Evolution Outlier in High Dimensional Functional Time Series

Antonio Elías, Juan Miguel Morales and Salvador Pineda

Outline

Evolution Outliers in High Dimensional Functional Time Series

Introduction

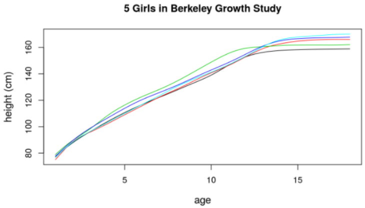
Functional Data Analysis

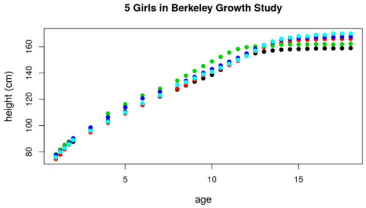
Functional Data Analysis (FDA) focuses on the analysis of smooth curves:

$$y_i(x), \quad x \in [a, b], \quad i = 1, \ldots, N.$$

However, in practise we observe discrete evaluations of functional data samples:

$$y_i(\tilde{x}_j), \quad \tilde{x} \in [a, b], \quad i = 1, \ldots, N \quad j = 1, \ldots, p.$$





- Ramsay, J. O. and Silverman, B. W. (2005). Functional Data Analysis, Springer series in Statistics.
- Special Issues on Functional Data Analysis and Related Topics, Journal of Multivariate Analysis.

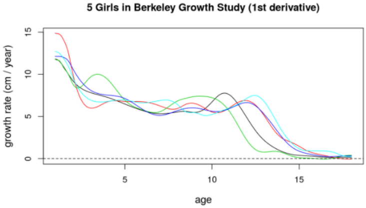
Functional Data Analysis

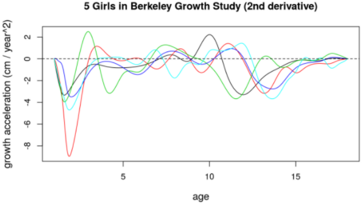
FDA propose to express discrete data by means of a basis expansion,

$$y_t(x) \approx \sum_{k=1}^{K} c_{t,k}\phi_k(x),$$

where $c_{t,k}$ are coefficients and ϕ_k a collection of basis functions with known derivatives

$$\partial_x^i y_t(x) \approx \sum_{k=1}^{K} c_{t,k}\partial_x^i \phi_k(x)$$





Ramsay, J. O. and Silverman, B. W. (2005). Functional Data Analysis, Springer series in Statistics.

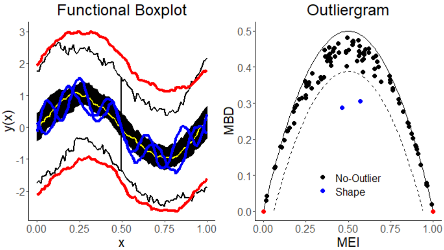
Special Issues on Functional Data Analysis and Related Topics, Journal of Multivariate Analysis.

Outliers in the context of Functional Data Analysis

Taxonomy of outliers IID samples [Hubert et al., 2005]:

Magnitude and shape.

Global or transitory.



Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational & Graphical Statistics*, 20(2):316–334.

Arribas-Gil, A. and Romo, J. (2014). Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics*, 15(4):603–619.

FTS and HDFTS

Functional Time Series

A sample of curves indexed in time $t = 1, \ldots, T$,

$$y^1(x), \ldots, y^T(x).$$

High Dimensional Functional Time Series

Multiple FTS indexed by $i = 1, \ldots, N$,

$$\mathbf{y}(x) = \begin{bmatrix} y_1^1(x) & y_1^2(x) & \ldots & y_1^N(x) \\ y_2^1(x) & y_2^2(x) & \ldots & y_2^N(x) \\ \vdots & \vdots & \ddots & \vdots \\ y_T^1(x) & y_T^2(x) & \ldots & y_T^N(x) \end{bmatrix}.$$

P. Raña, G. Aneiros, and J. M. Vilar, (2015) "Detection of outliers in functional time series", *Environmetrics*, 26(3):178–191.

