Electronic Journal of Applied Statistical Analysis
Vol. 00, Issue 00, September 2014, 1-10
DOI: DOI–xxxx

Big Data to Monitor Big Social Events: Analysing the mobile phone signals in the Brescia Smart City

Maurizio Carpita [*a] and Anna Simonetto [†a]

[a] *DMS StatLab - Data Methods and Systems Statistical Laboratory, Department of Economics and Management, University of Brescia, C.da Santa Chiara, 50 - 25122 Brescia (Italy)*

September 15, 2014

We present the first experience, carried out in 2013, devoted to the extraction of geo-referenced statistical information from mobile phone signals to monitor big social events in Brescia, a mid-sized city located near Milan (Lombardy, Italy). The statistical analysis is based on the Telecom Italia CityLive database, used to define raster structures and space-time profiles with the R open source language. The obtained results for the two big events held in Brescia in 2013 (the car race "Mille Miglia" and the bicycle race "Giro d'Italia") have proved very useful to this municipal administration to plan future events, and more generally to develop policies for the "smart city".

keywords: Mobile phone data, Raster objects, Geo-statistics with R.

1 Big social events monitoring in the smart city

Nowadays, thanks to the most recent information and communication technologies (ICT), it is possible to measure urban mobility with a high degree of precision, in order to improve the citizens' quality of life [(4), (6), (10), (12)]. The focus on this aspect falls within the concept of smart city, which has considerably grown in the last decade, together with the big data challenge [(2), (3), (7), (8), (9), (14), (15)].
Digital data from mobile phone networks and analytic tools for the analysis of mobility practices in the modern cities are useful to discover the dynamics and time-variation of

[*]Corresponding author: maurizio.carpita@unibs.it
[†]anna.simonetto@unibs.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

people mobility, and the use of urban spaces by permanent and temporary populations. The standard temporary (but periodic) population of a city is generated by home-work mobility behaviours. Other temporary (but episodic) populations rise in the case of big social events, as culture initiatives, musical concerts and sport competitions. In our study we considered data and statistical measures derived from this second type of temporary populations of a city. For both permanent and temporary populations, digital data from mobile phone networks can be merged with data from classical sample surveys, based on questionnaires: ICT and analytics are used for integration and extraction of useful statistical information from these very different data sources. The data science approach uses the results of these modern data analyses to increase the knowledge about the "live city" and to improve the quality of life of its (permanent and temporary) populations in a variety of ways: through measures that promote eco-friendly and sustainable environments, the delivery of connected public services and the effectiveness of the territorial marketing. All these benefits contribute to the development of the "smart city" idea.

Since 2013, the Municipality of Brescia has taken several initiatives related to smart city approach (osservatoriosmartcity.it/brescia/). In particular, it has been one of the first Italian municipalities to take an interest to use mobile phone data to improve the quality of the big social events that take place in the city. In this regard, the Municipality started a collaboration with DMS StatLab - Data Methods and Systems Statistical Laboratory of the University of Brescia to implement a system to analyse this type of data and develop of useful reports for territorial marketing. The interesting challenge concerns processing and analyzing a very really high mass of data.

2 Mobile phone signals, big data analysis and geo-statistics

Thanks to a special agreement between the Municipality of Brescia and "Telecom Italia" (the biggest Italian telephone operator), we could analyse the presence of users connected to the "Telecom Italia Mobile" (TIM) network (not only its customers, but all the users of its roaming service) in the city during the week of the "Mille Miglia" (13-19 May 2013) and the week that ended with the arrival of the "Giro d'Italia" (20-26 May 2013). The provided data are anonymous, aggregate, and collected in full respect of privacy and of national legislation in force.

Fig. 1 represents the big data-information flow related to our analysis. The geo-referenced mobile phone data are divided into pixels of about 150 metres, each of them identified by the latitude and longitude of the centroid (Fig. 1-left). For each pixel, a measure of the SIM (Subscriber Identity Module) density related to a certain period of time (in our case every half hour) is provided (for further details see (12) and (13)).

The estimate of the pixel SIM density is based on:

- knowledge of the last cell of the mobile network used by a SIM;
- the coverage of the mobile network, calculated by the planning software of TIM;
- the probability of telephone traffic in pixel.

