Authors' replies to Editor comments

Q1: "while it is – correctly – reported that increasing risk has at least two root causes (effects and impacts of CC as well as population dynamics in flood-prone areas) I kindly would like to suggest that you add one or two statements on the latter. Also in Northern Italy we do have distinct regions with decreasing population, and some others with population increase, which makes the general conclusion that values in flood-prone areas increase impossible (this is valid for many areas, but not for all)."

A1: We thank the Editor for this comment, which gives us the opportunity for a broader discussion on this crucial issue. We agree with the Editor in identifying the demographic growth as the most important cause for the urbanization sprawl, so that in the regions where emigration prevails, less marked urbanization growth rates are observed. Thus, on a global scale, the urbanization growth rates feature a strong regional variability. However, a fundamental driver for urban sprawl must be considered in addition to population growth in developed countries, such as Italy. Indeed, the touristic demand for accommodations (hotels, holiday homes) leisure facilities and shopping centers leads to significant soil consumption rates, even in those regions (Alpine and Apennine mountain valley, major islands, and Southern Italy) where resident population continues to emigrate towards other Italian regions or foreign countries, since several decades. Examples of population decline associated with high urbanization growth rates have been presented in the manuscript for the Italian case (lines 33-44) and two supporting references have been added.

CENSIS: 53° Rapporto sulla situazione sociale del Paese, Franco Angeli, https://www.censis.it/rapporto-annuale/sintesi-del-53%C2%B0-rapporto-censis, 2019. (in Italian)

ISPRA: Soil consumption in Italy – 2019 Edition, Report SNPA, 08/19, ISBN 978-88-448-0964-5, 2019. (in Italian)

Q2: "Moreover, as recently shown for Austria and Switzerland (Fuchs et al., 2015; Fuchs et al., 2017; Röthlisberger et al., 2017), elements at risk exposed to flood hazards are not always/not necessarily of the same building type and therefore exposure is highly dynamic in space and over time. To give an example, for Austria, it has been shown that a relatively higher share of buildings exposed to flood hazards in the lowlands belongs to the category of commercial buildings, while in the upper catchments this share is higher for residential buildings and hotels/buildings of the tourism industry. Consequently, if we include "population dynamics" or "dynamics in exposure" into a somehow dynamic risk concept, this should be mentioned. Obviously this does not necessarily mean that I expect a reference to my own works, this is only for illustration – there are also other works centered on these issues available."

A2: We strongly agree with comment. In our paper we actually distinguished between industrial-commercial areas and residential areas, in order to decrease the uncertainty of exposed people estimates. A further classification refine was not found to be necessary, in view of the homogeneity of fabrics in these two classes. The need to incorporate spatial heterogeneity of urbanization in dynamic risk mapping has been further underlined in the introduction section (lines 107-115) and further details on land use classification have been added in section 3.1 (hazard mapping) (lines 326-327) The following references which well suit the discourse have been added as examples of studies which taken into account the urbanization heterogeneity with reference to flood risk estimation.

Fuchs, S., Keiler, M., and Zischg, A.: A spatiotemporal multi-hazard exposure assessment based on property data, Natural Hazards and Earth System Sciences, 15, 2127-2142, doi:10.5194/nhess-15-2127-2015, 2015.
Vu, T. T., and Ranzi R.: Flood risk assessment and coping capacity of floods in central Vietnam. Journal of Hydro-Environment Research, 14, 44-60. doi:10.1016/j.jher.2016.06.001, 2017.

Q3: "Moreover, some of the items described in section 2 (methods) have overlaps with a paper by Metulini and Carpita (2020). As this concerns mainly the methods section, I personally can accept it (and also according to the overall NHESS guidelines this is somehow ok) but I kindly would like to ask you to change wording a bit so it is not 1:1 the same text."

A3: Sentences in method sections (in particular section 2) have been rephrased by keeping the same content but by using other words. Moreover, the whole paper has been revised by a native English speaker.

Authors' replies to short comment SC1 by Monego M.

We thank you for your general evaluation of the work and for the question that you arose, which gives us the opportunity of a deeper discussion on this aspect.

Q1: "However, if the methodology proposed correctly identifies human positions and people exposure, it does not take into account human disaster response behavior (during floods daily activities can drastically change). Some modeling techniques, including agent-based models (ABMs), have been recently introduced to the in the field of flood risk assessment to simulate the dynamic distribution of the population during flooding, while still introducing inevitable simplifications of the human behavior patterns and disaster responses. For an integrated flood risk management in the future, it will be increasingly essential to consider the feedback between floods and people in a dynamic way and I suggest to give a comment on this issue."

A1: Indeed, exposed people behaviors and habits can significantly change after hydro-climatic alarms or during flood event occurrences, as well as their ability to decrease vulnerability by implementing flood proofing practices. However, these virtuous behaviors are usually the result of extensive campaigns to raise public awareness against flood risk, coupled with trusted and effective warning systems. In the analyzed area the risk perception towards the secondary network is almost absent, as well as a capillary local warning system. In addition, the knowledge and a widespread application of flood proofing practices do not exist (both structural and non-structural). Flood risk perception is mainly related to the primary hydrographic network (i.e. Mella River). Despite the dramatic increase in flooding episodes and in consequent economic damages, the impairment of the conveyance capacity of the hydrographic network and the urbanization sprawl still continue. Therefore, in the regards of the specific test case, in this research the possibility of drastic changes in human behavior during heavy rainfalls are not expected.

On the other hand, a dynamic approach to the flood risk is becoming mandatory, especially in consideration of the auspicated-future application of non-structural practices to the risk mitigation. Actually, agent-based modelling falls into the framework of a dynamic assessment of flood risk. The methodology herein proposed has potential to monitor people mobility dynamic during crises, evidencing modifications of their spatiotemporal distribution. Mobile phone network hardly fails during floods, thus observations of people dynamic under crisis conditions could be beneficial for a better calibration of any dynamic model. Finally, mobile phone data are richer than the ones used in this study, since vector data allows provider to follow users along their path. This type of information was not available for this study. However, there are potentials to further improve this technique in order to assess mobility preferential ways and to change them to increase escape security.

The need for a dynamic approach to flood risk assessment, along with the following references to ABMs, has been remarked in the introduction section, to better set this work inside the most updated research (lines 98-106). Then, a brief discussion on people behavior change during flood alarms has been added in the final part of the conclusion section.

Dawson, R. J., Peppe, R., & Wang, M. (2011). An agent-based model for risk-based flood incident management. Natural Hazards, 59(1), 167-189. doi:10.1007/s11069-011-9745-4

Haer, T., Botzen, W. J. W., & Aerts, J. C. J. H. (2016). The effectiveness of flood risk communication strategies and the influence of social networks-insights from an agent-based model. Environmental Science and Policy, 60, 44-52. doi:10.1016/j.envsci.2016.03.006.

Q2: "Minor comments: at lines 285-289 and 297-299 there are little mistakes"

A2: Repetition will be removed from section 3.1 (lines 285-289 of the first submission).

Authors' replies to referee comment RC1 by Anonymous Referee #1

We thank Referee #1 for her/his general evaluation of the paper and her/his supportive comments. All requested revisions has been implemented in the revised paper. Individual replies are listed below.

Q1: "Abstract: The last sentence mentions the application of the method in real-time rescues and reliefs. However, this has not been demonstrated and the application in real-time lacks of a real-time data access.

A1: Actually the proposed exploitation of mobile phone data to estimate people exposed to floods is completely novel and it has not found verification during real-world episodes of inundations, yet. This issue will be remarked in the final sentence of the abstract (lines 18-20), that will be rewritten in order to make it clearer as follows: "This novel methodology still deserves verification during real-world flood episodes, even though it appears to be more reliable than crowdsourcing strategies, and seems to have potentials to better address real-time rescues and relief supplies".

Q2: "Moreover, real-time hazard maps (dynamic flood maps) are not available yet. Please discuss this need for dynamic hazard and exposure data and its accessibility in realtime in the discussion or conclusion section."

A2: A general need for a dynamic approach to flood risk assessment has been underlined by many authors, for instance those who authored the publications reported in the references below. Reasons are manifold and are related to the dynamic behavior of decision makers, exposed people or rescuer actions, and to the dynamic nature of climate and urbanization. For instance, effective campaigns devoted to increase people awareness to flood risk or their capability to undertake water proofing practices, as well as, the implementation of waring systems can dramatically diminish the flood risk over time, without adopting structural strategies to decrease the flood hazard. Furthermore, accounting for urban development trends could be beneficial, in order to assess the future increase in flood risk. A discussion on these aspects has been added in the introduction (lines 98-106) and the reference list will be improved with these updated references.

Dawson, R. J., Peppe, R., & Wang, M. (2011). An agent-based model for risk-based flood incident management. Natural Hazards, 59(1), 167-189. doi:10.1007/s11069-011-9745-4

Di Baldassarre, G., Viglione, A., Carr, G., Kuil, L., Salinas, J. L., & Blöschl, G. (2013). Socio-hydrology: Conceptualising human-flood interactions. *Hydrology and Earth System Sciences, 17*(8), 3295-3303. doi:10.5194/hess-17-3295-2013

Haer, T., Botzen, W. J. W., & Aerts, J. C. J. H. (2016). The effectiveness of flood risk communication strategies and the influence of social networks-insights from an agent-based model. Environmental Science and Policy, 60, 44-52. doi:10.1016/j.envsci.2016.03.006.

