TEST (2025) 34:502–507
https://doi.org/10.1007/s11749-025-00976-8

DISCUSSION

Rejoinder to the discussion on "Exploratory Functional Data Analysis"

Zhuo Qu[1,2] · Wenlin Dai[3] · Carolina Euan[4] · Ying Sun[1] · Marc G. Genton[1]

Received: 29 May 2025 / Accepted: 31 May 2025 / Published online: 23 June 2025

Keywords Clustering · Data visualization · Exploratory data analysis · Functional boxplot · Multivariate functional data · Outlier detection

1 Introduction

The authors thank all participants for their excellent discussions on various aspects of our paper. We were very pleased to see comments on the different aspects of exploratory functional data analysis that could provide a pathway to future research directions in this area. Throughout this rejoinder, we denote the discussion of Jeff Goldsmith as G, Rob Hyndman as H, Rosa Lillo as L, Sara López-Pintado as L-P, and Anna Maria Paganoni and Laura Sangalli as PS. These discussants have made foundational contributions to the development and application of EFDA tools. Their insights offer a balanced and constructive perspective that greatly enriches the discussion.

This research was supported by the King Abdullah University of Science and Technology (KAUST).

✉ Marc G. Genton
marc.genton@kaust.edu.sa

Zhuo Qu
zhuo.qu@kaust.edu.sa; zhuo.qu@stjude.org

Wenlin Dai
wenlin.dai@ruc.edu.cn

Carolina Euan
c.euancampos@lancaster.ac.uk

Ying Sun
ying.sun@kaust.edu.sa

1 Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

2 Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA

3 Center for Applied Statistics, Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, China

4 Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK

This rejoinder covers the different points raised by the discussants in different sections. Section 2 discusses enhancements to EFDA methodologies, including visualization strategies and alternative analytical workflows. Section 3 focuses on the central role of data depth, its application in visualization tools such as the functional boxplot, and various approaches to outlier detection. Section 4 addresses the challenges presented by complex functional data structures and the consequent need for robust and scalable methods. Finally, Section 5 highlights other specific future research directions identified by the discussants.

2 Enhancement of exploratory functional data analysis

We appreciate the discussants highlighting several valuable extensions and alternative points of view that enrich the EFDA toolkit presented in our review.

G suggested a practical enhancement: incorporating non-functional covariates directly into visualizations (e.g., rainbow plots for continuous variables, and color or facets for categorical variables) to facilitate comparisons across groups or covariate values. We agree that this can provide valuable initial insights in applied contexts. G also recommended the R package *tidyfun* (Scheipl et al. 2024) for quick and easy EFDA. It is true that *tidyfun* facilitates functional data analysis by providing a tidy data structure and functions that integrate seamlessly with established R workflows. This makes EFDA more accessible and efficient for practitioners.

H proposed a compelling alternative workflow: perform a robust principal component analysis (PCA, Croux et al. 2007) first and then apply anomaly detection tools (Dai and Genton 2019, Kandanaarachchi and Hyndman 2022) to the resulting low-dimensional principal component scores. As demonstrated with the annual US mortality data (1933–2022), this approach effectively leverages existing, well-understood multivariate methods for tasks like anomaly detection, offering computational efficiency and familiarity, although acknowledging that the PCA transformation might not perfectly preserve all original features. The latter issue was illustrated in a simulation study by Sun and Genton (2011); see their Figure 5 and Table 2.

3 Data depth, visualization, and outlier detection

Data depth, visualization, and outlier detection are central to EFDA, and the discussions provided valuable insights, particularly regarding depth-based approaches.

The concept of data depth was central in the discussions by L, L-P, and PS. They emphasized its role in providing a principled ordering for functional data but also acknowledged the diversity of depth notions. These include local depths (e.g., HMD, Cuevas et al. 2006; KFSD, Sguera et al. 2014), suitable for multimodal or asymmetric data, versus global depths. Further distinctions exist between infimal and integral depths, which have implications for properties such as band convexity and behavior with noisy data (L-P; Nagy et al. 2024). Additionally, based on the combined use of two indices, epigraphs and hypographs, of a function (López-Pintado and Romo 2009;

Franco-Pereira et al. 2011), along with those associated with the first and second derivatives of the functional data, alternatives to the use of depth measures (Franco-Pereira and Lillo 2020; Pulido et al. 2023) are considered to facilitate sorting, visualization, and clustering techniques with relatively low computation costs. We appreciate PS clarifying the capability of the integrated depth introduced by Elías et al. (2023) and further explored in Elías and Nagy (2025), to handle data observed over different domains without requiring full observation or a common domain. We also would like to highlight the work of Qu et al. (2022), who introduced global depths for *multivariate* functional data explicitly addressing diverse missing data scenarios.