The CityLive platform of Telecom Italia provides the extraction, transformation and loading (ETL) and storage systems, so that data are easily transferred to the server of the Municipality of Brescia (Fig. 1-centre). Finally, the DMS StatLab develops data analyses, maps and reports useful for city's policy decisions (Fig. 1-right).

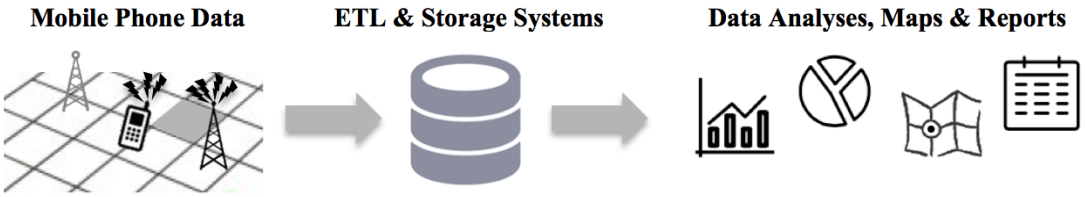

Figure 1: The big data-information flow

The proprietary algorithm used by Telecom to estimate the presence of active mobile phones ensures that the SIM density provides a good approximation of the number of users actually connected to the network in a certain time span only in the case of sufficiently large areas.

Our analysis focused on an area which includes the historic centre of Brescia (about 2.4 km to 1.8 km , Fig. 2-left), for which we have an estimate of the density of SIM (TIM or roaming on TIM) (Fig. 2-right) at regular intervals of 30 minutes (i.e. 48 time splits per day). Therefore, for the 26 days of May we considered, Telecom Italia provided 1,248 file (about 204 Mega bytes, also considering the missing data).

We used the open source R language, useful for geo-statistics [(1), (11)]. The structure that best allows the manipulation and analysis of these data is the raster. The raster defines a grid subdividing the space into rectangles, called 'cells' or 'pixels', and it permits to store one or more values for each of these cells. The R package used for the analysis is `raster` [(5)] and it is based on an object called, precisely, raster. This object consists mainly of two parts. The first is the header and it contains the metadata of the structure, such as the number of rows and columns, the spatial extension and, where applicable, the name and path of the files of the raw data.

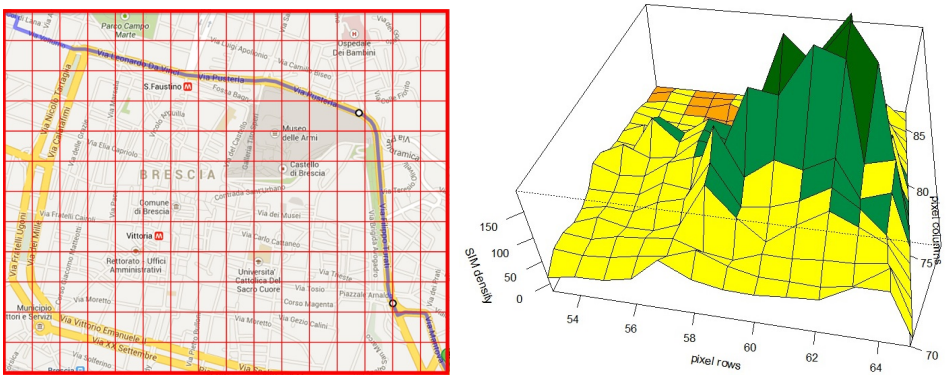

Figure 2: Left: Brescia historic centre; Right: SIM density on May, 16^{th} at 16:00

Fig. 3 shows the header of a raster containing 4 time instants (layers) recorded in May 16, 2013 (the day of the departure of the car race "Mille Miglia").

```
> Density
class       : RasterStack
dimensions  : 923, 607, 560261, 4  (nrow, ncol, ncell, nlayers)
resolution  : 0.001666667, 0.00125  (x, y)
extent      : 9.833334, 10.845, 45.20542, 46.35917  (xmin, xmax, ymin, ymax)
coord. ref. : NA
names       : Brescia_May_16_0800, Brescia_May_16_0830, Brescia_May_16_0900, Brescia_May_16_0930
min values  :                  0,                  0,                  0,                  0
max values  :              65535,              65535,              65535,              65535
```

