



# THÈSE

**En vue de l'obtention du  
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**  
Délivré par l'Institut National Polytechnique de Toulouse

---

**Présentée et soutenue par  
Rem-Sophia MOURADI**

Le 16 mars 2021

**Modélisation non-linéaire de champs multidimensionnels guidée  
par la donnée : application aux écoulements côtiers hydro-  
morphodynamiques**

---

Ecole doctorale : **SDUZE - Sciences de l'Univers, de l'Environnement et de  
l'Espace**

Spécialité : **Océan, Atmosphère, Climat**

Unité de recherche :

**CECI - Climat, Environnement, Couplages et Incertitudes / CERFACS**

Thèse dirigée par  
**Olivier THUAL et Cédric GOEURY**

Jury

**M. Olivier Le Maître**, Rapporteur  
**M. Pierre-Olivier Malaterre**, Rapporteur  
**Mme Clémentine Prieur**, Rapporteur  
**M. Florent Lyard**, Examinateur  
**Mme Hélène Roux**, Examinatrice  
**Mme Christine Keribin**, Examinatrice  
**M. Olivier THUAL**, Directeur de thèse  
**M. Cédric Goeury**, Directeur de thèse

**M. Pablo Tassi**, invité  
**M. Fabrice Zaoui**, invité



---

**Non-linear data-driven modelling on multidimensional fields: an application to hydro-morphodynamic coastal flows**

---

Rem-Sophia MOURADI

January 13, 2021



# Contents

<b>Introduction</b>	<b>1</b>
<b>1 Nearshore processes and morphodynamics</b>	<b>7</b>
1.1 Physical context: dynamics at regional, local and intake scales . . . . .	9
1.2 Available measurements for data-based and data-driven investigations . . .	21
1.3 Process-based hydro-morphodynamics for data-driven modelling . . . . .	27
1.4 State of the art of the intake dynamics investigations . . . . .	45
1.5 Summary . . . . .	49
<b>2 Data-driven modelling</b>	<b>51</b>
2.1 Uncertainty Quantification . . . . .	53
2.2 Data-based techniques . . . . .	69
2.3 Physically-based data-driven uncertainty reduction: Data Assimilation . .	78
2.4 Dimensionality Reduction . . . . .	88
2.5 Summary . . . . .	98
<b>3 Physically interpretable machine learning algorithm on multidimensional non-linear fields</b>	<b>99</b>
3.1 Motivations . . . . .	100
3.2 Published paper in Journal of Computational Physics . . . . .	100
3.3 Complementary data investigations using POD and variants . . . . .	142
3.4 Summary . . . . .	150
<b>4 Sensitivity of tidal modelling in coastal configurations</b>	<b>151</b>
4.1 Motivations . . . . .	152
4.2 Deterministic investigations . . . . .	152
4.3 Telemac User Conference article . . . . .	157
4.4 Intake's Boundary Condition and related uncertainties . . . . .	167
4.5 Summary . . . . .	171
<b>5 Model reduction for fast and accurate Data Assimilation</b>	<b>173</b>
5.1 Motivations . . . . .	174
5.2 Preprint for submission . . . . .	176
5.3 Attempt of calibration on different scenarios . . . . .	203
5.4 Summary . . . . .	204
<b>6 Sediment dynamics with uncertain parameters, boundaries and initial conditions</b>	<b>205</b>
6.1 Objective . . . . .	206
6.2 Uncertainty study . . . . .	207
6.3 Summary . . . . .	215

---

<b>7</b>	<b>Conclusions and outlook</b>	<b>217</b>
7.1	Main achievements . . . . .	218
7.2	Perspectives and open questions . . . . .	220
<b>A</b>	<b>Liste des acronymes</b>	<b>I</b>

