Analysis of spatio-temporal mobile phone data: a case study in the metropolitan area of Milan

Piercesare Secchi · Simone Vantini · Valeria Vitelli

P. Secchi · S. Vantini
MOX - Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy
e-mail: piercesare.secchi@polimi.it

S. Vantini
e-mail: simone.vantini@polimi.it

V. Vitelli (✉)

Oslo Center for Biostatistics and Epidemiology, Department of Biostatistics, University of Oslo, Sognsvannsveien 9, Domus Medica, 0372 Oslo, Norway
e-mail: valeria.vitelli@medisin.uio.no

1 Introduction

The metropolitan area of Milan is located in Northern-Italy and, with its 7.4 million inhabitants, is the fifth largest metropolitan area in Europe after the Ruhr, Moscow, Paris, and London. It includes the nine provinces of Milan, Bergamo, Como, Lecco, Lodi, Monza-e-Brianza, Novara, Pavia, and Varese. It is characterized by a very high concentration of working activities (nearly 10 % of the entire national Italian gross domestic product comes from this area, providing a per-capita GDP 50 % higher than the national average) but also of residential activities (the province of Milan is the most populated in Italy, with more than 1,000 inhabitants per km^2, more than five times the national average). The municipality of Milan, located in the center of the metropolitan area, and quite identifiable with the area included within the highway ring-road, is of course the main attractor of the region. Nearly 1.3 million people live there but every working day its population increases nearly 50 % since 600 thousand persons commute from the metropolitan area. This large number of commuters is mainly due to the lack of housing within the municipality; this, together with lack of important investments in improving the transport system in the last decades, has generated a critical situation in terms of population density dynamics. Indeed most roads connecting the municipality of Milan with the metropolitan area have reached their saturation level with peaks of the traffic/capacity ratio up to 150 % during rush hours (see OECD 2006a, b).

OECD identifies housing, transport, and congestion as the bottlenecks for the future growth of the Milan metropolitan area. These factors seem to badly affect the well-being of the city from many perspectives: (i) pollution (Milan is the second most air-polluted city in Europe after Moscow), (ii) economy (the difficulty in mobility of people and goods is estimated to damp the output of the area of more than 4 %), and, of course, (iii) demography (while the population of the metropolitan area is growing the population of the municipality of Milan is decreasing; in 1971 nearly 1/2 of the population of the province of Milan lived within the municipality, in 2001 only 1/3).

In recent years a congestion charge has been introduced, the regional railway network has been fully integrated, three new highways and two new subway lines are under construction, and a few bike- and car-sharing initiatives have been promoted. The Green Move project, which the present research is part of, is among these initiatives. Green Move is an interdisciplinary research project financed by Regione Lombardia involving different research groups at Politecnico di Milano and focused on the development of a vehicle sharing system based on the concept of "little, electric and shared vehicles". This work is a first attempt to gather information about population density dynamics in the metropolitan area of Milan from mobile network data belonging to the Telecom Italia database. In a long term perspective this information will

be used to optimally locate vehicles and docking stations of the car-sharing network. The possible use of this information is however much wider. Large scale quantitative information on human population density dynamics is of extreme interest to the urban planner, the traffic flow being functionally related to the porosity and the permeability of the urban texture. The same information may, of course, help the city manager, for instance to locate ambulances and police patrols at a certain time in the most suited locations, or to maximize the public exposition of alerts relevant to the well-being of the community.

In the Telecom Italia database, the metropolitan area of Milan is partitioned into a uniform lattice S_0 of 97×109 sites. In each site, the average number of mobile phones simultaneously using the network for calling is provided every 15 min for 14 days. This quantity is called Erlang and, at a first approximation, can be considered proportional to the number of active people in that site at that time, thus providing information about population density dynamics. Technically the Erlang $E_{\mathbf{x}j}$ relevant to the site $\mathbf{x} \in S_0$ and to the jth quarter of an hour is computed as

$$E_{\mathbf{x}j} = \frac{1}{15} \sum_{q=1}^{Q} \left| T_{\mathbf{x}j}^{q} \right|, \tag{1}$$

where $T_{\mathbf{x}j}^{q}$ indicates the time interval (or union of intervals) in which the qth mobile phone is using the network for calling while moving within site $\mathbf{x}$ and during the jth quarter of an hour; $|T_{\mathbf{x}j}^{q}|$ indicates its length in minutes. The number of potential phones using the network is indicated with Q. Equation (1) represents the formula actually in use by the mobile company for computing $E_{\mathbf{x}j}$, but its meaning is better captured by the equivalent representation

$$E_{\mathbf{x}j} = \frac{1}{15} \int_{15(j-1)}^{15j} N_{\mathbf{x}}(t)dt, \tag{2}$$

which shows that $E_{\mathbf{x}j}$ is the mean over the jth quarter of an hour of the number $N_{\mathbf{x}}(t)$ of mobile phones using the network within site $\mathbf{x}$ at time t, measured in minutes. The equivalence of representations (1) and (2) is easily proved through the following identities

$$\frac{1}{15} \sum_{q=1}^{Q} |T_{\mathbf{x}j}^{q}| = \frac{1}{15} \sum_{q=1}^{Q} \int_{15(j-1)}^{15j} \mathbf{1}_{\{T_{\mathbf{x}j}^{q}\}}(t)dt$$

$$= \frac{1}{15} \int_{15(j-1)}^{15j} \sum_{q=1}^{Q} \mathbf{1}_{\{T_{\mathbf{x}j}^{q}\}}(t)dt = \frac{1}{15} \int_{15(j-1)}^{15j} N_{\mathbf{x}}(t)dt.$$

The Erlang data we deal with are recorded every quarter of an hour, from March 18th, 2009, 00:15, till March 31st, 2009, 23:45. Indeed, in some sites of the lattice the entire temporal profile of the Erlang values is missing, while in other sites only some values are missing or non-admissible since they are negative (and they are treated as missing values). Hence, we restrict the analysis to a non-uniform time grid with
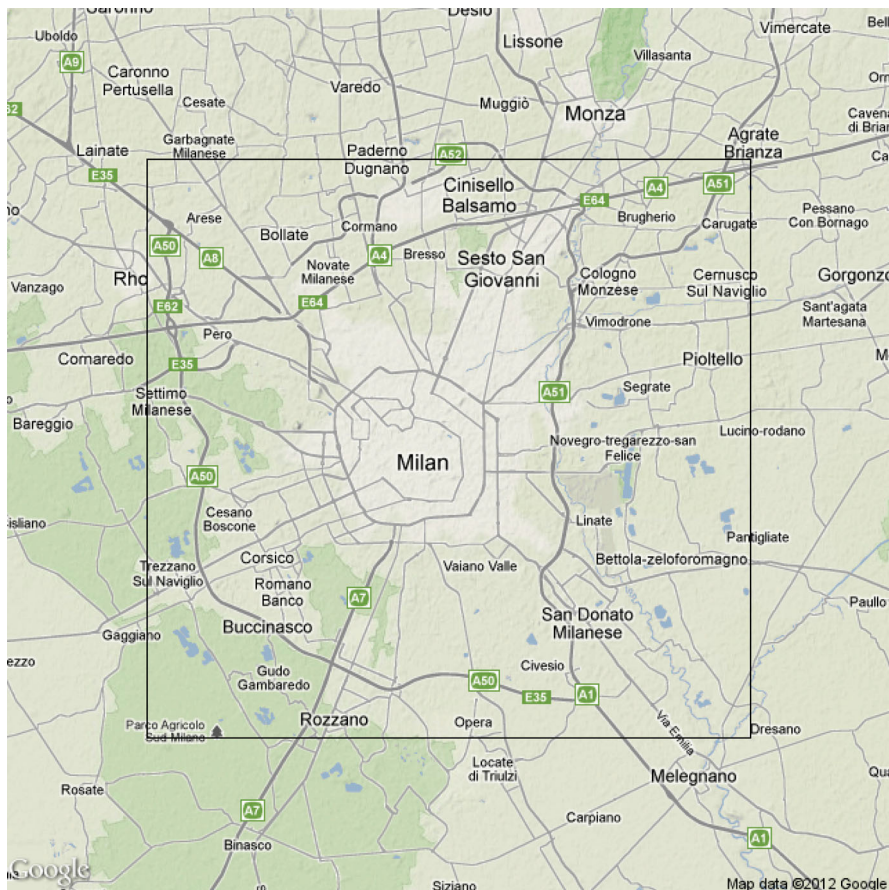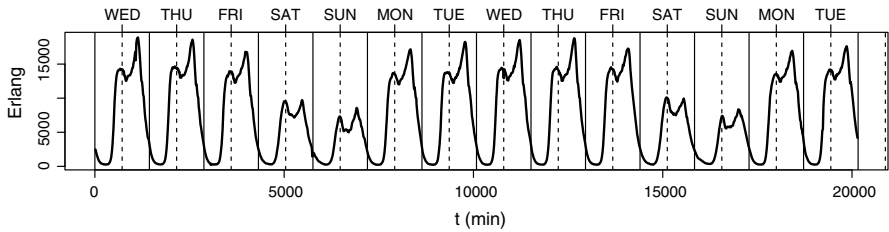
Fig. 1 In the *top panel*, the aggregated Erlang of the investigated area as a function of time. The *solid vertical lines* are drawn at midnight of each day, and the *dotted vertical lines* at noon. The first day is Wednesday March 18, 2009. In the *bottom panel*, a map of the region covered by the lattice of the Telecom Italia database

