The Annals of Applied Statistics
2016, Vol. 10, No. 2, 812–833
DOI: 10.1214/16-AOAS913

UNDERSTANDING RESIDENT MOBILITY IN MILAN THROUGH INDEPENDENT COMPONENT ANALYSIS OF *TELECOM ITALIA* MOBILE USAGE DATA

BY PAOLO ZANINI[1], HAIPENG SHEN[2] AND YOUNG TRUONG[3]

MOX at Politecnico di Milano, University of Hong Kong and University of North Carolina at Chapel Hill

We consider an urban planning application where *Telecom Italia* collected mobile-phone traffic data in the metropolitan area of Milan, Italy, aiming to retrieve meaningful information regarding working, residential, and mobility activities around the city. The independent component analysis (ICA) framework is used to model underlying spatial activities as spatial processes on a lattice independent of each other. To incorporate spatial dependence within the spatial sources, we develop a spatial colored ICA (scICA) method. The method models spatial dependence within each source in the frequency domain, exploiting the power of Whittle likelihood and local linear log-spectral density estimation. An iterative algorithm is derived to estimate the model parameters through maximum Whittle likelihood. We then apply scICA to the Italian mobile traffic application.

1. Introduction. Urban planning is the technical procedure for developing urban areas, designing buildings, streets and other infrastructures related to transportation/logistic distribution networks. Several processes affect the everyday life in an urban area. In particular, working, residential and mobility activities are crucial for the well-being of an urban area. For instance, as highlighted in Sheller and Urry (2006) and Kaufmann (2012), changes in management of resident mobility are crucial to understanding temporal and spatial modes of social life, thus structuring the urban areas.

The current paper concerns urban planning issues for the metropolitan area of Milan, located in the North of Italy, and in particular connects with the *Green Move* Initiative [Luè et al. (2012)], which aims at designing a 2nd generation electric vehicle sharing system for the city. Milan is the most populated province of

Received October 2015; revised January 2016.

[1]Supported in part by *Green Move*, a joint research program involving MOX—Laboratory for Modeling and Scientific Computing (Department of Mathematics, Politecnico di Milano) and funded by Regione Lombardia.

[2]Supported in part by US National Science Foundation (NSF), US National Institutes of Health (NIH) and University of Hong Kong Stanley Ho Alumni Challenge Fund.

[3]Supported in part by US NSF DMS-07-07090, DMS-11-06962 and NIH.

Key words and phrases. Spatial stochastic processes, periodogram, Whittle likelihood, mobile phone traffic, urban planning.

Italy and the fifth largest metropolitan area of Europe in terms of number of inhabitants, with a density of more than 1000 inhabitants per km^2. The urban area of Milan provides about 10% of the Italian gross domestic product. An *Organization for Economic Co-operation and Development* review [OECD (2006)] identified housing, transportation and congestion as the principal limitations for future development of Milan.

To help understand such issues, there is pressing need for a deep quantitative analysis of the main features about working, residential and mobility activities. Ratti et al. (2006) voiced interests in analytical answers to several long-standing questions in architecture and urban planning such as follows:

- How do you map vehicle origins and destinations?
- How do you understand the patterns of pedestrian movement?
- How do you highlight critical points in the urban infrastructure?
- What is the relationship between urban forms and flows?

Our research in this paper will help shed lights on such critical questions, and has general applicabilities for other large metropolitan areas as well.

Urban activities are difficult to be directly recorded due to their complexity. Several experimental studies [Ahas and Mark (2005), González, Hidalgo and Barabási (2008), Ratti et al. (2006)] have pointed out that mobile phone usage data offer indirect measurements with great potentials to study population behaviors and monitor urban dynamics in real time. The earlier studies are qualitative in nature. To complement them, we quantitatively analyze mobile usage data provided through the courtesy of *Telecom Italia*, the biggest mobile company in Italy, as part of the *Green Move* Initiative, through a research agreement between *Telecom* and *Politecnico di Milano*. The data describe the mobile traffic on the urban area of Milan, anonymously recorded as the average number of simultaneous contacts in a time unit on a tessellation of the territory in rectangular areas (i.e., pixels). Hence, the data can be regarded as a collection of spatially distributed signals varying along time. More details of the data are provided in Section 5.

As discussed above, the problems of interest are to identify interesting significant spatial-temporal features of resident mobility through analyzing the spatial-temporal mobile usage data. The resident mobility features of interest are unfortunately *hidden*. Hence, we view the application as a blind source separation (BSS) problem, and approach it using independent component analysis (ICA) [Comon and Jutten (2010), Hyvärinen and Oja (2000)]. The classical ICA application is the cocktail-party problem, where several microphones in a room record sound signals produced by independent sources. The goal then is to recover the original audio signals through such recordings.

By analogy, our current application is to decompose the mobile usage data as a mixture of *spatial sources that are independent of each other* (i.e., urban activities) mixed across time. More specifically, we anticipate to localize resident

activities to specific spatial regions, that are also distinct across components, describing different resident mobility behaviors that vary across time. For example, the numerical results (Section 5) identify the financial district in the city center, major railway stations in the suburban area and major outflow streets from the city center, respectively, as the three major spatial independent components.

The nature of our application suggests that each spatial source exhibits certain spatial dependence across the spatial lattice in the Milan area [Manfredini et al. (2015), Secchi, Vantini and Vitelli (2015)]. Such spatial dependence should be incorporated in the ICA procedure. Many ICA methods have been developed, as well as documented by Comon and Jutten (2010) and Lee et al. (2011). We highlight several major statistical approaches in the last decade: Chen and Bickel (2005, 2006), Eloyan, Crainiceanu and Caffo (2013), Guo (2011), Hastie and Tibshirani (2010), Matteson and Tsay (2016). As we are aiming at extracting independent spatial components, there exist some relevant spatial ICA methods in the fMRI literature that identify independent brain maps. [See, e.g., Calhoun et al. (2001) and van de Ven et al. (2004).]

However, to the best of our knowledge, no existing ICA methods (including the spatial ICA ones) incorporate spatial dependence *within* the (spatial) sources. The existing spatial ICA methods assume spatial independence within each independent component. The colored ICA (cICA) method of Lee et al. (2011) is the closest one that is motivated by the temporal ICA of fMRI data and takes into account temporal dependence within brain signals. To apply cICA to the mobile data, we need to first transpose the data matrix (for spatial ICA), and then concatenate the 2-dimensional (2D) spatial map recorded at each time point into a 1-dimensional vector (1D), before applying cICA to estimate the resulting 1D dependence structure. The concatenation destroys the inherent 2D spatial dependence in the original data, which is undesirable especially when the dependence is directional.

Hence, to address the spatial dependence challenge in the application, we develop spatial colored independent component analysis (scICA) to directly incorporate the 2D spatial dependence, which extends the (spatial) ICA literature. We model the spatial sources as independent spatial processes on a lattice, and derive the Whittle likelihood [Whittle (1952)] in the frequency domain. The model parameters are then estimated through maximizing the Whittle likelihood iteratively. The spectral densities of the sources are modeled nonparametrically using local linear log-spectral density estimation [Fan and Kreutzberger (1998)].

A similar iterative estimation algorithm is implemented in Lee et al. (2011). We point out two major differences between our algorithm and theirs. We model the spectral densities sources nonparametrically. This enables more flexible modeling of spatial dependence within each independent component, comparing to the parametric approach of Lee et al. (2011). This also removes the difficulty of order selection for parametric spatial processes. Furthermore, in the step of estimating the unmixing matrix, we can not use the Lagrange multiplier method of Lee et al.

(2011) due to ill-conditioning caused by the nonparametric spectral density estimation. Instead, we implement an alternative optimization algorithm that makes use of eigen decomposition. Details are provided in Section 3.

The rest of the paper is organized as follows. In Section 2, we introduce our model, providing some details about spatial processes on lattices and their spectral density functions, as well as Whittle likelihood. Then, in Section 3, we describe the iterative estimation algorithm for maximizing the Whittle likelihood. We apply scICA to two simulation studies to show the advantages of scICA in Section 4. As a comparison, we consider cICA and (spatial) fastICA [Hyvärinen and Oja (2000)], one of the most popular ICA methods. Section 5 concludes the paper analyzing the Milan mobile data. We also discuss the unique insights obtained by scICA. The numerical studies nicely highlight the advantage of incorporating spatial dependence. Given the popularity of ICA in the fMRI literature, one future research direction is to extend scICA to incorporate 3-dimensional spatial dependence so that the method can be applied to analyze fMRI data. (The motivating mobile data are distributed on a 2-dimensional spatial surface.)

All the numerical studies are carried out using the `R` statistical software [R Core Team (2013)]. Furthermore, we develop an accompanying `R` package that implements the cICA and scICA methods, named `coloredICA` and available on the CRAN website—http://cran.r-project.org/web/packages/coloredICA/index.html.

2. Spatial colored independent component analysis. This section presents the modeling details of our proposed spatial colored ICA (scICA) method. Under the ICA framework, we model the spatial independent sources as spatial fields on a spatial lattice that are independent of each other, where each spatial source exhibits certain dependence structures. We start with a brief mathematical formulation of ICA in Section 2.1. We then introduce some basic concepts and notation about spatial stochastic processes in Section 2.2. Finally, we derive the Whittle likelihood [Whittle (1952)] for the model based on all the independent sources in Section 2.3.

