MOX–Report No. 38/2014

Independent Component Analysis for Spatial Stochastic Processes on a Lattice

Shen, H.; Truong, Y.; Zanini, P.

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

Independent Component Analysis for Spatial Stochastic Processes on a Lattice.

Haipeng Shen[1], Young Truong[2], Paolo Zanini[3]

September 4, 2014

[1] Department of Statistics and Operations Research
University of North Carolina at Chapel Hill
Chapel Hill, 27599, NC
[2] Department of Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill, 27599, NC
[3] MOX – Department of Mathematics
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy
haipeng@email.unc.edu
truong@bios.unc.edu
paolo.zanini@polimi.it

Abstract

Independent Component Analysis (ICA) is a widespread data-driven methodology used to solve Blind Source Separation problems. A lot of algorithms have been proposed to perform ICA, but few of them take into account the dependence within the mixtures and not only the dependence between the mixtures. Some algorithms deal with the temporal ICA (tICA) approach exploiting the temporal autocorrelation of the mixtures (and the sources). In particular, colored ICA (cICA), that works in the spectral domain, is an effective method to perform ICA through a Whittle likelihood procedure assuming the sources to be temporal stochastic process. However spatial ICA (sICA) approach is becoming dominant in several field, like fMRI analysis or geo-referred imaging. In this paper we present an extension of cICA algorithm, called spatial colored ICA (scICA), where sources are assumed to be spatial stochastic processes on a lattice. We exploit the Whittle likelihood and a kernel based nonparametric algorithm to estimate the spectral density of a spatial process on a lattice. We illustrate the performance of the proposed method through different simulation studies and a real application using a geo-referred dataset about mobile-phone traffic on the urban area of Milan, Italy. Simulations and the real application showed the improvements provided by scICA method due to take into account the spatial autocorrelation of the mixtures and the sources.

Keywords: Independent Component Analysis, Spatial Stochastic Processes, Periodogram, Whittle likelihood, Mobile phone traffic

1 Introduction

Independent Component Analysis (ICA) is a data-driven methodology widely used to solve Blind Source Separation problems [6]. It can be expressed in terms of a latent variable linear model

$$X_j = c_{j1}S_1 + ... + c_{jK}S_K = \sum_{k=1}^{K} c_{jk}S_k, \quad \text{for } j = 1, ..., p, \tag{1}$$

where random variables $X_1, ..., X_p$ are observed linear combinations of K unknown (latent) random independent sources $S_1, ..., S_K$. Consider the case $K = p$. Using a vector-matrix notation equation (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) matrix of real numbers named mixing matrix. In applications, we observe $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathbb{R}^p$, 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 write:

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

where $\mathbb{S}$ is a (n, p) matrix containing the n unknown realizations of the p sources. A BSS problem consists in estimating the unmixing matrix $W = C^{-1}$, given $\mathbb{X}$, and then recover $\mathbb{S}$ through;

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

Being ICA a widespread approach for BSS problems, a lot of methods have been developed to approach it. Two widely used algorithms are infomax [5] and fastICA [15], where the unmixing matrix is estimated minimizing the mutual information (a measure of dependence) between the sources. This is equivalent to maximize the negentropy, a particular non-Gaussianity measure (indeed it is possible to show that the sources should not to be Gaussian distributed in order for the mixing matrix to be identifiable). These two algorithms rely only on the independence between the sources and try to estimate the marginal densities of the sources without any further assumption on the form of such densities.

Other methods, instead, make assumptions on the source densities. For example Independent Factor Analysis (IFA) [4, 21] models the independent components as mixtures of Gaussians, while Log-ICA and Lap-ICA [2] assume that the sources follow a Logistic and a Laplacian distribution, respectively.

All the above methods, while study the dependence structure between mixture variables trying to unmix the dependent signals in independent sources, do not exploit the possible correlation structure within the sources (and, then, within the mixtures). However in the real applications where ICA is commonly used, the signals are often autocorrelated, in time or space. For instance the typical framework where ICA has been introduced is the cocktail-party problem, where different microphones in a room register sounds produced by different sources. The goal here is to recover the original audio signals through the time signals registered by the microphones. In this case a time correlation within the sources (and the mixtures) is present.

ICA is a method widely used also in the analysis of fMRI data [9, 20]. This kind of data consists in the registration of the brain activity in a certain number of cuboid elements, called voxels, for a period of time. Then both a spatial (between voxels) and temporal (between instants of time) dependence is present and there are two different approaches that can be applied. We can consider each spatial brain map at every instant of time a mixture of independent image components, or each temporal signal at every voxel a mixture of independent temporal sources. The former approach is called spatial ICA (sICA) while the latter is named temporal ICA (tICA). Some methods have been developed taking into account the correlation within the sources for the tICA approach. The method described in [23] is the first algorithm that considers the temporal autocorrelation of the sources, through the analysis of their spectral densities. However this method is based on the assumption that the spectral densities of the sources are known up to a scale parameter and this assumption is unrealistic in the real applications. Other methods, like AMUSE or TDSEP algorithms (see [28, 30]), exploit the autocorrelation of the sources in the sense that they estimate the unmixing matrix W taking into account the independence between the sources at different lags. However they do not analyze the temporal structure within the single sources. Colored ICA (cICA) [18], instead, is an innovative procedure that takes into account the autocorrelation of the sources and it also works in the spectral domain, but in this case the knowledge of the spectral densities is not needed. Regarding the sICA approach in the literature there are no methods that involve the spatial autocorrelation of the sources in the evaluation of the independent components, imposing some stochastic spatial structure. In this paper we provide a method to fill this lack. Indeed, since the spatial independence is well suited to the sparse distributed nature of the spatial pattern for most cognitive activation paradigms (see, e.g., [20]), spatial approach is becoming dominant in the fMRI analysis. Furthermore, other fields where ICA analysis can provide interesting results are spreading over. For instance, spatiotemporal datasets that perfectly fit in the BSS framework are the geo-referred data, where the temporal changes of a certain quantity are measured on a specific geographic area. In this paper we present an interesting application of our method to the analysis of a mobile-phone traffic dataset related to the urban area of Milan, Italy. sICA approach, in this case, is particularly interesting because it allows to find out independent spatial maps related to different patterns that can be associated with specific activities within the city. The temporal profiles in the mixing matrix represent the temporal evolutions of such activities.

The metropolitan area of Milan, located in the North of Italy, is the fifth biggest metropolitan area of the entire Europe in terms of number of inhabitants. As all the large metropolitan areas, it is characterized by a consistent presence of working and residential/leisure activities. Indeed, the urban area of Milan provides nearly the 10% of the Italian gross domestic product and it is the most populated province of the country, with a density of more than 1000 inhabitants per km^2. An Organization for Economic Co-operation and Development (OECD) review of 2006 (see [22] for the complete report) identified housing, transport and congestion as the principal limitation for the future development of the area. In particular most of the principal roads connecting the city of Milan with its suburbs have reached their saturation during the rushing hours. These aspects cause a lot of problems, above all in terms of pollution and economy. Although in recent years something has been done to decrease the congestion stimulating the use of different means of transport, like the public transports or car and bike-sharing systems, a deep analysis of the main features regarding working, residential and mobility activities is crucial for the well-being of the city. Indeed, as highlighted in [17] and [27], changes in management of mobility are a key point to understand times, places and modes of social life, thus structuring the urban areas. Mobile phone network data are potentially an interesting tool to study population behaviors and for the real-time monitoring of the urban dynamics. Indeed they have been widely analyzed in several experimental studies (see e.g, [25, 1, 14]). Since these studies are quite qualitative, our aim is to analyze this kind of data through suitable statistical methods. Some first and recent statistical approaches to analyze these data are presented in [19] and [26]. Here we want to apply the method described in this paper in order to retrieve meaningful and useful information for urban planning.

