Mathematical Modelling and Analysis
Volume 27, Issue 1, 117–133, 2022
https://doi.org/10.3846/mma.2022.14043

Functional Modelling of Telecommunications Data*

Algimantas Birbilas and Alfredas Račkauskas

Vilnius University, Institute of Applied Mathematics
Naugarduko g. 24, LT-03225 Vilnius, Lithuania
E-mail(*corresp.*): algimantas.birbilas@mif.vu.lt
E-mail: alfredas.rackauskas@mif.vu.lt

Received December 10, 2020; revised December 15, 2021; accepted December 15, 2021

Abstract. This work deals with statistical modeling and forecasting of telecommunications data. Main mobile traffic events (SMS, Voice calls, Mobile data) are smoothed using B-spline functions and later analyzed in a functional framework. Functional linear auto-regression models are fitted using both bottom-up and top-down design methodologies. The advantages and disadvantages of both approaches for the prediction of mobile telephone users' habits are discussed.

Keywords: functional data analysis, functional linear regression, telecommunications data, prediction.

AMS Subject Classification: 62R10; 62M10.

1 Introduction

Modelling and predicting telecommunication parameters is a common challenge for over 20 years, see Frost and Melamed [12] and Meier-Hellstern et al. [14]. Any kind of new technology or new business approach requires new traffic measurement models addressed to different problems (lack of known requirements, not having enough data, etc.).

In this paper, telecommunications data are treated as observations of random curves. Instead of traditional statistical approach, we can exploit Functional Data Analysis (FDA) methods, which, according to Levitin et al. [13], can answer a number of different implementation questions. An overview of FDA is provided by Ramsay and Silverman [17], Ferraty and Vieu [11] and more recently by Wang et al. [19]. To the best of our knowledge, only a few

* The research supported by the Research Council of Lithuania, grant No. S-MIP-17-76

papers use functional approach for telecommunications data. Yu and Lambert [20] analysed records for completed international calls. The objective of the paper by Ben Slimen et al. [18] is to detect future malfunctions of a set of cells, by only observing key performance indicators that are considered as functional data. Aspirot et al. [2] study a non-parametric regression model, where the explanatory variable is non-stationary dependent functional and the response variable is scalar. There, this approach is applied to telecommunications, namely to estimate the quality of service for an end-to-end connection on a network.

The data set under investigation in this paper contains three different types of telecommunications data: the amount of SMS, Voice consumption, and Mobile data consumption. All three components can be named Call Data Records (CDR). This type of telecommunications data is used in a vast number of research papers, Calabrese et al. [7], Ozgul et al. [15], Cecaj et al. [1] to name a few.

Data set is obtained from Mobile Virtual Network Operator (MVNO), which is defined as an entity, who offers telecommunications services similar to a mobile network operator (MNO), however, the MVNO does not own any radio frequency spectrum (see e.g., www.yozzo.com/mvno-wiki/mvno-definitions). In other words, MVNO rents technical telecommunication tools (mobile network towers, CRM, Billing system, etc.) in order to provide services. Moreover, MVNOs purchase an amount of main mobile network products (SMS, Voice calls, Mobile data) at wholesale prices from MNO and sell them to the customers at their prices. Such business model is popular around the world (e.g., Brazile Telecom, Uno Mobile Italy, Carrefour Taiwan Mobile, Samatel Oman, Equitel Kenya, etc.). However, this business model has limited profitability so it is crucial to understand customer needs. Moreover, as price in the telecommunication business becomes more and more irrelevant, business owners must understand their customer habits to operate successfully.

In this paper, usage of three mobile products, Voice calls, SMS and Mobile data are investigated following functional data analysis methodology. Let $X = (X(t), t \geq 0)$ be a real valued continuous time process, e.g., consumption in continuous time of any of these mobile products. In order to study the behavior of X within the time interval of length d (one month in this research), we set

$$X_k(t) = X((k-1)d + td) - X((k-1)d), \quad k \geq 1, \ 0 \leq t \leq 1. \tag{1.1}$$

This construction generates a time series $X_k = (X_k(t), 0 \leq t \leq 1), \quad k \geq 0$, of random functions with values in a function space, say $\mathbb{E}$. The main problem discussed in the paper is forecasting of X_{n+1} from observations of $X_1, \ldots, X_n$. As a functional framework for paths of random processes under consideration we fix two function spaces. The first one is $\mathbb{E} = L_2(0,1)$ the classical Lebesgue space of square integrable functions on $[0,1]$, endowed with the inner product

$$\langle x, y \rangle = \int_0^1 x(t)y(t)dt, \quad x, y \in L_2(0,1),$$

and the corresponding norm $||x|| = \sqrt{\langle x, x \rangle}, \quad x \in L_2(0,1)$. The second is the space $\mathbb{E} = C[0,1]$ of continuous functions on $[0,1]$ endowed with the maximum norm $||x|| = \max_{0\le t\le 1}|x(t)|$. The choice of a functional framework in our context provides instruments to measure forecasting accuracy.

As a parallel aim, we deal with the aggregation problem which is sometimes referred to as the top-down versus bottom-up forecasting problem. It leads to the question: what is better, to forecast the aggregate of top level quantity directly, or to forecast the individual components separately and then aggregate them to form the forecast of the total? We associate for each user of a mobile product its own process $X_j = (X_j(t), t \ge 0)$, $j = 1, 2, \dots, N$, and construct the functional time series in same manner as above thus, obtaining N functional time series $X_{jk} = (X_{jk}(t), 0 \le t \le 1)$, $k \ge 1$; $j = 1, \dots, N$,

$$X_{jk}(t) = X_j((k-1)d + td) - X_j((k-1)d). \tag{1.2}$$

Clearly $X_k = \sum_{j=1}^{N} X_{jk}$. Hence, the forecast of X_{n+1} can be obtained as well from that of each $X_{j,n+1}$. The literature in statistical forecasting, and time series analysis suggests that the question what is better, to forecast directly X_{n+1} or to aggregate the forecasts of each $X_{j,n+1}$, is far from settled at either the theoretical or empirical levels. In our specific practical application, it is difficult to argue on theoretical grounds what the correct approach should be. Therefore, this question is settled empirically by trying both. Note, that normalized sums

$$Y_k = \frac{1}{N}\sum_{j=1}^{N} X_{jk} \tag{1.3}$$

model a behaviour of an ,,average" consumer.

