Outlier Detection in Functional Time Series with Applications to the Electricity Market

Paula Raña[*], Juan Vilar and Germán Aneiros
Departamento de Matemáticas, Universidade da Coruña

Outlier detection is an important task that has been largely studied in the statistical literature. Two main approaches have been developed in order to deal with inaccurate conclusions stemming from the presence of outliers. The first of them consists of developing robust methods, that are not affected by outliers. The second one consists of detecting outliers before apply any statistical methodology to the data.

In this paper, an outlier is defined as an observation (that may be scalar, vector o function observed over a continuum) that has been generated by a stochastic process with a distribution different from the vast majority of the remaining observations, which are assumed to be identically distributed.

In the context of functional data, outlier detection is not an easy task. On the one hand, the whole set of functional data (curves, images or functions) is not always possible to visualize. On the other hand, there are different kind of functional outliers: "magnitude outliers" (that arise when they lie outside the range of the vast majority of the data) and "shape outliers" (that are within the range of the rest of the data but differ in shape from them). Also a combination of both can give us another kind of functional outlier.

The aim of this work is to develop methods to detect any kind of outlier in functional time series. We propose three methods. The first one is a depth-based method that adapts the procedure of Febrero et al. (2008) to the functional time series setting, so that it takes the dynamics in the time series into account. Into this first proposal, one can choose also different options, depending on the functional depth or the bootstrap method (see Raña, Aneiros and Vilar (2015)). The other two main approaches are based on robust functional principal component analysis, following the ideas of Hyndman and Ullah (2007) to forecast functional time

[*]Corresponding author. E-mail: paula.rana@udc.es

series that we use to work with projections or with prediction errors in two different procedures.

A simulation study shows the performance of the proposed procedures and compare them with some methods in the statistical literature related to independent functional data. Finally, the procedures are applied on two real datasets from the Spanish Electricity Market, daily curves of electricity demand and price. Results conclude that dependence in the data must be taken into account to detect outliers in functional time series.

State of the art

Procedures for detecting functional outliers have been proposed over recent years. The first papers that have addressed outliers identification in the context of functional data are Hyndman and Ullah (2007) and Febrero, Galeano and González-Manteiga (2007, 2008).

Hyndman and Ullah (2007) proposed a method for robust estimation of functional principal components, which is the basis of their methodology for forecasting functional time series. As a by-product, they constructed a method for detecting outliers based on the integrated squared error (ISE) between each functional datum and its projection into a given number of robust principal components. The procedure in Febrero, Galeano, and González-Manteiga (2007) (Febrero, Galeano, and González-Manteiga (2008)) performs a depth-based test statistic for each curve, where the critical value is obtained with a bootstrap method after trimming the sample. We refer to this method as Depth-based Trimming (DBT).

Several procedures for detecting outliers in functional data have been proposed from these works. They are generally based on functional principal components analysis (Hyndman and Shang 2010; Sawant, Billor, and Shin 2012; Yu, Zou, and Wang 2012), functional depths (Sun and Genton 2011; Gervini 2012; Arribas-Gil and Romo 2014) or random projections (Fraiman and Svarc 2013).

Contributions

The previous mentioned papers deal with independent functional data. Nevertheless, some of them present applications to functional time series. This could lead to misleading conclusions. In fact, as clearly illustrated in Figure 1, the concept of magnitude outlier given earlier should be extended when applied to functional time series. Thus, it is necessary to develop approaches to detect outliers that take into account the features of functional time series. Looking at the Figure 1 one may suspect the possible presence of such three outliers; however, the same may not be said when observing the right panel in Figure 1. Local trends induced from the dependence structure could mask the presence of outliers; so, in functional time series, an observation could be an outlier despite being inside the range of the vast majority of the data.