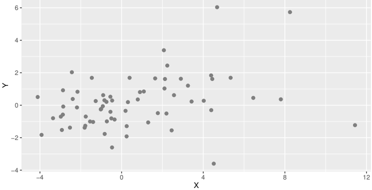
Gao, Y., Shang, HL. and Yang, Y. (2011) "High-dimensional functional time series forecasting: An application to age-specific mortality rates", *Journal of Multivariate Statistics*, 170:232-243.

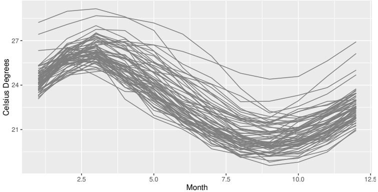
Depth-based methodology for grouped FTS

Depth Measures

Depth Measures

This simple problem becomes harder in other spaces





It is not straightforward → **Depth Measures**

Depth Measures

- The concept of **depth measure** was introduced by [Tukey, 1975].

Depth Measures

- The concept of **depth measure** was introduced by [Tukey, 1975].
- Given a datum $x \in \Omega$ from a distribution $P \in \mathcal{P}(\Omega)$ and $\Omega = \mathbb{R}^d$, a depth is a bounded and non-negative function

$$\begin{aligned} \mathrm{D} : \Omega \times \mathcal{P}(\Omega) &\rightarrow [0, 1] \\ (x, P) &\mapsto \mathrm{D}(x; P). \end{aligned}$$

Depth Measures

- The concept of **depth measure** was introduced by [Tukey, 1975].
- Given a datum $x \in \Omega$ from a distribution $P \in \mathcal{P}(\Omega)$ and $\Omega = \mathbb{R}^d$, a depth is a bounded and non-negative function

$$\begin{aligned} \mathrm{D} : \Omega \times \mathcal{P}(\Omega) &\rightarrow [0,1] \\ (x, P) &\mapsto \mathrm{D}(x; P). \end{aligned}$$

- It provides a center-outward ordering of points being the **deepest point** or **median**,

$$_{x \in \mathcal{P}} \mathrm{D}(x, P).$$

Functional Depth Measures

- The **first depth** was proposed by [Fraiman and Muniz, 2001].

Functional Depth Measures

- The **first depth** was proposed by [Fraiman and Muniz, 2001].
- **Properties**: [Nieto-Reyes and Battey, 2016] and [Nagy et al., 2017].

Functional Depth Measures

- The **first depth** was proposed by [Fraiman and Muniz, 2001].
- **Properties**: [Nieto-Reyes and Battey, 2016] and [Nagy et al., 2017].
- **Other definitions**:
  [Cuevas et al., 2007, Cuevas and Fraiman, 2009, López-Pintado and Romo, 2009, López-Pintado and Romo, 2011, Dutta et al., 2011, Mosler and Polyakova, 2012, Mosler, 2013, Sguera et al., 2014, López-Pintado et al., 2014, Claeskens et al., 2014, Nagy et al., 2021], etc.

- The **first depth** was proposed by [Fraiman and Muniz, 2001].
- **Properties**: [Nieto-Reyes and Battey, 2016] and [Nagy et al., 2017].
- **Other definitions**:
  [Cuevas et al., 2007, Cuevas and Fraiman, 2009, López-Pintado and Romo, 2009, López-Pintado and Romo, 2011, Dutta et al., 2011, Mosler and Polyakova, 2012, Mosler, 2013, Sguera et al., 2014, López-Pintado et al., 2014, Claeskens et al., 2014, Nagy et al., 2021], etc.

Two families of functional depths [Nagy et al., 2016]

1. **Integrated depth measures**
2. Non integrated depth measures

Integrated Depth for Functional Data Analysis

- $X : [0, 1] \to \mathbb{R}$.