Viglione, A., Di Baldassarre, G., Brandimarte, L., Kuil, L., Carr, G., Salinas, J. L., . . . Blöschl, G. (2014). Insights from socio-hydrology modelling on dealing with flood risk - roles of collective memory, risk-taking attitude and trust. *Journal of Hydrology, 518*(PA), 71-82. doi:10.1016/j.jhydrol.2014.01.018

Moreover, the approach proposed in this paper does not need real-time data, since the exposed people assessments can be carried out of emergency periods. Once functional box plots of exposed people are derived, as those reported in Figure 6 and in Figure 7, an assessment of people affected by floods can be expressed in terms of typical spatio-temporal patterns and their uncertainty, independently of the availability of real-time data. Thus the main aim of the paper is to pursue risk prevention by enhancing preparedness. The exploitation of mobile phone data in real-time will become more feasible after a widespread use of 5G and GPS technologies. A brief remark of this issue has been added to the final part of the conclusion section (lines 475-506). However, examples of almost real-time applications already exist, but in different contexts. For instance, Florence municipality uses mobile phone data to assess tourist crowding and citizen mobility (daily, weekly and monthly). In general, mobile phone data accessibility is getting easier, as mobile phone providers are more conscious of their market value (an example is given by API developed by TIM, the largest provider in Italy, which allows stakeholders to download data in almost-real-time).

Q3: “Line 105: Please explain the term “Erlang mobile phone measures” or give a reference to it.”

Raw data description at lines 125-131 has been improved as follows:

“The proposed geo-statistical approach relies on Erlang mobile phone measures. An Erlang is the unit of measure of traffic intensity in a telecommunication system or network and it is widely used to quantify load and efficiency. The name is a tribute to A. K. Erlang (1878-1929), a Danish mathematician and statistician who firstly worked on traffic engineering (Erlang, 1909). In this study, Erlang measures consist in two-dimensional matrices which provide the spatial distribution of the average number of mobile phone users (MPU) bearing a SIM connected to the network, within a temporal interval and inside a spatial region. These data are collected by mobile phone providers and recorded at constant time steps with reference to a georeferenced grid of square cells.”

Erlang. A.K (1909). The theory of probabilities and telephone conversations. Nyt Tidsskrift Matematika, B.20, 33-39.

Q4: “Figure 5: The figure is hard to understand. Figure 5a is referred in the figure caption as days a week but shows the months. Moreover, e.g., green color is refereed to days from July to September but the figure shows green also in October. The same is for blue and yellow. Please revise the figure accordingly or add an additional explanation.”

A3: The caption of Figure 5 has been corrected and improved. References to panels in Figure 5 were inverted in the first version of the caption. The cluster names refer to the period that they mainly, but not exclusively, belong to. Thus, the green cluster indicates days mainly belonging to the period July-September, even though a few days of this cluster can be observed in October. Additional explanations of the spine plots have been added to section 4.1 (lines 384-388) and in the figure caption to precise this aspect.

Q5: “Lines 1-61 can be shortened remarkably.”

Indeed, the first part of the introduction (lines 20-60) reports preliminaries which introduce known concepts. In our opinion this part is beneficial to place the work in a comprehensive-historical framework. On the other hand, considering the additional discussions required by reviews and comments, a better balance could be achieved by giving more room to the specific issues and research advances faced in the paper. Thus, we simplified this part by removing very well-known concepts and ancillary discussions and by shortening some paragraphs. More precisely: lines 21-23: shortened; Lines 23-26: deleted; Lines 33-35: shortened; Lines 39-45: shortened; Lines 48-57: shortened; Line 58: deleted (references were deleted consistently). In the revised version, the preliminary part of the introduction involves lines 22-59, including the additional discussion flowing Editor’s comments. This should be beneficial to shift the discussion focus towards the main aim of the paper and to better set the work in the most recent literature.

Authors’ replies to referee comment RC2 by Anonymous Referee #2

We thank Referee #2 for her/his thorough reading and general evaluation of the paper and her/his suggestions for future developments. All requested revisions has been implemented in the revised paper. Replies are listed below.

Q1: “In section 2.1, the Authors refer to a specific parameter: “k is a parameter that need to be chosen”. One reference for the choice of k should be reported, especially for non-statistician readers.”

A1: As we have said in lines 337-338, the criterion used in our application aims to maximize k subject to avoiding the presence of zeros in the vector of HOG features (i.e. avoiding empty bins). According to Salhi et al. (2013), the larger the number, the more accurate the results are. Moreover, according to them, parameter k usually ranges from 4 to 20 in the related literature. The text has been revised (lines 174-189).

We will add within the manuscript some details on it and the following reference: Salhi, A. I., Kardouchi, M., & Belacel, N. (2013). Histograms of fuzzy oriented gradients for face recognition. In 2013 International conference on computer applications technology (ICCAT) (pp. 1–5). IEEE

Q2: “Moreover, the motivation for the choice of Bouveyron and Come (2015)’s procedure, among all possible functional data techniques, should be (briefly) addressed.”

A1: We have choosen Bouveyron and Come (2015)’s model-based functional data clustering method because its better flexibility compared to alternative methods. As already reported in lines 167—171 of the original version of manuscript, we need a method in which to each cluster it corresponds an estimated functional curve with specific parameters. In fact, our aim is to consider the similarities in the functional form of the daily density profiles (DDPs), viewed as a curve of values (y-axis) with respect to time instants (x-axis). Adopting a model-based functional data clustering method, each group’s curves are modelled by their own set of distributional parameters.

Moreover, since we have high dimensional dataset (more variables, or features, than observations), as we wrote in lines 171—173, the chosen method is suitable for these special kinds of dataset, because it applies sub-space clustering (Agrawal et al., 1998) which is generally adopted to consider just the minimum number of variables needed for grouping objects, thus reducing the dimensionality.

In lights of the referee comment, we will rephrase the related part of the manuscript in order to be more clear and more exhaustive (lines 197-201).

Q3: “Concerning the Carpita and Metulini (2020)’s statistical matching approach, are there any kinds of test, procedure, etc., to evaluate representativeness and reliability of the final result of the population assessment step? This aspect should be (briefly) addressed.”

A3: There are no specific tests proposed in the literature suitable to our case. Moreover, a comparison test using official data is not possible since the lack of official data. However, the proposed procedure it is the result of a series of tests. In particular, we think that a key aspect related to the reliability of the population assessment step is the choice of the measure of central tendency used in the denominator of equation 5 (line 221). The choice of preferring the median instead of the mean is motivated by the strongly asymmetrical distribution of Estimated TIM Market Share (ETMS), as shown in Metulini, Carpita (2019b). The choice of the median has been then tested by Carpita (2019) for the case of the Lake Iseo during the Floating Piers (there official data are available). Results show that the estimated number of people is similar to that provided by official sources. Revised text at lines 338-343.

Q3: “Further developments: Due to the richness of mobile information and the heterogenous moving behaviour of individuals during the day, several further developments can be considered for future works. For example, it would be very interesting (from a prevention perspective) to restrict the sample of investigation and focus on the intraday mapping of individuals in meaningful time periods of the year, e.g. the months with highest probability of observing floods. Also the idea of considering the movement response of residents to floods may be strongly interesting.

A3: The project has recently been refinanced by Lombardy Region (the local authority of the administrative region where the study catchment is placed) and a new partner is TIM (the most important mobile phone provider in Italy). This allows us to have access to the full set of data that providers routinely collect. Datasets include a vector data reporting the individual user location along with their SIM identification number. On the one hand, it would be possible to track users down. As a perspective the ongoing research, just started, includes the collection of matrices of OD-Origin-Destination vectors in different seasons, days of the week and hour of the day. By knowing the density of vectors with origin and destination of the paths around critical traffic nodes it will be possible, more precisely, to forecast potential critical conditions for mobility and better manage traffic. Also coupling traffic management decision support systems with real-time rainfall-runoff-flooding modelling is a research perspective being considered. From a prevention perspective, this could make it possible the identification of preferential traffic flows evidencing potential risks during inundation onsets or emergency situations. Alternative safe pathways could be identified and enforced to exposed people, in order to facilitate their evacuation. On the other hand, it would be possible to profile the SIM users, even though keeping anonymousness and respecting their privacy. Thus, users could be categorized (with respect to age, gender, etc.) in order to isolate specific targets from the whole user set. Thus their behaviors could be statistically analyzed separately from the others. Unfortunately, such additional data were not available for this study, but in the forthcoming development these research advances could be addressed.

Another issue regards people behaviors during flood emergency or warnings, which could significantly change with respect to those of ordinary days. The geostatistic analysis proposed in our work is able to detect possible differences in the exposed people spatial distribution, that are statistically significant. A further objective of the research extension is the implementation of a warning system in the study area, coupled with a campaign to make people aware of flood risk associated with the analyzed stream network. Mobile phone data will be useful to evidence the actual response of expose people to these non-structural practices and to estimate the expected decrease in the flood risk. To have a statically robust assessment, a larger set of data is however needed, since emergency or warning occurrences are rare (a few days in a year). The extension of the project will increase the dataset size and will make this research advances easier. As dataset size increase, it would be possible to restrict the analysis to specific periods (maybe the rainy seasons)

A discussion on further developments has been added to the final part of the conclusion section (lines 475-506).

Q4: Similarly, possible insights may be evaluated for the statistical matching procedure in future works. For example, what happens sharing population in different classes? Assuming heterogeneity in the behaviour of individuals, can you include in the procedure the propensities of different classes of residents to the use of smartphones during the day?

A future development of the statistical matching procedure between mobile phone data and census data could use demographic and socio-economic information about the SC (sezioni di censimento) areas, for example the ISTAT ARCH.I.M.E.DE database (www.istat.it/it/archivio/190365). Since, it is likely to assume heterogeneous behaviors of individuals, we may think to use ARCH.I.M.E.DE. database in future works to share individuals in classes in terms of their age, gender, income or their job. In fact, different mobile phone companies have different costs, and this may affect differently the choice of different classes of individuals. This discussion has been added in the final part of the conclusion section (lines 475-506).