Sun and Genton (2012) constructed the Adjusted Functional Boxplot (AFB) by extending to the setting of spatio-temporal data the Functional Boxplot (FBox) they had previously proposed (Sun and Genton (2011)). This procedure is a fully functional method, based on depth, for outlier detection in spatio-temporal data. However, it is not designed to identify the magnitude outliers introduced here.

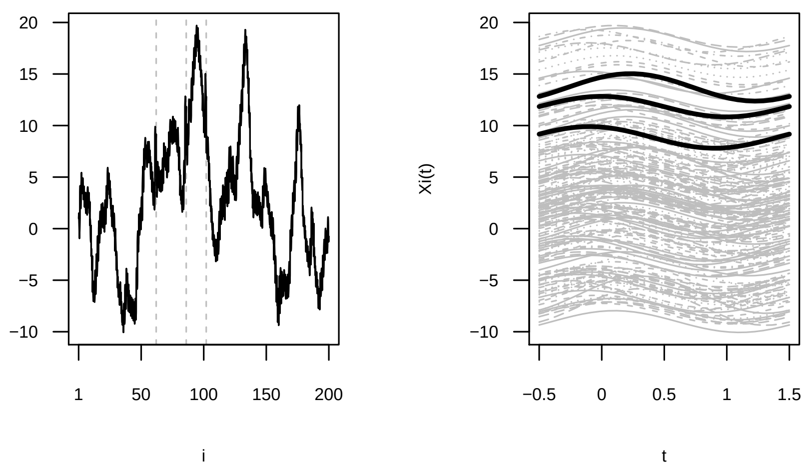

Figure 1: Left panel: functional time series (i denotes the temporal index) contaminated with three outliers; the vertical dashed lines indicate the positions where the outliers emerged. Right panel: the corresponding curves $X_i(t)$ (the black curves are the outliers).

As far as we know, the methods proposed in this study are the firsts dealing with the problem of outlier detection specifically addressed for functional time series. The importance of taking the dependence in the data into account, to avoid possible errors when methods for independent data are applied on functional time series, has been proved in the simulation study. First part of the presented work, involving depth-based method, is already published in *Environmetrics* (see Raña, Aneiros and Vilar (2015)).

Acknowledgments

This research was partially supported by Grants MTM2011–22392 and CN2012/130 from Spanish Ministerio de Economía y Competitividad and Xunta de Galicia, respectively.

References

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

[2] Febrero, M., Galeano, P., and González-Manteiga, W. (2007), "Functional analysis of NOx levels: location and scale estimation and outlier detection, *Computational Statistics*, 22, 411–427.

[3] Febrero, M., Galeano, P., and González-Manteiga, W. (2008), "Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels," *Environmetrics*, 19, 331–345.

[4] Fraiman, R., and Svarc, M. (2013), "Resistant estimates for high dimensional and functional data based on random projections," *Computational Statistics and Data Analysis*, 58, 326–338.

[5] Gervini, D. (2012), "Outlier detection and trimmed estimation for general functional data," *Statistica Sinica*, 22, 1639–1660.

[6] Hyndman, R.J., and Shang, H.L. (2010), "Rainbow plots, bagplots, and boxplots for functional data," *Journal of Computational and Graphical Statistics*, 19, 29–45.

[7] Hyndman, R.J., and Ullah, M.S. (2007), "Robust forecasting of mortality and fertility rates: A functional data approach," *Computational Statistics and Data Analysis* 51, 4942–4956.

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

[9] Sawant, P., Billor, N., and Shin, H. (2012), "Functional outlier detection with robust functional principal component analysis," *Computational Statistics*, 27, 83–102.

[10] Sun, Y., and Genton, M.G. (2011), "Functional boxplots," *Journal of Computational and Graphical Statistics*, 20, 316–334.

[11] Sun Y, Genton G. 2012. Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *Environmetrics* **23**: 54–64.

[12] Yu, G., Zou, C., and Wang, Z. (2012), "Outlier detection in functional observations with applications to profile monitoring," *Technometrics*, 54, 308–318.