

# GENERALIZED EXPONENTIAL SMOOTHING IN PREDICTION OF HIERARCHICAL TIME SERIES

Daniel Kosiorowski<sup>1</sup>, Dominik Mielczarek<sup>2</sup>,  
Jerzy P. Rydlewski<sup>3</sup>, Małgorzata Snarska<sup>4</sup>

## ABSTRACT

Shang and Hyndman (2017) proposed a grouped functional time series forecasting approach as a combination of individual forecasts obtained using the generalized least squares method. We modify their methodology using a generalized exponential smoothing technique for the most disaggregated functional time series in order to obtain a more robust predictor. We discuss some properties of our proposals based on the results obtained via simulation studies and analysis of real data related to the prediction of demand for electricity in Australia in 2016.

**Key words:** functional time series, hierarchical time series, forecast reconciliation, depth for functional data.

## 1. Introduction

The problem of optimal reconciliation of forecasts of complex economic phenomena partitioned into certain groups and/or levels of hierarchy has been considered in the economic and econometric literature many times and is still present in a public economic debate (see Kohn (1982), Weale (1988), Hyndman et al. (2011)). National import/export quantities or Gross National Product balances are important examples here. Discrepancies between forecasts prepared at the global level and obtained by aggregating regional forecasts or forecasts prepared according to certain hierarchy levels are usually thought to be caused by different methodologies or different precision of measurements used at different hierarchy levels or "prediction clusters". The issue is also very important from a particular company's point of view in a context of a product or a service lines management, consumers portfolio optimization and consumers segmentation. Let us take the equipment for running grouped in levels of hierarchy with respect to age, sex, competitive or re-creative usage and season designation as an example of material product line management. Let us take demand and supply of electricity within day and night optimized with respect to forecasted day and night demand, customers grouped with respect to

<sup>1</sup>Department of Statistics, Cracow University of Economics, Kraków, Poland.

E-mail: daniel.kosiorowski@uek.krakow.pl.

<sup>2</sup>AGH University of Science and Technology, Faculty of Applied Mathematics, al. A. Mickiewicza 30, 30-059 Krakow, Poland. E-mail: dmielcza@wms.mat.agh.edu.pl.

<sup>3</sup>AGH University of Science and Technology, Faculty of Applied Mathematics, al. A. Mickiewicza 30, 30-059 Krakow, Poland. E-mail: ry@agh.edu.pl.

<sup>4</sup>Department of Financial Markets, Cracow University of Economics, Kraków, Poland.

E-mail: malgorzata.snarska@uek.krakow.pl.

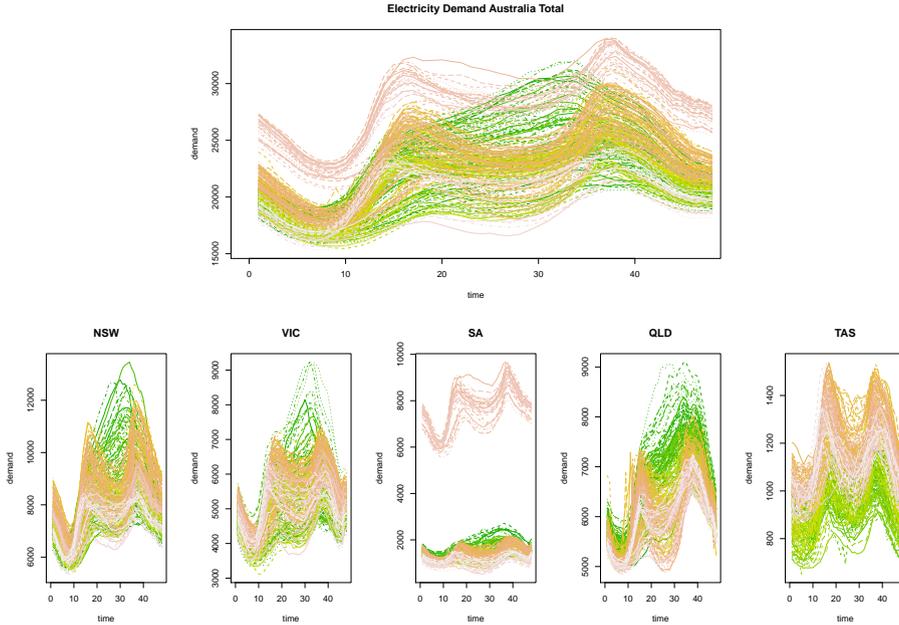


Figure 1: Electricity demand in regions of Australia in 2016 – hierarchical functional time series example

regions of living or residing and a degree of "consumer priority" as examples of im-material product sales optimization. Let us take the Internet holiday booking service divided into sub-services with respect to certain wealth or "an inclination to adventures" criterion as an example of a service management. In recent years a very interesting statistical methodology named functional data analysis (FDA) for analyzing functional data has been developed (see Bosq (2000), Ramsay et al. (2009), Horvath and Kokoszka (2012), Krzyśko et al. (2013), Shang and Hyndman (2017)). For applications of FDA in economics see Kosiorowski (2014), Kosiorowski (2016), Kosiorowski, Rydlewski and Snarska (2017a). Economic usefulness of outliers detection procedures in the FDA setup has been recently described and discussed in Nagy et al. (2017), Kosiorowski, Rydlewski and Zawadzki (2018a), Kosiorowski, Mielczarek and Rydlewski (2018c).

In this context, it is worth stressing, that many others economic phenomena may effectively be described by means of functions or their series (i.a. utility curves, yield curves, development paths of companies or countries).

Following the above cited authors we consider a random curve  $X = \{x(t), t \in [0, T]\}$ , where  $T$  is fixed, as a random element of the separable Hilbert space  $L^2([0, T])$  with the inner product  $\langle x, y \rangle = \int x(t)y(t)dt$ . The space is equipped with the Borel  $\sigma$ -algebra. Furthermore, in Bosq's (2000) monograph it is proved that

probability distributions do exist for functional processes with values in Hilbert space. We denote this probability distribution by  $\mathcal{F}$ . Functional time series (FTS) is a series of functions indexed by time (e.g. see Fig. 1, colors indicate time succession of sequence of the functional objects according to the base R *terrain.colors* color palette). A hierarchical functional time series is a series of functions grouped at specified levels (household, town, region, whole country), (i.e. see Fig. 1). At each level a forecast can be made. A natural problem arises: how to use information obtained at different levels to obtain a reconciliated prediction for all levels?