For the process $(X_k, k \ge 1)$ we fit first order functional auto-regressive models both with deterministic and random coefficient. Point-wise multiplication by a function and integral type operators are used on lagged time series. Predictions are made using rolling-window procedure and prediction residuals are compared afterwards.

The rest of the paper is organized as follows. In Section 2, we present the data under investigation. Section 3 contains the analysis of fitted two types auto-regressive models whereas in Section 4 we discuss the top-down versus bottom-up forecasting problem by using fitted models. Section 5 contains conclusions.

2 Data

Data consists of mobile products (SMS, Voice calls, Mobile data) usage by almost 1500 users. Data are filtered by only taking subscribers that consumed any of the three mobile products over 75% of all period under consideration. Moreover, data are aggregated on daily level into total count of SMS measured in units, total sum of Voice consumption measured in minutes, and total sum of Mobile data consumption in MB. Finally, after filtering out inactive subscribers we have data from 39 Mobile data consumers, 791 SMS consumers and 714 Voice calls consumers.

As the aim is to predict monthly patterns of consumers we use monthly segmentation of the data under consideration. The segmented data then are of the form

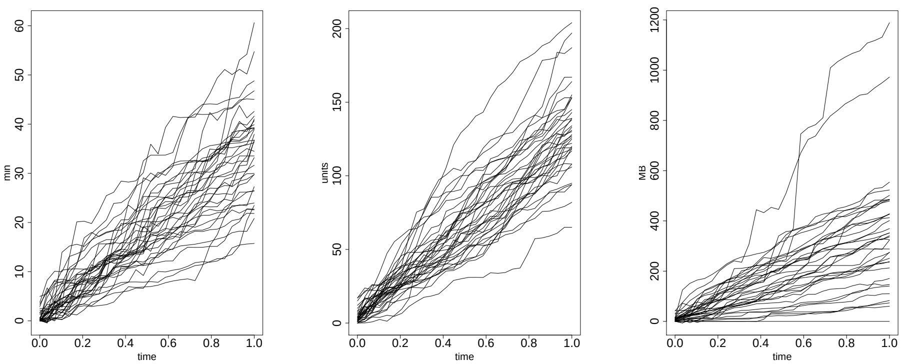
$$x_{jk}^{(\ell)} = (x_{jki}^{(\ell)},\ i = 1, \cdots, I), \quad k = 1, \dots, n; \quad j = 1, \dots, N_\ell,$$

where $\ell = 1, 2, 3$ corresponds to the three metrics (SMS, Voice, Mobile data), j corresponds to a consumer and k indicates month as time series unit. These data are interpreted as measurements of monthly curves $x_{jk}^{(\ell)} = (x_{jk}^{(\ell)}(t), t \in [0, 1])$,

$$x_{jki}^{(\ell)} = x_{jk}^{(\ell)}(t_i) + \varepsilon_{jk}^{(\ell)}(t_i),$$

where $t_1, \dots, t_I$ denotes the reference points (corresponds to days of a month in our case) and $\varepsilon_{jk}^{(\ell)}(t_i)$ are measurement errors.

In order to get a better understanding of consumption curves, the following graphs are presented. Figure 1 provides one subscriber consumption for all months.



Voice SMS Mobile data

Figure 1. Mobile product consumption: one subscriber all months

Figure 2 provides one month curve for 10 randomly selected subscribers.

As a next step in data preparation B-splines smoothing was applied in order to reconstruct functions $x_{jk}^{(\ell)}(t), t \in [0, 1]$. This gave the three functional samples

$$x_{jk}^{(\ell)} = (x_{jk}^{(\ell)}(t), t \in [0, 1]), \quad k = 1, \dots, n; \quad j = 1, \dots, N_\ell,$$

corresponding to SMS ($\ell = 1$), Voice ($\ell = 2$) and Mobile data ($\ell = 3$) of j'th consumer during k'th month. An example of data transformation from raw to functional sample is presented in Figure 3.

Although eight basis functions produced good enough approximation, in order to keep the roughness of curves, estimation with the penalty was appropriately used. The best penalty parameter was calculated by generalized

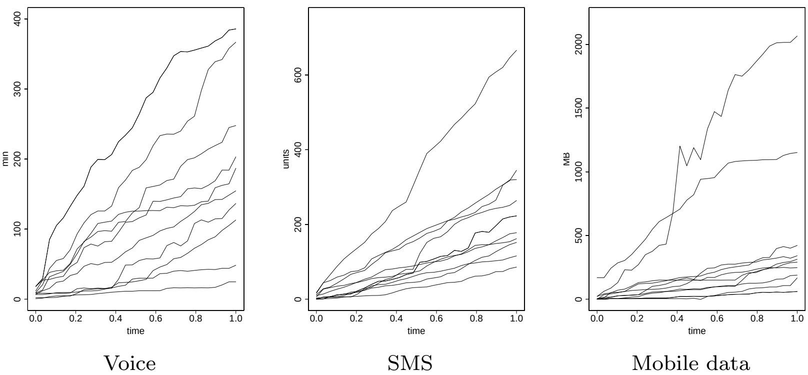


Figure 2. Mobile product consumption: ten subscribers during one month.