The rest of the paper is organized as follows. Firstly, in Section 2 we briefly describe spatial processes on lattices, introducing some simple models well known in literature and presenting a non-parametric method to estimate the spectral density of spatial stochastic processes. Then, in Section 3 we describe in details the algorithm we propose. Since it extends the cICA method to the spatial case we call it spatial colored Independent Component Analysis (scICA). In Section 4 we present some simulation to validate scICA and to show the improvement due to take into account the spatial correlation within the sources. Finally, in Section 5 we deeply analyze the mobile-phone traffic dataset by highlighting the significative results obtained by scICA algorithm. All the simulations and the analyses of real data are carried out using `R` statistical software [24]. Furthermore we are developing a `R` package implementing cICA and scICA algorithms.

2 Spatial models on lattices and their spectral representation

Let $\mathbf{s} \in \mathbb{R}^2$ be a generic location in a 2-dimensional Euclidean space and suppose that the potential datum $Z(\mathbf{s})$ at a spatial location $\mathbf{s}$ is a random quantity. If $\mathbf{s}$ varies over an index set $D \subseteq \mathbb{R}^2$, the spatial random field

$$\{Z(\mathbf{s}); \mathbf{s} \in D\} \tag{5}$$

is generated. A realization of (5) is denoted $\{z(\mathbf{s}); \mathbf{s} \in D\}$. We consider D a fixed regular collection of countably many points, say $D = \{\mathbf{s} = (u, v)' : u = \cdots, -1, 0, 1, \cdots; v = \cdots, -1, 0, 1, \cdots\}$. In this case (5) is called a spatial process on a lattice. We consider weakly-stationary processes, when the covariance $C(\mathbf{u})$ is defined, for every $\mathbf{u} \in \mathbb{Z}^2$, as

$$C(\mathbf{u}) = Cov(Z(\mathbf{s} + \mathbf{u}), Z(\mathbf{s})) \quad \forall \mathbf{s} \in D. \tag{6}$$

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

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^2} \sum_{\mathbf{u} \in \mathbb{Z}^2} C(\mathbf{u}) e^{-i\mathbf{u}'\boldsymbol{\omega}}, \tag{7}$$

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

$$C(\mathbf{u}) = \int_{\Pi^2} f(\boldsymbol{\omega}) e^{i\mathbf{u}'\boldsymbol{\omega}} d\boldsymbol{\omega}. \tag{8}$$

Therefore covariance and spectral density form a Fourier pair (a detailed description of spatial stochastic processes and their properties can be found, for instance, in [7, 13]).

2.1 Spatial Autoregressive Moving-Average Models

We now introduce the class of Spatial Autoregressive Moving-Average (SARMA) models (see [13] for a complete description). We start with the model for $Z(u, v)$ given by

$$Z(u, v) = \sum_{j=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} \phi_{jl} Z(u - j, v - l) + \epsilon(u, v), \tag{9}$$

or, by setting $\Phi(T_1, T_2) = 1 - \sum_{j=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} \phi_{jl} T_1^j T_2^l$ so that (9) can be abbreviated as

$$\Phi(T_1, T_2) Z(u, v) = \epsilon(u, v), \tag{10}$$

where $\phi_{00} = 0$, T_1 and T_2 are such that $T_1^{p_1} Z(u, v) = Z(u + p_1, v)$ and $T_2^{p_2} Z(u, v) = Z(u, v + p_2)$ and $\epsilon(u, v)$ is white noise with zero mean and variance σ^2. Model (9) or (10) is called Spatial Autoregressive (SAR) model. For example, if we consider a symmetric first-order model, $\Phi(T_1, T_2)$ reads:

$$\Phi(T_1, T_2) = 1 - \phi_1(T_1 + T_1^{-1}) - \phi_2(T_2 + T_2^{-1}). \tag{11}$$

We take into account now a finite lattice with $n = n_1 \cdot n_2$ sites $\{\mathbf{s}_1, ..., \mathbf{s}_n\}$. We also define the random vector $\mathbf{Z}$ in $\mathbb{R}^n$ as $\mathbf{Z} = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), ..., Z(\mathbf{s}_n))'$ and the random vector $\boldsymbol{\epsilon}$ in $\mathbb{R}^n$ as $\boldsymbol{\epsilon} = (\epsilon(\mathbf{s}_1), \epsilon(\mathbf{s}_2), ..., \epsilon(\mathbf{s}_n))'$, assuming that $\boldsymbol{\epsilon}$ is gaussian distributed with $\mathbf{0}$ mean and a (diagonal) covariance matrix Λ. Let $B = (b_{jl})$ be a matrix to be interpreted as the spatial-dependence matrix (the matrix that gathers the coefficients ϕ_{jl}) with $b_{jj} = 0$. Then, the SAR model for $\mathbf{Z}$ can be written as:

$$(I - B)\mathbf{Z} = \boldsymbol{\epsilon}. \tag{12}$$

Thus, it is easy to see that the distribution of $\boldsymbol{\epsilon}$ induces the distribution of $\mathbf{Z}$. Specifically:

$$\mathbf{Z} \sim N_n(\mathbf{0}, (I - B)^{-1}\Lambda(I - B')^{-1}). \tag{13}$$

By the analogy with time-series models, it is possible to introduce Spatial Moving-Average (SMA) or Spatial Autoregressive Moving-Average (SARMA) processes

$$Z(u, v) = \Theta(T_1, T_2)\epsilon(u, v) \tag{14}$$

$$\Phi(T_1, T_2)Z(u, v) = \Theta(T_1, T_2)\epsilon(u, v), \tag{15}$$

respectively, where $\theta_{00} = 1$. Defining $E = (e_{jl})$ the spatial dependence matrix that gathers the coefficients θ_{jl} such that $e_{jj} = 0$, the SMA model for $\mathbf{Z}$ can be written as:

$$\mathbf{Z} = (I - E)\boldsymbol{\epsilon}. \tag{16}$$

Hence

$$\mathbf{Z} \sim N_n(\mathbf{0}, (I - E)\Lambda(I - E')). \tag{17}$$

The SARMA model for $\mathbf{Z}$

$$(I - B)\mathbf{Z} = (I - E)\boldsymbol{\epsilon} \tag{18}$$

provides

$$\mathbf{Z} \sim N_n(\mathbf{0}, (I - B)^{-1}(I - E)\Lambda(I - E')(I - B')^{-1}). \tag{19}$$

2.2 Spectral representation for SARMA models

Consider the general expression for SARMA model

$$\Phi(T_1, T_2)Z(u, v) = \Theta(T_1, T_2)\epsilon(u, v) \tag{20}$$

that can be reduce to the SAR or SMA model if $\Theta(T_1, T_2) = 1$ or $\Phi(T_1, T_2) = 1$, respectively.

It can be shown that the spectral density $f(\boldsymbol{\omega})$ of the stochastic process Z at a frequency $\boldsymbol{\omega} \in \Pi^2$ is given by:

$$f(\boldsymbol{\omega}) = \frac{\left|\sum_{j=-\infty}^{+\infty}\sum_{l=-\infty}^{+\infty} \theta_{jl} e^{-i(j,l)\cdot\boldsymbol{\omega}}\right|^2}{\left|1 - \sum_{j=-\infty}^{+\infty}\sum_{l=-\infty}^{+\infty} \phi_{jl} e^{-i(j,l)\cdot\boldsymbol{\omega}}\right|^2} \frac{\sigma^2}{(2\pi)^2} = \frac{|A(\boldsymbol{\omega})|^2}{|B(\boldsymbol{\omega})|^2} f_\epsilon(\boldsymbol{\omega}), \tag{21}$$

where $f_\epsilon(\boldsymbol{\omega}) = \sigma^2/(2\pi)^2 \; \forall \boldsymbol{\omega} \in \Pi^2$ is the spectral density of the white noise and $(j, l) \cdot \boldsymbol{\omega} = j\omega_1 + l\omega_2$.