# List of Figures

1	Example of measured sedimentation in the intake of interest and comparison to siltation in Bray Harbor, Irish Sea (Muir Éireann, <i>Source: Afloat Magazine</i> .) . . . . .	3
2	Schematic drawing of the interest intake. . . . .	4
1.1	Wind exposition on the English Channel coast from 1981 to 2010, by Letortu [130] . . . . .	12
1.2	Wind rose for 2003-2007 National Centers for Environmental Prediction (NCEP) data at the vicinity of the power plant, by Latteux [122] . . . . .	13
1.3	Wave rose for 2003-2009 Numerical Atlas for Oceanic and Coastal Sea States (ANEMOC) data, 24 km off coast at the vicinity of the power plant, by Latteux [122] . . . . .	13
1.4	Sediment types in the English Channel, by Larssonneur et al. [120] . . . . .	15
1.5	Bottom types evolutions in the marine zone of the intake, by Costa et al. [46]	15
1.6	Cumulative distributions of the sediments granulometry realized for the 2010 campaign, by Latteux [122]. Points 1 to 6 are outside the intake, points 7 and 8 are inside. . . . .	16
1.7	Median sediment grains inside the intake on 1993, by Latteux [122]. <i>Sable</i> is french word for sand, and $D_{50}$ designates the median sand grain size. . . . .	17
1.8	Granulometry distributions inside the intake on 2008, by Latteux [122]. . . . .	17
1.9	2007 vs. 2018 mono-beam bathymetry measurements in the power plant's water intake. Values are positive when red, negative when blue and near-zero when white. . . . .	22
1.10	Measurements of Tidal Level (TL), Wind velocity (Wv), Wind direction (Wdir), Wave Height (WvH) and Wave Direction (WvD) on January 2016. (P: Percentile) . . . . .	23
1.11	Locations of measurement points for the two-months 2010 survey. . . . .	24
1.12	Superposition of 2010 measurements for the hydrodynamic variables on Point 1. . . . .	24
1.13	Plots of turbidity and wave heights measurements on Point 4, by [122] . . . . .	25
1.14	Measurement devices for the 2018 campaign. . . . .	26
1.15	Acoustic Doppler Current Profiler (ADCP) measurement transects in the intake represented in red color, and an example of velocity magnitude profile at a transect perpendicular to pumps, from left bank (pump) to right bank. . . . .	27
1.16	Examples of salinity and turbidity measurements with the SAMBAT buoy. . . . .	27
1.17	Variation of Manning's coefficient with bed forms, by Hassanzadeh [91] . . . . .	33
1.18	Bed forms illustrations by Guy et al. [86] . . . . .	39
1.19	Scheme of TELEMAC-MASCARET System (TMS) modules coupling [244]	45

1.20	Intake's daily deposition rate as a function of the average of $H_s^3 * T_p / z_s$ for each deposition period, colored by the sedimentation month, by Latteux [122] . . . . .	47
2.1	Comparison between Neural Networks (NN) and human brain. Images (a) vs. (b) for a single artificial vs. human neuron, and images (c) vs. (d) for NN vs. human brain, by Meng et al. [158] . . . . .	75
2.2	Some major Modal Decomposition (MD) techniques, by Taira et al. [241]. L (Linear), NL (Non-Linear), E (Experimental) and NS (Navier-Stokes) . . . . .	90
2.3	Graphical representation of Singular Value Decomposition (SVD) meaning, by Taira et al. [241]. . . . .	94
3.1	First four temporal coefficients of Proper Orthogonal Decomposition (POD) applied to the complete data set. Example of an outlier date circled in red in Figure (a). . . . .	143
3.2	First four spatial patterns of POD applied to the complete data set. Values are positive when red, negative when blue and near-zero when white. . . . .	143
3.3	The first four temporal coefficients of the POD applied to the filtered data set. . . . .	144
3.4	First four spatial patterns before and after filtering. Values are positive when red, negative when blue and near-zero when white. . . . .	144
3.5	Bathymetry variance maps over defined periods. . . . .	144
3.6	First four spatial patterns of the POD applied to the Time-Restrained 2016 – 2018 set (T-R). Values are positive when red, negative when blue and near-zero when white. . . . .	145
3.7	The first four temporal coefficients of POD applied to the Time-Restrained 2016–2018 set (T-R). Lines correspond to the 2016-2018 POD, while dashed lines represent the temporal coefficients obtained by projecting the 2007-2015 measurements on the 2016-2018 basis. . . . .	145
3.8	First four spatial patterns of the POD applied to the Space-Restrained set (S-R). Values are positive when red, negative when blue and near-zero when white. . . . .	146
3.9	Comparison of the 2007 – 2018 time averaged RMSE using the full, S-R and T-R sets for POD. Approximation rank corresponds to the used POD modes number. . . . .	146
3.10	Comparison of Explained Variance Rate (EVR) resulting from POD on the full, S-R and T-R sets. Approximation rank corresponds to the used POD modes number. . . . .	147
3.11	Comparison of time averaged relative RMSE between full field and field reconstruction from reduced vectors, provided by Principal Component Analysis (PCA), and Kernel PCA (KPCA) with cosine and polynomial kernels, using Scikit-learn [189]. . . . .	148
3.12	Spatial patterns comparison between PCA and Sparse PCA (SPCA) applied to the intake bathymetry, using Scikit-learn [189]. Values are positive when red, negative when blue and near-zero when white. . . . .	148
3.13	Pearson correlations between temporal coefficients of the whole intake bathymetry POD, and POD applied to each single-beam profile independently. . . . .	149
4.1	Example of meshed calculation domain for the intake with adequate Boundary Conditions (BC). . . . .	152