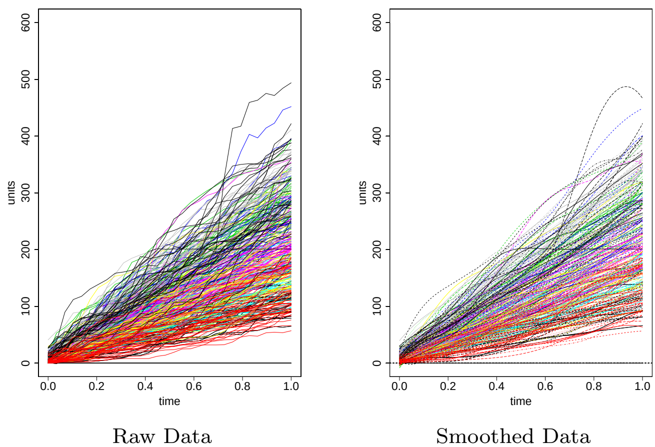


Figure 3. SMS consumption for 10 subscribers: raw vs smoothed. Different color represents different subscriber.

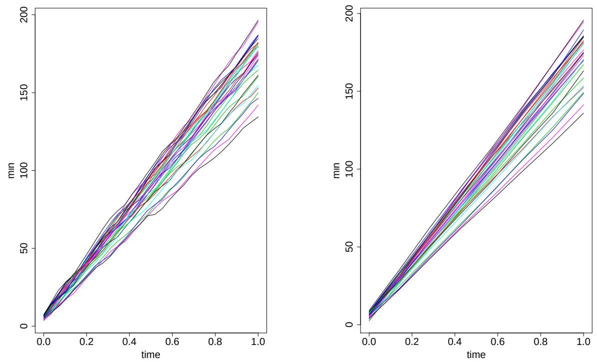
cross-validation measure as developed by Craven and Wahba [9]. Below we will simplify the notation by omitting the mobile product index ℓ.

As widely accepted in functional data analysis the obtained functional sample (x_{jk}) is interpreted as observations of a sequence of stochastic processes (X_{jk}), $X_{jk} = (X_{jk}(t), 0 \leq t \leq 1)$ defined above by (1.2). The aggregated sample $x_k = (x_k(t), 0 \leq t \leq 1)$, $k = 1, \ldots, n$, is obtained by setting

$$x_k(t) = \sum_{j=1}^{N} x_{jk}(t)$$

and is interpreted as observations of the functional time series (X_k) defined in (1.1).

Graphical illustration for aggregated Voice consumption is presented in Figure 4.



Raw Data Smoothed Data

Figure 4. Aggregated Voice consumption: raw vs smoothed. Different color represents different month.

3 Fitting the functional time series models

One of the most popular and frequently used functional time series models is a functional auto-regressive first order (FAR(1)) model introduced by Bosq [6]. This model has been successfully used by many authors, including Cavallini et al. [8], Besse and Cardot [4], Besse et al. [5], Bernard [3], and Damon and Guillas [10] for forecasting of electricity consumption, traffic, climatic variations, electrocardiograms, and ozone concentration respectively.

We found two auto-regressive models appropriate for the functional time series $(X_k, k \geq 1)$, defined in (1.1). The first one is a (possibly random coefficient) FAR(1) model

$$X_k - \mu = \rho(X_{k-1} - \mu) + \varepsilon_k, \quad k \geq 1, \tag{3.1}$$

where $X_0 = 0$, $\mu = E(X_k)$ is the mean function, ρ is a (possibly random) bounded linear operator in the space under consideration and (ε_k) is a white noise. ρ is used in equations only as the general model operator. The second model is a (possibly random coefficient) FAR(1) model for differences:

$$X_k - X_{k-1} = \rho(X_k - X_{k-1}) + \varepsilon_k, \quad k \geq 1, \tag{3.2}$$

where ρ is a (possibly random) bounded linear operator in the space under consideration. We use two types of operators ρ in (3.1) and (3.2). First one corresponds to $L_2(0,1)$ framework and is specified as a convolution kernel operator

$$(\rho x)(t) = \int_0^1 \kappa(t,s)x(s)ds, \quad t \in [0,1],$$

where the function $\kappa(t,s)$ satisfies

$$\int_0^1 \int_0^1 \kappa^2(t,s)dtds < \infty,$$

whereas the second type of ρ corresponds to $C[0,1]$ framework and is specified as multiplication by a function:

$$(\rho x)(t) = \beta(t)x(t), \quad t \in [0,1],$$

where $\beta(t), t \in [0,1]$, is a continuous function. The choice of a functional framework for functional sample in our context fixes a distance used to measure the forecasting accuracy. Note also that point-wise models for Hilbert space valued time series does not make sense in general. Since we have prepared the data as continuous curves we can exploit both functional spaces $L_2(0,1)$ and $C[0,1]$. For estimation procedures of FAR(1) models we refer to Ramsay et al. [16].

3.1 Point-wise autoregressive model

3.1.1 Non random coefficient

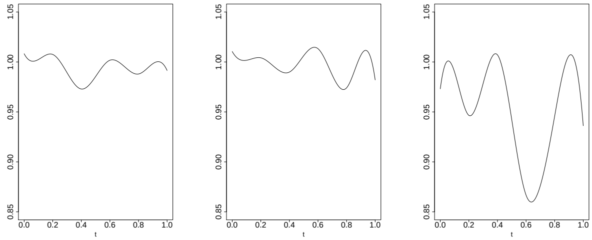
For the functional time series $(X_k, k \geq 1)$ defined in (1.1) consider first the point-wise FAR(1) model

$$X_k(t) - \mu(t) = \beta(t)(X_{k-1}(t) - \mu(t)) + \varepsilon_k(t), \quad t \in [0,1], \tag{3.3}$$

where $\beta(t), t \in [0,1]$ is an unknown continuous function. $\beta(t)$ is used only in point-wise non-differentiated models. We estimate $\mu(t)$ and $\beta(t)$ from the functional sample $x_0(t) = 0, x_1(t), x_2(t), \ldots, x_n(t)$ by

$$\widehat{\mu}(t) = \frac{1}{n}\sum_{k=1}^{n} x_k(t), \quad \widehat{\beta}(t) = \frac{\sum_{k=1}^{n}(x_{k-1}(t) - \widehat{\mu}(t))(x_k(t) - \widehat{\mu}(t))}{\sum_{k=1}^{n}(x_{k-1}(t) - \widehat{\mu}(t))^2}.$$

The function $\widehat{\beta}(t)$ is shown in Figure 5.



