



HAL
open science

Prédiction du futur de données fonctionnelles à partir de covariables

Selman Sezgin, Julien Jacques, Kahina Mokrani, Sylvain Allio

► To cite this version:

Selman Sezgin, Julien Jacques, Kahina Mokrani, Sylvain Allio. Prédiction du futur de données fonctionnelles à partir de covariables. 56èmes Journées de Statistique de la SFdS, Jun 2025, Marseille, France. <hal-05034942>

HAL Id: hal-05034942

<https://hal.science/hal-05034942v1>

Submitted on 15 Apr 2025

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.



HAL Authorization

PRÉDICTION DU FUTUR DE DONNÉES FONCTIONNELLES À PARTIR DE COVARIABLES

Selman Sezgin ¹, Julien Jacques ², Kahina Mokrani ³ & Sylvain Allio ⁴

¹ *Orange Innovation / Université de Lyon 2, Université Claude Bernard Lyon 1, ERIC, France, selman.sezgin@orange.com*

² *Université Lumière Lyon 2, Université Claude Bernard Lyon 1, ERIC, France, Julien.Jacques@univ-lyon2.fr*

³ *Orange Innovation, France, kahina.mokrani@orange.com*

⁴ *Orange Innovation, France, sylvain.allio@orange.com*

Résumé. Le présent travail aborde un problème de prédiction en analyse de données fonctionnelles. On considère la prédiction du futur d'une fonction Y à partir de l'observation de son passé ainsi que de celui de covariables fonctionnelles. Plus précisément, à partir d'une collection i.i.d. d'observations de Y et des covariables, nous développons un modèle qui, étant donné le début de Y et le début des covariables, prédit le futur de Y . Notre méthode repose sur la décomposition du futur de Y dans une base de fonctions, et sur la prédiction des coefficients d'expansion par une approche de régression scalaire-sur-fonction. Les performances de notre modèle sont évaluées sur des données fonctionnelles synthétiques. Les résultats expérimentaux montrent que notre modèle permet de construire des prédictions plus lisses et plus précises que les méthodes traditionnelles existants.

Mots-clés. Analyse de données fonctionnelles, prédiction de données longitudinales, décomposition en fonctions de base.

Abstract. Functional data analysis is a branch of statistics where data are treated as smooth functions rather than finite-dimensional vectors. The functional approach allows for a more accurate representation of dynamic and continuous phenomena and has numerous real-world applications, such as in healthcare and finance. In this work, we address the problem of predicting a function Y based on functional covariates. More specifically, given an i.i.d. collection of observations of Y and the covariates, we develop a model that, given the beginning of Y and the beginning of the covariates, predicts the future of Y . Our method relies on decomposing the future of Y into a basis of functions and predicting the expansion coefficients using a scalar-on-function regression approach. The performance of our model is evaluated on both synthetic and real functional data. Experimental results show that our model provides smoother and more accurate predictions than traditional existing methods.

Keywords. Functional data analysis, longitudinal data prediction, functional basis expansion.

1 Introduction

L'analyse de données fonctionnelles est une branche active des statistiques qui traitent les données sous la forme de courbes lisses plutôt que de vecteurs de dimension finie. Le traitement des données par analyse fonctionnelle a de nombreux domaines d'application comme dans la santé [14], la finance [7] ou la biologie [5]. Dans ce cadre, contrairement à une approche d'apprentissage statistique traditionnelle, une donnée ne correspond pas à la réalisation d'un vecteur aléatoire de dimension finie, mais à une fonction lisse à domaine continue appartenant à un espace de dimension infinie, ce qui permet de refléter plus fidèlement des phénomènes dynamiques continus. Le livre [10] est une référence clé pour une introduction à l'analyse de données fonctionnelles. Les domaines de recherche sont variés et incluent la régression [13], le lissage [2], la classification [12] et le clustering [9]. Les données fonctionnelles sont souvent représentées sous la forme de données dites *longitudinales*, c'est-à-dire où un (ou plusieurs) phénomène est observé sur un intervalle temporel pour plusieurs instances. Par exemple, une étude clinique pourrait suivre l'évolution dans le temps de paramètres physiologiques (poids, pression artérielle, température corporelle etc.) d'un ensemble d'individus. Dans le présent travail, nous étudions un problème de prédiction pour des données fonctionnelles longitudinales. Plus spécifiquement, nous considérons un ensemble de trajectoires d'une fonction d'intérêt, que l'on désignera par Y , et des trajectoires conjointes de covariables fonctionnelles. Notre but est de construire un modèle fonctionnel qui, à partir du début de la courbe de Y et des courbes des covariables, prédit la suite de la courbe de Y . Plus formellement, on considère une collection i.i.d. de données fonctionnelles

