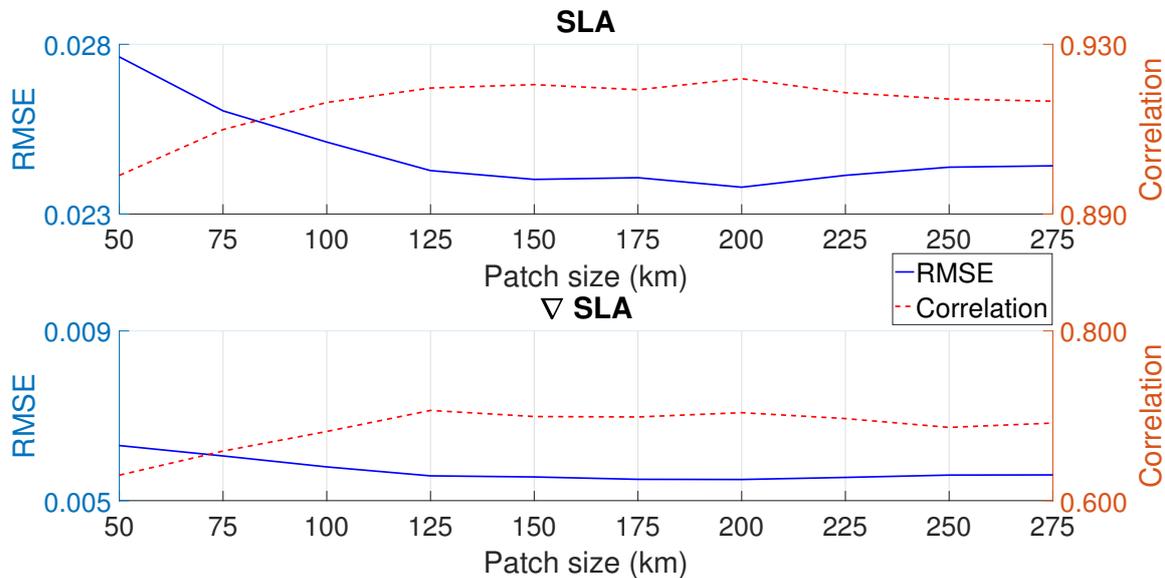


Figure 8.3 – AnDA sensitivity to patch size W_p (1 pixel \approx 5 km) (top) SLA reconstruction root mean squared error (RMSE) and correlation as a function of patch size (in km). (bottom) SLA gradient reconstruction root mean squared error (RMSE) and correlation as a function of patch size (in km). Taken from [171]. © 2018 IEEE.

Figure 8.3 presents SLA and ∇ SLA RMSE and correlation coefficient as a function of patch size W_p . From the reported results, the optimal patch size seems to lie within the range of 125-200 km, which roughly corresponds to ten times the region’s internal Rossby radius of deformation. This result is coherent with the range of scales we are trying to resolve, i.e., those between the smallest scale resolved by OI (\sim 100 km) and the size of the smallest mesoscale features present in the region (\sim 10-15 km). Additionally, it seems reasonable that, in order to capture mesoscale dynamics, the considered patch size should be at least twice as big as the size of the mesoscale features to be resolved. The obtained gain is in agreement with the expected potential gain of AnDA to better resolve mesoscale structures in the horizontal scale range of \sim 10-15 km to \sim 100 km. Additional experiments (not shown here) indicate that, for SWOT-like data, the best performance is indeed obtained with a patch size of $W_p = 150$ km.

8.5.2 Number of neighbours

Figure 8.4 presents SLA and ∇ SLA RMSE and correlation coefficient as a function of the number of nearest neighbours K used in the analog forecasting step. Interestingly, the best reconstruction performance is obtained for $K = 25$ neighbours. Increasing the number of neighbours seems to hinder reconstruction performance, which may be explained by the existing compromise between model locality in state space and the number of neighbours K used in the analog forecasting step. In this respect, a more local description in state space is obtained by using fewer neighbours, and

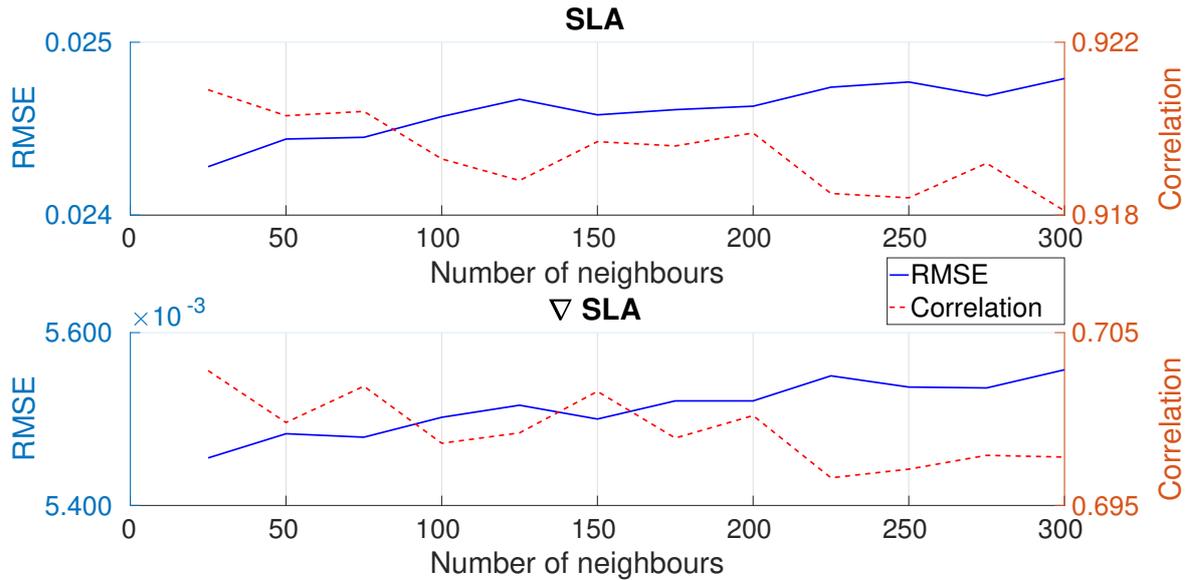


Figure 8.4 – AnDA sensitivity to number of neighbours K . (top) SLA reconstruction root mean squared error (RMSE) and correlation as a function of the number of neighbours used in the forecasting step of AnDA. (bottom) SLA gradient reconstruction root mean squared error (RMSE) and correlation as a function of the number of neighbours used in the forecasting step of AnDA. Taken from [171]. © 2018 IEEE.

thus considering a higher number of neighbours translates into a more global estimated model, which may negatively affect reconstruction performance.

8.5.3 Assimilation lag

Figure 8.5 presents SLA and ∇ SLA RMSE and correlation as a function of the assimilation lag δt considered in the assimilation step of the AnDA framework. Not surprisingly, the further away observations are in time, the less effective the correction introduced by the assimilation. In this respect, as observations are further away, the linear approximation introduced by the analog forecasting operator loses validity and becomes less accurate, which partially explains the algorithm’s decrease in performance.

8.5.4 Pseudo-observations half-window size

Figure 8.6 presents SLA and ∇ SLA RMSE and correlation coefficient as a function of the pseudo-observation half-window size D (with the total size of the $t_0 \pm D$ pseudo-observation window size being $W = 2D + 1$). The best performance is obtained for a half-window size of $D = 5$ days, which is coherent with the temporal correlation scales of the smaller mesoscale features we are interested in reconstructing. Additionally, this seems to be the parameter that has the biggest effect on performance (at least for the nadir along-track observations considered).

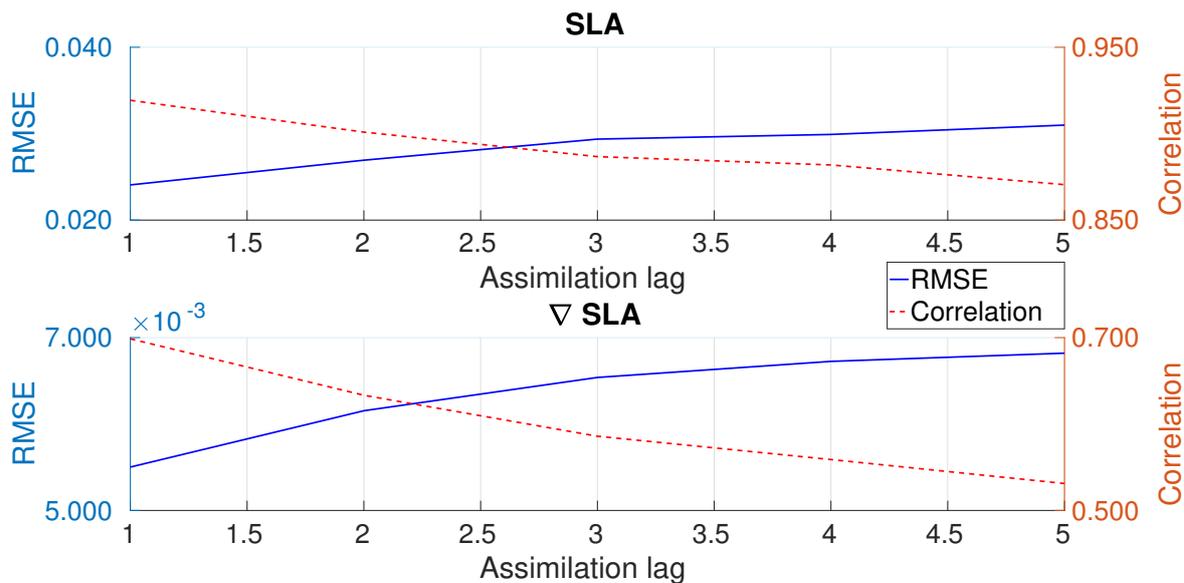


Figure 8.5 – AnDA sensitivity to assimilation lag δt . (top) SLA reconstruction root mean squared error (RMSE) and correlation as a function of the assimilation lag δt (in days) used for the assimilation of observations in AnDA. (bottom) SLA gradient reconstruction root mean squared error (RMSE) and correlation as a function of the assimilation lag δt (in days) used for the assimilation of observations in AnDA.

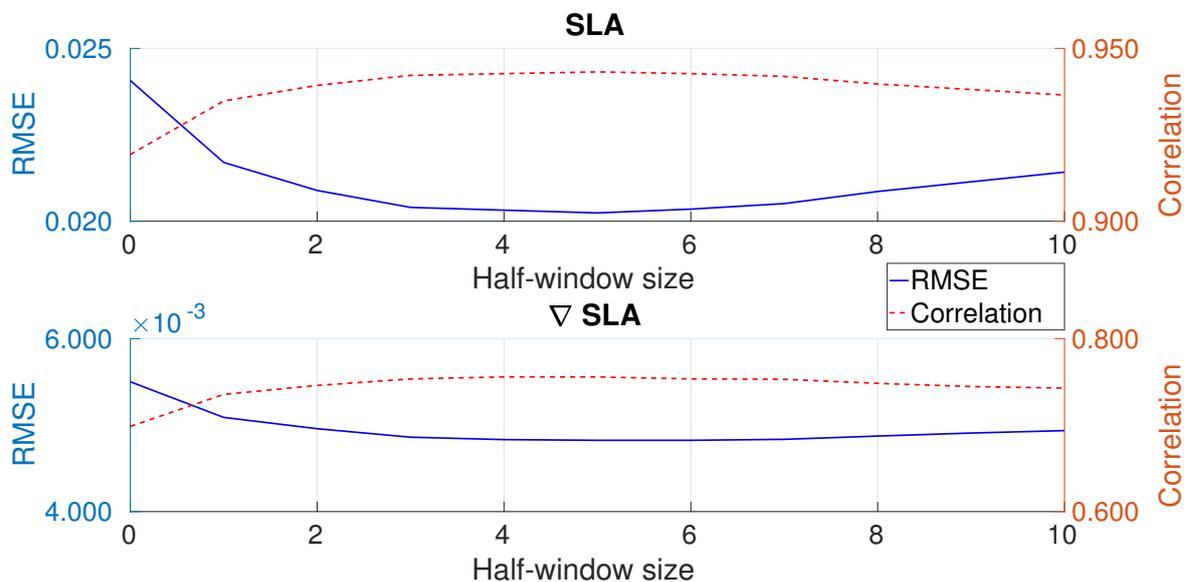


Figure 8.6 – AnDA sensitivity to observation half-window size D . (top) SLA reconstruction root mean squared error (RMSE) and correlation as a function of the half-window size D (in days) for observations accumulated on a time window $t_0 \pm D$ days. (bottom) SLA gradient reconstruction root mean squared error (RMSE) and correlation as a function of the half-window size D (in days) for observations accumulated on a time window $t_0 \pm D$ days. Taken from [171]. © 2018 IEEE.

Table 8.2 – Root mean squared error (Correlation) for AnDA SLA and SLA gradient (∇ SLA) reconstruction from nadir along-track observations accumulated on a window $t_0 \pm D$ with $D = 5$ days under different noise settings, considering a centered additive white Gaussian noise with variance $\sigma_n^2 = \{0, 1, 3, 5\} \text{ cm}^2$. Best result in **bold**. Results for OI [57] and NN-LLOD [169] given as reference.

Setting	SLA	∇ SLA
$\sigma_n^2 = 0 \text{ cm}^2$	0.01969 (0.9465)	0.004687 (0.7679)
$\sigma_n^2 = 1 \text{ cm}^2$	0.01975 (0.9459)	0.004705 (0.7664)
$\sigma_n^2 = 3 \text{ cm}^2$	0.01978 (0.9457)	0.004699 (0.7660)
$\sigma_n^2 = 5 \text{ cm}^2$	0.01992 (0.9448)	0.004714 (0.7651)
OI	0.02927 (0.8451)	0.006655 (0.6052)
NN-LLOD	0.02115 (0.6965)	0.004506 (0.5893)

8.6 AnDA Noise sensitivity analysis

A preliminary step towards the characterization of the performance of AnDA in the context of the interpolation of high-resolution SLA fields from altimetry data consists in evaluating the effect of acquisition noise on reconstruction performance. To this end, we evaluate the effect of different noise types and sources for nadir along-track and SWOT data independently.

8.6.1 Nadir along-track data

Data acquisition errors for conventional altimeter nadir along-track data involve uncorrelated noise sources, so that, as previously explained, they can be simulated by means of a centered white Gaussian random noise process of variance σ_n^2 . For current airborne altimetry missions, noise variance is in the $\sigma_n^2 = [3, 5] \text{ cm}^2$ range [204]. To study the effect of noise for AnDA reconstruction performance, we simulate along-track data with a variable observation error $\sigma_n^2 = \{0, 1, 3, 5\} \text{ cm}^2$. The base parameter settings considered are: $W_p = 30$ pixels (~ 150 km, 1 pixel ≈ 5 km), $K = 25$ neighbours, $\delta t = 1$ days, $D = 5$ days, $N_{EOF} = 9$ EOF components.

Table 8.2 presents SLA and ∇ SLA RMSE and correlation for nadir along-track observations under the considered noise settings. Results for OI [57] and NN-LLOD [169] are also given as reference. From these results, it seems clear that AnDA remains considerably robust to noise sources inherent to nadir along-track data. This can be explained by the use of an Ensemble Kalman Filter and Smoother (EnKFS) within the AnDA framework, since this stochastic assimilation technique relies on Gaussian noise hypothesis and thus works best under such conditions. It should also be noted that even under the harshest noise conditions AnDA still outperforms OI and NN-LLOD, which illustrates the potential of AnDA for improving the reconstruction of high-resolution SLA fields, specially considering both OI and NN-LLOD interpolate the high-resolution SLA field from noise-free nadir along-track SLA observations accumulated on a window $t_0 \pm D$ of size $D = 10$ days (i.e. double the size of the window considered here by AnDA).

8.6.2 SWOT data

The characterization of the effect of SWOT-like observations noise requires the distinction between correlated and uncorrelated noise sources. As previously explained, according to the SWOT mission error budget, SWOT will present both correlated noise relating to the sampling technique and geometry, and to atmospheric perturbations (e.g. roll, phase, timing, baseline dilation and wet troposphere errors) and uncorrelated noise sources relating to instrument error (e.g. KaRIn noise). We refer the interested reader to [59, 78] for a detailed description of the error sources involved in SWOT and their simulation implementation in the SWOT simulator. Taking this into account, we consider here three distinct situations to account for the different nature of error sources.

In Table 8.3, we present SLA and ∇ SLA RMSE and correlation when pseudo-SWOT observations are considered for these three distinct cases, namely an ideal case where no noise sources exist, a second case where only an uncorrelated (KaRIn) noise source is involved and a worst-case scenario where both correlated and uncorrelated noise sources are involved. We consider here real observations associated with the current day only ($D = 0$). From Table 8.3, we can conclude that correlated noise presence in SWOT data, in particular, seems to have a strong effect on performance, whereas AnDA seems relatively robust to uncorrelated KaRIn noise. We believe two distinct effects to be present here. On one hand, when considering SWOT data more pixels are present in observations (given the wide-swath nature of the altimeter track), which may enhance the effects of both noise types, even though correlated noise seems to produce the greater effect on performance. On the other hand, AnDA uses an EnFKS, which works best when noise is uncorrelated, and is unable to handle correlated noise properly without the use of additional techniques [224]. Our results also indicate that both effects seem to be enhanced when data is accumulated over several days (not shown here).

8.7 Nadir along-track vs. SWOT data

To evaluate the effect of different observation sampling patterns, we now consider both nadir along-track observations and pseudo-SWOT wide-swath high-resolution SLA observations obtained using the SWOT simulator [78], and compare the performance of AnDA when assimilating both types of observations separately. The experimental procedure and performance metrics considered are identical to those in Section 8.6. Given the sensitivity of the system to correlated noise sources, we consider only KaRIn noise for SWOT data and an intermediate noise level of $\sigma_n^2 = 3 \text{ cm}^2$ for nadir along-track data, under the assumption that a pre-processing scheme can be exploited to either filter correlated noise sources [84] or take them into account into the correlation structure of data assimilation schemes [224]. Additional alternative strategies to deal with correlated noise sources will be presented and briefly discussed in Section 8.10. The default parameter settings considered are: $W_p = 30$ pixels (~ 150 km, 1 pixel ≈ 5 km), $K = 25$

Table 8.3 – Root mean squared error (Correlation) for AnDA SLA and SLA gradient (∇ SLA) reconstruction from daily SWOT observations under different noise settings, namely ideal SWOT observations without noise sources, SWOT observations when considering only correlated noise sources (KaRIn noise), and SWOT observations when both correlated and uncorrelated noise sources are considered. Best result in **bold**.

Setting	SLA	∇ SLA
No noise	0.01802 (0.9548)	0.004410 (0.7894)
KaRIn noise	0.01810 (0.9543)	0.004436 (0.7857)
All noise sources	0.03765 (0.8143)	0.005791 (0.5900)

Table 8.4 – Root mean squared error (Correlation) for AnDA SLA and SLA gradient (∇ SLA) reconstruction from nadir along-track observations (AT_D) and wide-swath SWOT observations ($SWOT_D$). For each type of observations, both daily observations ($D = 0$) and observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days are considered. Best result in **bold**.

Setting	SLA	∇ SLA
AT_0	0.02395 (0.9186)	0.005507 (0.6989)
AT_5	0.01978 (0.9457)	0.004699 (0.7660)
$SWOT_0$	0.01810 (0.9543)	0.004436 (0.7857)
$SWOT_5$	0.01920 (0.9502)	0.004345 (0.7913)

neighbours, $\delta t = 1$ days, $N_{EOF} = 9$ EOF components.

Table 8.4 presents SLA and ∇ SLA RMSE and correlation results obtained when considering nadir along-track data and SWOT data independently, for current observations only and for observations accumulated on a time window $t_0 \pm D$, with $D = 5$ days. From Table 8.4, a first interesting result is that considering SWOT observations for the current day only (without accumulation) is already enough to outperform OI, MS-VE-DINEOF, NN-LLOD and AnDA when considering only along-track data. We report a mean normalized RMSE (nRMSE, computed by normalizing the RMSE by the standard deviation of the ground-truth high-resolution SLA fields) gain of around 2.8% for the whole year 2013 and 6.2% for days when at least one SWOT track is observed (with respect to AnDA applied to nadir along-track data accumulated on a time window $t_0 \pm D$ with $D = 5$ days). This clearly highlights that the performance gain relates to the availability of SWOT observations. This is in agreement with previous studies [32, 217] that suggest that SWOT observations should be able to provide as much information as four conventional altimeters for the reconstruction of mesoscale features. To further evaluate this result in detail, we present in Figure 8.7 the time series of the nRMSE for the AnDA assimilation of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (Figure 8.7, top) and for the AnDA assimilation of daily SWOT observations (Figure 8.7, middle). The nRMSE gain between them, i.e., the gain obtained when considering daily SWOT observations rather than nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days is also depicted (Figure 8.7, bottom). The red filling indicates periods during which there is a positive nRMSE gain when considering daily SWOT data rather than nadir along-track observations

accumulated on a time window $t_0 \pm D$ with $D = 5$ days (i.e. $nRMSE_{SWOT_0} \leq nRMSE_{AT_5}$). We also consider the time series of the proportion of the total SLA field covered by observations, depicted in Figure 8.8, where we present the observation coverage proportion (given as the ratio between the number of observed pixels and the total number of pixels in the considered region, excluding land pixels) for nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (Figure 8.8, top) and for daily SWOT observations (Figure 8.8, middle), and the coverage proportion gain between them, i.e., the coverage gained by considering daily SWOT observations rather than nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days. Similarly to Figure 8.7, the red filling indicates periods during which there is an nRMSE gain when considering daily SWOT data rather than nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (i.e. $nRMSE_{SWOT_0} \leq nRMSE_{AT_5}$).

Not surprisingly, the assimilation of daily SWOT data seems to outperform the assimilation of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days most of the time, with few days involving a loss in performance when considering SWOT data. Indeed, for days when a SWOT swath exists within the considered region, the surface coverage of such swath (and thus the amount of information contained within such swath) is considerably larger when compared to the surface coverage (and information contained) in nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days, as can be clearly seen in Figure (8.8, bottom). This increased surface coverage may then partially explain the improved reconstruction performance. By contrast, days where no SWOT observation exists involve a very small mean nRMSE gain of 0.8% w.r.t. the AnDA interpolation of nadir along-track data, with around 38% of such days actually involving a loss in reconstruction performance (with a mean nRMSE loss of around 2%). In this respect, days involving a loss in performance are, as expected, mostly related to a loss in observation coverage, i.e. days when no SWOT observation exist. For such days, AnDA relies solely on the Analog Forecasting scheme to produce a prediction when considering daily SWOT observations, whereas when considering nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days AnDA will always assimilate some information conveyed by observations.

