TEST (2025) 34:498–501
https://doi.org/10.1007/s11749-025-00969-7

DISCUSSION

Comments on: exploratory functional data analysis

Anna Maria Paganoni[1] · **Laura M. Sangalli**[1]

Received: 11 February 2025 / Accepted: 18 February 2025 / Published online: 8 April 2025

Abstract

The paper "Exploratory Functional Data Analysis" provides a comprehensive review of recent exploratory approaches for functional data, highlighting the challenges posed by the high dimensionality and complexity of these data objects. Here, we further comment on challenging aspects for the exploratory analysis of these data, which present opportunities for future research.

Keyword Phase variation of functional data · Partially observed functional data · Functional data over multidimensional and complex domains · Functional spaces beyond L^2

The paper by Qu, Dai, Euan, Sun and Genton provides a comprehensive review of recent techniques for Exploratory Functional Data Analysis (EFDA), which are fundamental in various real-world applications. It covers key statistical concepts of robust statistics, like quantiles and depth measures. The study examines visualization tools, including rainbow plots and functional boxplots, to enhance data interpretation. It also explores outlier detection methods integrated with visualization for anomaly identification. Additionally, the paper reviews clustering techniques for functional data, differentiating between dense and sparse observations. The authors also discuss future directions in EFDA, highlighting new challenges in visualizing and analyzing evolving functional data types, such as functional time series, spatial functional data, and wearable health data. We believe this paper serves as a highly valuable resource for researchers and practitioners, providing guidance for their research directions, while also offering a comprehensive literature review on EFDA.

Here, we aim to further explore the complexities of functional data, highlighting some key challenges that present opportunities for future research.

(1) Phase variation of functional data The authors appropriately discuss the problem of potential misalignment of functional data (Marron et al. 2015); indeed, if not accurately taken into account, the phase variability may blur subsequent

✉ Laura M. Sangalli
laura.sangalli@polimi.it

Anna Maria Paganoni
anna.paganoni@polimi.it

1 Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, Italy

analysis. Concerning functional outlier detection, while much interest has been dedicated to various kinds of amplitude outliers, as extensively discussed by the authors, less attention has so far been given to phase outliers (Vantini 2012). Even with effective registration procedures that may resolve most phase outliers, some observations may still show anomalous residual phase variability. This issue can be addressed either by detecting outlying warping functions or by developing appropriate methods based on robust statistics, such as depths, to identify curves exhibiting abnormal behavior in both phase and amplitude. However, much still remains to be done in this respect, especially in the case of multivariate functional data. Additionally, when present, misalignment may affect all phases of the exploratory data analysis, including, for instance, dimensional reduction and clustering, fueling a very active literature; see, e.g., the recent review in Olsen et al. (2018).

(2) Partially observed or incomplete functional data Partially observed functional data refer to the fact that each functional datum in the sample may lack any observation on extensive portions of the support. These data are also referred to as incomplete or fragmented functional data, or functional snippets. Likewise misalignment, also partial observability influences all phases of the exploratory data analysis and must be appropriately accounted for. This has spurred a highly active body of research, addressing sparse or partially observed functional data from multiple perspectives. Key approaches include functional principal component analysis [see, e.g., James et al 2000, Yao et al. 2005, Liu et al. 2017, Palummo et al. 2024], mean and covariance estimation [see, e.g., Kraus 2015, Liebl and Rameseder 2019, Lin and Wang 2022], imputation of missing data [see, e.g., Kraus 2015, Delaigle and Hall 2016, Kneip and Liebl 2020], supervised and unsupervised classification [see, e.g., James and Hastie 2001, Delaigle and Hall 2013, Stefanucci et al. 2018, Kraus and Stefanucci 2019], as well as functional depth measures, as discussed by the authors. In this respect, we would like to kindly correct the authors, by specifying that the depth measure introduced in Elías et al. (2023), and further explored in Elías and Nagy (2024), can handle functional data observed over different grids and different domains, without requiring any common domain of observation, nor data that are fully observed. While the setting partially observed and of sparse functional data has garnered significant interest, as the authors rightly discussed in their review, much work remains to be done on EFDA for this crucial data setting, which frequently arises in real-world applications.

(3) Functional data over multidimensional and complex domains The authors provide an extensive discussion on multivariate functional data. However, another increasingly relevant type of functional data in modern applications involves functions defined over multidimensional and complex supports. Such data are particularly common in medical imaging. For example, neuroimaging techniques like functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) capture biological signals from the brain. In these cases, the functional data are defined upon complex domains, such as the cortical surface or the gray matter volume [see, e.g., Lila et al. 2016, Arnone et al.2023, Clementi et al. 2023], which are non-convex and multidimensional, posing significant challenges for exploratory data analysis, particularly in terms of efficient visualization.