Dynamic maps of people exposure to floods based on mobile phone data

Matteo Balistrocchi[1], Rodolfo Metulini[2], Maurizio Carpita[3], Roberto ~~Ranzi[1]~~ Ranzi[4]

[1]Department of Engineering “Enzo Ferrari”, University of Modena and Reggio Emilia, Modena (MO), 41125, Italy

~~[1]Department of Civil, Environmental, Architectural Engineering and Mathematics, University of Brescia, Brescia (BS), 25123, Italy~~
[2]Department of Economics and Statistics, University of Salerno, Fisciano (SA), 84084, Italy
[3]Department of Economics and Management, University of Brescia, Brescia (BS), 25122, Italy
~~[1]~~[4]Department of Civil, Environmental, Architectural Engineering and Mathematics, University of Brescia, Brescia (BS), 25123, Italy

Correspondence to: Matteo Balistrocchi (matteo.balistrocchi@unimore.it)

ha formattato: Italiano (Italia)

ha formattato: Inglese (Regno Unito)

ha formattato: Italiano (Italia)

Abstract. Floods are acknowledged as one of the most serious threats to ~~human~~ people’s lives and properties worldwide. To mitigate the flood risk, it is possible to act separately on its components: hazard, vulnerability, exposure. Emergency management plans can actually provide effective non-structural practices to decrease both people exposure and vulnerability. Crowding maps depending on characteristic time patterns, herein referred to as dynamic exposure maps, ~~provide~~ represent a valuable tool to enhance the flood risk management plans. In this paper, the suitability of mobile phone data to derive crowding maps is discussed. A test case is provided by a strongly urbanized area subject to frequent floodings located ~~in~~ on the western outskirts of Brescia ~~town~~ (northern Italy). Characteristic exposure spatio-temporal patterns and their uncertainties were detected, with regard to land cover and calendar period. This novel methodology still deserves verification during real-world flood episodes, even though it appears to be more reliable than crowdsourcing strategies, and seems to have potentials to better address real-time rescues and relief~~s~~ supplies~~y~~~~This novel methodology appears to be more reliable than crowdsourcing strategies, and has potentials to better address real-time rescues and reliefs supply~~.

1 Introduction

Floods are natural phenomena whose hazards afflict nearly 20 million people worldwide (Kellens et al., 2013), posing a serious challenge to the protection of ~~human~~ people’s lives and the liveability of urban settlements. ~~Both low-income and high-income countries are strongly impacted by extreme weather events and a~~ A high-confidence increase trend in the ~~resulting~~ economic damages and social costs due to extreme weather events has globally been documented (Kreibich et al., 2019). ~~As reported by Munich RE (2020) over the period 1980-2019, flooding accounts for some 40% of all loss-related natural catastrophes, with~~

~~losses worldwide totalling more than US$ 1tn. For instance, floods and landslides in Thailand in 2011 resulted in 43 US$ bn, that is the highest flood losses of all time.~~

Two major factors can be advocated for justifying such a trend: climate change and increased urbanization ~~and people exposure~~ (Hartmann et al., 2013). These factors involve different components of the risk concept (UN ISDR, 2009), which is given by the combination of hazard, exposure and vulnerability. Climate change was popularly acknowledged as a leading cause for the increases in the frequency and intensity of heavy storms and, consequently, of the flood ~~episodes~~ hazard (Solomon et al., 2007). ~~This factor is therefore related to the hazard component of the flood risk.~~ However, according to the Intergovernmental Panel on Climate Change (IPCC) ~~IPCC~~ (Hartmann et al., 2013), and as confirmed also by up-to-date analyses of flood intensity in Europe (Blöschl et al., 2019), the absence of a global likely trend in the incidence of floods arises. R~~This can be due to various r~~easons lie in~~:~~ the high regional variability of heavy storm trends~~, in terms of type, magnitude and significance~~, as well as in the strong influence played by ~~the~~ watershed hydrologic characteristics and ~~the~~ local flood management practices ~~on the flood generation processes~~.

On the contrary, population urbanization represents a likely global trend, though characterized by a strong regional variability. M~~People m~~igration from countryside or mountain areas to cities is the main driver of urban sprawl. In 2008, for the first time in human history, more than half the world population was living in urban settlements and the percentage continues to augment (UN DESA Population Division, 2012). Touristic demand is an additional driver for urbanization growth, that plays a peculiar role in developed countries. For instance, in Italy many areas are affected by emigration, namely Alpine and Apennine valleys, southern regions and islands, and urbanization growth rates are equal to or greater than the national average rate. A clear example is provided by Southern Italy, where the annual rate of soil consumption between 2017 and 2018 was 0.23 %, ~~larger~~greater than the national average ~~national one~~ of 0.21 % (ISPRA, 2019), ~~while~~even if Southern Italy faced a population decrease of 1.5 % between 2015 and 2019 (the national average loss was estimated at 0.7 %) (CENSIS, 2019). An additional example is provided by the Sondrio province (in the mountainous part of the Adda River basin, Lombardy, Northern Italy), where the yearly soil consumption per capita is 1.11 m^2, whereas the regional one is 0.63 m^2.

~~On the contrary, the population urbanization represents a non-climaticlikely global trend, though characterized by a strong regional variability.. People migration from countryside or mountain areas to cities is the main driver of urban sprawl. In 2008, for the first time in human history, more than half the world population was living in urban settlements and the percentage continues to augment (UN DESA Population Division, 2012). Population urbanization determines dramatic increases in people exposure and vulnerability to floods, since most of recent urbanizations are developed disregarding the natural extension of floodplains. Therefore, urbanizations often lie in flood prone areas and local communities are not able to put in place effective flood defence practices.Touristic demand is an additional driver for urbanization growth in developed countries. For instance, in Italy many areas affected by people emigration, namely Alpine and Apennine valleys, southern regions and islands, urbanization growth rates are equal to or greater than the national average rate. A clear example is provided by Apulia Region (Southern Italy), where the soil consumption was about 8.43% in 2017 and 2018 (larger than the national one, equal to 7.63%) (ISPRA, 2019), while Southern Italy faced a population decrease of 1.5% between 2015 and 2019 (CENSIS, 2019).~~

U~~Population u~~rbanization determines dramatic increases in people exposure and vulnerability to floods in immigration areas, since most of recent urbanizations lie in flood prone areas and local communities are not able to put ~~in place~~ effective flood defence practices in place. Moreover, urbanization usually leads to the impairment of the conveyance capacity of the stream network, so that flooding areas are basically larger than in the undeveloped condition. T~~hus, t~~he urbanization sprawl consequently results in increased damage to communities, private properties and public infrastructures, and~~, which is defined as the product of exposure and vulnerability for a given hazard. Indeed, this second factor~~ must be regarded as the main cause for the likely increasing trend of ~~the~~ flood risk (Barredo et al., 2009). A number of researches on flood risk changes under economic and population growth scenarios indicate that this contribution is at least equal to, but commonly larger than~~,~~ the climate change one (Feyen et al., 2009; Maaskant et al., 2009; Bouwer et al., 2010; Te Linde et al., 2011; Rojas et al., 2013).

This flood risk trend forced an unavoidable shift in the paradigms of flood defence, by recognizing that not all events can be completely controlled and that structural practices have limits (Johnson and Priest, 2008).~~ For instance, the European Flood Risk Management Directive (European Union, 2007) acknowledges that floods cannot be stopped from occurring, and that~~Thus,~~ the~~ focus must be placed on how to mitigate the damages to flood prone communities (European Union, 2007). O~~Hence, o~~ver the last decades, flood risk management has evolved from a structural-based defence approach, aiming at decreasing the hazard component, towards a more holistic perspective (Merz et al., 2010; Arrighi et al., 2019) ~~,~~ taking into consideration ~~drivers and impacts of flood risk. V~~vulnerability and exposure~~ were thus investigated in a deeper manner than in the past.~~, ~~whilst n~~Novel concepts were introduced, such as residual risk (UN ISDR, 2009), accounting for the potential structural failure of the defence system (Vorogushin et al., 2009; Schumann, 2017; Balistrocchi et al., 2019), and resilience, that is the ability to recover from a damage or to absorb an impact (Liao, 2012).

~~Actually, vulnerability reduction plays an essential role for successful adaptation to flood risk (Kreibich et al., 2017).~~ Differently from the hazard, the mitigation of exposure and vulnerability ~~the expected damage's components~~ can be pursued by means of non-structural practices. Among them, a prime role is played by emergency management plans, which allow authorities responsible for the protection of local communities to dispatch timely and appropriate mitigation measures during the occurrence of flood episodes. The development of effective emergency management plans is closely related to the concept of authorities' preparedness (European Union, 2007). Such plans are actually intended to provide people with early warnings, reliable real-time information and ~~to~~ the ~~better address~~improvement of relief supplies and rescue efforts. In this regard, a detailed and reliable picture of the real-time spatiotemporal variability of the flood risk would be highly beneficial.

Presently, large amounts of geospatial data can be obtained from a number of sources, namely remote sensing or aircraft platforms. T~~However, t~~hese sources yield situation snapshots, but do not provide information at the spatiotemporal resolution needed for managing urban floodings and are hardly validated ~~in~~ on the field. To overcome these problems, the possibility of taking advantage of crowdsourcing techniques has recently attracted much attention (Mazumdar et al., 2017; Rosser et al., 2017; Hirata et al., 2018; Mazzoleni et al., 2018). These techniques have been made available by the widespread proliferation of smartphones and tablets, along with the success of social media. During emergency phases, the advantages in the real-time implementation of the emergency plan are twofold: ~~on the one hand~~firstly, a large amount of volunteered geographical information suitably georeferenced can be collected (Goodchild, 2007), guiding authorities in developing collaborative flooding maps or in estimating the number and location of exposed people; ~~on the other hand~~secondly, crucial information can be communicated to the ~~exposed~~ people exposed, making them aware of the actual risk magnitude and ~~supporting~~ enhancing their capability to face the situation. After the emergency phase, authorities can also exploit the collected data to enhance their preparedness and to better match the emergency management plans to specific real-world needs of the flood prone community.