$$\left\{ Y_i(t), X_{1,i}(t), \dots, X_{p,i}(t); t \in [0, T + H] \right\}_{i=1}^n \quad (1)$$

où le i -ème échantillon contient un fonction $Y_i(t)$ et des covariables $X_{i,\ell}(t), \ell = 1, \dots, p$, définies sur l'intervalle $[0, T + H]$, avec T et H strictement positifs. On considère ensuite un nouvel échantillon $Y_{n+1}(t), X_{n+1,\ell}(t), \ell = 1, \dots, p$ où les fonctions ne sont observées que jusqu'au temps T et sont inconnues entre les temps T et $T + H$. Le but de ce travail est de faire une prédiction lisse de la suite de $Y_{n+1}(t)$ sur $[T, T + H]$ à partir de son observation passée sur $[0, T]$, de l'observation passée des covariables et du comportement général des fonctions sur les n échantillons complets initiaux (1). Le problème étudié est illustré dans la figure 1.

Nous proposons un modèle fonctionnel en trois étapes :

1. La fonction Y sur l'intervalle futur est décomposée dans une base de fonctions,
2. Un modèle fonctionnel paramétrique est optimisé pour estimer les coefficients d'expansion du futur de Y à partir du début de la courbe de Y et des covariables,
3. L'inférence pour la prédiction de Y sur le nouvel $(n + 1)$ -ème échantillon est faite à partir des paramètres optimisées et de l'observation passée des fonctions.

Les parties suivantes développent notre approche.

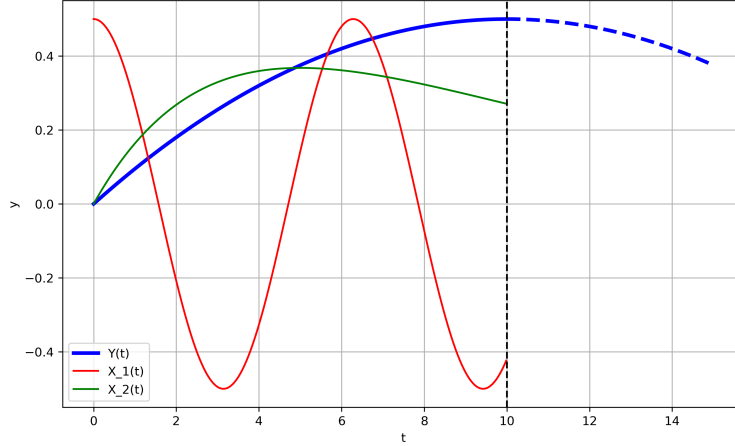


FIGURE 1 – Illustration du problème étudié. On cherche à prédire le futur de $Y(t)$ entre $T = 10$ et $T + H = 15$ à partir du passé de $Y(t)$ et du passé des covariables $X_1(t)$ et $X_2(t)$.