The problem of hierarchical time series prediction is solved with various ways. Bottom-up method relies on forecasting each of the disaggregated series at the lowest level of the hierarchy, and then using simple aggregation to obtain forecasts at a higher level of the hierarchy (see Kahn (1998)). Top-down method involves forecasting of aggregated series at the top level of the hierarchy, and then using disaggregation to obtain forecasts at a lower level of the hierarchy based on historical proportions. Shang and Hyndman (2017), extending the method of Hyndman et al. (2011), considered grouped functional time series forecasting as an optimal combination of individual forecasts using generalized least squares regression with level forecasts treated as regressors. In the context of hierarchical FTS prediction a general problem arises: which method of forecasting should be chosen (see Bosq (2000), Besse et al. (2000), Hyndman and Ullah (2007), Hyndman and Shang (2009), Aue et al. (2015), Kosiorowski, Mielczarek and Rydlewski (2017b, 2018b)). Shang and Hyndman (2017) proposed a grouped functional time series forecasting approach as a combination of individual forecasts obtained by means of their smart predicting method, in which functional time series is reduced to a family of one dimensional time series of principal component scores representing original functional series (see Kosiorowski, 2014). As a result of conducted simulation studies, we decided to modify their methodology. Instead of using principal component scores forecast methods, we decided to propose a certain functional generalization of exponential smoothing technique (see Hyndman et al. (2008) for a theoretical background of the exponential smoothing), i.e. we used moving local medians and moving local functional trimmed averages (Febrero-Bande and de la Fuente, 2012) for the most disaggregated series in order to obtain more robust predictor than Shang and Hyndman (2017). The main aim of the paper is to modify Shang and Hyndman (2017) predictor so that it could cope with functional outliers and/or it would be elastic enough to adapt to changes in data generating mechanism. The remainder of the paper is as follows: in the second section elements of depth concept for functional data are sketched and in the third section our proposals are introduced. Fourth section presents results of simulation as well as empirical studies. The paper ends with conclusions, references and a short appendix containing R script showing how to calculate forecasts using our proposals with free *DepthProc* and *fda.usc* R package (see Kosiorowski and Zawadzki, 2018; Febrero-Bande and de la Fuente, 2012).

## 2. Depths for functional data

For obtaining robust hierarchical FTS predictor we focused our attention on the functional data depth concept (Nagy et al. (2016) and Nieto-Reyes and Battey (2016)). We have chosen, in our opinion the best depth for the considered functional data, namely the corrected generalized band depth (cGBD, see López-Pintado and Jörnsten (2007)), but for computational reasons we restrict our considerations to the case of the band consisting of two functions.

If  $X_1$  and  $X_2$  are independent functional random variables generated by the functional time series, and generating the observations (real functions in the  $L^2([0, T])$  space), the cGBD of curve  $x$  with respect to  $\mathcal{F}$  is defined as

$$cGBD(x|\mathcal{F}) = \mathcal{F}(G(x) \subset A^c(x; X_1, X_2))$$

where  $G(x) = \{(t, x(t)) : t \in [0, T]\}$  is a graph of function  $x$ , and  $a_{1,2} = \{t \in [0, T] : X_2(t) - X_1(t) \geq 0\}$  and

$$A^c(x; X_1, X_2) = \{t \in a_{1,2} : X_1(t) \leq x(t) \leq X_2(t)\}, \text{ if } a_{1,2} \geq a_{2,1}$$

or

$$A^c(x; X_1, X_2) = \{t \in a_{2,1} : X_2(t) \leq x(t) \leq X_1(t)\}, \text{ if } a_{2,1} > a_{1,2}.$$

Now, let  $\mathbf{X}^N = \{x_1, \dots, x_N\}$  be a sample of continuous curves defined on the compact interval  $[0, T]$ . Let  $\lambda$  denote the Lebesgue measure and let  $a(i_1, i_2) = \{t \in [0, T] : x_{i_2}(t) - x_{i_1}(t) \geq 0\}$ , where  $x_{i_1}$  and  $x_{i_2}$  are band delimiting objects. Let  $L_{i_1, i_2} = \frac{\lambda(a(i_1, i_2))}{\lambda([0, T])}$ . The empirical cGBD of a curve  $x$  with respect to the sample  $\mathbf{X}^N$ , which estimates cGBD for curve  $x$  in the considered functional space with respect to  $\mathcal{F}$ , is defined as (see López-Pintado and Jörnsten, 2007)

$$cGBD(x|\mathbf{X}^N) = \frac{2}{N(N-1)} \sum_{1 \leq i_1 < i_2 \leq N} \frac{\lambda(A^c(x; x_{i_1}, x_{i_2}))}{\lambda([0, T])}$$

where

$$A^c(x; x_{i_1}, x_{i_2}) = \{t \in a(i_1, i_2) : x_{i_1}(t) \leq x(t) \leq x_{i_2}(t)\}, \text{ if } L_{i_1, i_2} \geq \frac{1}{2}$$

or

$$A^c(x; x_{i_1}, x_{i_2}) = \{t \in a(i_2, i_1) : x_{i_2}(t) \leq x(t) \leq x_{i_1}(t)\}, \text{ if } L_{i_2, i_1} > \frac{1}{2}.$$

Within this definition, the introduced earlier band depth (López-Pintado and Romo, 2009) is modified so that it only takes into account the proportion of the domain where the delimiting curves define a contiguous region which has non-zero width. Further in our proposals we use *the depth regions of order  $\alpha$*  for the considered cGBD, i.e.  $R_\alpha(\mathcal{F}) = \{x : cGBD(x, \mathcal{F}) \geq \alpha\}$ . Note, that  $\alpha$ -central regions  $R_\alpha(\mathcal{F}) = \{x \in L^2([0, T]) : D(x, \mathcal{F}) \geq \alpha\}$  may be defined for any statistical depth function  $D(x, \mathcal{F})$ , where  $\mathcal{F}$  denotes a probability distribution (Zuo and Serfling, 2000). Note also that various robust and nonparametric descriptive characteristics, like scatter, skew-

ness, kurtosis, may be expressed in terms of  $\alpha$ -central regions. These regions are nested and inner regions contain less and less probability mass. Following Paindaveine and Van Bever (2013), when defining local depth it will be more appropriate to index the family  $\{R_\alpha(\mathcal{F})\}$  by means of probability contents. Consequently, for any  $\beta \in (0, 1]$  we define the smallest depth region with  $\mathcal{F}$ -probability equal or larger than  $\beta$  as

