Functional Depths in the Context of Functional Time Series

Antonio Elías, Juan Miguel Morales and Salvador Pineda

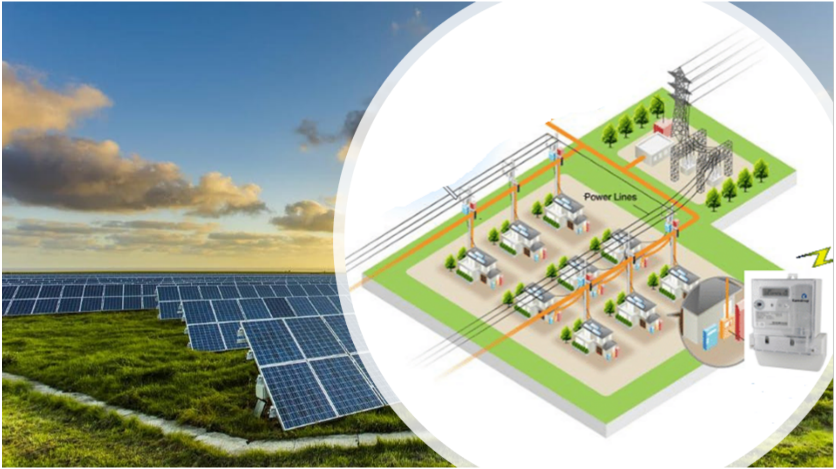
SEIO 2022

Outline

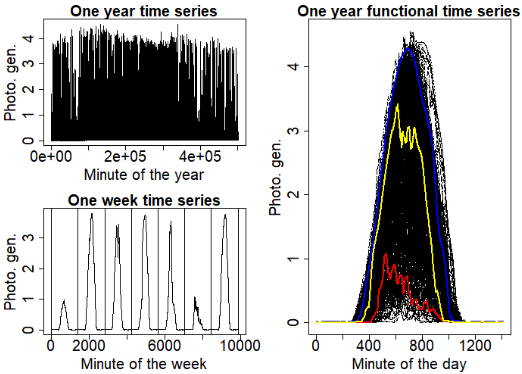
Functional Depths in the Context of Functional Time Series

1. Functional Time Series and Functional Depth Measures
2. Research lines
   - Independency Test for FTS
   - Evolution outliers for grouped HDFTS
3. Conclusions

Functional Time Series and Functional Depth Measures


Power Lines

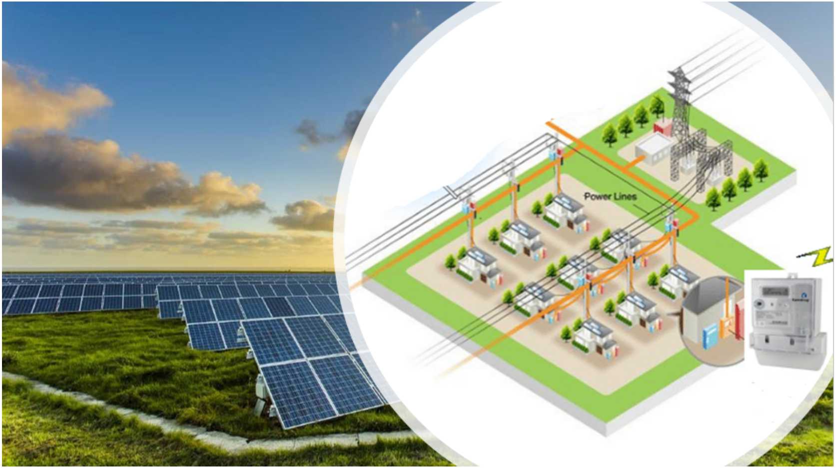
From one meter to FDA and Functional Time Series



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



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 $i = 1, \ldots, N$,

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

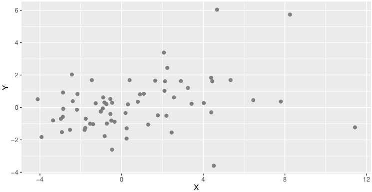
Hörmann, S. and Kokoszka, Piotr P. (2012) "Functional Time Series". Handbook of Statistics. Netherlands: Elsevier B.V., 2012, vol. 30, pp. 157–186.