2 Modèle fonctionnel

Une hypothèse classique en analyse de données fonctionnelles est que les fonctions sous-jacentes aux observations appartiennent à l'espace de Hilbert des fonctions de carré intégrable, qui est un espace de dimension infinie. Pour réduire la dimensionalité du problème et faciliter le traitement statistique, il est courant de décomposer les données dans des bases de fonctions [10]. Des exemples classiques de bases sont les bases de Fourier pour les phénomènes périodiques [8], les ondelettes [3] et les splines qui sont largement utilisées en pratique pour représenter des fonctions non périodiques [1]. À travers la décomposition dans une base, l'information contenue dans une fonction à domaine continu est encodée par un nombre restreint de coefficients. Notre modèle repose sur l'idée clé que, étant donnée une base de fonctions, prédire le futur de Y revient à prédire ses coefficients dans cette base.

Notons $Y_i^{(1)}(t)$ et $Y_i^{(2)}(t)$ la restriction de $Y_i(t)$ sur l'intervalle $[0, T]$ et $[T, T + H]$ respectivement. De même, on introduit les notations $X_{i,\ell}^{(1)}(t)$ et $X_{i,\ell}^{(2)}(t)$ pour les covariables. Fixons une base de fonctions $\{\psi_k : k \in \llbracket 1, K \rrbracket\}$ définies sur $[T, T + H]$. Alors la décomposition de $Y_i^{(2)}(t)$ dans cette base s'écrit

$$Y_i^{(2)}(t) = \sum_{k=1}^K b_{i,k} \psi_k(t) + \varepsilon_i(t), \quad (2)$$

où $\varepsilon_i(t)$ est une fonction résiduelle définie comme la composante de $Y_i^{(2)}(t)$ qui est linéairement indépendante de la base $\{\psi_k\}$. En pratique, $\varepsilon_i(t)$ correspond au bruit des mesures ou à des comportements indésirables de la fonction. Par hypothèse du problème, les coefficients d'expansion dans la décomposition (2) sont calculables pour $i = 1, \dots, n$ mais inconnus pour $i = n + 1$. Nous proposons alors de prédire les coefficients $\{\hat{b}_{n+1,k} : k \in \llbracket 1, K \rrbracket\}$ puis de

reconstruire la suite de $Y_{n+1}^{(2)}(t)$ avec

$$\hat{Y}_{n+1}^{(2)}(t) = \sum_{k=1}^K \hat{b}_{n+1,k} \psi_k(t). \quad (3)$$

Nous traitons le calcul des $\{\hat{b}_{n+1,k}\}$ comme un ensemble de problèmes de régression scalaire-sur-fonction indépendants. Pour chaque $k = 1, \dots, K$, les échantillons

$$\left\{ (Y_i^{(1)}(t), X_{1,i}^{(1)}(t), \dots, X_{p,i}^{(1)}(t)), b_{k,i} \right\}_{i=1}^n \quad (4)$$

induisent un problème de régression scalaire-sur-fonction où le coefficient $b_{k,i}$ est estimé à partir des prédicteurs fonctionnels $Y_i^{(1)}(t), X_{\ell,i}^{(1)}(t), \ell = 1, \dots, p$. Le modèle fonctionnel linéaire standard pour résoudre ce problème est le suivant [11] :

$$\hat{b}_{k,i} = \int_{u=0}^T Y_i^{(1)}(u) \beta_{k,0}(u) du + \sum_{\ell=1}^p \int_{u=0}^T X_{i,\ell}^{(1)}(u) \beta_{k,\ell}(u) du. \quad (5)$$

Les fonctions $\beta_{k,\ell}(t), \ell = 0, \dots, p$ sont des paramètres définies sur $[T, T + H]$ et de carré intégrable. À l'aide des intégrales dans (5), ces paramètres permettent d'extraire l'information utile des fonctions en entrée pour la prédiction des coefficients d'expansion du futur de Y . En injectant (5) dans (2), notre modèle fonctionnel s'écrit :

$$\hat{Y}_i^{(2)}(t) = \sum_{k=1}^K \left(\int_{u=0}^T Y_i^{(1)}(u) \beta_{k,0}(u) du + \sum_{\ell=1}^p \int_{u=0}^T X_{i,\ell}^{(1)}(u) \beta_{k,\ell}(u) du \right) \psi_k(t). \quad (6)$$

En particulier, la courbe recherchée $\hat{Y}_{n+1}^{(2)}(t)$ peut être prédite par inférence à partir des paramètres $\{\beta_{k,\ell}\}$ estimés sur les n premiers échantillons et des courbes observées $Y_{n+1}^{(1)}(t)$ et $X_{\ell,n+1}^{(1)}(t), \ell = 1, \dots, p$.