2.3 Estimation of the spectral density based on Whittle log-likelihood

We now approach the problem of estimating the spectral density. In particular we focus on a non-parametric estimation of the spectral density based on Whittle log-likelihood [29]. For this reason we briefly introduce the spatial periodogram, an essential tool for the Whittle estimator.

The periodogram (also called sample spectral density) is a classical non-parametric estimator of the spectral density. For spatial processes observed on a regular grid $D = \{\mathbf{s} = (s_1, s_2) : s_1 = 0, ..., n_1 - 1; s_2 = 0, ..., n_2 - 1\}$, $D \in \mathbb{R}^2$, $n = n_1 \cdot n_2$, the spatial periodogram at a frequency $\boldsymbol{\omega} \in \Pi^2$ is given by:

$$I(\boldsymbol{\omega}) = \frac{1}{(2\pi)^2 n} \left| \sum_{\mathbf{s} \in D} Z(\mathbf{s}) \exp(-i\mathbf{s}'\boldsymbol{\omega}) \right|^2 . \tag{22}$$

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

$$\omega_{k_1} = \frac{2\pi k_1}{n_1} \quad k_1 = 0, \pm 1, ..., \pm m_1 \quad \text{where} \quad m_1 = \lceil \frac{(n_1 - 1)}{2} \rceil$$

$$\omega_{k_2} = \frac{2\pi k_2}{n_2} \quad k_2 = 0, \pm 1, ..., \pm m_2 \quad \text{where} \quad m_2 = \lceil \frac{(n_2 - 1)}{2} \rceil.$$

If we define the Discrete Fourier Transform of the data as:

$$J(\boldsymbol{\omega}) = \frac{1}{2\pi\sqrt{n}} \sum_{\mathbf{s} \in D} Z(\mathbf{s}) e^{-i\mathbf{s}'\boldsymbol{\omega}}, \tag{23}$$

then the periodogram can be obtained as:

$$I(\boldsymbol{\omega}) = J(\boldsymbol{\omega})\overline{J(\boldsymbol{\omega})} = |J(\boldsymbol{\omega})|^2. \tag{24}$$

The spatial periodogram is an asymptotically unbiased estimator of the spectral density, but it is not consistent, since the variance at a specific frequency is proportional to the square of the spectral density at that frequency. Nevertheless, the periodogram values at different frequencies are asymptotically uncorrelated [11]. To avoid this inconsistency problem one of the most popular methods in the spectral parametric context is the Whittle estimation, based on an approximation to the Gaussian negative log-likelihood, and it uses the periodogram as a pilot estimate. For a parametric model of the spectral density f_θ, with $\theta \in \Theta \in \mathbb{R}^p$, the Whittle parameter estimator $\widehat{\theta}$ is given by:

$$\widehat{\theta} = \arg\min_{\theta} L(\theta, I), \tag{25}$$

where $L(\theta, I)$ denotes the Whittle log-likelihood

$$L(\theta, I) = \int_{\Pi^2} \left(\log f_\theta(\boldsymbol{\omega}) + \frac{I(\boldsymbol{\omega})}{f_\theta(\boldsymbol{\omega})} \right) d\boldsymbol{\omega}. \tag{26}$$

The log-likelihood (26) can be interpreted as the Kullback-Leibler divergence between I and f_θ. Note that, in practice, (26) is approximated by a discretized version:

$$\sum_k \left(\log f_\theta(\boldsymbol{\omega}_k) + \frac{I(\boldsymbol{\omega}_k)}{f_\theta(\boldsymbol{\omega}_k)} \right), \tag{27}$$

where the sum extends over all the Fourier frequencies.

Based on the discrete approximation (27), it is possible to obtain a nonparametric estimator for the log-spectral density $m_\theta = \log f_\theta$ [8, 10]. It is easy to see that, minimizing (27) is equivalent to maximize in θ

$$\sum_k \left(Y_k - m_\theta(\boldsymbol{\omega}_k) - e^{Y_k - m_\theta(\boldsymbol{\omega}_k)} \right) \tag{28}$$

where Y_k denotes the log-periodogram value at the Fourier frequency $\boldsymbol{\omega}_k$. We consider the estimator obtained for the log-spectral density function $m(\boldsymbol{\omega}_j) = \log f(\boldsymbol{\omega}_j)$ by a multidimensional local linear kernel estimator. At each Fourier frequency $\boldsymbol{\omega}_j$, this is carried out by approximating $m(\mathbf{x})$ using $a_j + \mathbf{b}_j'(\boldsymbol{\omega}_j - \mathbf{x})$ for $\mathbf{x}$ in a sufficiently small neighborhood of $\boldsymbol{\omega}_j$. The parameters a_j and $\mathbf{b}_j$ will then be estimated by maximizing the local likelihood function described by

$$\max_{a_j, \mathbf{b}_j} \sum_k \left(Y_k - a_j - \mathbf{b}_j'(\boldsymbol{\omega}_j - \boldsymbol{\omega}_k) - e^{Y_k - a_j - \mathbf{b}_j'(\boldsymbol{\omega}_j - \boldsymbol{\omega}_k)} \right) K_H(\boldsymbol{\omega}_j - \boldsymbol{\omega}_k), \tag{29}$$

where the function K_H is a rescaled bidimensional kernel, H is a bidimensional bandwidth matrix and $K_H(\mathbf{x}) = |H|^{-1/2} K(H^{-1/2}\mathbf{x})$. The local maximum likelihood estimator $\widehat{m}_{LK}(H, \boldsymbol{\omega}_j) \equiv \widehat{m}_{LK}(\boldsymbol{\omega}_j)$ of $m(\boldsymbol{\omega}_j)$ is $\widehat{a}_j$ in the maximizer $(\widehat{a}_j, \widehat{\mathbf{b}}_j)$ of (29).

3 Spatial colored Independent Component Analysis

We now consider the BSS problem (3) assuming the sources to be spatial processes defined on a finite lattice D with n sites. Let $\mathbf{S} = (S_1, ..., S_p)'$ be a random vector in $\mathbb{R}^p$. We can define the spectral density and the periodogram of the jth source as $f_{S_j}(\boldsymbol{\omega})$ and $I(\boldsymbol{\omega}, S_j)$ respectively. Then the sources Whittle log-likelihood is given by

$$L(f_{\mathbf{S}}; \mathbf{S}) = \sum_{j=1}^{p} \sum_{k=1}^{n} \left(\frac{I(\boldsymbol{\omega}_k, S_j)}{f_{S_j}(\boldsymbol{\omega}_k)} + \ln(f_{S_j}(\boldsymbol{\omega}_k)) \right) \tag{30}$$

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 we observed the mixed spatial processes. So the log-likelihood (30) can be rewritten as

$$L(W, f_{\mathbf{S}}; \mathbf{X}) = \sum_{j=1}^{p} \sum_{k=1}^{n} \left(\frac{\mathbf{e}_j' W' I(\boldsymbol{\omega}_k, \mathbf{X}) W \mathbf{e}_j}{f_{S_j}(\boldsymbol{\omega}_k)} + \ln(f_{S_j}(\boldsymbol{\omega}_k)) \right) + n \ln|\det(W)| \tag{31}$$

where $I(\boldsymbol{\omega}_k, \mathbf{X})$ is the matrix periodogram of the mixed signals at the Fourier frequency $\boldsymbol{\omega}_k$ and $\mathbf{e}_j = (0, ..., 0, 1, 0, ..., 0)'$ with the jth entry being 1. Then we basically need to estimate both the unmixing matrix W and the sources spectral density f_{S_j}, for $j = 1, ..., p$. Therefore we implement an iterative algorithm, alternating a step where sources spectral densities are estimated with a step where an estimate $\widehat{W}$ of W is obtained. The iterative algorithm stops when the difference between $\widehat{W}_{new}$ and $\widehat{W}_{old}$ is under a convergence threshold, where the difference is measured through the Amari error (see [3] for details), a criterion widely used in ICA framework.