- $X : [0, 1] \to \mathbb{R}$.
- P is the law of X where $P \in \mathcal{P}(\Omega)$ being $\Omega = C([0, 1])$.

Integrated Depth for Functional Data Analysis

- $X : [0, 1] \to \mathbb{R}$.
- P is the law of X where $P \in \mathcal{P}(\Omega)$ being $\Omega = C([0, 1])$.
- P_t is the marginal probability distribution of $X(t)$ and $t \in [0, 1]$.

- $X : [0, 1] \to \mathbb{R}$.
- P is the law of X where $P \in \mathcal{P}(\Omega)$ being $\Omega = C([0, 1])$.
- P_t is the marginal probability distribution of $X(t)$ and $t \in [0, 1]$.

$$X_1, \ldots, X_n \sim P$$

Integrated Depth for Functional Data Analysis

- $X : [0, 1] \to \mathbb{R}$.
- P is the law of X where $P \in \mathcal{P}(\Omega)$ being $\Omega = C([0, 1])$.
- P_t is the marginal probability distribution of $X(t)$ and $t \in [0, 1]$.

$$X_1, \dots, X_n \sim P$$

Definition: Integrated Functional Depths (FD) [Nagy et al., 2016]

Given an univariate depth measure D, and a weighting function $w : [0, 1] \to [0, \infty)$, with $\int_0^1 w(t) = 1$, the Integrated Functional Depth of x with respect to P is

Integrated Depth for Functional Data Analysis

- $X : [0,1] \to \mathbb{R}$.
- P is the law of X where $P \in \mathcal{P}(\Omega)$ being $\Omega = C([0,1])$.
- P_t is the marginal probability distribution of $X(t)$ and $t \in [0,1]$.

$$X_1, \ldots, X_n \sim P$$

Definition: Integrated Functional Depths (FD) [Nagy et al., 2016]

Given an univariate depth measure D, and a weighting function $w : [0,1] \to [0,\infty)$, with $\int_0^1 w(t) = 1$, the Integrated Functional Depth of x with respect to P is

$$\mathrm{FD}_w(x, P) = \int_0^1 \mathrm{D}(x(t), P_t) w(t) dt.$$

Integrated Depth for Functional Data Analysis

- $X : [0,1] \to \mathbb{R}$.
- P is the law of X where $P \in \mathcal{P}(\Omega)$ being $\Omega = C([0,1])$.
- P_t is the marginal probability distribution of $X(t)$ and $t \in [0,1]$.

$$X_1, \ldots, X_n \sim P$$

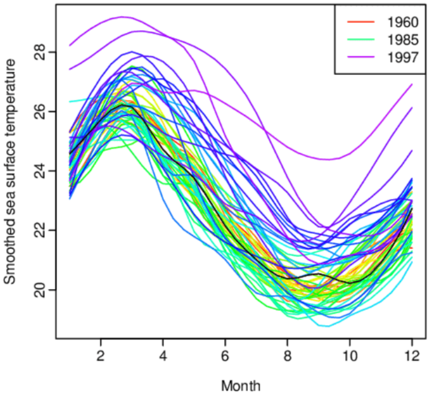
Definition: Integrated Functional Depths (FD) [Nagy et al., 2016]

Given an univariate depth measure D, and a weighting function $w : [0,1] \to [0,\infty)$, with $\int_0^1 w(t) = 1$, the Integrated Functional Depth of x with respect to P is

$$\mathrm{FD}_w(x, P) = \int_0^1 \mathrm{D}(x(t), P_t) w(t) dt.$$

Different D provides different FD. Examples are the **Fraiman and Muniz Depth** (FM), **Modified Band Depth** (MBD) and the **Modified Half Region Depth** (MHRD).