Voice SMS Mobile data

Figure 5. $\widehat{\beta}(t)$ function for model (3.3) with Voice, SMS, and Mobile data consumption data.

The values of $\widehat{\beta}(t)$ varies in the interval $[0.85, 1.05]$. This indicates either random walk behavior of the centered process $(X_k(t) - \mu(t), k \geq 1)$ or randomness of the function $\beta(t)$ in (3.3). The first option leads to the model

$$X_k(t) - X_{k-1}(t) = \gamma(t)(X_{k-1}(t) - X_{k-2}(t)) + \varepsilon_k(t), \quad t \in [0, 1] \tag{3.4}$$

with OLS estimator of $\gamma(t)$ ($\gamma(t)$ is used only in point-wise differentiated models) given by

$$\widehat{\gamma}(t) = \frac{\sum_{k=1}^{n}(x_{k-1}(t) - x_{k-2}(t))(x_k(t) - x_{k-1}(t))}{\sum_{k=1}^{n}(x_{k-1}(t) - x_{k-2}(t))^2}.$$

We see in Figure 6, that the values of $\widehat{\gamma}(t)$ are in the interval $[-0.7, 0.4]$ which means point-wise stationarity of the solution of (3.4).

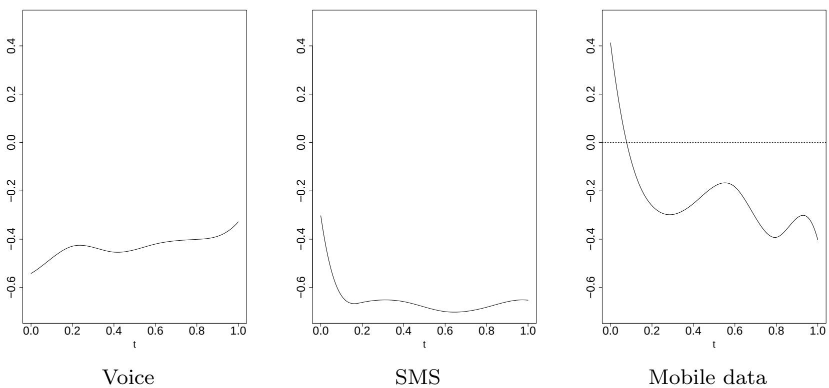


Figure 6. $\widehat{\gamma}(t)$ function for differentiated model (3.4) with Voice, SMS, and Mobile data consumption data.

3.1.2 Random coefficient

To introduce randomness for $\beta(t)$ in (3.3) we consider for each $j = 1, \ldots, N$, the point-wise FAR(1) model

$$X_{jk}(t) - \mu_j(t) = \beta_j(t)(X_{j,k-1}(t) - \mu_j(t)) + \varepsilon_{j,k}(t), \quad t \in [0, 1], \tag{3.5}$$

where $\mu_j = E(X_{jk})$ and $\beta_j(t)$ is unknown continuous function on $[0, 1]$.

Recall for each $k = 1, \ldots, n$ and each $j = 1, \ldots, N$ the curve $x_{jk} = (x_{jk}(t), 0 \leq t \leq 1)$ is interpreted as observation of the random process $X_{j,k} = (X_{j,k}(t), 0 \leq t \leq 1)$ considered as random element in the Banach space $C[0, 1]$. The mean μ_j then is estimated by sample mean $\widehat{\mu}_j(t)$:

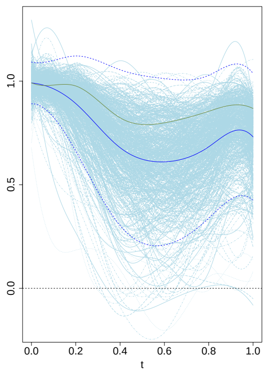
$$\widehat{\mu}_j(t) = \frac{1}{n}\sum_{k=1}^{n} x_{jk}(t)$$

for each $t \in [0,1]$. Next for each $j = 1, \dots, N$ we estimate the function β_j in (3.5) by OLS obtaining

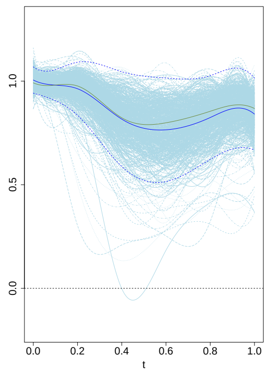
$$\widehat{\beta}_j(t) = \frac{\sum_{k=1}^{n}(x_{j,k-1}(t) - \widehat{\mu}_j(t))(x_{j,k}(t) - \widehat{\mu}_j(t))}{\sum_{k=1}^{n}(x_{j,k-1}(t) - \widehat{\mu}_j(t))^2},$$

where $x_{j,0}(t) = 0$ for any $j = 1, \dots, N$. We interpret these functions as observations of the random function $(\beta(t), t \in [0,1])$ in (3.3). Hence, all distributional parameters of $\beta(t)$ can be estimated from the sample $\widehat{\beta}_1(t), \dots, \widehat{\beta}_N(t)$.