3.1 The iterative algorithm

Firstly we imagine the unmixing matrix W to be fixed. Then, the log-periodogram $Y(\boldsymbol{\omega}_k, S_j)$ can be easily evaluated for every $j = 1, ..., p$ and every $k = 1, ..., n$. Hence, for every $j = 1, ..., p$, the spectral density f_{S_j} can be estimated through the nonparametric method (29). In the cICA algorithm [18], differently, the parameter of the spectral density of the temporal sources are estimated through a parametric procedure and then the spectral densities are evaluated. In our framework, where sources are assumed to be spatial stochastic processes, a parametric approach would be difficult to deal with, because of the nontrivial way to choose the order of the autoregressive and moving-average parts in SARMA models. Moreover, spatial parametric procedures would be too restrictive, because of the very different features that spatial sources could present in real applications. Hence, the nonparametric approach allows us to take into account many different structures for the sources.

We now fix f_{S_j} for $j = 1, ..., p$. A typical procedure in ICA methods is to prewhite data [16]. In this way W is orthogonal and this allows us to drop the last term in (31). However we need to impose an orthogonality constraint on the unmixing matrix. Then, for every $j = 1, ..., p$, we minimize

$$\widetilde{L}(W, f_{\mathbf{S}}; \mathbf{X}) = \mathbf{w}_j'(A_k + \tau C_j)\mathbf{w}_j \tag{32}$$

where $\mathbf{w}_j = W\mathbf{e}_j$ is the jth column of W, $A_k = \sum_{k=1}^{n} \frac{I(\boldsymbol{\omega}_k, \mathbf{X})}{f_{S_j}(\boldsymbol{\omega}_w)}$, $C_j = \sum_{k \neq j} \mathbf{w}_k \mathbf{w}_k'$ and τ is a positive tuning parameter. Matrix C_j provides an orthogonality constraint in the sense that $\mathbf{w}_j' C_j \mathbf{w}_j = \sum_{k \neq j} \langle \mathbf{w}_j, \mathbf{w}_k \rangle^2$. This representation provides a straightforward estimate for $\mathbf{w}_j$. Indeed it is easy to see that $(A_k + \tau C_j)$ is symmetric and positive-definite. Hence the argmin of (32) is the eigenvector of $(A_k + \tau C_j)$ corresponding to the lowest eigenvalue. However, the problem of setting the tuning parameter τ still remains. It is important to point out that orthogonality has to be a constraint and not simply a penalization. For this reason we set an initial (small) value for τ and then we proceed in an alternating way as follows:

a) we obtain $\widehat{W}$ from (32);

b) if the orthogonality error is under a certain threshold, we remain with this estimate for W. Unless we repeat the step a) with $\tau = 2\tau$.

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

We can finally summarize the iterative algorithm discussed in this section. Firstly we initialize $\widehat{W}$. Then, while the Amari error is greater than a certain threshold, we repeat the following steps:

1) we estimate the sources spectral density through the nonparametric algorithm (29);
2) we update $\widehat{W}$ according the minimization of (32), using the rule described above to impose the orthogonality constraint.

Remark 3.1 *Another possibility to involve the orthogonality constraint is to use the Newton-Raphson method with Lagrange multiplier as presented in [18]. However in the framework analyzed in this paper, the nonparametric estimate of the spectral density could lead to bad conditioned Hessian matrix in the Newton-Raphson update. For this reason we prefer to estimate the unmixing matrix W through the criterium (32). In any case we point out that, in those situations where the Hessian matrix does not present bad conditioning problems, the results of the two approaches do not show relevant differences.*

Remark 3.2 *We presented here the particular case when $K = p$. To consider $K < p$ a typical procedure adopted in ICA method is to project data in the K-dimensional space identified by the first K principal direction. Then proceed with the estimate of the unmixing and of the mixing matrix in this space and finally recover the original mixing matrix by the inverse of the first transformation.*

4 Simulation study

In this section we present some simulation studies, comparing the results obtained by scICA with those obtained by cICA and fastICA (the most popular ICA algorithm). To perform cICA algorithm, we vectorize the 2D processes and we consider them as 1D processes. We make this in order to compare cICA with scICA and to evaluate if taking into account the 2D dependence gives significative improvements with respect to consider the dependence only in one direction. fastICA algorithm, instead, is used as a benchmark algorithm to implement ICA, since it is the most widespread method used in the literature. All simulations are carried out on a $n_1 \times n_2$ grid, with $n_1 = n_2 = 20$.

4.1 First simulation study: symmetric SARMA processes of the first order

The simulation involves two sources and two mixtures. We perform 100 different runs and for each run the mixing matrix C is generated randomly. The first source is generated according the following symmetric SAR model of the first order:

$$Z(u,v) = \phi_1(Z(u-1,v)+Z(u+1,v))+\phi_2(Z(u,v-1)+Z(u,v+1))+\epsilon(u,v) \tag{33}$$

with $\phi_1 = 0.3$, $\phi_2 = 0.4$ and $\epsilon(u, v)$ a gaussian noise with zero mean and variance $\sigma^2 = 0.3^2$. The second source is generated according the following SMA model of the first order:

$$Z(u, v) = \epsilon(u, v) + \theta_1(\epsilon(u-1, v) + \epsilon(u+1, v)) + \theta_2(\epsilon(u, v-1) + \epsilon(u, v+1)) \quad (34)$$

with $\phi_1 = 0.25$, $\phi_2 = 0.3$ and $\epsilon(u, v)$ a gaussian noise with zero mean and variance $\sigma^2 = 0.3^2$. Then, the data matrix $\mathbb{X}$ is generated according to the model (3). In the left panel of Figure 1 the boxplots of the Amari errors for every method considered are shown. Both scICA and cICA significantly outperform fastICA algorithm. The two colored methods seem comparable. However, if we consider the differences between the two errors for every run, we can observe that scICA is significantly better. In the right panel of Figure 1 the boxplot of the differences is depicted. Furthermore we report the p-value of the test to verify if the mean of the difference can be considered less than zero. The p-value is very low, equal to 0.00304, providing statistical evidence that the difference is significant.

In BSS problems we are not interested only in a good estimate of the mixing matrix, but we also aim to reconstruct efficiently the sources. For this reason we evaluate for each run the error in estimating the sources, i.e. the mean of the absolute value of the difference between the true and the estimated sources over the 400 pixels of the 20×20 lattice. In Figure 2 we show the differences of the error over the 100 runs between scICA and cICA algorithm, both for the first and the second source. We also depict the p-value to verify if the mean of the differences could be considered lower than zero. We can observe that for the second source the p-value is around 0.05, providing us slight evidence to reject the null hypothesis, while for the first source the evidence is substantially stronger.

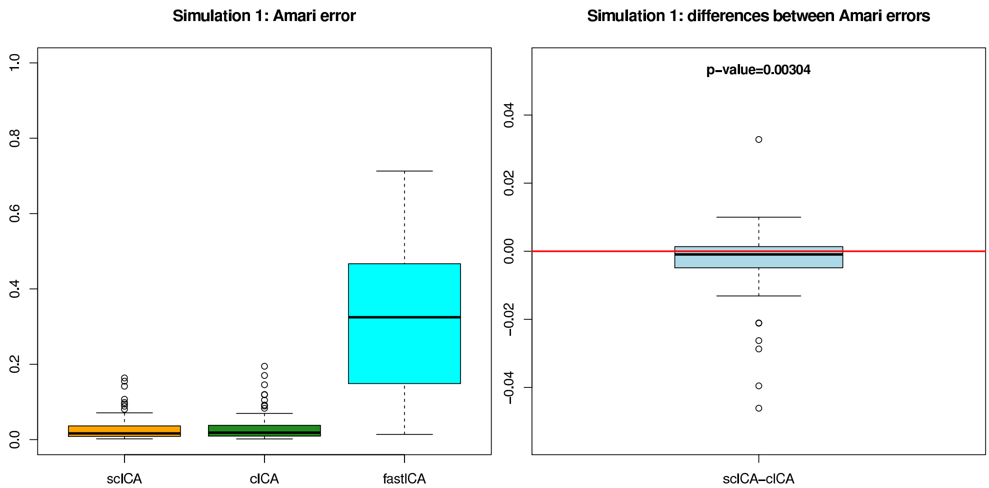

Figure 1: Simulation 1 - On the left panel: boxplot of the Amari error for the three methods considered. On the right panel: boxplot of the differences between scICA and cICA Amari error. The p-value to test if the mean of the difference can be considered lower than zero is shown above the boxplot.