The functional boxplot (Sun and Genton 2011) was discussed by L-P and L. L-P noted its limitations with respect to shape or phase outliers when using integrated depths and discussed recent work on decomposing variability or using alternative depths (such as the infimal depth of Nagy et al. (2024), which satisfies the convexity of the band but is sensitive to noise). We confirm that the construction of the function boxplot can be based on any measure of functional depth, as originally mentioned by Sun and Genton (2011), although some depths might be more suitable for the problem at hand. It is also worth noting that not all depth notions are equally robust—some may perform poorly in the presence of noise or outliers. Therefore, careful consideration must be taken when selecting a depth measure, especially in practical applications involving complex or irregular data. L mentioned the alternative boxplot for functional data (Martin-Barragan et al. 2016) with epigraphs and hypographs of a function. The consensus, with which we concur, is that the choice of depth depends on the data characteristics and analysis goals, and combining insights from multiple depth measures or exploratory tools is often beneficial.

Regarding outlier detection, PS highlighted the persistent issue of phase variation, where misalignment can obscure amplitude features, and noted the need for methods to address phase outliers (Vantini 2012), especially for multivariate functional data. H presented an effective alternative using anomaly detection algorithms on principal component scores (Kandanaarachchi and Hyndman 2022), which can identify different outliers compared to depth-based methods such as directional outlyingness (Dai and Genton 2019). L highlighted the utility of local depths and geometric indices (Azcorra et al. 2018; Ieva and Paganoni 2020; Ojo et al. 2022; Ojo et al. 2023) to identify outliers, particularly in large-scale datasets. These different perspectives (depth-based, PCA-based, index-based) enrich the approaches available for identifying abnormal observations in functional data.

4 Challenges in exploratory functional data analysis

The discussants highlighted several challenges, primarily concerning complex functional data structures and the corresponding need for efficient computational algorithms.

PS reviewed active research in partially observed functional data or functional snippets. Partial observations influence all phases of EFDA, spurring the development of specific approaches including functional principal component analysis, mean and covariance estimation, imputation of missing data, supervised and unsupervised depth measures, as well as adaptive depth measures in our review paper.

PS also underscored the difficulties arising from functional data defined over multi-dimensional and/or complex domains commonly appearing in medical imaging (e.g., functional magnetic resonance imaging, magnetoencephalography), where visualization and analysis are particularly challenging. Furthermore, they reminded us that functional data often reside in functional spaces beyond L^2, e.g., densities defined in the Bayes space (Van den Boogaart et al. 2014), necessitating tailored methods that respect the geometry of the specific data. Likewise, L-P broadened the scope by extending the exploratory analysis framework from functional data to non-Euclidean object data (e.g., directions, covariance matrices, trees). The development of tools such as metric half-space depth (Dai et al. 2023), which reflects the intrinsic geometry of such data, is crucial for exploring these increasingly common data types.

L highlighted the significant challenge of scalability when working with "big functional data and high-dimensional functional data," emphasizing the need for computationally efficient algorithms to handle processing, visualization, and outlier detection. This underscores the need for algorithmic advances to ensure that EFDA methods remain feasible in real-world settings. One notable example is the development of a fast algorithm for computing the modified band depth (MBD) used in functional boxplots; see Sun et al. (2012). It demonstrates how computational improvements can dramatically expand the applicability of EFDA tools. In the future, the development of statistically robust and computationally scalable methods will be critical to meet the demands of increasingly large and complex functional datasets. Such methods must also be geometrically appropriate for diverse and complex functional data types while maintaining computational efficiency.

5 Other future directions

Beyond the broader challenges discussed previously, the discussants also identified several specific and practical avenues for future EFDA research, including its application in new domains, improvements in clustering methodologies, and further investigation of the effects of noise.