$$R^\beta(\mathcal{F}) = \bigcap_{\alpha \in A(\beta)} R_\alpha(\mathcal{F}),$$

where  $A(\beta) = \{\alpha \geq 0 : P(R_\alpha(\mathcal{F})) \geq \beta\}$ . The depth regions  $R_\alpha(\mathcal{F})$  and  $R^\beta(\mathcal{F})$  provide only the deepest point neighborhood. In our considerations, we can replace probability distribution  $\mathcal{F} = \mathcal{F}^{\mathbf{X}}$  by its symmetrized version for any function  $x$ , namely  $\mathcal{F}_x = \frac{1}{2}\mathcal{F}^{\mathbf{X}} + \frac{1}{2}\mathcal{F}^{2x-\mathbf{X}}$ . For any depth function  $D(\cdot, \mathcal{F})$  the corresponding *sample local depth function at the locality level*  $\beta \in (0, 1]$  is  $LD^\beta(x, \mathcal{F}^{(N)}) = D(x, \mathcal{F}_x^{\beta(N)})$ , where  $\mathcal{F}_x^{\beta(N)}$  denotes the empirical probability distribution related to those functional observations that belong to  $R_x^\beta(\mathcal{F}^{(N)})$ , where  $\mathcal{F}^{(N)}$  denotes empirical probability distribution calculated from  $X^N$ . Thus  $R_x^\beta(\mathcal{F}^{(N)})$  is the smallest sample depth region that contains at least a proportion  $\beta$  of the  $2N$  random functions  $x_1, \dots, x_N, 2x - x_1, \dots, 2x - x_N$ . Depth is always well defined – it is an affine invariant from original depth. For  $\beta = 1$  we obtain global depth, while for  $\beta \simeq 0$  we obtain extreme localization. As in the population case, our sample local depth will require considering, for any  $x \in L^2([0, T])$ , the symmetrized distribution  $\mathcal{F}_x^{(N)}$ , which is empirical distribution associated with  $x_1, \dots, x_N, 2x - x_1, \dots, 2x - x_N$ . Sample properties of the (global) depths result from general findings presented in Zuo and Serfling (2000). Implementations of local versions of several depths including projection depth, Student, simplicial,  $L^p$  depth, regression depth and modified band depth can be found in free R package *DepthProc* (see Kosiorowski and Zawadzki, 2018). In order to choose the locality parameter  $\beta$  we recommend using expert knowledge related to the number of components or regimes in the considered data. Sample properties of the local versions of depths result from general findings presented in Paindaveine and Van Bever (2013). For other concepts of local depths see, e.g., Sguera et al. (2016).

### 3. Our proposals

We consider a sample of  $N$  functions  $\mathbf{X}^N = \{x_i(t) : t \in [0, T], i = 1, \dots, N\}$ . Let  $FD^\beta(y|\mathbf{X}^N)$  denote sample functional depth of  $y(t)$  with locality parameter  $\beta$ , e.g., the functional depth is equal to corrected generalized band depth:  $FD = cGBD$ . Sample  $\beta$ -local median is defined as

$$MED_{FD^\beta}(\mathbf{X}^N) = \arg \max_{i=1, \dots, N} FD^\beta(x_i|\mathbf{X}^N).$$

Assume now, that we observe a functional time series  $x_i^{cluster}(t)$ ,  $i = 1, 2, \dots$  for each considered level of hierarchy. We put our proposals forward.

PROPOSAL 1: We independently apply on each considered level of hierarchy and in each cluster of that level a moving median predictor to obtain a forecast for the cluster, and then generalized exponential smoothing takes the following form

$$\hat{x}_{n+1}^{cluster}(t) = MED_{FD^\beta}(W_{n,k}^{cluster}),$$

where  $W_{n,k}$  denotes a moving window of length  $k$  ending in a moment  $n$ , i.e.

$$W_{n,k}^{cluster} = \{x_{n-k+1}^{cluster}(t), \dots, x_n^{cluster}(t)\}.$$

Then joint reconciliation of forecasts is conducted and our reconciled predictor takes the form:

$$\hat{\mathbf{X}}_{n+1}(t) = \mathbf{F}(\hat{x}_{n+1}^{cluster_1}(t), \dots, \hat{x}_{n+1}^{cluster_M}(t)),$$

where  $\mathbf{F}$  denotes the reconciliation procedure. In Proposal 1,  $\mathbf{F}$  is a generalized least squares procedure originally proposed by Shang and Hyndman (2017) and  $M$  is the number of moving median predictors.

PROPOSAL 2: Sample  $\beta_{threshold}$ -trimmed mean with locality parameter  $\beta$  is defined as

$$ave(\beta_{threshold}, \beta)(\mathbf{X}^N) = ave(x_i : FD^\beta(x_i | \mathbf{X}^N) > \beta_{threshold}),$$

where  $ave$  denotes the sample functional average, and  $\beta_{threshold}$  is a pre-specified trimming threshold. In this setup a generalized exponential smoothing technique is applied independently on each considered level of hierarchy and in each cluster of that level as well. As a predictor for  $(n+1)th$  moment we take

$$\hat{x}_{n+1}^{cluster}(t) = z \cdot ave(\beta_{threshold}^1, \beta_1)(W_{n_1, k_1}^{cluster}) + (1-z) \cdot ave(\beta_{threshold}^1, \beta_1)(W_{n_2, k_2}^{cluster}),$$

where  $W_{n,k}^{cluster}$  denotes a moving window of length  $k$  ending in a moment  $n$ , i.e.  $W_{n,k}^{cluster} = \{x_{n-k+1}^{cluster}(t), \dots, x_n^{cluster}(t)\}$ ,  $z \in [0, 1]$  is a forgetting parameter and  $n_2 < n_1$ . Thus we consider a closer past represented by  $W_{n_1}^{cluster}$  and a further past represented by  $W_{n_2}^{cluster}$ .

Note that lengths of the moving windows  $k, k_1, k_2$  used in Proposals 1 – 2 relate to the analogous forgetting parameters  $\alpha$  in the classical exponential smoothing. Additionally, we have at our disposal the resolution parameter  $\beta$ , at which we predict a phenomenon. Such an approach allows us to accommodate expert knowledge and adjust forgetting and resolution parameters to the researcher's requirements. For comparison purpose, note that in original Shang and Hyndman (2017) paper, the authors made predictions using functional regression based on constant in time functional principal components scores modelled by means of the well known one-dimensional time series models (see Hyndman and Shang, 2009).

In the next step we consider the whole hierarchy structure of the phenomenon. Assume that a hierarchical structure is described by fixed hierarchy levels, which are also divided into sub-levels, which are divided into sub-levels, and so on – assume that we have  $M$  clusters in the whole hierarchical structure. Smart reconciliation of forecasts is conducted in this step and our reconciled predictor takes the form:

$$\hat{\mathbf{X}}_{n+1}(t) = \mathbf{F}(\hat{x}_{n+1}^{cluster_1}(t), \dots, \hat{x}_{n+1}^{cluster_M}(t)),$$

where  $\mathbf{F}$  is a Generalized Least Squares Estimator (see Shang and Hyndman, 2017).

