MOX–Report No. 25/2012

Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the Milan urban region

Manfredini, F.; Pucci, P.; Secchi, P.; Tagliolato, P.; Vantini, S.; Vitelli, V.

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

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

Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the Milan urban region

Fabio Manfredini [a], Paola Pucci [a], Piercesare Secchi [b], Paolo Tagliolato [a], Simone Vantini [b], Valeria Vitelli [c]

[a] Politecnico di Milano, Dipartimento di Architettura e Pianificazione, Milano, Italy
[b] Politecnico di Milano, MOX - Dipartimento di Matematica, Milano, Italy
[c] Ecole Centrale Paris and Supelec, Chair on Systems Science and the Energetic challenge, European Foundation for New Energy - Électricité de France, Paris, France

fabio.manfredini@polimi.it
paola.pucci@polimi.it
piercesare.secchi@polimi.it
paolo.tagliolato@polimi.it
simone.vantini@polimi.it
valeria.vitelli@supelec.fr

Keywords: Urban Studies, Mobility, Mobile Phone Network Data Analysis, Urban Computing, Geographic Analysis, Nonparametric Statistics, Spatial Databases and GIS.

2010 AMS Subject Classification: 62H11, 62H20, 62P30.

Abstract

The paper presents a novel geo-statistical unsupervised learning technique aimed at identifying useful information on hidden patterns of mobile phone use. These hidden patterns regard different usages of the city in time and in space which are related to individual mobility, outlining the potential of this technology for the urban planning community. The methodology allows to obtain a reference basis that reports the specific effect of some activities on the Erlang data recorded and a set of maps showing the contribution of each activity to the local Erlang signal. We selected some results as significant for explaining specific mobility and city usages patterns (commuting, nightly activities, distribution of residences, non systematic mobility) and tested their significance and their interpretation from an urban analysis and planning perspective at the Milan urban region scale.

1 Introduction

Interpretative tools for the identification of mobility practices in the contemporary metropolis are needed, not only for some known limitations of traditional data sources but also because new forms of mobility are emerging, describing new city dynamics and time-variations in the use of urban spaces by temporary populations. Traditional data sources for urban and mobility investigations (i.e. surveys, census) have some known limitations, including high cost of surveys, difficulty of data updating, difficulty of describing city dynamics and time dependent variations in intensity of urban spaces usages by temporary populations at different scales. These new forms of mobility, close to the daily mobility, are characterized both by being based on the use of transportation system, and by the efficient appropriation of information technologies (internet, mobile phone). They intensified the density of the moves with which we can read diversified uses of the city, that traditional sources of analysis are unable to return with continuity. As underlined by some authors Kaufmann (2002); Sheller and Urry (2006), changes in management of mobility in the contemporary cities are a useful key for understanding the transformations of times, places and modes of social life and work programs, structuring the metropolitan areas. In this perspective, mobility may represent a tool of both knowledge and project for urban planners, provided that a better understanding of different patterns of mobility in the form of "active biographies", which increase the range of "post-fordist living and labor styles" Nuvolati (2003), is available. Considering the role of mobility practices in social and spatial differentiation, it becomes important to formulate pertinent analytical approaches, aimed at describing the different densities of use of the city as a new challenge and a prerequisite for understanding the city and its dynamics. Hence, from an analytical point of view, it becomes important to accompany the traditional quantitative approaches referred to a geographic displacement that tends to focus on movement in space and time, in an aggregate way and for limited periods, with data sources able to describe fine grain over-time variation in urban movements.

In this direction, an interesting contribution may come from mobile phone network data as a potential tool for the development of real-time monitoring, useful to describe urban dynamics as it has been tested in several experimental studies Ratti et al. (2006); Ahas and Mark (2005); Gonzalez et al. (2008). The application researches focused on two different products. Some studies deal with aspects of representation of the data, emphasizing the aspects most directly evocative, to highlight how these data may represent the "Mobile landscapes" Ratti et al. (2006). Other studies focus on data-mining analysis to building methods for managing large amounts of data, and on the construction of instruments capable of deriving summary information and relevant data on cell-phone Ahas and Mark (2005). As opposed to the more traditional methods of urban surveys, the use of aggregated and anonymous mobile phone network data has shown promise for large-scale surveys with notably smaller efforts and costs Reades et al. (2007). If we consider the observed and aggregated telephone traffic as the result of individual behaviours and habits, we can treat mobile phone data as a useful source on the real use of the cities, capturing for example traces of temporary populations, which are difficult to intercept by traditional data sources, but which, at the same time, increasingly affect urban practices both quantitatively and qualitatively.

An increasing number of studies concerns the exploitation of mobile phone data in urban analysis and planning Becker et al. (2011). In particular an interesting issue regards the classification of urban spaces according to their users' practices and behaviours Reades et al. (2007); Soto and Frías-Martínez (2011a). In Soto and Frías-Martínez (2011b) the authors outline the fact that city areas are generally not characterized by just one specific use, and for this reason they introduce the use of c-means, a fuzzy unsupervised clustering technique for land use classification, which returns for each area a certain grade of membership to each class. In the same paper fuzziness is then abandoned to favour the identification of areas with a clearly defined use. We want to drive the reader's intuition on the interesting point that different "basic" profiles of city usages can concur in the same place and that the overall observed usage of a certain place is the superimposition of layers of these basic profiles.

In this article we experiment a novel geo-statistical unsupervised learning technique finalized to identify useful information on hidden patterns of mobile phone use regarding different usages of the city in time and in space which are related to individual mobility, outlining the potential of this technology in the urban planning community. The results return new maps of the region, each describing the intensity of one of the identified mobility pattern on the territory.

The territorial distribution of the intensity of these patterns allows us to reconstruct the density of use of urban spaces in different temporal, and territorial scales as a precondition:

- to identify temporary populations and different forms of mobility that structure the relationships in the contemporary city;
- to propose diversified management policies and mobility services that city users require, increasing the efficiency of the supply of public services.

