TEST (2025) 34:490–493
https://doi.org/10.1007/s11749-025-00966-w

DISCUSSION

Comments on: Exploratory functional data analysis

Rosa E. Lillo[1,2]

Received: 9 February 2025 / Accepted: 9 February 2025 / Published online: 4 March 2025

It is a great pleasure to contribute to the discussion of the paper *Exploratory Functional Data Analysis*, which provides a useful overview of recent methods in functional data analysis (FDA), emphasizing data visualization to better understand key analytical aspects that arise when extracting meaningful insights from the data. As a culmination of the review of available options in the literature, the implementation of these methods in the R package EFDA is particularly relevant, as it will be highly useful for the scientific community.

Given the wide range of application fields where data can be modeled as functional data (both univariate and multivariate), I fully agree with the authors on the importance of promoting, improving, and innovating in all aspects related to visualization, representation, and preliminary exploration of the data.

In my view, one of the key ideas emerging from the paper is how research in FDA has promoted the ordering of functional data to enable a subsequent representation that naturally highlights features such as outliers, while also emulating, in a somewhat parallel manner, typical statistical techniques available in $\mathbb{R}^n$ (e.g., clustering and supervised classification). This ordering, as is well known, is not unique and has given rise to various definitions of depth for functional data, which seek to rank the observations from the center outwards. These definitions of depths facilitate the visualization of central regions versus extreme observations.

I would like to complement the review presented in the paper by discussing local-oriented depths as an alternative to the more commonly used global-oriented depths. The main difference is that while the global approach implies that the depth of an observation depends equally on all other observations, the local approach makes the depth of an observation more dependent on nearby observations than on those further away. Two examples of these local depth measures are the h-modal depth (HMD, Cuevas et al. 2006) and the kernelized functional spatial depth (KFSD, Sguera et al. 2014). Local depth has been shown to be particularly useful for analyzing functional samples with structures that deviate from unimodality or symmetry, especially in the context of outlier detection (Sguera et al. 2016).

✉ Rosa E. Lillo
rosaelvira.lillo@uc3m.es

1 Department of Statistics, Universidad Carlos III de Madrid, Madrid, Spain

2 IBiDat, uc3m-Santander Big Data Institute, Universidad Carlos III de Madrid, Madrid, Spain

As evidenced in the paper, a significant amount of work in FDA has focused on both providing depth definitions that capture the topology of the data and enhancing their use as tools for detecting outliers, supervised classification, or clustering. In this sense, another interesting approach to data visualization is the ordering provided by other indices that try to mimic the natural ordering in $\mathbb{R}$ and are based on the geometric concepts of the epigraph and hypograph of a function. The definition of the epigraph and hypograph indices associated with a set of functional data (Franco-Pereira et al. 2011) is inspired by the Band Depth (López-Pintado and Romo 2009), and the combined use of these two indices allowed an alternative box–plot for functional data (Martín-Barragán et al. 2016) to that of Sun and Genton (2011).

On the other hand, the combination of these indices, along with those associated with the first and second derivatives of the functional data, enables: (1) a visualization of functional data in a multivariate space, (2) the application of multivariate inferential techniques (hypothesis testing, outlier detection, clustering) based on both statistical methodologies and machine learning approaches, and (3) an algorithmic implementation with relatively low computational times. Some examples of this alternative to the use of depth measures include Franco-Pereira and Lillo (2020) and Pulido et al. (2023). As a corollary of this idea, it is worth mentioning that these indices also allow for an adapted version for multivariate functional data (Ieva and Paganoni 2020; Pulido et al. 2024).

1 Future perspectives

The rapid advancement in sensor technology across diverse fields such as neuroscience and sports now allows us to obtain vast amounts of functional data (Big Functional Data and Hig-Dimensional Functional Data), which are challenging to process, among other reasons, due to the computational constraints of algorithms originally developed for smaller functional datasets. The scientific community in the field of functional data must address these situations by providing imaginative and, above all, efficient solutions. An initiative in this regard, arising from a real-world problem in social networks (Azcorra et al. 2018), suggests that the use of indices related to functional data with a geometric interpretation based on the classic Pearson correlation enables data visualization. Moreover, these indices enable the ordering of data to detect outliers in terms of magnitude, shape, or amplitude in a scalable manner, even when dealing with millions of functional data (Ojo et al. 2022). Additionally, they allow for an adapted version for multivariate functional data, where random projections play a significant role (Ojo et al. 2023).