Using Shang and Hyndman (2017) notation we can write our model in the form

$$R_n = S_n b_n,$$

where  $R_n$  is a vector of all series at all clusters,  $b_n$  is a vector of the most disaggregated data and  $S_n$  is a fixed matrix that shows a relation between them. In the considered example we have  $R_n = [R_{Australia}, R_{NSW}, R_{QLD}, R_{SA}, R_{TAS}, R_{VIC}]^T$ ,  $b_n = [R_{NSW}, R_{QLD}, R_{SA}, R_{TAS}, R_{VIC}]^T$ , where  $T$  denotes a vector transpose. The matrix  $S_n$  is an  $6 \times 5$  matrix, where the only non-zero elements are  $S_{1i} = 1$  for  $i = 1, 2, \dots, 5$  and  $S_{jj-1} = 1$  for  $j = 2, \dots, 6$ . We propose to do the base forecast:

$$\hat{R}_{n+1} = S_{n+1} \beta_{n+1} + \varepsilon_{n+1},$$

where  $\hat{R}_{n+1}$  is a matrix of the base forecast for all series at all levels,  $\beta_{n+1} = E[b_{n+1} | R_1, \dots, R_n]$  is the unknown mean of the forecast distribution of the most disaggregated series and  $\varepsilon_{n+1}$  is responsible for errors. We propose to use a generalized least square method as in Shang and Hyndman (2017)

$$\hat{\beta}_{n+1} = (S_{n+1}^T W^{-1} S_{n+1})^{-1} S_{n+1}^T W^{-1} \hat{R}_{n+1}$$

modified so that we use a robust estimator of the dispersion matrix  $W$ , i.e. instead of diagonal matrix, which contains forecast variances of each time series, we can use a robust measure of joint forecast dispersion taking into account dependency structure between the level forecasts. Note that a dynamic updating of the dispersion matrix estimates should be considered in further studies. Let us define our dispersion matrix:

$$W = \text{diag}\{v^{total}, v^{cluster_1}, \dots, v^{cluster_M}\}$$

where

$$v^{cluster} = V \left\{ \int_0^T \left( x_{nk}^{cluster}(t) - \hat{x}_n^{cluster}(t) \right)^2 dt, k = 1, 2, \dots, K_{cluster}, n = 1, 2, \dots, N \right\},$$

$K_{cluster}$  is the number of observations in the considered cluster in time  $n$ ,  $N$  is here the number of moments at which observations have been made and where  $V$  is some chosen robust measure of dispersion. We propose to substitute  $V = c \cdot MAD^2$

instead of Shang and Hyndman's (2017) variance. If we consider a hierarchy as in the above Figure 1, our dispersion matrix takes a simple form:

$$W = \text{diag}\{v^{\text{Australia}}, v^{\text{NSW}}, v^{\text{QLD}}, v^{\text{SA}}, v^{\text{TAS}}, v^{\text{VIC}}\}.$$

We propose to use  $c \cdot \text{MAD}^2$  instead of variance or take into account a dependency structure between level series using the well-known minimum covariance determinant (MCD) or recently proposed PCS robust matrix estimators of multivariate scatter (see Vakili and Schmitt (2014)).

#### 4. Properties of our proposals

Thanks to the kindness of prof. Han Lin Shang, who made his R script available for us, we calculated the optimal combination of forecast predictor and we compared Shang and Hyndman (2017) proposal with our Proposals 1 and 2 and with independent moving functional mean (no reconciliation was conducted for this predictor).

We generated samples of trajectories from one dimensional SV, GARCH, Wiener, Brownian bridge processes, functional FAR(1) processes tuned as in Didericksen et al. (2012), and various mixtures of them. In the cases of the considered mixtures, we treated one of the component as "good" and the rest as "bad" - outlying. In the simulations we considered several locality parameters which differed within the levels of hierarchy and several moving window lengths. We considered samples with and without functional magnitude as well as shape outliers (see Kosiorowski, Mielczarek, Rydlewski, 2018b, 2018c, and references therein). The outliers were defined with respect to the functional boxplot induced by cGBD, i. e. we replaced 1%, 5%, 10% of curves in the samples by means of arbitrary curves being outside a band determined by the functional boxplot whiskers, and compared medians and medians of absolute deviations from the medians (MAD) of integrated forecasts errors in these two situations. Fig. 2 presents simulated hierarchical functional time series consisting of three functional autoregression models of order 1 (i.e. FAR(1)) with Gaussian kernels and sine-cosine errors design (see Didericksen et al. 2012). Fig. 3 presents corresponding level forecasts obtained by means of our Proposal 1 and local moving median calculated from 15-obs. windows and locality parameters equal to 0.45. Fig. 4 presents simulated hierarchical time series consisting of three processes, each being mixtures of two one-dimensional stochastic volatility processes (SV). Fig. 5 presents corresponding level forecasts obtained by means of our Proposal 1 and local moving median calculated from 15-obs. window and locality parameters equal to 0.45. In Figures 2, 4, 5 and 6 colours indicate time sequence of the functional objects according to basic R package *terrain.colors* colour palette. We indicated the order of appearance of observations using colors palette starting from yellow and ending in blue. In the appendix we placed a simple script depending on *DepthProc* R package illustrating a general idea of the performed simulations.