Potentials and drawbacks of various crowdsourcing techniques have long been ~~long~~ debated, even though crowdsourcing has already found successful applications in some weather related disasters (Poser and Dransch, 2010; Hung et al., 2016; Guntha et al., 2018). Such researches have underlined that in many cases, during the emergency phase, crowdsourced data have at least the same quality ~~of~~ as the authoritative ones (Goodchild et al., 2017). Nevertheless, several concerns have been pointed out through analyses of crowdsourcing technique applications to real-world disasters: i) the raw data quality is generally poor because of malicious intentions of ~~rough~~ nasty elements ~~of~~ in the community or incompetence of stakeholders, so that spurious, erroneous, malformed, redundant or incomplete data must be purged out of the database to make them suitable, ii) the sample significance is basically low, owing to the limited number of exposed people actively participating in crowdsourcing, so that information of general interest must be extrapolated from an exiguous fraction of the whole population, iii) the communication network is not completely reliable, as it ~~frequently~~ could fail~~s~~ or malfunction~~s~~ during disaster occurrences.

Researches have therefore been addressed to filter such rumours from crowdsourced data (Han and Ciravegna, 2019). However, other approaches can be followed to develop effective dynamic information tools through the exploitation of mobile phone data collected by providers. Such data make it possible to geo-localize mobile phone users over ~~the~~ time, ~~that is~~in order to derive time-dependent crowding maps. When such maps are intersected with hazard maps, showing the flooding area extensions corresponding to a selected frequency of occurrence, dynamic exposure maps are obtained. As demonstrated by Carpita and Simonetto (2014) with reference to episodic crowd concentrations due to social events, recurrent spatiotemporal patterns can be derived from mobile phone data by means of ~~geostatistic~~geostatistical analyses. Herein, the application of this methodology to the periodic spatiotemporal variability of the resident population related to home-work mobility is investigated. To do so, additional tools are developed to extrapolate the real-world population from the crowding maps of provider's clients. Dynamic exposure maps can be set in a more general call for a dynamic approach to flood risk assessment and management (Viglione et al., 2014). Actually, flood risk varies over time not only with regard to climate non-stationarities and urban development trends, since hydrological, economical, political, technological and social processes are also involved. For instance, effective campaigns devoted to increase people awareness towards flood risk, or to promote their capability to undertake effective water proofing practices, or to exploit warning systems, can dramatically diminish the flood risk over time. The same occurs by keeping the memory of flood disasters. These processes are inter-related and evolve over time (Di Baldassare et al., 2013). In this regard, some Authors proposed dynamic agent-based models to assess the temporal change in flood risk (Dawson et al. 2011; Haer et al., 2016). Such models are capable to perform a spatially-distributed analysis of flood risk, accounting for multiple factors, or agents, and their action-feedback relationships.

It must be pointed out that dynamic risk maps should account for the spatial heterogeneity of urbanization, in order to obtain precise assessments. Flood prone urbanizations can feature very different characteristics relevant to both exposure and vulnerability, even inside the same watershed. Land use is the first discrimination level to be considered, as people densities and temporal patterns could significantly differ in commercial, industrial, service, transport and residential areas. Secondly, the fabric type may also have remarkable impacts on the overall risk. Indeed, Fuchs et al. (2015) evidenced significantly different exposures among various types of land use (tourist accommodation, commercial, recreational, residential) in Austria. Urbanization heterogeneity is also relevant for flood risk studies in developing countries as shown by Vu and Ranzi (2017); in their assessment of flood risk in central Vietnam, they estimated the exposure and vulnerability of building and people by collecting questionnaires including data on building types.

~~It must be pointed out that dynamic risk maps should account for the spatial heterogeneity of urbanization, to obtain precise assessments. Flood prone urbanizations can feature very different characteristics relevant to both exposure and vulnerability, even inside the same watershed. Land use is the first discrimination level to be considered, as people densities and temporal patterns could significantly differ in commercial, industrial, service, transport and residential areas. Secondly, the fabric type may also have relevant impacts on the overall risk. Indeed, Fuchs et al. (2015) evidenced significantly different exposures among various types of land use (tourist accommodation, commercial, recreational, residential) in Austria. Urbanization heterogeneity is also relevant in developing countries as shown by Vu and Ranzi (2017); in their assessment of flood risk in Vietnam, they estimated the exposure and vulnerability of building and people by collecting questionnaires including data on building types.~~

T~~In order t~~o demonstrate the potentials of the ~~geostatistic~~geostatistical analysis herein proposed, it is applied to ~~A~~a suitable case study, ~~has been~~ identified in the western outskirts of Brescia ~~town~~ (Lombardy, northern Italy).~~, for which~~ ~~a~~ A detailed knowledge of the flooding dynamics and a sizeable set of mobile phone data are available for this watershed. The suggested approach made it possible to derive reliable dynamic exposure maps with respect to the land coverage and the calendar time periods, obtaining estimates of the expected number of people affected by flood hazards along with its uncertainty.

Hence, the paper is organized according to the following sections: (*i*) firstly, the innovative aspects of the ~~geostatistic~~geostatistical analysis methodology herein utilized are illustrated, (*ii*) secondly, the main hydraulic-hydrologic features of the analysed study area are described along with the available mobile phone data, (*iii*) the methodology application and the results are finally discussed.

ha formattato: Tipo di carattere: (Predefinito) Times New Roman, Nessuna sottolineatura

2 Analysis methodology

The proposed geo~~-~~statistical approach relies on Erlang mobile phone measures. An Erlang is the unit of measure of traffic intensity in a telecommunication system or network and it is widely used ~~for~~to ~~measuring~~quantify load and efficiency. The name is a tribute to A. K. Erlang (1878-1929), a Danish mathematician and statistician who firstly worked on traffic engineering (Erlang, 1909). In this study, Erlang measures consist in two-dimensional matrices which provide~~s~~ the spatial distribution of the average number of mobile phone users (MPU) bearing a SIM connected to the network, within a temporal interval and inside a spatial region. These data are collected by mobile phone providers and recorded at constant time steps with reference to a georeferenced grid of square cells.~~The proposed geo-statistical approach relies on Erlang mobile phone measures, which consist in the average number of mobile phone users (MPU) bearing a connected SIM. These data are collected by mobile phone providers and recorded at constant time steps with reference to a georeferenced grid of square cells.~~ The availability of such a kind of data is progressively capturing the attention~~ly arousing the enthusiasm~~ of ~~the~~ urban planners~~'~~ ~~community~~ (Becker et al. 2011; Calabrese et al., 2015), as they offer a variety of potential applications. In this study, the MPU spatiotemporal variability was summarized by means of daily density profiles (*DDP*), that provide~~s~~ the variability within a day of the MPU referred to a spatial region of interest. Such regions are inundation areas, thus expressing the spatiotemporal variability of people exposed to the flood risk. To define *DDP*, let e_{it} be the number of MPU in the *i*-th grid cell in a generic time interval *t*. Let $I_r = \{i_1, \dots, i_m\}$ be the set of grid cells in~~related to~~ region *r* of interest. Furthermore, let define $T_d = \{t_1, \dots, t_o\}$ be the set of ~~time~~ intervals of time in a day *d*. The daily density profile (DDP_{rd}) can be defined according to Eq. (1), as a vector of the sums of MPU (a sum for each considered time instant) in region *r* and day *d* (length *o*~~)~~) ~~of values describing the sum of MPU in region *r* and day *d*:~~

$$DDP_{rd} = \{\sum_{l=1}^{m} e_{il,t1}, \sum_{l=1}^{m} e_{il,t2}, \dots, \sum_{l=1}^{m} e_{il,to}\}'. \quad (1)$$

Herein, the interest lies in ~~analyzing~~analysing and classifying the occurrences in a time series of DDP_{rd}, related to a set $d = \{d_1, \dots, d_n\}$ of *n* ~~analyzed~~analysed days. More precisely, the proposed approach firstly involves the clustering of similar DDP_{rd}, as discussed in detail in the following Section 2.1. The clustering procedure consists of two steps. In the first one, MPU spatial variability inside region *r* is considered by changing ~~the~~ index *i* in a R^2 *x-y* coordinate space; to do so a data reduction strategy is applied. In the second one, the DDP_{rd} temporal variability is evaluated by changing index *t* in a R^1 space.

The characteristics~~Clustering~~ of our ~~mobile phone~~ data raise some issues related to the choice of the clustering technique to be chosen. ~~according to the above described procedure is not straightforward.~~ In fact, t~~Nevertheless, t~~raditional techniques (Arabie and De Soete, 1996) may not produce robust results when the number of variables ~~are~~is larger than the number of observations. ~~cannot be applied. In fact,~~ Our d~~d~~ata amount to *n* observations and $p = m*o$ variables (number of MPU ~~informations~~ per~~values each~~ day). For instance, let us consider a case in which ~~if one has one year of~~ available data ~~refered~~refer to one year (i.e. $n = 365$): information~~s~~ in each cell of the grid ~~are~~is available 4 times per hour~~, repeated every 15 minutes~~ (~~so~~thus, $o = 96$) and the~~in a~~ region is covered by 500 grid cells ($m = 500$).~~,~~ It follows that t~~t~~he ~~variable~~ number of variables is ~~way~~far much larger than the number of observations ($p > n$) and so we refer to a~~n, depicting a~~ high-dimensional data setup ~~context~~ (Donoho, 2000). In ~~When analyzing~~ high-dimensionality~~data,~~ some~~several~~ issues need to be considered, such as those of the curse of dimensionality (Keogh and Mueen, 2017)~~, need to be faced~~. ~~With specific regard to data clustering, this issue has been addressed by~~ Bouveyron et al. (2007) addressed this issue with regard to clustering. However, as suggested by Jovi et al. (2015), a suitable solution is represented by ~~high-dimensional~~ data reduction ~~provides a suitable solution~~. To do so, ~~an~~ the

~~approach based on the~~ Histogram of Oriented Gradients (HOG) approach is used in this paper. Therefore, data reduction works~~acts~~ on index *i*, in order to convert the support from R^2 (*x-y* coordinate space) to R^1.

Once the DDP_{rd} are clustered in statistically similar groups,~~ the~~ the total number of people in set T_d and~~ in~~ region *r* can be estimated and associated with descriptive bands (DB), as discussed in Section 2.2. In this regard, there is~~a~~ crucial concern ~~is given by~~due the lack of MPU data from all companies providing phone services in northern Italy. To deal with this problem, as firstly suggested by Metulini and Carpita (2020), the approach proposed in this paper adopts a strategy to infer the total number of people by matching census data to available mobile phone data~~ to census data~~.