Another interesting, yet counterintuitive, result is that performance seems to decrease when accumulating SWOT observations over a time window $t_0 \pm D$. We hypothesize that wide-swath observations, given their two-dimensional nature and higher continuous spatial coverage, are prone to capturing moving/changing structures multiple times as observations are accumulated, thus creating inconsistent SLA observations leading to the reconstruction of fictitious structures. Accumulating SWOT observations will then tend to increase the sensitivity of AnDA to changes in the SLA field occurring during the days over which observations are accumulated. In this respect, accumulating observations over multiple days seems to be an appropriate strategy to improve AnDA SLA reconstruction performance for nadir along-track data, whereas for

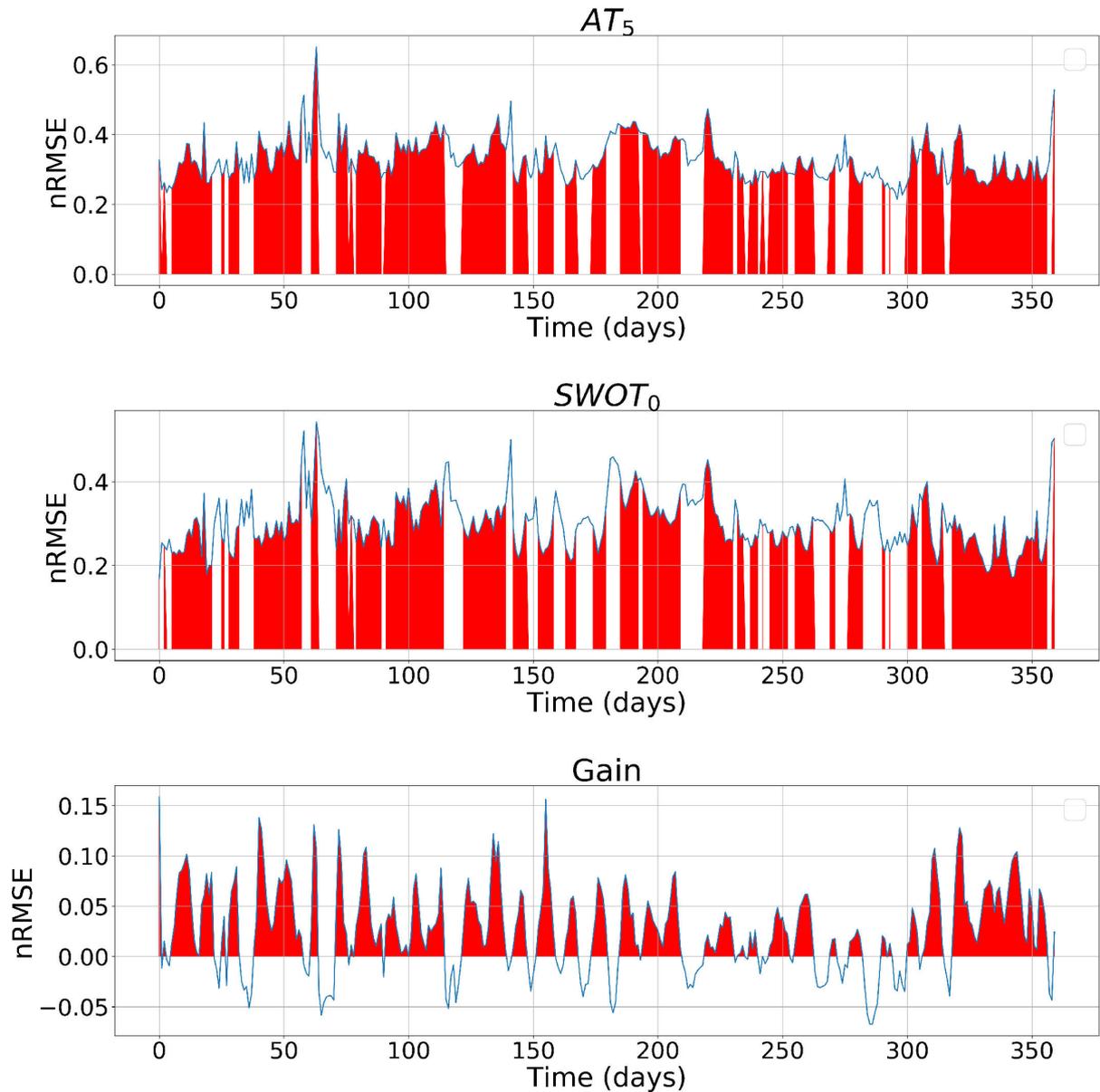


Figure 8.7 – Time series of the normalized root mean squared error (nRMSE) for the AnDA assimilation of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (top, noted as AT_5) and for the AnDA assimilation of daily SWOT observations (middle, noted as $SWOT_0$). The time series of the nRMSE gain obtained when considering daily SWOT observations rather than nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (given by $nRMSE_{AT_5} - nRMSE_{SWOT_0}$) is also depicted (bottom, noted as Gain). The red filling indicates periods during which there is a positive nRMSE gain when considering daily SWOT data rather than nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (i.e. $nRMSE_{SWOT_0} \leq nRMSE_{AT_5}$).

SWOT data AnDA seems to be prone to issues arising from inconsistencies between accumulated observations. An illustration of this principle is presented in Figure 8.9, where we depict the ground-truth SLA fields (8.9a), as well as the observations and interpolated SLA fields for the AnDA assimilation of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (Figures 8.9c and 8.9d), daily SWOT observations (Figures 8.9e and 8.9f), and SWOT observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (Figures 8.9g and

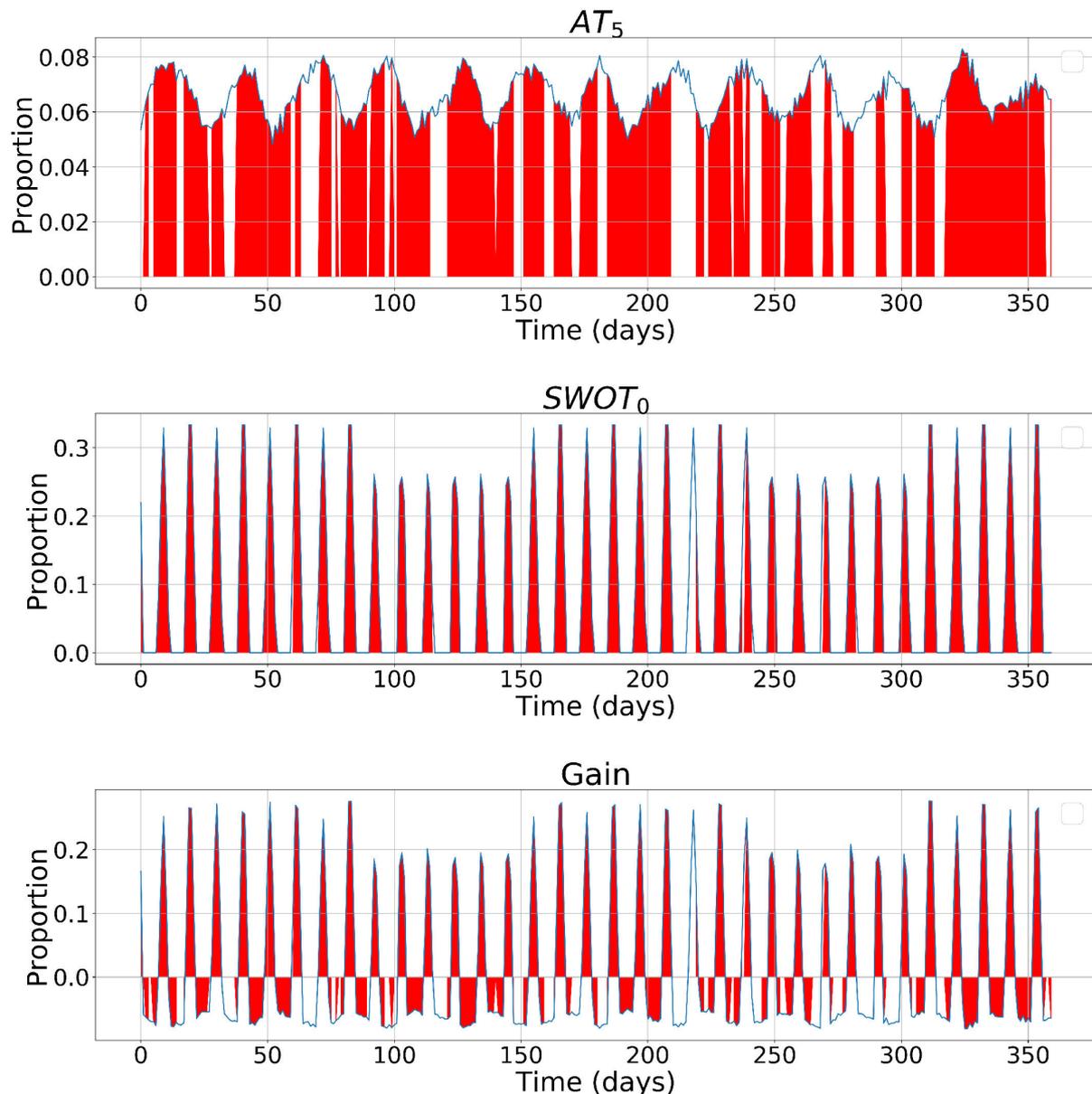


Figure 8.8 – Time series of the observation coverage (given as the ratio between the number of observed pixels and the total number of pixels in the considered region, excluding land pixels) for nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (top, noted as AT_5) and for daily SWOT observations (middle, noted as $SWOT_0$). The time series of the coverage gain obtained when considering daily SWOT observations rather than nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days is also depicted (bottom, noted as Gain). *The red filling indicates periods during which there is a positive nRMSE gain when considering daily SWOT data rather than nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (i.e. $nRMSE_{SWOT_0} \leq nRMSE_{AT_5}$).*

8.9h). We also include the SLA field interpolated by OI as reference (Figure 8.9b). Figure 8.10 presents analogue results for the gradient of the interpolated fields ∇SLA . In this example, we can see that the accumulation of SWOT observations leads to the reconstruction of false structures along the high front in the southern part of the region, on the Algerian coast (Figure 8.9h), whereas the reconstruction obtained from nadir along-track data accumulated on a time window $t_0 \pm D$ with $D = 5$ days and from daily SWOT observations remain closer to the ground-truth

Table 8.5 – Root mean squared error (Correlation) for AnDA SLA and SLA gradient (∇ SLA) reconstruction from the fusion of nadir along-track observations (AT_D) and SWOT observations ($SWOT_D$). For each type of observations, both daily observations ($D = 0$) and observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days are considered. Best result in **bold**.

Setting	SLA	∇ SLA
$AT_0 + SWOT_0$	0.01742 (0.9576)	0.004375 (0.7934)
$AT_5 + SWOT_5$	0.01876 (0.9523)	0.004318 (0.7952)
$AT_5 + SWOT_0$	0.01687 (0.9607)	0.004286 (0.8051)

SLA field along the coast (Figures 8.9d and 8.9f). Indeed, no daily SWOT observation exists for this particular day (Figure 8.9e), so that all SWOT swaths accumulated come from days around the actual reconstruction date. In this way, the SLA fields sampled by these swaths slightly differ from the real SLA field. Even if the differences between the sampled and real SLA fields are small, the accumulation of SWOT observations will still involve the assimilation of inconsistent SLA fields. A similar, yet milder, effect can be observed in the eastern part of the considered region, where the shape of the structures captured is slightly deformed southwards due to the information conveyed by the accumulated SWOT observations.

8.8 Altimetric data fusion

We further evaluate the potential of SWOT data to improve the reconstruction of high-resolution SLA fields by considering the joint analysis of nadir along-track and SWOT observations (which may be regarded as equivalent to complementing SWOT wide-swath observations with the nadir along-track data that is part of the SWOT configuration). We report in Table 8.5 SLA and ∇ SLA RMSE and correlation results when jointly considering these two data sources in the AnDA scheme. As expected, this joint analysis leads to a relative SLA RMSE gain of 6.80% w.r.t. the sole use of SWOT data and 14.71% w.r.t. the sole use of nadir along-track data. Not surprisingly, when considering the fusion of nadir along-track and SWOT data, the best strategy involves the fusion of accumulated nadir along-track observations over a time window of $D = 5$ days, thus increasing the spatial coverage of nadir along-track data, and SWOT observations captured on the current day only, thus fully exploiting the two-dimensional nature and added information provided by SWOT without being negatively affected by the increased sensitivity of AnDA to inconsistencies between accumulated SWOT observations.

This result is further validated by Figure 8.11, depicting the time series of the nRMSE for the AnDA assimilation of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (Figure 8.11, top) and for the AnDA assimilation of the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (Figure 8.11, middle). The nRMSE gain between them, i.e., the gain obtained when complementing nadir along-track observations accumulated on a time window $t_0 \pm D$ with

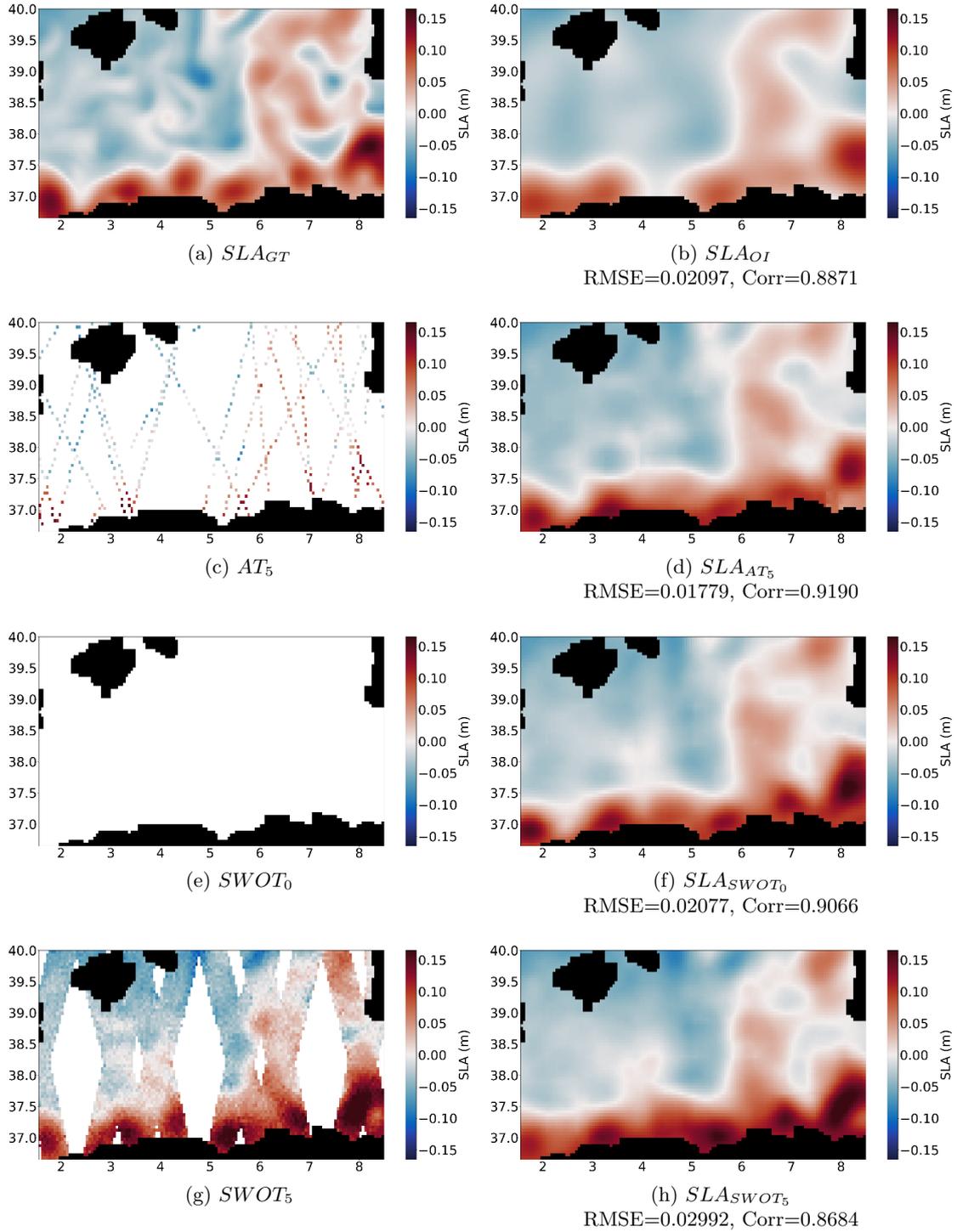


Figure 8.9 – AnDA SLA reconstruction results for the 2nd May, 2013. (8.9a) Real high-resolution ground-truth SLA field. (8.9b) Optimal Interpolation SLA reconstruction obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of 10 days and a spatial correlation scale of 100 km. (8.9c) Nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as AT_5). (8.9d) AnDA SLA reconstruction from nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as AT_5 and depicted in Figure (8.9c)). (8.9e) Daily SWOT observations (noted as $SWOT_0$). (8.9f) AnDA SLA reconstruction from daily SWOT observations (noted as $SWOT_0$ and depicted in Figure (8.9e)). (8.9g) SWOT observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as $SWOT_5$). (8.9h) AnDA SLA reconstruction from SWOT observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as $SWOT_5$ and depicted in Figure (8.9g)). RMSE and correlation values given for each reported result.

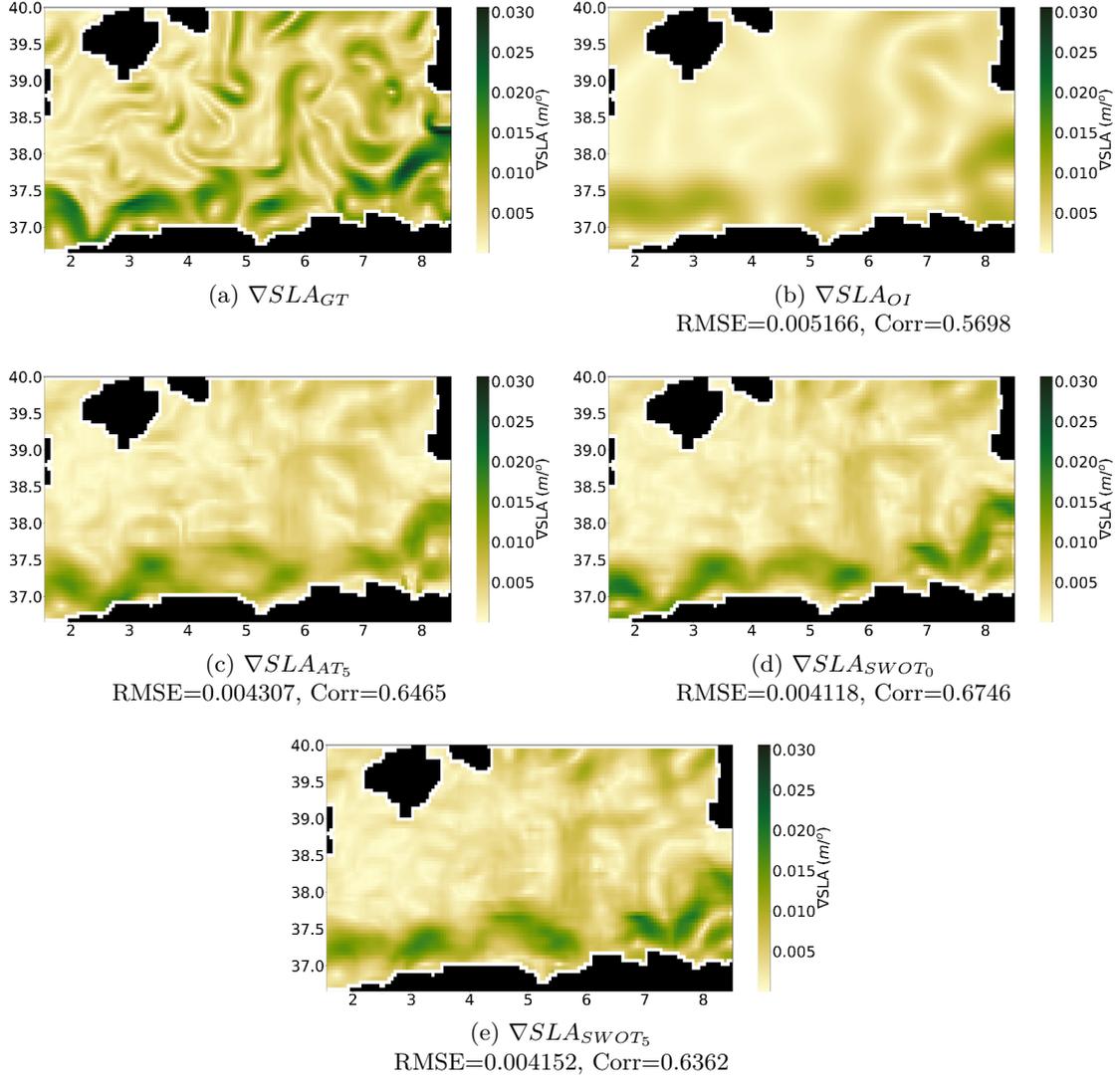


Figure 8.10 – AnDA SLA gradient reconstruction results for the 2nd May, 2013. (8.10a) Real high-resolution ground-truth SLA gradient field. (8.10b) Optimal Interpolation SLA gradient reconstruction obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of 10 days and a spatial correlation scale of 100 km. (8.10c) AnDA SLA gradient reconstruction from nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as AT_5 and depicted in Figure (8.9c)). (8.10d) AnDA SLA gradient reconstruction from daily SWOT observations (noted as $SWOT_0$ and depicted in Figure (8.9e)). (8.10e) AnDA SLA gradient reconstruction from SWOT observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as $SWOT_5$ and depicted in Figure (8.9g)). RMSE and correlation values given for each reported result.

$D = 5$ days with daily SWOT data, is also depicted (Figure 8.11, bottom). The red filling indicates periods during which the nRMSE gain is positive, i.e. when the additional use of daily SWOT data improves interpolation performance ($nRMSE_{AT_5+SWOT_0} \leq nRMSE_{AT_5}$). From the reported results, it is clear that considering additional SWOT data improves reconstruction performance considerably. Indeed, for most days, considering additional SWOT observations helps improve reconstruction performance. Furthermore, for the few days where a loss in reconstruction nRMSE exists, such loss is considerably small relative to the nRMSE gain from other

days.