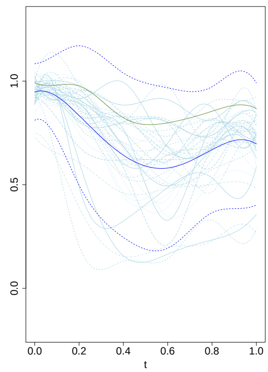
Functions $\widehat{\beta}_j(t), j = 1, \dots, N$, are shown in Figure 7 together with the sample mean, median and mean $\pm$ two standard deviations. As the mean and standard deviation are estimated point-wise, the estimation of median uses the integration of a univariate depth along the time axis.



Voice



SMS



Mobile data

Figure 7. Functions $\widehat{\beta}_j(t)$ in light blue for Voice, SMS, and Mobile data consumption data presented in (3.5) together with sample mean in blue, two standard deviations in dotted blue and median in green.

We estimate the function $\beta(t), t \in [0,1]$, in (3.3) by

$$\widehat{\beta}(t) = \frac{1}{N}\sum_{j=1}^{N}\widehat{\beta}_j(t), \quad t \in [0,1].$$

Similarly randomness of the function $\gamma(t), t \in [0,1]$, in (3.4) comes from processes $(X_{jk}, k \geq 1), j = 1, \dots, N$. So that we first fit the model

$$X_{j,k}(t) - X_{j,k-1}(t) = \gamma_j(t)(X_{j,k-1}(t) - X_{j,k-2}(t)) + \varepsilon_{j,k}(t), \; t \in [0,1], \; j = 1, \dots, N$$

with OLS estimator of $\gamma_j(t)$ given by

$$\widehat{\gamma}_j(t) = \frac{\sum_{k=1}^{n}(x_{j,k-1}(t) - x_{j,k-2}(t))(x_{j,k}(t) - x_{j,k-1}(t))}{\sum_{k=1}^{n}(x_{j,k-1}(t) - x_{j,k-2}(t))^2}.$$

All $\widehat{\gamma}_j(t)$ values are in the interval $[-0.6, 0.5]$. This indicates the point-wise stationarity of the first differences of each of the process $(X_{jk}, k \geq 1)$.

Now we estimate the function $\gamma(t), t \in [0,1]$, in (3.4) by

$$\widehat{\gamma}(t) = \frac{1}{N} \sum_{j=1}^{N} \widehat{\gamma}_j(t), \quad t \in [0,1].$$

A number of insights can be brought from $\widehat{\beta}_j(t)$ and $\widehat{\gamma}_j(t)$ graphs. First of all, socio-economic insights - subscribers tend to live on calendar month cycles. The main activity is monitored at the beginning and the end of the month while the middle is less active. Second insight - cultural. The most active subscribers from this operator are ethnic minorities from Middle East, Africa, Central America where communication intensity is higher. These groups also tend to be more active on Voice and SMS consumption rather than Mobile data due to sufficient bundle offerings (free or lower price calls and SMS to African countries, Cuba, etc.) from the operator. Moreover, these groups tend to transfer money to their relatives in their home countries after they collect them - during the start or end of the month. This leads to communication increase before and after money transfers.

$\widehat{\beta}_j(t)$ graphs in Figure 7 indicates that consumption is not connected to any mobile bundle (fixed amount of mobile products for a certain price with one month validity) renewal time. Despite the small number of more active months for a small number of subscribers, the majority values of Mobile data consumption are relatively small (up to 1024 MB) and there is no proof that the bundle is on average fully exhausted during the month. Regarding Voice and SMS, consumption exceeds bundle limits so most probably subscriber purchased add-on (one time add-on of certain mobile product) or continue actions with non-bundled prices. This can be identified in graphs where $\widehat{\beta}_j(t)$ values fluctuate several times. Despite if an add-on is purchased or not after the bundle is exhausted, $\widehat{\beta}_j(t)$ stays under a stable shape of one fluctuation. To conclude, the bundle can be renewed any time of the month so stable shape whole month shows no direct link between bundle renewal and consumption.

Voice and SMS consumption habits are quite similar as it is seen from $\widehat{\gamma}_j(t)$ and $\widehat{\beta}_j(t)$ graphs. The main difference is the start and end of the calendar month. It can be explained by an increase in arriving technical SMS on certain periods. In other words, subscribers receive info about certain services (ex. low balance, utility cost, etc.) or commercial offerings (ex. new bundle, new offer, etc.) during the start and end of the month. This info is usually provided by SMS, not by Voice call. On the contrary, Mobile data consumption during that period is used more for entertainment. It can seen from stable average function of $\widehat{\gamma}_j(t)$ for Mobile data in Figure 8.

To conclude, cultural and socio-economic insights can explain the shape of $\widehat{\beta}_j(t)$ and $\widehat{\gamma}_j(t)$ graphs. Moreover, technical communication features can explain certain moments of the month and the difference between mobile product consumption.

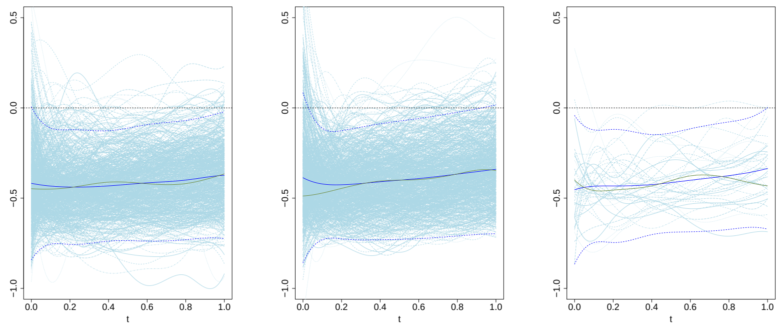


Figure 8. Functions $\widehat{\gamma}_j(t)$ in light blue for Voice, SMS, and Mobile data consumption data presented in (3.4) together with sample mean in blue, two standard deviations in dotted blue and median in green.