$p = 1,308$ elements, each element of the time grid being relative to a quarter of an hour for which an Erlang measurement has been observed in at least one site of the lattice. The lattice S_0 covers an area of $757 \, \text{km}^2$, reported in the bottom panel of Fig. 1, and included between latitude $45.37°$ and $45.57°$ North and longitude $9.05°$ and $9.35°$ East. It is divided in $|S_0| = N = 10{,}573$ approximately rectangular sites of

size 232 m $\times$ 309 m. Overall, 13,829,484 records are available, among which 110,475 are missing. The data set at hand can be genuinely considered an instance of spatially-dependent *functional data*, because of the high within-unit sample size and the very high signal-to-noise ratio (see Ke and Wang 2001). To our knowledge, this is the first attempt at the exploration of Erlang data with the methods provided by Functional Data Analysis (see Ramsay and Silverman 2005). To have a first idea of these data, in the top panel of Fig. 1 the aggregated Erlang for the investigated area, $\sum_{\mathbf{x} \in S_0} E_{\mathbf{x} j}$, is reported as a function of time, measured in minutes. A first inspection shows some global features such as the day/night effect and working/weekend day effect. The aim of the analysis is to identify these global features together with the local ones, more subtle to detect and possibly associated to particular subregions of the investigated area.

The Erlang data are progressively arousing the enthusiasm of the urban planners' community (Becker et al. 2011; Calabrese et al. 2011; Manfredini et al. 2015). In this work we aim at using Erlang data for segmenting the metropolitan area of Milan into subregions that share the same activity pattern along time in terms of population density dynamics. To this end, Erlang data need to be properly decomposed to point out the relevant spatial and temporal dynamics. More specifically, we propose to integrate a Treelet analysis for dimensional reduction (see Lee et al. 2008) with a Bagging Voronoi strategy for the exploration of spatial dependence (Secchi et al. 2013). Even though the previous work described in Secchi et al. (2013) shares the same philosophical approach, both the purpose of the analysis and its methodological details are different: indeed in Secchi et al. (2013) the focus is on clustering spatially dependent functional data, while here we aim at dimensional reduction of spatially dependent signals. In the present context some specific methodological challenges arise, whose solution has to be integrated to enrich the previously proposed method.

The rest of the paper is structured in five sections. In Sect. 2 the methodology used to perform dimensional reduction of spatially dependent functional data is presented. In Sect. 3 the site-wise temporal smoothing of time-varying Erlang data through a suitable Fourier expansion is described. In Sect. 4 the results of the analysis of the Telecom Italia database are shown. In Sect. 5, we draw some conclusions and we trace possible directions for future research. Finally, in Sect. 6, the Supplementary material available online is briefly described: it consists in a simulation study conducted to address some specific methodological issues.

2 Data analysis: methodology

Erlang data can give insight on different aspects of the urban area they are referred to, and their analysis can be developed with various scopes: the segmentation of the area into districts characterized by homogeneous telephonic patterns; the identification of a set of "reference signals" able to describe the different temporal patterns of use of the mobile phone network; the description of the influence of each detected telephonic pattern in each site of the lattice.

Let $E_{\mathbf{x}}(t)$ be the random Erlang value at site $\mathbf{x} \in S_0$ and time t. We consider an additive model where $E_{\mathbf{x}}(t)$ is represented by the value at t taken by a limited number K of time-varying functions $\{\psi_1, ..., \psi_K\}$, common to the entire lattice S_0, and coupled

with the values at $\mathbf{x}$ taken by K latent random fields $\{D_1, \ldots, D_K\}$ indexed by the sites of S_0:

$$E_{\mathbf{x}}(t) = \sum_{k=1}^{K} D_k(\mathbf{x})\psi_k(t) + \epsilon. \tag{3}$$

The random error term ϵ is assumed to be independent of $\{D_1, \ldots, D_K\}$, with zero mean and bounded variance.

Model (3) is often assumed in the context of reduced basis representation methods, like Functional Principal Component Analysis (FPCA) and Independent Component Analysis (ICA). The important difference here stands in the fact that we introduce spatial dependence, through the random fields $\{D_1, \ldots, D_K\}$ generating the scores of the basis expansion.

Model (3) implies the following regression model for the collection $\{E_{\mathbf{x}}(t)\}_{\mathbf{x} \in S_0}$ of time profiles belonging to our Erlang dataset,

$$\mathbb{E}\left[E_{\mathbf{x}}(t)|D_1 = d_1, \ldots, D_K = d_K\right] = \sum_{k=1}^{K} d_k(\mathbf{x})\psi_k(t), \tag{4}$$

for $\mathbf{x} \in S_0$ and $t \in [0, T]$. Details on the smoothing preprocessing of the data are described in Sect. 3. The time-varying functions $\{\psi_1, \ldots, \psi_K\}$ are unknown, each function describing a time profile for mobile phone activity. The surfaces $\{d_1, \ldots, d_K\}$ are the *unobserved* realizations of the random fields $\{D_1, \ldots, D_K\}$. The K values $\{d_1(\mathbf{x}), \ldots, d_K(\mathbf{x})\}$ represent the contributions to the Erlang time profile $E_{\mathbf{x}}$ of their coupled time-varying functions. The other way round, the K time-varying functions $\{\psi_1, \ldots, \psi_K\}$ express the evolution in time of the coupled surfaces. Estimation and interpretation of these two sets of functions is of interest to the urban planner, since they jointly describe both people behavior in time and population density dynamics in space.

2.1 Dimensional reduction: a short tutorial on Treelet analysis (TA)

We first consider the problem of estimating the set of functions $\{\psi_1, \ldots, \psi_K\}$. This means finding a parsimonious description of the sample of Erlang time profiles $\{E_{\mathbf{x}}(t)\}_{\mathbf{x} \in S_0}$ via a finite set of reference profiles. For the moment we are not considering the spatial dependence which is intrinsic in our data set: we will deal with it in the next subsection.

Many approaches are commonly considered to perform dimensional reduction of functional data, the most common being undoubtedly FPCA. Even though FPCA is a powerful method to find optimal subspaces for representing data, it presents the drawback of being a global method, not suitable for multi-resolution analysis, since each basis element most often results in a linear combination of all the primitive variables. On the other hand, wavelets (see Mallat 1989) are commonly used to generate a localized and multi-scale basis for data representation. Their main limitation is that

the wavelet basis is not data-driven, since basis elements are fixed, regardless of the data.