Figure 3: The header of a raster containing 4 time istants (layers)

The information in Fig. 3 allows us to locate the raster. The object belongs to the class Rasterstack, because it includes 4 layers (in our case, 4 half-hours). Each layer is composed of 923 pixels horizontally and 607 pixels vertically, for a total of 560,261 cells. Extent represents the spatial extension of the raster, showing the coordinates of the extreme North-West and the extreme South-East. Furthermore, the step for the length and width of each cell / pixel is visible in resolution. Finally, the header reports the names of the files containing the raw data, and for each file (ie each layer), the minimum and maximum value (in our case the minimum and maximum of SIM density, where 65535 represents a missing value). The second part of raster object is given by the raw data, i.e. the values related to each single cell (for example, in our case the SIM density).

These data can be saved jointly with the structure or kept separate. The `raster` package offers two possibilities. In the first case it is possible to save the "raster" object in "gri-grd" format. As a result we obtain two files containing both header and all the raw data. With this choice we will get files rather heavy, but easily movable from one device to another. Alternatively, it is possible to save the "raster" object as "RData" format, containing only the header. This file is very small and it targets to external files containing data (a file for each layer of the raster, in our case one for each time instant of detection). With this option, it is possible to work with very large datasets, without taking up too much RAM, but the possibility of use from different locations is limited.

3 The analysis of two big social events in Brescia, May 2013: "Mille Miglia" and "Giro d'Italia"

In the first instance we evaluated the daily profiles of SIM density during the month of interest (May 2013) to see whether it was a recognizable (weekly or daily) trend. Fig. 4 represents the density of SIM that are in the historic centre of Brescia, measured every half hour.

The breaks in the profiles due to times with no data collected from the mobile network. Given the considerable size of the historic centre of Brescia respect to the extension of the entire municipality, the SIM density can be considered a good estimate

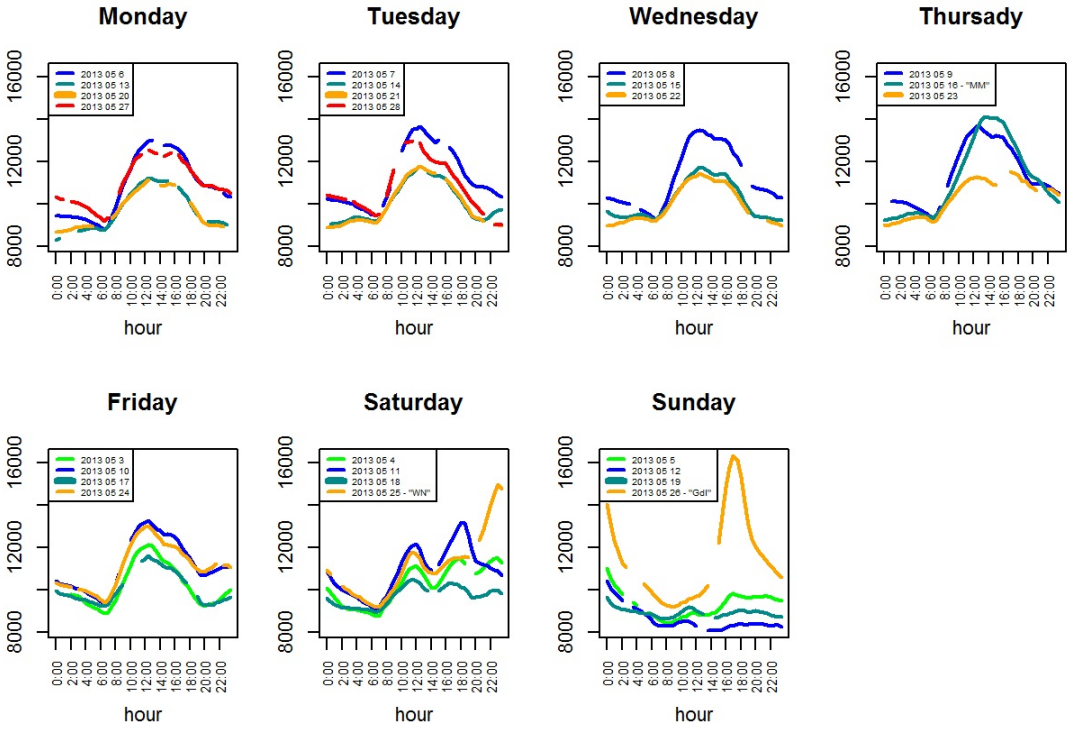

Figure 4: Daily profiles of SIM density in Brescia historic centre on May 2013. The days of "Mille Miglia" (MM), "White Night" (WN), and "Giro d'Italia" (GdI) are highlighted

of the actual number of users connected to the network during the period in the entire municipality.

