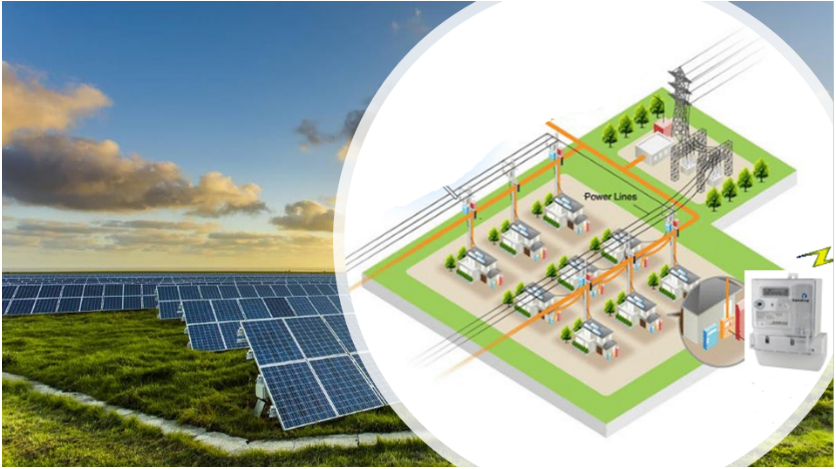
Depth-based Outlier Detection for Grouped Smart Meters: a Functional Data Analysis Toolbox

Antonio Elías, Juan Miguel Morales and Salvador Pineda


Power Lines

Outline

Depth-based Outlier Detection for Grouped Smart Meters:
a **Functional Data Analysis** Toolbox

1. Materials and Methods

   What is **Functional Data Analysis?**

   What is **Depth-based**?

2. Outlier detection toolbox

   Magnitude

   Shape

   Evolution

3. Numerical results

4. Conclusions

Materials and Methods

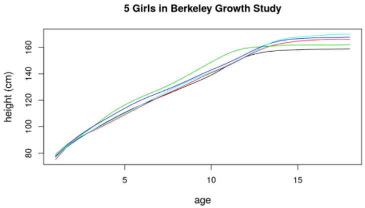
What is Functional Data Analysis?

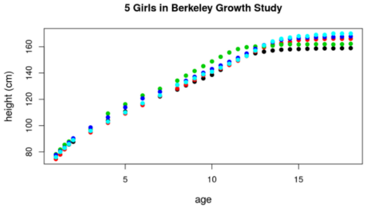
Functional Data Analysis (FDA) focuses on the analysis of smooth curves:

$$y_t(x), \quad x \in [a, b], \quad t = 1, \ldots, n$$

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

$$y_t(\tilde{x}_j), \quad \tilde{x} \in [a, b], \quad t = 1, \ldots, T \quad j = 1, \ldots, p.$$





- 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.

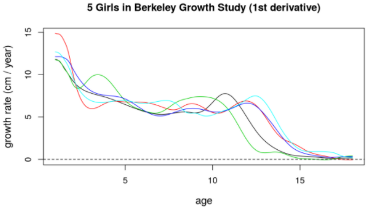
What is Functional Data Analysis?

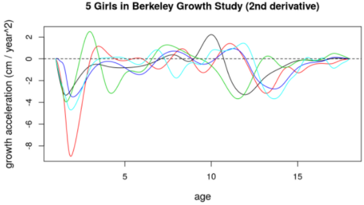
FDA propose to express discrete data by means of a basis expansion,

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

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

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

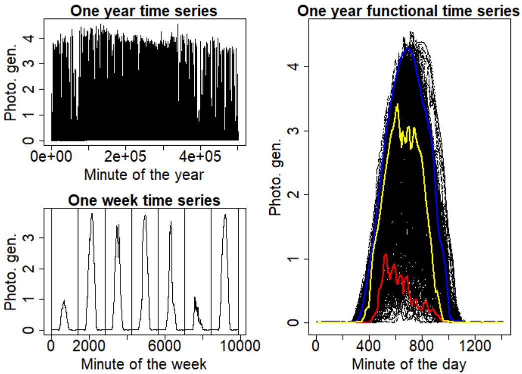




Ramsay, J. O. and Silverman, B. W. (2005). Functional Data Analysis, Springer series in Statistics.