The use of a *treelet basis*, introduced in Lee et al. (2008), is an efficient and recent approach that avoids these drawbacks: treelets are a multi-scale data-driven basis yielding a hierarchical tree that, at each level, represents data through an orthonormal basis. This data-driven basis seems the most suited to the analysis of Erlang data, which present extremely localized functional features, since treelets have been originally designed and developed for treating sparse unordered data. Their property is to have a hierarchical structure, since they are a multiscale orthonormal basis indexed on a hierarchical tree. Indeed, as in multi-resolution analysis, treelets provide a set of "scaling functions" defined on the nested subspaces $\mathbb{R}^J = V_0 \supset V_1 \supset \cdots \supset V_J$, and a set of orthogonal "detail functions" defined on residual spaces $\{W_j\}_{j=1}^J$, where $V_j \oplus W_j = V_{j-1}$ for all $j = 1, \ldots, J$. We remark that treelets are very close to wavelets, even though they are not a wavelet basis. Indeed, in treelet computation, the wavelet approach is mixed with principal component analysis, which is hierarchically performed on the couple of most correlated variables at any given level. At each level of the tree, these are identified and replaced by a coarse-grained *sum variable*, and by a residual *difference variable*: the new variables are computed by a local principal component analysis in two dimensions. Difference variables are then stored, and only sum variables are processed at higher levels of the tree.

More precisely, consider a generic functional sample $\chi_1, \ldots, \chi_N$ and J time instances $t_1, \ldots, t_J$. The algorithm described in Lee et al. (2008) is initialized with the sample design matrix $\mathbb{X} \in \mathbb{R}^{N \times J}$. In our functional specification, $\mathbb{X}$ is the evaluation matrix obtained by setting $\mathbb{X}_{ij} = \chi_i(t_j)$, for $i = 1, \ldots, N$ and $j = 1, \ldots, J$. In the language of Treelet analysis, for each $j = 1, \ldots, J$, we interpret $\chi_1(t_j), \ldots, \chi_N(t_j)$ as a sample from the variable $\chi(t_j)$. Note that in most functional data analyses, each function χ is observed only at discrete time points, often with error. It is thus common to identify the set of functions $\chi_1, \ldots, \chi_N$ by properly smoothing these discrete data and then evaluating each function on the same suitable time grid $t_1, \ldots, t_J$. We will give more details on the smoothing of Erlang data in Sect. 3.

After initializing the set of sum variables with the original variables $\chi(t_1), \ldots \chi(t_J)$, the algorithm proceeds in the construction of the tree by removing at each iteration the two most correlated variables from the set of sum variables, and by replacing them with the associated first principal component. The second principal component, i.e. the difference variable, is stored along iterations. The algorithm is stopped when the set of sum variables is empty, thus returning the set of difference variables $\{\varphi_1, \ldots, \varphi_J\}$, each represented by a vector in $\mathbb{R}^J$. In our functional specification, the vectors of this set are interpreted as the evaluation of a set of functions $\{\varphi_1(t), \ldots, \varphi_J(t)\}$ at time instances $t_1, \ldots, t_J$. For further details on treelet decomposition see Lee et al. (2008).

The output of the algorithm allows for the choice of any subset of difference variables, which in turn will generate a proper linear subspace of $\mathbb{R}^J$. For instance, the first $L \leq J$ difference variables generate the space $W_1 \oplus \ldots \oplus W_L$. In our application, the estimation of the set of functions $\{\psi_1(t), \ldots, \psi_K(t)\}$ is indeed accomplished by using the complete treelet basis $\{\varphi_1(t), \ldots, \varphi_J(t)\}$, according to a criterion detailed in the next subsections.

2.2 Bagging Voronoi Treelet analysis (BVTA)

The model expressed in Eq. (4) relates the observed functional signal to a linear combi-nation of a set of time-varying functions, each time-varying function contributing to the signal observed in a specific site of the lattice according to the value assumed in that site by a coupled surface. We can exploit the strategy described in the previous subsection and based on treelet decomposition for decoupling observed functional data into their constitutive parts. Indeed, we can directly apply the treelet basis decomposition to the N-dimensional sample of Erlang data $\{E_{\mathbf{x}}(t)\}_{\mathbf{x} \in S_0}$, and then select a K-dimensional subset of the complete treelet basis as an estimate for $\{\psi_1(t), \dots, \psi_K(t)\}$. The cou-pled surfaces $\{d_1(\mathbf{x}), \dots, d_K(\mathbf{x})\}$ will then be obtained by site-wise projection of the Erlang data on the estimates of $\{\psi_1(t), \dots, \psi_K(t)\}$. In the rest of the paper, we will refer to this strategy as *Treelet analysis* (TA).

The drawback of this approach is that it does not take into account spa-tial dependence, neither in the estimation of $\{\psi_1(t), \dots, \psi_K(t)\}$, nor in that of $\{d_1(\mathbf{x}), \dots, d_K(\mathbf{x})\}$. Due to the continuity in space of the phenomenon they capture, spatial dependence is intrinsic to the Erlang data. Hence, we develop a novel approach for the identification of the functions ψ and the coupled surfaces d by integrating the treelet decomposition with a proper treatment of spatial dependence. A comparison between this novel approach and TA for dimensional reduction of functional data indexed by a lattice is discussed in the Supplementary material (see the description of the Supplementary material in Sect. 6) in the light of the results of simulation studies.

We will take into account spatial dependence by following a Bagging Voronoi strategy along the lines illustrated in Secchi et al. (2013). The rationale beyond this strategy is simple, but effective: (i) replace the original data set with a reduced one, composed by local representatives of subsets of data belonging to neighborhoods covering the entire investigated area; (ii) analyze the local representatives; (iii) repeat the previous analysis many times for different reduced data sets associated to different randomly generated systems of neighborhoods, thus obtaining many different weak formulations of the analysis; (iv) finally, bag together the weak analyses to obtain a conclusive strong analysis.

At each iteration of the first part of the algorithm, called Bootstrap Step, we generate a partition of the considered region in random neighborhoods, which are used to com-pute local representatives. Each representative is a summary of the data belonging to the same element of the partition, and it is computed as a weighted mean with Gaussian isotropic weights [i.e., by performing spatial smoothing Banerjee et al. (2004)], even though other strategies are conceivable (Secchi et al. 2013). The sample of functional local representatives exploits a specific structure of spatial dependence, and it is usu-ally less noisy and less spatially dependent (Secchi et al. 2013). By applying the TA strategy to the sample of local representatives, one obtains a coarse estimate of a refer-ence basis. The coarse estimate of the coupled surfaces is then obtained by projecting each local representative on the estimated basis, and by assigning the corresponding scores to all sites of the lattice belonging to the element of the partition associated to the considered representative. After B replicates of this weak analysis, the intermediate output of the algorithm consists of:

– a collection of reference bases $\left\{\varphi_1^b(t), \ldots, \varphi_J^b(t)\right\}_{b=1}^B$;
– a collection of sets of surfaces $\left\{d_1^b(\mathbf{x}), \ldots, d_J^b(\mathbf{x})\right\}_{b=1}^B$.

The second part of the algorithm, the Aggregation Step detailed in the next subsection, *bags* together this intermediate output obtaining a final reference basis, estimate of the time-varying functions $\{\psi_1(t), \ldots, \psi_K(t)\}$, and an estimate of the coupled surfaces $\{d_1(\mathbf{x}), \ldots, d_K(\mathbf{x})\}$. Larger values of B imply a higher accuracy of the final estimate.

The proposed procedure is sketched in the pseudocode scheme in Fig. 2. We named this procedure *Bagging Voronoi Treelet Analysis* (BVTA), since it is based on bagging, it uses Voronoi tessellations to compute random partitions of the considered area, and it uses treelets to perform dimensional reduction.