2.1. *Mathematical formulation of ICA.* Mathematically, ICA can be expressed using the following latent variable linear model:

$$X_j = c_{j1}S_1 + \cdots + c_{jp}S_p, \qquad \text{for } j = 1, \ldots, p, \tag{1}$$

where the random variables $X_1, \ldots, X_p$ are observed linear combinations of p unknown (latent) random independent sources $S_1, \ldots, S_p$. In vector-matrix form, model (1) becomes

$$\mathbf{X} = C\mathbf{S}, \tag{2}$$

where $\mathbf{X}$ and $\mathbf{S}$ are random vectors in $\mathbb{R}^p$, and C is a (p, p) *mixing* matrix.

In applications, we observe $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p$, realizations of the random vector $\mathbf{X}$, while the corresponding realizations of $\mathbf{S}$ are unknown as well as the mixing

matrix C. Let $\mathbb{X}$ be the (n, p) data matrix whose rows are the n observations in the sample. Then we can rewrite (2) as

$$\mathbb{X} = \mathbb{S}C', \tag{3}$$

where $\mathbb{S}$ is a (n, p) matrix containing the n unknown realizations of the p sources. For blind source separation, one usually first estimates the unmixing matrix $W = C^{-1}$, given $\mathbb{X}$, and then recovers the hidden sources $\mathbb{S}$ through

$$\mathbb{S} = \mathbb{X}W'. \tag{4}$$

For our motivating mobile-usage data, the matrix $\mathbb{X}$ contains the traffic intensity measurements at p spatial grids across n time points, while the matrix $\mathbb{S}$ includes the p underlying spatial independent sources.

2.2. *Spatial stochastic processes on a lattice.* We assume the sources $S_1, \ldots, S_p$ as weakly stationary spatial processes. For a specific source S_j, let $\mathbf{s} \in \mathbb{R}^2$ be a spatial generic location in a 2-dimensional Euclidean space. Suppose that the potential datum $S(\mathbf{s})$ at the location $\mathbf{s}$ is a random quantity. Note that, for simplicity, we omit the subindex j here and beyond in places where the omission doesn't cause ambiguity.

Consider a spatial index set $D \subseteq \mathbb{R}^2$, a fixed regular collection of countably many spatial points, say, $D = \{\mathbf{s} = (u, v)' : u = \ldots, -1, 0, 1, \ldots; v = \ldots, -1, 0, 1, \ldots\}$. As $\mathbf{s}$ varies over the index set D, we obtain the spatial random field

$$\big\{S(\mathbf{s}); \mathbf{s} \in D\big\}. \tag{5}$$

In this case (5) is called a spatial process on a lattice [Cressie (1993)]. A realization of (5) is denoted as $\{s(\mathbf{s}); \mathbf{s} \in D\}$.

For a weakly stationary process $\{S(\mathbf{s})\}$, its covariance function $C_S(\mathbf{u})$ is defined, for every $\mathbf{u} \in \mathbb{Z}^2$, as

$$C_S(\mathbf{u}) = \operatorname{Cov}\big(S(\mathbf{s}+\mathbf{u}), S(\mathbf{s})\big), \qquad \forall \mathbf{s} \in D.$$

If the covariance values form an absolutely summable sequence, then we can define its Fourier Transform as

$$f_S(\boldsymbol{\omega}) = \frac{1}{(2\pi)^2} \sum_{\mathbf{u} \in \mathbb{Z}^2} C(\mathbf{u}) e^{-i\mathbf{u}'\boldsymbol{\omega}},$$

with $(\omega_1, \omega_2)' = \boldsymbol{\omega} \in \Pi^2 = [-\pi, \pi] \times [-\pi, \pi]$. The function $f_S(\boldsymbol{\omega})$ is the spectral density of the stochastic process $S(\mathbf{s})$. The covariance function at lag $\mathbf{u}$ can be recovered by the Inverse Fourier Transform of the spectral density as

$$C_S(\mathbf{u}) = \int_{\Pi^2} f_S(\boldsymbol{\omega}) e^{i\mathbf{u}'\boldsymbol{\omega}}\, d\boldsymbol{\omega}.$$

Therefore, the covariance and spectral density functions form a Fourier pair. A detailed description of spatial stochastic processes and their properties can be found, for instance, in Gebizlioğlu (1988) and Cressie (1993).

2.3. *The Whittle likelihood.* The approach of the Whittle likelihood was first introduced by Whittle (1952) in the context of time series, which was later extended into spatial processes. The Whittle likelihood is defined in the frequency domain through the use of the periodogram, that is, sample spectral density [Box and Jenkins (1970)]. Under the ICA framework, below we first introduce the Whittle likelihood for a single source, which is then used to derive the complete Whittle likelihood based on all sources.

Consider a spatial process S observed on a regular grid $D = \{\mathbf{s} = (s_1, s_2) : s_1 = 0, \ldots, n_1 - 1; s_2 = 0, \ldots, n_2 - 1\}$, $D \in \mathbb{R}^2$, $n = n_1 \cdot n_2$. The corresponding spatial periodogram at a frequency $\boldsymbol{\omega} \in \Pi^2$ is given by

$$I(\boldsymbol{\omega}, S) = \frac{1}{(2\pi)^2 n} \left| \sum_{\mathbf{s} \in D} S(\mathbf{s}) \exp(-i\mathbf{s}'\boldsymbol{\omega}) \right|^2.$$

The periodogram is usually computed at the set of bidimensional Fourier frequencies $\boldsymbol{\omega}_k = (\omega_{k_1}, \omega_{k_2})$:

$$\omega_{k_1} = \frac{2\pi k_1}{n_1} \qquad k_1 = 0, \pm 1, \ldots, \pm m_1 \qquad \text{where } m_1 = \left\lceil \frac{(n_1 - 1)}{2} \right\rceil,$$

$$\omega_{k_2} = \frac{2\pi k_2}{n_2} \qquad k_2 = 0, \pm 1, \ldots, \pm m_2 \qquad \text{where } m_2 = \left\lceil \frac{(n_2 - 1)}{2} \right\rceil,$$

where $\lceil \cdot \rceil$ denotes the smallest integer not less than the operand.

As discussed in Whittle (1952), the spatial periodogram $I(\boldsymbol{\omega}, S)$ has an asymptotic distribution of $f_S(\boldsymbol{\omega})\chi^2_2/2$. Furthermore, through an approximation of the quadratic form of the log-likelihood, the negative Whittle log-likelihood can be defined as

$$L(f_S; S) = \int_{\Pi^2} \left(\frac{I(\boldsymbol{\omega}, S)}{f_S(\boldsymbol{\omega})} + \log f_S(\boldsymbol{\omega}) \right) d\boldsymbol{\omega}. \tag{6}$$

We refer to Whittle (1952) for more details on the derivation of (6) and to Crujeiras (2006) for a detailed description of the 2D case. In Crujeiras (2006) it is also highlighted that the negative log-likelihood (6) can be interpreted as the Kullback–Leibler divergence between the periodogram function $I(\omega, S)$ and the spectral density function $f_S(\omega)$. Note that, taking into account the asymptotic independence of $I(\omega, S)$ across different Fourier frequencies $\boldsymbol{\omega}_k$, (6) in practice is approximated by the following discretized version:

$$\sum_k \left(\frac{I(\boldsymbol{\omega}_k, S)}{f_S(\boldsymbol{\omega}_k)} + \log f_S(\boldsymbol{\omega}_k) \right), \tag{7}$$

where the sum extends over all the Fourier frequencies.

We now consider the random vector of the p sources, $\mathbf{S} = (S_1, \ldots, S_p)'$ in $\mathbb{R}^p$, defined on a finite lattice D with n sites. For the jth source, define its spectral density and the periodogram as $f_{S_j}(\boldsymbol{\omega})$ and $I(\boldsymbol{\omega}, S_j)$, respectively. Then, due to

the assumption of independence, we can write the complete Whittle negative log-likelihood as the sum of the Whittle negative log-likelihood for each individual source:

$$L(f_{\mathbf{S}}; \mathbf{S}) = \sum_{j=1}^{p} \sum_{k=1}^{n} \left(\frac{I(\boldsymbol{\omega}_k, S_j)}{f_{S_j}(\boldsymbol{\omega}_k)} + \ln\big(f_{S_j}(\boldsymbol{\omega}_k)\big) \right), \tag{8}$$

where $f_{\mathbf{S}}$ is the diagonal spectral density matrix of the sources (diagonal because the sources are assumed independent).