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

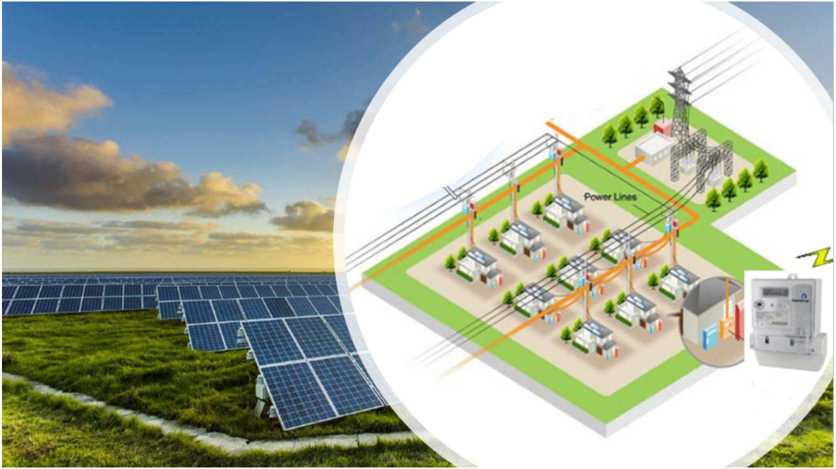
From one meter to FDA and Functional Time Series



Being again t be the index of days,

$$y_1(\tilde{x}), \ldots, y_T(\tilde{x}), \quad \tilde{x} = \{0, 1, \ldots, 1440\}, \quad t = \{1, \ldots, 365\}.$$

From multiple meters to High Dimensional Functional Time Series



From multiple meters to High Dimensional Functional Time Series

Let $i = 1, \dots, N$ be the index of the meters. Then, a sample of high dimensional functions takes the following form:

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

We denote by $y_{\cdot}^i(x)$ the i-th column of $\mathbf{y}(x)$. This is the FTS of one smart meter that is composed of $y_1^i(x), \dots, y_T^i(x)$ daily functions.

In contrast, $y_t^{\cdot}(x)$ is the t-th row corresponding to N functions for a given day t. This is a sample of N daily functions $y_t^1(x), \dots, y_t^N(x)$ where each represents a meter.

Univariate data:

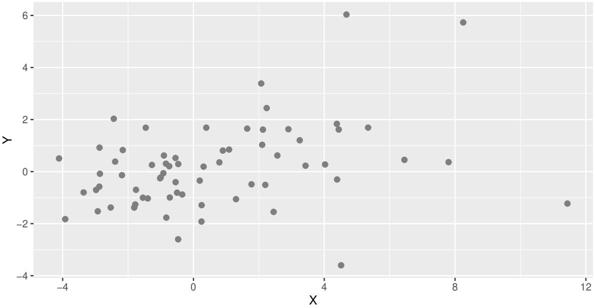
$$X : \Omega \to \mathbb{R}$$

Univariate data:

$$X : \Omega \to \mathbb{R}$$

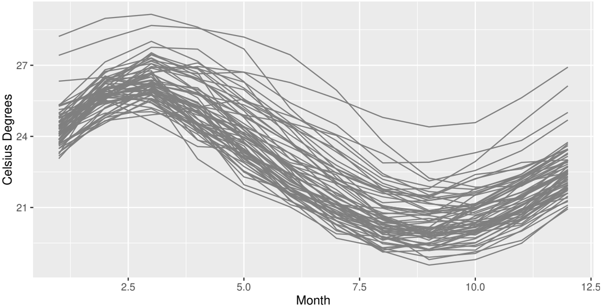
Bivariate data:

$$X = (X_1, X_2)^T : \Omega \to \mathbb{R}^2$$



Functional data:

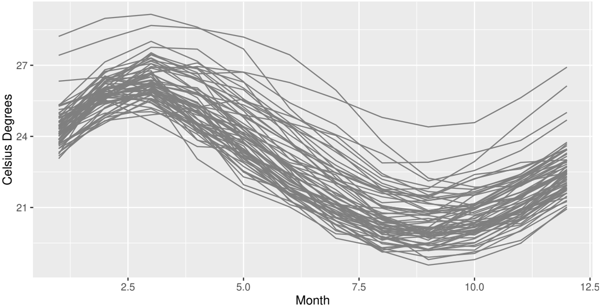
$$X(t) : t \in [a, b] \in \mathcal{C}[(a, b)]$$