2.1 Data reduction and clustering

To cluster similar *DDP* a technique for high-dimensional data reduction is ~~firstly~~ adopted first. Then, reduced data are ~~analyzed~~analysed by using a high-dimensional data clustering. Separately for each element of set T_d (i.e. for a given *t*), let $\varepsilon_{it} = \{e_{1,t}, e_{2,t}, \dots, e_{im,t}\}'$ be the~~ vector of dimension m of~~ MPU vector of~~in~~ region *r* in time instant *t* (dimension *m*). The aim is to reduce ε_{it} to the vector of ~~a new set of~~ values κ_{it} (~~, of~~ dimension $m' < m$), which contains~~that preserves~~ the relevant information contained in ~~the~~ set ε_{it}. To do so, set ε_{it}, separately for each *t*, undergoes a histogram of oriented gradients (HOG) feature extraction (Dalal and Triggs, 2005; Tomasi, 2012). Traditionally, HOG is~~ generally~~ applied to red-green-blue (RGB)~~, or grey scale,~~ images, in order to find similarities among images and to classify them. Vector ε_{it} can be considered a raster that expresses the magnitude of its elements through ~~colors~~colours, s~~S~~ince each value has~~is~~ associated~~ with~~ a location in the R^2 *x-y* coordinate space.~~, vector ε_{it} can be viewed as a raster featuring colors expressing the magnitude of its elements.~~ In consideration of our setting, t~~T~~he application of the HOG method looks~~, having defined our setting as above, is~~ straightforward. According to first step clustering~~When clustering in terms of the spatial distribution of users~~, the interest lies in the relative distribution of the MPU~~MPU~~ in region *r*, not in the~~e MPU~~ absolute amount of users. To be consistent with this aim, the HOG was applied to a normalized vector of ε_{it}. Vector z_{it} is thus defined as $z_{it} = \{e_{i,t} / max_{i \in I_r}(e_{i,t}), \forall\ i \in I_r\}$~~.~~. In order to obtain~~generate~~ the vector of features κ_{it}, by displaying~~containing~~ the relevant information in z_{it}, the HOG method firstly divides the~~ grid's~~ cells of the grid in~~to a number of~~ *S* smaller grids $G_1, \dots, G_S$ ($G_i \cap G_j = \emptyset, \forall\ i = 1,\dots,S$ *and* $\forall\ j = 1,\dots,S$ *with* $i \neq j$), where $\sqrt{S}$ is a parameter that needs to be chosen. Then, ~~The *direction* and the *magnitude* matrices~~ by using two different *gradient* matrices of each grid G_i (see for details in Dalal and Triggs, 2005) the *direction* and the *magnitude* matrices are~~ then~~ obtained. These matrices are used to derive~~derive~~ the histogram of- gradients with *k* -equal~~equal~~ bins, ~~,~~ where~~where~~ *k* -is a parameter that ~~is a parameter that~~ needs to be chosen. The ~~L~~larger ~~is~~ the parameter *k* is, the better ~~are~~ the results are; moreover, in related literature *k* usually ranges from 4 to 20 (Salhi et al., 2013).~~These matrices are used to derive the histogram of gradients with *k* equal bins, where *k* is a parameter that need to be chosen~~. The vector of features κ_{it} is ~~the vector of features~~ given by all the elements of the derived~~obtained~~ histogram of oriented gradients, ~~s when~~ stacked ∀ *S*. Considering that ~~The~~ length o~~f~~f ~~vector~~ κ_{it} is $S*k$,~~. Subsequently,~~ the vector κ_{it} is stacked over the subscript *t*, in order to derive (for region *r* and day *d*)~~obtaining~~ the vector of features κ_d ~~for region *r* and day *d*, (of~~ dimension $S*k*o$).

ha formattato: Tipo di carattere: Non Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Non Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo

ha formattato: Tipo di carattere: Corsivo

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo

ha formattato: Tipo di carattere: Non Corsivo

κ_d is used ~~contains the features~~ s ~~(variables) undergoing the~~ in the first clustering step of the method, where days represents~~in which~~ the objects to be clustered, in terms of~~, according to~~ how the MPU are distributed~~distribute~~ over region *r* according to index *i*~~, are represented by days. For all days in the data set~~ ~~$d = \{d_1, \ldots, d_n\}$,~~ κ_d is computed for all the days in ~~the~~ data set $d = \{d_1, \ldots, d_n\}$, and a *k*-mean cluster analysis (Hartigan and Wong, 1979), in which the objects to be clustered are the *n* days and κ_d contains the values of the *S*k*o* (with *S*k*o* < *m*o*) variables for day *d* to be attributed to a cluster~~, is used~~. ~~is performed, where the *n* days are the objects to be clustered.~~ ~~κ_d contains the values of the *S*k*o* (with *S*k*o* < *m*o*) variables for day *d* to be attributed to a cluster.~~ According to the Hartigan and Wong criterion, the ~~number of~~ clusters' number *H* is chosen by looking ~~to~~~~analyzing~~at the ~~decreasing trend of the~~ ratio between the total within sum of squares ($Tot\ within\ SS_H$) and the total sum of squares ($Tot\ SS_H$) for different values of *H*~~,~~ that needs~~s~~ to be minimized with respect to ~~the number of groups~~ *H*. For a certain *H*, the total within sum of squares is defined as $Tot\ within\ SS_H = \sum_{i=1}^{H} W_{SS}(C_i) = \sum_{i=1}^{H} \sum_{\kappa_d \in C_i} (\kappa_d - \mu_i)^2$, where μ_i is the centroid vector (length *S*k*o*) for cluster *i*; the total sum of squares is defined as $Tot\ SS_H = \sum_d (\kappa_d - \mu)^2$, where μ is the centroid vector for the full set of data.~~.~~ At this point the elements in $d = \{d_1, \ldots, d_n\}$ (the days) have been assigned to a number of clusters $C_1, \ldots, C_H$ ($C_i \cap C_j = \emptyset$, $\forall\ i = 1,\ldots,H$ and $\forall\ j = 1,\ldots,H$ with $i \neq j$).

In the second step, ~~that in which~~ ~~To account for~~when the MPU variability over time is accounted for, we consider the~~,~~ ~~for a given region *r*, as objects, the~~ vector DDP_{rd}~~,~~ (for a given region *r*)~~considered~~ as the collection of functional observations x_{rd} (T_d), $T_d \subseteq (t_1, \ldots, t_o)$ of length *o*, with *d* varying in $d = \{d_1, \ldots, d_n\}$ (i.e. $\sum_{l=1}^{m} e_{il,t1}$ in t_1). To do so, we adopt a model-based functional data clustering method (Becker et al., 2011) since it is more flexible than the alternatives: to each cluster it provides an estimated functional curve with specific parameters. We group~~, separately for each cluster of the previous step,~~ days *d* (cluster's objects) in terms of the *o* DDP_{rd} values (cluster's variables)~~,~~. separately for each cluster of the previous step. ~~We The aim atis to considering the s~~Similarities in the functional form of the DDP_{rd}~~,~~ are considered, if viewed in terms of~~as a~~ curves of values ~~of values~~ (*y*-axis) with respect to time instants (*x*-axis). ~~In doing so,~~ The curves of each group~~each group's curves~~ are modelled by using a~~by~~ group-specific~~their own~~ set of distributional parameters (see Becker et al., 2011 for details). The adopted method ~~we adopt~~ can be used~~t is suitable for~~ in high-dimensionality~~-data~~, as~~since~~ the clustering process employs~~applies~~ the criterion~~a~~ of the sub-space clustering (Agrawal et al., 1998) which is adopted~~, adopted~~ when one is only interested~~to~~ in considering ~~just~~ the minimum number of variables ~~needed~~ needed for grouping objects, thus to reduce the dimensionality~~To do so, a model-based functional data clustering method (Bouveyron and Come, 2015) was adopted, and days *d* (cluster's objects) were grouped in terms of the *o* DDPrd values (cluster's variables), separately for each cluster defined in the previous step. The aim is to consider the similarities in the functional form of the DDPrd, seen as a curve of values (*y*-axis) with respect the *x*-axis (time instants). In doing so, each group's curves are modelled by their own set of distributional parameters (see for details Bouveyron and Come, 2015). The method adopted in this work is suitable for high-dimensional dataset, since the clustering process applies the criteria of the sub-space clustering (Agrawal et al., 1998), adopted to consider just the minimum number of variables needed for grouping objects, thus reducing the dimensionality~~. In detail, it is herein proposed to adopt the following path: i)~~-~~ functional data outlier detection by likelihood ratio test (LRT) is adopted to remove

ha formattato: Tipo di carattere: Corsivo

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Non Corsivo

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito), Pedice

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito), Pedice

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Inglese (Regno Unito)

ha formattato: Tipo di carattere: Non Corsivo

the anomalous DDP_{rd} ~~are removed using functional data outlier detection by likelihood ratio test (LRT)~~, as proposed by Febrero-Bande et al. (2008); ii) the ~~clustering method developed by~~ Bouveyron et al. (2015) clustering method is applied, using~~along with~~ *funFEM* package in *R*.