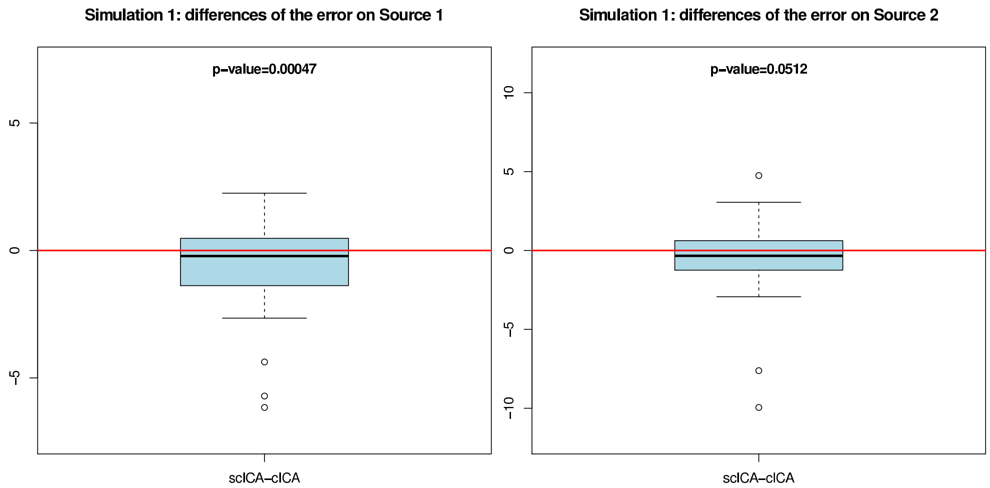

Figure 2: Simulation 1 - On the left panel: boxplot of the differences of the errors between scICA and cICA algorithm in estimating the first source. The p-value to test if the mean of the difference can be considered lower than zero is shown above the boxplot. On the right panel: boxplot of the differences of the errors between scICA and cICA algorithm in estimating the second source. The p-value to test if the mean of the difference can be considered lower than zero is shown above the boxplot.

4.2 Second simulation study: spatial sources with irregular structure

We now take into account two sources, say S_1 and S_2 created artificially and showed in Figure 3.

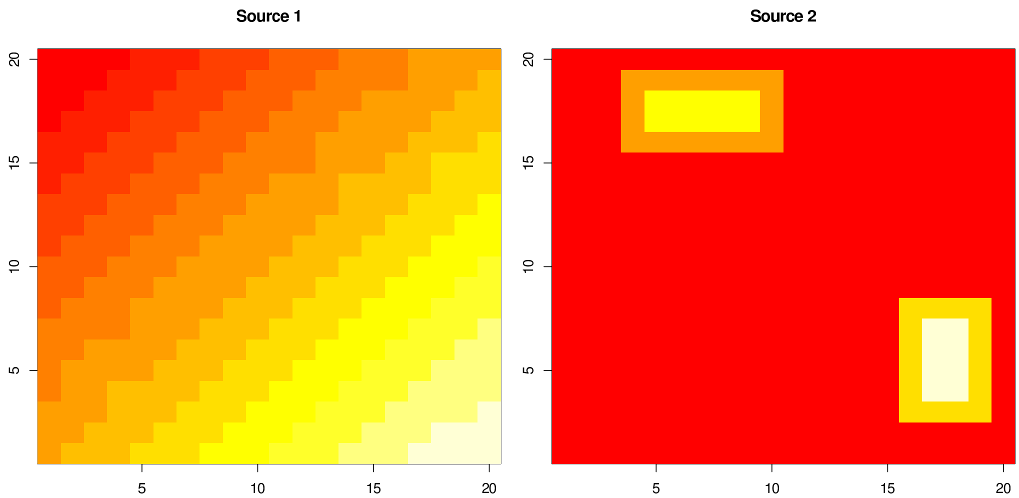

Figure 3: The two sources considered in the second simulation.

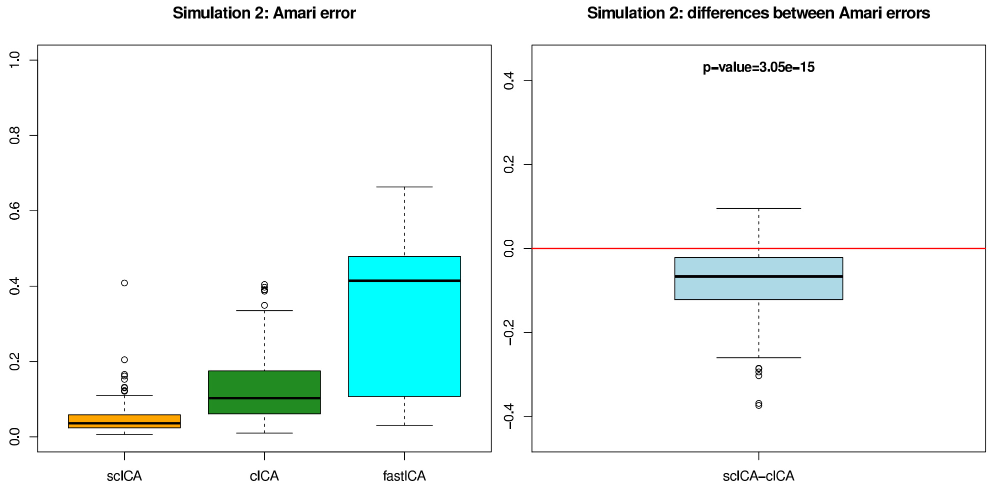

Figure 4: Simulation 2 - On the left panel: boxplot of the Amari error for the three methods considered. On the right panel: boxplot of the differences between scICA and cICA Amari error. The p-value to test if the mean of the difference can be considered lower than zero is shown above the boxplot.

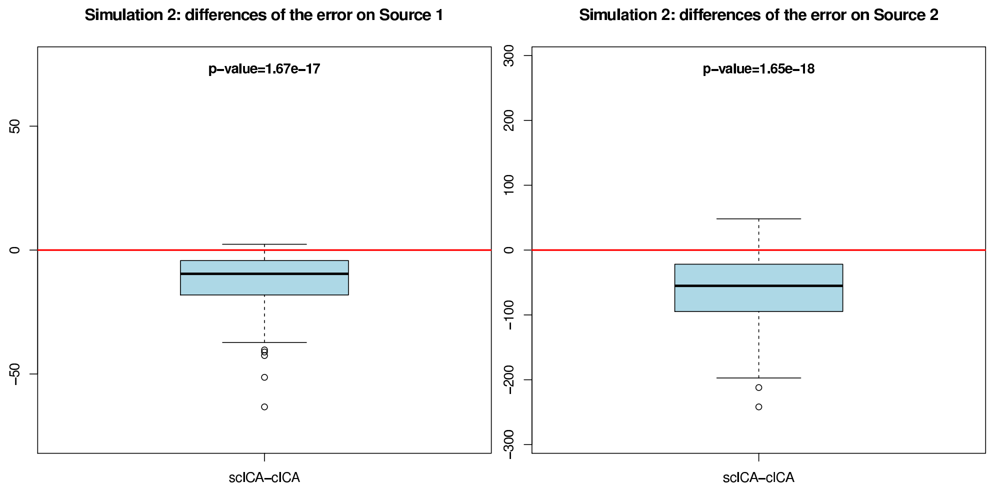

Figure 5: Simulation 2 - On the left panel: boxplot of the differences of the errors between scICA and cICA algorithm in estimating the first source. The p-value to test if the mean of the difference can be considered lower than zero is shown above the boxplot. On the right panel: boxplot of the differences of the errors between scICA and cICA algorithm in estimating the second source. The p-value to test if the mean of the difference can be considered lower than zero is shown above the boxplot.