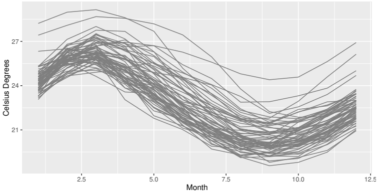
Gao, Y., Shang, HL. and Yang, Y. (2011) "High-dimensional functional time series forecasting: An application to age-specific mortality rates", *Journal of Multivariate Statistics*, 170:232-243.

Depth Measures

Depth Measures

This simple problem becomes harder in other spaces





It is not straightforward → **Depth Measures**

Functional Depth Measures

- Concept of **depth measure** [Tukey, 1975, Zuo and Serfling, 2000] and first **functional depth measure** [Fraiman and Muniz, 2001].

Elías, A. Jímenez, R., Paganoni, A. P. and Sangalli, L. M. (2022) "Integrated depths for partially observed functional data", *Journal of Computational & Graphical Statistics*.

Functional Depth Measures

- Concept of **depth measure** [Tukey, 1975, Zuo and Serfling, 2000] and first **functional depth measure** [Fraiman and Muniz, 2001].

- Given a datum $y \in \Omega$ from a distribution $P \in \mathcal{P}(\Omega)$, a depth is a bounded and non-negative function

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

Elías, A. Jímenez, R., Paganoni, A. P. and Sangalli, L. M. (2022) "Integrated depths for partially observed functional data", *Journal of Computational & Graphical Statistics*.

Functional Depth Measures

- Concept of **depth measure** [Tukey, 1975, Zuo and Serfling, 2000] and first **functional depth measure** [Fraiman and Muniz, 2001].

- Given a datum $y \in \Omega$ from a distribution $P \in \mathcal{P}(\Omega)$, a depth is a bounded and non-negative function

$$\begin{array}{rcl} \mathrm{FD} : \Omega \times \mathcal{P}(\Omega) & \rightarrow & [0, 1] \\ (y, P) & \mapsto & \mathrm{FD}(y; P). \end{array}$$

- Many different definitions of functional depth [Gijbels and Nagy, 2017].
  - Integrated depth measures (FMD and MBD).
  - Non integrated depth measures (EXTD and INFD).

Elías, A. Jímenez, R., Paganoni, A. P. and Sangalli, L. M. (2022) "Integrated depths for partially observed functional data", *Journal of Computational & Graphical Statistics*.

Time Series of Depths

- The sample of curves is indexed in time

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

Time Series of Depths

- 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

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

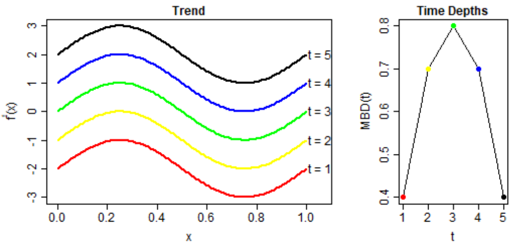
Time Series of Depths

- The sample of curves is indexed in time

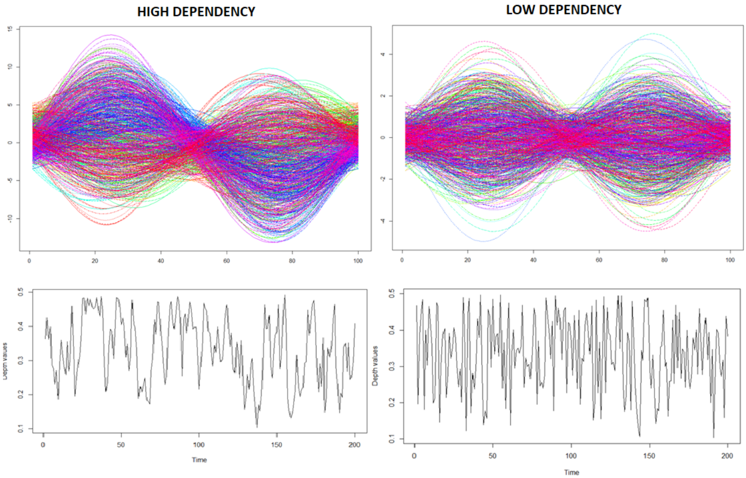
$$y^1(x), \ldots, y^T(x) \text{ for } t = 1, \ldots, T.$$

- Time series of depths

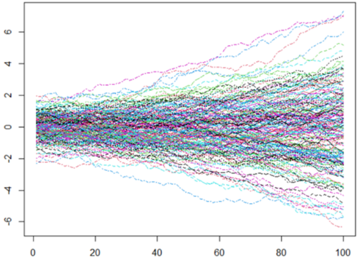
$$\{\mathrm{FD}(y^t, P_T) = \mathrm{FD}(t), \quad t \in (1, \ldots, T)\}.$$

