TEST (2025) 34:494–497
https://doi.org/10.1007/s11749-025-00968-8

DISCUSSION

Comments on: Exploratory functional data analysis

Sara Lopez-Pintado[1]

Received: 13 February 2025 / Accepted: 18 February 2025 / Published online: 18 March 2025

Abstract
In this invited paper we highlight some of the exploratory functional data methods described in the systematic review paper by Qu et al. (TEST, 2024. 10.1007/s11749-024-00952-8). We discuss recent developments related to functional boxplots and consider possible extensions of exploratory methods to non-Euclidean object data.

Keywords Data depth · Boxplot · Object data

1 Data depth

The paper by Qu et al. (2024) provides a thorough and systematic review of recent exploratory functional data analysis (EDA) methods. It offers a broad overview of descriptive tools developed over the past few decades for various types of functional data, including multivariate and irregularly observed functional data. The paper discusses summary statistics, functional data depth notions, and visualization tools, such as different versions of the functional boxplot. Additionally, it provides a detailed review of outlier detection methods for different types of functional data, as well as clustering approaches. The EDA methods are illustrated using real data applications in medicine and weather analyses.

Qu et al. (2024) is a timely contribution, as the increasing complexity and multi-faceted nature of observational data in many studies have heightened the need for an agnostic exploratory data analysis approach. Exploring data helps reveal distributional properties and distinguishes extreme observations from typical ones. In this context, a crucial first step is to address the absence of a canonical ordering for general functional data by proposing principled definitions of rank, median, order statistics, and outliers. Additionally, the introduction of visualization tools aids in understanding the data before diving into modeling.

✉ Sara Lopez-Pintado
s.lopez-pintado@northeastern.edu

1 Department of Public Health and Health Sciences, Northeastern University, Boston, USA

Data depth has proven to be a powerful exploratory and data-driven tool for ranking observations and uncovering features of the underlying data distribution. Originally introduced for multivariate Euclidean data, data depth provides a measure of how "representative" or "outlying" an observation is relative to a probability distribution (see, e.g., Tukey 1975; Zuo and Serfling 2000). Over the past several decades, this concept has been extended to functional data (Lopez-Pintado and Romo 2009; Narisetty and Nair 2016, among others). Qu et al. (2024) provide a detailed review of depth notions proposed for general functional data, including multivariate functional data—where functions take values in a multivariate space—and sparse functional data, where the grid of observed time points varies across functions (e.g., Sguera and Lopez-Pintado 2021; Elias et al. 2023).

2 Functional boxplot

One of the most widely recognized visualization tools based on functional data depth is the functional boxplot, introduced by Sun and Genton (2011). This tool has been extended to various settings, including functional boxplots for sparse functional data and the two-stage functional boxplot (Qu and Genton 2022, 2022; Dai and Genton 2018, among others). The original functional boxplot has gained popularity for summarizing the distribution of functional data and identifying potential outliers. However, it has limitations, such as its inability to detect shape-based or misaligned/shifted outliers when using an integrated type of depth. These issues have been addressed in recent works (Qu and Genton 2022; Xie et al. 2017) and are discussed in detail in Qu et al. (2024). In particular, Xie et al. (2017) decomposed functional data variability into three components—amplitude, phase, and vertical translation—using curve registration and proposed separate visualizations for each component. This approach is especially useful for analyzing misaligned curves.

Another limitation of the original functional boxplot is the loss of interpretability within the band representing the 50% central region**.** Nagy et al. (2024) recently demonstrated that when using an integral-type depth, the central region (or "box") in the functional boxplot does not satisfy a band convexity property. This means that some curves within the band may have lower depth than those used to construct the region, making interpretation challenging. However, the band convexity property holds when using an infimal-type depth, which ensures that all curves within the band are at least as deep as those defining it. On the other hand, infimal-type depths are highly sensitive to noise in functional data. Only a few recent studies have examined the impact of noise on functional depths and their implications for depth-based exploratory tools (see Nagy et al. 2024). Further research in this area would be valuable.

Overall, a key takeaway regarding functional data depth methods is that no single depth measure outperforms all others in every aspect**.** The choice of functional depth and visualization tool depends on the nature of the functional data and the objectives of the study. In practice, combining multiple methods or depth measures often provides complementary insights and leads to a more comprehensive exploratory analysis.

3 Extensions to object non-Euclidean data

In addition to the infinite-dimensional functional data examined in detail by Qu et al. (2024), other complex data objects are increasingly being generated across scientific disciplines and are rapidly gaining relevance. One important class of such data is finite-dimensional non-Euclidean data (Marron and Alonso 2014), which models objects such as directions, covariance matrices, and trees. Significant progress has been made in developing methods and theory to address the complexity of these data, including location measures (Fréchet 1948), statistical inference (Bhattacharya and Patrangenaru 2005), and classification (Dai and Müller 2018). However, exploratory data analysis (EDA) for nonstandard data remains an underdeveloped area. To address this gap, Dai and Lopez-Pintado (2023) proposed a metric halfspace depth for object data in general metric spaces, extending the celebrated Tukey's depth for Euclidean data. This new depth measure possesses desirable theoretical properties and adapts to the intrinsic geometry of the data. As a result, it can serve as a foundation for developing novel exploratory and visualization tools for this type of object data.

Funding Open access funding provided by Northeastern University Library.

References

Bhattacharya R, Patrangenaru V (2005) Large sample theory of intrinsic and extrinsic sample means on manifolds - II. Ann Stat 33:1225–1259

Dai W, Genton MG (2018) Multivariate functional data visualization and outlier detection. J Comput Graph Stat 27(4):923–934

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

Dai X, Müller H-G (2018) Principal component analysis for functional data on Riemannian manifolds and spheres. Ann Stat 46:3334–3361

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

Fréchet M (1948) Les éléments aléatoires de nature quelconque dans un espace distancié. Ann L'inst Henri Poincaré 10:215–310

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

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

Nagy S, Ferraty F (2019) Data depth for measurable noisy random functions. J Multivariate Analysis 170:95–114

Nagy S, Mrkvicka T, Elias A (2024) Which depth to use to construct functional boxplots? arXiv:2409.18603v1

Narisetty NN, Nair VN (2016) Extremal depth for functional data and applications. J Am Stat Assoc 111(516):1705–1714

Qu Z, Genton MG (2022) Sparse functional boxplots for multivariate curves. J Comput Graph Stat 31(4):976–989

Qu Z, Dai W, Euan C, Sun Y, Genton MG (2024) Exploratory functional data analysis. TEST. https://doi.org/10.1007/s11749-024-00952-8

Sguera C, Lopez-Pintado S (2021) A notion of depth for sparse functional data. TEST 30(3):630–649

Sun Y, Genton MG (2011) Functional boxplots. J Comput Graph Stat 20(2):316–334

Tukey JW (1975) Mathematics and the picturing of data. In: Proceedings of the international congress of mathematicians, vol 2, pp 523–531

Xie W, Kurtek S, Bharath K, Sun Y (2017) A geometric approach to visualization of variability in functional data. J Am Stat Assoc 112(519):979–993

Zuo Y, Serfling R (2000) General notions of statistical depth function. Ann Stat 28:461–482