3 Optimisation

En adoptant le principe classique de minimisation du risque empirique [4], l'optimisation du modèle (6) repose sur le calcul des paramètres fonctionnels $\{\beta_{k,\ell}\}$ de manière à minimiser l'erreur de prédiction sur une collection i.i.d. de données. Jusqu'à présent, nous avons considérés nos données fonctionnelles comme des courbes observées de manière continue dans le temps. Mais en pratique les fonctions sont observées sur un ensemble discret de points. Supposons que Y ainsi que les covariables sont mesurées aux temps $t_1 < \dots < t_{m_1}$ sur l'intervalle $[0, T]$ et aux instants $t_{m_1+1} < \dots < t_{m_1+m_2}$ sur $[T, T + H]$. L'optimisation des paramètres $\{\beta_{k,\ell}\}$ repose sur la minimisation de l'erreur quadratique suivante :

$$\min_{\{\beta_{k,\ell}\}} \sum_{i=1}^n \sum_{j=m_1+1}^{m_1+m_2} \left(Y_i^{(2)}(t_j) - \hat{Y}_i^{(2)}(t_j) \right)^2. \quad (7)$$

Remarquons que dans le présent travail, la base de prédiction $\{\psi_k\}$ dans laquelle la suite de Y est estimée n'est pas optimisée. Pour une fonction Y générale, une base de B-splines est un choix convenable. La taille de la base est un hyperparamètre qui peut par exemple être choisi par validation croisée. Le problème (7) est difficile à résoudre directement : d'abord les paramètres recherchés $\{\beta_{k,\ell}\}$ appartiennent à un espace de dimension infinie, puis le modèle (6) ne reflète pas la nature discrète des données fonctionnelles en pratique. Pour surmonter ces difficultés, nous développons une formulation en dimension finie du problème (7). Pour ce faire, nous décomposons dans des bases fonctionnelles les fonctions qui interviennent dans le modèle (6), à savoir $Y_i^{(1)}(t)$, les covariables $X_{i,\ell}^{(1)}(t)$ ainsi que les paramètres $\{\beta_{k,\ell}\}$:

$$Y_i^{(1)}(t) = \sum_{j=1}^{h_0} c_{0,i,j} \theta_{0,j}(t); \quad X_{i,\ell}^{(1)}(t) = \sum_{j=1}^{h_\ell} c_{\ell,i,j} \theta_{\ell,j}(t); \quad \beta_{k,\ell}(t) = \sum_{j=1}^{q_\ell} b_{k,\ell,j} B_{\ell,j}(t) \quad (8)$$

Puis en injectant ces décompositions dans le modèle (6) et après quelques manipulations algébriques, on obtient une représentation en dimension finie de (6) qui s'écrit :

$$\hat{Y}_i^{(2)}(t) = \boldsymbol{\xi}_i(t)^\top \mathbf{b}. \quad (9)$$

avec \mathbf{b} un vecteur qui encode l'ensemble des paramètres $\{\beta_{k,\ell}\}$ et un vecteur à dépendance temporelle $\boldsymbol{\xi}_i(t)$ qui encode les fonctions $Y_i^{(1)}(t)$ et $X_{i,\ell}^{(1)}(t)$, $\ell = 1, \dots, p$. Avec l'écriture (9), le problème (7) est réduit à un problème des moindres carré ordinaire :

$$\min_{\mathbf{b}} \sum_{i=1}^n \sum_{j=m_1+1}^{m_1+m_2} \left(Y_i^{(2)}(t_j) - \boldsymbol{\xi}_i(t_j)^\top \mathbf{b} \right)^2. \quad (10)$$

En arrangeant les vecteurs ligne $\boldsymbol{\xi}_i(t_j)^\top$ dans une matrice \mathbf{R} et les scalaires $Y_i^{(2)}(t_j)$ dans un vecteur \mathbf{y} , le problème (10) se réécrit de manière plus conventionnelles :

$$\min_{\mathbf{b}} \|\mathbf{y} - \mathbf{R}\mathbf{b}\|^2. \quad (11)$$

Les coefficients optimaux \mathbf{b}^* sont obtenus via l'inverse (ou la pseudo-inverse) de la matrice de Gram $\mathbf{R}^\top \mathbf{R}$.