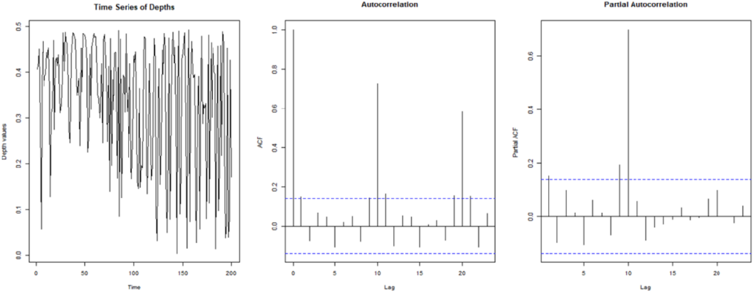


FTS from dynamic factor models [Gao et al., 2019]



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





Research lines

Independency Test for FTS

Motivation

- **Premise:** IID FDA must result in random time series of functional depths.

Gabrys, R. and Kokoszka, P. (2007), "Portmanteau test of independence for functional observations", *Journal of the American Statistical Association* 102(480), 1338–1348.

Motivation

- **Premise:** IID FDA must result in random time series of functional depths.
- Test [Ljung and Box, 1978]:

Gabrys, R. and Kokoszka, P. (2007), "Portmanteau test of independence for functional observations", *Journal of the American Statistical Association* 102(480), 1338–1348.

Motivation

- **Premise:** IID FDA must result in random time series of functional depths.
- Test [Ljung and Box, 1978]:
  H_0: The data are independently distributed (i.e. the correlations are 0, so that any observed correlations in the data result from randomness of the sampling process).

Gabrys, R. and Kokoszka, P. (2007), "Portmanteau test of independence for functional observations", *Journal of the American Statistical Association* 102(480), 1338–1348.

Motivation

- **Premise:** IID FDA must result in random time series of functional depths.
- Test [Ljung and Box, 1978]:
  - H_0: The data are independently distributed (i.e. the correlations are 0, so that any observed correlations in the data result from randomness of the sampling process).
  - H_1: The data are not independently distributed; they exhibit serial correlation.

Gabrys, R. and Kokoszka, P. (2007), "Portmanteau test of independence for functional observations", *Journal of the American Statistical Association* 102(480), 1338–1348.

Motivation

- **Premise:** IID FDA must result in random time series of functional depths.
- Test [Ljung and Box, 1978]:
  - H_0: The data are independently distributed (i.e. the correlations are 0, so that any observed correlations in the data result from randomness of the sampling process).
  - H_1: The data are not independently distributed; they exhibit serial correlation.

$$Q = T(T+2)\sum_{k=1}^{h} \frac{\hat{\rho}_k{}^2}{n-k},$$

T is the sample size, $\hat{\rho}_k{}^2$ is the sample autocorrelation at lag k, and h is the number of lags being tested.

Gabrys, R. and Kokoszka, P. (2007), "Portmanteau test of independence for functional observations", *Journal of the American Statistical Association* 102(480), 1338–1348.

Motivation

- **Premise:** IID FDA must result in random time series of functional depths.
- Test [Ljung and Box, 1978]:

  H_0: The data are independently distributed (i.e. the correlations are 0, so that any observed correlations in the data result from randomness of the sampling process).

  H_1: The data are not independently distributed; they exhibit serial correlation.

$$Q = T(T+2)\sum_{k=1}^{h} \frac{\hat{\rho}_k{}^2}{n-k},$$

  T is the sample size, $\hat{\rho}_k{}^2$ is the sample autocorrelation at lag k, and h is the number of lags being tested.

- Under the null hypothesis the statistic:

$$Q \sim \chi_h,$$

Gabrys, R. and Kokoszka, P. (2007), "Portmanteau test of independence for functional observations", *Journal of the American Statistical Association* 102(480), 1338–1348.

Motivation

- **Premise:** IID FDA must result in random time series of functional depths.
- Test [Ljung and Box, 1978]:

  H_0: The data are independently distributed (i.e. the correlations are 0, so that any observed correlations in the data result from randomness of the sampling process).