4.2	Velocity magnitude at intake's entrance cross-shore profile with different meshes, at HT. . . . .	153
4.3	Velocity magnitude at intake's entrance cross-shore profile using different meshes, at Low Tide (LT). . . . .	154
4.4	Velocity magnitude at intake's entrance cross-shore profile using different meshes, at Half Falling Tide (HFT). . . . .	154
4.5	Velocity magnitude at intake's entrance cross-shore profile using different meshes, at Half Rising Tide (HRT). . . . .	154
4.6	Velocity magnitude at intake's entrance cross-shore profile using different domain extensions, for simulations without Friction, at High Tide (HT). . .	155
4.7	Velocity magnitude at intake's entrance cross-shore profile using different domain extensions, for simulations without Friction, at Low Tide (LT). . .	156
4.8	Velocity magnitude at intake's entrance cross-shore profile using different domain extensions, for simulations without Friction, at Half Falling Tide (HFT). . . . .	156
4.9	Velocity magnitude at intake's entrance cross-shore profile using different domain extensions, for simulations without Friction, at Half Rising Tide (HRT). . . . .	156
4.10	Velocity magnitude at intake's entrance cross-shore profile using different domain extensions, for simulations with Friction, at High Tide (HT). . . .	157
4.11	Velocity magnitude at intake's entrance cross-shore profile using different domain extensions, for simulations with Friction, at Low Tide (LT). . . . .	157
4.12	Examples of Uncertainty Quantification (UQ) realizations (in colors) for the X-velocity $u$ profile at intake's entrance, at Half Rising Tide (HRT), for the 8 km domain. . . . .	167
4.13	Comparison of POD patterns of intake's entrance velocity profiles, with the different domain sizes at Half Rising Tide (HRT). . . . .	168
4.14	Comparison of Sobol' indices for the two first modes of velocity components, with the different domain sizes at Half Rising Tide (HRT). The full bar plot represents the total Sobol' indices. The dashed portion corresponds to the first Sobol' index, and the remaining to the interaction with other variables. . . . .	168
4.15	Comparison of POD patterns of intake's entrance x-velocity profiles, with two friction formulas, at Half Rising Tide (HRT). . . . .	169
4.16	Comparison of POD patterns of intake's entrance y-velocity profiles, with two friction formulas, at Half Rising Tide (HRT). . . . .	170
4.17	Comparison of Sobol' indices for the two first modes of the velocity components, with different friction formulas at characteristic time HRT. The full bar plot represents the total Sobol' indices. The dashed portion corresponds to the first Sobol' indice, and the remaining to the interaction with other variables. . . . .	170
5.1	Example of measurement reconstruction on a given tidal period (approx. 12 h 25 mn) for point 1 using 2 POD modes extracted from 38 occurrences. . . . .	174
5.2	Scatter plot of POD coefficients for Mode 1 of free surface, at points 1, 2 and 3 from the 2010 measurement campaign in front of the intake, and at nearest harbor (Hydrographic and Oceanographic Marine Service (SHOM)) . . . . .	175
5.3	EVR for the three hydrodynamic variables after measurement reduction on point 1. . . . .	176