4 Résultats expérimentaux

Notre modèle a été évaluée sur des données fonctionnelles synthétiques. Dans nos expériences, les fonctions sont mesurées toutes les secondes aux temps

$$t_1 = 1, \dots, t_{m_1} = T, t_{m_1+1} = T + 1, \dots, t_{m_1+m_2} = T + H \quad (12)$$

On suppose que $T + H = 300$ (i.e. 5 minutes) et que la fonction $Y(t)$ ne dépend que d'une seule covariable $X(t)$. Notons $G(t; \mu, \sigma) = e^{-(t-\mu)^2/(2\sigma)^2}$ la densité gaussienne (non normalisée). Les réalisations de $X(t)$ et $Y(t)$ sont générées à partir des équations suivantes :

$$X(t) = \sum_{u=1}^{10} c_u G(t; \mu_u, \sigma); \quad Y(t) = X(t) + A \sin(2\pi ft) \quad (13)$$

Modèle	$T = 240, H = 60$ ($\tau = 80\%$)	$T = 180, H = 120$ ($\tau = 60\%$)	$T = 120, H = 180$ ($\tau = 40\%$)
Lin	40.2 (59.5)	20.2 (19.1)	1.52 (0.08)
Ridge	0.403 (0.0083)	0.608 (0.00641)	0.69 (0.00715)
Lasso	0.402 (0.00683)	0.619 (0.00763)	0.702 (0.00892)
XGB	0.461 (0.0112)	0.658 (0.00691)	0.734 (0.0118)
Notre modèle	0.369 (0.00641)	0.602 (0.015)	0.691 (0.0105)

TABLE 1 – Performances de prédiction pour trois valeurs de τ . La moyenne et l'écart-type de la RMSE sont calculés sur les $n_{\text{simu}} = 30$ simulations.

où $c_u \sim \mathcal{U}[-1, 1]$, $\mu_1 < \dots < \mu_{10}$ sont équidistants sur l'intervalle $[0, T + H]$, $\sigma \sim \mathcal{U}[20, 50]$, $A \sim \mathcal{U}[0.3, 0.7]$ et $f = 10^{-2}$. Pour chaque simulation, on génère $n_{\text{train}} = 300$ échantillons d'entraînement et $n_{\text{test}} = 2000$ échantillons de test. Un bruit gaussien de SNR = 15 dB est ajouté aux valeurs de $Y(t)$ et $X(t)$. Pour rappel, pour un SNR donné, le bruit gaussien est défini par $\varepsilon(t) \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ avec SNR = $10 \log_{10}(P/\sigma_\varepsilon^2)$ où P est la puissance de la fonction, c'est-à-dire sa moyenne quadratique.

Le vecteur de paramètre \mathbf{b} du modèle (9) est optimisé sur les échantillons d'entraînement puis l'inférence consiste à calculer les vecteurs $[\hat{Y}_i(t_{m_1+1}), \dots, \hat{Y}_i(t_{m_1+m_2})]$ pour $i = 1, \dots, n_{\text{test}}$. Des bases de B-spline cubiques ont été utilisées pour représenter les fonctions dans (8).