  H_1: The data are not independently distributed; they exhibit serial correlation.

$$Q = T(T+2)\sum_{k=1}^{h} \frac{\hat{\rho_k}^2}{n-k},$$

  T is the sample size, $\hat{\rho_k}^2$ is the sample autocorrelation at lag k, and h is the number of lags being tested.
- Under the null hypothesis the statistic:

$$Q \sim \chi_h,$$

- For a significant level α, the critical region for rejection is

$$Q > \chi_{1-\alpha,h}.$$

Gabrys, R. and Kokoszka, P. (2007), "Portmanteau test of independence for functional observations", *Journal of the American Statistical Association* 102(480), 1338–1348.

Empirical Size

- Size = $P(\text{Reject H}_0 | H_0 \text{ is TRUE})$
- Brownian motion, Brownian bridge and Gaussian.

	FMD			MBD			EXTD			INFD		
LAG	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
T = 50												
1	10.6	5.5	1.5	10.5	5.8	1.2	11.5	6.5	0.7	10.1	5.1	1
3	9.7	5	1.6	9.6	4.2	1	10.9	5.1	0.9	11.2	5.4	1.1
5	9.7	5.3	1.7	9.3	4.9	1.6	10.4	5.8	1.6	9.7	4.2	1
T = 100												
1	9.5	5	0.6	9.7	4.3	0.4	11.2	5.7	0.6	7.5	3.5	0.5
3	9.2	4.3	0.8	8.1	4.3	0.5	11.9	4.6	0.8	9	4.4	0.8
5	8.8	4.8	1.2	9	4.9	1	10.1	5.3	0.7	9.3	4.4	1.2
T = 300												
1	9.3	4.2	0.7	8.4	4.1	1.1	11	5.4	1.3	9.1	4.6	0.6
3	7.5	3.7	0.8	7.9	3.6	0.7	10.9	6.1	1.3	9.7	4.9	0.9
5	8.3	3.6	1.1	7.9	4.7	1	11.6	6	1.3	10.5	5.4	0.9

Table 1: Size BM

Empirical Power

- Power= $P(Reject \mathrm{H}_0 | \mathrm{H}_1 \text{ is TRUE})$
- Scalar and integral operator FAR(1), FARMA(1,1), FMA(q), SFAR(8).

	FMD			MBD			EXTD			INFD		
LAG	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
T = 50												
1	44.9	32.7	13.8	47.3	37.1	16.8	63.4	49.9	24.3	26.1	17.6	5.9
3	32.5	22.2	8.5	34.6	23.9	9.9	48	35.1	14.9	19	10.9	3.4
5	30.3	20.6	8.2	32.8	22.4	8.9	42.3	30.2	14.3	16.7	10.5	2.9
T = 100												
1	74.7	64	39.1	78.6	68.3	46.1	91.6	86.8	66.7	42.2	31.7	16.7
3	59.6	47	27.5	65.2	53.2	31.5	82.7	71.6	48.6	31.5	22.3	10
5	53	40.1	22.7	58.1	45.7	27.2	74.1	62.1	40.5	28.2	19.2	8.3
T = 300												
1	99.3	98.4	95.4	99.7	99.2	97	100	100	99.9	84.6	79	60.5
3	97.5	96.3	88.8	98.5	97.2	92.8	99.9	99.9	99.8	74.3	67	46.6
5	96	93.8	83.2	97.4	95.8	89.6	99.9	99.9	99.2	70.2	58.2	37.6

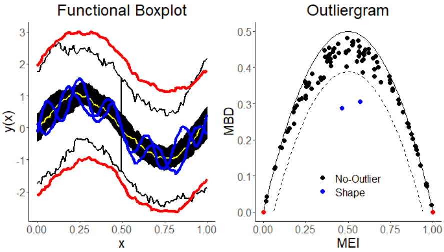
Table 2: Power FAR integral operator.

Evolution outliers for grouped HDFTS

Outliers in the context of Functional Data Analysis

Taxonomy of outliers: [Hubert et al., 2015, Ojo et al., 2021]:

Magnitude and shape.