We perform 100 different runs, generating the mixing matrix randomly at each run and the data matrix according to the model (3), where the sources matrix is composed by the sources of Figure 3 plus some gaussian noise with zero mean and different variances for the two sources. Specifically $\sigma_1^2 = 2^2$ and $\sigma_2^2 = 0.1^2$.

The boxplots of the Amari errors for every method considered are depicted in the left panel of Figure 4. The two colored methods clearly outperform fastICA algorithm, as well as in the first simulation. Furthermore, in this case the improvements by accounting for the spatial structure of the sources seem even more evident. Indeed the p-value is significantly lower, as shown in the right panel of Figure 4. Comparing the estimate of the sources for the two colored method, is evident how scICA strongly outperform cICA, as highlighted by the extremely low p-values in Figure 5.

5 Analysis of Telecom data

In this section we analyze a real mobile-phone traffic dataset, related to the metropolitan area of Milan (Italy). We refer to the data analyzed in [19] and [26], where the analyses are carried out through basis representations that do not consider any probabilistic assumptions on the model. In this paper we want to exploit probabilistic assumptions through the ICA framework, taking into account the spatial structure of this dataset using scICA algorithm.

5.1 Telecom dataset: description

The dataset describes the mobile phone traffic on the metropolitan area of Milan. Data are courtesy of Telecom Italia, the biggest mobile phone Italian company, thanks to a research agreement between Telecom and the Politecnico di Milano. Telephone traffic is anonymously recorded as the average number of simultaneous contacts in a time unit. Then, Telecom elaborates these measurements with a weighted interpolation, thus obtaining an evaluation of the phone traffic on a tessellation of the territory in rectangular areas (i.e., pixels). We analyze here the municipality of Milan divided into a lattice D_0 of 25×28 pixels (232m $\times$ 309m each). For each pixel of the covered area we observe the Erlang every 15 minutes for 14 days. 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, that, as a first approximation, can be considered proportional to the number of active people in that area at that time. The Erlang x_{ij} related to the pixel $l_i \in L_0$ and to the time interval t_j (i.e., the jth quarter of a hour) is evaluated as

$$x_{ij} = \sum_{r=1}^{R} T_{ij}^r$$

where T_{ij}^r indicates the length in minutes of the time interval (or union of intervals) in which the rth mobile phone is calling while moving in the pixel l_i during the time interval t_j. R indicates the total number of potential network users. Hence these data describe a phenomenon in a 2D-space at different instants of time. This may be represented by a surface varying along time, as depicted in Figure 6.

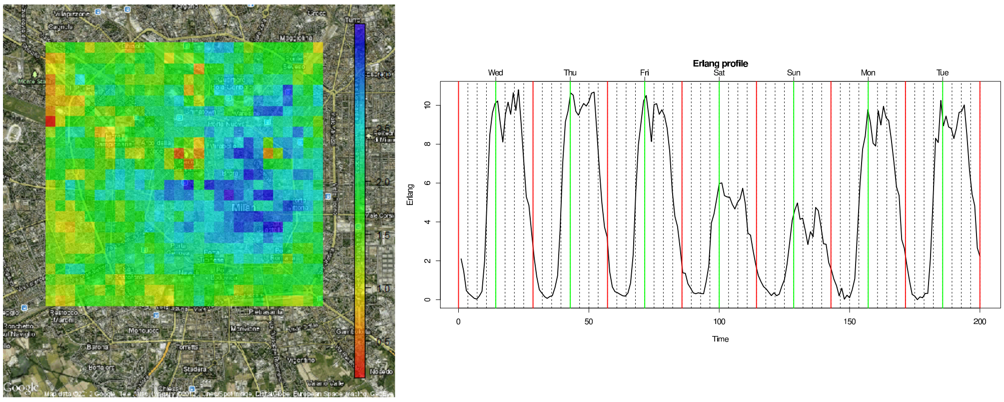

Figure 6: On the left: Erlang distribution on the lattice at a fixed instant of time. On the right: Erlang profile at a fixed pixel.

Aim of the analysis is to decompose the observed signal as a time-varying linear combination of a reduced number, say K, of time-invariant source surfaces. Specifically, for a fixed pixel l_i and a fixed time interval t_j:

$$x_{ij} = s_{i1}a_{j1} + ... + s_{iK}a_{jK},$$

where s_{ik} represents the contribution of the kth source in the pixel l_i and a_{jk} is the intensity of the kth source at the jth time interval. This problem fits in the BSS framework, indeed the purpose of the analysis is to represent $\mathbb{X}$ as the product of two matrices, a p x K matrix A and a n x K matrix $\mathbb{S}$, where each column of $\mathbb{S}$ represents the evaluation at the n pixels of the corresponding source surface and the element a_{jk} indicates the contribution of the kth surface at time j.

The Erlang data we deal with are recorded from March 18th to March 31st, 2009. Due to discontinuities in the information provided by the Telecom antennas, a pre-processing step is needed. For this purpose we follow the analysis presented in [19] and [26]. We do not go into the details of the pre-processing step here and we simply say that the dataset we used in our analysis is represented by the Erlang measurements in the lattice D_0 at $p = 200$ instants of time regularly spaced in the time interval of one week.

5.2 Independent Component Analysis: results obtained through fastICA and scICA algorithms

We now perform an Independent Component Analysis on the Telecom dataset. In this case the sources (i.e, the columns of $\mathbb{S}$) are spatial maps. Classical ICA methods, such as fastICA, do not take into account this information. Here we propose to apply the scICA algorithm to exploit this information from the estimate of $\mathbb{S}$ and C. Furthermore we compare it with the well-known fastICA algorithm. In Figures 7, 8 and 9 we present three dominant components identified by the two algorithms. Figure 7 seems to catch working activities. Indeed the temporal profiles, that are quite similar, are turned on during the daily hours of the working days more than during the daily hours of the weekend (the first day shown is Wednesday) and turned off during the nights. The spatial sources highlight the financial districts in the center of the city (i.e., the areas devoted to working activities). Figure 8 catches the behavior of the railway stations. Indeed both temporal profiles present a peak every working day around 6pm, when people take the train to come back home after work. However, while fastICA highlights in the spatial map only the Central railway station, that is the biggest Milanese station, scICA also catches the Garibaldi station (in the top central part of the map), another large station of the city.