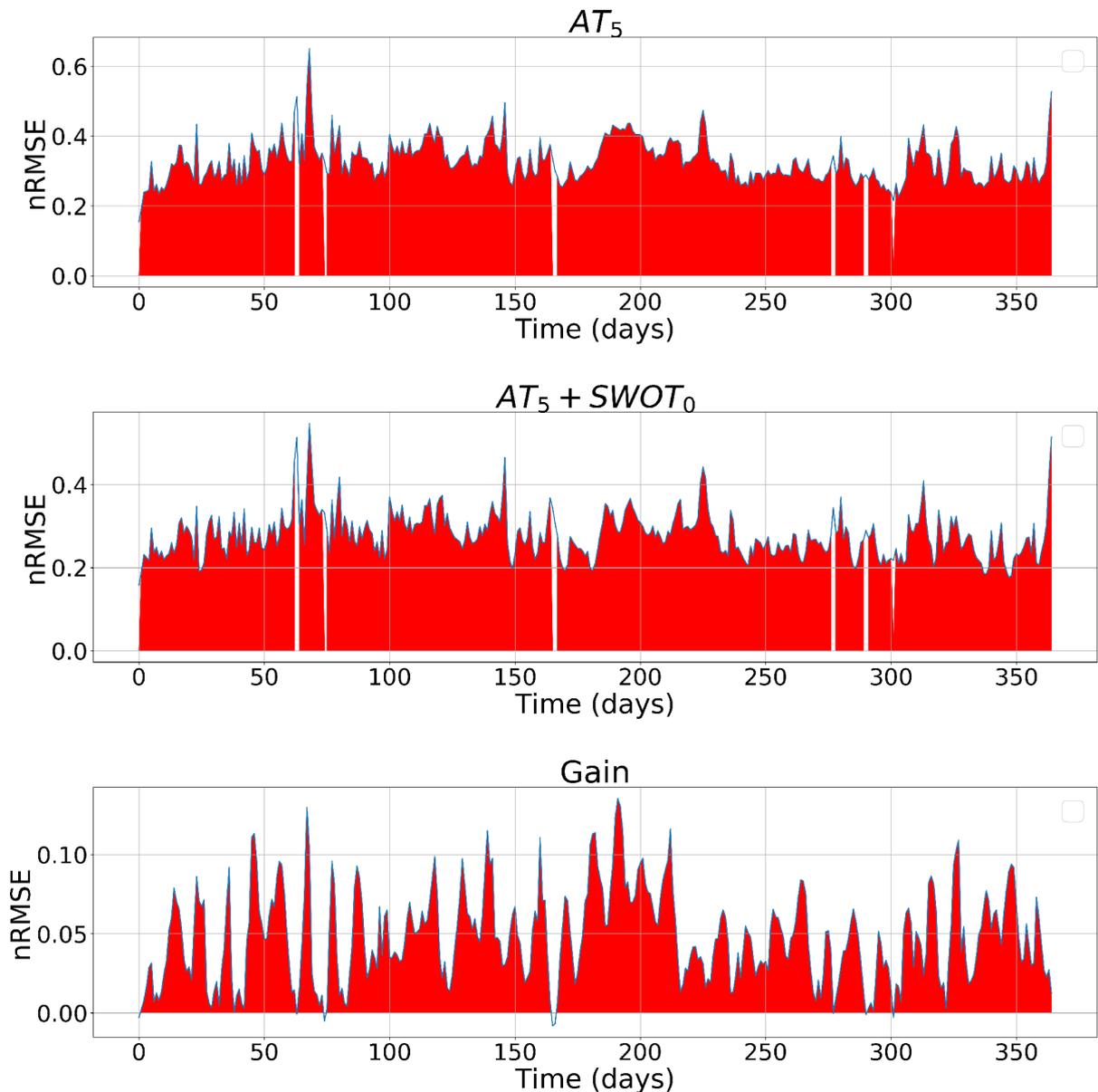


Figure 8.11 – Time series of the normalized root mean squared error (nRMSE) for the AnDA assimilation of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (top, noted as AT_5) and for the AnDA assimilation of the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (middle, noted as $AT_5 + SWOT_0$). The time series of the nRMSE gain obtained when complementing nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days with daily SWOT data (given by $nRMSE_{AT_5} - nRMSE_{AT_5+SWOT_0}$) is also depicted (bottom, noted as Gain). The red filling indicates periods during which the nRMSE gain is positive, i.e. when the additional use of daily SWOT data improves interpolation performance ($nRMSE_{AT_5+SWOT_0} \leq nRMSE_{AT_5}$).

To further highlight the performance of AnDA-based interpolations considering different observation sampling patterns and illustrate the gain brought forward by considering additional SWOT data, Figure 8.12 depicts the ground-truth SLA fields (8.12a), as well as the observations and interpolated SLA fields for the AnDA assimilation of nadir along-track observations

accumulated on a time window $t_0 \pm D$ with $D = 5$ days (Figures 8.12c and 8.12d), for the AnDA assimilation of daily SWOT observations (Figures 8.12e and 8.12f) and for the AnDA assimilation of the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (Figures 8.12g and 8.12h). We also include the SLA field interpolated by OI as reference (Figure 8.12b). Figure 8.13 presents analogue results for the gradient of the interpolated fields ∇SLA . In agreement with Table 8.4, the SLA field reconstructed from daily SWOT data appears closer to the ground-truth SLA field than the reconstruction obtained when considering nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days. Moreover, for the considered date, SWOT data provides additional information in the western part of the case-study region, south of Mallorca. Visually, this leads for instance to the improved reconstruction of an eddy-like structure north of the Algerian coast (38°N , 4.5°E), which is hardly recovered when considering the OI or AnDA reconstruction from nadir along-track data only. For this region, the joint assimilation of nadir along-track and SWOT observations seems to introduce slight differences in the delineation of small-scale structures, when compared to the sole assimilation of SWOT data. By contrast, in the eastern part of the case-study region, the absence of SWOT data and the scarce sampling of nadir along-track data lead to a greater spatial smoothing of fine-scale patterns. The gradient fields further illustrate the improvement issued from simultaneously considering SWOT and nadir along-track data, specially compared with OI. We may nonetheless notice that a significant fraction of the fine-scale structures could not be recovered.

If we compute the radially averaged power spectra of the ground-truth and reconstructed SLA fields (Figure 8.14) we can see that, effectively, considering additional SWOT data helps improve the reconstruction performance of AnDA, with the joint AnDA assimilation of nadir along-track and SWOT data outperforming the AnDA assimilation of nadir along-track observations only for most spatial scales in the mesoscale/sub-mesoscale range. Moreover, despite the similarity between the spectra of the assimilation of daily SWOT data and the joint assimilation of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations, we can observe that the joint assimilation helps better constraint the spectrum of the reconstructed SLA field, particularly for the smaller scales, where the joint assimilation produces a spectrum closer to the real ground-truth SLA spectrum. As expected, OI clearly underestimates the high-resolution finer details of the ground-truth SLA fields, whereas results obtained with AnDA approximate the ground-truth considerably better. For scales closer to the pixel size (~ 5 km), AnDA tends to over-estimate the spectrum, which we believe relates to both some slight border effects due to the patch formulation (which we have observed during our tests) and to the KaRIn pixel-wise Gaussian noise (since our experiments show that this effect is considerably less strong when noise-free observations are considered (not shown here)). The joint assimilation of nadir-along track and SWOT observations seems to help limit this

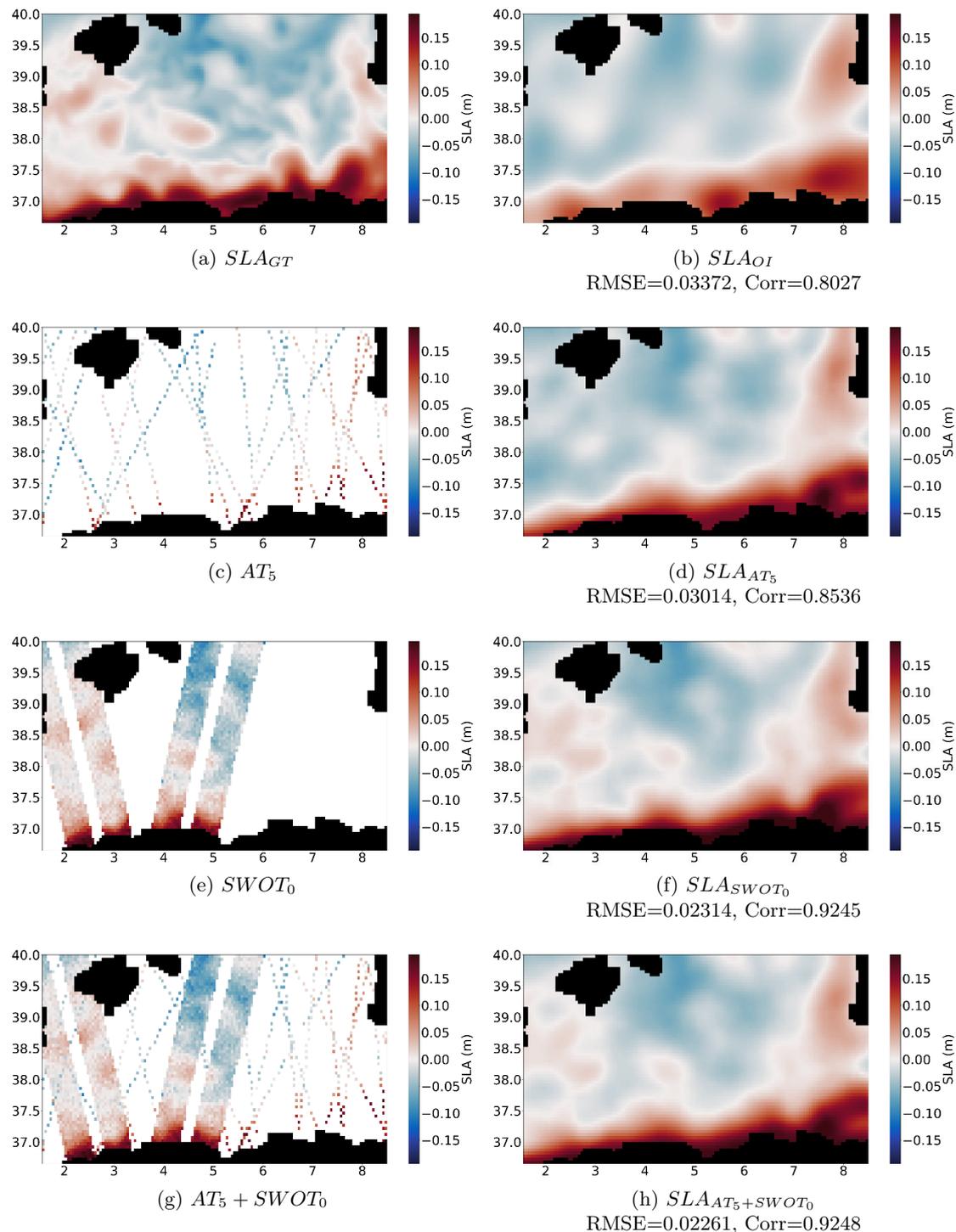


Figure 8.12 – AnDA SLA reconstruction results for the 9th March, 2013. (8.12a) Real high-resolution ground-truth SLA field. (8.12b) Optimal Interpolation SLA reconstruction obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of 10 days and a spatial correlation scale of 100 km. (8.12c) Nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as AT_5). (8.12d) AnDA SLA reconstruction from nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as AT_5 and depicted in Figure (8.12c)). (8.12e) Daily SWOT observations (noted as $SWOT_0$). (8.12f) AnDA SLA reconstruction from daily SWOT observations (noted as $SWOT_0$ and depicted in Figure (8.12e)). (8.12g) Fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (noted as $AT_5 + SWOT_0$). (8.12h) AnDA SLA reconstruction from the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (noted as $AT_5 + SWOT_0$ and depicted in Figure (8.12g)). RMSE and correlation values given for each reported result. Adapted from [170]. © 2019 IEEE.

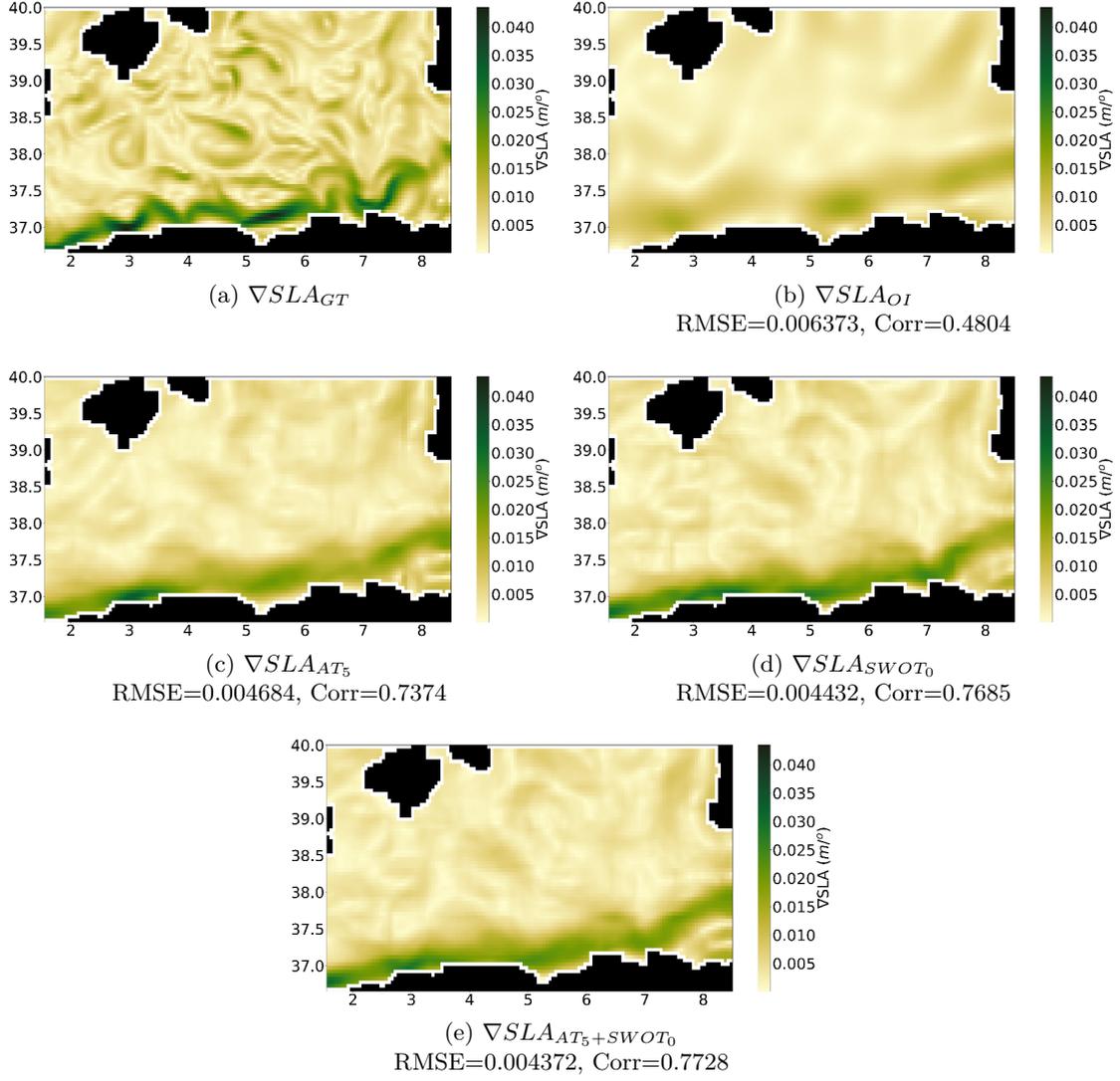


Figure 8.13 – AnDA SLA gradient reconstruction results for the 9th March, 2013. (8.13a) Real high-resolution ground-truth SLA gradient field. (8.13b) Optimal Interpolation SLA gradient reconstruction obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of 10 days and a spatial correlation scale of 100 km. (8.13c) AnDA SLA gradient reconstruction from nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as AT_5 and depicted in Figure (8.12c)). (8.13d) AnDA SLA gradient reconstruction from daily SWOT observations (noted as $SWOT_0$ and depicted in Figure (8.12e)). (8.13e) AnDA SLA gradient reconstruction from the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (noted as $AT_5 + SWOT_0$ and depicted in Figure (8.12g)). RMSE and correlation values given for each reported result. Adapted from [170]. © 2019 IEEE.

effect to some extent.

Furthermore, we analyze the time series of the proportion of the total SLA field covered by observations, depicted in Figure 8.15, where we present the observation coverage proportion for nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (Figure 8.15, top) and for the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (Figure 8.15, middle). The

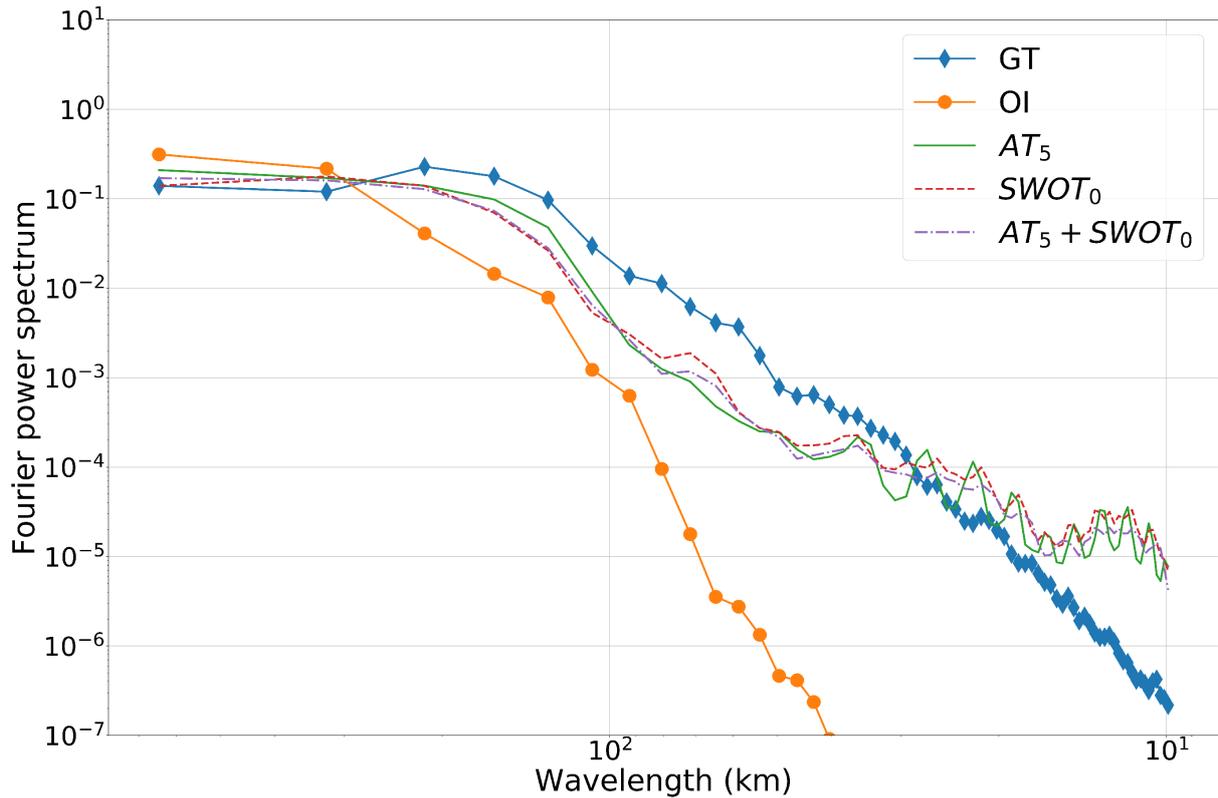


Figure 8.14 – Radially-averaged power spectra for AnDA reconstruction results on the 9th March, 2013. Reported power spectra correspond to the high-resolution ground-truth SLA field (GT), the Optimal Interpolation SLA reconstruction obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of 10 days and a spatial correlation scale of 100 km (OI), the AnDA SLA reconstruction from nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (AT_5), the AnDA SLA reconstruction from daily SWOT observations ($SWOT_0$) and the AnDA SLA reconstruction from the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations ($AT_5 + SWOT_0$). Adapted from [170]. © 2019 IEEE.

coverage proportion gain between them, i.e., the coverage gain obtained when complementing nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days with daily SWOT data, is also depicted (Figure 8.15, bottom). Similarly to Figure 8.11, the red filling indicates periods during which the nRMSE gain is positive, i.e. when the additional use of daily SWOT data improves interpolation performance ($nRMSE_{AT_5+SWOT_0} \leq nRMSE_{AT_5}$). As expected, the few days depicting a loss in reconstruction performance relate to periods with maximal nadir along-track observation coverage and minimal SWOT coverage, i.e., days when most of the region is sampled by nadir along-track observations and no SWOT observations exist. In this respect, one may indeed expect that for days of maximal nadir along-track observation coverage, the assimilation of nadir along-track observations only may perform nearly as well as or even slightly better than the assimilation of the fusion of nadir along-track and SWOT data, specially if no SWOT observations exists during such days of maximal nadir along-track observation coverage. We hypothesize that this behaviour may also relate to inconsistencies between SWOT and nadir along-track observations.

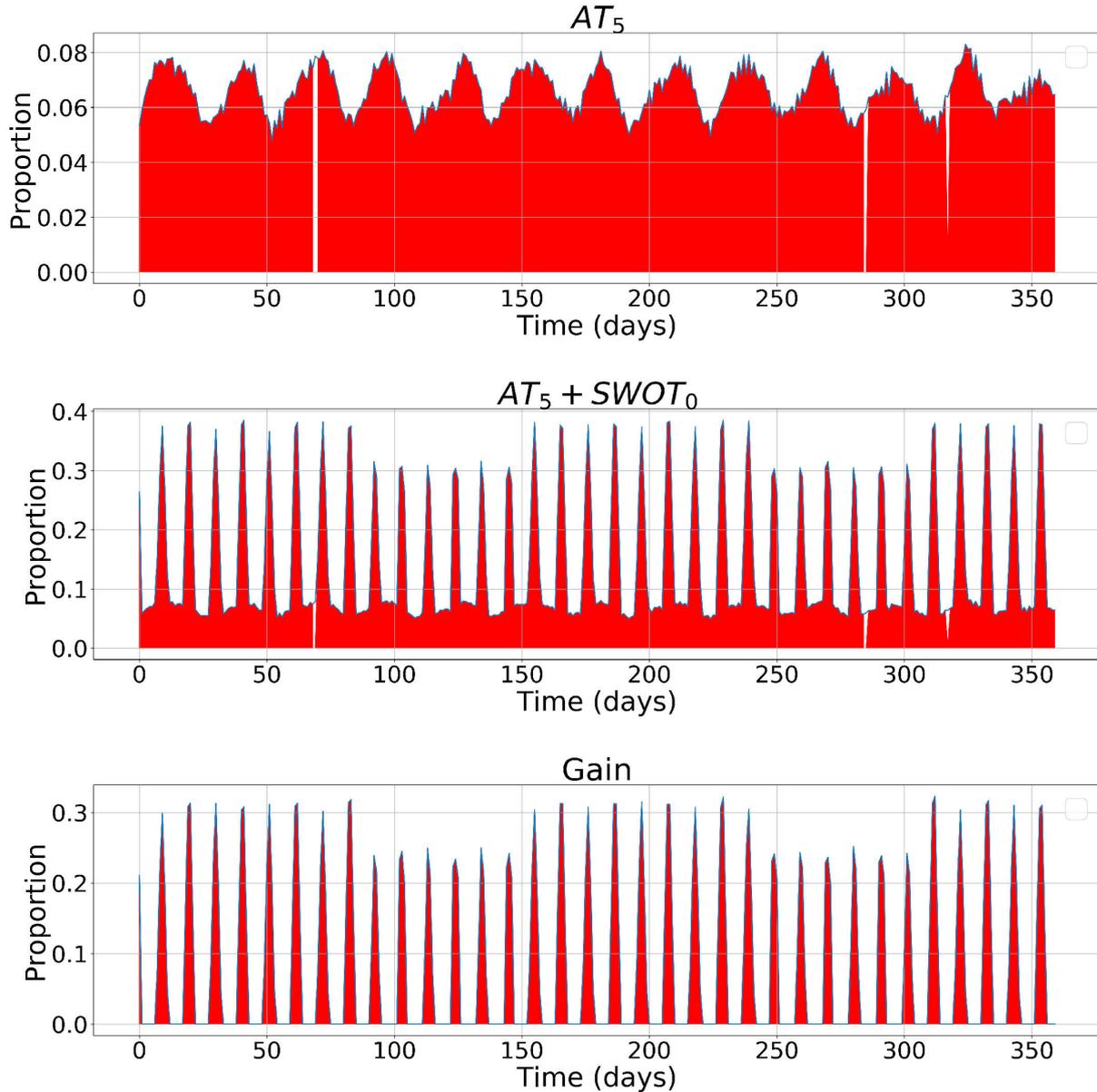


Figure 8.15 – Time series of the observation coverage (given as the ratio between the number of observed pixels and the total number of pixels in the considered region, excluding land pixels) for nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (top, noted as AT_5) and for the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (middle, noted as $AT_5 + SWOT_0$). The time series of the coverage gain obtained when complementing nadir along-track data accumulated on a time window $t_0 \pm D$ with $D = 5$ days with daily SWOT data is also depicted (bottom, noted as Gain). The red filling indicates periods during which the nRMSE gain is positive, i.e. when the additional use of daily SWOT data improves interpolation performance ($nRMSE_{AT_5+SWOT_0} \leq nRMSE_{AT_5}$).