Note that one has to fix some parameters in advance: n, the number of elements of each random partition and of the sample of functional local representatives; B, the number of bootstrap replicates; $d(\cdot, \cdot)$, the most proper metric to measure distances in the considered region. While B can be tuned in order to achieve the desired accuracy, and $d(\cdot, \cdot)$ has to be chosen according to the particular application at hand, the choice of n is somehow more tricky and has great influence on the algorithm performances. In Secchi et al. (2013) the tuning of this parameter has been extensively studied in the light of simulations, and for the specific purpose of minimizing misclassification error. Nevertheless, the study pointed out a quite general conclusion, which is valid also when the Bagging Voronoi strategy is employed for purposes different from clustering, as it is the case for the present paper. The optimal choice of n is the one that finds a good compromise between variance and bias: as n decreases, noise is reduced in the local representatives sample, since local representatives are weighted sample means calculated on sub-samples that are larger on average (minimal variance), but at the same time the associated Voronoi tessellation becomes coarser, thus including less homogeneous data in the computation of local representatives (maximal bias). On the other hand, as n increases, the resulting Voronoi tessellation becomes more and more refined, being able to catch very localized spatial features (minimal bias), but at the same time the variability reduction due to spatial smoothing is smaller (maximal variance). The optimal value of n determined by this trade-off depends both on the strength of spatial dependence, and on the distribution of the spatial signal generating the functional data. In Secchi et al. (2013) a tuning criterion for this parameter had been proposed for the purposes of clustering; in Sect. 4 we propose a criterion based on the total average variance, and more suited to the purposes of dimensional reduction. More insight on this novel criterion is given in the Supplementary material (see Sect. 6), where its performance is discussed and analysed in the light of the results of simulation studies.

2.3 Aggregation step: 1-median alignment for basis matching

We will here give the details of the Aggregation Step in the BVTA algorithm sketched in Fig. 2, whose aim is to aggregate together the B coarse results obtained in the Bootstrap Step. In the context of the present analysis, this means aggregating sets of treelet basis

Fig. 2 Pseudocode scheme of the BVTA algorithm

functions and of their coupled surfaces. The aggregation strategy illustrated in the following lines is a discrete variation of the Procrustes alignment procedures described for instance in Ramsay and Li (1998), James (2007), Kaziska and Srivastava (2007), Sangalli et al. (2009).

The BVTA algorithm is centered on data-driven bases, i.e., treelets. Hence, a match-ing of the elements of the B different bases generated by the Bootstrap Step of the

algorithm is in order before computing the final reference basis. Different approaches to basis matching are possible. We develop a procedure for 1-median basis alignment, which jointly computes the reference basis from the B coarse bases, while also reorder-ing their elements. This procedure is inspired by the joint clustering and alignment method described in Sangalli et al. (2010), where a Procrustes continuous alignment is integrated in a k-mean clustering strategy, to jointly meet the two tasks of assigning curves to a group, while simultaneously aligning them to the corresponding group prototype. In the context of bases matching, each object is a multivariate functional data (one of the coarse bases), and we look for the unique prototype (the reference basis) which best describes the set of functional objects, while also aligning their components, by permutations in the order of basis functions.

Consider the collection of all bases obtained in the Bootstrap Step, $\{\varphi_1^b, \ldots, \varphi_J^b\}_b^B$, and choose a proper measure $\tilde{d}(\cdot, \cdot)$ for the distance (or dissimilarity) between two bases, which in our application will be the $L^1([0, T]; \mathbb{R}^J)$ distance. Both the choice of the L^1 measure for computing the distance between bases, and the choice of the mul-tivariate functional median as reference basis, are motivated by the need of preserving in the aggregation step the sparse nature of treelets.

The 1-median basis alignment is an iterative algorithm, which is initialized by ran-domly selecting a reference basis $\{\tilde{\varphi}_1^{[0]}, \ldots, \tilde{\varphi}_J^{[0]}\}$, among the B coarse bases generated by the Bootstrap Step of the BVTA algorithm. The following two basic steps are then iterated until convergence (consider the l-th iteration, $l > 0$):

1. *Alignment step* For each of the B coarse bases, by permutation of their components, find the best matching to the reference basis $\{\tilde{\varphi}_1^{[l-1]}, \ldots, \tilde{\varphi}_J^{[l-1]}\}$ according to the measure $\tilde{d}(\cdot, \cdot)$. For $b = 1, \ldots, B$, let $\{k_1^{b,[l]}, \ldots, k_J^{b,[l]}\}$ be the permutation of the order of the elements in the basis $\{\varphi_1^b, \ldots, \varphi_J^b\}$ minimizing the distance to the current reference basis;

2. *Estimation step* Given the B reordered bases, let the new reference basis be that whose j-th element is

$$\tilde{\psi}_j^{[l]} = \operatorname*{argmin}_{\varphi \in L^1(T)} \sum_{b=1}^B \tilde{d}(\psi_{k_j^{b,[l]}}^b, \varphi), \quad j = 1, \ldots, J.$$

Since $\tilde{d}(\cdot, \cdot)$ is the $L^1([0, T]; \mathbb{R}^J)$ distance, the reference basis $\{\tilde{\varphi}_1^{[l]}, \ldots, \tilde{\varphi}_J^{[l]}\}$ is the functional median of the B reordered bases.

The algorithm is stopped at iteration $\bar{l}$ after two subsequent iterations with no reordering of the basis elements, for all bases. The final reference basis is thus identified as $\{\tilde{\varphi}_1, \ldots, \tilde{\varphi}_J\} \equiv \{\tilde{\varphi}_1^{[\bar{l}]}, \ldots, \tilde{\varphi}_J^{[\bar{l}]}\}$.

For $b = 1, \ldots, B$, set $\{\pi_1^b, \ldots, \pi_J^b\}$ to be the final permutation $\{k_1^{b,[\bar{l}]}, \ldots, k_J^{b,[\bar{l}]}\}$ and let $\{\tilde{d}_1^b(\mathbf{x}), \ldots, \tilde{d}_J^b(\mathbf{x})\} \equiv \{d_{\pi_1^b}^b(\mathbf{x}), \ldots, d_{\pi_J^b}^b(\mathbf{x})\}$, for $\mathbf{x} \in S_0$. For $j = 1, \ldots, J$, we now compute the sample variance $\tilde{s}_j^2$ of the sample $\{\tilde{d}_j^b(\mathbf{x})\}_{\mathbf{x} \in S_0, b=1,\ldots,B}$, whose size is $N \times B$. For estimating the time-varying functions $\{\psi_1, \ldots, \psi_K\}$ we take the K elements of the basis $\{\tilde{\varphi}_1, \ldots, \tilde{\varphi}_J\}$ associated to the K largest variances among $\{\tilde{s}_1^2, \ldots, \tilde{s}_J^2\}$, and we call these elements $\{\hat{\psi}_1, \ldots, \hat{\psi}_K\}$.
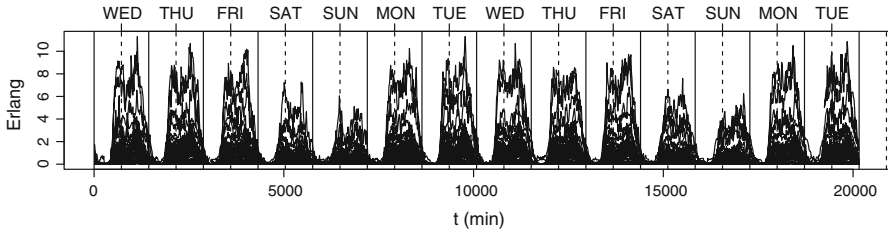
Fig. 3 Selection of 100 Erlang data, randomly drawn among the sites of the lattice, as a function of time. The *solid vertical lines* are drawn at midnight of each day, and the *dotted vertical lines* at noon. The first day is Wednesday March 18, 2009

Indeed, for each given $\mathbf{x} \in S_0$, the same indexes identifying $\{\hat{\psi}_1, \ldots, \hat{\psi}_K\}$ among the elements of the basis $\{\tilde{\varphi}_1, \ldots, \tilde{\varphi}_J$, also point to a collection of K surface sets among the sequence of surface sets $\{\tilde{d} \ \}$ $(\mathbf{x})\}_{b=1}^B$. Let $\hat{d}_k(\mathbf{x})$ be the mean of the k-th selected surface set, for $_1^b(\mathbf{x})\}_{b=1}^B, \ldots, \{\tilde{d}_J^b$ $k = 1, \ldots, K$. We take the surface $\{\hat{d}_k(\mathbf{x})\}_{\mathbf{x} \in S_0}$ to be an estimate of the surface $\{d_k(\mathbf{x})\}_{\mathbf{x} \in S_0}$ coupled with the time-varying function ψ_k (t).

