GRASPA 2015
BARI 15-16 JUNE

Multi-resolution and spatial Independent Component Analysis approaches for geo-referred and time-varying mobile phone data

P. Zanini[1]

[1] *MOX – Department of Mathematics, Politecnico di Milano, via Bonardi 9, 20133, Milano (Italy); paolo.zanini@polimi.it*

Abstract. *The aim of this work is to provide different statistical tools to catch meaningful and useful information from geo-referred quantities varying along time. In particular a mobile-phone traffic dataset is analyzed to decompose the spatiotemporal information in order to identify spatial and temporal patterns. Two different approaches have been followed. The first one is an Independent Component Analysis (ICA) approach, where sources are assumed to be spatial stochastic processes on a lattice, in order to take into account the spatial dependence between pixels. This method is called spatial colored Independent Component Analysis (Shen, Truong, Zanini, 2014). The second one is a multi-resolution approach, where a temporal sparsity to the final representation is imposed through a wavelet-inspired data-driven procedure. This method is called Hierarchical Independent Component Analysis (Secchi, Vantini, Zanini, 2014). Results highlight urban features related to residential, leisure and mobility activities.*

Keywords. *Mobile-phone Data; Independent Component Analysis; Spatial Stochastic Processes; Multi-resolution Analysis*

---

1 Introduction

The aim of this work is to provide different statistical tools to catch meaningful and useful information from geo-referred quantities varying along time. This kind of data is present in a wide range of different applications.
In environmental analysis, for instance, the measure of pollution in a geo-referred area (e.g. a city, a river, etc...) at different instants of time need to be studied in order to face the environmental pollution problem. In meteorology, often, the analysis of temperatures or wind velocities in a geographic zone across time is considered for a lot of purposes. This work is focused on a case-study where the quantity of interest is a measure of mobile-phone traffic intensity evaluated on a rectangular grid over the city of Milan (Italy) across two weeks. The main purpose of the analysis is to decompose this spatiotemporal dataset in order to identify spatial and temporal patterns characterizing specific locations and/or specific periods. These patterns will be associated to different population behaviors related to residential, leisure and mobility activities. This can be useful, for instance, for urban planning and for real-time monitoring of the urban dynamics.

2 Data description and methods

The analyzed dataset is courtesy of a research agreement between Telecom Italia and the Politecnico di Milano. It consists in the evaluation of Erlang, a dimensionless intensity measure of the use of the mobile network, in a rectangular lattice L_0 of $n = 10573$ pixels covering the metropolitan area of Milan for a global surface of more than 700 km^2. Measurements are taken every 15 minutes for a period of two weeks. Data have been preprocessed exploiting a Fourier basis expansion, as described in [2]. Then, processed data consist in the Erlang measurements in the lattice L_0 at $p = 200$ instants of time regularly spaced in the time interval of one week. An example for a specific pixel and for a fixed instant of time is shown in Figure 1.