3.2 Integral type autoregressive model

In this subsection we consider the functional time series $(Y_k, k \geq 1)$ where Y_k is defined by (1.3). Eventually, we fit the model

$$Y_k(t) - Y_{k-1}(t) = \int_0^1 \kappa(t,s)(Y_{k-1}(s) - Y_{k-2}(s))ds + \varepsilon_k(t), \quad t \in [0,1].$$

There exist several methods to estimate the kernel $\kappa(t,s)$. We used fit of fully functional linear model. Fit was implemented using functional basis representation:

$$\kappa(t,s) = \sum_k \sum_l b_{kl} v_k(s) \theta_l(t) = v(s)' B \theta(t)$$

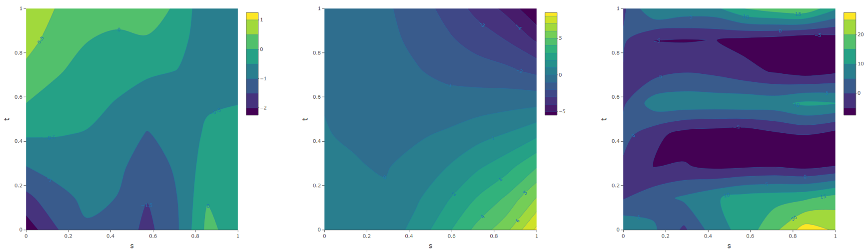
with K basis functions v_k and L basis functions θ_l. Furthermore, least squares method is used to estimate coefficients b_{kl}. This method was briefly presented by Ramsey et al. [17].

The model provides sufficient insights about Voice consumption. Estimated $\kappa(t,s)$ indicated that accumulated history s has less power than current moment t. Moreover, the effect on consumption changes from negative to positive (or stays close to zero depending on accumulated consumption history) at the end of the month.

On contrary, estimated $\kappa(t,s)$ of SMS consumption indicates that the history of consumption needs to be accumulated at least until the second part of the month to have an actual effect. When a significant part of consumption is accumulated, it can be indicated that accumulated history has a positive impact during the start of the month ($t < 0.2$) and a negative impact at the end of the month ($t > 0.8$).

The estimated model provides sufficient insights into Mobile data consumption as well. Accumulated history influences consumption during three periods.

After the start of the month ($t < 0.2$) $\kappa(t, s)$ is rising as more history is accumulated. During the month ($t \in [0.2, 0.8]$), the accumulated consumption history effect fluctuates between positive and negative values but remains more or less stable. The end of the month ($t > 0.8$) is similar to the start of the month where accumulated history is rising. $\kappa(t, s)$ surfaces are presented in Figure 9.



Voice SMS Mobile data

Figure 9. $\widehat{\kappa}(t, s)$ for Voice, SMS, and Mobile data consumption.

4 Forecasting

The focus of this section is forecasting. We seek to predict the future values x_{n+1} of the functional process $(X_k, k \geq 1)$ as a function of data $x_1, \ldots, x_n$, so as to minimize the mean squared forecast error. Using the auto-regressive model (3.1) under the assumption that ε_k is a strong white noise, the conditional expectation of $X_{n+1} - \mu$ given $X_1, X_2, \ldots, X_n$ equals to $\rho(X_n - \mu)$. Hence, for known ρ, the best predictor is given by $\rho(X_n - \mu)$. Since ρ is unknown an approximation to the solution is $\widehat{\rho}(X_n - \widehat{\mu})$, where $\widehat{\rho}$ is a consistent estimator of ρ and $\widehat{\mu}$ is the empirical mean. Thence, model (3.1) leads to

$$\widehat{X}_{n+1} = \widehat{\mu} + \widehat{\rho}(X_n - \widehat{\mu}).$$

If ρ is deterministic, we take its estimator as discussed above. If ρ is random, we take $\widehat{\rho}$ an estimator of the mean $E\rho$. Likewise, using the model (3.2) we have

$$\widehat{X}_{n+1} = X_n + \widehat{\rho}(X_n - X_{n-1}). \tag{4.1}$$

By de-aggregated (bottom-up) approach we have

$$\widehat{X}_{n+1} = \widehat{\mu} + \sum_{j=1}^{N} \widehat{\rho}_j(X_{j,n} - \widehat{\mu}_j),$$

where $\widehat{\mu}_j = n^{-1} \sum_{k=1}^{n} X_{jk}$, and

$$\widehat{X}_{n+1} = X_n + \sum_{j=1}^{N} \widehat{\rho}_j(X_{j,n} - X_{j,n-1}). \tag{4.2}$$

The corresponding mean square prediction errors are

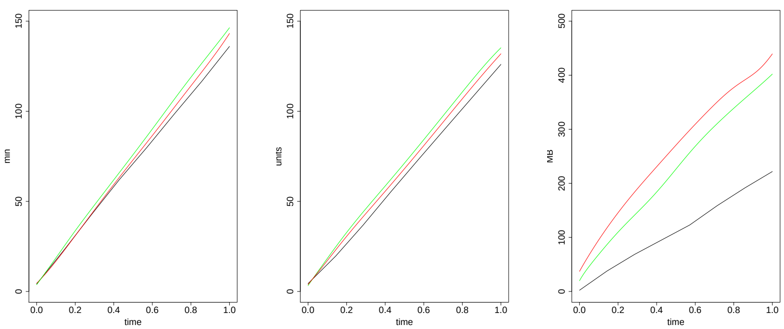
$$\Delta_1 := E\Big(\int_0^1 (\widehat{X}_{n+1}(t) - X_{n+1}(t))^2 dt\Big)^{\frac{1}{2}}, \quad \Delta_2 := E[\max_{0\le t\le 1} |\widehat{X}_{n+1}(t) - X_{n+1}(t)|].$$