Finally, we further analyze interpolation performance in relation to the spatial coverage of the available observations. We directly compare, in Figure 8.16, the spatial coverage of the joint nadir along-track/SWOT dataset (depicted previously in Figure (8.15, middle)) to the nRMSE interpolation gain when complementing nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days with daily SWOT data (given by $nRMSE_{AT_5} - nRMSE_{AT_5+SWOT_0}$, previously depicted in Figure (8.11, bottom)). The red filling indicates periods during which the

nRMSE gain is positive, i.e. when the additional use of daily SWOT data improves interpolation performance ($nRMSE_{AT_5+SWOT_0} \leq nRMSE_{AT_5}$). Whereas the spatial coverage of nadir along-track observations is characterized by a relatively constant baseline around a 5% coverage, the SWOT sampling results in periodic spikes in the coverage time series, with a characteristic period of 10 days and strong coverage variations from 0% to 40%. The nRMSE gain time series depicted in Figure (8.16, bottom) is significantly correlated to this periodic spikes pattern ($R=0.60$, $p<0.001$). This indicates that the improvement in interpolation performance is mainly observed for dates where at least one SWOT track passes through the case-study region. Overall, whereas the mean nRMSE gain is around 5.0% for the whole year 2013, it increases to 7.0% for days when at least one SWOT track is observed in the case-study region. We may also notice that peaks in Figure (8.16, bottom) are visually wider than those in Figure (8.16, top), which supports some ability of AnDA to propagate SWOT information over a few days. Moreover, a mean nRMSE gain of 3.5% w.r.t. the AnDA interpolation of nadir along-track data only is observed for days where no SWOT observation exist, while a very small fraction of these days (around 1.4%) involves a loss in reconstruction performance (with a mean nRMSE loss of around 0.2%). This further highlights the performance gain associated with the fusion of along-track and SWOT data, specially when compared to equivalent results obtained for the AnDA assimilation of SWOT data alone.

It should also be noted that both accumulating nadir along-track observations over a time window and the use of wide-swath SWOT observations (or a combination of both) have proven to be effective strategies to outperform OI, MS-VE-DINEOF and NN-LLOD. Indeed, when compared to OI, MS-VE-DINEOF and NN-LLOD, AnDA presents lower RMSE levels and higher correlation coefficients, which indicates that mesoscale structures are being better recovered.

8.9 Additional regressors for analog forecasting

One of the main advantages of using the locally-linear analog forecasting model (with respect to both classical model-based forecasting and other analog forecasting formulations) is that one may consider additional sources of information as regression variables when computing the analog forecast. In this way, the linear regression model relating the K nearest analogs and their successors may involve not only the high-resolution SLA fields in the catalog, but also additional data sources, such as low-resolution OI interpolated SLA fields, numerically-resolved gradients, sea surface temperature (SST) fields, etc.

In this section, we explore the effect of considering both the low-resolution OI-interpolated SLA fields as well as the numerically-resolved high-resolution SLA gradients as additional regressors. Following previous results, we consider the best case scenario identified earlier and consider the assimilation of the fusion of nadir along-track data accumulated on a time window $t_0 \pm D$ with $D = 5$ days and observation noise variance $\sigma_n = 3 \text{ cm}^2$ and daily SWOT observa-

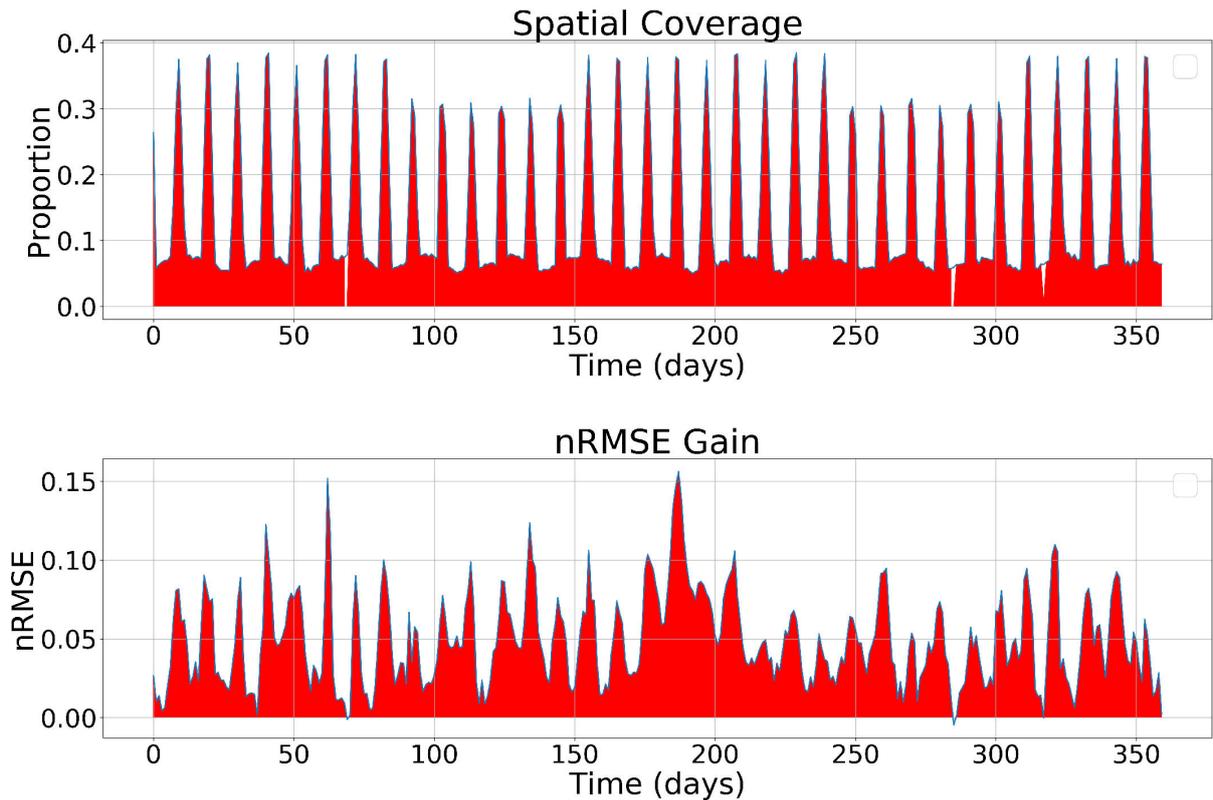


Figure 8.16 – (Top) Time series of the observation coverage (given as the ratio between the number of observed pixels and the total number of pixels in the considered region, excluding land pixels) for the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations. (Bottom) Time series of the nRMSE gain obtained when complementing nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days with daily SWOT data (given by $nRMSE_{AT_5} - nRMSE_{AT_5+SWOT_0}$). The red filling indicates periods during which the nRMSE gain is positive, i.e. when the additional use of daily SWOT data improves interpolation performance ($nRMSE_{AT_5+SWOT_0} \leq nRMSE_{AT_5}$).

tions with correlated KaRIn noise only. We also consider the following base parameter settings: $W_p = 30$ pixels (~ 150 km, 1 pixel ≈ 5 km), $K = 25$ neighbours, $\delta t = 1$ days, $N_{EOF} = 9$ EOF components. Obtained root mean squared error (RMSE) and correlation results are summarized in Table 8.6.

From Table 8.6, the contribution of considering OI-resolved low-resolution SLA fields as an additional regressor seems to be marginal, while considering numerically-resolved high-resolution SLA gradient fields as an additional regressor seems to provide a slight improvement in reconstruction performance.

To further validate these conclusions, we present in Figure 8.17 the radially averaged power spectra of the ground-truth and reconstructed SLA fields when considering no additional regressors, and when considering low-resolution SLA fields and SLA gradient fields as regressors. We also depict the power spectrum of the OI-resolved SLA field as reference. Reported power spectra seem to suggest no noticeable improvement in reconstruction performance is produced

Table 8.6 – Root mean squared error (Correlation) for AnDA SLA and SLA gradient (∇ SLA) reconstruction from the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations ($AT_5 + SWOT_0$), and considering different additional regressors, namely no additional regressors, OI-resolved low-resolution SLA fields, and numerically-resolved high-resolution SLA gradient fields. Best result in **bold**.

Setting	SLA	∇ SLA
$AT_5 + SWOT_0$	0.01687 (0.9607)	0.004286 (0.8051)
$AT_5 + SWOT_0 + \text{OI regressors}$	0.01679 (0.9612)	0.004277 (0.8060)
$AT_5 + SWOT_0 + \nabla$SLA regressors	0.01656 (0.9626)	0.004218 (0.8088)

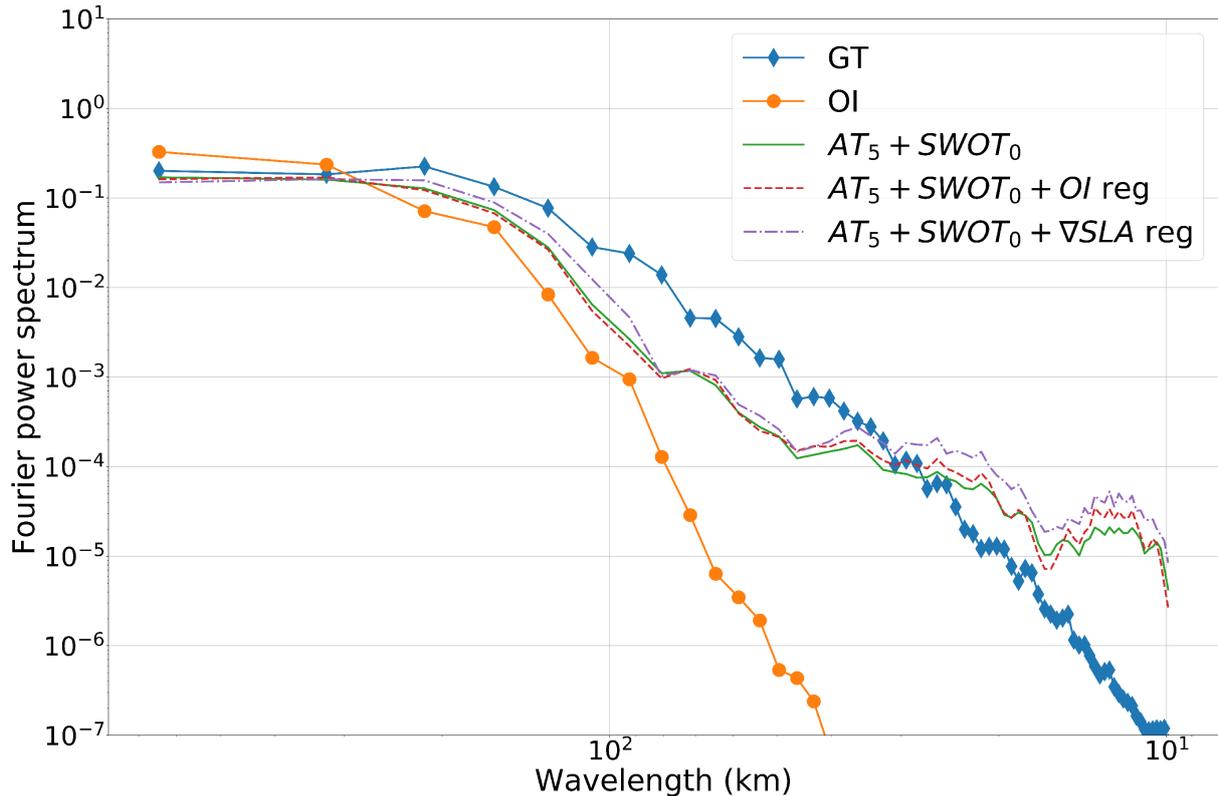


Figure 8.17 – Radially-averaged power spectra for AnDA reconstruction results when considering additional regressors on the 9th March, 2013. Reported power spectra correspond to the high-resolution ground-truth SLA field (GT), the Optimal Interpolation SLA reconstruction obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of 10 days and a spatial correlation scale of 100 km (OI), and the AnDA SLA reconstruction from the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations when no additional regressors are considered ($AT_5 + SWOT_0$), when OI-resolved low-resolution SLA fields are considered as additional regressors ($AT_5 + SWOT_0 + \text{OI reg}$), and when numerically-resolved high-resolution SLA gradient fields are considered as additional regressors ($AT_5 + SWOT_0 + \nabla \text{SLA reg}$).

by the inclusion of low-resolution SLA fields as additional predictors. The introduction of high-resolution SLA gradient fields as additional regressors, on the contrary, seems to provide a slight improvement in reconstruction performance for all spatial scales.

Figure 8.18 illustrates some reconstruction results by presenting ground-truth SLA and ∇ SLA fields (Figures 8.18a and 8.18b), as well as interpolated SLA fields and ∇ SLA fields for

the base case considering the AnDA assimilation of the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (Figures 8.18c and 8.18d) and for alternative cases considering as additional regressors the OI-resolved low-resolution SLA fields (Figures 8.18e and 8.18f) and the numerically-resolved high-resolution SLA gradient fields (Figures 8.18g and 8.18h). These results are in agreement with the power spectra reported in Figure 8.17. Indeed, no noticeable improvement in reconstruction can be observed when low-resolution SLA fields are used as additional regressors, while the introduction of numerically-resolved high-resolution SLA gradient fields seems to help produce results that are closer to the ground-truth fields, particularly for the reconstruction of SLA gradient fields.

8.10 Assimilation of observation gradients

As previously demonstrated, the presence of correlated noise sources in SWOT observations has a considerable impact on the performance of AnDA. In this respect, the exploration of techniques to filter or take such noise into account seem of the utmost importance to be able to fully exploit the scientific potential within SWOT observations when the mission goes airborne in 2021. Currently, efforts being made to deal with SWOT noise include developing filtering strategies [84] and considering SWOT noise structure within the covariance structure of model (6.1) in data assimilation issues [224].

As an alternative to these methods, we present here a proof of concept involving the assimilation of numerically-resolved observation gradients. Given that SWOT observations present a wide-swath two-dimensional nature, we may indeed compute and assimilate the gradient field of high-resolution SLA SWOT observations. The assimilation of true two-dimensional gradient fields from nadir along-track observations, on the other hand, is not possible given that only gradients along the satellite track can be computed from such observations. In the context of AnDA, we achieve this by modifying observation model \mathcal{H} so that it approximates the gradient of observations by means of a simple finite difference method, thus introducing a relationship between numerically-resolved gradients and system states that allows for the direct assimilation of SLA gradients. In this way, by modifying observation operator \mathcal{H} we may choose to assimilate both SLA observations and their gradients, or only SLA observation gradients.

To illustrate this, we present in Table 8.7 the root mean squared error (RMSE) and correlation for the AnDA assimilation of daily SWOT observations and numerically-resolved daily SWOT observation gradients, and considering both correlated and uncorrelated noise sources. We analyze three distinct cases, namely the assimilation of SWOT observations, the joint assimilation of SWOT observations and their numerically-resolved gradients, and the assimilation of numerically-resolved SWOT observation gradients only. For the reported experiments, we consider the following base parameter settings: $W_p = 30$ pixels (~ 150 km, 1 pixel ≈ 5 km), $K = 25$ neighbours, $\delta t = 1$ days, $D = 0$ days, $N_{EOF} = 9$ EOF components. From the reported results, a

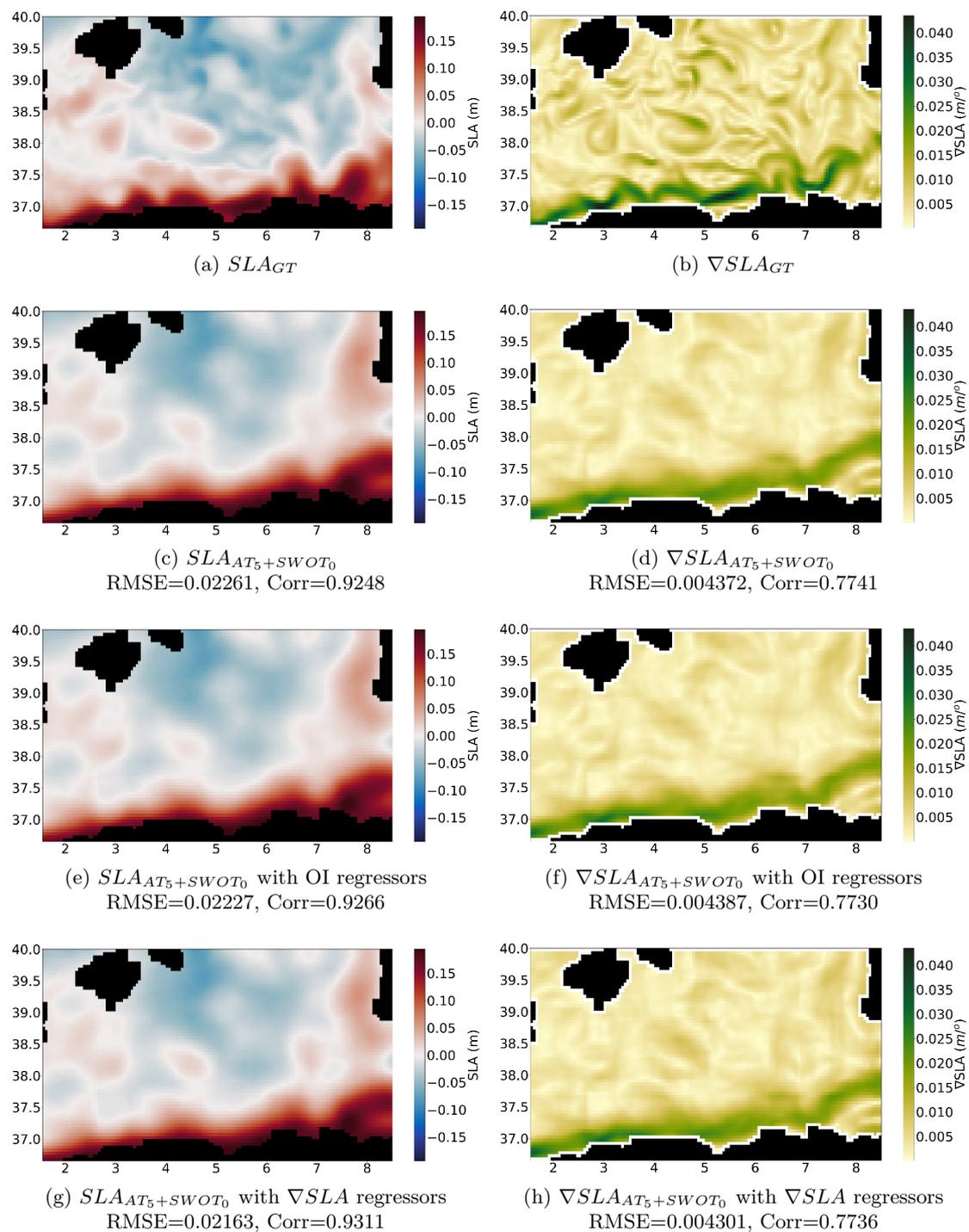


Figure 8.18 – AnDA SLA and SLA gradient (∇SLA) reconstruction results when considering additional regressors for the 9th March, 2013. Reported results depict the real high-resolution ground-truth SLA field (8.18a) and SLA gradient field (8.18b), as well as the AnDA SLA and SLA gradient reconstruction from the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations and considering: ((8.18c) and (8.18d)) no additional regressors, ((8.18e) and (8.18f)) OI-resolved low-resolution SLA fields as additional regressors, and ((8.18g) and (8.18h)) numerically-resolved high-resolution SLA gradient fields as additional regressors. RMSE and correlation values given for each reported result.

Table 8.7 – Root mean squared error (Correlation) for AnDA SLA and SLA gradient (∇ SLA) reconstruction when assimilating daily SWOT observations and/or their corresponding numerically-resolved SWOT observation gradients. Best result in **bold**.

Setting	SLA	∇ SLA
All noise sources + SLA assimilation	0.03765 (0.8143)	0.005791 (0.5900)
All noise sources + SLA and ∇ SLA assimilation	0.03749 (0.8134)	0.005756 (0.5931)
All noise sources + ∇SLA assimilation	0.03126 (0.8621)	0.006789 (0.5367)

small marginal gain can be observed with the joint assimilation of both SWOT observations and their corresponding gradients. The highest gain, however, involves the assimilation of SWOT observation gradients only.

These results may be explained by the fact that the assimilated observation gradients are computed using finite differences, which implicitly implements a mean filtering scheme. This implies gradients are thus robust to correlated noise sources, and may suggest indeed that considering observation gradients could be useful for the assimilation of SWOT observations contaminated by correlated noise. Even though the best obtained performance falls below that of previous results (even OI), this proof of concept intends to illustrate the potential of gradient assimilation as a means to deal with the correlated noise sources associated with the SWOT mission. In this respect, this suggest either a combination of gradient assimilation with the filtering or covariance modeling approaches mentioned before, as well as the development of more complex observation models capable of simultaneously assimilating multiple sources of SLA observations and their gradients; one may consider, for example, the joint assimilation of nadir along-track SLA observations and numerically-resolved SWOT observation gradients.

8.11 Conclusion

In the present chapter, we explored the potential of SWOT observations to overcome the limitations of currently available altimetry products for the reconstruction of high-resolution sea level anomaly (SLA) fields from satellite observations. We focused on a case-study region in the Western Mediterranean Sea, characterized by small Rossby radii, and performed observing system simulation experiments (OSSE) considering state-of-the-art data-driven interpolation schemes. Indeed, the ever-increasing availability of remote sensing, in situ and simulation datasets truly supports the exploration of data-driven approaches as a powerful alternative to classical model-driven schemes. These data-driven schemes have been shown to provide computationally-efficient and highly-flexible approaches and to significantly outperform OI. In this respect, they may be regarded as a means to exploit high-resolution numerical simulation datasets for the reconstruction of SLA fields from partial satellite observations. Among the evaluated schemes, the Analog Data Assimilation (AnDA) appeared as particularly relevant to explicitly model upper ocean

dynamics temporal dependencies, and led to the best interpolation performance.