Another interesting area requiring further research, both in terms of visualization and analysis, is the application of FDA in engineering (Yildirim et al. 2025), where the ability to capture data from sensors once again presents a valuable opportunity for research in functional data analysis. Specifically, there are problems related to survival analysis and predictive maintenance, where the nature of the data suggests modeling them as functional data with variable domains. Although some studies have already addressed these situations by recording curves in the interval [0,1] (Aguilera-Morillo et al. 2019), adapting concepts such as MPCA (Happ and Greven 2018) represents a promising starting point in this field.

Finally, I would like to mention a topic that is not only relevant in FDA but also in statistics in general: the development of techniques to determine the optimal number of clusters in a clustering exercise. Most references in this field assume a predefined number of clusters, and there is a lack of studies specifically focused on finding the optimal number of clusters for functional data.

To conclude, I would like to congratulate the authors (Zhuo, Wenlin, Carolina, Ying, and Marc) for the excellent work done in this invited paper, and I encourage them to continue contributing with their efforts to make FDA research more visible and, above all, to make the methodology accessible for its potential, multiple, and exciting applications.

Acknowledgements This research is part of the I+D+i projects PDC2022-133359 and PID2022-137243OB-I00 funded by - MCIN/AEI/ and European Union NextGenerationEU/PRTR. This initiative has also been partially carried out within the framework of the Recovery, Transformation, and Resilience Plan funds, financed by the European Union (Next Generation) through the grant ANTICIPA (INCIBE) and The ENIA 2022 Chairs for the creation of university-industry chairs in AI-AImpulsa: UC3M-Universia.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

References

Aguilera-Morillo MC, Aguilera AM, Jiménez-Molinos F, Roldán JB (2019) Stochastic modeling of random access memories reset transitions. Math Comput Simul 159:197–209

Azcorra A, Cuevas R, Chiroque L, Fernández A, Laniado H, Lillo RE, Romo J, Sguera C (2018) Unsupervised scalable statistical method for identifying influential users in online social networks. Scientific Rep 8:6995

Cuevas A, Febrero M, Fraiman R (2006) On the use of the bootstrap for estimating functions with functional data. Comput Stat Data Anal 51(2):1063–1074

Franco-Pereira A, Lillo RE (2020) Rank tests for functional data based on the epigraph, the hypograph and associated graphical representations. Adv Data Anal Classif 14(3):651–676

Franco-Pereira AM, Lillo RE, Romo J (2011) Extremality for functional data. In: Ferraty F (ed) Recent advances in functional data analysis and related topics. Physica-Verlag HD, pp 131–134

Happ C, Greven S (2018) Multivariate functional principal component analysis for data observed on different (dimensional) domains. J Am Stat Association 113(522):649–659

Ieva F, Paganoni AM (2020) Component-wise outlier detection methods for robustifying multivariate functional samples. Stat Papers 61:595–614

López-Pintado S, Romo J (2009) On the concept of depth for functional data. J Am Stat Association 104(486):718–734

Martín-Barragán B, Lillo RE, Romo J (2016) Functional boxplots based on epigraphs and hypographs. J Appl Stat 43(6):1088–1103

Ojo O, Fernández-Anta A, Genton M, Lillo RE (2023) Multivariate functional outlier detection using the fast massive unsupervised outlier detection indices. Stat
Ojo OT, Fernández-Anta A, Lillo RE, Sguera C (2022) Detecting and classifying outliers in big functional data. Adv Data Anal Classification 16:725–760
Pulido B, Franco A, Lillo RE (2023) A fast epigraph and hypograph-based approach for clustering functional data. Stat Comput 33:36
Pulido B, Franco A, Lillo RE (2024) Clustering multivariate functional data using the epigraph and hypograph indices: a case study on madrid air quality. Submitted
Sguera C, Galeano P, Lillo RE (2014) Spatial depth-based classification for functional data. Test 23:725–750
Sguera C, Galeano P, Lillo RE (2016) Functional outlier detection by a local depth with application to nox levels. Stochastic Environ Res Risk Assessment 30(4):1115–1130
Sun Y, Genton M (2011) Functional boxplots. J Comput Gr Stat 20(2):316–334
Yildirim C, Franco-Pereira A, Lillo RE (2025) Condition monitoring and multi-fault classification of hydraulic systems using multivariate functional data analysis. Heliyon 11:e41251