In practice, we do not observe the sources, but rather the mixed spatial processes $\mathbf{X}$. So, exploiting equation (2), the Whittle negative log-likelihood (8) can be rewritten as

$$\begin{aligned} L(W, f_{\mathbf{S}}; \mathbf{X}) = &\sum_{j=1}^{p} \sum_{k=1}^{n} \left(\frac{\mathbf{e}_j' W' I(\boldsymbol{\omega}_k, \mathbf{X}) W \mathbf{e}_j}{f_{S_j}(\boldsymbol{\omega}_k)} + \ln\big(f_{S_j}(\boldsymbol{\omega}_k)\big) \right) \\ &- n \ln\big|\det(W)\big|, \end{aligned} \tag{9}$$

where $I(\boldsymbol{\omega}_k, \mathbf{X})$ is the matrix periodogram of the mixed signals at the Fourier frequency $\boldsymbol{\omega}_k$, and $\mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)'$ with the jth entry being 1. As a side comment, a typical preprocessing procedure in ICA is to prewhite the data [Hyvarinen, Karhunen and Oja (2001)]. It follows then that W is orthogonal, which allows us to drop the last term in (9).

As one can see, the complete Whittle likelihood (9) is a function of both the unmixing matrix W and the source spectral densities f_{S_j}, for $j = 1, \ldots, p$ (as well as their parameters). Below in Section 3, we discuss the algorithm that derives the parameter estimates through minimizing the negative Whittle log-likelihood (9), that is, maximizing the Whittle likelihood.

3. Parameter estimation. We implement an iterative algorithm, alternating between a step where the source spectral densities are estimated and a step where an estimate $\widehat{W}$ of W is obtained. We consider here an adaptation to the 2D case of the cICA algorithm presented in Lee et al. (2011). We use the same convergence criterion, stopping the algorithm when the difference between $\widehat{W}_{\text{new}}$ and $\widehat{W}_{\text{old}}$ is below a convergence threshold, where the difference is measured through the Amari error, a criterion widely used in the ICA framework.

Consider two $p \times p$ matrices W_1 and W_2 where W_2 is invertible. According to Amari, Cichocki and Yang (1996), we define the Amari error between W_1 and W_2 as

$$\frac{1}{p} \sum_{i=1}^{p} \left(\frac{\sum_{j=1}^{p} |w_{ij}|}{\max_j |w_{ij}|} - 1 \right) + \frac{1}{p} \sum_{j=1}^{p} \left(\frac{\sum_{i=1}^{p} |w_{ij}|}{\max_i |w_{ij}|} - 1 \right),$$

where the w_{ij}'s are the entries of $W_1 W_2^{-1}$.

3.1. *Spectral density estimation.* We consider the unmixing matrix W as fixed and estimate the spectral densities for the sources. Since (9) involves a sum over the p sources, we can obtain estimates of the p spectral densities separately.

We start with describing how to obtain an estimate for a generic f_S through minimizing (7). Different from the cICA algorithm of Lee et al. (2011), we perform nonparametric local linear estimation of the spectral densities based on the discrete approximation (7) [Crujeiras and Fernández-Casal (2010), Fan and Kreutzberger (1998)]. The nonparametric approach offers us greater flexibility to incorporate a wide range of spatial dependence. On the other hand, parametric approaches can be too restrictive and involve a large number of parameters even for spatial processes of low order, not to mention the difficulty of order estimation for the parametric models.

It can be shown that minimizing (7) is equivalent to maximizing

$$\sum_k \bigl(Y(\boldsymbol{\omega}_k, S) - m_S(\boldsymbol{\omega}_k) - e^{Y(\boldsymbol{\omega}_k, S) - m_S(\boldsymbol{\omega}_k)}\bigr),$$

where $Y(\boldsymbol{\omega}_k, S) = \log(I(\boldsymbol{\omega}_k, S)$ and $m_S(\boldsymbol{\omega}_k) = \log f_S(\boldsymbol{\omega}_k)$. We use a multidimensional local linear regression to estimate the log-spectral density function $m_S(\boldsymbol{\omega}_j)$. At each Fourier frequency $\boldsymbol{\omega}_l$, we approximate $m_S(\boldsymbol{\omega}_l)$ using $a_l + \mathbf{b}_l'(\boldsymbol{\omega}_l - \boldsymbol{\omega})$ for $\boldsymbol{\omega}$ in a sufficiently small neighborhood of $\boldsymbol{\omega}_l$. The parameters a_l and $\mathbf{b}_l$ are then estimated by the solution to the following optimization problem:

$$\begin{aligned}\max_{a_l, \mathbf{b}_l} \sum_k \bigl(&Y(\boldsymbol{\omega}_k, S) - a_l - \mathbf{b}_l'(\boldsymbol{\omega}_l - \boldsymbol{\omega}_k) \\ &- e^{Y(\boldsymbol{\omega}_k, S) - a_l - \mathbf{b}_l'(\boldsymbol{\omega}_l - \boldsymbol{\omega}_k)}\bigr) K_H(\boldsymbol{\omega}_l - \boldsymbol{\omega}_k),\end{aligned} \tag{10}$$

where $K_H(\boldsymbol{\omega}) = |H|^{-1/2} K(H^{-1/2}\boldsymbol{\omega})$ is a rescaled bidimensional kernel, with H being a bidimensional bandwidth matrix. Then $\widehat{a}_l$ is the local linear estimate of $m_S(\boldsymbol{\omega}_l)$.

It then follows that, to minimize (9) for fixed W to obtain the estimates for f_{S_j}, $j = 1, \ldots, p$, we can solve (10) by letting $Y(\boldsymbol{\omega}_k, S_j) = \log(I(\boldsymbol{\omega}_k, S_j)) = \log(\mathbf{e}_j' W' I(\boldsymbol{\omega}_k, \mathbf{X}) W \mathbf{e}_j)$, which is known once W is fixed.

3.2. *Unmixing matrix estimation.* We now fix f_{S_j} for $j = 1, \ldots, p$ and discuss how to estimate the unmixing matrix W. Note that one needs to impose orthogonality constraints on W to make the problem identifiable. For that purpose, Lee et al. (2011) use the method of the Lagrange multiplier and minimize a penalized version of (9) through Newton–Raphson updating. Their optimization algorithm involves the inverse of the Hessian matrix, which can be ill-conditioned in our setting due to the nonparametric spectral density estimates.

Hence, we propose an alternative optimization algorithm. In particular, for every $j = 1, \ldots, p$, we minimize

$$\widetilde{L}_j(W, f_{\mathbf{S}}; \mathbf{X}) = \mathbf{w}_j'(A_j + \tau C_j)\mathbf{w}_j, \tag{11}$$

where $\mathbf{w}_j = W\mathbf{e}_j$ is the jth column of W, $A_j = \sum_{k=1}^{n} \frac{I(\boldsymbol{\omega}_k, \mathbf{X})}{f_{S_j}(\boldsymbol{\omega}_k)}$, $C_j = \sum_{l=1}^{j-1} \mathbf{w}_l \mathbf{w}_l'$, and τ is a positive tuning parameter. The matrix C_j provides an orthogonality constraint in the sense that $\mathbf{w}_j' C_j \mathbf{w}_j = \sum_{l=1}^{j-1} \langle \mathbf{w}_j, \mathbf{w}_l \rangle^2$, which is non-negative, with equality when $\mathbf{w}_j$ is orthogonal to $\mathbf{w}_1, \ldots, \mathbf{w}_{j-1}$. Hence, $\mathbf{w}_1$ is estimated first without considering the constraint term C_1, and $\mathbf{w}_2, \ldots, \mathbf{w}_p$ are then estimated taking into account the corresponding orthogonality penalty term.

The alternative criterion (11) provides a straightforward way of estimating $\mathbf{w}_j$. Indeed, it is easy to observe that the matrix $A_j + \tau C_j$ is symmetric and positive-definite. Hence, the argmin of (11) is the eigenvector of $A_j + \tau C_j$ that corresponds to the smallest eigenvalue.

One still needs to choose the tuning parameter τ. To ensure the orthogonality, we begin with an initial (small) value for τ and then proceed in the following alternating manner:

(a) obtain $\widehat{W}$ from minimizing (11);
(b) if the orthogonality error is below a certain threshold, we retain $\widehat{W}$ as the estimate for W; otherwise, repeat Step (a) with the tuning parameter being 2τ.

Note that the orthogonality error is measured by $\|\widehat{W}\widehat{W}' - I_p\|_F$, with $\|\cdot\|_F$ being the Frobenius norm.

REMARK 3.1. The numerical results shown in the rest of our paper are obtained through minimizing the criterion (11). In any case, we point out that, in those situations where the Hessian matrix is not ill-conditioned, the two approaches give similar estimates. In our `R` package, we allow the user to choose either algorithm to estimate the unmixing matrix.

3.3. *Summary of the iterative algorithm.* Our iterative algorithm can be summarized as follows. We begin with some initial estimate $\widehat{W}$. Then, while the Amari error is greater than a certain threshold, we alternate between the following steps:

(1) estimate the source spectral densities through the nonparametric algorithm (10);
(2) update $\widehat{W}$ by minimizing (11).

The iterative algorithm considerably simplifies the minimization of the Whittle negative log-likelihood (9). The same iterative idea has been successfully used in Lee et al. (2011). In all our numerical studies, the algorithm always converges, usually in less than 10 iteration steps. We demonstrate empirical convergence of our algorithm in more details via the simulation studies in Section 4.3. Our estimation algorithm and the one in Lee et al. (2011) are both implemented in our `coloredICA` `R` package. We are currently working on updating the package, exploiting the seamless interaction of `R` with `C++` [see Eddelbuettel and François (2011)]. The update expects to speed up computation by up to 4 or 5 times.