The aim of this strategy is to~~With this strategy we aim at~~ assign the elements in $d = \{d_1, \ldots, d_n\}$ (the days) to ~~a number of~~ final clusters $C_{F,1}, \ldots, C_{F,Z}$ ($C_{F,i} \cap C_{F,j} = \emptyset, \forall i = 1,..,Z, \forall j = 1, \ldots, Z$), with $Z \geq H$. Thus, ~~the~~ adopting~~on of~~ these ~~two~~ steps ~~would permit~~ makes it possible to represent~~a representation of~~ the dynamic of the MPU's presences, in terms~~the form~~ of a representative *DDP* for each group of days~~in region r~~.~~, where, with, with r~~Representative here ~~we~~ means~~intend~~ that ~~to~~ each group ~~belongs~~ includes days that are similar, in terms of index *i* (spatial distribution of MPU) and index *t* (temporal dynamic of MPU), ~~each other simil~~ar~~ar each other and~~and dissimilar~~in between,~~ from those included in other groups. ~~in terms of index i (spatial distribution of MPU) and index t (temporal dynamic of MPU)~~

2.2 Population assessment

With the~~If the~~ clustering strategy ~~makes~~ it is possible to display~~represent the dynamic of~~ the amount~~presence~~ of mobile phone users in region *r* for~~over~~ a set of time instants in clusters of days~~for a group of "representative" days.~~, ~~an additional strategy is needed~~However, ~~to~~the estimate of the total amount of people is needed for developing dynamic exposure maps. ~~Indeed~~Unfortunately, ~~usually~~~~in most times,~~ data ~~are~~ availability~~le~~ regards ~~just for~~only one mobile phone company. To have a reliable estimation~~ed~~ of the total number of people ~~that are~~ actually present in the study area, users of other mobile phone providers must be considered, as well. Collecting all these data is either unfeasible or ~~an~~ unsustainably expensive. To perform ~~an national level analys~~analysis ~~is at~~on a national ~~level~~scale, a convenient solution is represented by the use of the market share of the provider company, that can be applied to "correct" the DDP_{rd}. Hence, an estimation~~e~~ of the total number of people can be obtained (e.g. let s_n to be the national level share assuming~~than can assume~~ values in the range [0,1], the correct *DDP* would~~may~~ be $DDP_{correct} = \frac{DDP_{rd}}{s_n}$) ~~can be obtained~~. Country-level estimates are available from~~through~~ *Il Sole 24 Ore* newspaper (Il_sole_24_ore, 2017). However, the market share usually varies significantly among cities, according to different social-economic characteristics of users. For instance, per-capita revenues are on average 19,514 €/year in Italy and 23,418 €/year in the Brescia Municipality (data by Ministry of Economy and Finance, Department of Finance, 2016), whereas the percentages of foreigners are 8.5 % and 18.5 %, respectively (data by Italian National Institute of Statistics ISTAT, 2017). Furthermore, families featuring more than 4 people are about 21.0 % in Italy and 16 % in the Brescia Municipality, while the percentage of people older than 65 is quite near to the national average of about 22.0 % (ISTAT, 2017).

Thus, to suitably estimate the market share, the smallest level of aggregation, represented by the "Sezioni di Censimento" (SC) (i.e. population census districts), was used in this study (ISTAT, 2017). T~~We suggedt the~~~~The~~ following strategy is suggested ~~is suggested~~: ~~we compare~~ the~~data on the~~ number of residents from administrative archives is compared with ~~were compared to~~ the number of TIM users ~~on~~ in a residential area in the late evening hours. Bering in mind the characteristics of the social dynamics of the analysed~~is~~ residential areas, it is~~s~~ reasonable to assume that, in the late evening hours, ~~during these hours~~

ha formattato: Tipo di carattere: Non Corsivo

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito) residential SCs are only populated~~ only~~ by residents. The ~~Such ~~comparisons, using data from ISTAT (*Anagrafe Comunale*), were performed separately for each SC. ~~using ISTAT data (*Anagrafe Comunale*).~~

S~~First, s~~ince the MPU grid is made of square cells while SCs are irregular polygons, the number of TIM users belonging to each SC was estimated by intersecting these spatial data. Thus, ~~to count the number of TIM users in each polygon ~~the portion of the cell belonging to the SC polygon ~~were ~~was calculated in order to count how many TIM users are present in each polygon, by using the function *extract* in *raster* package, *R*. Let $Cell_j$; $j = 1, 2, \dots, J_{SC}$ be the TIM cells (pixels) which overlaps ~~overlapping~~ a chosen SC, the ratio A_j in Eq. (2)

$$A_j = \frac{area(SC) \cap area(Cell_j)}{area(Cell_j)}, \qquad (2)$$

~~which ~~represents how much of ~~the portion of ~~$Cell_j$ is included~~covered~~ in~~by~~ the chosen SC (for each SC, $A_j > 0; j = 1, 2, \dots, J_{SC}$). If $Cell_j$ is completely covered~~included~~ by~~in~~ SC, then $A_j = 1$, otherwise $A_j < 1$. Let TUC_j be the density of TIM Users in $Cell_j$, the estimation~~ed~~ of the number of TIM users in SC ETU_{SC} is computed as shown in Eq. (3).

$$ETU_{SC} = \sum_j TUC_j * A_j \qquad (3)$$

The Estimated TIM Market Share in SC $ETMS_{SC}$ is thus given by the ratio in Eq. (4), where P_{SC} is the resident number assessed by ~~the ~~population census ~~the ~~for the SC.

$$ETMS_{SC} = \frac{ETU_{SC}}{P_{SC}} \qquad (4)$$

Differently from s_n, the $ETMS_{SC}$ range is not necessarily in [0,1], since P_{SC} ~~ETU_{SC}~~ could be smaller than ~~larger than~~ ETU_{SC} ~~P_{SC}~~. An application example of this procedure can be found in Metulini and Carpita (2019b). In the count of P_{SC}, ~~elderly people (>80 years) and ~~children (younger than ~~<~~ 11 years) and elderly people (older than 80 years) ~~and ~~were excluded, aiming at taking into consideration only people bearing a smartphone. The distribution of $ETMS_{SC}$ can be used as a proxy for the TIM market share ~~at ~~on a city level. More specifically, it appears to be convenient to use the median of the distribution of $ETMS_{SC}$. The median is preferable to the mean in those cases when the distribution is symmetric. Metulini and Carpita (2019b) showed the presence of a strongly asymmetric distribution of TMS_{SC} for the case study of the city of Brescia. Moreover, Carpita (2019) showed good results in terms of the comparison between estimated people and official data by using the median for the case study of the Lake of Iseo during the Floating Piers.~~More specifically, it appears to be convenient to use the median of the distribution of $ETMS_{SC}$, which is preferable to the mean in those cases when the distribution is asymmetric~~. Let $me(.)$ be the median statistics, the estimate $\widehat{DDP}_{rd}$ for a given region r for a given day d is finally given by Eq. (5).

$$\widehat{DDP}_{rd} = \frac{DDP_{rd}}{me(ETMS_{SC})} \qquad (5)$$

2.3 Result representation

For the sake of result interpretation, a graphical representation is herein adopted. Let us consider the~~the vector~~ DDP_{rd} vector to be a functional curve~~ of functional observations~~ x_{rd} (T_d) displaying~~representing~~, in the *y*-axis, the sum of MPU in region *r* and day *d* ~~(in y-axis)~~ with respect to~~, in the x-axis,~~ time instants $T_d \in (t_1, \dots, t_o)$ in the *x*-axis ~~(in the x-axis)~~. Functional box

ha formattato: Tipo di carattere: Corsivo

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito), Pedice

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito), Pedice

ha formattato: Tipo di carattere: Corsivo

plots (FBP) (Sun and Genton, 2011; 2012) can be used to display t~~T~~he profile for representative days.~~ can be displayed by using functional box plots (FBP), the analogue of the traditional box plot for curves (Sun and Genton, 2011; 2012).~~

In FBP a cluster~~group~~ of curves is ordered in term of its~~using the concept of~~ "band depth".~~,~~ A~~a~~ "median" value and an "envelope" ~~is generated, that is~~are generally~~can be~~ used to define the~~a~~ functional counterpart~~version~~ of traditional descriptive bands. Moreover, ~~it is possible to~~one can ~~assign a curve to the outlier group~~detect outlier curves. ~~,~~Specifically, a curve is an outlier ~~-~~if it exceeds ~~by 1.5~~ the margins of the envelope~~ margins~~ by 1.5 in at least one considered time instant.

Separately for each final cluster, a~~A~~ FBP ~~strategy~~ is derived~~performed separately for each final cluster~~. Let us consider cluster h, ~~and~~ let $d_h = \{d_{1h}, \dots, d_{nh}\}$ be the group of days that belong~~ing~~ to cluster h, and let $\widehat{DDP}_{rd,h} = [\widehat{DDP}_{r,d1h}, \dots, \widehat{DDP}_{r,dnh}]$ be the matrix of dimension $o*n_h$ with a $\widehat{DDP}_{rd}$ in each column. ~~Let~~ By ~~c~~considering each vector as a~~to be a~~ curve, ~~-~~the FBP representing the profile plot of the total number of people (that we will call "city users", or simply "users") ~~in~~at different hours (with DB) ~~in~~on representative days is~~ applied~~ computed using~~to~~ matrix $\widehat{DDP}_{rd,h}$~~ to generate the profile plot estimating the dynamic of the total number of people (that we will call "city users", or simply "users") in different hours (with DB) in representative days.~~.

3 Case study description

The study area was selected as an emblematic and widespread situation of the unacceptable high flood risk affecting the foothill zone of the Po River Valley (Lombardy Region, northern Italy). It lies ~~in~~ on the western outskirts of Brescia ~~town~~ (Figure 1) and is overall included in the watershed of the Oglio River, a primary left-bank tributary of the Po River. The main drainage is supplied by the Mella River, bounding the eastern side of the study area, that is a left-bank tributary of the Oglio River. As can be seen in Figure 1, the study area features five natural streams originating from the southern boundary of the Alpine chain. From West to East, they are: Laorna, Gandovere, Vaila, La Canale and Solda.

Before the area was anthropized, most of such streams had probably ~~had been~~ flooded~~ing~~ the alluvial plain swamping into marshes, without a main outlet in the main river network. This was the result of both their almost ephemeral regime and the ~~terrain~~ endorheic morphology of their watersheds. As the agricultural use of the alluvial plain ~~grooved~~grew, these streams were connected to the constructed irrigation-drainage network, in order to exploit their low flow for irrigation purposes and to drain the flood flow into the Mella River. These constructed downstream reaches feature two main drainage canals, the Gandoverello canal and the Mandolossa canal, depicted in Figure 1. They have become the artificial downstream reaches of the five watersheds, providing a drainage capacity in the South direction both for the mountain watersheds and low lands located in the alluvial plain. The total catchment area amounts to 112.3 km^2, featured by an average imperviousness of about 22%; further details on the watershed hydrologic characteristics are available in the supplementary material.