Specifically, we developed an OSSE to evaluate the performance of the considered data-driven approaches for the interpolation of SLA fields from altimetry data. We further evaluated the sensitivity of AnDA to its most relevant internal parameters and to different noise sources, and benchmarked its performance for the assimilation of irregularly-sampled altimetry observations. With a view towards the upcoming SWOT altimetry mission, we considered different observation sampling patterns, namely satellite nadir along-track observations and wide-swath SWOT observations, and explored the joint analysis of both sources of altimetry data. Finally, we explored the use of alternative data sources, such as OI-resolved low-resolution SLA fields or numerically-resolved SLA gradient fields, as additional sources of information to enhance the reconstruction performance of AnDA. Reported results support the relevance of AnDA with respect to state-of-the-art approaches (OI [57, 204], MS-VE-DINEOF [161, 213] and NN-LLOD [169]) for the high-resolution reconstruction of sea level anomaly (SLA) fields. We report a clear improvement of 32% (12%) in terms of SLA RMSE (correlation) and 29% (27%) in terms of ∇ SLA RMSE (correlation) with respect to OI when considering the AnDA assimilation of nadir along-track data. Our experiments also suggest an additional SLA reconstruction performance improvement of 42% (12%) in terms of SLA RMSE (correlation) and 34% (29%) in terms of ∇ SLA RMSE (correlation) with respect to OI for the joint AnDA assimilation of SWOT and nadir along-track observations.

Importantly, these results highlight the potential of SWOT observations to provide more information than currently available nadir along-track altimetry observations. Overall, our experiments truly support the potential of SWOT observations to better constraint AnDA and improve the reconstruction of mesoscale features in satellite-derived altimetry fields. We may draw three main conclusions:

- The joint use of nadir along-track and SWOT in AnDA naturally arises as the best strategy, with a relative interpolation gain of about 14.71% (resp. 42.36%) in terms of SLA RMSE w.r.t. AnDA (resp. OI) applied solely to nadir along-track data.
- Contrary to nadir along-track, the accumulation of SWOT data over consecutive days does not help, and even negatively affects, interpolation performance.
- The irregular time-sampling of SWOT observations for a given region, typically involving a revisit period of 10 days for mid-latitudes, results in a relatively large time variability of interpolation performance.

Throughout most of this chapter, the considered OSSE involved realistic uncorrelated noise levels for both nadir along-track and SWOT data. For SWOT data, however, correlated noise sources are also expected. Given that AnDA has proven to be considerably sensitive to these correlated noise sources, our results might then be regarded as upper-bounds for the improvement

to be expected from the SWOT mission for the reconstruction of satellite-derived altimetry fields. Importantly, such noise sources need to be appropriately accounted for in the interpolation model considered [171], as previously illustrated in this chapter. In this respect, we introduced the assimilation of numerically-resolved SWOT observation gradients as a possible strategy to deal with correlated noise sources contaminating SWOT data. However, additional preprocessing should still be carried out to filter correlated noise in SWOT data [84] or take its sources into account in the correlation structure of data assimilation schemes [224]. Regarding the valorization of these contributions, reported results have been presented in both national and international conferences [171,172], and constitute the main subject of a journal article [170].

As far as future work perspectives are concerned, future developments should focus on combining the above mentioned SWOT noise processing strategies with the AnDA framework, in order to specifically address correlated noise sources either within the assimilation models [224] or through a dedicated preprocessing step [84]. In this regard, the joint assimilation of SWOT observation gradients and nadir along-track SLA data appears as a promising alternative to deal with the correlated noise sources present in SWOT data. Hopefully, the combination of the above mentioned strategies will help develop useful tools to process real observations from the future SWOT altimetry mission. Other interesting research avenues include the combination of additional sources of altimetry data, as well as the exploitation of the synergy with alternative ocean dynamical tracers (e.g. sea surface temperature, sea surface salinity, ocean color, etc.). Further exploring the exploitation of structural information present in wide-swath observations, for example by means of finite size Liapunov exponents (FSLE), is also an appealing research direction.

Part III

Conclusion and Perspectives

Conclusion and Perspectives

*"This is the end
Beautiful friend
This is the end
My only friend
The end"*

Jim Morrison

9.1	Conclusions	177
9.2	Perspectives	179
9.3	Final thoughts	181

9.1 Conclusions

Throughout this thesis dissertation we aimed at exploring data-driven approaches to tackle a variety of inverse problems in the context of physical oceanography and geosciences, with a special focus on the reconstruction of upper ocean dynamics from satellite observations. In this regard, we explored the exploitation of data-driven methodologies for the resolution of both segmentation and reconstruction issues in multiple scientific contexts. Importantly, this work contributes to ongoing joint efforts between the geosciences and signal processing communities to exploit the ever growing availability of remote sensing, in situ and modeling/reanalysis data to improve our understanding of ocean dynamics and ocean-atmosphere interactions. In this respect, this thesis dissertation is, given its subject matter, at the interface between data science and geosciences. As such, we attempted throughout this work to partially bridge the gap between signal processing and geosciences issues to improve current ocean dynamics segmentation and reconstruction models.

Inspired by the recent success of non-negative and sparse blind source separation formulations in the context of signal and image processing applications, in the first part of this work we extended constrained blind source separation models to the problem of the observation-based

characterization and decomposition of operators. We developed relevant models and proposed efficient and mathematically-sound algorithms for the estimation of model parameters. Besides the benchmarking and evaluation of model robustness and algorithm complexity and performance for a variety of ideal and non-ideal setting, we also explored a number of applications for both the segmentation of upper ocean dynamics from remote sensing data and the reconstruction/forecasting of dynamical systems, most notably for the well-known Lorenz '96 dynamical model. Our results support the relevance of the proposed formulations, with respect to more classical formulations involving orthogonal priors and spatio-temporally invariant operators, for both the segmentation and reconstruction of system dynamics.

In a second part of this dissertation, we shifted our focus towards a particular inverse problem, namely the interpolation of high-resolution geophysical fields from irregularly-sampled data, with a particular interest on the interpolation of sea level anomaly (SLA) fields from satellite-derived altimetry observations. In preparation for the launch of the SWOT mission in 2021, we explored the effect of considering different sampling patterns, more particularly those relating to classic along-track nadir altimetry and wide-swath off-nadir radar interferometry.

Initially, we tackled the problem of the interpolation of SLA fields from satellite observations by reformulating it as an image super-resolution issue exploiting convolutional operators. We resorted to the operator decomposition models introduced in the first part of this work as a means to locally adapt the super-resolution model in smaller spatial domains, thus achieving a more performant reconstruction. In this regard, the proposed model and algorithms constitute a novel methodological contribution to the problem of the spatio-temporal interpolation of SLA fields from irregularly-sampled data. The convolutional operator formulation also us allowed to benefit from additional sources of complementary information, such as high-resolution sea surface temperature (SST) fields. In particular, we evaluated different model constraints for the decomposition of convolutional operators, and demonstrated the relevance of non-negative issues, with respect to orthogonal or sparse-base priors, for the data-driven interpolation of SLA fields from satellite-derived, irregularly-sampled altimetry observations.

Despite the relevance of non-negative operator decomposition formulations, we subsequently evolved towards alternative data-driven formulations exploiting analog forecasting in the context of data assimilation, given their recently demonstrated suitability for the interpolation of geophysical fields from irregularly-sampled, gap-ridden datasets. Following previous efforts, we applied the Analog Data Assimilation (AnDA) framework to our problem of interest. In this respect, AnDA can be regarded as a means to exploit high-resolution simulation/reanalysis datasets to the reconstruction of SLA fields from partial satellite observations by emulating numerical models in the context of data assimilation, thus obtaining a more performant reconstruction not subject to restrictive model hypothesis, as may be the case for classical data

assimilation. To optimize the application of the proposed framework to our particular situation, we investigated the sensitivity of AnDA to its most relevant internal parameters and to noise and error sources affecting satellite-derived altimetry datasets. With a view towards the future SWOT mission, we explored, in particular, the effect of considering both nadir along-track and SWOT observations independently, as well as a fusion of both data sources in order to exploit the synergy between nadir along-track and SWOT observations. Moreover, the proposed formulation permitted the exploitation of complementary information sources, such as OI-resolved low resolution SLA fields or numerically-resolved high-resolution SLA gradient fields, as regressors in the forecasting step of the proposed data-driven data assimilation framework. In particular, we identified a clear limitation relating to the application of stochastic data assimilation schemes to assimilate observations contaminated by correlated noise, as is the case for SWOT data. In this respect, we also explored the assimilation of numerically-resolved gradients as a possible alternative to deal with such correlated noise source, even though reported results suggest more research in this direction.

9.2 Perspectives

As far as possible future work perspectives are concerned, many interesting research avenues have been identified throughout the present thesis work.

From a methodological point of view, more work is needed in order to improve the robustness and performance of the operator decomposition models proposed in the first half of this thesis work, in particular for non-ideal cases involving noisy mixing parameters where model identifiability may become compromised due to the relaxation of the parameter sharing hypothesis necessary for the computation of local linear operators. Considering robust estimation approaches for local linear operators and alternative or additional constraints on either mixing coefficients α_{nk} or regression matrices β_k may prove to be effective means of increasing model robustness and/or rendering it more suitable for alternative applications. In particular, imposing additional constraints on the regression matrices may be of particular interest for the interpretation of the identified modes.

Additionally, non-local extensions of operator decomposition models to combine spatio-temporal and similarity-based neighbour selection for the estimation of local linear operators, as considered in regression-based super-resolution models [2, 253], seem of particular relevance for geophysical field interpolation applications. Finally, the extension of the proposed models to non-linear decompositions, for example by exploiting kernel-based formulations [230, 262] or convolutional neural networks (CNN) based models [49], and to non-Gaussian settings also appear as a relevant research directions. In particular, non-linear dictionary-based decompositions may prove useful for combining non-linear mappings and locally-adapted models.

As far as practical applications for the proposed observation-based operator decomposition framework are concerned, future work should aim at identifying possible applications where the decomposition of local linear operators may be suitable to help produce more relevant or easier to interpret decompositions, as well as increase reconstruction performance. It may also involve further development of previously studied applications, such as the analysis and reconstruction of upper ocean dynamics from joint SSH-SST observations or other alternative ocean tracers. Further exploring the geophysical interpretation of model parameters and their relation to other geophysical quantities should also be considered. In this respect, the application of the proposed data-driven analysis to numerical simulations associated with idealized geophysical conditions is of key interest [131]. The use of additional local (potential) descriptors of upper ocean dynamics within the proposed methodology, such as SST gradient, associated quantities like SST gradient divergence and curl, or other geophysical tracers, such as salinity and chlorophyll-a [227], could also be considered. Moreover, the application of the proposed model at a global scale appears promising to reveal shared or differentiating dynamical modes.

As far as data-driven geophysical field interpolation applications are concerned, applying the NN-LLOD image super-resolution framework to the interpolation of SLA fields from SWOT data might prove useful to gain further insight into the strengths and limitations of the proposed operator decomposition model and further compare it to the Analog Data Assimilation framework. Indeed, as illustrated by NN-LLOD results, using operator decomposition models for the reconstruction of high-resolution SLA fields from simultaneous partial SLA observations and high-resolution SST observations appears as a promising approach in the context of current and future altimeter missions, and should be further explored. With regards to the Analog Data Assimilation application considered, we once again stress the importance of combining AnDA with present efforts to filter correlated noise sources in SWOT or, alternatively, take them into consideration for the definition of covariance structures in data assimilation issues. In particular, combining these approaches with the assimilation of numerically-resolved SLA observation gradients is of particular interest for the development of relevant tools for the analysis and processing of SWOT observations once the mission is airborne. Moreover, the possibility of combining multiple sources of altimetry data and considering additional oceanic tracers as regressors in the AnDA framework opens up a number of interesting research avenues that should be explored. Last but not least, as mentioned earlier, the two-dimensional nature of wide-swath off-nadir interferometry observations will enable the exploitation of two-dimensional information within altimetry observations by means of numerically-resolved gradients or finite size Liapunov exponents (FSLE), among other techniques. In this respect, the unparalleled observational capabilities of the SWOT mission encompass a rich scientific potential for an improved understanding of ocean dynamics, ocean-atmosphere interactions and the role of the ocean in critical earth system processes, such as climate change, and should thus be investigated thoroughly.

Besides oceanography issues, other application fields, such as the blind identification of MIMO (Multiple Input Multiple Output) systems that can be applied, for example, to the multi-path problem in communication systems [1, 28, 56], might also be explored.

Finally, one may also consider the integration of the proposed operator decomposition approaches into analog data assimilation methods. In particular, this may be achieved by considering the decomposition of local linear analog forecasting operators onto a dictionary computed from the catalog of past observations. Preliminary experiments on this technique did not yield an increase in performance when tested on the considered SLA datasets. However, given the strong dependence of the operator decomposition reconstruction performance on the considered application and on dataset characteristics, we believe this hybrid approach may prove useful in non-ideal settings involving noisy or small catalogs, which have not been explored in the present work due to time constraints. In this respect, further research is needed to verify this hypothesis and identify relevant applications for this approach.

9.3 Final thoughts

To conclude, we would like to emphasize once again the potential of observation-based operator decomposition models as a tool for analyzing ocean tracer couplings. In particular, we hope the developed models will allow for a deeper understanding of the interactions between ocean quantities, the processes governing ocean dynamics and the role of the ocean within the global earth system.

As mentioned in the introductory chapter of this work, we also hope our findings will help new developments in terms of upper ocean dynamics segmentation, high-resolution interpolation of geophysical fields from multi-source high-resolution observations and related applications, such as the high-resolution reconstruction of sea surface currents, among others.

List of Figures

4.1	Probability density function (PDF) for the normalized mean squared estimation error (nMSE) for mixing coefficients α_{nk} . Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD). All presented probability distributions were computed using a Gaussian kernel.	66
4.2	Probability density function (PDF) for the normalized mean squared estimation error (nMSE) for linear modes β_k . Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD). For ALS, two different initialization schemes for α_{nk} are presented: a random initialization and an initialization based on the binary clustering of observations \mathbf{x}_n into K classes. All presented probability distributions were computed using a Gaussian kernel.	67
4.3	Normalized mean squared estimation error (nMSE) median evolution for mixing coefficients α_{nk} . Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).	69
4.4	Normalized mean squared estimation error (nMSE) median evolution for linear modes β_k . Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).	70
4.5	Normalized mean squared estimation error (nMSE) final median value (at convergence) for mixing coefficients α_{nk} as a function of the number of classes K considered. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).	70

4.6	Normalized mean squared estimation error (nMSE) final median value (at convergence) for linear modes β_k as a function of the number of classes K considered. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).	71
4.7	Normalized mean squared \mathbf{y}_n reconstruction error (nMSE) final median value (at convergence) as a function of the number of classes K considered. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).	71
4.8	Normalized mean squared estimation error (nMSE) final median value (at convergence) for mixing coefficients α_{nk} as a function of mixing coefficient signal-to-noise ratio (SNR) when Gaussian noise is added to cluster-specific mixing coefficients. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).	73
4.9	Normalized mean squared estimation error (nMSE) final median value (at convergence) for linear modes β_k as a function of mixing coefficient signal-to-noise ratio (SNR) when Gaussian noise is added to cluster-specific mixing coefficients. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).	74
4.10	Normalized mean squared \mathbf{y}_n reconstruction error (nMSE) final median value (at convergence) as a function of mixing coefficient signal-to-noise ratio (SNR) when Gaussian noise is added to cluster-specific mixing coefficients. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).	75
4.11	Normalized mean squared estimation error (nMSE) final median value (at convergence) for mixing coefficients α_{nk} as a function of the ratio between cluster standard deviation σ_x and minimal distance d_{min} (parameter $\lambda = 6\frac{\sigma_x}{d_{min}}$), for initially non-overlapping clusters. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD).	76

4.12 Normalized mean squared estimation error (nMSE) final median value (at convergence) for linear modes β_k as a function of the ratio between cluster standard deviation σ_x and minimal distance d_{min} (parameter $\lambda = 6\frac{\sigma_x}{d_{min}}$), for initially non-overlapping clusters. Results presented for the ALS algorithm using a gradient descent approach (ALS-gradient), the ALS algorithm using a direct minimization (ALS-direct) and the dictionary-based decomposition of local linear operators (LLOD). 77

4.13 Normalized mean squared estimation error (nMSE) final median value (at convergence) for mixing coefficients α_{nk} as a function of the ratio between cluster standard deviation σ_x and minimal distance d_{min} (parameter $\lambda = 6\frac{\sigma_x}{d_{min}}$), for initially non-overlapping clusters. Results presented for the original dictionary-based local linear operator decomposition (LLOD) and for two robust variants, namely Filtered LLOD, a filtering of local linear operators with mean value deviation higher than $f_c = \sigma_{\Phi}$ (where σ_{Φ}^2 is the local linear operator mean value variance), and Robust LLOD, which involves the iterative re-weighted least squares estimation of local linear operators Θ_n 78

4.14 Normalized mean squared estimation error (nMSE) final median value (at convergence) for linear modes β_k as a function of the ratio between cluster standard deviation σ_x and minimal distance d_{min} (parameter $\lambda = 6\frac{\sigma_x}{d_{min}}$), for initially non-overlapping clusters. Results presented for the original dictionary-based local linear operator decomposition (LLOD) and for two robust variants, namely Filtered LLOD, a filtering of local linear operators with mean value deviation higher than $f_c = \sigma_{\Phi}$ (where σ_{Φ}^2 is the local linear operator mean value variance), and Robust LLOD, which involves the iterative re-weighted least squares estimation of local linear operators Θ_n 79

4.15 Normalized mean squared parameter estimation errors (nMSE) for the single observation version of the ALS-direct algorithm (Algorithm 1) applied to a single-cluster ground-truth synthetic dataset without parameter sharing between observations. *Results presented for both the EM and pseudo-EM initialization variants of the considered algorithm. Reported initialization variants are based on model (3.4), a latent class simplification of the proposed decomposition model (3.3). Vertical lines indicate the iteration number at which the change between the EM (pseudo-EM) procedure for the latent class model (3.4) and the ALS procedure for general model (3.3) occurs. Taken from [166]. © 2016 IEEE.* 80

4.16	Relative mean squared reconstruction error (rMSE) histograms for the single observation version of the ALS-direct algorithm (Algorithm 1) applied to a single-cluster ground-truth synthetic dataset without parameter sharing between observations. <i>Results presented for both the EM and pseudo-EM initialization variants of the considered algorithm. Reported initialization variants are based on model (3.4), a latent class simplification of the proposed decomposition model (3.3).</i> . . .	81
4.17	Log-likelihood $\mathcal{L}^{(i)}$ (<i>top</i>) and relative change in log-likelihood $\Delta\mathcal{L}^{(i)}$ (<i>bottom</i>) evolution for the single observation version of the ALS-direct algorithm (Algorithm 1) applied to a single-cluster ground-truth synthetic dataset without parameter sharing between observations. <i>Results presented for both the EM and pseudo-EM initialization variants of the considered algorithm. Reported initialization variants are based on model (3.4), a latent class simplification of the proposed decomposition model (3.3).</i>	83
5.1	Illustration of the complete dictionary-based local linear analog forecasting operator non-negative decomposition. <i>Nearest neighbours are searched (in the train dataset) for observations in the train dataset to estimate local linear operators, which are then decomposed to obtain a dictionary of modal matrices β_k. Local linear operators are subsequently estimated for observations in the test dataset (new observations) from nearest neighbours in the train dataset, and then projected onto the manifold spanned by modal matrices β_k under non-negativity constraints. The forecasting for the next time step simply amounts to the application of the projected linear operator to the current state.</i>	88
5.2	Forecasting performance of local linear analog forecasting operators: we depict the normalized forecasting RMSE vs. the number of analogs used to estimate the locally-linear analog forecasting operator for Lorenz '96 dynamics for: locally-linear operators (<i>Full line</i>), the proposed non-negative decomposition of locally-linear operators with $K = 4$ modes (<i>Dashed line</i>). We consider three experimental settings: a noise-free scenario with a large catalog of 2×10^5 exemplars, a noisy scenario with a large catalog of 2×10^5 exemplars, a noise-free scenario with a smaller catalog of 2×10^3 exemplars.	89
5.3	Illustration of the reconstruction of Lorenz '96 dynamics using the proposed dictionary-based local linear analog forecasting operator non-negative decomposition. <i>Real Lorenz '96 time series alongside with the reconstructed Lorenz '96 for the base case involving a large, noise-free catalog and considering $K = 250$ neighbours.</i>	90
5.4	Correlation time series of the considered ground-truth sea surface salinity (SSS) and sea surface temperature (SST) fields in the Alboran Sea.	91

5.5	Non-negative decomposition of SST-SSS relationships in the Alboran Sea: (a) Ground-truth SST and SSS fields on March 22 nd , 2011, (b) SST-derived predictions of the SSS fields for each mode of the considered dictionary-based local linear operator decomposition considering $K = 2$ modes.	93
5.6	Distribution of SST-SSS correlation when mode 1 (resp. mode 2) of the two-class dictionary-based local linear operator decomposition dominates, i.e. $\alpha_{n1} > \alpha_{n2}$ (resp. $\alpha_{n2} > \alpha_{n1}$).	94
5.7	Ocean surface dynamics segmentation. <i>Predicted SSH per mode for the 1st January, 2004. Reported results correspond to both considered models, namely our proposed non-negative decomposition model and a latent class regression model introduced in [244], considering $K = 2$ classes. Taken from [166]. © 2016 IEEE.</i>	95
5.8	Sketch of the considered patch-based representation of the SST field (in degrees represented in false colors), denoted by \mathbf{x} , and the SSH field (in meters represented by contour lines), denoted by \mathbf{y} , at locations s_i and s_j and time t_i . <i>Adapted from [244]. © 2014 IEEE.</i>	97
5.9	SSH reconstruction error for the 1 st January, 2004. <i>Reported results correspond to our proposed non-negative decomposition model considering $K = 2$ classes. The Jet colormap was explicitly chosen to enhance contrast between positive and negative reconstruction error values.</i>	99
5.10	Time series of the daily FVE (Fraction of Variance Explained) for model (3.3): top, SSH prediction for the first mode (SQG-like mode); middle, global SSH prediction; bottom, global SSH gradient prediction. <i>FVE is computed as $1 - FVU$, with FVU being the fraction of variance unexplained, i.e., the ratio between the reconstruction error variance and the ground-truth field variance.</i>	101
5.11	Mean annual SSH field prediction for each dynamical mode of model (3.3): top, first mode (SQG-like mode); bottom, second mode.	102
5.12	Mean annual SSH prediction error standard deviation for the first mode (SQG-like mode) of model (3.3).	102
5.13	Mean annual mapping of mixing coefficients α for each dynamical mode of model (3.3): top, mode 1 (SQG-like mode); bottom, mode 2.	103
5.14	Time series of the daily mean of mixing coefficients α of model (3.3): top, mode 1 (SQG-like mode); bottom, mode 2.	104
5.15	Spatio-temporal variability of mixing coefficients α of model (3.3) with respect to four contrasted zones.	105
6.1	Comparison of conventional pulse-limited radar altimetry (left) and synthetic aperture radar (SAR) altimetry (right). <i>Credits: R.K. Raney, Johns Hopkins University Applied Physics Laboratory¹. Used with permission.</i>	112