2 Data

For the present research we had the opportunity to use the same data that feeds the CityLive platform developed by Telecom Italia for the real time evaluation of urban dynamics based on the anonymous monitoring of mobile phone networks. Telephone traffic is anonimously recorded by each cell of the network as the average number of concurrent contacts in a time unit. Telecom Italia elaborate then these measurements obtaining their distribution by means of weighted interpolations, throughout a tessellation of the territory in squared areas (pixels).

In the Telecom Italia database, the metropolitan area of Milan is divided into a uniform grid (lattice) S_0 of 97×109 pixels. For each pixel, Telecom Italia made available the average number of mobile phones simultaneously using the network for calling, for every 15-minute time interval along a period of 14 days. This quantity is called Erlang and, to a first approximation, can be considered proportional to the number of active people in that pixel at that time interval, hence providing information about people density and mobility. Technically the Erlang $E_{\mathbf{x}j}$ relevant to the pixel $\mathbf{x} \in S_0$ and to the jth quarter of an hour is computed as:

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

where $T^q_{\mathbf{x}j}$ indicates the time interval (or union of intervals) in which the qth mobile phone is using the network for calling within pixel $\mathbf{x}$ and during the jth quarter of an hour. $|T^q_{\mathbf{x}j}|$ indicates the length of $T^q_{\mathbf{x}j}$ expressed in minutes. The number of potential phones using the network is indicated with Q. Even though the phone company uses equation (1) to compute $E_{\mathbf{x}j}$, the meaning of this quantity is better understood from its equivalent representation:

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

where $N_{\mathbf{x}}(t)$ indicates the number of mobile phones using the network within the pixel $\mathbf{x}$ at time t. Equation (2) shows that $E_{\mathbf{x}j}$ is the temporal mean over the jth quarter of an hour of the number of mobile phones using the network within pixel $\mathbf{x}$.

The Erlang data we deal with are recorded, with missing values, from March 18th, 2009, 00:15 am, till March 31st, 2009, 23:45 pm, providing $p = 1308$ records per pixel. The lattice of pixels S_0 covers an area of 757 km^2 included between latitudes 45.37 and 45.57 and longitude 9.05 and 9.35 which corresponds to the Milan core city and to the first ring of municipalities surrounding Milan, located along the ring road. It is divided in $N = 97 \times 109 = 10573$ approximately rectangular pixels. On the whole, 13829484 records are available. To have a first idea of these data, in the top panel of Figure 1 the aggregated Erlang, i.e. the sum of the Erlang measures for each pixel in the investigated area, $\sum_{\mathbf{x} \in S_0} E_{\mathbf{x}j}$, is represented as a function of the corresponding quarter of an hour $j = 1, \ldots, p$. Some peculiar features are already noticeable, such as the day/night effect and working/weekend day effect. The aim of the analysis is indeed to identify these global features together with those that are more local, both in terms of time and space.

3 Methodology

3.1 Data Preprocessing: Fourier Expansion

In each pixel, we may consider the process of the Erlang measures over time, which can be thought as a continuous process in time describing the average number of mobile phones using the network in that site (see equation 2). An example of the observed Erlang profiles along time is shown in Figure 1 (middle): 100 sites have been randomly selected in the lattice, and the Erlang measures recorded in each selected site have been plotted as a function of time. It is clear from the picture that, beside macro periodic behaviors due to week and daily-seasonality in the average use of mobile phone, Erlang data present strongly localized features.

Indeed, in each site of the lattice we observe a discrete version of the Erlang continuous process, recorded every quarter of an hour: due to discontinuities in information provided by the network antennas, the Erlang measure is missing at some time intervals, and hence the time grid of the Erlang measurements is non–uniform. Moreover, some records are negative, due to measurement errors,