REMARK 3.2. ICA methods are known to have permutation and scale ambiguity problems [Hyvärinen and Oja (2000)]. Scale ambiguity is often avoided thanks to the pre-whitening procedure, which results in sources with unitary variance. Permutation ambiguity is not of concern in a lot of applications. Otherwise one can use the procedure described in Lee et al. (2011) to order the sources during each updating.

4. Simulation studies. In this section we present two simulation studies, comparing scICA with fastICA and cICA. The numerical comparisons clearly show that scICA performs the best. Both simulation studies are carried out on a $n_1 \times n_2$ grid, with $n_1 = n_2 = 20$. In the first simulation, we model the dependence within each spatial source using spatial autoregressive moving-average (SARMA) models [Gebizlioğlu (1988)], while, in the second one, we consider spatial sources with irregular dependence structures.

4.1. *Simulation study* 1: *Symmetric SARMA processes of order one*. The simulation involves two sources and two mixtures. We perform 100 different runs, and, for each run, the entries of the mixing matrix C are randomly generated from the standard uniform distribution.

A SARMA model for $S(\mathbf{s}) \equiv S(u, v)$ is defined as

$$S(u,v) = \sum_{l=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \phi_{lm} S(u-l, v-m) + \sum_{l=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \theta_{lm} \varepsilon(u-l, v-m),$$

where $\phi_{00} = \theta_{00} = 0$, and $\varepsilon(\cdot, \cdot)$ is white noise with zero mean and variance σ^2. Note that a SARMA model can be reduced to a spatial autoregressive (SAR) or spatial moving-average (SMA) model by respectively setting the MA coefficients $\theta_{lm} = 0$ or the AR coefficients $\phi_{pm} = 0$. Certain SAR and SMA models are used in this simulation study as discussed below.

The n realizations of the first spatial source are generated according to the following SAR model of order one:

$$S_1(u,v) = \phi_1 S_1(u-1, v) + \phi_2 S_1(u+1, v) + \varepsilon(u, v),$$

with $\phi_1 = -0.35$, $\phi_2 = 0.7$, and $\varepsilon(u, v)$ being Gaussian with zero mean and variance $\sigma^2 = 0.3^2$. The n realization of the second spatial source are generated according to the following SMA model of order one:

$$S_2(u,v) = \varepsilon(u, v) + \theta_1 \varepsilon(u-1, v) + \theta_2 \varepsilon(u+1, v),$$

with $\theta_1 = 0.38$, $\theta_2 = -0.45$, and Gaussian noise $\varepsilon(u, v)$ with zero mean and variance $\sigma^2 = 0.3^2$. Then we generate the mixing matrix C randomly from the standard uniform distribution, and the data $\mathbb{X}$ according to model (3).

As such, we consider two spatial sources with spatial dependence only in the horizontal direction. Note that, to run cICA on the simulated data, the spatial lattices are vectorized in the vertical direction; hence, this configuration enables us

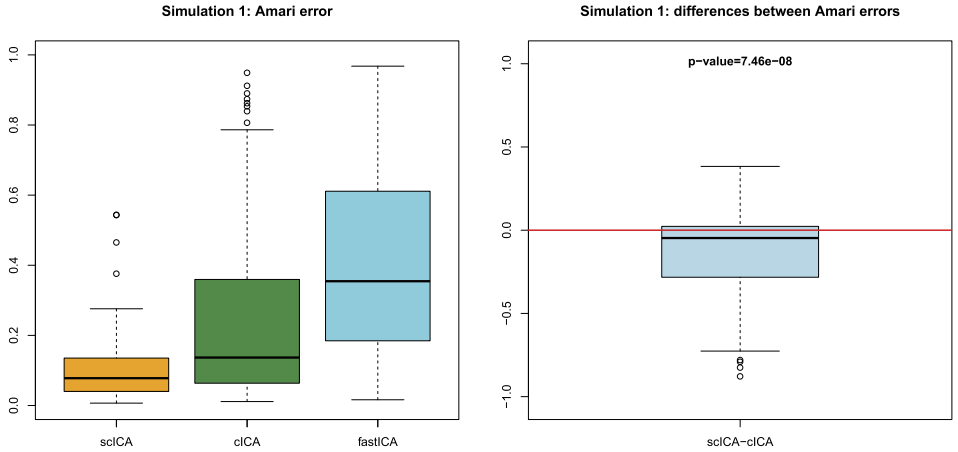

FIG. 1. *Simulation* 1*—Left panel*: *boxplots of the Amari error for the three methods*. *Right panel*: *density plot of the difference of the Amari error between scICA and cICA*.

to investigate whether the unidirectional vectorization necessary for cICA distorts its estimation performance when the spatial dependence exists in a direction that does not align with the vectorization direction. The following results confirm that this is indeed the case, which further supports the necessity of incorporating 2D dependence directly, without vectorization, as in scICA.

The left panel of Figure 1 shows the boxplots of the Amari errors for every method considered. Indeed, scICA outperforms cICA (and both colored methods outperform fastICA). If we consider the pairwise comparison between the two errors for every run, then scICA is indeed significantly better than cICA—the pairwise t-test has a p-value of $\simeq 7 \cdot 10^{-8}$.

In BSS problems, we are interested in not only accurate estimation of the mixing matrix, but also efficient reconstruction of the sources. For this reason, we compare the absolute difference between the recovered sources and the true sources, respectively, by scICA and cICA. (We focus only on the colored methods because fastICA clearly performs worse than the other methods.) Specifically in Figure 2, we plot the mean absolute reconstruction error on the entire lattice, averaged over the 100 runs, for scICA (on the left) and cICA (on the right), using the same color scheme. The top row is for reconstructing Source 1, while the bottom row is for Source 2.

It is evident that scICA has uniformly better accuracy than cICA. For every run, we also evaluate the mean absolute reconstruction error over the 400 pixels of the 20×20 lattice. The p-value for the corresponding pairwise t-test is around 10^{-8} for each source, which provides strong evidence that scICA reconstructs the sources better than cICA.

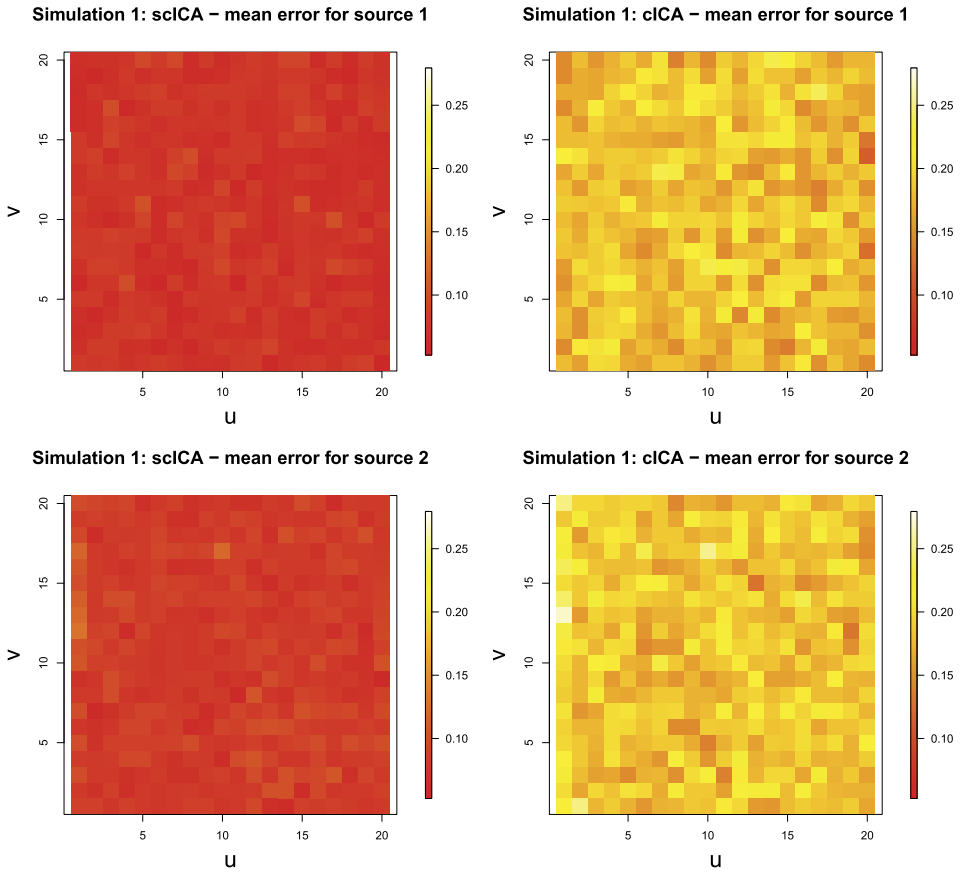

FIG. 2. *Simulation* 1*—Mean absolute reconstruction error of the sources over the* 100 *runs made by scICA* (*on the left*) *and cICA* (*on the right*).