Considering the weekdays (from Monday to Friday) and holidays (from Saturday to Sunday), it is possible to see that the daily trend is similar between the two groups, even if there is a density level variation at the beginning and at the end of the month. There are three outlier periods, corresponding to the departure of the historical car race "Mille Miglia" (May 16^{th}), the "Notte Bianca" (White Night) in occasion of the cycle race "Giro d'Italia" (May, 25^{th}) and the Sunday of the final award ceremony of the "Giro d'Italia" (May 26^{th}). Starting from these considerations, for the analysis of the impact of the "Mille Miglia" departure, we decided to evaluate the differences in density of that day compared to the mean value of the workdays of that week. A similar analysis was also conducted with reference to the maximum value, but because the results are quite comparable, we present only the first ones.

Looking to Fig. 5, showing the evolution of the SIM density in the whole area of interest, it is useful to assess the impact of major events on the presence in the historic centre of Brescia. We note 3 anomalous peaks. The first occurs on May 16^{th}, days of the "Mille Miglia" departure. We record an increase in SIM density by around 25% compared to the mean of the other weekdays, and 23% compared to the maximum value of them.

The second peak was recorded at the "Notte Bianca" (May 25^{th}, from 20.30 to 22.30).

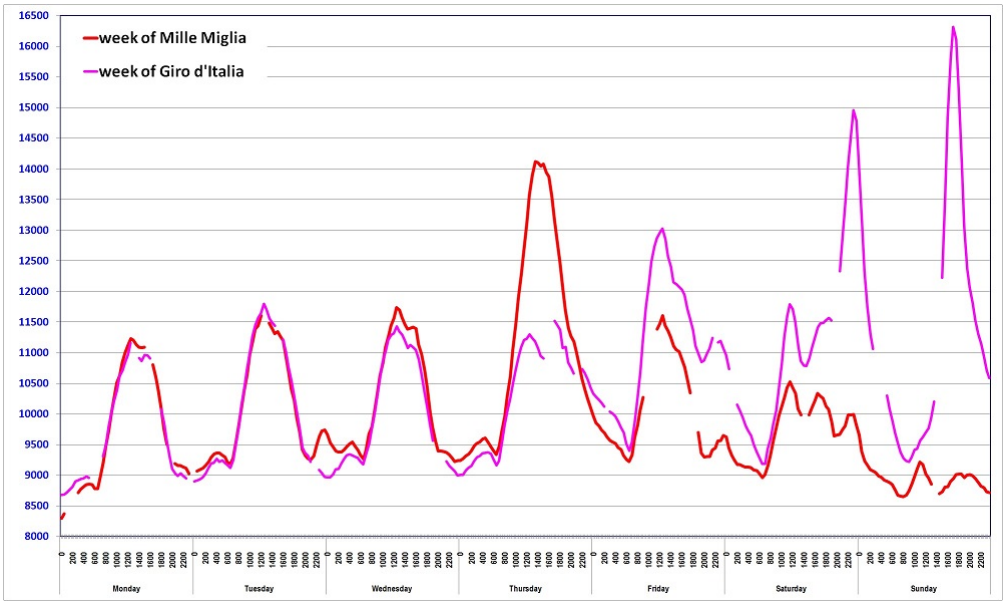

Figure 5: SIM density profiles in the week of the "Mille Miglia" and "Giro d'Italia"

The increase in attendance is estimated at about 26% above the mean of the Saturdays of May and 26% above the maximum of the other days of the same week.
Finally, the last peak corresponds to the final award of the cycle race "Giro d'Italia", in the afternoon of May 25^{th} (from 15.30 to 18.30). In this case the deviation from the mean of the Sundays of May is estimated at about 71% above the mean and 35% above the maximum of the other days of the same week.