6.2	Different quantities derived from altimetric measurements with reference to the Earth's Geoid and the Ellipsoid. <i>Credits: CLS. Source: duacs.cls.fr². Used with permission.</i>	113
6.3	Comparison of classic nadir along-track observations (6.3a) and off-nadir wide-swath SWOT observations (6.3b). <i>Adapted from [170]. © 2019 IEEE.</i>	115
6.4	SWOT satellite scheme. A 10 m boom separates two SAR antennas. Each antenna receives the backscatter from both swaths, which are illuminated by a single antenna with two different polarizations (to have two separate simultaneous measurements). Interferometry is used to estimate the OST for both swaths from the received backscatter on both antennas. <i>Taken from [76]. © American Meteorological Society. Used with permission.</i>	116
6.5	Comparison of along-track nadir-looking altimeters (left) and off-nadir wide-swath interferometry altimeters (right). <i>The off-nadir track allows for a wider footprint of the instrument, while a higher resolution can be achieved by exploiting interferometry between two off-nadir radiometers. Credits: AVISO³. Used with permission.</i>	117
7.1	Illustration of the irregular sampling of high-resolution observations associated with ocean remote sensing data: sea level anomaly image with the sampled along-track positions by satellite altimeters (cyan squares) in a ± 10 -day time window around April 20 th , 2012. <i>Adapted from [169]. © 2017 IEEE.</i>	134
7.2	Probability distribution for the normalized root mean square reconstruction error (nRMSE) for daily high-resolution SLA images $\{Y(t)\}_t$, for a global convolutional model and for locally-adapted decompositions of a global convolutional model using principal component analysis (PCA) [206], KSVD [3] and non-negative decomposition (NN) and considering $K = 10$ classes. The probability distribution of the nRMSE for daily low-resolution SLA images $\{Y_{LR}(t)\}_t$ is given as reference (noted as SLA_{LR}).	136
7.3	High-resolution SSH image Y reconstruction, April 20 th , 2012: first row, from left to right, real high-resolution SSH image Y , low-resolution SSH image Y_{LR} (noted as SSH_{LR}), reconstruction of high-resolution SSH image Y using global convolutional model (7.1); second row, reconstruction of high-resolution SSH image Y using a 10-class locally-adapted decomposition (7.4) of global convolutional model (7.1) using, from left to right, principal component analysis (PCA) [206], KSVD [3] and non-negative decomposition (NN). <i>Adapted from [169]. © 2017 IEEE.</i>	136
8.1	Comparison of nadir along-track observations (8.1b) and SWOT observations (8.1c) generated from ground-truth high-resolution SLA fields (8.1a) using real satellite tracks spatio-temporal locations and the SWOT simulator, respectively. <i>Adapted from [170]. © 2019 IEEE.</i>	147

8.2	<p>Pseudo-observations generated from the synthetic ground-truth SLA fields by considering either observations at current day t_0 or observations accumulated over a time window $t_0 \pm D$, with $D = 5$ days, for both nadir along-track altimetry (noted as AT_D with $D = \{0, 5\}$) and pseudo-SWOT observations (noted as $SWOT_D$ with $D = \{0, 5\}$) (8.2a) Daily nadir along-track observations. (8.2b) Nadir along-track observations accumulated on a window $t_0 \pm D$ with $D = 5$ days. (8.2c) Daily SWOT observations. (8.2b) SWOT observations accumulated on a window $t_0 \pm D$ with $D = 5$ days. (8.2e) Fusion of daily nadir along-track and SWOT observations. (8.2f) Fusion of nadir along-track and SWOT observations accumulated on a window $t_0 \pm D$ with $D = 5$ days. Adapted from [170]. © 2019 IEEE.</p>	147
8.3	<p>AnDA sensitivity to patch size W_p (1 pixel ≈ 5 km) (top) SLA reconstruction root mean squared error (RMSE) and correlation as a function of patch size (in km). (bottom) SLA gradient reconstruction root mean squared error (RMSE) and correlation as a function of patch size (in km). Taken from [171]. © 2018 IEEE.</p>	149
8.4	<p>AnDA sensitivity to number of neighbours K. (top) SLA reconstruction root mean squared error (RMSE) and correlation as a function of the number of neighbours used in the forecasting step of AnDA. (bottom) SLA gradient reconstruction root mean squared error (RMSE) and correlation as a function of the number of neighbours used in the forecasting step of AnDA. Taken from [171]. © 2018 IEEE.</p>	150
8.5	<p>AnDA sensitivity to assimilation lag δt. (top) SLA reconstruction root mean squared error (RMSE) and correlation as a function of the assimilation lag δt (in days) used for the assimilation of observations in AnDA. (bottom) SLA gradient reconstruction root mean squared error (RMSE) and correlation as a function of the assimilation lag δt (in days) used for the assimilation of observations in AnDA.</p>	151
8.6	<p>AnDA sensitivity to observation half-window size D. (top) SLA reconstruction root mean squared error (RMSE) and correlation as a function of the half-window size D (in days) for observations accumulated on a time window $t_0 \pm D$ days. (bottom) SLA gradient reconstruction root mean squared error (RMSE) and correlation as a function of the half-window size D (in days) for observations accumulated on a time window $t_0 \pm D$ days. Taken from [171]. © 2018 IEEE.</p>	151

- 8.7 Time series of the normalized root mean squared error (nRMSE) for the AnDA assimilation of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (top, noted as AT_5) and for the AnDA assimilation of daily SWOT observations (middle, noted as $SWOT_0$). The time series of the nRMSE gain obtained when considering daily SWOT observations rather than nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (given by $nRMSE_{AT_5} - nRMSE_{SWOT_0}$) is also depicted (bottom, noted as Gain). *The red filling indicates periods during which there is a positive nRMSE gain when considering daily SWOT data rather than nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (i.e. $nRMSE_{SWOT_0} \leq nRMSE_{AT_5}$).* 156
- 8.8 Time series of the observation coverage (given as the ratio between the number of observed pixels and the total number of pixels in the considered region, excluding land pixels) for nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (top, noted as AT_5) and for daily SWOT observations (middle, noted as $SWOT_0$). The time series of the coverage gain obtained when considering daily SWOT observations rather than nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days is also depicted (bottom, noted as Gain). *The red filling indicates periods during which there is a positive nRMSE gain when considering daily SWOT data rather than nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (i.e. $nRMSE_{SWOT_0} \leq nRMSE_{AT_5}$).* 157
- 8.9 AnDA SLA reconstruction results for the 2nd May, 2013. (8.9a) Real high-resolution ground-truth SLA field. (8.9b) Optimal Interpolation SLA reconstruction obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of 10 days and a spatial correlation scale of 100 km. (8.9c) Nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as AT_5). (8.9d) AnDA SLA reconstruction from nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as AT_5 and depicted in Figure (8.9c)). (8.9e) Daily SWOT observations (noted as $SWOT_0$). (8.9f) AnDA SLA reconstruction from daily SWOT observations (noted as $SWOT_0$ and depicted in Figure (8.9e)). (8.9g) SWOT observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as $SWOT_5$). (8.9h) AnDA SLA reconstruction from SWOT observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as $SWOT_5$ and depicted in Figure (8.9g)). RMSE and correlation values given for each reported result. 159

8.10 AnDA SLA gradient reconstruction results for the 2nd May, 2013. (8.10a) Real high-resolution ground-truth SLA gradient field. (8.10b) Optimal Interpolation SLA gradient reconstruction obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of 10 days and a spatial correlation scale of 100 km. (8.10c) AnDA SLA gradient reconstruction from nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as AT_5 and depicted in Figure (8.9c)). (8.10d) AnDA SLA gradient reconstruction from daily SWOT observations (noted as $SWOT_0$ and depicted in Figure (8.9e)). (8.10e) AnDA SLA gradient reconstruction from SWOT observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as $SWOT_5$ and depicted in Figure (8.9g)). RMSE and correlation values given for each reported result. 160

8.11 Time series of the normalized root mean squared error (nRMSE) for the AnDA assimilation of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (top, noted as AT_5) and for the AnDA assimilation of the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (middle, noted as $AT_5 + SWOT_0$). The time series of the nRMSE gain obtained when complementing nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days with daily SWOT data (given by $nRMSE_{AT_5} - nRMSE_{AT_5+SWOT_0}$) is also depicted (bottom, noted as Gain). The red filling indicates periods during which the nRMSE gain is positive, i.e. when the additional use of daily SWOT data improves interpolation performance ($nRMSE_{AT_5+SWOT_0} \leq nRMSE_{AT_5}$). 161

- 8.12 AnDA SLA reconstruction results for the 9th March, 2013. (8.12a) Real high-resolution ground-truth SLA field. (8.12b) Optimal Interpolation SLA reconstruction obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of 10 days and a spatial correlation scale of 100 km. (8.12c) Nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as AT_5). (8.12d) AnDA SLA reconstruction from nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as AT_5 and depicted in Figure (8.12c)). (8.12e) Daily SWOT observations (noted as $SWOT_0$). (8.12f) AnDA SLA reconstruction from daily SWOT observations (noted as $SWOT_0$ and depicted in Figure (8.12e)). (8.12g) Fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (noted as $AT_5 + SWOT_0$). (8.12h) AnDA SLA reconstruction from the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (noted as $AT_5 + SWOT_0$ and depicted in Figure (8.12g)). RMSE and correlation values given for each reported result. Adapted from [170]. © 2019 IEEE. 163

- 8.13 AnDA SLA gradient reconstruction results for the 9th March, 2013. (8.13a) Real high-resolution ground-truth SLA gradient field. (8.13b) Optimal Interpolation SLA gradient reconstruction obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of 10 days and a spatial correlation scale of 100 km. (8.13c) AnDA SLA gradient reconstruction from nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (noted as AT_5 and depicted in Figure (8.12c)). (8.13d) AnDA SLA gradient reconstruction from daily SWOT observations (noted as $SWOT_0$ and depicted in Figure (8.12e)). (8.13e) AnDA SLA gradient reconstruction from the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (noted as $AT_5 + SWOT_0$ and depicted in Figure (8.12g)). RMSE and correlation values given for each reported result. Adapted from [170]. © 2019 IEEE. 164

8.14 Radially-averaged power spectra for AnDA reconstruction results on the 9th March, 2013. Reported power spectra correspond to the high-resolution ground-truth SLA field (GT), the Optimal Interpolation SLA reconstruction obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of 10 days and a spatial correlation scale of 100 km (OI), the AnDA SLA reconstruction from nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (AT_5), the AnDA SLA reconstruction from daily SWOT observations ($SWOT_0$) and the AnDA SLA reconstruction from the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations ($AT_5 + SWOT_0$). Adapted from [170]. © 2019 IEEE. 165

8.15 Time series of the observation coverage (given as the ratio between the number of observed pixels and the total number of pixels in the considered region, excluding land pixels) for nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days (top, noted as AT_5) and for the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations (middle, noted as $AT_5 + SWOT_0$). The time series of the coverage gain obtained when complementing nadir along-track data accumulated on a time window $t_0 \pm D$ with $D = 5$ days with daily SWOT data is also depicted (bottom, noted as Gain). The red filling indicates periods during which the nRMSE gain is positive, i.e. when the additional use of daily SWOT data improves interpolation performance ($nRMSE_{AT_5+SWOT_0} \leq nRMSE_{AT_5}$). 166

8.16 (Top) Time series of the observation coverage (given as the ratio between the number of observed pixels and the total number of pixels in the considered region, excluding land pixels) for the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations. (Bottom) Time series of the nRMSE gain obtained when complementing nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days with daily SWOT data (given by $nRMSE_{AT_5} - nRMSE_{AT_5+SWOT_0}$). The red filling indicates periods during which the nRMSE gain is positive, i.e. when the additional use of daily SWOT data improves interpolation performance ($nRMSE_{AT_5+SWOT_0} \leq nRMSE_{AT_5}$). 168

-
- 8.17 Radially-averaged power spectra for AnDA reconstruction results when considering additional regressors on the 9th March, 2013. *Reported power spectra correspond to the high-resolution ground-truth SLA field (GT), the Optimal Interpolation SLA reconstruction obtained from nadir along-track observations with a standard covariance parametrization considering a temporal correlation scale of 10 days and a spatial correlation scale of 100 km (OI), and the AnDA SLA reconstruction from the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations when no additional regressors are considered ($AT_5 + SWOT_0$), when OI-resolved low-resolution SLA fields are considered as additional regressors ($AT_5 + SWOT_0 + OI$ reg), and when numerically-resolved high-resolution SLA gradient fields are considered as additional regressors ($AT_5 + SWOT_0 + \nabla$ SLA reg).* 169
- 8.18 AnDA SLA and SLA gradient (∇ SLA) reconstruction results when considering additional regressors for the 9th March, 2013. *Reported results depict the real high-resolution ground-truth SLA field (8.18a) and SLA gradient field (8.18b), as well as the AnDA SLA and SLA gradient reconstruction from the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations and considering: ((8.18c) and (8.18d)) no additional regressors, ((8.18e) and (8.18f)) OI-resolved low-resolution SLA fields as additional regressors, and ((8.18g) and (8.18h)) numerically-resolved high-resolution SLA gradient fields as additional regressors. RMSE and correlation values given for each reported result.* 171

List of Tables

2.1	Synthesis of constrained blind source separation algorithms. <i>Comparison criteria include enforced constraints, convergence properties, flexibility and code availability.</i>	48
3.1	Computational cost of the different steps of the proposed algorithms, in number of operations.	63
4.1	Error statistics (after convergence) for the different algorithms considered, computed over 100 runs of the algorithms with randomly generated data.	68
4.2	Estimation features for the single observation version of the ALS-direct algorithm (Algorithm 1) applied to a single-cluster ground-truth synthetic dataset without parameter sharing between observations. Reported results include mixing coefficients α_{nk} normalized mean squared estimation error ($e_{\alpha_{nk}}$), modal matrices β_k normalized mean squared estimation error (e_{β_k}), final log-likelihood ($\mathcal{L}(\alpha_{nk}, \beta_k)$) for the latent class initialization using model (3.4), and final relative mean squared error (rMSE, presented as a percentage) for the reconstruction of variables $\{\mathbf{y}_n\}$. <i>Results presented for both the EM and pseudo-EM initialization variants of the considered algorithm. Reported initialization variants are based on model (3.4), a latent class simplification of the proposed decomposition model (3.3). Taken from [166]. © 2016 IEEE.</i>	81
5.1	Relative mean square error (rMSE, given as a percentage) for the different models considered. [<i>Reported results correspond to a simple global linear regression; the two-class latent regression model introduced in [244] and our proposed non-negative decomposition model considering $K = 1$ and $K = 2$ classes. Taken from [166]. © 2016 IEEE.</i>	99
5.2	Normalized root mean squared error (nRMSE) values for global SSH and SSH gradient predictions for a two-mode model with and without external SQG forcing, and for a single-mode model without SQG forcing. <i>nRMSE values where calculated by normalizing the spatio-temporal global prediction RMSE for SSH and SSH gradient fields by the spatio-temporal standard deviation of the real SSH and SSH gradient fields.</i>	100

7.1	Normalized root mean square reconstruction error (nRMSE) for daily high-resolution SLA images $\{Y(t)\}_t$, for a global convolutional model and for locally-adapted decompositions of a global convolutional model using principal component analysis (PCA) [206], KSVD [3] and non-negative decomposition (NN), considering $K = 2$, $K = 5$ and $K = 10$ classes. The nRMSE value for daily low-resolution SLA images $\{Y_{LR}(t)\}_t$ is given as reference (noted as SLA_{LR}). Best results for each number of classes K considered are presented in bold. Results that outperform a global convolutional model are underlined.	135
8.1	Root mean squared error (Correlation) for SLA and SLA gradient (∇ SLA) reconstruction from nadir along-track observations for the different algorithms considered, namely OI [57], NN-LLOD [169], MS-VE-DINEOF [161, 213] and AnDA [161, 162]. Best result in bold	148
8.2	Root mean squared error (Correlation) for AnDA SLA and SLA gradient (∇ SLA) reconstruction from nadir along-track observations accumulated on a window $t_0 \pm D$ with $D = 5$ days under different noise settings, considering a centered additive white Gaussian noise with variance $\sigma_n^2 = \{0, 1, 3, 5\} \text{ cm}^2$. Best result in bold . Results for OI [57] and NN-LLOD [169] given as reference.	152
8.3	Root mean squared error (Correlation) for AnDA SLA and SLA gradient (∇ SLA) reconstruction from daily SWOT observations under different noise settings, namely ideal SWOT observations without noise sources, SWOT observations when considering only correlated noise sources (KaRIn noise), and SWOT observations when both correlated and uncorrelated noise sources are considered. Best result in bold	154
8.4	Root mean squared error (Correlation) for AnDA SLA and SLA gradient (∇ SLA) reconstruction from nadir along-track observations (AT_D) and wide-swath SWOT observations ($SWOT_D$). For each type of observations, both daily observations ($D = 0$) and observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days are considered. Best result in bold	154
8.5	Root mean squared error (Correlation) for AnDA SLA and SLA gradient (∇ SLA) reconstruction from the fusion of nadir along-track observations (AT_D) and SWOT observations ($SWOT_D$). For each type of observations, both daily observations ($D = 0$) and observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days are considered. Best result in bold	158
8.6	Root mean squared error (Correlation) for AnDA SLA and SLA gradient (∇ SLA) reconstruction from the fusion of nadir along-track observations accumulated on a time window $t_0 \pm D$ with $D = 5$ days and daily SWOT observations ($AT_5 + SWOT_0$), and considering different additional regressors, namely no additional regressors, OI-resolved low-resolution SLA fields, and numerically-resolved high-resolution SLA gradient fields. Best result in bold	169

- 8.7 Root mean squared error (Correlation) for AnDA SLA and SLA gradient (∇ SLA) reconstruction when assimilating daily SWOT observations and/or their corresponding numerically-resolved SWOT observation gradients. Best result in **bold**. 172

Bibliography

- [1] K. Abed-Meraim, P. Loubaton, and E. Moulines. A subspace algorithm for certain blind identification problems. *Information Theory, IEEE Transactions on*, 43(2):499–511, Mar 1997.
- [2] E. Agustsson, Timofte R., and L. Van Gool. Regressor Basis Learning for Anchored Super-Resolution. In *IEEE International Conference on Pattern Recognition (ICPR)*, Cancun, Mexico, December 2016.
- [3] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, Nov 2006.
- [4] M. M. Ali, D. Swain, and R. A. Weller. Estimation of ocean subsurface thermal structure from surface parameters: A neural network approach. *Geophysical Research Letters*, 31(20):n/a–n/a, 2004. L20308.
- [5] O. Aliagnat, J.M. Boucher, D.C. He, W. Pieczynski, and Universitc De Sherbrooke. Hidden markov fields and unsupervised segmentation of images. In *IAPR 11th International Conference on Pattern Recognition (ICPR 92)*, 1992.
- [6] J. T. Allen, D. A. Smeed, J. Tintoré, and S. Ruiz. Mesoscale subduction at the Almeria-Oran front: Part 1: Ageostrophic flow. *Journal of Marine Systems*, pages 263–285, 2001.
- [7] T. Amemiya. *Advanced econometrics*. Harvard university press, 1985.
- [8] J. M. Beckers and M. Rixen. EOF Calculations and Data Filling from Incomplete Oceanographic Datasets. *Journal of Atmospheric and Oceanic Technology*, 20(12):1839–1856, December 2003.
- [9] E. Berger, M. Sastuba, D. Vogt, B. Jung, and H. B. Amor. Estimation of perturbations in robotic behavior using dynamic mode decomposition. *Advanced Robotics*, 29(5):331–343, 2015.
- [10] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics & Data Analysis*, 52(1):155 – 173, 2007.

-
- [11] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics & Data Analysis*, 52(1):155 – 173, 2007.
- [12] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding.
- [13] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi Morel. K-WEB: Nonnegative dictionary learning for sparse image representations. In *IEEE International Conference on Image Processing (ICIP)*, Melbourne, Australia, September 2013.
- [14] J. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Technical Report TR-97-021, ICSI, 1997.
- [15] A. Birnbaum, I. M. Johnstone, B. Nadler, and D. Paul. Minimax bounds for sparse PCA with noisy high-dimensional data. *Annals of statistics*, 41(3):1055, 2013.
- [16] A. L. Boulesteix and K. Strimmer. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, 8(1):32–44, 2007.
- [17] G. C. Bourantas, M. Ghommem, G. C. Kagadis, K. Katsanos, V. C. Loukopoulos, V. N. Burganos, and G. C. Nikiforidis. Real-time tumor ablation simulation based on the dynamic mode decomposition method. *Medical Physics*, 41(5):053301, 2014.
- [18] T. Bouwmans and E. H. Zahzah. Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 122:22–34, 2014.
- [19] C. S. Bretherton, C. Smith, and J. M. Wallace. An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate*, 5(6):541–560, 1992.
- [20] F. P. Bretherton, R. E. Davis, and C. B. Fandry. A technique for objective analysis and design of oceanographic experiments applied to mode-73. In *Deep Sea Research and Oceanographic Abstracts*, volume 23, pages 559–582. Elsevier, 1976.
- [21] K. H. Brink and R. D. Muench. Circulation in the point conception-santa barbara channel region. *Journal of Geophysical Research: Oceans*, 91(C1):877–895, 1986.
- [22] D. S. Broomhead and G. P. King. Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, 20(2):217 – 236, 1986.
- [23] D. S. Broomhead and G. P. King. Nonlinear phenomena and chaos, edited by s. *Sarkar Hilger, Bristol*, page 113, 1986.

- [24] B. W. Brunton, L. A. Johnson, J. G. Ojemann, and J. N. Kutz. Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of Neuroscience Methods*, 258:1 – 15, 2016.
- [25] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *IEEE Conf. on Computer Vision and Pattern Recognition, CVPR'05*, volume 2, pages 60–65 vol. 2, June 2005.
- [26] Budisić, M. and Mohr, R. and Mezić, I. Applied koopmanism. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4):047510, 2012.
- [27] G. Burgers, P.r Jan van Leeuwen, and G. Evensen. Analysis scheme in the ensemble kalman filter. *Monthly Weather Review*, 126(6):1719–1724, 1998.
- [28] A. G. Burr. The multipath problem: an overview. In *Multipath Countermeasures, IEE Colloquium on*, pages 1/1–1/7, May 1996.
- [29] C. Sidney Burrus. Iterative reweighted least squares, December 2012. OpenStax CNX.
- [30] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.
- [31] X. Capet, J. C. McWilliams, M. J. Molemaker, and A. F. Shchepetkin. Mesoscale to submesoscale transition in the california current system. part i: Flow structure, eddy flux, and observational tests. *Journal of Physical Oceanography*, 38(1):29–43, 2008.
- [32] M. J. Carrier, H. E. Ngodock, S. R. Smith, I. Souopgui, and B. Bartels. Examining the Potential Impact of SWOT Observations in an Ocean Analysis–Forecasting System. *Monthly Weather Review*, 144(10):3767–3782, 2016.
- [33] K. S. Casey and D. Adamec. Sea surface temperature and sea surface height variability in the north pacific ocean from 1993 to 1999. *Journal of Geophysical Research: Oceans*, 107(C8):14–1–14–12, 2002.
- [34] D. B. Chelton and M. G. Schlax. The accuracies of smoothed sea surface height fields constructed from tandem satellite altimeter datasets. *Journal of Atmospheric and Oceanic Technology*, 20(9):1276–1302, 2003.
- [35] K. K. Chen, J. H. Tu, and C. W. Rowley. Variants of dynamic mode decomposition: Boundary condition, koopman, and fourier analyses. *Journal of Nonlinear Science*, 22(6):887–915, December 2012.
- [36] P. C. Chu, C. Fan, and W. T. Liu. Determination of vertical thermal structure from sea surface temperature. *Journal of Atmospheric and Oceanic Technology*, 17(7):971–979, 2016/01/21 2000.