Integrated Depth for Functional Data Analysis



Functional Depths in the context of FTS

- The sample of curves is indexed in time

$$y^1(x), \ldots, y^T(x) \text{ for } t = 1, \ldots, T.$$

Functional Depths in the context of FTS

- The sample of curves is indexed in time

$$y^1(x), \ldots, y^T(x) \text{ for } t = 1, \ldots, T.$$

- Time series of depths

$$\{\mathrm{FD}(y^t, P_T) = \mathrm{FD}(t), \quad t \in (1, \ldots, T)\}.$$

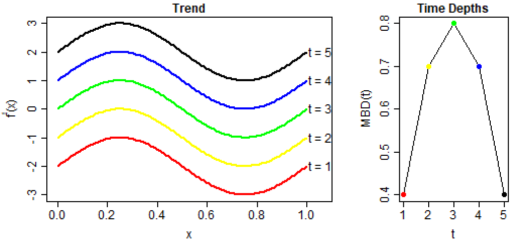
Functional Depths in the context of FTS

- The sample of curves is indexed in time

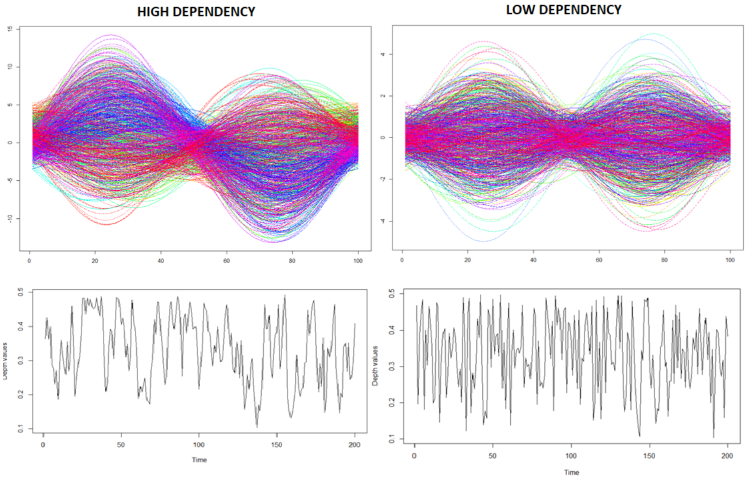
$$y^1(x), \ldots, y^T(x) \text{ for } t = 1, \ldots, T.$$

- Time series of depths

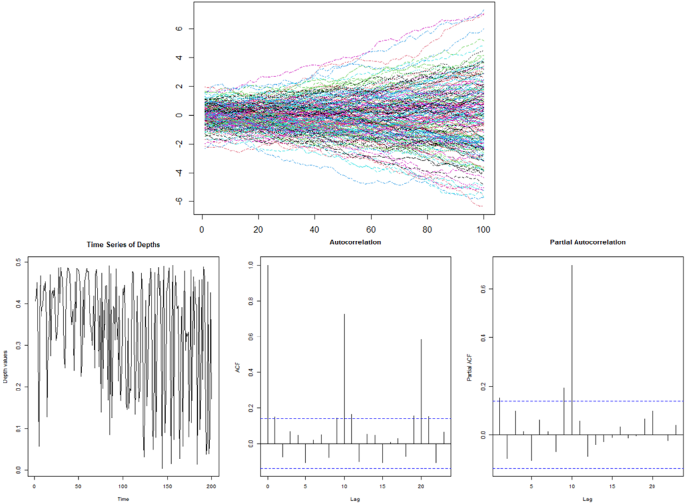
$$\{\text{FD}(y^t, P_T) = \text{FD}(t), \quad t \in (1, \ldots, T)\}.$$



FTS from dynamic factor models [Gao et al., 2019]

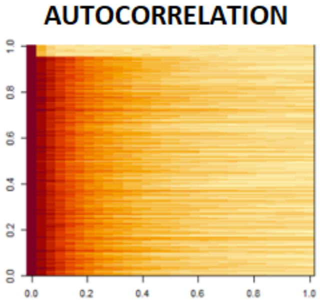


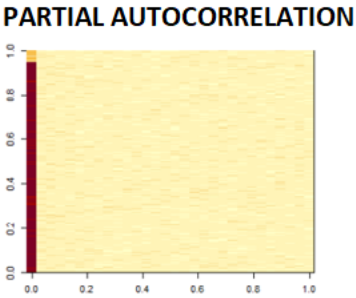
FTS from Seasonal Functional Autorregresive model [Zamani et al., 2021]