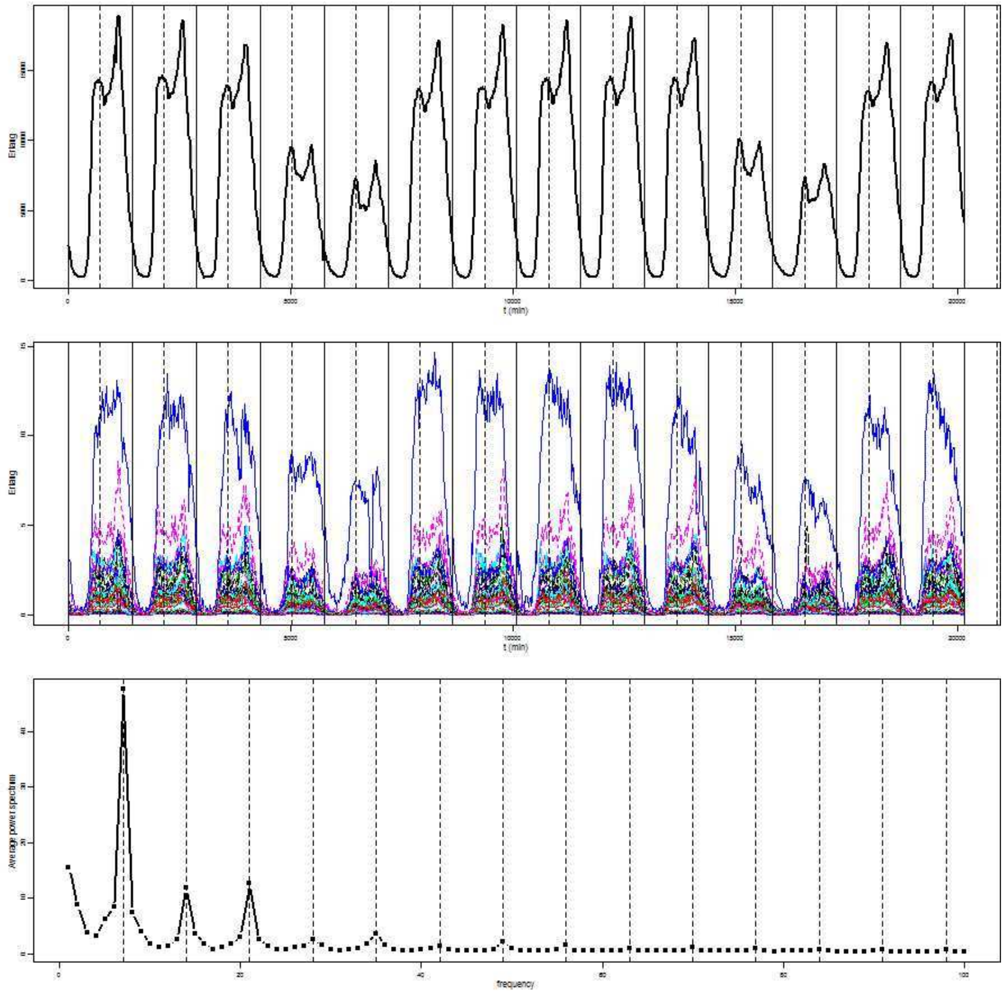

Figure 1: Erlang data: (top) the aggregated Erlang of the investigated area as a function of time; (middle) a random selection of 100 Erlang data, drawn at random among the sites of the lattice, as a function of time; the solid vertical lines are drawn at midnight of each day, and the dotted vertical lines at midday. The first day is Wednesday March 18, 2009. (bottom) average power spectrum; dotted vertical lines are drawn for multiples of 7.

and should be treated as missing values. We thus need to choose a proper basis expansion to reconstruct the functional form of the time-varying Erlang data on a common grid of time values.

To this purpose, we perform a pixel-wise smoothing of the Erlang data via a Fourier basis expansion of period one-week. The idea is to represent each Erlang profile as a function in time obtained as a weighted sum of sinusoids of increasing frequency. Formally, the reconstructed functional form of the Erlang profile relative to the pixel $\mathbf{x} \in S_0$ is a function $E_{\mathbf{x}}(t)$ such that

$$E_{\mathbf{x}}(t) = \frac{c_0^{\mathbf{x}}}{2} + \sum_{h=1}^{H} \left[a_h^{\mathbf{x}} \cos(h\omega t) + b_h^{\mathbf{x}} \sin(h\omega t)\right], \tag{3}$$

where $t \in [0; T]$, $\omega = 2\pi/T$ and $T = 60 \cdot 24 \cdot 7$ is the period expressed in minutes with the coefficients $c_0^{\mathbf{x}}$, $a_h^{\mathbf{x}}$ and $b_h^{\mathbf{x}}$ estimated through ordinary least squares.

In Figure 1 (bottom) the average power spectrum of the Telecom Italia database is reported. This plot shows, for each frequency, the relevant average contribution of the corresponding sinusoid to the Erlang profiles observed within the investigated area. From a graphical inspection of the plot, it is clear that the frequencies significantly contributing to the Erlang time variation are the smaller ones (all less than 7), capturing the difference among days or blocks of days (e.g., the working and weekend days variation), and the frequencies multiple of 7, capturing the recurring daily dynamics. Note that if only the frequencies multiple of 7 were present, the Erlang profiles would be daily-periodic. For an extensive description of smoothing procedures for functional data we refer to Ramsay and Silverman (2005).

3.2 Dimensional Reduction: Treelet Decomposition

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

In this work, we assume that a limited number of time–varying basis functions, common to the entire area under investigation, are sufficient to describe pixel-by-pixel all the corresponding Erlang profiles. Indeed we will interpret the basis functions as describing the Erlang profile associated to a specific temporal dynamic related to a human activity. More formally, we assume the following model for the generation of Erlang data

$$E_{\mathbf{x}}(t) = \sum_{k=1}^{K} d_{\mathbf{x}k}\psi_k(t) + \epsilon_{\mathbf{x}}(t), \tag{4}$$

where $\{\psi_1(t), \ldots, \psi_K(t)\}$ is the set of time–varying basis functions and $d_{\mathbf{x}1}, \ldots, d_{\mathbf{x}K}$ describe their contribution to the Erlang profile relative to the pixel $\mathbf{x}$. The quantity $\epsilon_{\mathbf{x}}(t)$ represents an error term describing unstructured variability of the Erlang data.

In the statistical literature, the process leading to the identification of the finite dimensional basis $\{\psi_1(t), \ldots, \psi_K(t)\}$ and of the coefficients $d_{\mathbf{x}1}, \ldots, d_{\mathbf{x}K}$ is known as *dimensionality reduction*. A very common procedure for dimension reduction is Principal Component Analysis Ramsay and Silverman (2005). In this work we use a method known as *treelet analysis* introduced in Lee et al. (2008). Before describing the details, it is worth noticing that differently from Becker et al. (2011) and Calabrese et al. (2011) where CDR data are analyzed and studied from a Lagrangian perspective (i.e, the atoms of the analysis are users and the focus is in the identification of a limited number of prototypical users associated to interpretable time-patterns), here Erlang data are analyzed and studied from an Eulerian perspective (i.e, the atoms of the analysis are pixels and the focus is identifying a limited number of prototypical pixels associated to interpretable time-patterns).

Treelets (i.e., the estimates of the basis functions $\psi_1(t), \ldots, \psi_K(t)$) have been originally proposed as a surrogate of wavelets for dealing with unordered variables. Nevertheless, we found them to be an effective dimension reduction technique for Erlang profiles and, more generally, for data with peculiar functional features, like spikes, periodicity, outliers.

Similarly to wavelets, the treelet decomposition has the property of following a hierarchical structure interpretable in a multiscale framework. Differently from wavelets, treelets are data-driven. More specifically, the treelet analysis generates a sparse multiscale orthonormal set on functions iteratively detected through nested pairwise Principal Component Analysis. See Lee et al. (2008) for further details.

Once the treelets $\psi_1(t), \ldots, \psi_K(t)$ have been identified, for each pixel $\mathbf{x} \in S_0$ their respective contributions $d_{\mathbf{x}1}, \ldots, d_{\mathbf{x}K}$ to the Erlang profile $E_{\mathbf{x}}(t)$ are obtained by orthogonal projection.