La métrique de performance du modèle est la RMSE (Root Mean Square Error) sur les échantillons de test :

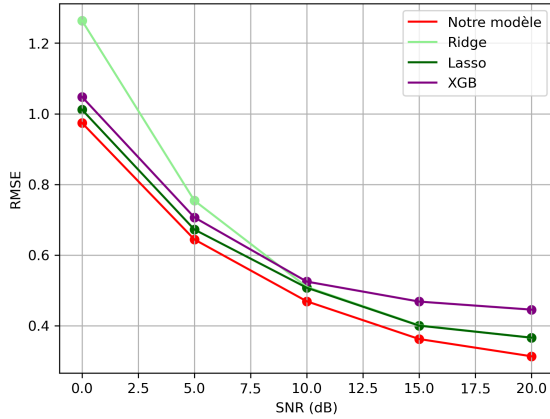
$$\text{RMSE} = \sqrt{\frac{1}{n_{\text{test}} m_2} \sum_{i=1}^{n_{\text{test}}} \sum_{j=m_1+1}^{m_1+m_2} \left(Y_i(t_j) - \hat{Y}_i(t_j) \right)^2} \quad (14)$$

Le problème étudié peut être résolu par une approche discrète avec une régression de \mathbb{R}^{2m_1} (passé de $Y(t)$ et $X(t)$) vers \mathbb{R}^{m_2} (futur de $Y(t)$). La régression linéaire (Lin), la régression ridge (Ridge), la régression lasso (Lasso) et XGBoost [6] (XGB) ont été choisis comme modèles compétiteurs.

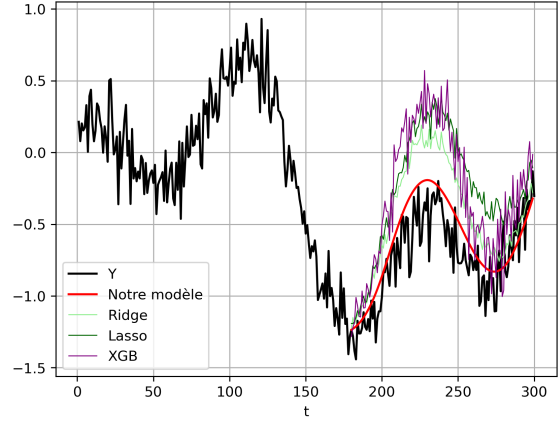
La première expérience consiste à évaluer l'impact des valeurs de T et H sur la précision des prédictions, sachant que la longueur totale $T + H = 300$ est fixée. Soit $\tau = \frac{T}{T+H} \times 100\%$ la portion de courbe observée. En général, la tâche de prédiction est d'autant plus difficile que τ diminue. Le tableau 1 synthétise les performances des modèles pour plusieurs valeurs de τ avec $n_{\text{simu}} = 30$ simulations.

La deuxième expérience consiste à fixer $T = 240$ et $H = 60$ et faire varier la force du bruit appliqué aux données. En général, la tâche de prédiction est d'autant plus difficile que le SNR est faible. Le graphique 2a compare les performances des modèles en fonction du niveau de bruit injecté, sur une seule simulation des données (des résultats moyens sur plusieurs simulation seront présentés lors de la conférence). Enfin, la figure 2b est un exemple d'inférence des modèles pour une simulation des données.

Les résultats expérimentaux dans 1 et 2a suggèrent que les prédictions de notre modèle sont plus précises que les modèles compétiteurs. Une meilleure robustesse vis-à-vis de la



(a) Performances de prédiction pour plusieurs niveaux de bruit.



(b) Exemple d'inférence sur un échantillon de test.

FIGURE 2 – Test de robustesse des modèles vis-à-vis du bruit et exemple d'inférence. La régression linéaire a été omise pour plus de visibilité car elle présente des erreurs de prédiction significativement supérieures.

fenêtre de prédiction et du bruit sont observées. De plus, le graphique 2b montre que notre modèle produit une prédiction lisse, ce qui a l'avantage d'une part de refléter la nature fonctionnelle des données, d'autre part d'avoir des prédictions plus simples à interpréter visuellement. Cela s'explique par le fait que la courbe prédite est par construction exprimée dans une base de fonctions (voir 3) tandis que les modèles concurrents sont discrets.