HDFTS from dynamic factor models [Gao et al., 2019]

- N = 500.
- T = 1000.
- Contamination = 5%





Methodology for grouped FTS

- Time depths of each FTS as,

$$\mathbf{FD}(t) = [\mathrm{FD}^1(t), \mathrm{FD}^2(t), \dots, \mathrm{FD}^N(t)],$$

- **Prototype evolution** is a trimmed mean [Fraiman and Muniz, 2001],

$$\mu\mathrm{FD}(t) = \frac{1}{\lceil \alpha N \rceil} \sum_{r=1}^{\lceil \alpha N \rceil} \mathrm{FD}^{[r]}(t),$$

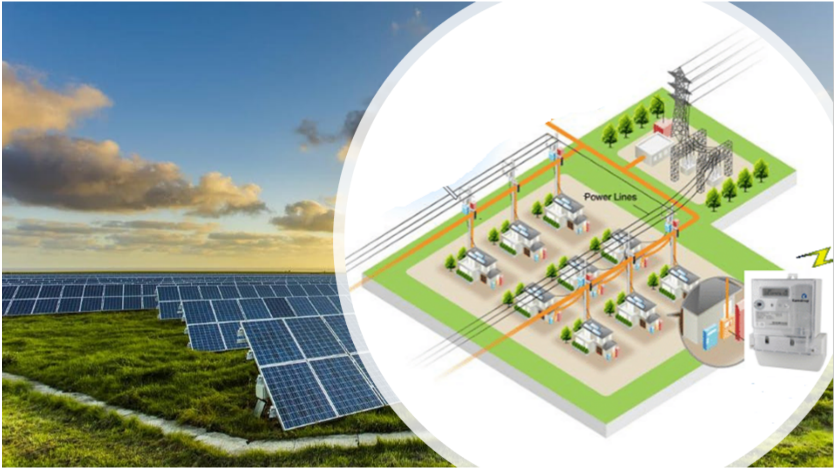
- **Euclidean distance** between each $\mathbf{FD}(t)$ and the prototype $\mu\mathrm{FD}(t)$, i.e.,

$$d(\mathrm{FD}^i(t), \mu\mathrm{FD}(t)) = \sqrt{\sum_{t=1}^{T} \left(\mathrm{FD}^i(t) - \mu\mathrm{FD}(t)\right)^2},$$

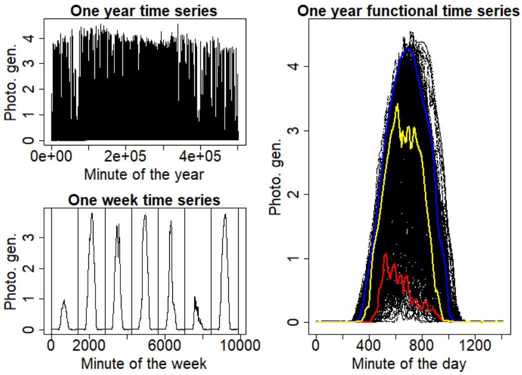
- **Outlier detection rule** for skewed distributions [Hubert and Vandervieren, 2008],

$$d(\mathrm{FD}^i(t), \mu\mathrm{FD}(t)) > Q_3(\mathbf{d}) + \gamma \times \exp^{3MC} \times IQR(\mathbf{d}).$$