3.3 Spatial Smoothing: Bagging Voronoi Tessellations

The contribution $d_{\mathbf{x}r}$ of the r-th treelet $\psi_r(t)$ to the local Erlang profile $E_{\mathbf{x}}(t)$ is expected to vary smoothly in space because of the spatial dependence between Erlang profiles recorded in close sites which is induced by the arbitrary segmentation of the area and by the mobility of phone users. Thus an improved estimate of $d_{\mathbf{x}r}$ for pixel $\mathbf{x}$ can be obtained by borrowing information from neighboring pixels. In an urban setting, the identification of an optimal neighborhood system is not a trivial issue because of the unisotropic and dishomogeneous nature of the urban matrix. Indeed, detecting close sites is more an aim of the analysis than a starting point.

For this reason we decided to exploit spatial dependence in a fully non parametric setting using a Bagging strategy based on Voronoi tessellations proposed in Secchi et al. (2012). We refer to this paper for a deeper understanding of the rationale behind the Bagging Voronoi Tessellation strategy, which is however easily described:

(i) we build a neighborhood system by randomly generating a Voronoi tessellation covering the entire area under study;

(ii) for each neighborhood (i.e. for each element of the tessellation) we exploit spatial dependence by computing the median of the values $d_{\mathbf{x}r}$ relative to the pixels $\mathbf{x}$ within the neighborhood. We attribute the value of that median to each pixel $\mathbf{x}$ within the neighborhood;

(iii) we then repeat steps (i) and (ii) B times (known as bootstrap replicates).

At the end of the B bootstrap iterations, to each pixel $\mathbf{x}$ corresponds a sample of B medians; this sample is summarized by its median $\hat{d}_{\mathbf{x}r}$ which provides an improved estimate of $d_{\mathbf{x}r}$, taking into account spatial dependence. By plotting, for each pixel in the lattice S_0, the value $\hat{d}_{\mathbf{x}r}$ we obtain a smooth surface describing the variation in space of the contribution of the treelet $\psi_r(t)$ to the local Erlang profiles and thus identifying regions within the urban matrix that are similar with respect to the human activity characterized by the treelet $\psi_r(t)$.

On the whole, the entire methodology allows us to identify:

- a reference basis $\{\psi_1(t), \ldots, \psi_K(t)\}$, i.e the set of basis functions describing the specific effects on the Erlang data of some human activities recorded in the area. In the top panels of Figures 2-7 a selection of the most easily interpretable treelets is reported;
- a set of maps, i.e. the set of spatially–varying functions $\{\hat{d}_1(\mathbf{x}), \ldots, \hat{d}_K(\mathbf{x})\})$ showing pixel-by-pixel the contribution of each treelet to the local Erlang profile. In the top panels of Figures 2-7 the maps corresponding to the treelets illustrated on top are reported.

4 Case study: experimenting Milan mobility patterns

4.1 Case study: Milan

In our work, urban planning expert knowledge showed the potential of this methodology. We discuss in the present paragraph the specific case of the Milan urban region. We analyzed and mapped "hidden mobile phone use patterns" derived from the treelets analysis in order to verify the potential of this method for explaining spatial urban usage and mobility patterns.

Milan is an urban region which goes far beyond its administrative boundaries. The core city and the whole urban area have been affected in the last 20 years by relevant changes in their spatial structures and have generated new relationships between centre and suburbs. Daily mobility patterns are now even more complex than in the past when a hierarchical structures of cities was present and the physical relationship between jobs and homes was the main reason of mobility. Now the commuter flows describe only a minor part of the overall urban movements (about the 29 percent, excluding returning home). Daily mobility is generated by many other reasons which are becoming increasingly relevant. These non-systematic flows are related to individual habits and are the effects of diversified and complex uses of the Milan urban region. For the intrinsic characteristics of this kind of mobility, it is difficult to measure its dimension and its intensity, in space and in time and systematic studies or sources which provide this information in Italy do not exist. Within the Milan urban region services and activities are distributed in a wide territory and there is a plurality of places with specific meanings for mobile populations. At the moment, the urban region of Milan is a densely populated, integrated area where 4.000.000 inhabitants live, where there are 370.000 firms, covered by huge flows of people moving daily in this wide area. Mapping overall mobility in space and in time therefore requires new data sources, able to adequately describe mobility patterns.

4.2 Testing the treelets

Among the dozens of treelets produced applying the methodology explained in the previous section, we selected some results as significant for explaining specific mobility and city usages patterns and tested their significance and their interpretation from an urban analysis and planning perspective at the Milan scale.