In order to check the statistical properties of our proposals we considered em-

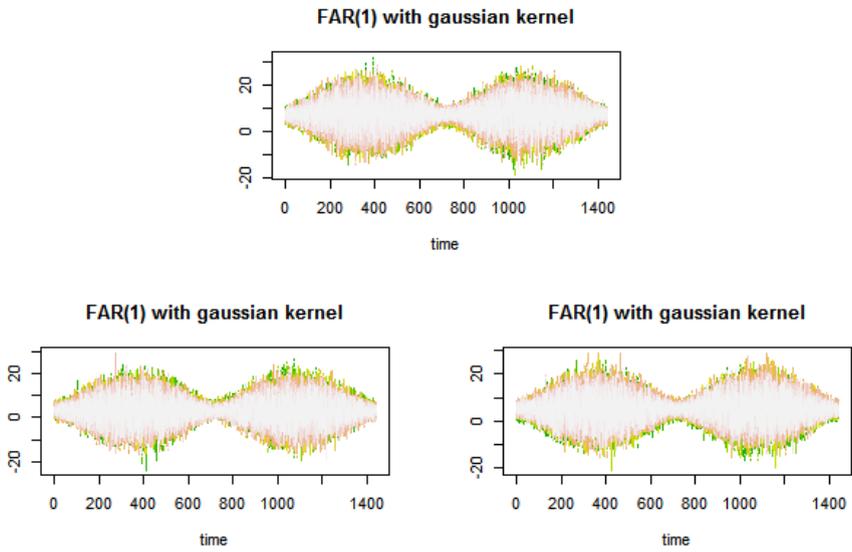


Figure 2: Simulated HFTS consisting of FAR(1) processes

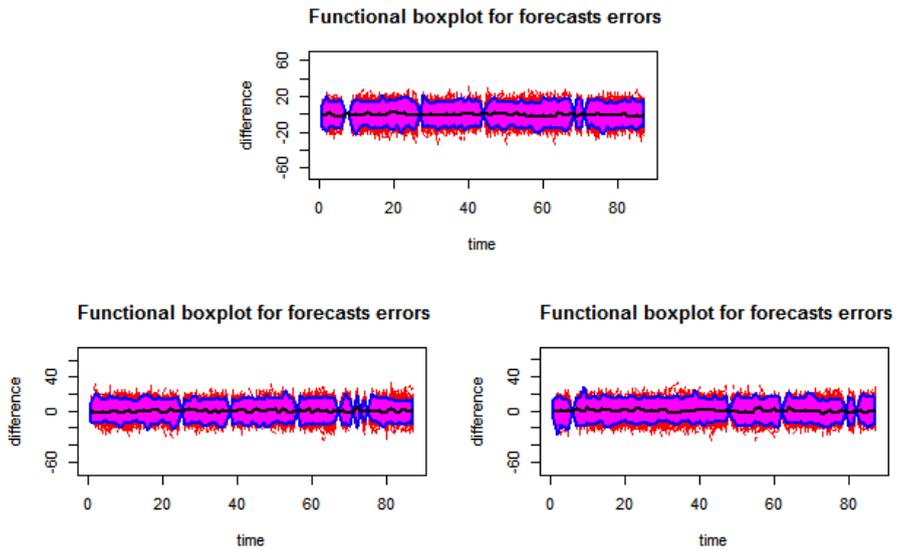


Figure 3: HFTS prediction using our Proposal 1

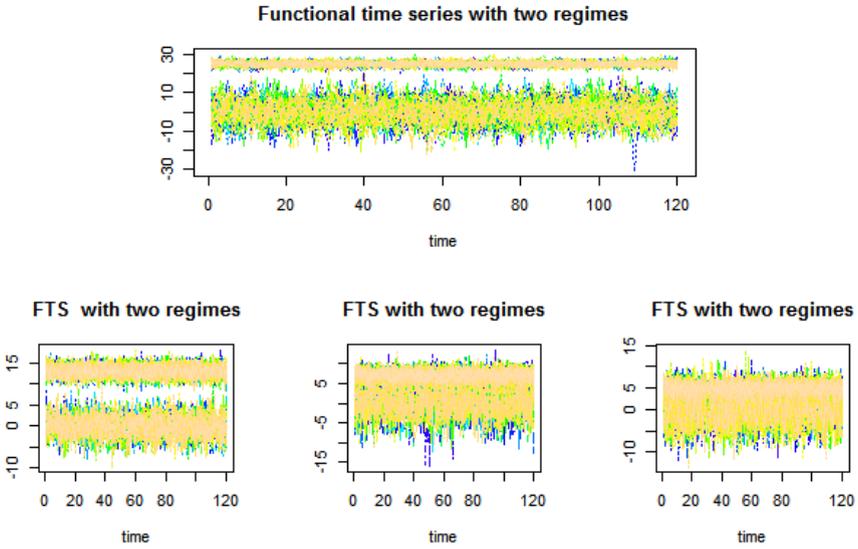


Figure 4: Simulated HFTS consisting of two regime FTS processes

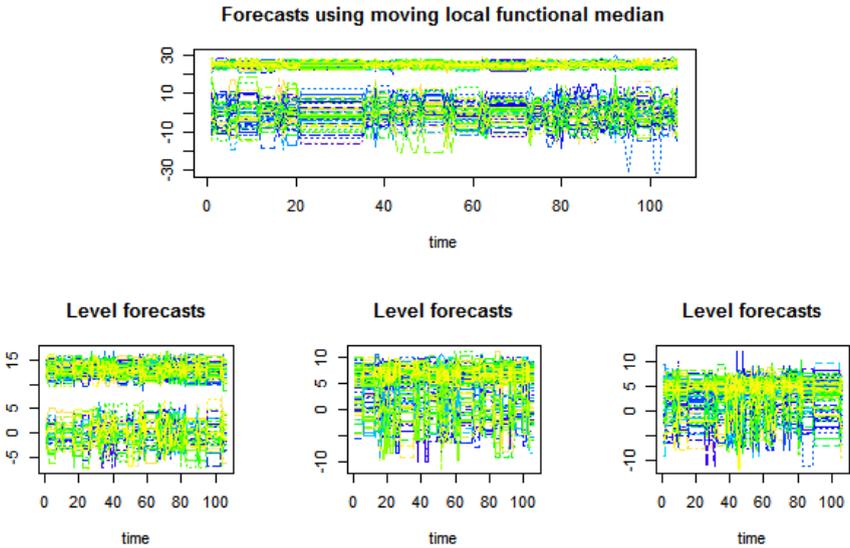


Figure 5: HFTS prediction using our Proposal 1

pirical data set related to an electricity demand in the period from 1 to 31 January 2016 in Australia. The data come from five regions of Australia, denoted with the following symbols: NSW, QLD, SA, TAS, VIC. All the considered data were taken from Australian Energy Market Operator <https://www.aemo.com.au/>. Fig. 6 presents 291 predicted electricity demand curves obtained by means of our Proposal 1 using moving local median to calculate level forecasts, window lengths  $k = 10$  observations and locality parameter equal to  $\beta = 0.2$ . Fig. 7 presents boxplots for integrated prediction errors in each region and in the whole Australia. Fig. 8 presents functional boxplot for the prediction of errors for the whole Australia in 2016, where the predictions were obtained using Proposal 1. Fig. 9 presents functional boxplots for prediction of errors of electricity demand in NSW, QLD, VIC and TAS in 2016 obtained using Proposal 1. Note that the functional median of prediction errors is close to 0 for Australia and its regions (see Fig. 8 and 9). The boxplots in Fig. 7 show that a prediction error is small as well. The volumes of central regions in Fig. 8 and 9 may be treated as the predictor effectiveness measure. We could therefore deduce that our predictor is median-unbiased (for more details on median-unbiasedness for functional data see Kosiorowski, Mielczarek and Rydlowski, 2017b, and references therein).

The performance of our proposals was compared with Shang and Hyndman (2017) proposal and with the independent moving functional mean predictor (without the reconciliation of forecasts), in terms of the median of absolute deviation of integrated prediction errors in each region and in the whole Australia. Table 1 summarizes results of this comparison. In general, the obtained results lead us to the conclusion that our proposals seem to be more robust to functional outliers than Hyndman and Shang proposal. It is not surprising, as the authors made their forecasts on the basis of nonrobust generalized least squares method. Admittedly, Shang and Hyndman (2017) claimed that their proposal performed better in comparison with bottom-up approach based on moving medians, but note that they considered Fraiman and Muniz global depths only. Moreover, thanks to the locality parameter adjustment, our proposals are more appropriate for detecting the change of regimes in the HFTS setup. In the cases of data sets without outliers, simple functional moving means, where the reconciliation procedure is not conducted, seem to outperform all other proposals. In this "clean data" situation, the performance of our both proposals and of Shang and Hyndman (2017) proposal is comparable. Our second proposal is more computationally demanding, however.