Case Study: Smart meters data


Power Lines

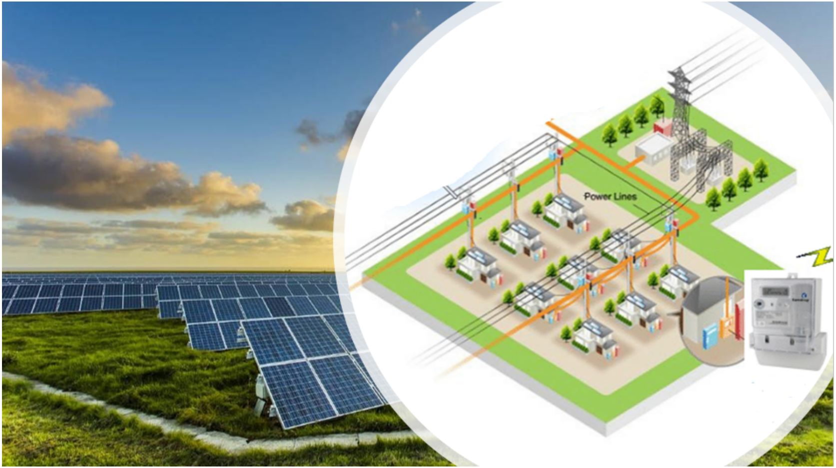
From one meter to FDA and Functional Time Series



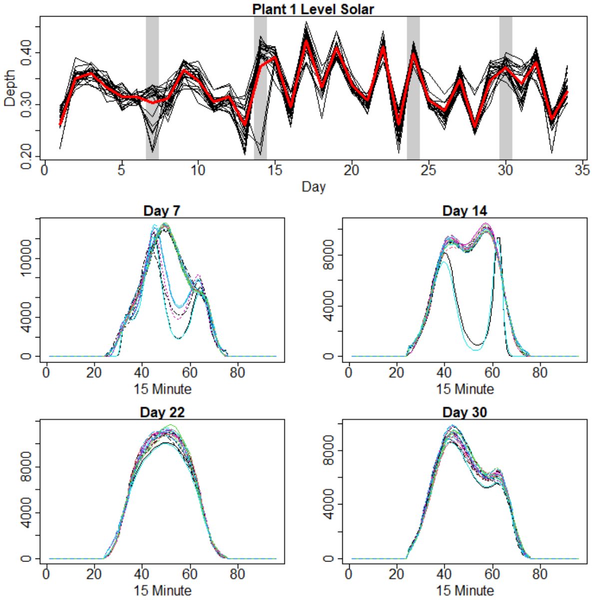
Being t the index of days,

$$y^1(x), \dots, y^T(x), \quad x = \{0, 1, \dots, 1440\}, \quad t = \{1, \dots, 365\}.$$

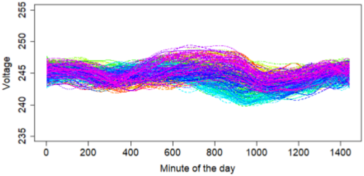
FDA in the Context of Smart Meters



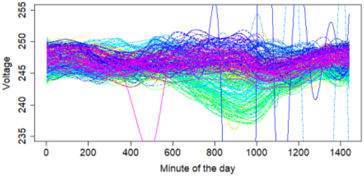
Solar farm. N = 22, T = 34.



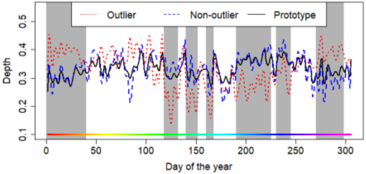
Household voltage circuit. N = 22, T = 365.



(a) Non-outlier vol_{2818}

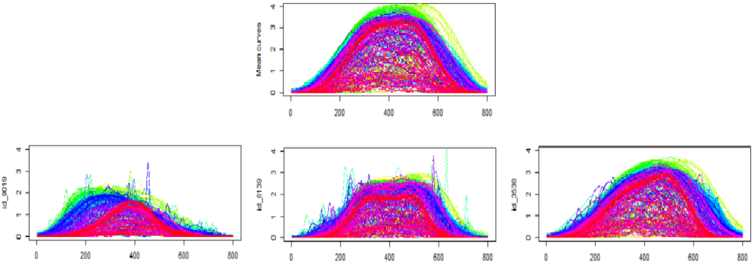


(b) Outlier vol_{9019}



(c) Time series of depths

Household photo-voltaic energy generation. N = 22, T = 365.