(4) Functional spaces beyond L^2 The authors primarily focus on functional data within an L^2 space, but in many applications, functional data reside in different functional spaces, requiring tailored exploratory data analysis techniques. For instance, densities are a form of constrained functional data that belong to the Bayes space (Van den Boogaart et al. 2014), necessitating specialized methods for classification (Nerini and Ghattas 2007) and dimensional reduction (Hron et al. 2016; Delicado 2011). Beyond densities, various other types of functional data require alternative embeddings, as discussed in Marron and Alonso (2014); Marron and Dryden (2021) and related works. To ensure accurate analysis and meaningful insights, it is essential to align exploratory data analysis methods with the specific functional space in which the data naturally reside.

We conclude by commending the authors for their work, which offers a thorough review and valuable insights into EFDA. Finally, we extend our gratitude to the Editors of TEST for the opportunity to comment on this work.

Funding Open access funding provided by Politecnico di Milano within the CRUI-CARE Agreement.

References

Arnone E, Negri L, Panzica F, Sangalli LM (2023) Analyzing data in complicated 3d domains: smoothing, semiparametric regression, and functional principal component analysis. Biometrics. https://doi.org/10.1111/biom.13845

Boogaart KG, Egozcue JJ, Pawlowsky-Glahn V (2014) Bayes hilbert spaces. Aust & N Z J Stat 56(2):171–194

Clementi L, Arnone E, Santambrogio MD, Franceschetti S, Panzica F, Sangalli LM (2023) Anatomically compliant modes of variations: new tools for brain connectivity. PLoS ONE 18(11):0292450

Delaigle A, Hall P (2013) Classification using censored functional data. J Am Stat Assoc 108(504):1269–1283. https://doi.org/10.1080/01621459.2013.824893

Delaigle A, Hall P (2016) Approximating fragmented functional data by segments of markov chains. Biometrika 103(4):779–799

Delicado P (2011) Dimensionality reduction when data are density functions. Comput Stat & Data Anal 55(1):401–420

Elías A, Jiménez R, Paganoni AM, Sangalli LM (2023) Integrated depths for partially observed functional data. J Comput Graph Statist 32(2):341–352. https://doi.org/10.1080/10618600.2022.2070171

Elías A, Nagy S (2024) Statistical properties of partially observed integrated functional depths. TEST, 1–26

Hron K, Menafoglio A, Templ M, Hrůzová K, Filzmoser P (2016) Simplicial principal component analysis for density functions in bayes spaces. Comput Stat & Data Anal 94:330–350

James GM, Hastie TJ (2001) Functional linear discriminant analysis for irregularly sampled curves. J R Stat Soc: Ser B (Stat Methodol) 63(3):533–550

James G, Hastie T, Sugar C (2000) Principal component models for sparse functional data. Biometrika 87(3):587–602

Kneip A, Liebl D (2020) On the optimal reconstruction of partially observed functional data. Ann Stat 48(3):1692–1717

Kraus D (2015) Components and completion of partially observed functional data. J R Stat Soc: Ser B (Stat Methodol) 77(4):777–801. https://doi.org/10.1111/rssb.12087

Kraus D, Stefanucci M (2019) Classification of functional fragments by regularized linear classifiers with domain selection. Biometrika 106(1):161–180

Liebl D, Rameseder S (2019) Partially observed functional data: the case of systematically missing parts. Computational Statistics & Data Analysis 131, 104–115 https://doi.org/10.1016/j.csda.2018.08.011 . High-dimensional and functional data analysis

Lila E, Aston JAD, Sangalli LM (2016) Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. Ann Appl Stat 10(4):1854–1879

Lin Z, Wang J-L (2022) Mean and covariance estimation for functional snippets. J Am Stat Assoc 117(537):348–360

Liu C, Ray S, Hooker G (2017) Functional principal component analysis of spatially correlated data. Stat Comput 27(6):1639–1654

Marron JS, Alonso AM (2014) Overview of object oriented data analysis. Biom J 56(5):732–753

Marron JS, Dryden IL (2021) Object Oriented Data Analysis. Chapman and Hall/CRC, NY

Marron JS, Ramsay JO, Sangalli LM, Srivastava A (2015) Functional data analysis of amplitude and phase variation. Statist Sci 30(4):468–484. https://doi.org/10.1214/15-STS524

Nerini D, Ghattas B (2007) Classifying densities using functional regression trees: applications in oceanology. Comput Stat & Data Anal 51(10):4984–4993

Olsen NL, Markussen B, Raket LL (2018) Simultaneous inference for misaligned multivariate functional data. J R Stat Soc: Ser C: Appl Stat 67(5):1147–1176

Palummo A, Arnone E, Formaggia L, Sangalli LM (2024) Functional principal component analysis for incomplete space-time data. Environ Ecol Stat 31(2):555–582

Stefanucci M, Sangalli LM, Brutti P (2018) Pca-based discrimination of partially observed functional data, with an application to aneurisk65 data set. Stat Neerl 72(3):246–264. https://doi.org/10.1111/stan.12137

Vantini S (2012) On the definition of phase and amplitude variability in functional data analysis. TEST 21:676–696

Yao F, Muller H-G, Wang J-L (2005) Functional data analysis for sparse longitudinal data. J Am Stat Assoc 100(470):577–590