4.2. *Simulation study* 2: *Spatial sources with irregular structures*. We now consider two sources following irregular patterns as shown in Figure 3.

We perform 100 simulation runs. For each run, the source matrix is composed by adding the sources in Figure 3 with independent Gaussian noises of zero mean and different variances for each source: $\sigma_1^2 = 2^2$ for Source 1, and $\sigma_2^2 = 0.1^2$ for Source 2. We then generate the mixing matrix randomly at each run from the standard uniform distribution and obtain the data matrix according to model (3).

The left panel of Figure 4 presents the boxplots of the Amari errors for every method considered. Similar to the first simulation, the two colored methods clearly outperform fastICA. What is different is that, in this case, the improvement of scICA over cICA is much bigger, as evidenced from the right panel of Figure 4 and the extremely small p-value (of $3.05 \cdot 10^{-15}$) from the corresponding pairwise t-test for the Amari error.

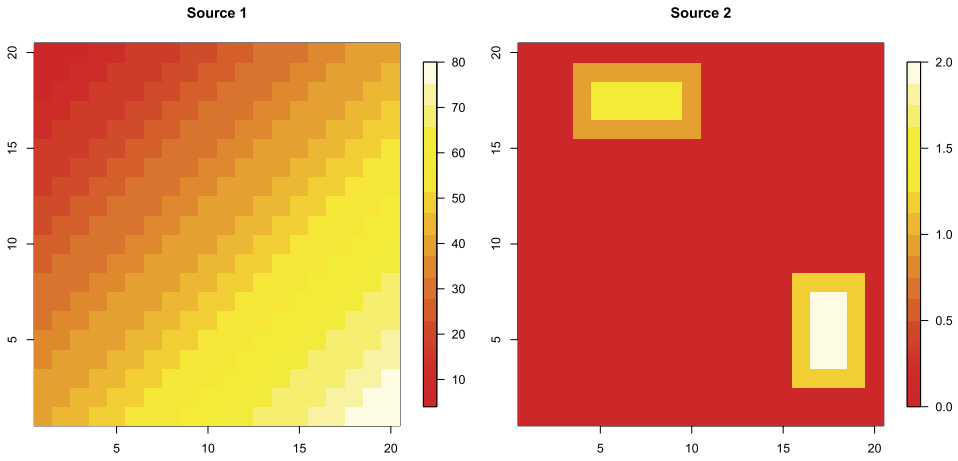

FIG. 3. *Simulation* 2*—The two spatial sources.*

Furthermore, Figure 5 compares the performance for reconstructing the sources, which again shows that scICA significantly outperforms cICA. Note that the two bright rectangle regions in Source 2, shown in Figure 3, are evident in the cICA reconstruction error plots, especially for reconstructing Source 1 (the top row of Figure 5). This suggests that cICA has trouble separating the two spatial sources. Pairwise t-tests also yield extremely small p-values (of $1.67 \cdot 10^{-17}$ and $1.65 \cdot 10^{-18}$, respectively), which again shows that scICA recovers the sources significantly better than cICA.

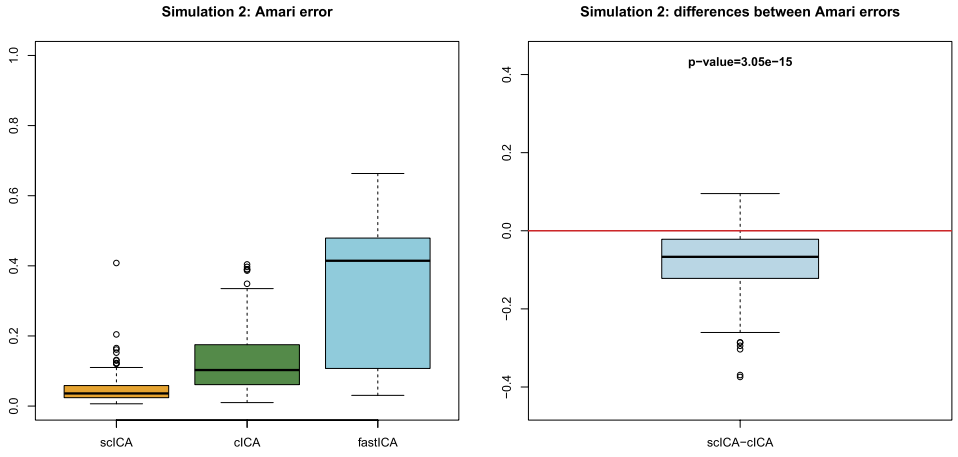

FIG. 4. *Simulation* 2*—Left panel*: *boxplots of the Amari error for the three methods. Right panel*: *density plot of the difference of the Amari error between scICA and cICA.*

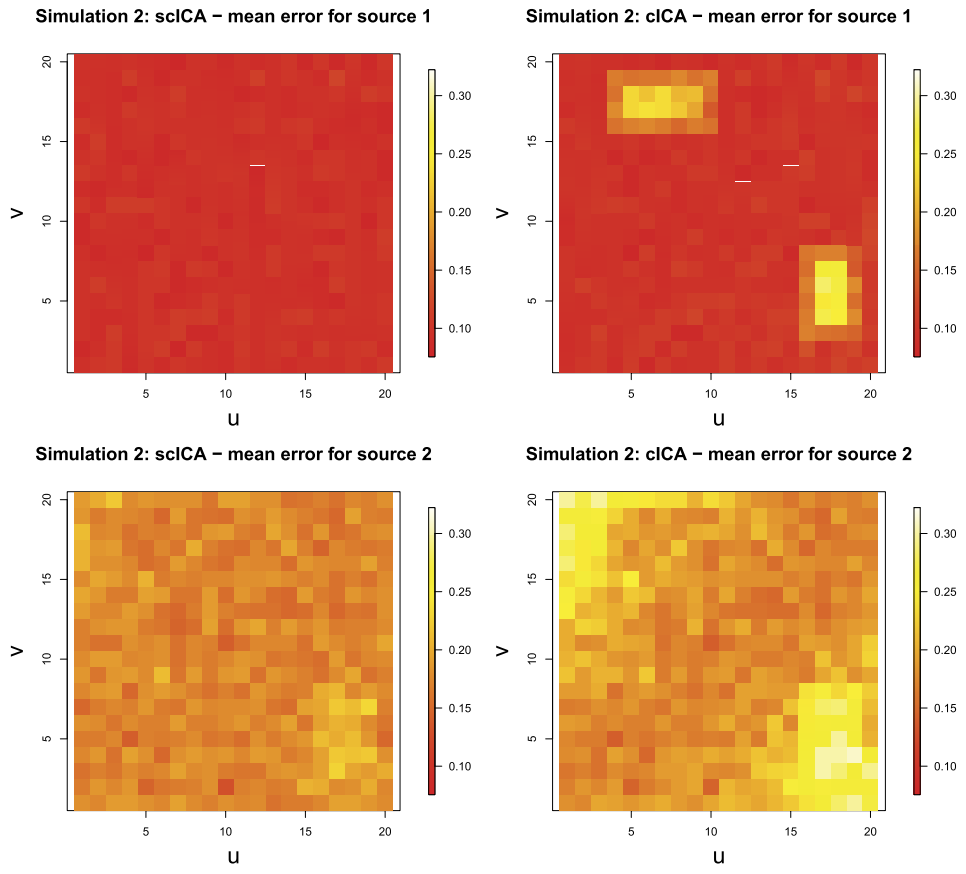

FIG. 5. *Simulation* 2*—Mean absolute reconstruction error of the sources over the* 100 *runs made by scICA* (*on the left*) *and cICA* (*on the right*).

There are two reasons for the great improvement of scICA over cICA for this simulation study. First of all, scICA is specially designed for 2D spatial structures, while cICA is for 1D data. Furthermore, scICA models the dependence through the nonparametric local spectrum approach, which is another difference from the parametric cICA method.

4.3. *Empirical convergence.* We conclude the simulated studies with a brief analysis of the empirical convergence of the scICA method, as mentioned at the end of Section 3. In both simulation studies, using a tolerance level of 10^{-3} for the Amari error, the algorithm usually converges in very few iterations: the average number of iterations needed for convergence is 7 and 3 (rounded), respectively. Furthermore, we show in Figure 6, for both studies, a plot of the Whittle log-likelihood at each iteration for a representative run in order to better illustrate empirical convergence of the algorithm. Each representative run is chosen where

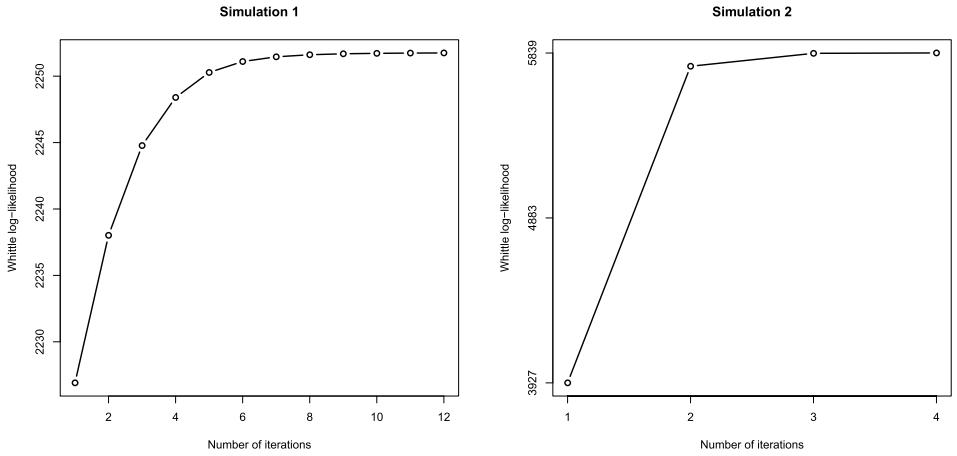

FIG. 6. *Empirical convergence*: *Whittle log-likelihood versus number of iterations*, *for one representative run in Simulation* 1 (*left panel*) *and Simulation* 2 (*right panel*).

the number of iterations needed for convergence is higher than the average in order to avoid showing a particularly "lucky" simulation run.