**Table 1.** MAD of integrated forecasts errors

<i>Predictor</i>	<i>Australia</i>	<i>NSW</i>	<i>VIC</i>	<i>SA</i>	<i>QLD</i>	<i>TAS</i>
Proposal 1	1126	470	401	146	224	49
Proposal 2	1311	628	452	147	181	52
H & S Proposal	1275	627	1004	176	230	51

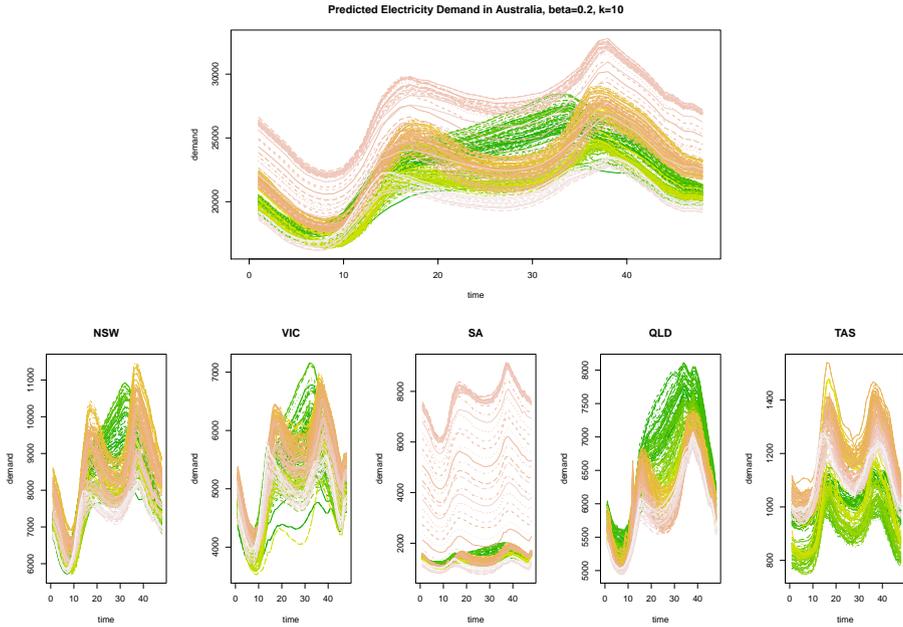


Figure 6: Predicted electricity demand in Australia in 2016

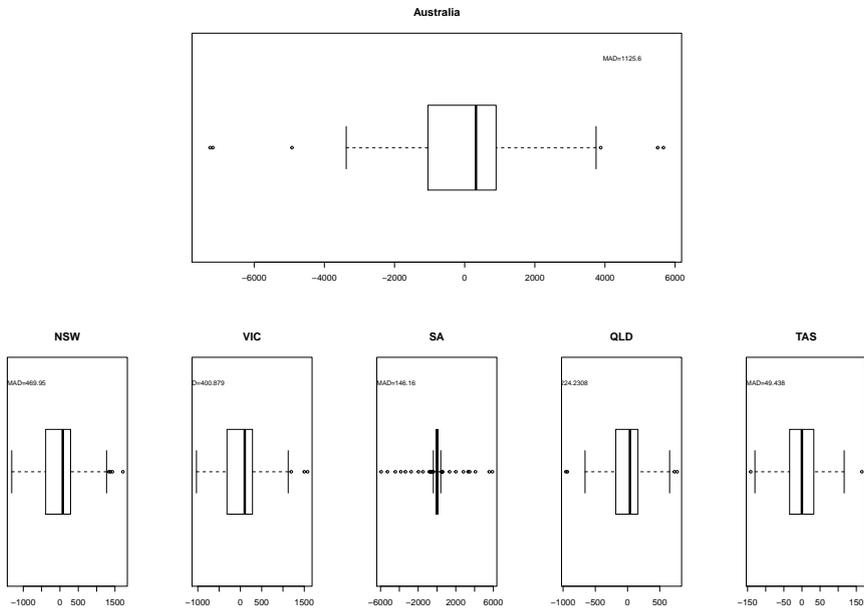


Figure 7: Boxplots for integrated forecasts errors for electricity demand in Australia and its regions in 2016, predictions obtained using Proposal 1

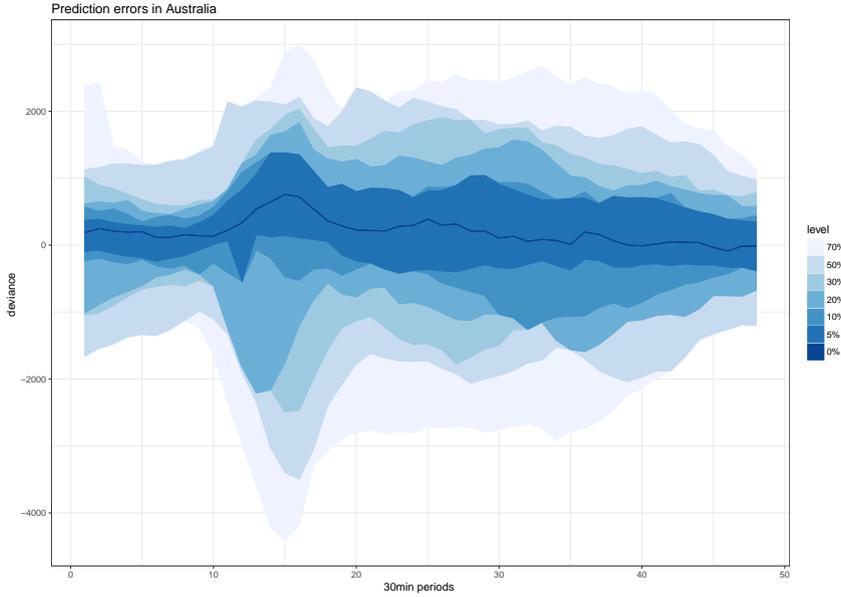


Figure 8: Functional boxplot for prediction of errors for the whole Australia in 2016, predictions obtained using Proposal 1

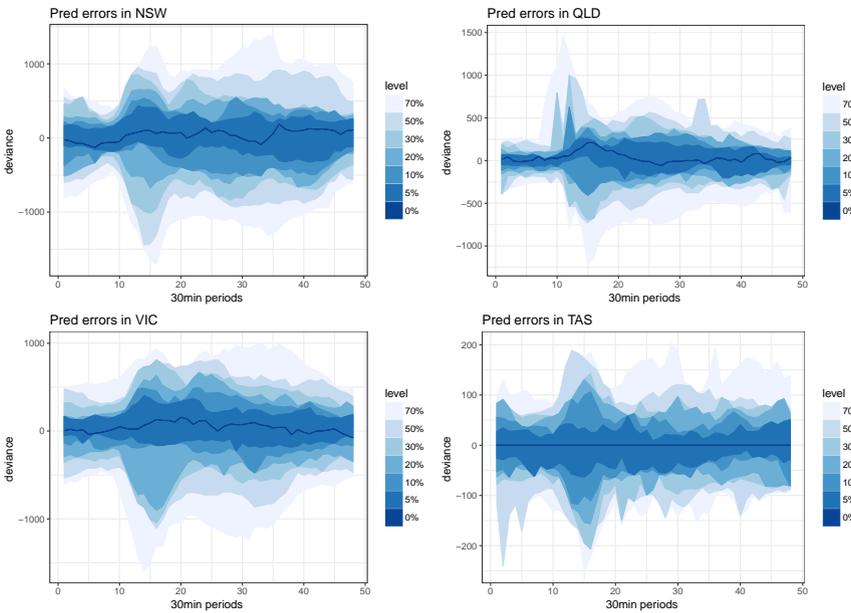


Figure 9: Functional boxplots for prediction errors of electricity demand in NSW, QLD, VIC and TAS in 2016 obtained using Proposal 1