-
- [37] P. L. Combettes and J. C. Pesquet. Proximal Splitting Methods in Signal Processing. In H.H. Bauschke, R.S. Burachik, P.L. Combettes, V. Elser, D.R. Luke, and H. (Eds.) Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011.
- [38] P. L. Combettes and V. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.
- [39] P. Comon. Independent component analysis, A new concept? *Signal Processing*, 36(3):287 – 314, 1994. Higher Order Statistics.
- [40] P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.
- [41] D. Dai, R. Timofte, and L. Van Gool. Jointly optimized regressors for image super-resolution. In *Eurographics*, 2015.
- [42] A. d’Aspremont, L. E. Ghaoui, M. I. Jordan, and G. R. Lanckriet. A direct formulation for sparse PCA using semidefinite programming. In *Advances in neural information processing systems*, pages 41–48, 2005.
- [43] G. M. Davis, S. G. Mallat, and Z. Zhang. Adaptive time-frequency decompositions. *Optical Engineering*, 33:33 – 33 – 9, 1994.
- [44] S. T. M. Dawson, M. S. Hemati, M. O. Williams, and C. W. Rowley. Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, 57(3):42, 2016.
- [45] F. De la Torre and M. J. Black. A framework for robust subspace learning. *International Journal of Computer Vision*, 54(1):117–142, Aug 2003.
- [46] F. Dellaert. The expectation maximization algorithm. Technical report, Georgia Institute of Technology, 2002.
- [47] W.S. DeSarbo and W.L. Cron. A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, 5(2):249–282, 1988.
- [48] C. Dicle, H. Mansour, D. Tian, Mo. Benosman, and A. Vetro. Robust low rank dynamic mode decomposition for compressed domain crowd and traffic flow analysis. In *Multimedia and Expo (ICME), 2016 IEEE International Conference on*, pages 1–6. IEEE, 2016.
- [49] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, Feb 2016.

- [50] N. Ducet and P.-Y. Le Traon. A comparison of surface eddy kinetic energy and Reynolds stresses in the Gulf Stream and the Kuroshio Current systems from merged TOPEX/Poseidon and ERS-1/2 altimetric data. *Journal of Geophysical Research: Oceans*, 106(C8):16603–16622, 2001.
- [51] N. Ducet, P.-Y. Le Traon, and G. Reverdin. Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. *Journal of Geophysical Research: Oceans*, 105(C8):19477–19498, 2000.
- [52] D. Duke, J. Soria, and D. Honnery. An error analysis of the dynamic mode decomposition. *Experiments in Fluids*, 52(2):529–542, 2012.
- [53] M. Duran-Moro, J.-M. Brankart, P. Brasseur, and J. Verron. Exploring image data assimilation in the prospect of high-resolution satellite oceanic observations. *Ocean Dynamics*, 67(7):875–895, Jul 2017.
- [54] M. Durand, L. Fu, D. P. Lettenmaier, D. E. Alsdorf, E. Rodriguez, and D. Esteban-Fernandez. The Surface Water and Ocean Topography Mission: Observing Terrestrial Surface Water and Oceanic Submesoscale Eddies. *Proceedings of the IEEE*, 98(5):766–779, May 2010.
- [55] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, Sep 1936.
- [56] E. Eiding and A. Yeredor. Blind mimo identification using the second characteristic function. *Signal Processing, IEEE Transactions on*, 53(11):4067–4079, Nov 2005.
- [57] R. Escudier, J. Bouffard, A. Pascual, P.-M. Poulain, and M.-I. Pujol. Improvement of coastal and mesoscale observation from space: Application to the northwestern Mediterranean Sea. *Geophys. Res. Lett.*, 40(10):2148–2153, 2013.
- [58] M. Espig, W. Hackbusch, and A. Khachatryan. On the Convergence of Alternating Least Squares Optimisation in Tensor Format Representations. *ArXiv e-prints*, May 2015.
- [59] D. Esteban-Fernandez. SWOT project mission performance and error budget document, JPL Doc. JPL D-79084. Technical report, Jet Propulsion Laboratory (JPL), National Aeronautics and Space Administration (NASA), 2014.
- [60] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.
- [61] G. Evensen. *Data Assimilation*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

- [62] R. Fablet, M. Lopez-Radcenco, J. Verron, B. Moure, B. Chapron, and A. Pascual. Learning multi-tracer convolutional models for the reconstruction of high-resolution SSH fields. In *IGARSS 2017: 2017 IEEE International Geoscience and Remote Sensing Symposium*, Fort Worth, Texas, USA, July 2017.
- [63] R. Fablet and F. Rousseau. Missing data super-resolution using non-local and statistical priors. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 676–680, September 2015.
- [64] R. Fablet and F. Rousseau. Joint interpolation of multisensor sea surface temperature fields using nonlocal and statistical priors. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(6):2665–2675, June 2016.
- [65] R. Fablet, J. Verron, B. Moure, B. Chapron, and A. Pascual. Improving mesoscale altimetric resolution/data in the Mediterranean Sea: multi-tracer convolutional retreatment of standard products. working paper or preprint, September 2016.
- [66] R. Fablet, J. Verron, B. Moure, B. Chapron, and A. Pascual. Improving mesoscale altimetric data from a multi-tracer convolutional processing of standard satellite-derived products. *IEEE Transactions on Geoscience and Remote Sensing*, 2017.
- [67] R. Fablet, P. H. Viet, and R. Lguensat. Data-driven Models for the Spatio-Temporal Interpolation of satellite-derived SST Fields. *IEEE Trans. Comput. Imag.*, 2017.
- [68] J. H. Faghmous, I. Frenger, Y. Yao, R. Warmka, A. Lindell, and V. Kumar. A daily global mesoscale ocean eddy dataset from satellite altimetry. *Scientific Data*, 2:150028 EP –, Jun 2015. Data Descriptor.
- [69] C. Fevotte, N. Bertin, and J. L. Durrieu. Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis. *Neural Computation*, 21(3):793–830, March 2009.
- [70] M. Foster. An application of the wiener-kolmogorov smoothing theory to matrix inversion. *Journal of the Society for Industrial and Applied Mathematics*, 9(3):387–392, 1961.
- [71] K. Fraedrich. Estimating the dimensions of weather and climate attractors. *Journal of the Atmospheric Sciences*, 43(5):419–432, 1986.
- [72] L.-L. Fu and A. Cazenave. *Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications*. International Geophysics. Elsevier Science, 2000.
- [73] L.-L. Fu, D. B. Chelton, P.-Y. Le Traon, and R. Morrow. Eddy dynamics from satellite altimetry. *Oceanography*, 23(4):14–25, 2010.
- [74] L.-L. Fu and R. Ferrari. Observing oceanic submesoscale processes from space. *Eos, Trans. AGU*, 89(48):488–488, 2008.

- [75] L.-L. Fu, D. Stammer, B.B. Leben, and D.B. Chelton. Improved spatial resolution of ocean surface topography from the TOPEX/Poseidon - Jason-1 tandem altimeter mission. *EOS, Transaction, American Geophysical Union*, 84:247–248, 2003. cited By 2.
- [76] L.-L. Fu and C. Ubelmann. On the transition from profile altimeter to swath altimeter for observing global ocean surface topography. *Journal of Atmospheric and Oceanic Technology*, 31(2):560–568, 2014.
- [77] A. Fukuoka. The central meteorological observatory, a study on 10-day forecast (a synthetic report). *Geophysical Magazine*, 22(3):177–208, 1951.
- [78] L. Gaultier and C. Ubelmann. SWOT Simulator Documentation. Technical report, JPL, NASA, 2010.
- [79] L. Gaultier, C. Ubelmann, and L.-L. Fu. The Challenge of Using Future SWOT Data for Oceanic Field Reconstruction. *Journal of Atmospheric and Oceanic Technology*, 33(1):119–126, November 2015.
- [80] M. E. Gheche, J. F. Aujol, Y. Berthoumieu, C. A. Deledalle, and R. Fablet. Texture synthesis guided by a low-resolution image. In *2016 IEEE 12th Image, Video, and Multi-dimensional Signal Processing Workshop (IVMSP)*, pages 1–5, July 2016.
- [81] P. V. Giampouras, A. A. Rontogiannis, and K. D. Koutroumbas. Low-rank and sparse NMF for joint endmembers’ number estimation and blind unmixing of hyperspectral images. *CoRR*, abs/1703.05785, 2017.
- [82] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *2009 IEEE 12th International Conference on Computer Vision*, pages 349–356, Sept 2009.
- [83] G. H. Golub and C. Reinsch. Singular value decomposition and least squares solutions. *Numer. Math.*, 14(5):403–420, April 1970.
- [84] L. Gomez-Navarro, R. Fablet, E. Mason, A. Pascual, B. Mourre, E. Cosme, and J. Le Sommer. SWOT Spatial Scales in the Western Mediterranean Sea Derived from Pseudo-Observations and an Ad Hoc Filtering. *Remote Sensing*, 10(4), 2018.
- [85] C. Gonzalez-Haro and J. Isern-Fontanet. Global ocean current reconstruction from altimetric and microwave SST measurements. *Journal of Geophysical Research: Oceans*, 119(6):3378–3391, 2014.
- [86] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 107–113. IET, 1993.

-
- [87] H. Arbabi and I. Mezić. Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the koopman operator. *SIAM J. Applied Dynamical Systems*, 16:2096–2126, 2017.
- [88] M. Hallin, D. Paindaveine, and T. Verdebout. Efficient r-estimation of principal and common principal components. *Journal of the American Statistical Association*, 109(507):1071–1083, 2014.
- [89] F. Hamilton, T. Berry, and T. Sauer. Ensemble Kalman Filtering without a Model. *Physical Review X*, 6(1):011021, March 2016.
- [90] A. Hannachi, I. T. Jolliffe, and D. B. Stephenson. Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology*, 27(9):1119–1152, 2007.
- [91] A. Hannachi, I. T. Jolliffe, D. B. Stephenson, and N. Trendafilov. In search of simple structures in climate: simplifying EOFs. *International Journal of Climatology*, 26(1):7–28, 2006.
- [92] A. Hannachi and A. O’Neill. Atmospheric multiple equilibria and non-gaussian behaviour in model simulations. *Quarterly Journal of the Royal Meteorological Society*, 127(573):939–958, 2001.
- [93] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12):2639–2664, 2004.
- [94] D. M. Hardy and J. J. Walton. Principal components analysis of vector wind measurements. *Journal of Applied Meteorology*, 17(8):1153–1162, 1978.
- [95] H. H. Harman. *Modern factor analysis*. University of Chicago press, 1976.
- [96] T. Hastie, R. Mazumder, J. Lee, and R. Zadeh. Matrix completion and low-rank svd via fast alternating least squares. *ArXiv e-prints*, October 2014.
- [97] U. Hausmann and A. Czaja. The observed signature of mesoscale eddies in sea surface temperature and the associated heat transport. *Deep Sea Research Part I: Oceanographic Research Papers*, 70:60 – 72, 2012.
- [98] P. Héas and C. Herzet. Optimal low-rank dynamic mode decomposition. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pages 4456–4460. IEEE, 2017.
- [99] I. M. Held, R. T. Pierrehumbert, S. T. Garner, and K. L. Swanson. Surface quasi-geostrophic dynamics. *Journal of Fluid Mechanics*, 282:1–20, 1 1995.

- [100] M. S. Hemati, C. W. Rowley, E. A. Deem, and L. N. Cattafesta. De-biasing the dynamic mode decomposition for applied koopman spectral analysis of noisy datasets. *Theoretical and Computational Fluid Dynamics*, 31(4):349–368, 8 2017.
- [101] G. E. Hinton and S. T. Roweis. Stochastic neighbor embedding. In *Advances in neural information processing systems*, pages 857–864, 2003.
- [102] A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- [103] AE Horel. Applications of ridge analysis to regression problems. *Chem. Eng. Progress.*, 58:54–59, 1962.
- [104] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933.
- [105] H. Hotelling. The most predictable criterion. *Journal of Educational Psychology*, 26(2):139–142, 1935.
- [106] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936.
- [107] P. O. Hoyer. Non-negative matrix factorization with sparseness constraints. *J. Mach. Learn. Res.*, 5:1457–1469, December 2004.
- [108] P. J. Huber. Robust statistics. new york, chichester, brisbane, 1981.
- [109] P. J. Huber. *Robust statistical procedures*, volume 68. Siam, 1996.
- [110] P. J Huber and E. M. Ronchetti. Robust statistics. NY: Wiley., 10(1002):9780470434697, 2009.
- [111] A. Hyvärinen. Survey on independent component analysis. In *Neural Computing Surveys*. Citeseer, 1999.
- [112] A. Hyvärinen, J. Karhunen, and E. Oja. Independent Component Analysis. Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control, 2001.
- [113] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.
- [114] J. Isern-Fontanet, B. Chapron, G. Lapeyre, and P. Klein. Potential use of microwave sea surface temperatures for the estimation of ocean currents. *Geophysical Research Letters*, 33(24):n/a–n/a, 2006. L24608.
- [115] J. Isern-Fontanet, G. Lapeyre, P. Klein, B. Chapron, and M. W. Hecht. Three-dimensional reconstruction of oceanic mesoscale currents from surface information. *Journal of Geophysical Research: Oceans*, 113(C9):n/a–n/a, 2008. C09005.

-
- [116] J. Isern-Fontanet, M. Shinde, and C. González-Haro. On the transfer function between surface fields and the geostrophic stream function in the Mediterranean Sea. *Journal of Physical Oceanography*, 44:1406–1423, 2014.
- [117] R. I. Jennrich. A simple general procedure for orthogonal rotation. *Psychometrika*, 66(2):289–306, Jun 2001.
- [118] R. I. Jennrich. A simple general method for oblique rotation. *Psychometrika*, 67(1):7–19, Mar 2002.
- [119] I. M. Johnstone and A. Y. Lu. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 104(486):682–693, 2009.
- [120] I. T. Jolliffe and J. Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 374(2065), 2016.
- [121] I. T. Jolliffe, N. T. Trendafilov, and M. Uddin. A Modified Principal Component Technique Based on the LASSO. *Journal of Computational and Graphical Statistics*, 12(3):531–547, 2003.
- [122] I. T. Jolliffe, M. Uddin, and S. K. Vines. Simplified EOFs - three alternatives to rotation. *Climate Research*, 20(3):271–279, April 2002.
- [123] Jovanović, M. R. and Schmid, P. J. and Nichols, J. W. . Sparsity-promoting dynamic mode decomposition. *Physics of Fluids*, 26(2):024103, 2014.
- [124] M. Juza, B. Mourre, L. Renault, S. Gómara, K. Sebastián, S. Lora, J. P. Beltran, B. Frontera, B. Garau, C. Troupin, M. Torner, E. Heslop, B. Casas, R. Escudier, G. Vizoso, and J. Tintoré. SOCIB operational ocean forecasting system and multi-platform validation in the Western Mediterranean Sea. *Journal of Operational Oceanography*, 9(sup1):s155–s166, February 2016.
- [125] M. Juza, B. Mourre, L. Renault, S. Gómara, K. Sebastián, S. Lora, J. P. Beltran, B. Frontera, B. Garau, C. Troupin, M. Torner, E. Heslop, B. Casas, R. Escudier, G. Vizoso, and J. Tintoré. Socib operational ocean forecasting system and multi-platform validation in the western mediterranean sea. *Journal of Operational Oceanography*, 9(sup1):s155–s166, 2016.
- [126] H. F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200, Sep 1958.
- [127] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, Mar 1960.

- [128] T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, and A.Y. Wu. An efficient k-means clustering algorithm: analysis and implementation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):881–892, Jul 2002.
- [129] Y. Kawahara. Dynamic mode decomposition with reproducing kernels for koopman spectral analysis. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 911–919. Curran Associates, Inc., 2016.
- [130] H. A. L. Kiers. Simplimax: Oblique rotation to an optimal target with simple structure. *Psychometrika*, 59(4):567–579, Dec 1994.
- [131] P. Klein, B.L. Hua, G. Lapeyre, et al. Upper ocean turbulence from high-resolution 3d simulations, 2008.
- [132] P. Klein, J. Isern-Fontanet, G. Lapeyre, G. Roulet, E. Danioux, B. Chapron, S. Le Gentil, and H. Sasaki. Diagnosis of vertical velocities in the upper ocean from high resolution sea surface height. *Geophysical Research Letters*, 36(12):n/a–n/a, 2009. L12603.
- [133] C. J. Koblinsky, P. Gaspar, and G. Lagerloef. The future of spaceborne altimetry oceans and climate change: A long-term strategy. Technical report, National Aeronautics and Space Administration (NASA), United States, 1992. N-92-26121.
- [134] B. O. Koopman. Hamiltonian systems and transformations in hilbert space. *Proceedings of the National Academy of Sciences of the United States of America*, 17(5):315–318, 1931.
- [135] B. Kramer, P. Grover, P. Boufounos, S. Nabi, and M. Benosman. Sparse sensing and DMD-based identification of flow regimes and bifurcations in complex flows. *SIAM Journal on Applied Dynamical Systems*, 16(2):1164–1196, 2017.
- [136] D. Kuang, S. Yun, and H. Park. Symnmf: Nonnegative low-rank approximation of a similarity matrix for graph clustering. *Journal of Global Optimization*, 62(3):545–574, 2015.
- [137] P. K. Kundu and J. S. Allen. Some three-dimensional characteristics of low-frequency current fluctuations near the oregon coast. *Journal of Physical Oceanography*, 6(2):181–199, 1976.
- [138] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor. Chapter 3: Koopman analysis. In *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*, chapter 3, pages 39–53. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2016.

-
- [139] J. N. Kutz, X. Fu, S. L. Brunton, and N. B. Erichson. Multi-resolution dynamic mode decomposition for foreground/background separation and object tracking. In *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 921–929, Dec 2015.
- [140] J. E. Kutzbach. Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over north america. *Journal of Applied Meteorology*, 6(5):791–802, 1967.
- [141] L. S. Gandin. Objective analysis of meteorological fields. *Quarterly Journal of the Royal Meteorological Society. Translated from the Russian. Jerusalem (Israel Program for Scientific Translations)*, 92(393):447–447, 1965.
- [142] J. H. LaCasce and A. Mahadevan. Estimating subsurface horizontal and vertical velocities from sea-surface temperature. *Journal of Marine Research*, 64(5):695–721, 2006-09-01T00:00:00.
- [143] G. Lapeyre and P. Klein. Dynamics of the upper oceanic layers in terms of surface quasi-geostrophy theory, 2006.
- [144] G. Lapeyre and P. Klein. Impact of the small-scale elongated filaments on the oceanic vertical pump. *Journal of Marine Research*, 64(6):835–851, 11 2006.
- [145] C. L. Lawson and R. J. Hanson. *Solving least squares problems*, volume 15, chapter 23, page 161. SIAM, 1995.
- [146] F. Le Gland, V. Monbet, and V. D. Tran. Large sample asymptotics for the ensemble Kalman filter. Research Report RR-7014, INRIA, 2009.
- [147] C. Le Goff, P. Tandeo, B. Chapron, and R. Fablet. Spatio-temporal decomposition of satellite-derived SST-SSH fields: links between surface data and ocean interior dynamics in the aghulas region. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, Submitted.
- [148] P.-Y. Le Traon and G. Dibarboure. Velocity mapping capabilities of present and future altimeter missions: The role of high-frequency signals. *Journal of Atmospheric and Oceanic Technology*, 19(12):2077–2088, 2002. cited By 39.
- [149] P.-Y. Le Traon and G. Dibarboure. An illustration of the contribution of the TOPEX/Poseidon - Jason-1 tandem mission to mesoscale variability studies. *Marine Geodesy*, 27(1-2):3–13, 2004. cited By 39.
- [150] P.-Y. Le Traon, G. Dibarboure, and N. Ducet. Use of a high-resolution model to analyze the mapping capabilities of multiple-altimeter missions. *Journal of Atmospheric and Oceanic Technology*, 18(7):1277–1288, 2001.

- [151] P.-Y. Le Traon, Y. Faugère, F. Hernandez, J. Dorandeu, F. Mertz, and M. Ablain. Can We Merge GEOSAT Follow-On with TOPEX/Poseidon and ERS-2 for an Improved Description of the Ocean Circulation? *Journal of Atmospheric and Oceanic Technology*, 20(6):889–895, 2003.
- [152] P.-Y. Le Traon, F. Nadal, and N. Ducet. An improved mapping method of multisatellite altimeter data. *Journal of Atmospheric and Oceanic Technology*, 15(2):522–534, 1998.
- [153] P.-Y. Le Traon, F. Nadal, and N. Ducet. An improved mapping method of multisatellite altimeter data. *Journal of Atmospheric and Oceanic Technology*, 15(2):522–534, 1998.
- [154] D. Lee, W. Lee, Y. Lee, and Y. Pawitan. Super-sparse principal component analyses for high-throughput genomic data. *BMC bioinformatics*, 11(1):296, 2010.
- [155] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, October 1999.
- [156] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Proceedings of the 13th International Conference on Neural Information Processing Systems, NIPS'00*, pages 535–541, Cambridge, MA, USA, 2000. MIT Press.
- [157] S. Y. Lee. Blind source separation and independent component analysis: A review. *Neural Information Processing-Letters and Reviews*, 6(1):1–57, 2005.
- [158] T.-W. Lee. Independent component analysis. In *Independent component analysis*, pages 27–66. Springer, 1998.
- [159] J. Lei and V. Q. Vu. Sparsistency and agnostic inference in sparse PCA. *The Annals of Statistics*, 43(1):299–322, 2015.
- [160] E. W. Leuliette and J. M. Wahr. Coupled pattern analysis of sea surface temperature and TOPEX/Poseidon sea surface height. *Journal of Physical Oceanography*, 29(4):599–611, 2016/01/22 1999.
- [161] R. Lguensat, P. Tandeo, P. Aillot, and R. Fablet. The Analog Data Assimilation. *Mon. Weather Rev.*, 2017.
- [162] R. Lguensat, P.H. Viet, M. Sun, G. Chen, T. Fenglin, B. Chapron, and R. Fablet. Data-driven Interpolation of Sea Level Anomalies using Analog Data Assimilation. Technical report, Lab-STICC, Centre Nationale de la Recherche Scientifique (CNRS), IMT Atlantique, October 2017.
- [163] Q. Li, F. Dietrich, E. M. Bollt, and I. G. Kevrekidis. Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the koopman operator. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(10):103111, 2017.