Analysis of the first derivatives emphasis solar panel orientation and tilt differences.

Conclusions

Conclusions

- We propose a method to detect evolution outliers in grouped FTS.
- Theoretical links between the temporal structure of the FTS and the temporal structure of the time series of depths require more research.
- Visit our website for more details `oasys.uma.es`.

  Article: `https://arxiv.org/abs/2107.01144`.
  Code: `smartOASYS` an R-package available at our Github Organization.
- Contact: `aelias@uma.es`.

Evolution Outlier in High Dimensional Functional Time Series

Antonio Elías, Juan Miguel Morales and Salvador Pineda

Claeskens, G., Hubert, M., Slaets, L., and Vakili, K. (2014).
Multivariate functional halfspace depth.
Journal of the American Statistical Association, 109(505):411–423.

Cuevas, A., Febrero-Bande, M., and Fraiman, R. (2007).
Robust estimation and classification for functional data via projection-based depth notions.
Computational Statistics, 22(3):481–496.

Cuevas, A. and Fraiman, R. (2009).
On depth measures and dual statistics. a methodology for dealing with general data.
Journal of Multivariate Analysis, 100(4):753 – 766.

Dutta, S., Ghosh, A. K., and Chaudhuri, P. (2011).
Some intriguing properties of tukey's half-space depth.
Bernoulli, 17(4):1420–1434.

Fraiman, R. and Muniz, G. (2001).
Trimmed means for functional data.
Test, 10(2):419–440.

Gao, Y., Shang, H. L., and Yang, Y. (2019).
High-dimensional functional time series forecasting: An application to age-specific mortality rates.
Journal of Multivariate Analysis, 170:232 – 243.
Special Issue on Functional Data Analysis and Related Topics.

Hubert, M., Rousseeuw, P. J., and Segaert, P. (2005).
Multivariate functional outlier detection.
Statistical Methods Applications, 24(1):177–202.

Hubert, M. and Vandervieren, E. (2008).
An adjusted boxplot for skewed distributions.
Computational Statistics & Data Analysis, 52(12):5186–5201.

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

López-Pintado, S. and Romo, J. (2011).
A half-region depth for functional data.
Computational Statistics & Data Analysis, 55(4):1679–1695.

López-Pintado, S., Sun, Y., Lin, J. K., and Genton, M. G. (2014).
Simplicial band depth for multivariate functional data.
Advances in Data Analysis and Classification, 8(3):321–338.

Mosler, K. (2013).
Depth statistics.
In *Robustness and complex data structures*, pages 17–34. Springer, Heidelberg.

Mosler, K. and Polyakova, Y. (2012).
General notions of depth for functional data.
arXiv:1208.1981.

Nagy, S., Gijbels, I., and Hlubinka, D. (2017).
Depth-based recognition of shape outlying functions.
Journal of Computational and Graphical Statistics, 26(4):883–893.

Nagy, S., Gijbels, I., Omelka, M., and Hlubinka, D. (2016).
Integrated depth for functional data: statistical properties and consistency.
ESAIM. Probability and Statistics, 20.

Nagy, S., Helander, S., Bever, G. V., Viitasaari, L., and Ilmonen, P. (2021).
Flexible integrated functional depths.
Bernoulli, 27(1):673 – 701.

Nieto-Reyes, A. and Battey, H. (2016).
A topologically valid definition of depth for functional data.
Statistical Science, 31(1):61–79.

Sguera, C., Galeano, P., and Lillo, R. (2014).
Spatial depth-based classification for functional data.
TEST, 23(4):725–750.

References v

Tukey, J. W. (1975).
Mathematics and the picturing of data.
In *Proceedings of the International Congress of Mathematics (Vancouver, 1974)*, volume 2, pages 523–531.

Zamani, A., Haghbin, H., Hashemi, M., and Hyndman, R. J. (2021).
Seasonal functional autoregressive models.
Journal of Time Series Analysis, n/a(n/a).