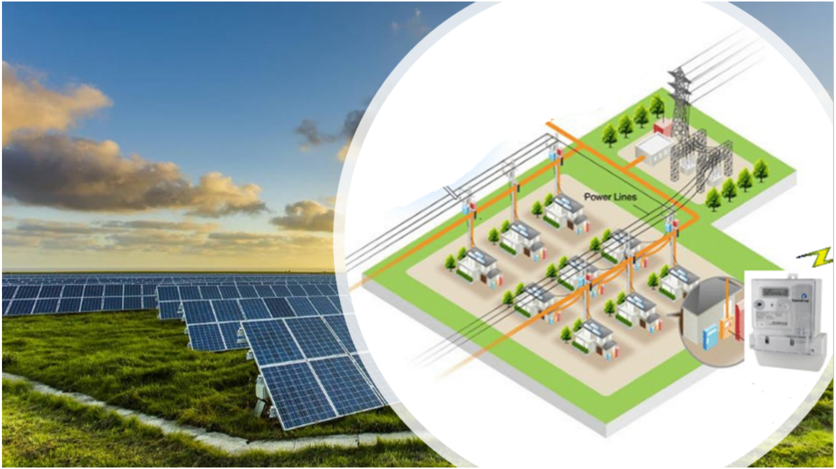


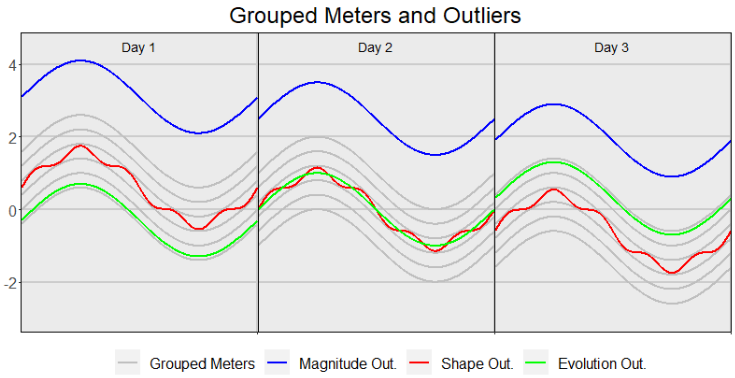
Temporal dependency.

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.

P. Raña, G. Aneiros, and J. M. Vilar, "Detection of outliers in functional time series," Environmetrics, vol. 26, no. 3, pp. 178–191, 2015.


Power Lines


Grouped Meters and Outliers
Day 1
Day 2
Day 3
4
2
0
-2
Grouped Meters
Magnitude Out.
Shape Out.
Evolution Out.

Evolution outlier detection for grouped HDFTS

- Time depths of each FTS as,

$$\mathbf{FD}(t) = [\mathrm{FD}_1(t), \mathrm{FD}_2(t), \dots, \mathrm{FD}_N(t)],$$

Evolution outlier detection for grouped HDFTS

- 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],

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

Evolution outlier detection for grouped HDFTS

- 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.,

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

Evolution outlier detection for grouped HDFTS

- 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.,

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

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

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

Simulation results

- Simulate N Grouped FTS, i.e., same evolution.

Simulation results

- Simulate N Grouped FTS, i.e., same evolution.
- Simulate outlying FTS modifying the group evolution. Two models.

Simulation results

- Simulate N Grouped FTS, i.e., same evolution.
- Simulate outlying FTS modifying the group evolution. Two models.
- Benchmarks from FDA and non-FDA literature.

Simulation results

- Simulate N Grouped FTS, i.e., same evolution.
- Simulate outlying FTS modifying the group evolution. Two models.
- Benchmarks from FDA and non-FDA literature.

Table 3: Simulation results for Model 1.

Outliers		1%		5%		10%	
		TPR	**TNR**	**TPR**	**TNR**	**TPR**	**TNR**
KNN		0.000	0.899	0.010	0.903	0.014	0.900
AKNN		0.000	0.930	0.000	0.9316	0.003	0.924
LOF		0.250	0.884	0.198	0.8914	0.197	0.892
COF		0.180	0.950	0.102	0.9477	0.092	0.961
INFLO		0.240	0.961	0.194	0.960	0.181	0.963
ONESVM		0.510	0.582	0.392	0.5772	0.398	0.590
FEA		0.450	0.506	0.522	0.5083	0.248	0.478
PCA		0.450	0.506	0.522	0.5083	0.248	0.478
TDEPTH							
	MBD	1.000	1.000	1.000	1.000	1.000	1.000
	FMD	1.000	1.000	1.000	1.000	1.000	1.000
	EXTD	1.000	1.000	1.000	1.000	1.000	1.000
	INFD	1.000	0.936	1.000	0.978	1.000	0.987
STDEPTH							
	MBD	1.000	1.000	1.000	1.000	1.000	1.000
	FMD	1.000	1.000	1.000	1.000	1.000	1.000
	EXTD	1.000	1.000	1.000	1.000	1.000	1.000
	INFD	1.000	0.950	1.000	0.976	1.000	0.986
FBOX		0.000	0.988	0.000	0.986	0.000	0.986
OUTGRAM		0.000	0.953	0.000	0.952	0.005	0.943
GEOM							
	AMP	0.000	1.000	0.000	1.000	0.000	1.000
	PHASE	0.000	1.000	0.000	1.000	0.000	1.000
ELASTIC							
	AMP	0.000	1.000	0.000	1.000	0.000	1.000
	PHASE	0.000	1.000	0.000	1.000	0.000	1.000