In particular, the Laorna drainage path was strongly manipulated by a constructed straight canal 7 kilometres long, that diverts its streamflow towards the Gandovere stream and intercepts additional surface runoff produced by the northward low land. Both streams thus confluence into the Gandoverello canal. This was formerly the downstream reach of the Gandovere stream, whose limited conveyance capacity made it necessary to decrease the hydrologic load. Thus, a flow divider was constructed upstream of the Laorna confluence, so that almost half of the Gandovere streamflow is diverted towards the Mandolossa canal inlet by means of a diversion constructed canal that intercepts the Vaila stream. All these flow discharges, along with those coming from the La Canale stream and the Solda stream, converge into the inlet of the Mandolossa canal, which is characterized by the largest conveyance capacity in the study area.

This drainage network is also exploited by the irrigation system of Franciacorta, a vineyard-agricultural district located West of the study area, for the final disposal of the residual flow discharges. In Figure 1 two of the most important irrigation canals belonging to this system, Seriola Castrina and Seriola Nuova, are reported. Their discharges mainly affect the Gandovere and Laorna streamflows. Further contributions come from the West Brescia outskirt, in terms of both urban stormwaters and irrigation excess flows, which directly drains into the Mandolossa canal. Finally, water table resurgences of the high Po River Valley are also present downstream of the Mandolossa canal inlet. Their fresh waters are however intercepted and drained by the downstream reach of the Mandolossa canal.

3.1 Hazard mapping

Since the late 50s of the last century, this area has been subject to a deep urbanization sprawl, yielding the present land cover condition depicted in Figure 2. The dramatic increase in both the urban fabric and the industrial-commercial coverages has occurred at the expenses of ~~the ~~croplands and ~~the ~~permanent crops, so that sparse and isolated fabrics have evolved in a continuous and heterogeneous urbanization. In addition, various transport units have been upgraded to speedways and two highways were constructed. The flood risk perception in this this area ~~was ~~has historically been related to the Mella River inundations, which affected the Brescia outskirts ~~since ~~in the late 60s of the last century, until its riverbed underwent ~~severe~~ important engineering works. Conversely, with reference to the secondary stream network, the absence of a clear risk perception allowed such an urbanization sprawl to occur regardless of the floodplain extents. As Figure 2 clearly shows, most of the urban fabric areas and the industrial and commercial settlements are adjacent to the stream network.

The increase in the land coverage ~~yielde~~yields~~d~~ huge impermeability degrees for the plain watersheds. Except for a combined sewer overflow located in the West Brescia watershed ~~and ~~that discharges~~ing~~ into the Mella River, all the stormwaters produced by these urbanizations discharge into this secondary stream network, as well as those produced by many settlements in the Franciacorta district, which improperly exploits the irrigation system as a final receipt of combined sewer systems. Moreover, the low risk perception has led to a significant impairment of the functionality of these canals. The stream flow is now constrained into a number of narrow culverts and bridges with low decks and large piers. In addition, urbanized canals are no longer maintained, so that the riparian vegetation grow~~ve~~s in an uncontrolled manner. The combination of the increase in the peak flow discharges and in the exposure, along with the decrease in the stream network conveyance capacity has led to a dramatic increase in the flood risk. Flood episodes explained by the secondary hydrographic network insufficiency have been observed since the late 90s, evidencing an empirical frequency of occurrence far less than 20 years. The hazard mapping was thus referred to return periods spanning from 5 years to 20 years, which are significantly less than those conventionally required in Italy for a secondary stream network to be considered verified (20-50 years). ~~The urban areas exposed to floodings are estimated in 160 ha, 231 ha and 330 ha, with respect to 5 years, 10 years and 20 years return periods. About 30% of those areas are residential fabrics whereas the remaining 70% is given by industrial and commercial settlements.~~

The hazard analysis was herein conducted by using a design event method. As demonstrated by Balistrocchi et al. (2013) in this climatic context, a design event method is capable ~~to~~ of providing~~e~~ results comparable to those of more sophisticated continuous approaches, if it is based on the Chicago synthetic hyetograph with a duration equal to the double of the catchment time of concentration. A classical leaf hydrologic model was developed in accordance with the sub-catchment~~s~~ subdivision illustrated in Figure 1. Extensive surveys of the stream cross sections and the inline structures were carried out to assess the actual conveyance capacities of the stream reaches. Flooding volumes were hence estimated by limiting the flood hydrographs generated through the hydrologic model to the overflow threshold discharges. Surveys of the historical flooding extensions that occurred during the last three decades addressed the delimitation of the flood prone areas. The total amount of exposed urban areas ~~amount to~~is about 160 ha (return period 5 yr), 230 ha (return period 10 yr) and 330 ha (return period 20 yr). Most of such areas are devoted to industrial-commercial settlements (70 %), whereas the remaining part ~~is~~ includes residential ~~fabrics~~ areas of medium-low density featuring similar fabric types (detached or semi-detached houses). ~~To decrease the uncertainty of people estimates, d~~Distinguishing between these two classes was found to be useful in decreasing the uncertainty of people estimates. The resulting flood hazard map is reported in Figure 2 along with the land cover, and it highlights the large amount of residential fabrics, industrial and commercial settlements potentially affected by ~~storm~~ flood events featuring low-medium return periods. An unacceptably high flood risk is therefore evidenced for the study area. Most of such areas were too dispersed or small to have suitable intersections with MPU grid cells. Therefore, they were grouped into four macro-areas referred to the ~~river network~~individual canals and the land use, that was ~~distinguished~~ classified in~~between~~ urban fabrics (red) and commercial-industrial settlements (dark green). As shown in Figure 2 they are: the Laorna and Gandovere streams confluence (areas 1 and 5), the La Canale and Solda streams (areas 2 and 6), the southern Gandovere canal (areas 3 and 7), the Mandolossa canal in Roncadelle Municipality (areas 4 and 8).

3.2 Available mobile phone data

In t~~T~~his work we use ~~focuses on~~ mobile phone data provided by Telecom Italia Mobile (TIM), that~~which~~ is currently the largest mobile phone operator in Italy. ~~in this sector.~~ According to the national economic newspaper, TIM's national share amounted to 30.2 % in December 2016 (*Il Sole 24 ore*). In our analysis, Erlang measure data represent the average number of both calling and non-~~t~~calling mobile phone SIMs (~~both calling and not calling) that are~~ assigned to a certain ~~that~~cell of the considered grid ~~grid's cell~~ in a certain~~that~~ quarter. Statistical research in the area of urban planning using Erlang measure data is increasing: for example~~,~~ ~~have already been used in the context of urban planning along with statistical methods by~~ Carpita and Simonetto (2014) and Metulini and Carpita (2019a)~~, who~~ studied~~analyzed~~ the dynamics of people's presence ~~inof people during~~at big events in the city of Brescia;~~, by~~ Zanini et al. (2016)~~, who find,~~ ~~by~~ applied~~yingmean of~~ an Independent Component Analysis (ICA) method, for separatinge the city of Milan in a few main areas. ~~a number of spatial components that separate main areas~~

~~of the city of Milan~~ , ~~and by o~~Other works~~s~~ (Manfredini et al., 2015, Secchi et al., 2015) used Erlang measure data to study the dynamics of people's presence in Milan.~~.~~

In this study, reliable Erlang measures of MPU recorded by the TIM company are available. The investigated area is marked in a black solid line in Figure 1 (WGS 84 UTM 32 N coordinates: 5,040,920–5,049,980N, 585,970–592,970E, area about 64 km^2) and is ~~centered~~centred on the Mandolossa-Gandovere network. The area is covered by a georeferenced grid of square cells with 150 m sides, which provides the number of TIM users every 15 minutes. ~~In details~~More precisely, for each ~~grid's~~ cell of the grid and for each~~h time~~ interval of time, the ~~corresponding~~ recorded data refer~~s~~ to an~~the~~ average measure~~number~~ of the number of mobile phones simultaneously connected to the network. For instance, Figure 3 shows a detail of ~~–~~the spatiotemporal distribution of TIM MPUs occurred on Wednesday, November 18th 2015, in a sample area of 20x20=400 cells, near ~~to~~ the Mandolossa inlet. Therein, exposed areas, obtained by intersecting the urban covers with the flooding areas, were also reported. Thus, Figure 3 provides a sequence of snapshots of a dynamic map of people exposure to floods. As can be seen, the spatial distribution of raw data is realistic, as major densities ~~suitably concentrate~~are observed along the main street network and in the urban areas. The temporal variability is also reasonable; for instance, lower densities are evidenced during nighttime in industrial sectors and main streets; see, for instance, the industrial settlement near ~~to~~ the confluence of the La Canale stream in the Mandolossa canal (flooding area marked with 6 in Figure 2). ~~The mobility feature~~ In ~~of~~ these data, the information about the user mobility is hidden, meaning that one ~~can not~~cannot~~it is not possible to~~ trace the path followed by a single MPU over time. Measures are available ~~in~~ for the period 2014–2016, even though after data inspection a more limited subset was found to be suitable for the analysis (from July 1st 2015 to August 11th 2016), due to data collection issues.