---

5.4	Example of measurement reconstruction on a given tidal period for point 1 using 2 POD modes. . . . .	203
5.5	Example of measurement reconstruction on a given tidal period for point 1 using 2 POD modes. . . . .	204
6.1	Mesh computational domain for a preliminary UQ study of the intake's hydro-morphodynamics, with prescribed Boundary Conditions (BC). . . . .	207
6.2	Tidal levels characteristics Probability Density Functions (PDFs) inferred by Kernel smoothing from the 2007-2018 data. . . . .	208
6.3	POD expansion coefficients for the first four modes of velocity Boundary Conditions (BC) profile at intake's entrance. . . . .	209
6.4	Volume deposit and sediment flux PDFs inferred by Kernel smoothing from the 2007-2018 data. . . . .	210
6.5	POD expansion coefficients for the first four modes of velocity BC profile at intake's entrance. . . . .	211
6.6	Statistics of sedimentation in the intake obtained from preliminary UQ investigation. Quantities are expressed in meters. . . . .	212
6.7	Interest points inside the intake for UQ investigations. . . . .	213
6.8	First and total Sobol' indices of the studied uncertain parameters at the entrance of the intake. . . . .	213
6.9	First and total Sobol' indices of the studied uncertain parameters at the bending portion of the intake. . . . .	214
6.10	First and total Sobol' indices of the studied uncertain parameters downstream of the intake at point 5. . . . .	215

# List of Tables

1.1	Frequency (‰) of observed winds on 1968 by origin and force scale at Dieppe (France), by Loic [136]. . . . .	11
1.2	Summary of measurements, their frequencies, periods, sources and spatial coverage. . . . .	23
1.3	Measurements uncertainties. . . . .	25
1.4	Waves characteristic quantities . . . . .	38
2.1	Orthonormal Askey-scheme hypergeometric polynomials for parametric PDFs [269] . . . . .	63



# Introduction

*Au 21ème siècle, la recherche scientifique dans de nombreux domaines de la physique implique de traiter diverses sources de données. Les observations deviennent de plus en plus nombreuses et accessibles, grâce aux avancées technologiques de mesure, stockage de données, capacité de transmission et traitement. Par ailleurs, la puissante avancée historique des ressources computationnelles, à partir de la fin des années 90, a encouragé l'usage des modèles à base de processus physiques, permettant d'améliorer notre compréhension des phénomènes observés, et de générer de la donnée supplémentaire. C'est ainsi que les approches guidées par la donnée sont devenues une des pierres angulaires de la physique, allant de l'Assimilation de Données (AD) au Machine Learning (ML).*

*Cette thèse se concentre sur le ML interprétable, en utilisant les techniques de Réduction de Dimension (DR) et de régression probabiliste non-linéaire et multivariée, en combinant deux approches classiques : la Décomposition en modes Propres Orthogonaux (POD), et l'Expansion par Polynômes du Chaos (PCE). La méthodologie proposée est appliquée à différentes étapes de la modélisation guidée par la donnée : (i) l'apprentissage à base de données mesurées, (ii) la Quantification des Incertitudes (UQ) efficace et (iii) l'Assimilation de Données rapide et précise.*