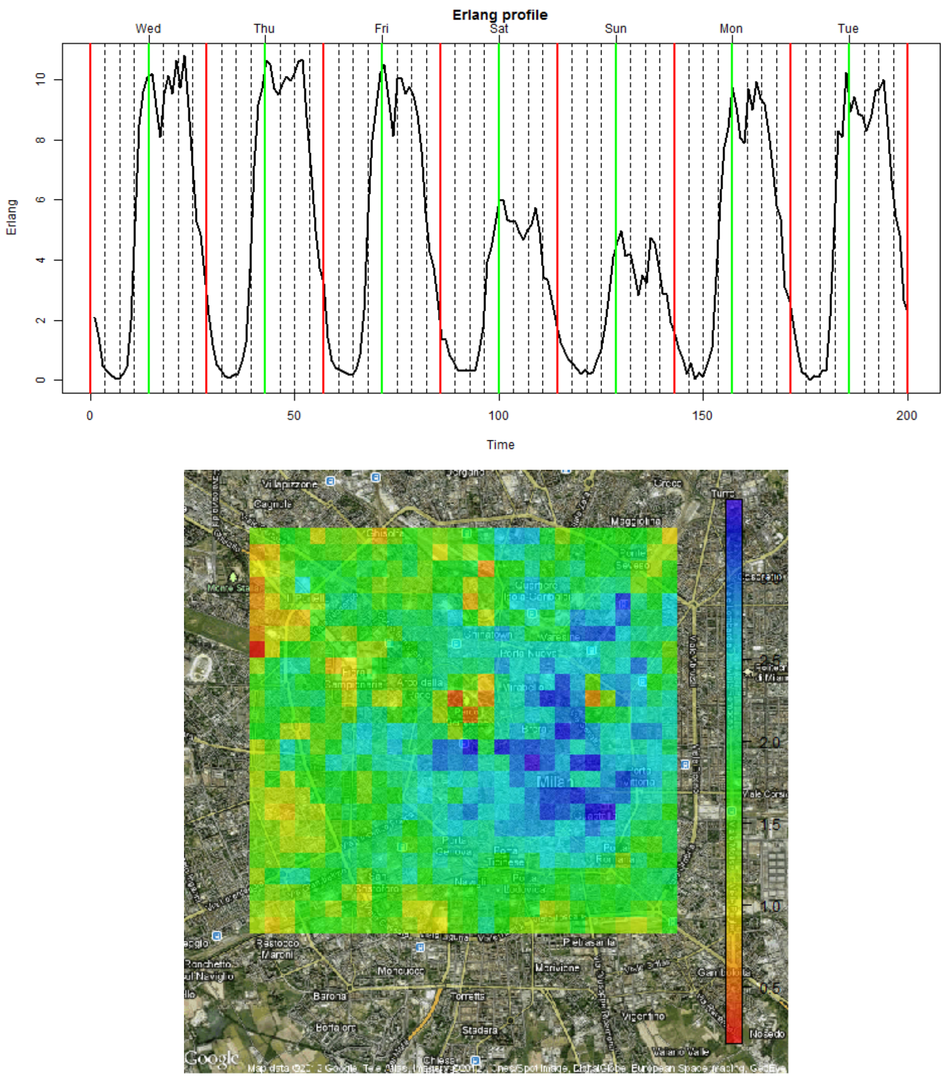

Figure 1: On the top: Erlang signal for a specific pixel. On the bottom: Erlang surface at a fixed instant of time.

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:

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

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 Blind Source Separation framework

$$X = SA.$$

Indeed the purpose of the analysis is to represent the $n \times p$ data matrix X as the product of two matrices, the $p \times K$ basis (loadings) matrix A, and the $n \times K$ source (scores) matrix S, where A gathers the temporal profiles and S the spatial maps of the decomposition. Then, spatial maps can be associated to different urban features, while temporal profiles describe the activation periods of such features.
Two different approaches have been considered:

- an Independent Component Analysis (ICA [1]) approach, where sources are modeled as independent random variables. In fact, sources are assumed to be stochastic processes on a lattice, in order to take into account the spatial dependence between pixels. This analysis is performed through the algorithm named spatial colored Independent Component Analysis (scICA [4]), which works in the frequency domain. It exploits the Whittle likelihood and a kernel based nonparametric algorithm in order to estimate the spatial processes and their spectral densities;
- a multi-resolution approach, where the interest is in finding a sparse and multi-resolution estimate for the basis matrix A. The method used is named Hierarchical Independent Component Analysis (HICA [3]). It provides a multi-resolution (wavelet inspired) data-driven basis, through a hierarchical procedure with the application of ICA on pairs of variables at each step. In this way a temporal sparsity to the final representation is imposed.

The two approaches have been applied to the Telecom dataset, providing interesting results in terms of phenomenological interpretation. The comparison is done not with the purpose to establish which method is better, but to show the different features catched by the two approaches considered, and how these results can contribute to highlight patterns related to urban activities.

References

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

[2] Manfredini, F., Pucci, P., Secchi, P., Tagliolato, P., Vantini, S., Vitelli, V. (2015). Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the Milan urban region. In Paganoni, A., Secchi, P. (eds.). *Advances in Complex Data Modeling and Computational Methods in Statistics Contributions to Statistics*, Springer, 133–147.

[3] Secchi, P., Vantini, S., Zanini, P. (2014). Hierarchical Independent Component Analysis: a multi-resolution non-orthogonal data-driven basis. *Mox Report 01/2014*, Dipartimento di Matematica, Politecnico di Milano.

[4] Shen, H., Truong, Y., Zanini, P. (2014). Independent Component Analysis for Spatial Stochastic Processes on a Lattice. *Mox Report 38/2014*, Dipartimento di Matematica, Politecnico di Milano.