5. Analysis of the *Telecom Italia* data. In this section, we present the motivating application of analyzing the real mobile-phone traffic data collected in the metropolitan area of Milan (Italy). The data have been previously analyzed in Manfredini et al. (2015) and Secchi, Vantini and Vitelli (2015) through basis representations, where a first attempt to consider spatial dependence is done through a bagging Voronoi approach. In this paper we analyze the dataset in the ICA framework, taking into account the spatial structure of the sources using scICA.

The aim of the analysis is to decompose the traffic patterns as a time-varying linear combination of independent spatial pattern maps. We demonstrate the comparison of scICA with fastICA [Hyvärinen and Oja (2000)] and cICA [Lee et al. (2011)]. The benchmark fastICA is the most popular ICA method in the literature. To perform cICA, we vectorize the 2D observations binding the columns and use the software from Lee et al. (2011). The comparison between cICA and scICA suggests that directly incorporating the 2D dependence gives more meaningful results than only incorporating 1D dependence as in cICA.

It takes about 100 seconds for our current `R` implementation of scICA to produce results. We expect to speed up the computation by a factor of 4 or 5 with the next update, incorporating integration of `R` with `C++`. On the other hand, it takes fastICA and cICA 0.5 and 2 seconds, respectively, to analyze the data.

The reported analysis of this mobile-phone dataset clearly highlights the benefits achieved from using scICA. As we will see, taking into account potential spatial dependence within the sources, scICA obtains more meaningful spatial maps.

This is crucial so that one can exploit these results for urban planning purposes—the clearer the pattern shown in the spatial maps, the easier the association to a specific type of urban activity within the city. Then the information can be used to understand times and areas where, for instance, public transportation should be enhanced or which areas are occupied by more people during specific hours of a day and, hence, more services are needed.

5.1. *Data description*. We analyze here the municipality of Milan divided into a lattice D_0 of 25×28 pixels (232 m $\times$ 309 m each). For each pixel of the covered area we observe the Erlang every 15 minutes for 14 days from March 18th to March 31st, 2009. The Erlang is a dimensionless unit calculated as the sum of the length of every call in a given time interval divided by the length of the interval (i.e., 15 minutes). For each pixel and for each quarter of an hour, this measure represents the average number of mobile phones simultaneously calling through the network. The Erlang can be approximately considered as proportional to the number of active subscribers in that area at that time. Hence, the data describe the number of subscribers in a 2D-space at different instants of time, which can be depicted as a spatial surface varying along time in the two panels of Figure 7.

For illustration purposes, Figure 8 highlights three major spatial structures in Milan that we will discuss later in connection with the ICA results. The biggest circular white curve represents a circular big road that is very crucial for Milan mobility. Indeed, the center of Milan (indicated by the white big circle, also in Figure 9) is a congestion charge area, hence, very few vehicles can go inside. For this reason, the circular big road is the most common way to "pass through" the city. This crucial ring road is identified in the 3rd independent component (IC) extracted by our scICA method (Figure 11). Furthermore, the small white circle in the upper right corner corresponds to the central railway station of Milan (Fig-

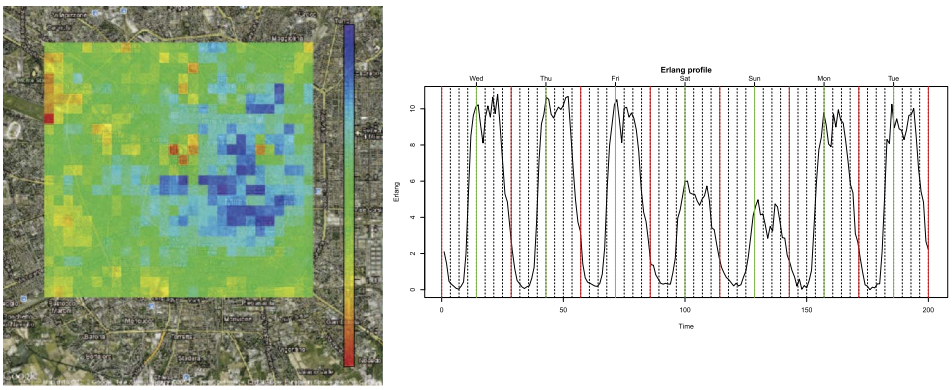

FIG. 7. *Telecom Italia, spatial and temporal pattern of Erlang—Left: Erlang distribution on the lattice at a fixed instant of time. Right: Erlang profile at a fixed spatial pixel.*

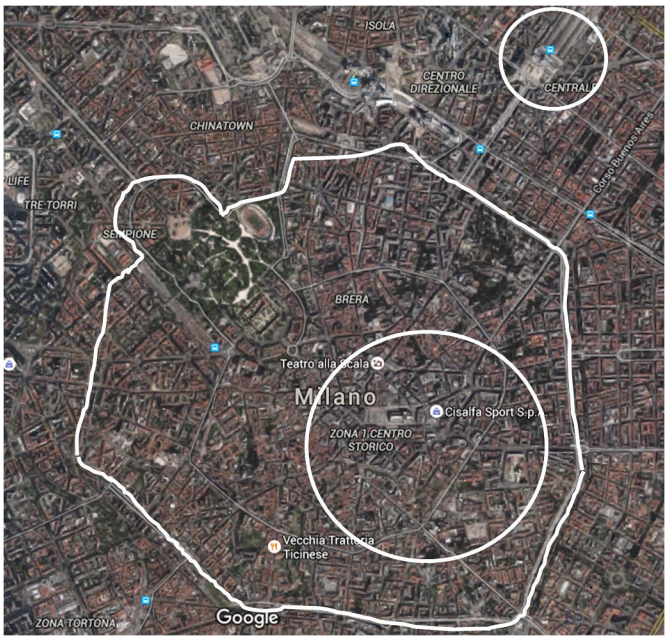

FIG. 8. *Telecom Italia, three major spatial structures*: *Circular road, center city* (*circle in the middle*), *central railway station* (*circle in the upper right corner*).

ure 10), which is the main railway station of the city and the main transportation center of the entire northern part of Italy for both workers and occasional visitors.

Due to discontinuities in the information provided by the Telecom antennas, preprocessing is needed before formal data analysis. For this purpose, we follow the steps described in Manfredini et al. (2015) and Secchi, Vantini and Vitelli (2015). Without going too much into the details, a pixel-wise smoothing of the Erlang through a Fourier basis expansion of a period of one week is performed. We then sample the measurements at $p = 200$ instants of time regularly spaced in the time span of the data collection.

5.2. *Comparison between fastICA, cICA and scICA.* In Figures 9 to 11, we respectively present the first three dominant ICA components identified by scICA (top panels), fastICA (middle panels) and cICA (bottom panels). With each panel, the recovered spatial sources are depicted on the left, while the corresponding temporal profiles are shown on the right. It is worth to point out that, in the temporal profile graphs, each week goes from Wednesday to Tuesday, while the vertical red and green solid lines represent midnights and middays, respectively.