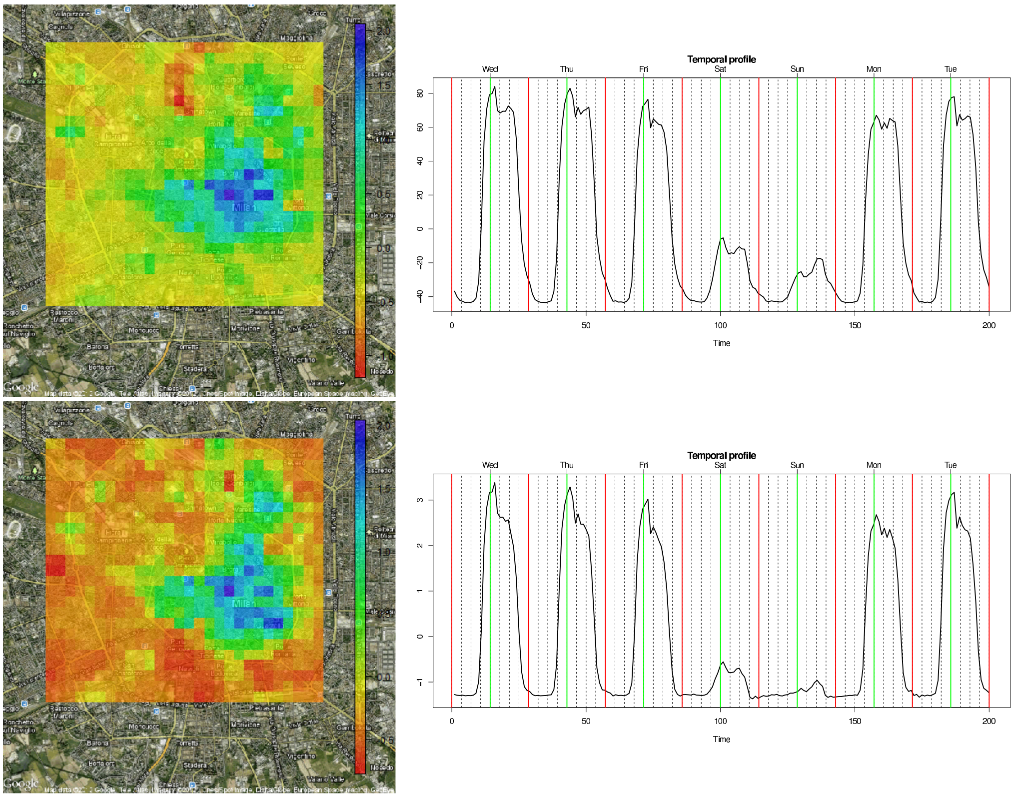

Figure 7: Working activities: the top panel presents surface (on the left) and temporal profile (on the right) identified by scICA. The bottom panel presents surface (on the left) and temporal profile (on the right) identified by fastICA. The surfaces catch the areas devoted to working activities. The temporal profiles are quite similar and they are turned on during the daily hours of the working days more than during the daily hours of the weekend and turned off during the nights.

Figure 9 presents the more interesting component to compare the two methods, where the improvements by incorporating the spatial dependence seem clear. The temporal profiles, indeed, are turned on during the daily hours of the working days more than during the weekend, with a peak around 6pm. The surfaces identify the areas around the center. This component seems to speak about the traffic after the work activities. However scICA component presents a more interesting surface, highlighting the big outflow streets, while fastICA seems able to highlight only the big ring around the center of the city.

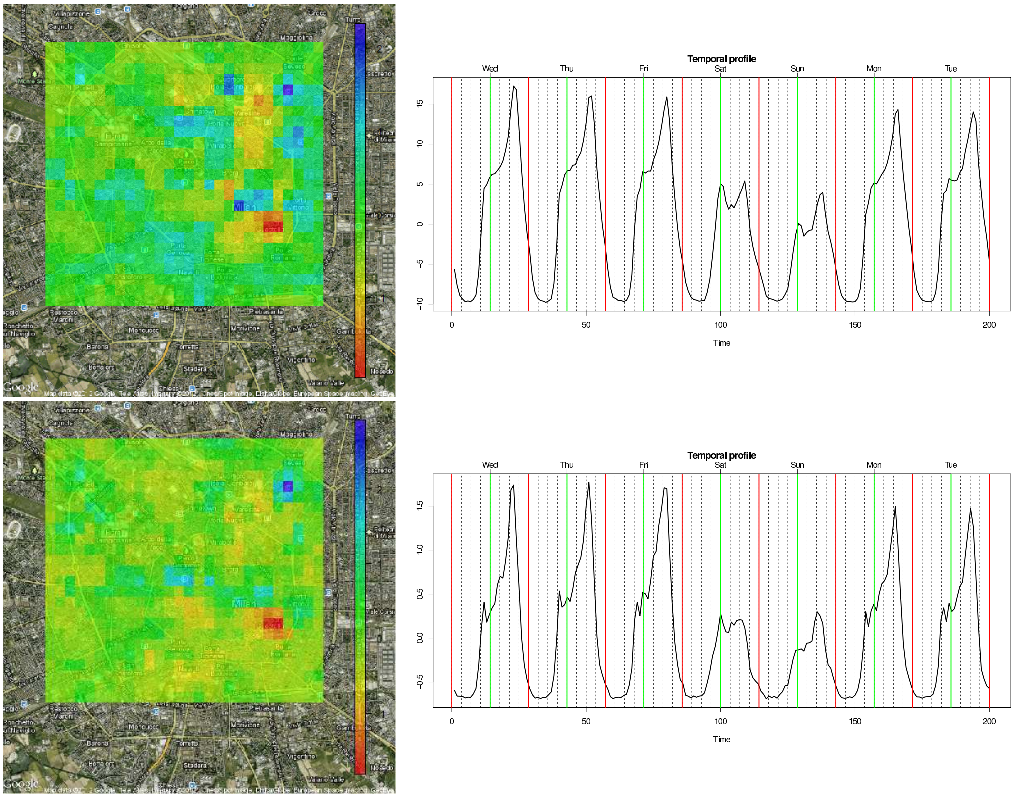

Figure 8: Railway stations: the top panel presents surface (on the left) and temporal profile (on the right) identified by scICA. The bottom panel presents surface (on the left) and temporal profile (on the right) identified by fastICA. Both temporal profiles present a peak around the 6pm of the working days. The fasICA source (bottom panel on the left) shows a single pixel with a high value on the Central railway station, the biggest railway station of Milan. The scICA source (top panel on the left) highlights the Central railway station, but also highlights Garibaldi railway station (in the central top part of the map), another large railway station of the city.

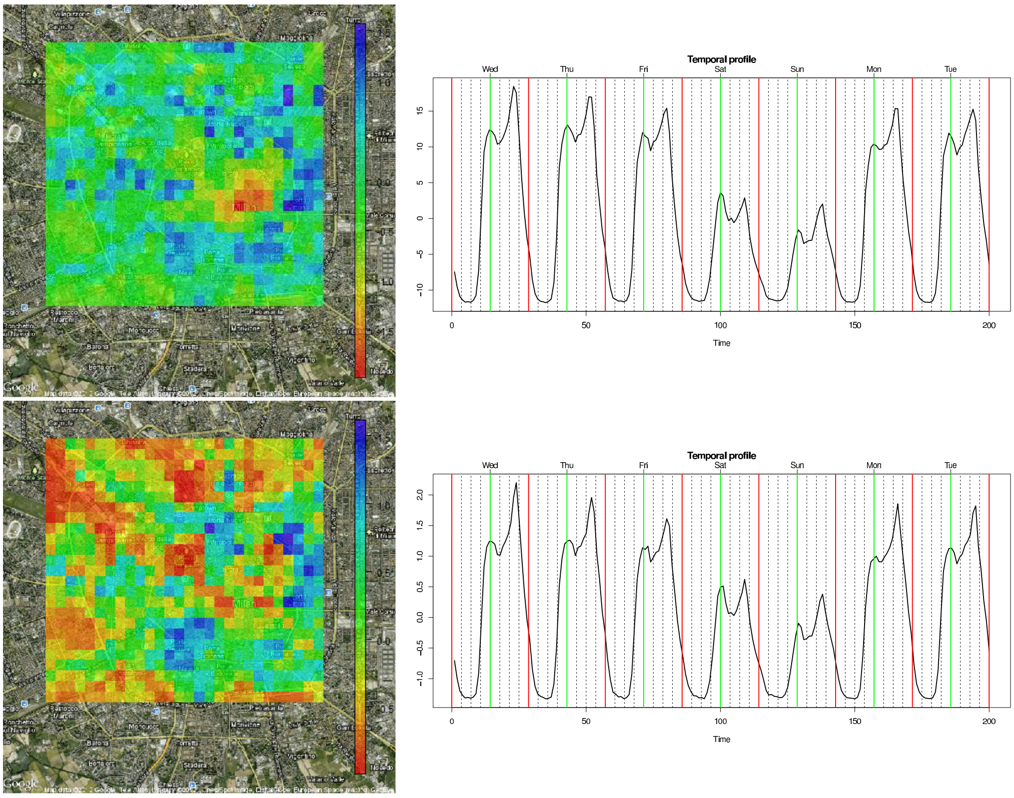

Figure 9: Traffic: the top panel presents surface (on the left) and temporal profile (on the right) identified by scICA. The bottom panel presents surface (on the left) and temporal profile (on the right) identified by fastICA. The temporal profiles are on during the daily hours of the working days more than the weekend, with a peak around 6pm. The surfaces identify the areas around the center. This component seems to speak about the traffic after the work activities. The scICA component presents a more interesting surface, highlighting the big outflow streets of the city.