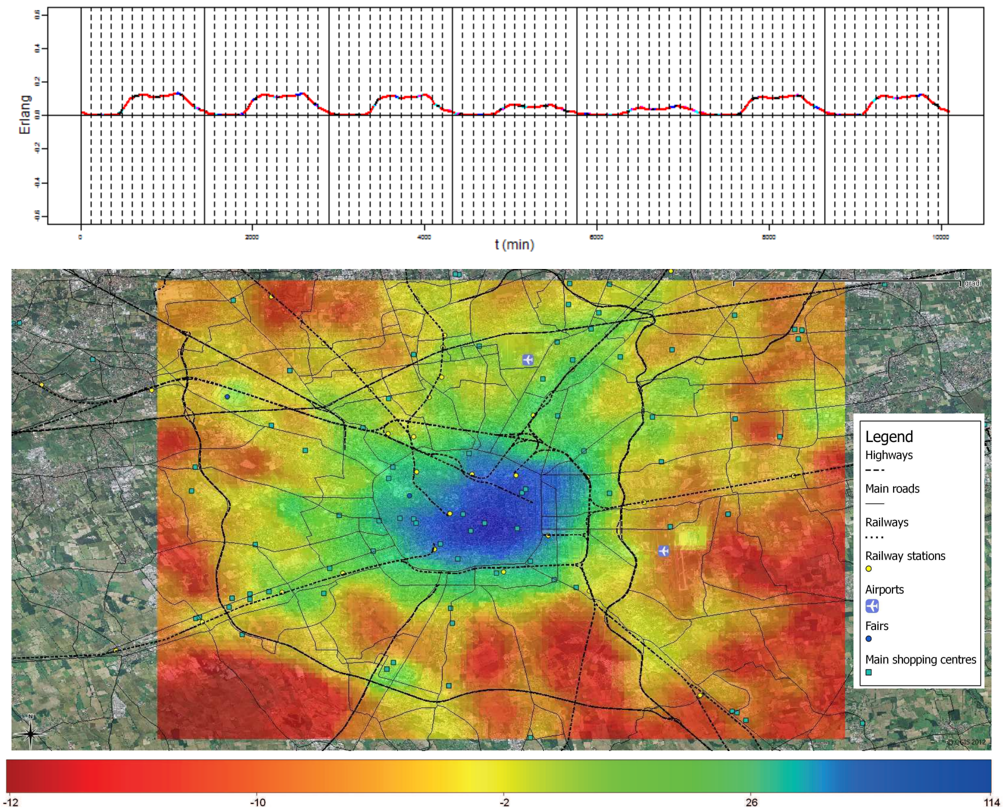

Figure 2: Treelet 1 - The “average use” treelet map. The treelet contains different temporal patterns of mobile phone activity (i.e. daily, working day versus week end) that fit with actual city usage.

On each image we added infrastructures (railways and main roads), main shopping centres, railway stations, localization of the city airport and of the fair trade centre in order to facilitate the interpretation of the map.

The “average use” treelet map (Figure 2) highlights some urban districts characterized by specific telephonic patterns that are compatible with the real urban structure of the region. The treelet contains different temporal patterns of mobile phone activity (i.e. daily, working day versus week end) that fit with actual city usage. In particular we can observe the highest values in the Milan city centre and in others neighbourhoods where there is a strong attraction of urban populations during working day and, with minor intensity, during week end.

In other suburban districts, the intensity is lower, due to the presence of a less relevant mobile phone activity. In general we can conclude that the emerging spatial patterns represent well the highly populated areas versus the poorly populated areas. The mobile phone activity and the urbanized area produce in fact a similar image of the region.

The proposed methodology shows its advantages when we try to face with other, less evident spatial patterns which are difficult to intercept through traditional data sources.

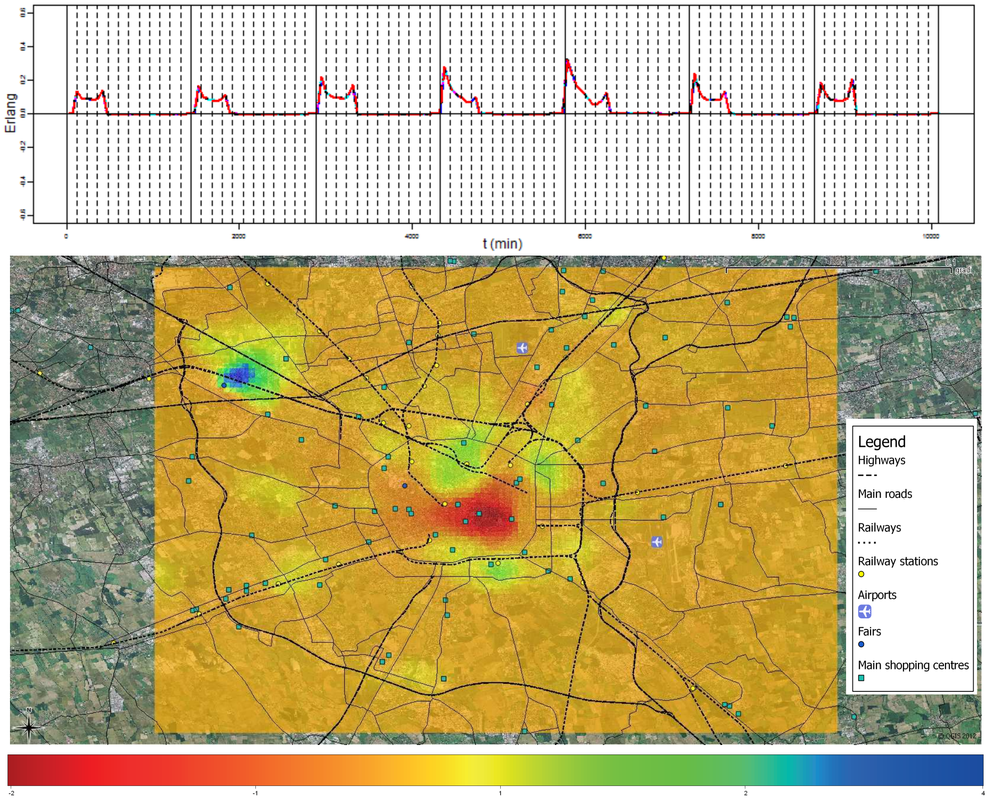

Figure 3: Treelet 2 - Nightly activity. Hot spots highlight the presence of night work: organization of the International Milan Design Week Exhibition district (North-West); delivering and distributing products in the Fruit and Vegetable Wholesale market (second circular ring of the city - South East); scarce nighly activity inside the city centre.

Figure 3 (Treelet 2) is about the density of mobile phone activity late at night (in particular from midnight until 8 am). We can observe here some interesting hot spots where the values are very high. For example, the exhibition district in the Northern Western side of the map. In the considered period an important Fair (the 2009 International Milan Design Week) was held and the peak fits well with the nightly activities necessary for the mounting and the organization of the site. Another point of interest is the Fruit and Vegetable Wholesale market in the South Eastern part of the region where consistent night work happens for delivering and distributing products that come from whole Italy and abroad. The city centre is characterized by a relative low value, according to the absence of relevant nightly activity inside it.