Regarding the first independent component (IC), Figure 9 suggests that it captures working activities according to the spatial/temporal profiles. All three methods give quite similar temporal profiles, which are higher for the weekdays than

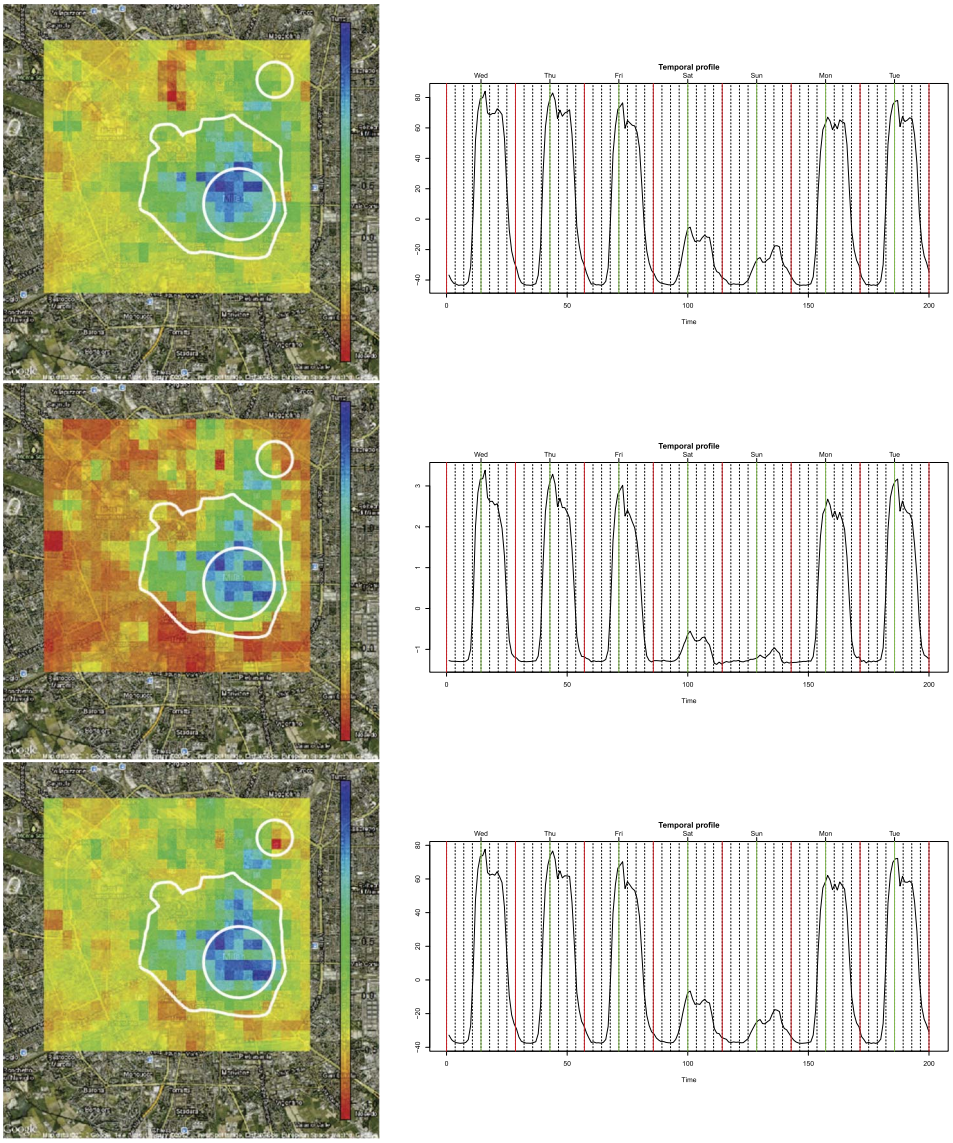

FIG. 9. *IC*1—*Working activities*: *spatial maps* (*on the left*) *and temporal profiles* (*on the right*) *identified by scICA* (*top*), *fastICA* (*middle*) *and cICA* (*bottom*). *The green spatial regions highlight the areas with lots of working activities*.

the weekend, and, within each day, high during the daily hours and low during the nights. The blue areas (inside the internal circle) in the corresponding spatial sources highlight the financial districts in the center of the city (i.e., the areas with high mobile activities during working hours); also see Figure 8.

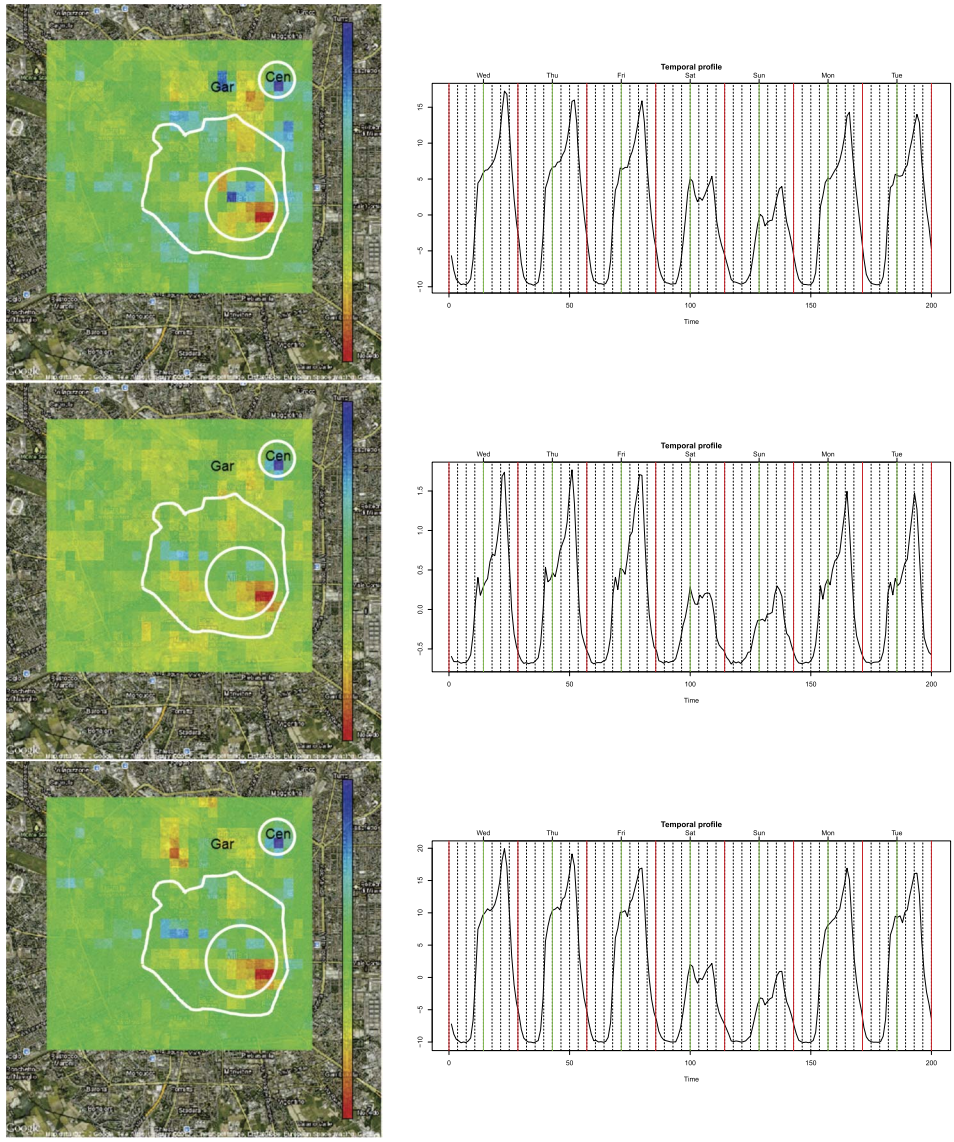

FIG. 10. *IC2—Railway stations*: *spatial maps* (*on the left*) *and temporal profiles* (*on the right*) *identified by scICA* (*top*), *fastICA* (*middle*) *and cICA* (*bottom*). *All temporal profiles present a peak around* 6 *pm of the working days*. *fasICA and cICA highlight only the central railway station* (*"Cen"*). *scICA identifies both the Central station and the large Garibaldi station* (*"Gar"*).

For the second IC, Figure 10 corresponds to the subscriber behavior at the railway stations. Indeed, the temporal profiles present a peak every working day around 6 pm, when people take the train to go home after work. Milan has two major railway stations—the biggest Central station (Figure 8) and the Garibaldi