To see how accurately predictive models perform we use rolling-window procedure. The procedure starts with choosing rolling window set (sequence of observations) and its length p :$\{ x_1, x_2, \ldots, x_p\}$. Rolling-window is taken from data sample, so p depends on sample size T: $\{x_1, x_2, \ldots, x_p\} \subset \{x_1, x_2, \ldots, x_T\}$. Afterwards, forecast horizon length h is defined. Horizon is the last period in rolling-window and it is used to measure the quality of predictions. As data under consideration contains 36 months of mobile products consumption (T=36), we choose rolling-window length p equal to 31 months and horizon length h equal to 1 (horizon set is $\{x_p\}$). Rolling-window is implemented by going month forward for every partition. Hence, six partitions of data set are constructed : $\{x_{1+p}, x_{2+p}, \ldots, x_{31+p}\}, p = 0, \ldots, 5$. These date sets are used to fit the models. Residuals are presented in the Table 1 below.

Table 1. Summary of prediction errors Δ_1 and Δ_2. SMS are in units, Voice in min, Mobile Data in MB.

	Δ_1 SMS	Δ_2 SMS	Δ_1 Voice	Δ_2 Voice	Δ_1 Mobile data	Δ_2 Mobile data
Eq. (4.1)	**2**	**4**	**5**	**8**	125	176
Eq. (4.2)	4	7	6	11	**106**	**152**

Prediction visualization for example month is provided in Figure 10.



Voice SMS Mobile data

Figure 10. Point-wise model predictions for all three consumption metrics. Red line indicates top-down approach (4.1), green line indicates bottom-up approach (4.2), black line - actual data. Presented predictions are made from models with $1 - 30$ window months.

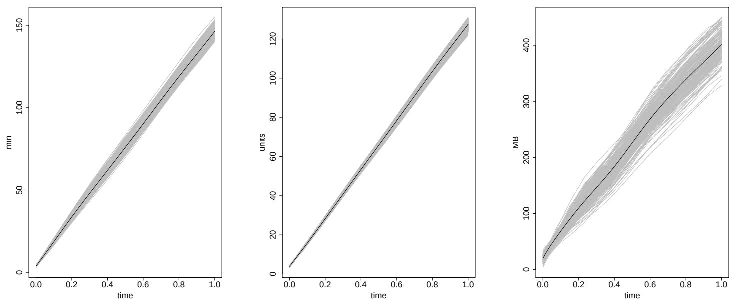
In order to understand the variation of prediction, bootstrap intervals were created. Bootstrapping for top-down approach was implemented by three steps:

1. Randomly sampling subscribers with replacement

2. Averaging consumption for sampled subscribers

3. Creating model and prediction for one step ahead

Sampling was done 200 times. An example graph is provided in Figure 11.



Voice calls SMS Mobile data

Figure 11. Prediction (black line) with bootstrap values (gray lines) for top-down (4.1) approach using $1 - 30$ months window

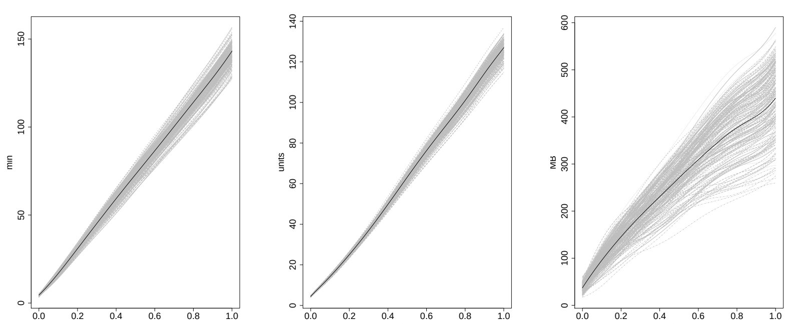
Bootstrapping for bottom-up approach was almost identical to top-down. The only difference - 2. and 3. swap their positions:

1. Randomly sampling subscribers with replacement

2. Creating model and prediction for one step ahead

3. Averaging consumption for sampled subscribers

An example graph is provided below in Figure 12.



Voice calls SMS Mobile data

Figure 12. Prediction (black line) with bootstrap values (gray lines) for bottom-up approach (4.2) model using $1 - 30$ months window.

Models predict consumption quite well for most cases. Few models for Mobile data consumption prediction did perform poorly, which enlarged Δ_1 and Δ_2 errors. This led to the conclusion, that the higher number of mobile product users, the better is accuracy. Moreover, forecasts are stable for the whole month, no significant fluctuations can be identified during the month.

Overall, forecasts manage to recognize socio-economic and cultural effects presented before. Point-wise models with top-down design performed better for SMS and Voice cases. On the contrary, Mobile data consumption predictions were better with the bottom-up design. Finally, bootstrapped intervals suggested that the bottom-up approach is less robust in comparison to the top-down.

5 Conclusions

Analysis manages to identify models and make predictions for mobile product consumption. Two types of strategies, Top-down and Bottom-up, help to guide through model creation. Two types of models, point-wise and integral-wise, provided insights about mobile product consumption.