Figure 4 (Treelet 33) puts in evidence some locations with high concentration of mobile phone activity during the evening of the working days and during daytime (from 8 am until 8 pm) of the week end. It shows a significant correspondence with main residential districts of the Milan urban region. It highlights a relevant concentration of homes along the second circular ring of the city, where the density of resident population reaches the highest value of Milan, but also in some municipalities with a residential profile and social housing in the south,

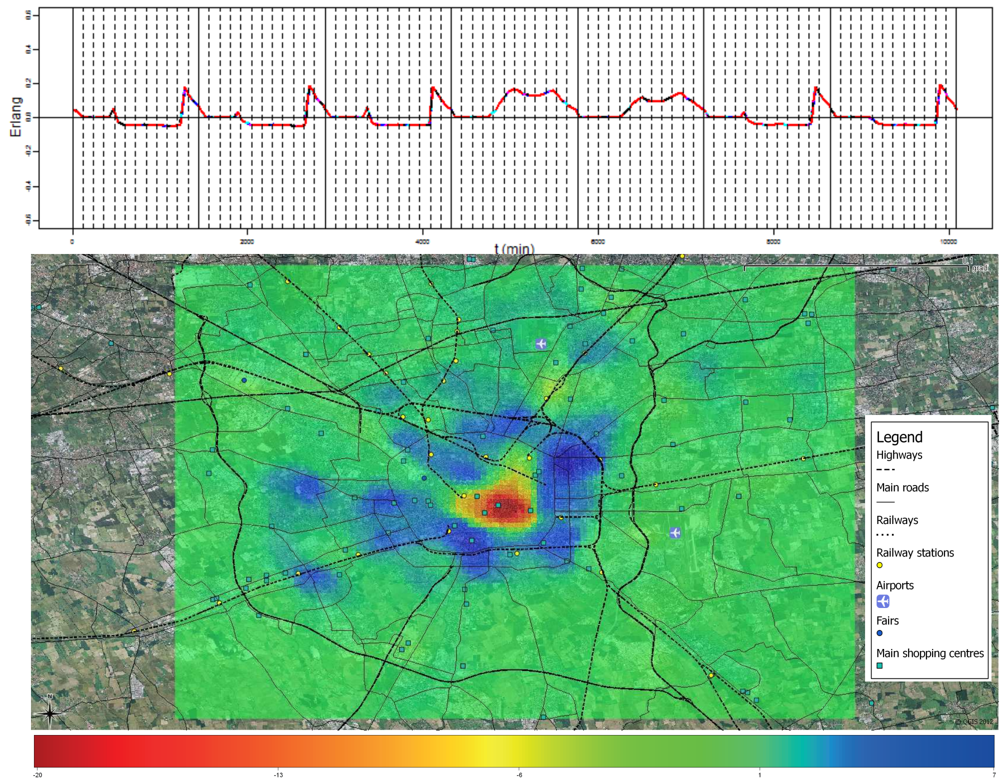

Figure 4: Treelet 33 - Concentration of activities during the evenings of working days and during daytime (from 8 am until 8 pm) of the week end: residential districts of the Milan urban region.

south-west and in the north of the metropolitan area (Corsico, Rozzano, Sesto S.G). The Milan city centre appears as a void and this is consistent with the changes that occurred in the last decades, namely a gradual replacement of the residents with activities mainly related to the service and the commercial sectors.

Figure 5 (Treelet 78) shows places with high density of activity during Saturday evening, from 8 pm until midnight. Focusing on the core city area, we notice several interesting patterns: a high activity in some places where there are many pubs and restaurants near the Milan Central Station, in the Navigli District, in the Isola Quarter and in other ambits characterized by the presence of leisure spaces (Filaforum Assago in the south of Milan) but also of activities in a continuous cycle as the hospitals. This treelet has proven to be effective in describing the temporal profile of the city lived by night populations during Saturday.

Figure 6 (Treelet 82 and Treelet 83) concerns more directly the representation of specific mobility patterns evident at the Milan urban region scale during working days, i.e. commuting flows. In fact, the Treelet 82 map is about the concentration of mobile phone activity during the morning rush hours and the Treelet 83 regards the activity during the evening rush hours. The emerging spatial patterns are quite different and show several interesting mobility practices within and outside the city. In fact, during the morning, we observe a concentra-

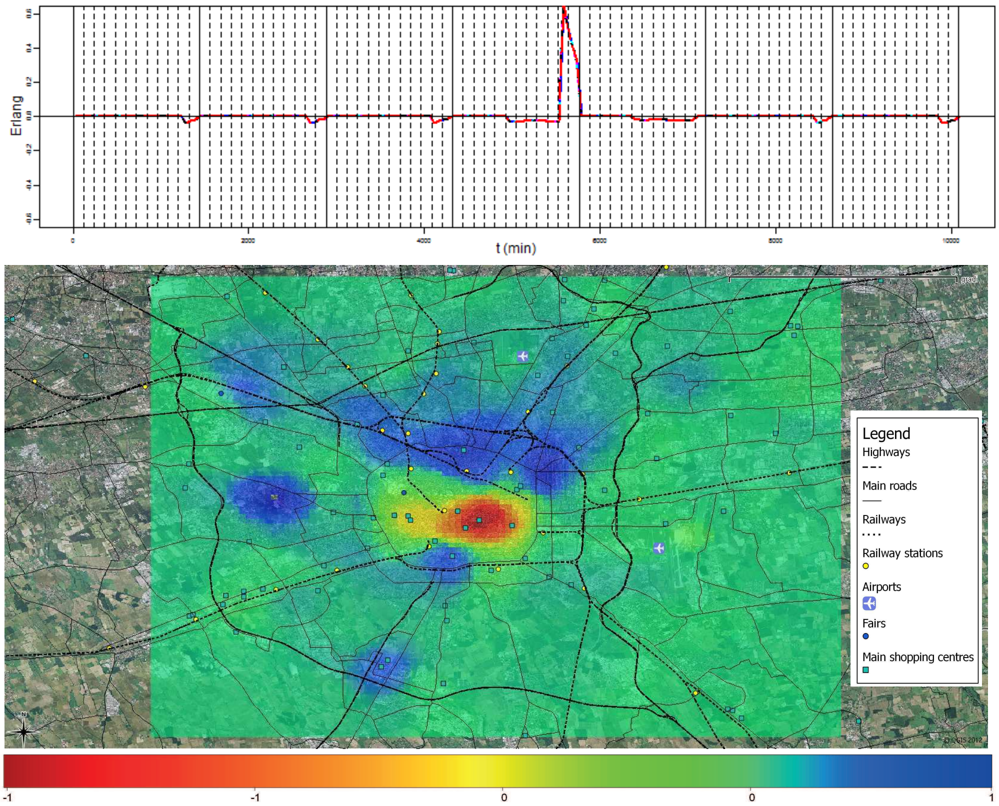

Figure 5: Treelet 78 - Density of activity during Saturday evening (8pm-midnight). Saturday night population: leisure and hospitals.