#### 4.1. Uncertainty evaluation

Series of functional principal component scores are considered as surrogates of original functional time series (see Aue et al. (2015), Hyndman and Shang (2009)). Several authors postulate using the dynamic functional principal components approach in order to take into account the time changing dependency structure of the described phenomenon (Aue et al., 2015). Note that such modification may drastically increase computational complexity of the HFTS procedure. In a context of uncertainty evaluation of our proposals, we suggest considering Vinod and López-de-Lacalle (2009) maximum entropy bootstrap for time series approach. Bootstrap methods for FTS were studied by Hörmann and Kokoszka (2012) and Shang (2018), among others. Similarly, as in Shang and Hyndman (2017) we propose to use maximum entropy bootstrap methodology to obtain confidence regions and to conduct statistical inference. The *meboot* and *DepthProc* R packages give the appropriate computational support for these aims.

### 5. Conclusions

The hierarchical functional time series methodology opens new areas of statistical as well as economic research. E-economy provides a great deal of HFTS data. Our HFTS predictor proposal, which is based on local moving functional median, performs surprisingly well in comparison with Shang and Hyndman (2017) proposal. The lengths of the moving windows used in our proposals relate to the forgetting parameters  $\alpha$ 's in the classical exponential smoothing. Moreover, we have at our disposal a "data resolution parameter" -  $\beta$ , at which we predict the phenomenon. When using the locality parameter, we can take into account different sensitivity to details, e.g. the number of different regimes of the considered phenomena. Further economic applications of the HFTS methodology may be found, for example, in Kosiorowski, Mielczarek, Rydlewski (2018b).

### Acknowledgements

JPR research has been partially supported by the AGH UST local grant no. 15.11.420.038. DM and JPR research has been partially supported by the Faculty of Applied Mathematics AGH UST statutory tasks within subsidy of the Ministry of Science and Higher Education, grant no. 11.11.420.004. MS research was partially supported by NSC Grant no. OPUS.2015.17.B.HS4.02708, DK research has been partially supported by the grant awarded to the Faculty of Management of CUE for preserving scientific resources for 2016, 2017 and 2018.

## REFERENCES

- AUE, A., DUBABART NORINHO, D., HÖRMANN, S., (2015). On the prediction of stationary functional time series, *Journal of the American Statistical Association*, 110 (509), pp. 378–392.
- BESSE, P. C., CARDOT, H., STEPHENSON, D. B., (2000). Autoregressive forecasting of some functional climatic variations, *Scandinavian Journal of Statistics*, 27 (4), pp. 673–687.
- BOSQ, D. (2000). *Linear processes in function spaces*. Springer.
- DIDERICKSEN, D., KOKOSZKA, P., ZHANG, X. (2012). Empirical properties of forecasts with the functional autoregressive model, *Computational Statistics*, 27 (2), pp. 285–298.
- FEBRERO-BANDE, M. O., DE LA FUENTE, M., (2012). Statistical computing in functional data analysis: the R package *fda.usc*, *Journal of Statistical Software*, 51 (4), pp. 1–28.
- HORVATH, L., KOKOSZKA, P., (2012). *Inference for functional data with applications*, Springer-Verlag.
- HÖRMANN S., KOKOSZKA, P., (2012). Functional Time Series, in *Handbook of Statistics: Time Series Analysis – Methods and Applications*, 30, pp. 157–186.
- HYNDMAN, R. J., AHMED R. A., ATHANASOPOULOS, G., SHANG, H. L., (2011). Optimal combination forecasts for hierarchical time series, *Computational Statistics & Data Analysis*, 55 (9), pp. 2579–2589.
- HYNDMAN, R.J., KOEHLER, A.B., ORD, J. K., SNYDER, R. D., (2008). *Forecasting with exponential smoothing – the state space approach*, Springer-Verlag.
- HYNDMAN, R. J., SHANG, H., L., (2009). Forecasting functional time series, *Journal of the Korean Statistical Society*, 38 (3), pp. 199–221.
- HYNDMAN, R. J., ULLAH, M., (2007). Robust forecasting of mortality and fertility rates: A functional data approach, *Computational Statistics & Data Analysis*, 51 (10), pp. 4942–4956.
- HYNDMAN, R. J., KOEHLER, A. B., SNYDER, R.D., GROSE, S., (2002). A state space framework for automatic forecasting using exponential smoothing methods, *International Journal of Forecasting*, 18 (3), pp. 439–454.

- KAHN, K. B., (1998). Revisiting top-down versus bottom-up forecasting, *The Journal of Business Forecasting Methods & Systems*, 17 (2), pp. 14–19.
- KOHN, R., (1982). When is an aggregate of a time series efficiently forecast by its past, *Journal of Econometrics*, 18 (3), pp. 337–349.
- KOSIOROWSKI, D., ZAWADZKI, Z., (2018). DepthProc: An R package for robust exploration of multidimensional economic phenomena, *arXiv: 1408.4542*.
- KOSIOROWSKI, D., (2014). Functional regression in short term prediction of economic time series, *Statistics in Transition*, 15 (4), pp. 611–626.
- KOSIOROWSKI, D. (2016). Dilemmas of robust analysis of economic data streams, *Journal of Mathematical Sciences (Springer)*, 218 (2), pp. 167–181.
- KOSIOROWSKI, D., RYDLEWSKI, J. P., SNARSKA, M., (2017a). Detecting a structural change in functional time series using local Wilcoxon statistic, *Statistical Papers*, pp. 1–22, URL <http://dx.doi.org/10.1007/s00362-017-0891-y>.
- KOSIOROWSKI, D., MIELCZAREK, D., RYDLEWSKI, J. P., (2017b). Double functional median in robust prediction of hierarchical functional time series – an application to forecasting of the Internet service users behaviour, available at: [arXiv:1710.02669v1](https://arxiv.org/abs/1710.02669v1).
- KOSIOROWSKI, D., RYDLEWSKI, J.P., ZAWADZKI Z., (2018a). Functional outliers detection by the example of air quality monitoring, *Statistical Review (in Polish, forthcoming)*.
- KOSIOROWSKI, D., MIELCZAREK, D., RYDLEWSKI, J. P., (2018b). Forecasting of a Hierarchical Functional Time Series on Example of Macromodel for the Day and Night Air Pollution in Silesia Region - A Critical Overview, *Central European Journal of Economic Modelling and Econometrics*, 10 (1), pp. 53–73.
- KOSIOROWSKI, D., MIELCZAREK, D., RYDLEWSKI, J. P., (2018c). Outliers in Functional Time Series – Challenges for Theory and Applications of Robust Statistics, In M. Papież & S. Śmiech (eds.), *The 12<sup>th</sup> Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, Conference Proceedings, Cracow: Foundation of the Cracow University of Economics, pp. 209–218.
- KRZYŚKO, M., DEREGOWSKI, K., GÓRECKI, T., WOŁYŃSKI, W., (2013). Kernel and functional principal component analysis, *Multivariate Statistical Analysis 2013 Conference, plenary lecture*.