Like in many other dimensional reduction methods, the right value for K is decided by giving consideration to the fraction of total variance explained by each component.

3 Data analysis: preprocessing

The Erlang data described in Sect. 1 are an instance of spatially dependent functional data, indexed by the sites of a spatial lattice. In each given site, the discrete sequence of Erlang values can be considered as a sampling of a continuous process in time (see Ke and Wang 2001), describing the average number of mobile phones using the network in that site, as expressed in Eq. (2). An example of the observed Erlang profiles along time is shown in Fig. 3: 100 sites have been randomly selected in the lattice, and the Erlang measurements recorded in each selected site have been plotted as a function of time. It can be observed in the picture that, beside a periodic behavior due to night/day alternation in the average use of mobile phone, Erlang data present strongly localized features. Moreover, the average intensity of the Erlang profile can be very different from one site to another.

Indeed, in each site of the lattice we observe a discrete version of the Erlang continuous process, recorded approximately every quarter of an hour: due to discontinuities in the information provided by the network antennas, the Erlang measure is missing at some time instances, and hence the time grid of Erlang measurements is non-uniform. Moreover, some Erlang recordings are negative due to measurement errors, and should be treated as missing values. We thus need to choose a proper basis expansion to reconstruct the functional form of the time-varying Erlang data, and to evaluate them on a common uniform grid of time values of dimension $J = 200$, before applying the methodology presented in Sect. 2.

For an extensive description of smoothing procedures for functional data we refer to Ramsay and Silverman (2005). In our application, we perform a site-wise smoothing
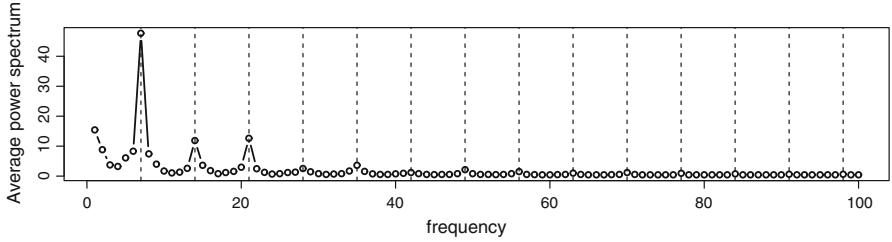
Fig. 4 Average power spectrum $\overline{P}(h)$ obtained via site-wise smoothing of the Erlang measures with a Fourier basis of dimension $H = 200$. Only the values of $\overline{P}(h)$ for $h = 1, \ldots, 100$ are shown in the *plot*. *Dotted vertical lines* are drawn for multiples of 7

of the Erlang data via a Fourier basis expansion, due to the evident seasonality in the Erlang profiles. We set the period of the Fourier basis equal to 1 week: hence, the reconstructed functional form of the Erlang profile for site $\mathbf{x} \in S_0$ is a function $E_{\mathbf{x}}(t)$ such that

$$E_{\mathbf{x}}(t) = \frac{c_0^{\mathbf{x}}}{2} + \sum_{h=1}^{H} \left[a_h^{\mathbf{x}} \cos(h\omega t) + b_h^{\mathbf{x}} \sin(h\omega t) \right], \tag{5}$$

where $t \in [0; T]$, $\omega = 2\pi/T$ and $T = 60 \times 24 \times 7$ is the period expressed in minutes. In the following, the periodic terms in the Fourier basis expansion oscillating at frequency $\omega, 2\omega, 3\omega, \ldots$ will be referred to as first, second, third, ... harmonic, respectively. The coefficients $c_0^{\mathbf{x}}$, $a_h^{\mathbf{x}}$ and $b_h^{\mathbf{x}}$, are estimated by means of least squares.

The basis dimension H should be carefully tuned: it has to be chosen large enough to ensure that the very localized features (sudden peaks, oscillations, ...) which characterize this kind of data (see Fig. 3) are properly caught by the smoothing procedure. To select the basis dimension, we analyze the *power spectrum* associated to the site-wise smoothing of the Erlang data with a Fourier basis of large dimension $H = 200$. The power spectrum of the Fourier expansion of a signal represents the amplitude of the signal as a function of the frequency, and at the h-th frequency it is related to the amplitude of the h-th harmonic

$$P_{\mathbf{x}}(h) = \sqrt{(a_h^{\mathbf{x}})^2 + (b_h^{\mathbf{x}})^2}. \tag{6}$$

Hence, the more the h-th harmonic is relevant in the explanation of features occurring in the data, the more $P_{\mathbf{x}}(h)$ will be large. A local maximum in the power spectrum detects the frequency of a harmonic explaining relevant features in the data. When the power spectrum vanishes towards zero, there is no need to include higher frequency harmonics.

In each site we thus obtain a power spectrum from the site-wise smoothing of the Erlang measurements. We choose the most proper value of H by inspecting the shape of the average power spectrum over all sites of the lattice, i.e., $\overline{P}(h) = \frac{1}{N} \sum_{\mathbf{x} \in S_0} P_{\mathbf{x}}(h)$. The average power spectrum of the Telecom Italia database is reported in Fig. 4.

A graphical inspection of the plot makes it clear that the frequencies significantly contributing to the Erlang time variation are the smaller ones (all less than 7), capturing differences among days or blocks of days (e.g., the working and weekend days variation), and the multiples of 7, capturing the recurring daily dynamics. Due to the huge dimension of the Telecom Italia database, we choose a basis of very high dimension, in order to be reasonably sure to catch all relevant localized features. The picture indicates that, for frequencies higher than 100ω, the average power spectrum is negligible: we thus set $H = 100$ for subsequent analyses.

4 Data analysis: results

The smooth functions obtained by the preprocessing of the Erlang data are then analyzed along the method presented in Sect. 2. In particular, the analysis has been performed for different values of the dimension n of the Voronoi tessellation, ranging from 50 to 2,500 elements. For each value of n, $B = 50$ random Voronoi maps have been used in the Bootstrap Step of the BVTA algorithm. The metric $d(\cdot, \cdot)$ for generating the random Voronoi maps is the Euclidean distance on the plane, after having flattened the inspected area using the international WGS84 UTM 32N geographical system map. Local representatives are identified as weighted means with Gaussian isotropic weights.

As discussed in Sect. 2, choosing the right value for n is a delicate issue, since the optimal value of this parameter is strongly related to the spatial dependence occurring between data recorded in different sites. In Secchi et al. (2013), the choice of n is driven by the idea that a good value for n would be the one providing stable results of the performed analysis across bootstrap replicates. In that work a cluster analysis is performed, and thus the concept of bootstrap stability refers to cluster assignment of each site across replicates. To measure stability the authors introduced a criterion that averages over the entire area a pixel-wise measure of the entropy of the cluster assignment distribution along bootstrap replicates. A thorough simulation study on the use of this entropy-based criterion as a strategy for the choice of n has been conducted in Secchi et al. (2013), where its performances are discussed and explored in the light of simulation studies.