tion of traffic along the main roads, in proximity of relevant high roads junctions and in the surrounding areas. We can put this trend in relation with the daily commuting flows of people moving from homes, located in a wide area around Milan to job places. From an urban analysis point of view, it can be seen as a representation of the overall traffic generated mainly by cars, from 8 am to 10 am from Monday to Friday. Treelet 83 measures the mobile phone activity during the evening rush hours, which are longer than the morning ones, since they last more than 4 hours, from 4 pm to 8 pm. In this case, the hot spots are mainly located within the road ring. The map well represents the complex mobility pattern related to the exit from workplaces, when, before going home, chains of daily shifts take place, linked to a number of social practices (shopping, going to the gym, go get a family member or friend). The chain of daily moves becomes more articulated, and the daily rush hours are dilated. As it emerges from traditional sources Regione Lombardia, Direzione Generale Infrastrutture e mobilità (2002): the individual daily displacements in the Province of Milan are 2,55 moves/person, with an average of 2 moves in sequence. The spatial pattern puts in evidence a relative low intensity in the city centre and an increasing density of activity in proximity of the main shopping centres, commercial streets (Vigevanese), and radial connections moving outward.

Figure 7 (Treelet 93) highlights another relevant mobility pattern, which is

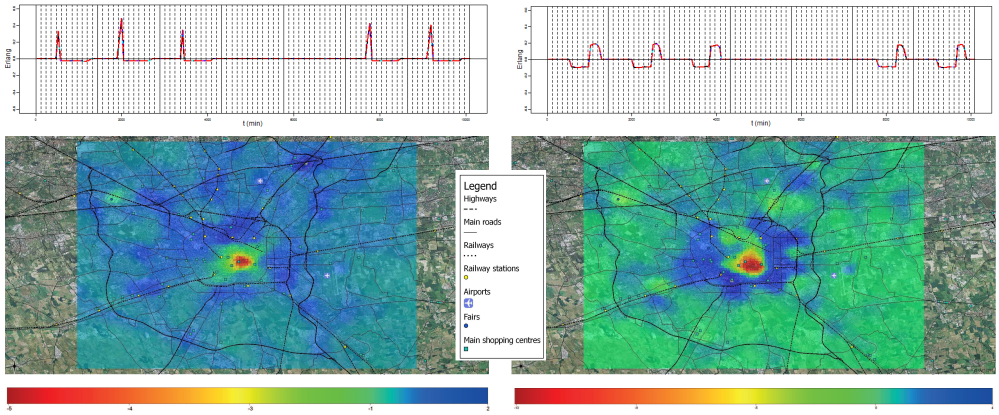

Figure 6: Treelet 82 (left) and 83 (right) - Mobility practices. Weekdays commuting flows at the Milan urban region scale: morning rush hours (letf) vs to evening rush hours (right).

difficult to intercept through database traditionally used in urban studies: the shopping activity and in general the leisure activity. The map represents the density of mobile phone use during Saturday, from 10 am to 8 pm. Shopping and leisure are two of the main reasons of mobility in contemporary cities: they belong to the category of non systematic mobility, and they significantly contribute to the even more complex mobility patterns that can be observed in the Milan urban region due to the distribution of commercial centres, commercial streets and, in general, of activities (museums, touristic sites, cinemas, just to cite some) inside and outside the city. These places attract, especially in certain days of the week, a huge amount of population coming from a vast territory that goes far beyond the administrative boundaries of the city. The map is the result of this spatial pattern and shows an important concentration of mobile phone traffic in the city centre and in other several places outside the city (most of them corresponding to the presence of commercial centres). The mainly residential areas, recognized in the previous Figure 6, are consequently characterized by the lowest value.

5 Future works

The research allowed us to test the potential of the treelet decomposition analysis in explaining relevant urban usage and mobility patterns at the Milan urban scale. We plan to improve the integration of traditional database (i.e. land cover maps, distribution of activities, infrastructures and transport junctions) with mobile phone data pattern in order to reach a less descriptive and a more synthetic classification of the urban space according to its temporal and spatial usages, that could be useful for understanding the dynamic of temporary populations and of mobility patterns and for promoting more specific urban policies. This task may be possible from the recognition of mobile populations which are given the opportunity to choose among alternative forms of available mobility which can offer the greatest flexibility, range of connections, reversibility and the best means of accessing the various resources and destinations possible, but also which offer more oriented services.