6 Conclusion

In this paper we presented a new algorithm, named scICA, to solve the Independent Component Analysis problem, considering the unobserved sources as spatial stochastic processes on a lattice. Therefore this algorithm is particularly suitable to face the spatial ICA approach and allows to take into account the spatial dependence between the sources (and the mixtures). Our method works in the spatial spectral domain and follows the idea of the cICA method, where the Whittle likelihood is exploited in order to estimate the unmixing matrix and the spectral densities. However, in the spatial case, a parametric estimate of the spectral density might present several drawbacks. Hence we considered a non-parametric approach for the spectral density estimation. Through different simulated examples we clearly showed the improvements due to take into account the spatial dependance, both with respect a benchmark algorithm like fastICA and the cICA method that, differently, considers the dependence only in one direction. Then we applied scICA algorithm to a real dataset regarding the mobile-phone traffic in the metropolitan area of Milan (Italy) along time, obtaining very interesting and meaningful results useful, for instance, for urban planning.

Future works and improvements can be considered. In particular methods for the estimation of the spectral density on irregular lattices or for nonstationary spatial processes (see, for example, [11, 12]) can be integrate with the proposed algorithm in order to allow the analysis of more general spatial datasets.

References

[1] Ahas, R., Mark, U. (2005). Location based services-new challenges for planning and public administration. *Futures* **37** 547–561.

[2] Allassoniere, S., Younes, L. (2012). A stochastic algorithm for probabilistic independent component analysis. *The Annals of Applied Statistics 6* 125–160

[3] Amari, S., Cichocki, A., Yang, H. H. (1996). A new Learning Algorithm for Blind Signal Separation. *Neural Computation* **11** 1875–1883.

[4] Attias, H. (1999). Independent Factor Analysis. *Neural Computation* **11** 803–851.

[5] Bell, A., Sejnowski, T. J. (1995). An Information-Maximization Approach to Blind Source Separation and Blind Deconvolution. *Neural Computation* **7** 1129–1159.

[6] Comon, P., Jutten, C. (2010). *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press.

[7] Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.

[8] Crujeiras, R. M., Fernandez-Casal, R. (2010). On the estimation of the spectral density for continuos spatial processes. *Statistics: A Journal of Theoretical and Applied Statistics* **44** 587–600.

[9] De Luca, M., Beckmann, C. F., De Stefano, N., Matthews, P. M., Smith, S. M. (2006). fMRI resting state networks define distinct modes of long-distance interactions in the human brain. *Neuroimage* **29** 1359–1367.

[10] Fan, J., Kreutzberger, E. (1998). Local smoothing for Spectral Density Estimation. *Scandinavian Journal of Statistics* **25** 359–369.

[11] Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika* **89** 197–210.

[12] Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of American Statistical Association* **102** 321–331.

[13] Gebizlioglu, O. L. (1998). Spatial processes: modelling, estimation, and hypothesis testing. *Commun. Fac. Sci. Univ. Ank. Ser. A1* **37** 67–94.

[14] Gonzalez, M. C., Hidalgo, C. A., Barabasi, A. L. (2008). Understanding individual human mobility patterns. *Nature* **453** 779–782.

[15] Hyvarinen, A., Oja, E. (2000). Independent Component Analysis: Algorithms and Applications. *Neural Networks* **13** 411–430.

[16] Hyvarinen, A., Karhunen, J., Oja, E. (2001). *Independent Component Analysis*. Wiley, New York.

[17] Kaufmann, V. (2002). *Re-thinking mobility contemporary sociology*. Ashgate.

[18] Lee, S., Shen, H., Truong, Y., Lewis, M., Huang, X. (2011). Independent Component Analysis Involving Autocorrelated Sources With an Application to Functional Magnetic Resonance Imaging. *Journal of American Statistical Association* **106** 1009–1024.

[19] Manfredini, F., Pucci, P., Secchi, P., Tagliolato, P., Vantini, S., Vitelli, V. (2014). Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the Milan urban region. In Paganoni, A., Secchi, P. (eds.) *Advances in Complex Data Modeling and Computational Methods in Statistics*, Springer, Milano, forthcoming.

[20] McKeown, M. J., Sejnowski, T. J. (1998). Independent Component Analysis of fMRI Data: Examining the Assumptions. *Human Brain Mapping* **6** 368–372.

[21] Moulines, E., Cardoso, J., Gassiat, E. (1997). Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. *IEEE International Conference on Acoustics, Speech and Signal Processing, 1997* **5** 3617–3620.

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

[23] Pham, D., Garat, P. (1997). Blind Separation of Mixture of Independent Sources Through a Quasi-Maximum Likelihood Approach. *IEEE Transactions on Signal Processing* **45** 1712–1725.

[24] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

[25] Ratti, C., Pulselli, R. M., Williams, S., Frenchman, D. (2006). Mobile Landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design* **33** 727–748.

[26] Secchi, P., Vantini, S., Vitelli, V. (2012). A Case Study on Spatially Dependent Functional Data: the Analysis of Mobile Network Data for the Metropolitan Area of Milan. *MOX report 43/2012*, Department of Mathematics, Politecnico di Milano.

[27] Sheller, M., Urry, J. (2006). The new mobilities paradigm. *Environment and Planning A* **38** 207–226.

[28] Tong, L., Liu, R., Huang, Y. (1991). Indeterminacy and Identifiability of Blind Identification. *IEEE Transactions on Circuits and Systems* **38** 499–509.

[29] Whittle, P. (1952). Some results in time series analysis. *Skandivanisk Aktuarietidskrift* **35** 48–60.

[30] Ziehe, A., Muller, K. (1998). TDSEP – an efficient algorithm for blind separation using time structure. *Proceedings of ICANN 98* 675–680.

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

38/2014 Shen, H.; Truong, Y.; Zanini, P.
Independent Component Analysis for Spatial Stochastic Processes on a Lattice

37/2014 Giuliani, N.; Mola, A.; Heltai, L.; Formaggia,L.
FEM SUPG stabilisation of mixed isoparametric BEMs: application to linearised free surface flows

36/2014 Abb,A.;Bonaventura,L.; Nini, M.;Restelli,M.;
Anisotropic dynamic models for Large Eddy Simulation of compressible flows with a high order DG method

35/2014 Tricerri, P.; Ded, L.; Deparis, S.; Quarteroni, A.; Robertson, A.M.; Sequeira, A.
Fluid-structure interaction simulations of cerebral arteries modeled by isotropic and anisotropic constitutive laws

34/2014 Antonietti, P.F.; Pacciarini, P.; Quarteroni, A.
A discontinuous Galerkin Reduced Basis Element method for elliptic problems

33/2014 Canuto, C.; Simoncini, V.; Verani, M.
Contraction and optimality properties of an adaptive Legendre-Galerkin method: the multi-dimensional case

32/2014 Agosti, A.; Formaggia, L.;Scotti, A.
Analysis of a model for precipitation and dissolution coupled with a Darcy flux

31/2014 Corno, J.; de Falco, C.; De Gersem, H.; Schps, S.
Isogeometric Simulation of Lorentz Detuning in Superconducting Accelerator Cavities

30/2014 Ferroni, A.; Formaggia, L.; Fumagalli, A.;
Numerical analysis of Darcy problem on surfaces

29/2014 Arioli, G.; Koch, H.
Some symmetric boundary value problems and non-symmetric solutions