What is Depth-based

Functional data:

$$X(t) : t \in [a, b] \in \mathcal{C}[(a, b)]$$



It is not straightforward! $\rightarrow$ **Depth Measures**

What is Depth-based

- It is a statistic that provides a real number for each datum that assigns its level of centrality.

$$\{\mathrm{FD}(y_t, P_T) = \mathrm{FD}(t), \quad t \in (1, \dots, T)\}.$$

- Depth-based is said to refer to a method that uses **depth measure**.

Romo, J. and Lopez-Pintado S. (2009). "On the concept of depth for functional data", JASA.

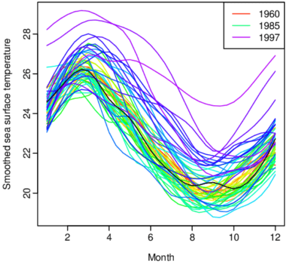
Gijbels, I. and Nagy, S. (2017). "On a general definition of depth for functional data", Statist. Sci..

What is Depth-based

- It is a statistic that provides a real number for each datum that assigns its level of centrality.

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

- Depth-based is said to refer to a method that uses **depth measure**.



- We use the Modified Band Depth (MBD).

Romo, J. and Lopez-Pintado S. (2009). "On the concept of depth for functional data", JASA.

Gijbels, I. and Nagy, S. (2017). "On a general definition of depth for functional data", Statist. Sci..

Outlier detection toolbox

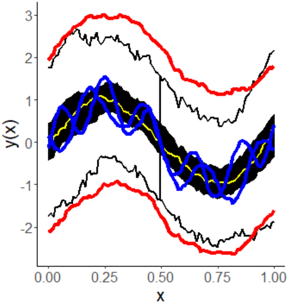
Outlier detection toolbox

We propose a toolbox based on the concept of depth to highlight:

- **Magnitude outliers**: the feature of analysis for the meter is higher or lower than the majority.
- **Shape outliers**: the shape of the profile for the meter is different than the majority.
- **Evolution outliers**: the day-to-day dependency for the meter is different for than the majority.

Magnitude Outlier

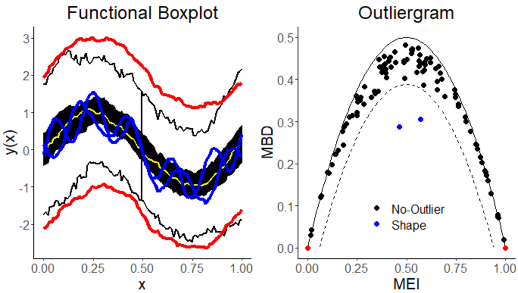
- **Functional Boxplot**. Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational & Graphical Statistics*, 20(2):316–334.



- In the context of smart meters, we apply this method for every single day, where we have one curve for each meter.
- Then we determine as **magnitude outlier** those meters that have been unmasked as magnitude outlier the majority of the days.

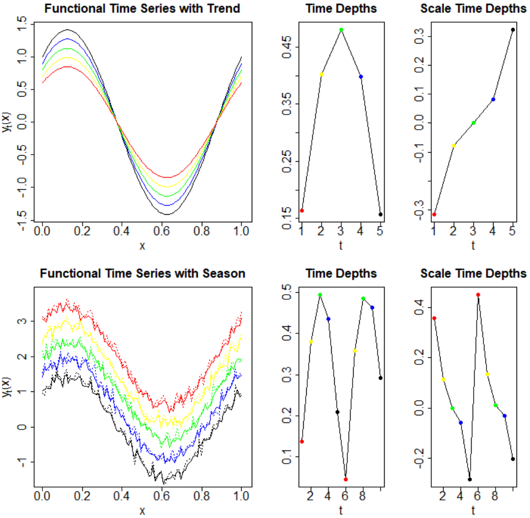
Shape Outlier

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



- In the context of smart meters, we apply this method for every single day, where we have one curve for each meter.
- Then we determine as **shape outlier** those meters that have been unmasked as shape outlier the majority of the days.