5 Conclusion

Nous avons développé un modèle fonctionnel pour prédire la suite d'une fonction à partir de son passé et du passé de covariables fonctionnelles. Notre approche repose sur la décomposition du futur de la fonction dans une base fonctionnelle appropriée, puis sur la prédiction des coefficients dans cette base. Des paramètres fonctionnels sont introduits et un problème des moindres carrés ordinaire est formulé. Le modèle intègre de manière naturelle la fonction à prédire et les covariables. Les résultats expérimentaux montrent que l'approche fonctionnelle produit des prédictions lisses, plus précises et plus robustes que les modèles traditionnels existants. La suite de ce travail consistera à appliquer notre modèle dans un cas pratique avec des données réelles.

Références

- [1] Ana M AGUILERA et MC AGUILERA-MORILLO. “Comparative study of different B-spline approaches for functional data”. In : *Mathematical and Computer Modelling* 58.7-8 (2013), p. 1568-1579.

- [2] Wan Anis Farhah Wan AMIR, Md Yushalify MISRO et Mohd Hafiz MOHD. “Flexible functional data smoothing and optimization using beta spline”. In : *AIMS Mathematics* 9.9 (2024), p. 23158-23181.
- [3] Anestis ANTONIADIS et al. “Clustering functional data using wavelets”. In : *International Journal of Wavelets, Multiresolution and Information Processing* 11.01 (2013), p. 1350003.
- [4] Francis BACH. *Learning theory from first principles*. MIT press, 2024.
- [5] Veerabhadran BALADANDAYUTHAPANI et al. “Bayesian Hierarchical Spatially Correlated Functional Data Analysis with Application to Colon Carcinogenesis”. In : *Biometrics* 64.1 (fév. 2008), p. 64-73. ISSN : 0006-341X. DOI : 10.1111/j.1541-0420.2007.00846.x.
- [6] Tianqi CHEN et Carlos GUESTRIN. “XGBoost : A Scalable Tree Boosting System”. In : *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA : Association for Computing Machinery, 2016, p. 785-794. ISBN : 9781450342322. DOI : 10.1145/2939672.2939785. URL : <https://doi.org/10.1145/2939672.2939785>.
- [7] Sonali DAS et al. “The effect of global crises on stock market correlations : Evidence from scalar regressions via functional data analysis”. In : *Structural Change and Economic Dynamics* 50 (2019), p. 132-147.
- [8] Clara HAPP et Sonja GREVEN. “Multivariate functional principal component analysis for data observed on different (dimensional) domains”. In : *Journal of the American Statistical Association* 113.522 (2018), p. 649-659.
- [9] Julien JACQUES et Cristian PREDÀ. “Functional data clustering : a survey”. In : *Advances in Data Analysis and Classification* 8 (2014), p. 231-255.
- [10] James O. RAMSAY et Bernard Walter SILVERMAN. *Functional Data Analysis*. Springer, 2005. ISBN : 9780387400808. URL : <http://www.worldcat.org/isbn/9780387400808>.
- [11] Philip T REISS et al. “Methods for scalar-on-function regression”. In : *International Statistical Review* 85.2 (2017), p. 228-249.
- [12] Fabrice ROSSI et Nathalie VILLA. “Support vector machine for functional data classification”. In : *Neurocomputing* 69.7-9 (2006), p. 730-742.
- [13] Jean Steve TAMO TCHOMGUI et al. “A mixture of experts regression model for functional response with functional covariates”. In : *Statistics and Computing* 34.5 (juill. 2024). ISSN : 0960-3174. DOI : 10.1007/s11222-024-10455-z. URL : <https://doi.org/10.1007/s11222-024-10455-z>.
- [14] Yuyan YI et al. “Classification of EEG signals : An interpretable approach using functional data analysis”. In : *Journal of Neuroscience Methods* 376 (2022), p. 109609. ISSN : 0165-0270. DOI : <https://doi.org/10.1016/j.jneumeth.2022.109609>. URL : <https://www.sciencedirect.com/science/article/pii/S0165027022001364>.