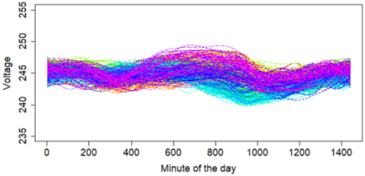
Real data results

- Pecan street data set.
- One-minute records of smart meters from Austin over one year. 25 households.
- Voltage circuit and photo-voltaic energy generation.

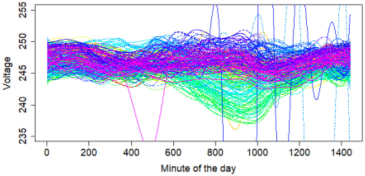
		Zero derivative				First derivative			
	Meter id	**M**	**S**	**E**	$\widetilde{\mathbf{E}}$	**M**	**S**	**E**	$\widetilde{\mathbf{E}}$
Voltage	vol5746			✓		✓			
	vol6139			✓		✓			
	vol7901			✓	✓	✓			
	vol9019			✓					
	vol9922					✓			
	vol7951					✓			
Solar	sol9019							✓	
	sol6139							✓	
	sol3538								✓

Key learnings

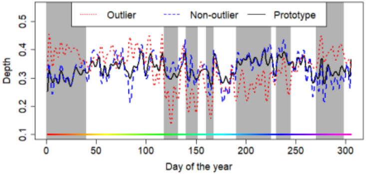
1. Evolution outliers are not detected by other methods.
2. First derivatives allow detecting those outliers not detected with level data.
3. Scaled depths unmask those outliers which are not detected with regular depths.
4. Computational efficiency of Integrated Depth Measures.



(a) Non-outlier vol_{2818}

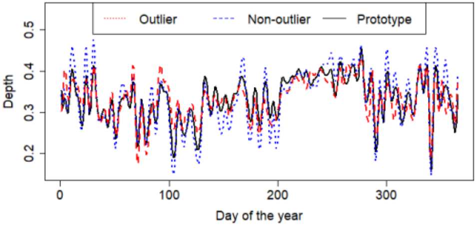


(b) Outlier vol_{9019}

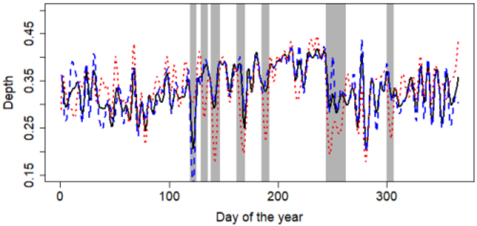


(c) Time series of depths

Case study: 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.

Conclusions

Conclusions

- Time series of functional depths to cope with time dependent functional data.

- Two applications:
  - Independency test for FTS.
  - Evolution outlier detection in grouped HDFTS.

- Further research:
  - Distribution of the test statistics under the null.
  - Links between the temporal structure of the FTS and the time series of depths.

- Visit our website for more details `oasys.uma.es`

  "A FDA Approach to Evolution Outlier Mining for Grouped Smart Meters".

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

- Contact: `aelias@uma.es`.

Functional Depths in the Context of Functional Time Series

Antonio Elías, Juan Miguel Morales and Salvador Pineda

SEIO 2022

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.

Gijbels, I. and Nagy, S. (2017).
On a general definition of depth for functional data.
Statist. Sci., 32(4):630–639.

Hubert, M., Rousseeuw, P. J., and Segaert, P. (2015).
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.

Ljung, G. M. and Box, G. E. P. (1978).
On a measure of lack of fit in time series models.
Biometrika, 65(2):297–303.

Ojo, O. T., Fernández Anta, A., Lillo, R. E., and Sguera, C. (2021).
Detecting and classifying outliers in big functional data.
Advances in data analysis and classification.

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).

Zuo, Y. and Serfling, R. (2000).
General notions of statistical depth function.
The Annals of Statistics, 28(2):461–482.