Evolution Outlier



Evolution Outlier

- Lets define the time depths of each meter as,

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

- **Prototype evolution** is a trimmed mean,

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

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

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

- Right-tailed **outlier detection rule** for skewed distributions,

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

Numerical results

Numerical results

- Pecan street data set.
- One-minute records of smart meters from Austin over one year.
- Voltage circuit and photo-voltaic energy generation.

		Zero derivative				First derivative			
	Meter id	**M**	**S**	**E**	$\widetilde{\mathbf{E}}$	**M**	**S**	**E**	$\widetilde{\mathbf{E}}$
Voltage	vol_{5746}			✓		✓			
	vol_{6139}			✓		✓			
	vol_{7901}			✓	✓	✓			
	vol_{9019}			✓					
	vol_{9922}					✓			
	vol_{7951}					✓			
Solar	sol_{9019}							✓	
	sol_{6139}							✓	
	sol_{3538}								✓

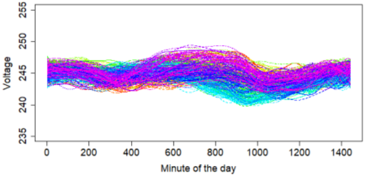
Key learnings

1 Evolution outliers are not detected by other methods.

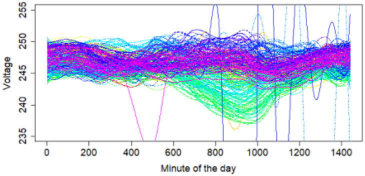
2 First derivatives allow detecting those outliers not detected with level data.

3 Scaled depths unmask those outliers which are not detected with regular depths.

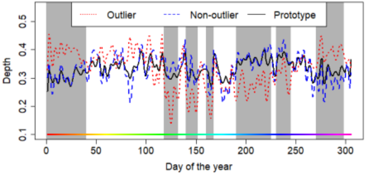
Household voltage circuit



(a) Non-outlier vol_{2818}

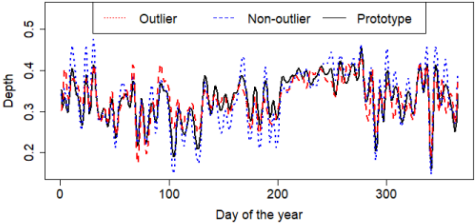


(b) Outlier vol_{9019}

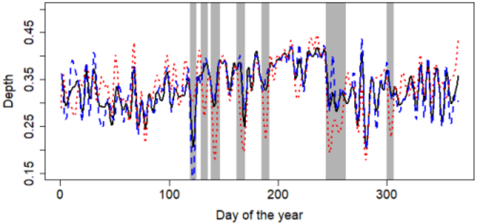


(c) Time series of depths

Household photo-voltaic energy generation

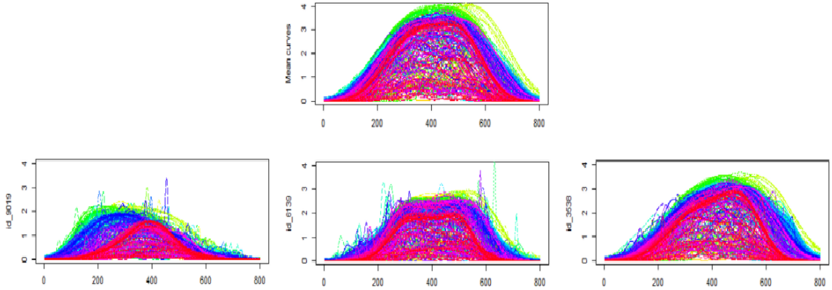


(a) Time series of depths computed on the zero derivative.



(b) Time series of depths computed on the first derivative.

Household photo-voltaic energy generation



Conclusions

Conclusions

- We propose a toolbox to highlight three types of outliers.
- Daily evolution is an important concern for outlier detection.
- 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`.

OASYS
Optimization and Analytics for Sustainable energY Systems

European Union
European Social Fund
Investing in your future

Junta de Andalucía

Depth-based Outlier Detection for Grouped Smart Meters: a Functional Data Analysis Toolbox

Antonio Elías, Juan Miguel Morales and Salvador Pineda