-
- [164] M. Loève. *Probability theory: foundations, random sequences*. van Nostrand Princeton, NJ, 1955.
- [165] M. Lopez-Radcenco, A. Aissa-El-Bey, P. Ailliot, and R. Fablet. Non-negative decomposition of geophysical dynamics. In *ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, April 2017.
- [166] M. Lopez-Radcenco, A. Aissa-El-Bey, P. Ailliot, R. Fablet, and P. Tandeo. Non-negative decomposition of linear relationships: application to multi-source ocean remote sensing data. In *ICASSP 2016 : 41st IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4179–4183, 2016.
- [167] M. Lopez-Radcenco, A. Aissa-El-Bey, P. Tandeo, and R. Fablet. Décomposition Non-négative de Dynamiques Géophysiques. In *GRETSI 2017: XXVIème colloque du GRETSI*, Juan-Les-Pins, France, September 2017.
- [168] M. Lopez-Radcenco, R. Fablet, and A. Aissa-El-Bey. Non-negative observation-based decomposition of operators. *IEEE Transactions on Signal Processing*, Submitted.
- [169] M. Lopez-Radcenco, R. Fablet, A. Aissa-El-Bey, and P. Ailliot. Locally-adapted convolution-based super-resolution of irregularly-sampled ocean remote sensing data. In *ICIP 2017 : IEEE International Conference on Image Processing*, Beijing, China, September 2017.
- [170] M. Lopez-Radcenco, A. Pascual, L. Gomez-Navarro, A. Aissa-El-Bey, B. Chapron, and R. Fablet. Analog Data Assimilation of Along-track Nadir and Wide-swath SWOT Altimetry Observations in the Western Mediterranean Sea. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Forthcoming.
- [171] M. Lopez-Radcenco, A. Pascual, L. Gomez-Navarro, A. Aissa-El-Bey, and R. Fablet. Analog data assimilation for along-track nadir and SWOT altimetry data in the Western Mediterranean Sea. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Valencia, Spain, July 2018.
- [172] M. Lopez-Radcenco, A. Pascual, L. Gomez-Navarro, A. Aissa-El-Bey, and R. Fablet. Assimilation par Analogues de Données Altimétriques Nadir et SWOT dans la Mer Méditerranée Occidentale. In *Conférence Française de Photogrammétrie et de Télédétection (CFPT)*, Marne-la-Vallée, France, June 2018.
- [173] A. C. Lorenc, S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, and F. W. Saunders. The met. office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126(570):2991–3012, 2000.

- [174] E. N. Lorenz. *Empirical Orthogonal Functions and Statistical Weather Prediction*. Scientific report. Statistical Forecasting Project (Massachusetts Institute of Technology). Massachusetts Institute of Technology, Department of Meteorology, 1956.
- [175] E. N. Lorenz. Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric sciences*, 26(4):636–646, 1969.
- [176] E. N. Lorenz. Predictability: a problem partly solved. In *Seminar on Predictability, 4-8 September 1995*, volume 1, pages 1–18, Shinfield Park, Reading, 1995. ECMWF, ECMWF.
- [177] L. Lorenzi, F. Melgani, and G. Mercier. Inpainting strategies for reconstruction of missing data in VHR images. *IEEE Geoscience and Remote Sensing Letters*, 8(5):914–918, 2011.
- [178] J. L. Lumley. The structure of inhomogeneous turbulent flows. *Atmospheric Turbulence and Radio Wave Propagation*, 1967.
- [179] J. L. Lumley. Coherent structures in turbulence. In R. E. Meyer, editor, *Transition and Turbulence*, pages 215–242, 1981.
- [180] J. B. Macqueen. Some Methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Math, Statistics, and Probability*, volume 1, pages 281–297. University of California Press, 1967.
- [181] P. Malanotte-Rizzoli, V. Artale, G. L. Borzelli-Eusebi, S. Brenner, A. Crise, M. Gacic, N. Kress, S. Marullo, M. Ribera d’Alcalà, S. Sofianos, T. Tanhua, A. Theocharis, M. Alvarez, Y. Ashkenazy, A. Bergamasco, V. Cardin, S. Carniel, G. Civitarese, F. D’Ortenzio, J. Font, E. Garcia-Ladona, J. M. Garcia-Lafuente, A. Gogou, M. Gregoire, D. Hainbucher, H. Kontoyannis, V. Kovacevic, E. Kraskapoulou, G. Kroskos, A. Incarbona, M. G. Mazzocchi, M. Orlic, E. Ozsoy, A. Pascual, P.-M. Poulain, W. Roether, A. Rubino, K. Schroeder, J. Siokou-Frangou, E. Souvermezoglou, M. Sprovieri, J. Tintoré, and G. Triantafyllou. Physical forcing and physical/biochemical variability of the Mediterranean Sea: a review of unresolved issues and directions for future research. *Ocean Sci.*, 10(3):281–322, May 2014.
- [182] S. G. Mallat. *A wavelet tour of signal processing, second edition*. Academic Press, 1999.
- [183] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, Dec 1993.
- [184] J. Mann and J. N. Kutz. Dynamic mode decomposition for financial trading strategies. *Quantitative Finance*, 16(11):1643–1655, 2016.
- [185] G. Maze, H. Mercier, R. Fablet, P. Lenca, M. Lopez-Radcenco, P. Tandeo, C. Le Goff, and Ch. Feucher. Ocean heat content structure revealed by un-supervised classification of hydrographic profiles. In *OSD 2016: Ocean Sciences Meeting*, New Orleans, United States, February 2016.

-
- [186] G. Maze, H. Mercier, R. Fablet, P. Tandeo, M. Lopez-Radencio, P. Lenca, Ch. Feucher, and C. Le Goff. Coherent heat patterns revealed by unsupervised classification of Argo temperature profiles in the North Atlantic Ocean. *Progress in Oceanography*, 151:275 – 292, 2017.
- [187] I. Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1):309–325, August 2005.
- [188] I. Mezić. Analysis of fluid flows via spectral properties of the koopman operator. *Annual Review of Fluid Mechanics*, 45(1):357–378, 2013.
- [189] A. H. Monahan, J. C. Fyfe, M. H. P. Ambaum, D. B. Stephenson, and G. R. North. Empirical orthogonal functions: The medium is the message. *Journal of Climate*, 22(24):6501–6514, 2009.
- [190] T. K. Moon and W. C. Stirling. *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, 2000.
- [191] R. Morrow and P.-Y. Le Traon. Recent advances in observing mesoscale ocean dynamics with satellite altimetry. *Advances In Space Research*, 50(8):1062–1076, Oct 2012.
- [192] S. Na, L. Xumin, and G. Yong. Research on k-means clustering algorithm: An improved k-means clustering algorithm. In *Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on*, pages 63–67, April 2010.
- [193] R. S. Nerem, D. P. Chambers, E. W. Leuliette, G. T. Mitchum, and B. S. Giese. Variations in global mean sea level associated with the 1997–1998 ENSO event: Implications for measuring long term sea level change. *Geophysical Research Letters*, 26(19):3005–3008, 1999.
- [194] R. S. Nerem, E. Leuliette, and A. Cazenave. Present-day sea-level change: A review. *Comptes Rendus Geoscience*, 338(14):1077 – 1083, 2006. La Terre observée depuis l’espace.
- [195] A. Newson, A. Almansa, M. Fradet, Y. Gousseau, and P. Pérez. Video inpainting of Complex Scenes. *SIAM Journal on Imaging Sciences*, 7(4):1993–2019, January 2014.
- [196] A. Y. Ng. Feature selection, L1 vs. L2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning*, page 78. ACM, 2004.
- [197] A. M. Obukhov. Statistically homogeneous fields on a sphere. *Usp. Mat. Nauk*, 2(2):196–198, 1947.
- [198] A. M. Obukhov. The statistically orthogonal expansion of empirical functions. *Bulletin of the Academy of Sciences of the USSR. Geophysics Series (English Transl.)*, 1:288–291, 1960.

- [199] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37(23):3311 – 3325, 1997.
- [200] A. Ozerov and C. Févotte. Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. *Audio, Speech, and Language Processing, IEEE Transactions on*, 18(3):550–563, March 2010.
- [201] M. Pal, R. Roy, J. Basu, and M.S. Bepari. Blind source separation: A review and analysis. In *Oriental COCODA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCODA/CASLRE), 2013 International Conference*, pages 1–5, Nov 2013.
- [202] C. Pan, D. Xue, and J. Wang. On the accuracy of dynamic mode decomposition in estimating instability of wave packet. *Experiments in Fluids*, 56(8):164, Aug 2015.
- [203] A. Pascual, J. Bouffard, S. Ruiz, B. Buongiorno Nardelli, E. Vidal-Vijande, R. Escudier, J. M. Sayol, and A. Orfila. Recent improvements in mesoscale characterization of the western Mediterranean Sea: synergy between satellite altimetry and other observational approaches. *Scientia Marina*, 77(1):19–36, March 2013.
- [204] A. Pascual, M.-I. Pujol, G. Larnicol, P.-Y. Le Traon, and M.-H. Rio. Mesoscale mapping capabilities of multisatellite altimeter missions: First results with real data in the Mediterranean Sea. *J. Mar. Syst.*, 65(1–4):190–211, March 2007.
- [205] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pages 40–44 vol.1, Nov 1993.
- [206] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572, 1901.
- [207] G. Peyré, S. Bogleux, and L.D. Cohen. Non-local Regularization of Inverse Problems. *Inverse Probl. and Imag.*, 5(2):511–530, 2011.
- [208] D. T. Pham. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, 129(5):1194–1207, 2001.
- [209] D. L. Phillips. A technique for the numerical solution of certain integral equations of the first kind. *J. ACM*, 9(1):84–97, January 1962.
- [210] W. Pieczynski. Parameter estimation in the case of hidden data. In *Proceedings 16th Biennial Symposium on Communications, Kingston, Canada, 1992*.
- [211] W. Pieczynski. Champs de markov cachés et estimation conditionnelle itérative, 1994.

- [212] W. Pieczynski. Convergence of the iterative conditional estimation and application to mixture proportion identification. In *Statistical Signal Processing, 2007. SSP '07. IEEE/SP 14th Workshop on*, pages 49–53, Aug 2007.
- [213] B. Ping, F. Su, and Y. Meng. An Improved DINEOF Algorithm for Filling Missing Values in Spatio-Temporal Sea Surface Temperature Data. *PLOS ONE*, 11(5):e0155928, May 2016.
- [214] G. Plaut and R. Vautard. Spells of low-frequency oscillations and weather regimes in the northern hemisphere. *Journal of the Atmospheric Sciences*, 51(2):210–236, 1994.
- [215] R. W. Preisendorfer. Principal component analysis in meteorology and oceanography,(ed) curtis. D., *Mobley*, page 419, 1988.
- [216] J. L. Proctor and P. A. Eckhoff. Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *International Health*, 7(2):139–145, 2015.
- [217] M.-I. Pujol, G. Dibarboure, P.-Y. Le Traon, and P. Klein. Using high-resolution altimetry to observe mesoscale signals. *Journal Of Atmospheric And Oceanic Technology*, 29(9):1409–1416, 2012.
- [218] M.-I. Pujol and G. Larnicol. Mediterranean sea eddy kinetic energy variability from 11 years of altimetric data. *Journal of Marine Systems*, 58(3–4):121–142, December 2005.
- [219] R. K. Raney. The delay/doppler radar altimeter. *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1578–1588, Sept 1998.
- [220] W.R. Reed and H. Ye. Which panel data estimator should i use? *Applied Economics*, 43(8):985–1000, 2011.
- [221] M. B. Richman. Rotation of principal components. *Journal of Climatology*, 6(3):293–335, 1986.
- [222] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
- [223] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and DAN S. Henningson. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641:115–127, 2009.
- [224] G. A. Ruggiero, E. Cosme, J.M. Brankart, J. Le Sommer, and C. Ubelmann. An Efficient Way to Account for Observation Error Correlations in the Assimilation of Data from the Future SWOT High-Resolution Altimeter Mission. *J. Atmos. Oceanic Tech.*, 33(12):2755–2768, 2016.
- [225] S. Ruiz, A. Pascual, B. Garau, I. Pujol, and J. Tintoré. Vertical motion in the upper ocean from glider and altimetry data. *Geophysical Research Letters*, 36(14), 2009.

- [226] F. H. Ruymgaart. A robust principal component analysis. *Journal of Multivariate Analysis*, 11(4):485 – 497, 1981.
- [227] M. Saraceno, C. Provost, and A. R. Piola. On the relationship between satellite-retrieved surface temperature fronts and chlorophyll a in the western south atlantic. *Journal of Geophysical Research: Oceans*, 110(C11):n/a–n/a, 2005. C11016.
- [228] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.
- [229] P. J. Schmid, L. Li, M. P. Juniper, and O. Pust. Applications of the dynamic mode decomposition. *Theoretical and Computational Fluid Dynamics*, 25(1):249–259, Jun 2011.
- [230] B. Schölkopf, A. Smola, E. Smola, and K.R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299–1319, 1998.
- [231] W. C. Siu and K. W. Hung. Review of image interpolation and super-resolution. In *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–10, Dec 2012.
- [232] P. Smyth. Finite mixture models and the EM algorithm. University Lecture, 2015.
- [233] K. Stadlthanner, F. J. Theis, C. G. Puntonet, and E. W. Lang. Extended sparse nonnegative matrix factorization. In J. Cabestany, A. Prieto, and F. Sandoval, editors, *Computational Intelligence and Bioinspired Systems*, volume 3512 of *Lecture Notes in Computer Science*, pages 249–256. Springer Berlin Heidelberg, 2005.
- [234] G. W. Stewart. On the early history of the singular value decomposition. *SIAM Review*, 35(4):551–566, 1993.
- [235] H. von Storch and F. W. Zwiers. *Statistical Analysis in Climate Research*. Cambridge University Press, 1999.
- [236] L. Sun, S. Ji, S. Yu, and J. Ye. On the equivalence between canonical correlation analysis and orthonormalized partial least squares. In *Twenty-First International Joint Conference on Artificial Intelligence*, 2009.
- [237] Y. Susuki and I. Mezić. Nonlinear koopman modes and power system stability assessment without models. *IEEE Transactions on Power Systems*, 29(2):899–907, March 2014.
- [238] G. Takács and D. Tikk. Alternating least squares for personalized ranking. In *Proceedings of the Sixth ACM Conference on Recommender Systems*, RecSys ’12, pages 83–90, New York, NY, USA, 2012. ACM.
- [239] N. Takeishi, Y. Kawahara, Y. Tabei, and T. Yairi. Bayesian dynamic mode decomposition. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 2814–2821, 2017.

-
- [240] N. Takeishi, Y. Kawahara, and T. Yairi. Learning Koopman Invariant Subspaces for Dynamic Mode Decomposition. *ArXiv e-prints*, October 2017.
- [241] N. Takeishi, Y. Kawahara, and T. Yairi. Sparse Nonnegative Dynamic Mode Decomposition. In *IEEE International Conference on Image Processing (ICIP)*, Beijing, China, September 2017.
- [242] N. Takeishi, Y. Kawahara, and T. Yairi. Subspace dynamic mode decomposition for stochastic koopman analysis. *Phys. Rev. E*, 96:033310, Sep 2017.
- [243] P. Tandeo, P. Ailliot, J. Ruiz, A. Hannart, B. Chapron, A. Cuzol, V. Monbet, R. Easton, and R. Fablet. Combining Analog Method and Ensemble Data Assimilation: Application to the Lorenz-63 Chaotic System. In V. Lakshmanan, E. Gilleland, A. McGovern, and M. Tingley, editors, *Machine Learning and Data Mining Approaches to Climate Science*, pages 3–12. Springer, 2015.
- [244] P. Tandeo, B. Chapron, S. Ba, E. Autret, and R. Fablet. Segmentation of mesoscale ocean surface dynamics using satellite SST and SSH observations. *Geoscience and Remote Sensing, IEEE Transactions on*, 52(7):4227–4235, July 2014.
- [245] P. Tandeo, M. Pulido, and F. Lott. Offline parameter estimation using EnKF and maximum likelihood error covariance estimates: Application to a subgrid-scale orography parametrization. *Quarterly Journal of the Royal Meteorological Society*, 141(687):383–395, 2015.
- [246] R. Tibshirani. Regression Shrinkage and Selection Via the LASSO. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.
- [247] A. N. Tikhonov. On the stability of inverse problems. In *Dokl. Akad. Nauk SSSR*, volume 39, pages 195–198, 1943.
- [248] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Math.*, 4:1035–1038, 1963.
- [249] A. N. Tikhonov and V.Y. Arsenin. Solution of ill-posed problems. *Winston & Sons, Washington*, 1977.
- [250] A. N. Tikhonov, A. Goncharsky, V. V. Stepanov, and A. G. Yagola. *Numerical methods for the solution of ill-posed problems*, volume 328. Springer Science & Business Media, 2013.
- [251] A. N. Tikhonov, A. S. Leonov, and A. G. Yagola. *Nonlinear ill-posed problems*. London: Chapman & Hall, 1998, 1998.

- [252] R. Timofte, V. De Smet, and L. Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2013.
- [253] R. Timofte, V. De Smet, and L. Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian Conference on Computer Vision*, pages 111–126. Springer, 2014.
- [254] C. Tofallis. Model building with multiple dependent variables and constraints. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 48(3):371–378, 1999.
- [255] M. Tonani, A. Teruzzi, G. Korres, N. Pinardi, A. Crise, M. Adani, P. Oddo, S. Dobricic, C. Fratianni, M. Drudi, et al. The Mediterranean Monitoring and Forecasting Centre, a component of the MyOcean system. In *Proceedings of the Sixth International Conference on EuroGOOS*, pages 4–6, 2011.
- [256] Jo. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics*, 1:391, 2014.
- [257] C. Ubelmann, P. Klein, and L.-L. Fu. Dynamic Interpolation of Sea Surface Height and Potential Applications for Future High-Resolution Altimetry Mapping. *Journal of Atmospheric and Oceanic Technology*, 32(1):177–184, October 2014.
- [258] D. S. Ullman and P. C. Cornillon. Evaluation of Front Detection Methods for Satellite-Derived SST Data Using In Situ Observations. *Journal of Atmospheric and Oceanic Technology*, 17(12):1667–1675, 2000.
- [259] J.-P. Vert and M. Kanehisa. Graph-driven feature extraction from microarray data using diffusion kernels and kernel CCA. In *Advances in neural information processing systems*, pages 1449–1405, 2003.
- [260] C. R Vogel. Computational methods for inverse problems, 2002.
- [261] W. Härdle and L. Simar. *Canonical Correlation Analysis*, pages 321–330. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.
- [262] Q. Wang. Kernel principal component analysis and its applications in face recognition and active shape models. *CoRR*, abs/1207.3538, 2012.
- [263] B. C. Weare and J. S. Nasstrom. Examples of extended empirical orthogonal function analyses. *Monthly Weather Review*, 110(6):481–485, 1982.
- [264] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley. A data-driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, Dec 2015.

-
- [265] D. M. Witten, R. Tibshirani, and T. Hastie. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 10(3):515–534, 2009.
- [266] J. M. Wooldridge. *Econometric analysis of cross section and panel data*. MIT Press, Cambridge and London, 2002.
- [267] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in neural information processing systems*, pages 2080–2088, 2009.
- [268] X. Wu, X. H. Yan, Y. H. Jo, and W. T. Liu. Estimation of subsurface temperature anomaly in the north atlantic using a self-organizing map neural network. *Journal of Atmospheric and Oceanic Technology*, 29(11):1675–1688, 2016/01/21 2012.
- [269] A. Wynn, D. S. Pearson, B. Ganapathisubramani, and P. J. Goulart. Optimal mode decomposition for unsteady flows. *Journal of Fluid Mechanics*, 733:473–503, 2013.
- [270] Y. Xu and W. Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences*, 6(3):1758–1789, 2013.
- [271] Y. Susuki and I. Mezić. A prony approximation of koopman mode decomposition. *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 7022–7027, 2015.
- [272] D. Yang, Z. Li, Y. Xia, and Z. Chen. Remote sensing image super-resolution: Challenges and approaches. In *2015 IEEE International Conference on Digital Signal Processing (DSP)*, pages 196–200, July 2015.
- [273] J. Yang, J. Wright, T.S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, Nov 2010.
- [274] D. Zachariah, M. Sundin, M. Jansson, and S. Chatterjee. Alternating least-squares for low-rank matrix reconstruction. *IEEE Signal Processing Letters*, 19:231–234, April 2012.
- [275] R. Zadeh, H. Li, B. He, M. Lublin, and Y. Perez. Cme 323: Distributed algorithms and optimization, spring 2015. University Lecture, 2015.
- [276] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang. Robust principal component analysis with complex noise. In *International conference on machine learning*, pages 55–63, 2014.
- [277] G. Zhou, A. Cichocki, and S. Xie. Fast nonnegative matrix/tensor factorization based on low-rank approximation. *IEEE Transactions on Signal Processing*, 60(6):2928–2940, June 2012.

Titre : Approches pilotées par les données pour la télédétection océanique : De la décomposition non négative d'opérateurs à la reconstruction des dynamiques de la surface de l'océan à partir de données satellitaires

Mots clés : Télédétection satellitaire, Masses de données, Décomposition d'opérateurs, Non-négativité, Dynamiques de l'océan, Interpolation

Résumé : Au cours des dernières années, la disponibilité toujours croissante de données de télédétection multi-source de l'océan a été un facteur clé pour améliorer notre compréhension des dynamiques de la surface de l'océan. A cet égard, il est essentiel de mettre au point des approches efficaces pour exploiter ces ensembles de données. En particulier, la décomposition des processus géophysiques en modes pertinents est une question clé pour les problèmes de caractérisation, de prédiction et de reconstruction. Inspirés par des progrès récents en séparation aveugle des sources, nous visons, dans la première partie de cette thèse, à étendre les modèles de séparation aveugle de sources sous contraintes de non-négativité au problème de la caractérisation et décomposition d'opérateurs ou fonctions de transfert entre variables d'intérêt. Nous développons des schémas computationnels efficaces reposant sur des fondations mathématiques solides.

Nous illustrons la pertinence des modèles de décomposition proposés dans différentes applications impliquant l'analyse et la prédiction de dynamiques géophysiques. Par la suite, étant donné que la disponibilité toujours croissante d'ensembles de données multi-sources supporte l'exploration des approches pilotées par les données en tant qu'alternative aux formulations classiques basées sur des modèles, nous explorons des approches basées sur les données récemment introduits pour l'interpolation des champs géophysiques à partir d'observations satellitaires irrégulièrement échantillonnées. De plus, en vue de la future mission SWOT, la première mission satellitaire à produire des observations d'altimétrie par satellite complètement bidimensionnelles et à large fauchée, nous nous intéressons à évaluer dans quelle mesure les données SWOT permettraient une meilleure reconstruction des champs altimétriques.

Title: Data-driven Approaches for Ocean Remote Sensing: From the Non-negative Decomposition of Operators to the Reconstruction of Satellite-derived Sea Surface Dynamics

Keywords : Satellite remote sensing, Massive datasets, Operator decomposition, Non-negativity, Ocean dynamics, Interpolation

Abstract: In the last few decades, the ever-growing availability of multi-source ocean remote sensing data has been a key factor for improving our understanding of upper ocean dynamics. In this regard, developing efficient approaches to exploit these datasets is of major importance. Particularly, the decomposition of geophysical processes into relevant modes is a key issue for characterization, forecasting and reconstruction problems. Inspired by recent advances in blind source separation, we aim, in the first part of this thesis dissertation, at extending non-negative blind source separation models to the problem of the observation-based characterization and decomposition of linear operators or transfer functions between variables of interest. We develop mathematically sound and computationally efficient schemes.

We illustrate the relevance of the proposed decomposition models in different applications involving the analysis and forecasting of geophysical dynamics. Subsequently, given that the ever-increasing availability of multi-source datasets supports the exploration of data-driven alternatives to classical model-driven formulations, we explore recently introduced data-driven models for the interpolation of geophysical fields from irregularly sampled satellite-derived observations. Importantly, with a view towards the future SWOT mission, the first satellite mission to produce complete two-dimensional wide-swath satellite altimetry observations, we focus on assessing the extent to which SWOT data may lead to an improved reconstruction of altimetry fields.