Analogously to Secchi et al. (2013), we here define a measure of stability across replicates which is coherent to the aims of the present analysis. An intermediate output of the BVTA algorithm consists of the collection of the B surface sets $\{\tilde{d}_1^b(\mathbf{x}), \ldots, \tilde{d}_J^b(\mathbf{x})\}_{b=1}^B$, each surface being coupled with an element of the reference basis $\{\tilde{\varphi}_1(t), \ldots, \tilde{\varphi}_J(t)\}$. Indeed the final estimates of the time-varying functions $\{\psi_1(t), \ldots, \psi_K(t)\}$ and of their coupled surfaces $\{d_1(\mathbf{x}), \ldots, d_K(\mathbf{x})\}$ are exclusively based on this intermediate output, that we therefore require to be stable with respect to the choice of n. To measure stability, for each site $\mathbf{x} \in S_0$ and $j = 1, \ldots, J$, we compute the bootstrap variance of the sample $\{\tilde{d}_j^b(\mathbf{x})\}_{b=1}^B$. We then average over $\mathbf{x} \in S_0$, and sum over $j = 1, \ldots, J$. We call this quantity *Total Average Variance (TAV)*: its minimization is the criterion driving the choice for n. A small value of TAV implies that for each site and for each element of the reference basis we attain stable scores across bootstrap replicates. A more thorough simulation study to show the validity
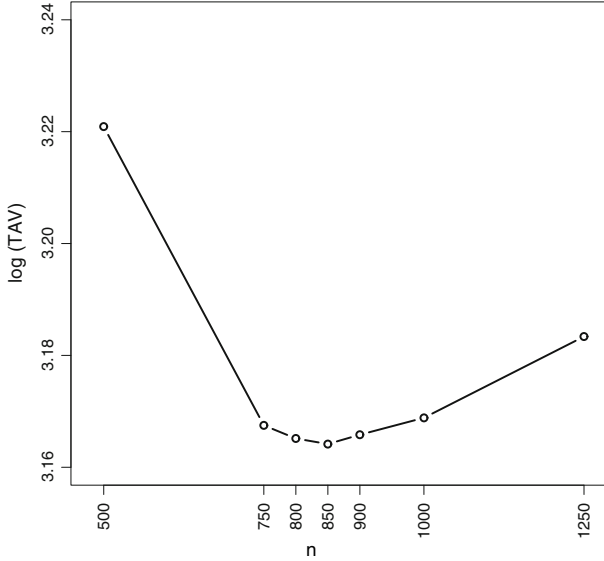
Fig. 5 Total average variance (log scale) for the BVTA of the Erlang data as a function of n

in the use of the TAV criterion as a strategy for the choice of n in the context of the BVTA strategy is included in the Supplementary material (see Sect. 6). The results of the simulations show the goodness of this criterion for selecting n, and they also prove the better estimate of the time-varying functions $\{\psi_1(t), \ldots, \psi_K(t)\}$ obtained with the selected n.

In Fig. 5 the logarithm of TAV is reported as a function of n. A minimum is observed for $n = 850$, that is thus the dimension of the Voronoi tessellation used to run the BVTA algorithm for the analysis of the Erlang data. This dimension of the Voronoi tessellation is associated to an average area of the Voronoi elements equal to $0.77\,\text{km}^2$, that corresponds to the area of a circle of diameter nearly equal to 1 km. This indicates that spatial dependence is relevant up to this distance, and thus reveals 1 km to be the practical spatial range of our data.

The approach described at the end of Sect. 2.3 for selecting the right number K of treelets is visually captured by Fig. 6: the height of the bars in the figure corresponds to fraction of the total variance $\tilde{s}_j^2$ explained by each treelet score, for $j \leq 60$; for $j > 60$ we may safely assume that treelets are representing only noise. By inspecting the plot, one is reasonably supported in setting a threshold for the fraction of total variance, leading to a possible value of K: for instance, if we set the threshold to 0.2 %, this is overtaken by a couple of dozens treelets, which is therefore the selected dimension for the reference basis. It is worthy of notice that the time-varying functions thus selected, and their coupled surfaces, are also easily interpretable. We here discuss just four of them, that we deem to be particularly interesting for illustrating the peculiar properties of the analysis conducted by means of the BVTA algorithm:

- the general mobile phone activity along time function $\hat{\psi}_1$;
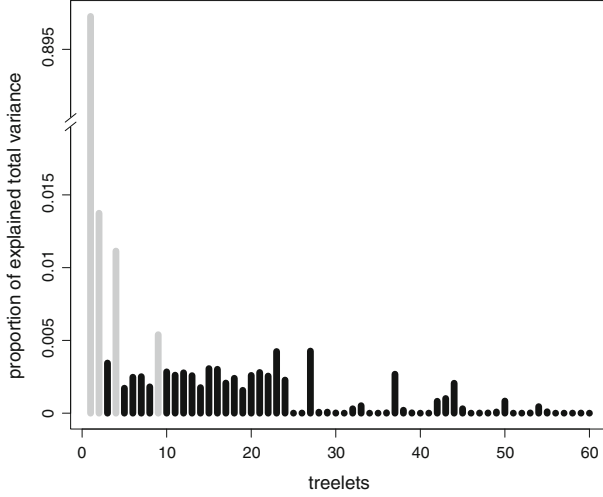- the working/non-working time function $\hat{\psi}_2$;

Fig. 6 Proportion of the total variance $\tilde{s}_j^2$ explained by each treelet score, for $j \leq 60$, output of the BVTA algorithm for analyzing the Erlang data. The *grey bars* correspond to the four treelets whose interpretation is the most relevant, and on which we focus in the rest of the paper

– the rush-hour time function $\hat{\psi}_4$;
– the International Fair at the Rho-Fiera exhibition site time function $\hat{\psi}_9$.

The first function corresponds to a global feature of our dataset, the second illustrates a static activity, the third one a dynamic activity, and the fourth a spot event concentrated in space and time. Figure 7 reports the temporal pattern of the basis elements $\{\hat{\psi}_1(t), \hat{\psi}_2(t), \hat{\psi}_4(t), \hat{\psi}_9(t)\}$ over a week-period starting from Wednesday 00:00 and ending with Tuesday 24:00. Full vertical lines separate different days while dotted lines are reported every two hours to help the reader. Their coupled surfaces $\{\hat{d}_1(\mathbf{x}), \hat{d}_2(\mathbf{x}), \hat{d}_4(\mathbf{x}), \hat{d}_9(\mathbf{x})\}$ are represented in Fig. 8. A value close to 0 in a particular site of the map means that the corresponding reference basis element does not significantly contribute to the Erlang signal measured in that site. On the contrary, a positive/negative large value on the map means that the corresponding reference basis element significantly contributes to the Erlang signal in that site, with sign coherent to the score sign. The 0-level contour lines are traced in bold. Figure 9 zooms on the city center of the previous maps. In the remaining part of the Section, we give more details on the interpretations of $\{\hat{\psi}_1(t), \hat{\psi}_2(t), \hat{\psi}_4(t), \hat{\psi}_9(t)\}$ and of their coupled surfaces, with the aim of illustrating the type of information about the city dynamics that can be drawn from our analysis.

4.1 General mobile phone activity along time function $\hat{\psi}_1$

The estimated surface $\{\hat{d}_1(\mathbf{x})\}$ coupled with $\hat{\psi}_1$ is reported in the top-left panel of Fig. 8. This map catches the urbanization of the area, clearly pointing out day-time low-density population areas and day-time highly-populated areas. We thus relate $\{\hat{d}_1(\mathbf{x})\}$ to the average population density, and the reference basis element $\hat{\psi}_1$ to the
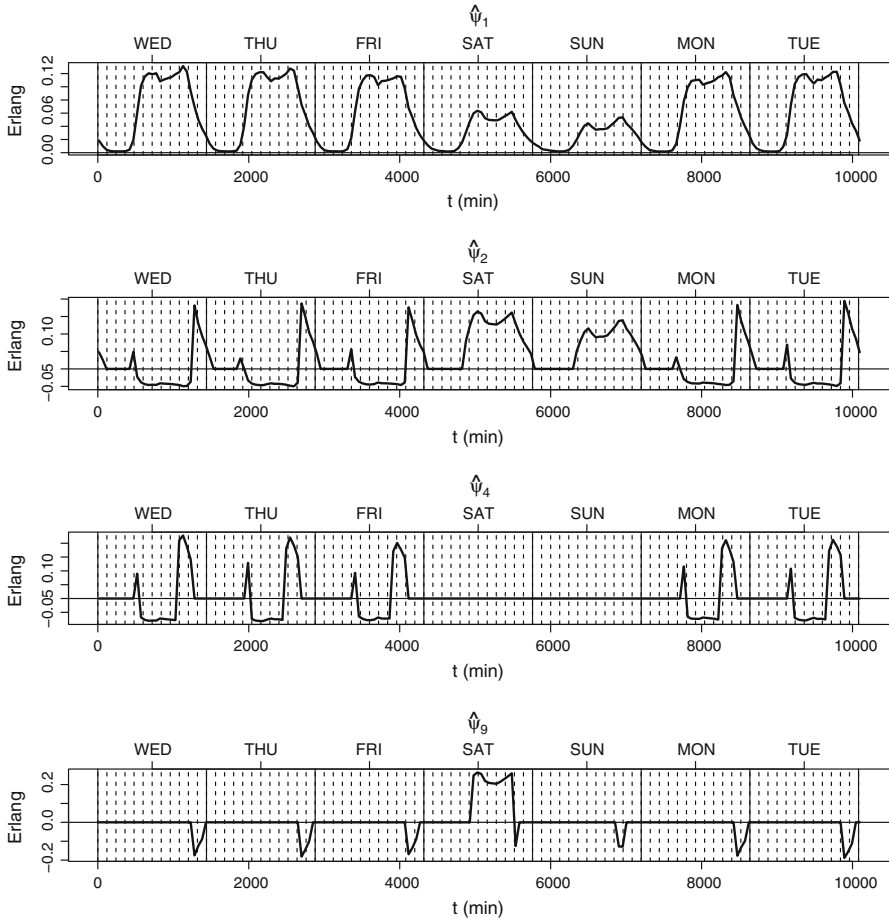
Fig. 7 Some selected elements of the reference basis (from top to bottom: $\hat{\psi}_1$, $\hat{\psi}_2$, $\hat{\psi}_4$, and $\hat{\psi}_9$) output of the BVTA algorithm analyzing the Erlang data