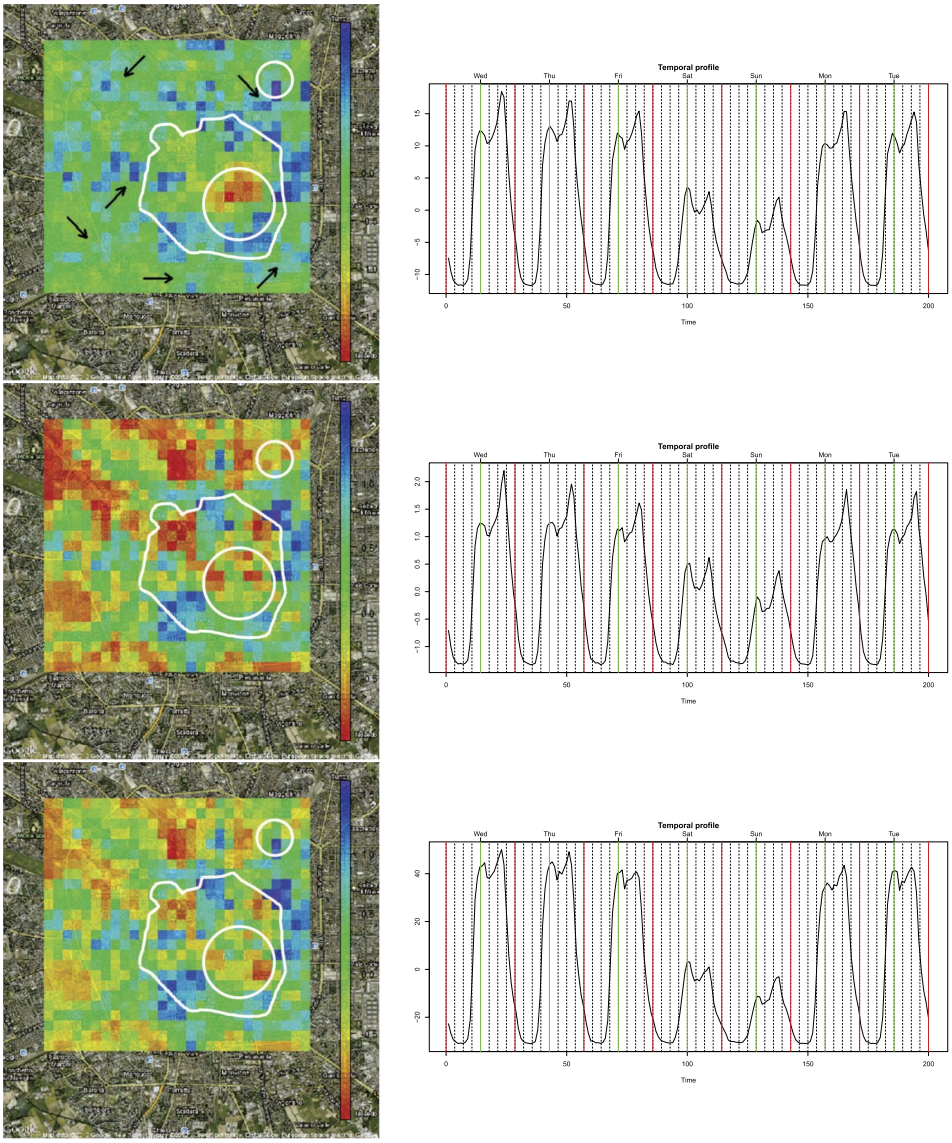

FIG. 11. *IC3—Traffic after work*: *spatial maps* (*on the left*) *and temporal profiles* (*on the right*) *identified by scICA* (*top*), *fastICA* (*middle*) *and cICA* (*bottom*). *The temporal profiles have a peak around* 6 *pm. The scICA spatial component clearly highlights the big outflow streets of the city.*

station. In the spatial map, all three methods highlight the Central station (labeled as "Cen"); however, only scICA is able to identify the Garibaldi station (labeled as "Gar"). We also want to comment on the red region shown inside the circle in the bottom right quadrant of all three spatial maps. This area is not particularly

interesting with respect to the neighborhood, and the negative large absolute value is caused by unusual high usage activity recorded at one particular pixel in the middle of the red region.

The third IC presents the most interesting comparison among the three methods, as depicted in Figure 11. This comparison shows where the improvement from incorporating the spatial dependence is the most evident. The temporal profiles are similar to those in the first IC, except with a peak around 6 pm. This component seems to describe the traffic pattern after working hours. The scICA spatial map clearly identifies the areas around the city center and highlights the big outflow streets (identified by the black arrows in the top panel). These are part of the crucial ring road depicted in Figure 8 which are major roads for people to go through the city. As a comparison, fastICA and cICA only identify a less clear spatial pattern.

Acknowledgment. We thank the Editor and the referee team for their insightful comments that led to great improvement of the paper. Thanks are also due to Convenzione di Ricerca DIAP—Politecnico di Milano and Telecom Italia that provided the data.

REFERENCES

AHAS, R. and MARK, U. (2005). Location based services—new challenges for planning and public administration. *Futures* **37** 547–561.

AMARI, S., CICHOCKI, A. and YANG, H. H. (1996). A new learning algorithm for blind signal separation. *Adv. Neural Inf. Process. Syst.* 757–763.

BOX, G. E. P. and JENKINS, G. M. (1970). *Times Series Analysis. Forecasting and Control*. Holden-Day, San Francisco. MR0272138

CALHOUN, V. D., ADALI, T., PEARLSON, G. D. and PEKAR, J. J. (2001). Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Hum. Brain Mapp.* **13** 43–53.

CHEN, A. and BICKEL, P. J. (2005). Consistent independent component analysis and prewhitening. *IEEE Trans. Signal Process.* **53** 3625–3632. MR2239886

CHEN, A. and BICKEL, P. J. (2006). Efficient independent component analysis. *Ann. Statist.* **34** 2825–2855. MR2329469

COMON, P. and JUTTEN, C. (2010). *Handbook of Blind Source Separation*: *Independent Component Analysis and Applications*. Academic Press, San Diego.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York. MR1239641

CRUJEIRAS, R. M. (2006). Contribution to spectral spatial statistics. Ph.D. dissertation, Universidade de Santiago de Compostela.

CRUJEIRAS, R. M. and FERNÁNDEZ-CASAL, R. (2010). On the estimation of the spectral density for continuous spatial processes. *Statistics* **44** 587–600. MR2739414

EDDELBUETTEL, D. and FRANÇOIS, R. (2011). Rcpp: Seamless R and C++ integration. *J. Stat. Softw.* **40** 1–18.

ELOYAN, A., CRAINICEANU, C. M. and CAFFO, B. S. (2013). Likelihood-based population independent component analysis. *Biostatistics* **14** 514–527.

FAN, J. and KREUTZBERGER, E. (1998). Automatic local smoothing for spectral density estimation. *Scand. J. Statist.* **25** 359–369. MR1649039

GEBIZLIOĞLU, Ö. L. (1988). Spatial processes: Modelling, estimation, and hypothesis testing. *Comm. Fac. Sci. Univ. Ankara Ser. A_1 Math. Statist.* **37** 67–94 (1991). MR1203412

GONZÁLEZ, M. C., HIDALGO, C. A. and BARABÁSI, A.-L. (2008). Understanding individual human mobility patterns. *Nature* **453** 779–782.

GUO, Y. (2011). A general probabilistic model for group independent component analysis and its estimation methods. *Biometrics* **67** 1532–1542. MR2872404

HASTIE, T. and TIBSHIRANI, R. (2010). ProDenICA: Product Density Estimation for ICA using tilted Gaussian density estimates. R package version 1.0.

HYVARINEN, A., KARHUNEN, J. and OJA, E. (2001). *Independent Component Analysis*. Wiley, New York.

HYVÄRINEN, A. and OJA, E. (2000). Independent component analysis: Algorithms and applications. *Neural Netw.* **13** 411–430.

KAUFMANN, V. (2012). *Re-Thinking Mobility*: *Contemporary Sociology*. Ashgate, Aldershot.

LEE, S., SHEN, H., TRUONG, Y., LEWIS, M. and HUANG, X. (2011). Independent component analysis involving autocorrelated sources with an application to functional magnetic resonance imaging. *J. Amer. Statist. Assoc.* **106** 1009–1024. MR2894760

LUÈ, A., COLORNI, A., NOCERINO, R. and PARUSCIO, V. (2012). Green move: An innovative electric vehicle-sharing system. *Procedia—Social and Behavioral Sciences* **48** 2978–2987.

MANFREDINI, F., PUCCI, P., SECCHI, P., TAGLIOLATO, P., VANTINI, S. and VITELLI, V. (2015). Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the Milan urban region. In *Advances in Complex Data Modeling and Computational Methods in Statistics* (A. Paganoni and P. Secchi, eds.) 133–147. Springer, Berlin.

MATTESON, D. S. and TSAY, R. S. (2016). Independent component analysis via distance covariance. *J. Amer. Statist. Assoc.* To appear. DOI:10.1080/01621459.2016.1150851.

OECD (2006). *Territorial Reviews*: *Milan*, *Italy*. OECD Publishing.

R CORE TEAM (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

RATTI, C., PULSELLI, R. M., WILLIAMS, S. and FRENCHMAN, D. (2006). Mobile landscapes: Using location data from cell phones for urban analysis. *Environ. Plann. B*, *Plann. Des.* **33** 727–748.

SECCHI, P., VANTINI, S. and VITELLI, V. (2015). Analysis of spatio-temporal mobile phone data: A case study in the metropolitan area of Milan. *Stat. Methods Appl.* 1–22.

SHELLER, M. and URRY, J. (2006). The new mobilities paradigm. *Environ. Plann. A.* **38** 207–226.

VAN DE VEN, V. G., FORMISANO, E., PRVULOVIC, D., ROEDER, C. H. and LINDEN, D. E. J. (2004). Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. *Hum. Brain Mapp.* **22** 165–178.

WHITTLE, P. (1952). Some results in time series analysis. *Skand. Aktuarietidskr.* **35** 48–60. MR0049539

P. ZANINI
MOX—DEPARTMENT OF MATHEMATICS
POLITECNICO DI MILANO
VIA BONARDI 9
20133 MILANO
ITALY
E-MAIL: paolo.zanini@polimi.it

H. SHEN
INNOVATION AND INFORMATION MANAGEMENT
FACULTY OF BUSINESS AND ECONOMICS
UNIVERSITY OF HONG KONG
POKFULAM
HONG KONG
E-MAIL: haipeng@hku.hk

Y. TRUONG
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
CHAPEL HILL, NORTH CAROLINA 27599
USA
E-MAIL: truong@bios.unc.edu