For example, L pointed to the growing relevance of EFDA in engineering as another key research direction (Yildirim et al. 2025). Data captured from sensors in engineering contexts, for instance, in survival analysis and predictive maintenance, often manifest as functions defined on variable domains. This characteristic presents both a unique challenge and a clear opportunity for advancing EFDA techniques. L also raised a general question regarding the optimal number of clusters in a clustering exercise. The development of techniques to determine the optimal number of clusters may be an exciting area for both general EDA and specific EFDA research. Additionally, L-P touched on the practical issue of the impact of noise on functional depth measures and derived tools, an area that deserves further investigation.

In conclusion, the discussions reinforced that EFDA is a dynamic and essential area. We are grateful for the insightful contributions that highlight the progress made and chart important directions for future work, focusing on robustness, adaptability, scalability to complex data types, and user-friendly implementations.

Funding Open access publishing provided by King Abdullah University of Science and Technology (KAUST).

References

Azcorra A, Chiroque LF, Cuevas R, Fernández Anta A, Laniado H, Lillo RE, Romo J, Sguera C (2018) Unsupervised scalable statistical method for identifying influential users in online social networks. Sci Rep 8(1):6955

Croux C, Filzmoser P, Oliveira M (2007) Algorithms for projection-pursuit robust principal component analysis. Chemom Intell Lab Syst 87:218–225

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

Dai W, Genton MG (2019) Directional outlyingness for multivariate functional data. Comput Stat Data Anal 131:50–65

Dai X, López-Pintado S, Initiative ADN (2023) Tukey's depth for object data. J Am Stat Assoc 118(543):1760–1772

Elías A, Jiménez R, Paganoni AM, Sangalli LM (2023) Integrated depths for partially observed functional data. J Comput Graph Stat 32(2):341–352

Elías A, Nagy S (2025) Statistical properties of partially observed integrated functional depths. TEST 34(1):125–150

Franco-Pereira AM, 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: Recent advances in functional data analysis and related topics, pp 131–134. Springer

Ieva F, Paganoni AM (2020) Component-wise outlier detection methods for robustifying multivariate functional samples. Stat Pap 61(2):595–614

Kandanaarachchi S, Hyndman RJ (2022) Leave-one-out kernel density estimates for outlier detection. J Comput Graph Stat 31(2):586–599

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

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

Nagy S, Mrkvička T, Elías A (2024) Which depth to use to construct functional boxplots? arXiv preprint arXiv:2409.18603

Ojo OT, Fernández Anta A, Genton MG, Lillo RE (2023) Multivariate functional outlier detection using the fast massive unsupervised outlier detection indices. Stat 12(1):e567

Ojo OT, Fernández Anta A, Lillo RE, Sguera C (2022) Detecting and classifying outliers in big functional data. Adv Data Anal Classif 16(3):725–760

Pulido B, Franco-Pereira AM, Lillo RE (2023) A fast epigraph and hypograph-based approach for clustering functional data. Stat Comput 33(2):36

Qu Z, Dai W, Genton MG (2022) Global depths for irregularly observed multivariate functional data. arXiv preprint arXiv:2211.15125

Scheipl F, Goldsmith J, Wrobel J (2024) tidyfu: tools for tidy functional data. R package version 0.0.98. https://tidyfun.github.io/tidyfun/
Sguera C, Galeano P, Lillo R (2014) Spatial depth-based classification for functional data. TEST 23(4):725–750
Sun Y, Genton MG (2011) Functional boxplots. J Comput Graph Stat 20(2):316–334
Sun Y, Genton MG, Nychka DW (2012) Exact fast computation of band depth for large functional datasets: How quickly can one million curves be ranked? Stat 1:68–74
Van den Boogaart KG, Egozcue JJ, Pawlowsky-Glahn V (2014) Bayes hilbert spaces. Aust New Zealand J Stat 56(2):171–194
Vantini S (2012) On the definition of phase and amplitude variability in functional data analysis. TEST 21(4):676–696
Yildirim C, Franco-Pereira AM, Lillo RE (2025) Condition monitoring and multi-fault classification of hydraulic systems using multivariate functional data analysis. Heliyon 11(1)