Focusing on the day of the "Mille Miglia" departure, Fig. 6 shows the daily trend of the differences of mobile phone traffic of that day from the mean and maximum values of the other weekdays.

It is clear that this event has notably increased the presence of people in the historic centre than the other weekdays because, at the time of maximum affluence, there was an increase of about 25% of population with respect to the mean level.

At this point, however, it becomes essential to deepen the analysis at the pixel level in order to assess how these people are concentrated and which areas are more or less attractive. Also in this case, it seemed useful to provide the municipal administration a graphical tool of analysis to monitor the flow of tourists during the day.

We opted for a three-dimensional representation of the density on this area and we created a chart for each half hour of the day. For simplicity and clarity, we present only 3 graphs (Fig. 7), which however give an idea of the movement of people within the area of the historical centre during the day. This representation, easily overlappable on the map of the historical centre, it is useful to the municipality to plan territorial marketing initiatives or to properly program the changes of road network and the location of public officials, or health care. For example, in the third graph we see a small peak in the South-West area of the map, which corresponds to the point of crowding and departure

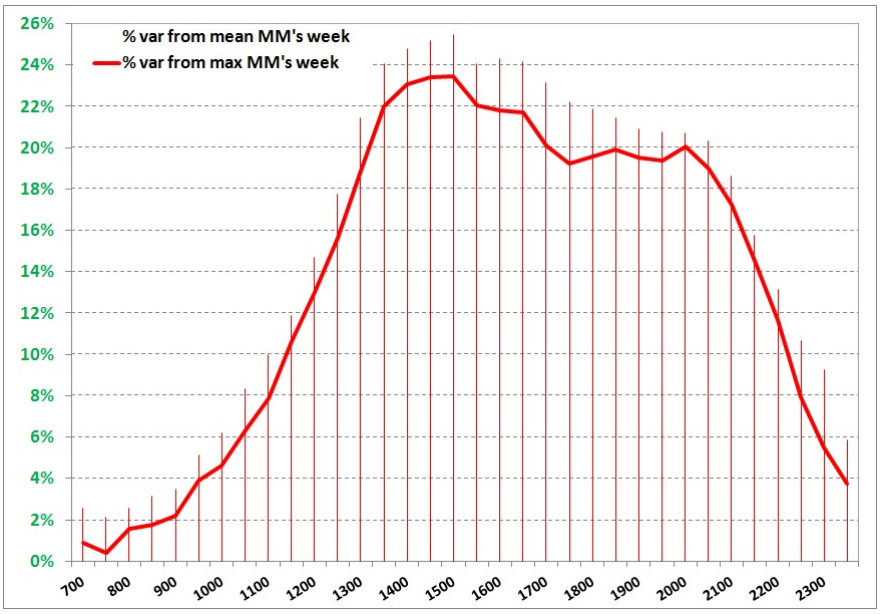

Figure 6: % differences of SIM density on the day of the "Mille Miglia" (MM) from the mean/maximum values of the other weekdays

of historical cars. These considerations are also important for businesses and restaurants that can monitor the flow of potential customers.

In addition, for the "Mille Miglia" the municipality carried out three surveys aimed at tourists (through direct interviews on the day of the race departure), business owners and residents (via paper questionnaires collected in the following weeks). The mobile phone data analysis, together with the results of these surveys, have been used for planning the 2014 edition of the same race. Scarce attendance recorded on the evening of Saturday, the day of race arrival, have provided information support to the decision to move the awards ceremony on Sunday, so as to facilitate the organization of collateral events (in particular the "Notte Bianca" on Saturday).