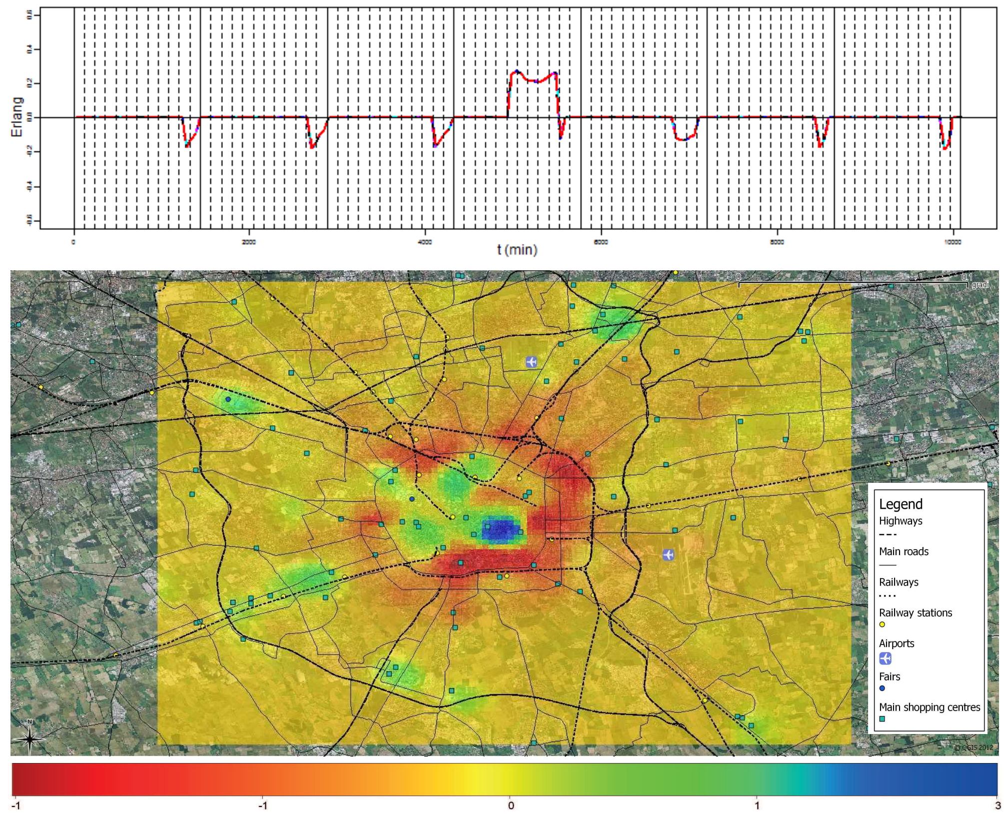

Figure 7: Treelet 93 - Mobility practices. Saturday (10am- 8pm), shopping and leisure activity.

Acknowledgment

The authors would like to acknowledge Piero Lovisolo, Dario Parata and Massimo Colonna, Tilab - Telecom Italia for their collaboration during the research project. This work was supported by Telecom Italia. We also thank Paolo Dilda for helping us in the preparation of the figures.

References

Ahas, R. and Mark, Ü. (2005), "Location based services–new challenges for planning and public administration?" *Futures*, 37, 547–561.

Becker, R. A., Caceres, R., Hanson, K., Loh, J. M., Urbanek, S., Varshavsky, A., and Volinsky, C. (2011), "A Tale of One City: Using Cellular Network Data for Urban Planning," *IEEE - Pervasive Computing*, 10, 18–26.

Calabrese, F., Lorenzo, G. D., Liu, L., and Ratti, C. (2011), "Estimating Origin-Destination Flows Using Mobile Phone Location Data," *IEEE - Pervasive Computing*, 10, 36–44.

Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008), "Understanding individual human mobility patterns," *Nature*, 453, 779–782, m3: 10.1038/nature06958; 10.1038/nature06958.

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

Lee, A. B., Nadler, B., and Wasserman, L. (2008), "Treelets – An adaptive multi-scale basis for sparse unordered data," *The Annals of Applied Statistics*, 2, 435–471.

Nuvolati, G. (2003), "Resident and Non-resident Populations: Quality of Life, Mobility and Time Policy," *The journal of regional analysis & policy*, 33(2), 67 –83.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, Springer.

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

Reades, J., Calabrese, F., Sevtsuk, A., and Ratti, C. (2007), "Cellular Census: Explorations in Urban Data Collection," *IEEE Pervasive Computing*, 6, 30–38.

Regione Lombardia, Direzione Generale Infrastrutture e mobilità (2002), "Indagine Origine/Destinazione regionale 2002 - sintesi," Tech. rep., Regione Lombardia.

Secchi, P., Vantini, S., and Vitelli, V. (2012), "Bagging Voronoi classifiers for clustering spatial functional data," *International Journal of Applied Earth Observations and Geoinformation*, doi: 10.1016/j.jag.2012.03.006.

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

Soto, V. and Frías-Martínez, E. (2011a), "Automated land use identification using cell-phone records," in *Proceedings of the 3rd ACM international workshop on MobiArch*, ACM, pp. 17–22.

— (2011b), "Robust Land Use Characterization of Urban Landscapes using Cell Phone Data," The First Workshop on Pervasive Urban Applications (PURBA).

MOX Technical Reports, last issues

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

23/2012 Fabio Nobile, Christian Vergara
Partitioned algorithms for fluid-structure interaction problems in haemodynamics

25/2012 Manfredini, F.; Pucci, P.; Secchi, P.; Tagliolato, P.; Vantini, S.; Vitelli, V.
Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the Milan urban region

24/2012 Antonietti, P.F.; Giani, S.; Houston, P.
hp Version Composite Discontinuous Galerkin Methods for Elliptic Problems on Complicated Domains

22/2012 Ettinger, B.; Passerini, T.;Perotto, S.; Sangalli, L.M.
Regression models for data distributed over non-planar domains

21/2012 Guglielmi, A.; Ieva, F.; Paganoni, A.M.; Ruggeri, F.; Soriano, J.
Semiparametric Bayesian models for clustering and classification in presence of unbalanced in-hospital survival

20/2012 Ieva, F.; Paganoni, A. M.; Zanini, P.
Statistical models for detecting Atrial Fibrillation events

19/2012 Faggiano, E.; Antiga, L.; Puppini, G.; Quarteroni, A.; Luciani G.B.; Vergara, C.
Helical Flows and Asymmetry of Blood Jet in Dilated Ascending Aorta with Normally Functioning Bicuspid Valve

18/2012 Formaggia, L.; Vergara, C.
Prescription of general defective boundary conditions in fluid-dynamics

17/2012 Manzoni, A.; Quarteroni, A.; Gianluigi Rozza, G.
Computational reduction for parametrized PDEs: strategies and applications

16/2012 Cutri', E.; Zunino, P.; Morlacchi, S.; Chiastra, C.; Migliavacca, F.
Drug delivery patterns for different stenting techniques in coronary bifurcations: a comparative computational study