*Les contributions présentées découlent d'investigations menées au sein de la communauté scientifique des géosciences, qui connaît par ailleurs une augmentation constante des sources de données. Les méthodologies proposées ont pour but de fournir un outil prédictif dans un contexte industriel, avec des défis sous-jacents, notamment concernant les contraintes liées au temps de calcul. Plus précisément, la modélisation de la morphodynamique dans un chenal bord-de-mer de centrale électrique est visée. Des données de surveillance du chenal, collectées durant plusieurs années dans le but d'optimiser sa gestion, ainsi qu'un modèle numérique hydro-morphodynamique, sont disponibles. L'objectif principal de cette thèse est donc d'établir une méthodologie de couplage optimal entre données de terrain et modélisation numérique, en utilisant des outils statistiques adaptés. La finalité est de prédire de manière rapide et précise l'élévation du lit sous-marin, aussi appelée bathymétrie. Cette méthodologie est appliquée dans une configuration côtière, avec pour objectif de mieux comprendre la morphodynamique (évolution des bathymétries). Cet aspect est crucial pour plusieurs applications, en particulier pour la prédiction de l'écoulement résultant, ce qui peut être d'intérêt socio-économique (par exemple pour prédire des inondations), ou d'intérêt industriel comme pour l'application proposée.*

In the 21st century, scientific research in many physical fields involves dealing with a variety of data sources. Observations are becoming plentiful and accessible, due to technological advances in measurement devices, data storage, transmission and treatment capacities. The powerful historical jump of computational resources since late 90's has encouraged the use of process-based models, allowing to improve our understanding of physical phenomena, and to generate additional data. This is how data-driven approaches have become one of the cornerstones of physics, from [Data Assimilation \(DA\)](#) to [Machine Learning \(ML\)](#). As an example, [ML](#) has gained interest from classical physics (fluid mechanics [32, 116, 145, 174]; aerodynamics [265, 275]; plasma physics [83, 188], astrophysics and astronomy [115, 258]) to quantum physics (particle physics [4]; quantum mechanics [164]). Perhaps, physicists nourish a hope about exploring the "chasm of ignorance" using data-based techniques, by pushing the boundaries of classical approaches [105]. Although time has not yet come for drastic change [105], and believing that data may come with added value to previously established theories, one may ask the following: what are the optimal combinations between all information sources? and how to help physical modelling advances using statistical tools? These are open research questions that we do not pretend to solve in one thesis, but represent a guideline for the presented discussions.

The thesis work focuses on interpretable [ML](#) using [Dimensionality Reduction \(DR\)](#) and non-linear multivariate probabilistic regression, by combining two classical approaches: [POD](#) and [Polynomial Chaos Expansion \(PCE\)](#). The proposed methodology is applied at various steps of the data-driven modelling: (i) pure measurement based learning; (ii) efficient [UQ](#) and (iii) fast and accurate [DA](#). In particular, the work takes place in the geosciences community, which also registers a constant increase in data sources [109]. Some interesting programs can be cited as the new SWOT satellite mission [166, 176], or the Sentinel satellite missions (Copernicus program) [66, 147]. Consequently, the community is also keen on data-based works, with increasingly represented contributions in [ML](#) [109, 205, 216], [DA](#) [39, 72], and [UQ](#) [19, 169]. These methods are of particular interest for example, in a context of climate change, where new data constantly need to be taken into account [204].

The proposed steps of this thesis aim at providing a predictive tool in an industrial context with inherent challenges, for example concerning computational time constraints. More precisely, the modelling of *morphodynamics* in a coastal power plant's water intake was targeted. Data were collected during many years of monitoring, in order to optimize the intake's management, and a *hydro-morphodynamic* numerical model is available. The main objective of this thesis is therefore to enable an optimal coupling methodology between field data and numerical modelling using appropriate statistical tools, for fast and accurate prediction of underwater topography (also called *bathymetry* or *bottom/bed* elevation). This methodology is applied in a coastal set-up, with the goal of better understanding sea bed temporal evolution, known as *morphodynamics*. This topic is crucial for many applications, in particular for the prediction of the resulting flow, which can be of socio-economical interest (e.g floods prediction) or industrial interest (e.g. current application).

## **Industrial context: power plants monitoring**

Water intakes are a crucial component of power plants, as they ensure their cooling process via a pumping system. The plants are therefore constructed near to natural water