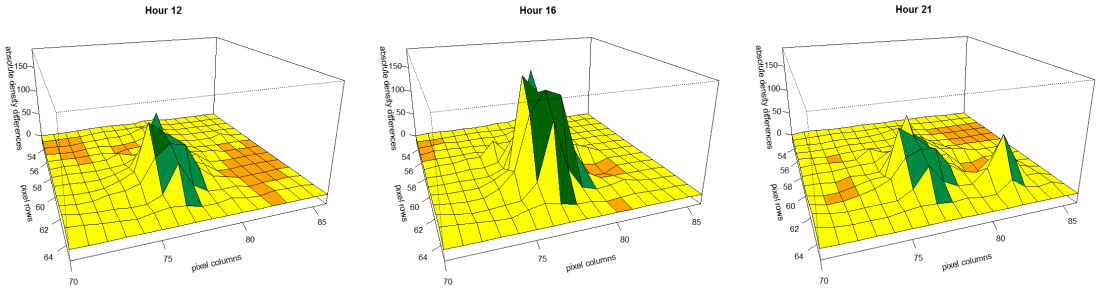

Figure 7: Differences in SIM density on the day of the "Mille Miglia" with reference to the mean value of the other weekdays of the week

3.1 The focus-pixel analysis

In order to improve the transport services and to provide a tool for the detailed analysis to businesses, we carried out in-depth analyses for certain areas of the city centre. In particular, we focused on the "San Faustino" subway Station (the stop for tourists who came from North-West, Fig. 8-left blue), the "Piazza Victoria" Square subway Station (the bus stop for tourists who came from South-East, Fig. 8-left violet) and "Piazza Loggia" (where there was the gathering of historical cars and the majority of exhibition stands, Fig. 8-left green).

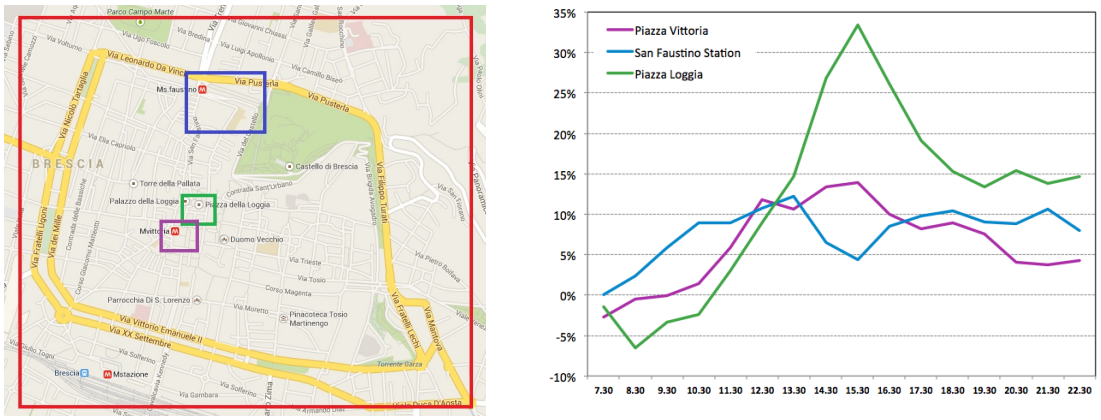

Figure 8: Left: Location of the three focus areas on the map of the historical centre; Right: % differences of SIM density the day of the "Mille Miglia" with respect to other weekdays

Analysing the percentage differences of SIM density on the day of the "Mille Miglia" departure compared to other weekdays (Fig. 8-right), we note that, for the analyzed focus, there is an increase in attendance recorded during all the day. However, this increase is not constant and it follows diversified trends in the three areas, with a considerable peak in "Piazza Loggia".

Another interesting aspect in the analysis of these focus area concerns the possibility of evaluating the coverage of the SIM estimated density offered by TIM in some special moments. Considering for example the Sunday of the awarding of "Giro d'Italia" (from 14 to 22), we know that there were thousands of people for the ceremony in "Piazza Loggia" (Fig. 9-left, photo). Through the appropriate statistical tools we could estimate the maximum deviation at the same time on the other days of the same week (Fig. 9-right).

4 Future Prospects

After this first interesting and useful experience in 2013, the Municipality of Brescia decided to confirm the collaboration with DMS StatLab of the University of Brescia,

Figure 9: Left: Focus-pixel on "Piazza della Loggia" during the award ceremony of the "Giro d'Italia" on May, 16 2013. Right: the hourly profiles (from 14 to 22) of the SIM density compared to the maximum of the other days of the same week

and to signed with Telecom Italia the real-time mobile phone data provision each 15 minutes time interval for the next two years. This first year is the beginning of an important journey that will continue with the monitoring of events planned in the city from May to October 2015, in conjunction with the World Exposition EXPO 2015.
It is worth to recall that the analysis of these type of data (in addition to the analysis of territorial marketing) could be also useful to develop spatio-temporal models dedicated to studies of urban mobility and environmental impact. The premises are therefore very positive. Digital technologies offer the ability to exploit statistical methodologies to transform data into information to facilitate decisions that are, therefore, awarer, more efficient and more effective.