general mobile phone use. The general mobile phone activity along time function is the most important in terms of magnitude, in the sense that it is the one presenting the largest contribution to the Erlang signals of many highly active sites. It can be indeed detected even through much simpler analyses, that do not take into account spatial dependence, or even by simply looking at a random sample of curves (e.g., Fig. 3). As it is evident by inspection of Fig. 7 (top panel), this reference basis element is always switched on with positive sign with significant values between 7:00 a.m. and 2:00 a.m. in working days and between 8:00 a.m. and 2:00 a.m. in weekend days. It describes daily and weekly periodicity. In particular it points out a larger activity during day-time with respect to night-time, a bi-modal behavior of the daily signal, and confirms Milan to be an attractor during the day-time of working days. This is clear from the lower level observed during the weekend.
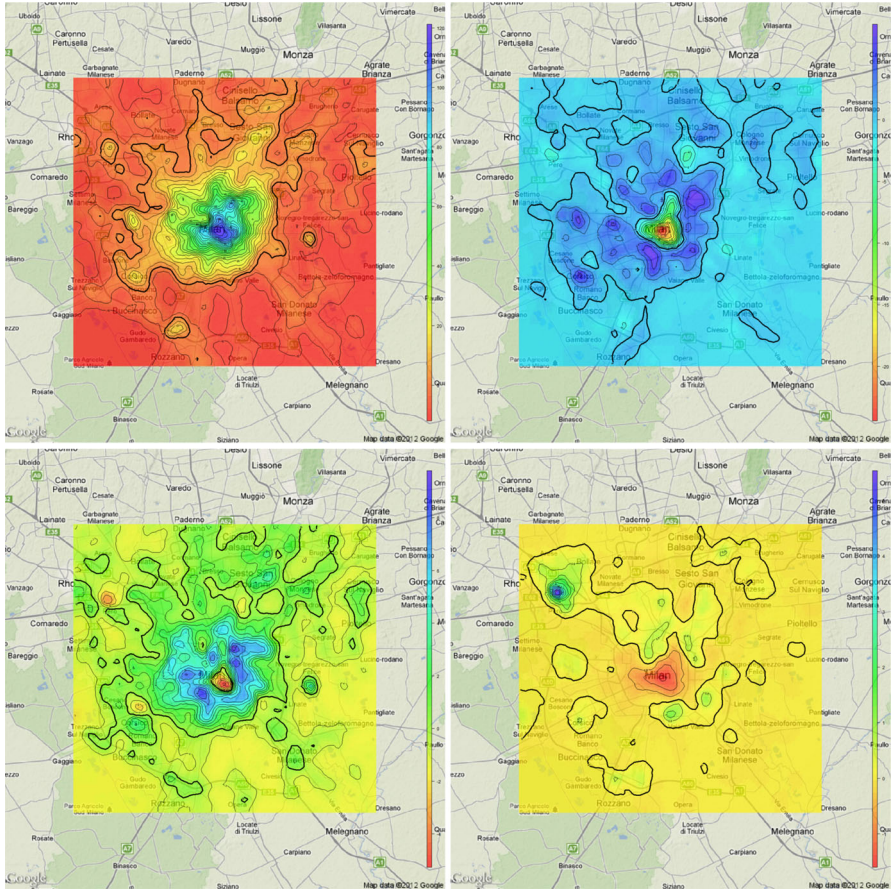
Fig. 8 From *top* to *bottom*, and then from *left* to *right*, maps of the estimated surfaces $\{\hat{d}_1(\mathbf{x}), \hat{d}_2(\mathbf{x}), \hat{d}_4(\mathbf{x}), \hat{d}_9(\mathbf{x})\}$ coupled to the reference basis elements reported in Fig. 7. The 0-level *contour lines* are reported in *bold*

4.2 Working/non-working time function $\hat{\psi}_2$

Looking at Fig. 7 (second panel from the top), we notice that this function contrasts working-time (i.e., from 8:30 a.m. to 8:00 p.m. of working days) against non-working time (i.e., from 7:00 a.m. to 8:30 a.m. and from 8:00 p.m. to 2:00 a.m. of working days, and day-time of week-end days). Positive values on the relevant map (top-right panel in Fig. 8) indicate high activity during non-working-time and a reduced activity during working-time, and viceversa for negative values. The map clearly spots the historical center connected with a northeast offshoot toward the Central Railway Station, areas mostly devoted to tertiary activities, and where the resident population density is very low. Then, a donut-shaped area around the city center, mostly covering residential or leisure areas, emerges with high positive values. Moving further from the city center, the values of $\hat{d}_2$ tend to vanish except for some non-working hours spots corresponding
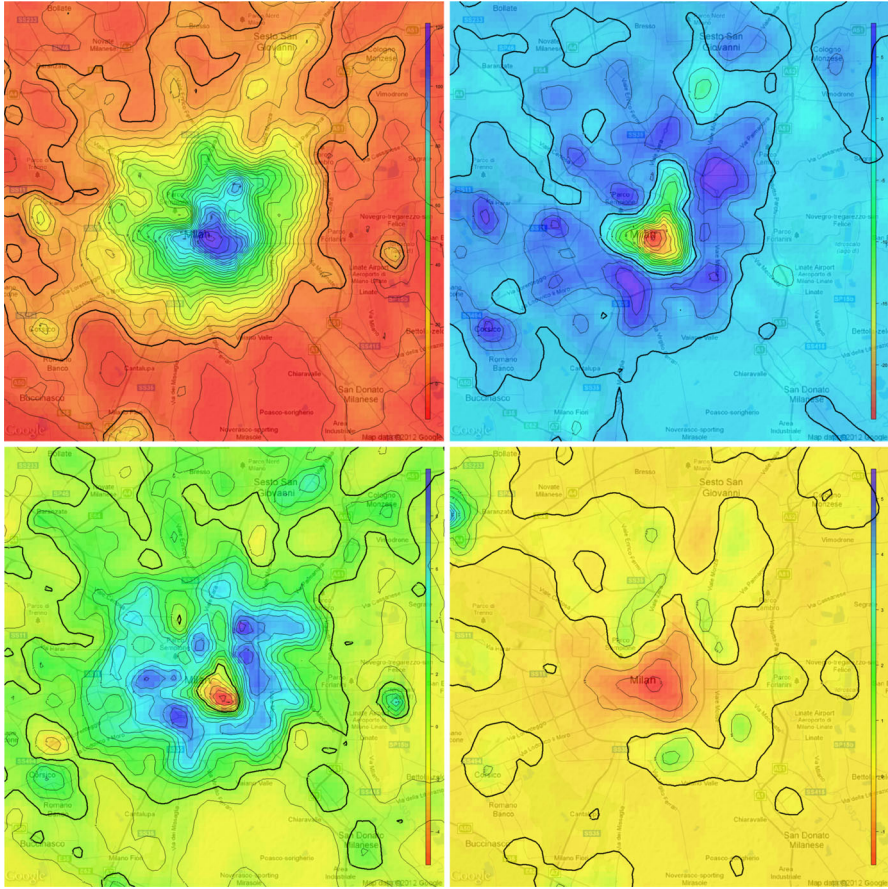
Fig. 9 Zooms on the city center of the maps already shown in Fig. 8

to satellite towns right outside the city of Milan. Moreover, one can observe a working hours spot in the north direction corresponding to the Bicocca neighborhood, that is a renewed area in the outskirts of Milan mostly devoted to tertiary and to university-related activities. This basis element presents the city center as an attractor during the working hours and the outskirts and the satellite towns as attractors during the non-working hours. This can possibly be explained by the daily population density dynamics, determined by movements of people from their residence to their working place and backward.