- LÓPEZ-PINTADO, S., ROMO, J., (2009). On the concept of depth for functional data, *Journal of the American Statistical Association*, 104, pp. 718–734.
- LÓPEZ-PINTADO, S., JÖRNSTEN, R., (2007). Functional analysis via extensions of the band depth, *IMS Lecture Notes–Monograph Series Complex Datasets and Inverse Problems: Tomography, Networks and Beyond*, Vol. 54, pp. 103–120, Institute of Mathematical Statistics.
- NAGY, S., GIJBELS, I., OMELKA, M., HLUBINKA, D., (2016). Integrated depth for functional data: Statistical properties and consistency, *ESIAM Probability and Statistics*, 20, pp. 95–130.
- NAGY, S. GIJBELS, I., HLUBINKA, D., (2017). Depth-Based Recognition of Shape Outlying Functions, *Journal of Computational and Graphical Statistics*, DOI: 10.1080/10618600.2017.1336445.
- NIETO-REYES, A., BATTEY, H., (2016). A topologically valid definition of depth for functional data, *Statistical Science* 31 (1), pp. 61–79.
- PAINDAVEINE, D., G. VAN BEVER, G., (2013). From depth to local depth: a focus on centrality, *Journal of the American Statistical Asssociation*, Vol. 108, No. 503, *Theory and Methods*, pp. 1105–1119.
- RAMSAY, J.O., G. HOOKER, G., GRAVES, S., (2009). *Functional data analysis with R and Matlab*, Springer-Verlag.
- SGUERA, C., GALEANO, P., LILLO, R. E., (2016). Global and local functional depths, arXiv 1607.05042v1.
- SHANG, H., L., HYNDMAN, R. J., (2017). Grouped functional time series forecasting: an application to age-specific mortality rates, *Journal of Computational and Graphical Statistics*, 26(2), pp. 330–343.
- SHANG, H., L., (2018). Bootstrap methods for stationary functional time series, *Statistics and Computing*, 28(1), pp. 1–10.
- WEALE, M., (1988). The reconciliation of values, volumes and prices in the national accounts, *Journal of the Royal Statistical Society A*, 151(1), pp. 211–221.
- VAKILI, K., SCHMITT, E., (2014). Finding multivariate outliers with FastPCS, *Computational Statistics & Data Analysis*, 69, pp. 54–66.
- VINOD, H.D., LÓPEZ-DE-LACALLE, J. L., (2009). Maximum entropy bootstrap for time series: the meboot R package, *Journal of Statistical Software*, 29 (5).

ZUO, Y., SERFLING, R., (2000). Structural properties and convergence results for contours of sample statistical depth functions, *Annals of Statistics*, 28 (2), pp. 483–499.

Australian Energy Market Operator, <https://www.aemo.com.au/>

## APPENDIX

#Simple R script, example showing how to calculate base forecasts for three hierarchy levels  
#using moving functional median implemented within the DepthProc R package.

```
require(DepthProc)
```

```
require(fda.usc)
```

```
require(RColorBrewer)
```

```
require(zoo)
```

```
wrapMBD = function(x){ depthMedian(x, depth_params = list(method="Local", beta=0.45,  
depth_params1 = list(method = "MBD"))) }
```

```
#Simple stochastic volatility 1D process simulator
```

```
SV j- function(n, gamma, fi, sigma, delta) {
```

```
epsilon j- rnorm(n)
```

```
eta j- rnorm(2*n, 0, delta)
```

```
h j- c()
```

```
h[1] j- rnorm(1)
```

```
for (t in 2:(2*n)) {
```

```
h[t] j- exp(gamma+fi*(h[t-1]-gamma)+sigma*eta[t]) }
```

```
Z j- sqrt(tail(h,n)) * epsilon
```

```
return(Z)}
```

```
example j- SV(100, 0, 0.2, 0.5, 0.1)
```

```
plot(ts(example))
```

```
#functional time series simulator
```

```
m.data1 j- function(n,a,b) {
```

```
M j- matrix(nrow=n,ncol=120)
```

```
for (i in 1:n) M[i,]j- a*SV(120,0,0.3,0.5,0.1)+b
```

```
M }
```

```
m.data.out1 j- function(eps,m,n,a,b,c,d){
```

```
H j- rbind(m.data1(m,a,b),m.data1(n,c,d))
```

```
ind=sample((m+n),eps)
```

```
H1=H[ind,]
```

```
H1 }
```

```
m j- matrix(c(1, 0, 1, 3, 2, 3, 2, 0), nrow = 2, ncol = 4) m[2,]=c(2,2,3,3) m[1,]=c(0,1,1,0)
```

```
#below three functional time series
```

```
M2A= m.data.out1(150,3000,7000,5,0,1,25)
```

```
M2B= m.data.out1(150,3000,7000,2,0,1,15)
```

```
M2C= m.data.out1(150,3000,7000,3,0,1,10)
```

```
matplot(t(M2A),type="l",col=topo.colors(151),xlab="time", main="Functional time series with  
two regimes")
```

```
matplot(t(M2B),type="l",col=topo.colors(151),xlab="time", main="FTS with two regimes")
```

```
matplot(t(M2C),type="l",col=topo.colors(151), xlab="time", main="FTS with two regimes")
```

```
#below moving local medians applied to the above series, window lengths = 15 obs.,
```

```
#locality parameters betas = 0.45
```

```
result4A = rollapply(t(M2A),width = 15, wrapMBD, by.column = FALSE)
```

```
result4B = rollapply(t(M2B),width = 15,wrapMBD, by.column = FALSE)
```

```
result4C = rollapply(t(M2C),width = 15, wrapMBD, by.column = FALSE)
```

```
matplot(result4A,type="l",col=topo.colors(87), xlab="time",main="local 15-obs moving func-  
tional median, beta=0.45")  
matplot(result4B,type="l",col=topo.colors(87), xlab="time",main="local 15-obs moving func-  
tional median, beta=0.45")  
matplot(result4C,type="l",col=topo.colors(87), xlab="time",main="local 15-obs moving func-  
tional median,  
beta=0.45")
```

#basic function for calculating  $\beta\_treshold$  – trimmed  $\beta$  – local MBD functional mean

```
beta_tresh_meanj-function(x,beta_tresh,beta)  
depths= depth(x, depth_params = list(method="Local", beta=beta, depth_params1 = list(method  
= "MBD")))  
ind=which(depths $\beta_tresh$ )  
wyn = func.mean(x[ind,])  
wyn$data
```