Aknowledgment

Many thanks for their support and data provision to Dr. Marco Trentini, Head of the Sector "Statistics, Audit, Planning and Control Management" of Municipality of Brescia, and to Dr. Piero Lovisolo of Telecom Italia, Technology, Smart Pipe & NetAPI.

References

[1] Bivand R.S., Pebesma E.J., Gmez-Rubio V. (2008) Applied Spatial Data Analysis with R. Use R!, Springer.

[2] Caragliu A., Del Bo C., Nijkamp P. (2011) Smart Cities in Europe, Journal of Urban Technology, 18:2, 65-82, DOI: 10.1080/10630732.2011.601117

[3] Giffinger R., Fertner C., Kramar H., Kalasek R., Pichler-Milanovic N., Meijers E. (2007). Smart Cities - Ranking of European medium-sized cities. Vienna University of Technology

[4] Grauwin S., Sobolevsky S., Moritz S., Gdor I., Ratti C. (2014) Towards a comparative science of cities: using mobile traffic records in New York, London and Hong Kong, in Computational Approaches for Urban Environments. Springer Ed., Senseable City Lab, MIT (http://senseable.mit.edu/papers/pdf/2014-06-Grauwin-ComparativeScience.pdf)

[5] Hijmans R.J. (2014) raster: raster: Geographic data analysis and modeling. R package version 2.2-31. http://CRAN.R-project.org/package=raster

[6] Hoteit S., Secci S., Sobolevsky S., Ratti C., Pujolle G. (2014) Estimating human trajectories and hotspots through mobile data. Computer Networks, in press, (http://dx.doi.org/10.1016/j.comnet.2014.02.011)

[7] ICSU (2011) Health and Wellbeing in the Changing Urban Environment: a Systems Analysis Approach. International Council for Science, Paris. (http://www.icsu.org/publications/reports-and-reviews/health-and-wellbeing/health-and-wellbeing-in-the-changing-urban-environment)

[8] Kamel Boulos M.N., Al-Shorbaji N.M. (2014) On the Internet of Things, smart cities and the WHO Healthy Cities. International Journal of Health Geographics, 13:10, 1-6

[9] Kitchin R. (2014) The real-time city? Big data and smart urbanism. GeoJournal, 79:1, 1-14

[10] Kung K.S., Greco K., Sobolevsky S., Ratti C. (2014) Exploring Universal Patterns in Human Home-Work Commuting from Mobile Phone Data. PLoS ONE 9(6): e96180. doi:10.1371/journal.pone.0096180

[11] Lamigueiro O.P. (2014) Diplaying Time Series, Spatial, and Space-Time with R. The R Series, CRC Press, Taylor and Francis Group.

[12] Louail T., Lenormand M., Cantu Ros O.G., Picornell M., Herranz R., Frias-Martinez E., Ramasco J.J., Barthelemy M.M. (2014) From mobile phone data to the spatial structure of cities. Scientific Reports 4, Article number: 5276 doi:10.1038/srep05276

[13] Manfredini F., Pucci P., Secchi P., Tagliolato P., Vantini S., Vitelli V. (2012) Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the Milan urban region. MOX report 25/2012 (http://mox.polimi.it/it/progetti/pubblicazioni/view.php?id=345&en=en)

[14] Resch B., Zipf A., Beinat E., Breuss-Schneeweis P., Boher M. (2012) Towards the Live City Paving the Way to Real-time Urbanism. International Journal on Advances in Intelligent Systems, vol 5 no 3 & 4, 470- 482

[15] Resch B., Zipf A., Breuss-Schneeweis P., Beinat E., Boher M. (2012) Live Cities and Urban Services - A Multi-dimensional Stress Field between Technology, Innovation and Society, Senseable City Lab, MIT (http://senseable.mit.edu/papers/publications.html)