4.3 Rush-hour time function $\hat{\psi}_4$

Positive scores with respect to this function point out areas where an high activity is present between 8:00 a.m. to 10:00 a.m. and 5:00 p.m. to 9:00 p.m., which correspond to the morning and evening rush hours (see the second panel from the bottom in Fig. 7).

Inspection of the coupled surface $\hat{d}_4$ in the bottom-left panel of Fig. 8, shows that areas particularly active during rush hours are concentrated around the third ring-road within the city (Circonvallazione Esterna), at the Central Railway Station, along arteries connecting the city with the satellite towns, along some segments of the highway ring-road, and in Linate Airport (the eastern spot on the map). It is also interesting to note the hole in the very city center corresponding to the congestion charge area, which is restricted only to local traffic during weekdays.

4.4 International fair at the Rho-Fiera exhibition site time function $\hat{\psi}_9$

This function contrasts the Saturday activity carried on between 10:00 a.m. and 8:00 p.m. and everyday dinner time (see the bottom panel in Fig. 7), and seems strongly related to the activity, during non-working time, connected to the International Fair at the Rho-Fiera exhibition site. This event has been held between the 18th and the 23rd March, 2009, at the Fiera Milano Exhibition Complex North-West of the city; the activities related to this occasional but highly attractive event clearly affect the Erlang measurements in the time period covered by our data. Indeed, in the bottom-right panel of Fig. 8, the Fiera Milano Exhibition Complex can be easily located by the positive peak in $\hat{d}_9$, while a corresponding negative peak is observed in the city center. This is possibly explained in the light of the interpretation given to $\hat{d}_2$, by a flow of people spending Saturday at the Exhibition site and dinner and after-dinner in the city center.

5 Conclusions

In this work, a real case study concerning the dimensional reduction of spatially dependent functional data, describing the number of mobile phones simultaneously using the Telecom Italia mobile network for calling at a given time, has been described. This work is a first innovative attempt to gather information about population density dynamics in Milan from mobile network data belonging to the Telecom Italia database. We believe that the applicability and impact of the proposed analysis are broad, both to the purposes of the Green Move project and, more generally, for future urban planning and development.

The methodology developed to perform dimensional reduction of spatially-dependent functional data is an innovative integration between a treelet analysis (see Lee et al. 2008) and the Bagging Voronoi strategy for the exploration of spatial dependence (see Secchi et al. 2013), and it is thus named *Bagging Voronoi Treelet Analysis*. We exploit the potentialities of both techniques, improving on the results that can be obtained using the original methods alone. The method is proven useful in the applicative context of interest, and in a simulated scenario close to the real one (see the Supplementary material).

Further research developments concern both the treatment of spatial dependence, and the dimensional reduction technique. The former can be improved by considering, either in the random generation of the set of nuclei for the tessellation, or in the distance $d(\cdot, \cdot)$ used to compute the Voronoi elements, relevant information concerning the area under investigation. For instance, the diffusion tensor describing the traffic flow

could be used to define a city-adapted measure of distance, thus obtaining Voronoi elements capable of "following" the flow of people. For what concerns the latter aspect, the dimensional reduction strategy can be modified by removing the orthogonality constraint implicit in the use of treelets for the construction of the reference basis. Indeed we are currently working on the analysis of mobile network data by means of Independent Component Analysis (Secchi et al. 2014), and on the integration of ICA within the Bagging Voronoi strategy: this will be the focus of future work.

6 Supplementary material

The supplementary material to the present paper contains a detailed description of two simulation studies, conducted in order to investigate the properties and performances of the method detailed in Sect. 2.

The simulations are aimed at supporting the use of TAV as a criterion for selecting the optimal value for n, and at checking whether the best estimate of the time-varying functions is obtained for the selected value of n. Moreover, a comparison with the results obtained by a standard TA, which does not take into account spatial dependence, is also shown.

Both simulation studies are designed to support these claims in a simulated scenario as close as possible to the case study at hand, and where the functional patterns generating the data are complex enough to be untractable with standard parametric models. In the first simulation, in order to evaluate the estimates provided by the TA and the BVTA strategies in optimal conditions, the time-varying component are set orthogonal. The second simulation, instead, is aimed at testing the robustness of the BVTA strategy when the set of time-varying components is no longer a set of orthogonal functions. In both cases, the coupled spatial surfaces are generated according to a Hidden Markov Random Field Model (Kunsch et al. 1995).

References

Banerjee S, Carlin B, Gelfand A (2004) Hierarchical modeling and analysis for spatial data. Monographs on statistics and applied probability. Chapman & Hall, London

Becker RA, Caceres R, Hanson K, Loh JM, Urbanek S, Varshavsky A, Volinsky C (2011) A tale of one city: using cellular network data for urban planning. IEEE Pervasive Comput 10(4):18–26

Calabrese F, Lorenzo GD, Liu L, Ratti C (2011) Estimating origin-destination flows using mobile phone location data. IEEE Pervasive Comput 10(4):36–44

James GM (2007) Curve alignment by moments. Ann Appl Stat 1:480–501

Kaziska D, Srivastava A (2007) Gait-based human recognition by classification of cyclostationary processes on nonlinear shape manifolds. J Am Stat Assoc 102:1114–1128

Ke C, Wang Y (2001) Semiparametric nonlinear mixed-effects models and their applications. J Am Stat Assoc 96:1272–1298

Kunsch H, Geman S, Kehagias A (1995) Hidden markov random fields. Ann Appl Probab 5(3):577–602

Lee AB, Nadler B, Wasserman L (2008) Treelets—an adaptive multi-scale basis for sparse unordered data. Ann Appl Stat 2(2):435–471

Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 11:674–693

Manfredini F, Pucci P, Secchi P, Tagliolato P, Vantini S, Vitelli V (2015) Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the Milan urban region. In: Paganoni AM, Secchi P (eds) Advances in complex data modeling and computational methods in statistics., Contributions to statisticsSpringer, Berlin, pp 133–147

OECD (2006a) OECD Territorial reviews: competitive cities in the global economy. OECD Publishing, Paris

OECD (2006b) OECD Territorial reviews: Milan, Italy. OECD Publishing, Paris

Ramsay JO, Li X (1998) Curve registration. J R Stat Soc Ser B Stat Methodol 60:351–363

Ramsay JO, Silverman BW (2005) Functional data analysis. Springer, Berlin

Sangalli LM, Secchi P, Vantini S, Veneziani A (2009) A case study in exploratory functional data analysis: geometrical features of the internal carotid artery. J Am Stat Assoc 104:37–48

Sangalli LM, Secchi P, Vantini S, Vitelli V (2010) K-mean alignment for curve clustering. Comput Stat Data Anal 54:1219–1233

Secchi P, Vantini S, Vitelli V (2013) Bagging voronoi classifiers for clustering spatial functional data. Int J Appl Earth Obs Geoinf 22:53–64

Secchi P, Vantini S, Zanini P (2014) Hierarchical independent component analysis: a multi-resolution non-orthogonal data-driven basis. In: Tech Rep 01/2014, MOX—Dipartimento di Matematica, Politecnico di Milano