The data set is analyzed as a functional data object. All mobile products acted more or less stable despite which rolling window was used. Almost all coefficient $\widehat{\beta}_j(t)$ and $\widehat{\gamma}_j(t)$ functions are in the interval $(-1, 1)$ identifying itself as trend stationary for bottom-up design. Top-down design $\widehat{\gamma}(t)$ functions also fluctuate in the same interval after amendments. Coefficient functions $\widehat{\beta}_j(t)$ and $\widehat{\gamma}_j(t)$ provide socio-economic (monthly business cycle) and cultural (ethnic minorities consumption habits) insights about subscriber habits. Analysis of $\widehat{\beta}_j(t)$ explains stable consumption shape while analysis of $\widehat{\gamma}_j(t)$ explains the difference between Voice and SMS consumption for our data set. Insights from $\widehat{\beta}_j(t)$ and $\widehat{\gamma}_j(t)$ suggest that mobile operator successfully targeted ethnic minorities during the period. After averaging consumption data, integral-wise models with stationary $\widehat{\kappa}(t, s)$ are created. $\widehat{\kappa}(t, s)$ surfaces identifies different history effect on averaged mobile product consumption. The surface analysis reveals that accumulated history affects Mobile data consumption more than it does for SMS and Voice consumption.

Predictions calculated for both point-wise model designs. Top-down models outperform bottom-up models in SMS and Voice consumption cases. For Mobile data cases, the bottom-up model performs better. It indicates that the Top-down approach is suitable when the subscriber base is larger. Moreover, the Top-down approach requires less data preparation and computational power. Finally, both approaches provide stable results during the month.

To conclude, functional data analysis can be a sufficient tool to investigate telecommunications data. Both designs can be used to analyze and predict the consumption of main telecommunication products.

Acknowledgements

The authors thank reviewers and an associated editor for helpful comments and suggestions on the paper.

References

[1] A.Cecaj, M. Mamei and F.J. Zambonelli. Re-identification and information fusion between anonymized CDR and social network data. *Journal of Ambient Intelligence and Humanized Computing*, **7**(1):83–96, 2016. https://doi.org/10.1007/s12652-015-0303-x.

[2] L. Aspirot, K. Bertin and G. Perera. Asymptotic normality of the Nadaraya – Watson estimator for nonstationary functional data and applications to telecommunications. *Journal of Nonparametric Statistics*, **21**(5):535–551, 2009. https://doi.org/10.1080/10485250902878655.

[3] P. Bernard. Analyse de signaux physiologiques. *Memoire Univ. Catholique Angers*, 1997.

[4] P. Besse and H. Cardot. Approximation spline de la prevision d'un processus fonctionnel autoregressif d'ordre 1. *The Canadian Journal of Statistics*, **24**(4):467–487, 1996. https://doi.org/10.2307/3315328.

[5] P. Besse, H. Cardot and D.B. Stephenson. Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, **27**(4):673–687, 2000. https://doi.org/10.1111/1467-9469.00215.

[6] D. Bosq. *Linear processes in function spaces: theory and applications, vol. 149.* Springer-Verlag New York, 2011. https://doi.org/10.1007/978-1-4612-1154-9.

[7] F. Calabrese, Z. Smoreda, V.D. Blondel and C. Ratti. Interplay between telecommunications and face-to-face interactions: A study using mobile phone data. *PLOS One*, **6**(7):1–6, 2011. https://doi.org/10.1371/journal.pone.0020814.

[8] A. Cavallini, G.C. Montanari, M. Loggini, O. Lessi and M. Cacciaris. Nonparametric prediction of harmonic levels in electrical networks. *Proceed. IEEE ICHPS VI*, pp. 165–171, 1994.

[9] P. Craven and G. Wahba. Smoothing noisy data with spline functions. *Numerische Mathematik*, **31**(4):377–403, 1978. https://doi.org/10.1007/BF01404567.

[10] J. Damon and S. Guillas. The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics*, **13**(7):759–774, 2002. https://doi.org/10.1002/env.527.

[11] F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice.* Springer Series in Statistics Springer, New York, 2006. https://doi.org/10.1007/0-387-36620-2.

[12] V.S. Frost and B. Melamed. Traffic modeling for telecommunication networks. *IEEE Communications Magazine*, **32**(3):70–81, 1994. https://doi.org/10.1109/35.267444.

[13] D.J. Levitin, R.L. Nuzzo, B.W. Vines and J.O. Ramsay. Introduction to functional data analysis. *Canadian Psychology/Psychologie canadienne*, **48**(3):135–155, 2007. https://doi.org/10.1037/cp2007014.

[14] K.S. Meier-Hellstern, P.E. Wirth, Y. Yan and D.A. Hoeflin. Traffic models for ISDN data users: office automation application. *ITC-13*, pp. 167–172, 1991. Available from Internet: `http://132.187.12.7/fileadmin/ITCBibDatabase/1991/meierhellstern911.pdf`.

[15] F. Özgül, A. Çelik, C. Atzenbeck and N. Gergin. *Investigating Terrorist Attacks Using CDR Data: A Case Study.* Lecture Notes in Social Networks. Springer, Vienna, 2011. https://doi.org/10.1007/978-3-7091-0388-3_17.

[16] J. Ramsay, G. Hooker and S. Graves. *Functional Data Analysis with R and MATLAB, 1st Edition.* Springer, New York, 2009. https://doi.org/10.1007/978-0-387-98185-7_1.

[17] J. Ramsay and B.W. Silverman. *Functional Data Analysis, 2nd Edition.* Springer Series in Statistics Springer, New York, 2005. https://doi.org/10.1007/b98888.

[18] Y. Ben Slimen, S. Allio and J. Jacques. Anomaly prevision in radio access networks using functional data analysis. In *GLOBECOM 2017 – 2017 IEEE Global Communications Conference*, pp. 1–6, 2017. https://doi.org/10.1109/GLOCOM.2017.8254087.

[19] J.L. Wang, J.-M. Chiou and H.-G. Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, **3**(1):257–295, 2016. https://doi.org/10.1146/annurev-statistics-041715-033624.

[20] Y. Yan and D. Lambert. Fitting trees to functional data, with an application to time-of-day patterns. *Journal of Computational and Graphical Statistics*, **8**(4):749–762, 1999. https://doi.org/10.1080/10618600.1999.10474847.