4 Analysis procedure application

4.1 Procedure parameterization

The application of the HOG procedure to reduce the dataset dimensionality was performed for each quarter of a day in T_d, by dividing the full~~original~~ grid in 9 smaller grids G_i, $i = 1, \ldots, 9$. The parameter $\sqrt{S}$ was thus set at 3. For each G_i, gradients and direction were then computed and the histogram of oriented gradients choosing~~with~~ k=4 bins corresponding, respectively, to angles 0°–45°, 45°–90°, 90°–135° and 135°–180° was obtained. In general, one can improve the recognition of the ~~analyzed~~analysed grid~~object is improved~~ by increasing the number of bins. This value was chosen in order to maximize k but, at the same time, to ~~avoid~~avoid~~ing~~ the presence of HOG features with ~~a~~ zero values~~ of zero~~~~zeros~~ ~~among~~~~in the vector of HOG features~~. In each quarter, the ~~e~~Extracted features are~~count for~~ $3^2*4 = 36$ ~~in each quarter.~~ ~~,So~~Therefore,~~ which correspond to a~~ the order of dimensionality reduction is of ~~n the order of~~ 400/36 ≈ 11. The final vector κ_d for the sample area near ~~to~~ the Mandolossa inlet (sample area evidenced in Figure 3), with ~~stacked~~ all the ~~all~~ quarters stacked ~~of~~ for the same day~~y stacked for sample area near to the Mandolossa inlet (sample area evidenced in Figure 3)~~ and for day d~~-~~, ~~counts~~~~amounts to~~contains 36*96 = 3456 features.

ha formattato: Tipo di carattere: Corsivo

The hierarchical *k*-means cluster analysis, with days used ~~here~~as the objects of the cluster~~ are the days *d*~~ and the features of κ_d used as cluster variables, ~~are represented by~~ ~~the features of κ_d,~~ was performed on a total amount of 360 days (*d* = 360) from July 1st 2015 to August 11th 2016. After data inspection, only the days of the last available year (from July 1st 2015 to August 11th 2016) were included in the analysis, since the first year (April, 2014 to June 30th 2015) features some collection problems. In effect, a configuration with 3 clusters sharply separating the days of the first year (till June 2015) and the days of the second year (by July 2015) was estimated, by performing the cluster analysis by using the full set of data. In the final sample all holidays were removed, in consideration of their specific characteristics with respect to typical days. More precisely, August, 15th, 1st and 2nd November, 8th December, 24th to 26th December, 31th December, January 1st and January 6th. 27th and 28th March (Easter), 25th April, May 1st, June 2nd were removed. In addition, those days where a large amount of data (>10%) were missing were removed, as well. Conversely, data in those days where missing data were less than 10% were maintained. A test for the possible presence of curse of dimensionality, based on the distribution of the distances among pairs of objects, ~~has been~~was performed. A unimodal distribution that suggests the absence of such a problem was derived. On the whole, the amount of suitable data appears to be sufficient to get reliable estimates of people exposed to the flooding risk in the study area.

The number of first-step clusters was chosen according to the relative decreasing trend of the total within sum of squares with the ~~increase in the number of groups~~group number increase. Figure 4 shows this trend, evidencing that a splitting in 4 clusters would decrease ~~by half~~ the total within sum of squares by half with respect to a 1 cluster splitting. Since this decrease appears to be satisfactory, the sample of days was split in H = 4 clusters, where, C_1 ~~mainly~~ corresponds to all the days ~~of~~ mostly occurring in July, August and September (green spine-plots shown in Figure 5), C_2 corresponds to working days mostly occurring from February to June (blue spine-plots shown in Figure 5), C_3 corresponds to working days mostly occurring from October to January (red spine-plots shown in Figure 5) and C_4 corresponds to the weekends except for those ~~of summer~~included in cluster C_1 (yellow spine-plots shown in Figure 5).

Hence, ~~SMU~~ MPU variability over time instants was accounted for by considering the Mandolossa's *DDP* of each day as a functional curve. Firstly, those days that have to be considered outliers were removed by using the curve outlier detection method (Febrero-Bande et al., 2008) separately for each first-step cluster. Secondly, it was evaluated whether days should be further grouped in terms of dissimilarity in the *DDP* functional curve dynamic. To do so, the assumption of independence of our functional data was tested by using Portmanteau (Gabrys and Kokoszka, 2007) and distance correlation (Székely and Rizzo, 2013) tests. Model based functional data clustering techniques (Bouveyron et al., 2015) suggests ~~to split~~splitting the "summer" group in 3 sub-groups, containing, respectively, the days of July, the days of August and the days of September. This second-step splitting leads to Z=6 final clusters, where $C_{F,1}$ includes days of July, $C_{F,2}$ includes days of August, $C_{F,3}$ includes days of September and $C_{F,4}$, $C_{F,5}$ and $C_{F,6}$ match, respectively, C_2, C_3 and C_4.

Illustrations of *DDPs* for representative days by using functional box plots are reported in Figure 6 (residential fabrics) and in Figure 7 (commercial and industrial settlements), for each of the 8 flooding areas shown in Figure 2 for the 10 year~~s~~ return period. Although functional data clustering suggests ~~the~~ splitting into six groups, for the sake of clarity, the summer months

ha formattato: Tipo di carattere: Corsivo

ha formattato: Tipo di carattere: Corsivo, Pedice

ha formattato: Tipo di carattere: Corsivo

(July, August and September) were combined in a single final cluster (Cluster 1, C1), as well as all the working-days from October to June in a single final group (Cluster 2, C2) and the weekends from October to June (Cluster 3, C3), thus leading to 3 clusters.

To extract the number of people in each quarter of each day from the grid's cells to the irregular polygon of each area (i.e. to find $DDP_{ri,d}$ for area r_i , i = 1, .., 8, and day *d*) the procedure described in Section 2.2 was applied. Hence, MPUs were firstly divided by a constant c = 0.85, in order to consider children (>12 years) and old people~~s~~ (>80 years), who likely do not have smartphones (i.e. about 85% of the people are in the age range [12,80] in Brescia). Then, by estimating the median value of the market share ratio at the SC level by adopting the methodological strategy~~strategy~~ in Section 2.2, which amounts to about 20%, the estimated DDP_s for each area and for each day were derived by applying Eq. (5). The estimated market share is also consistent with that found by Carpita and Fabbris (2019). Estimated $DDP_{ri,d}$ underwent the functional box plot strategy, separately for ~~the~~ days *d* in the 3 clusters (with outliers excluded) and for r_i corresponding to the 8 areas illustrated in Figure 2.

4.2 Results and discussion

Figure 6 and Figure 7 ~~report~~show, respectively, the resulting functional box plots for residential and ~~for~~ productive areas, ~~reporting~~ where the estimated number of city users in different time instants is indicated. ~~, separately for t~~The three clusters of days are reported separately. Overall, the number of city users is lower during the first hours of the day and ~~it~~ increases in the morning, reaching a peak during working hours (9 am–1 pm and 2–6 pm), both in residential and in productive areas. In the Moie di Sotto residential area located at the confluence of the Laorna stream and the Gandovere stream (flooding area 1 in Figure 2), the people number is estimated at about 200, during the first hours of the day and during the night and increases to about 250 ~~in~~ during working hours (inhabitant density is about 25–~~-~~30 ha^{-1}). The dynamic, similar in all the three clusters, shows irrelevant differences among different periods of the year.

In the Villaggio Badia residential area located North of the Mandolossa canal inlet (flooding area 2 in Figure 2), the city user number varies between a minimum of 1200 people and a maximum of 1400 people, during an~~d~~ average day (inhabitant density is about 30–~~-~~35 ha^{-1}). During the working-days of the months from October to June (cluster C2), the peak reaches 1600 users. Moreover, the descriptive bands appear to be wider in summer (cluster C1) and on weekends (cluster C3) as compared to cluster 2, where bands are narrower (*i.e.* lower variability between days).

Residential areas along the southern Gandovere canal (flooding area 3 in Figure 2) are ~~little~~ not very populated. Only 50–70 users are there during an average summer day ~~of summer~~ (cluster C1) or during the week~~-~~end (cluster C3) (inhabitant density is about 18–~~-~~23 ha^{-1}). The number amounts to more than 80 people on working hours of working days (cluster C2). The n~~N~~umber of city users in Roncadelle's residential area located along the Mandolossa canal (flooding area 7 in Figure 2) is less sensitive to working hours, especially during summer. In summer~~, the number of~~ city users var~~i~~ies from a minimum of about 600 up to a maximum of 700. ~~City users~~There are about 800 city users during working hours of working days and weekends (days belonging to clusters C2 and C3).

ha formattato: Tipo di carattere: Non Corsivo

ha formattato: Tipo di carattere: Non Corsivo

ha formattato: Tipo di carattere: Non Corsivo

ha formattato: Tipo di carattere: Non Corsivo

ha formattato: Tipo di carattere: Non Corsivo

ha formattato: Tipo di carattere: Non Corsivo

ha formattato: Tipo di carattere: Non Corsivo

ha formattato: Tipo di carattere: Non Corsivo

Industrial and commercial settlements of Moie di Sotto (flooding area 5 in Figure 2) feature 1000–1500 people (night, first hours of the day–working hours) in summer. These numbers increase up to about 1200–1800 ~~in~~ on working days and to about 1100–1600 ~~in~~ at weekends. This high density is mainly due to the presence of a commercial outlet of regional interest in this area. Industrial and commercial settlements along the La Canale and Solda stream network (flooding area 6 in Figure 2) are very highly populated by city users and the difference between the number of people in summer and ~~in~~ on working days is significant. Daily minimum and maximum values are included between 2000 and 3000 in summer, and between 2500 and 3500 during working days. Weekends follow a stable dynamic (about 2500 people ~~all along the~~throughout the entire day). Flooding areas related to the southern Gandovere canal (flooding area 8 in Figure 2) presents a productive area with an average number of users varying from 250 to 380 in summer and from 300 to 420 during working days. In the same way ~~of~~ as Villaggio Badia (southern part of flooding area 2 in Figure 2), the number of users ~~along~~ during the weekends is stable and it stands ~~to~~ at about 320/330. The industrial sector of Roncadelle (flooding area 4 in Figure 2) is another highly populated area, featuring more than 2000 people during the day. ~~In~~ On some particular working days and weekends this number reaches about 3000 (red dashed lines of outliers in Figure 7). In this area, a remarkable difference in the number of users between working days, summer days and weekends was not detected.

5 Conclusions

In this paper a novel approach to the assessment of the risk related to people exposure to floodings based on a geostatistical analysis of Erlang measures was proposed. Such a procedure takes advantage of data reduction (histogram of oriented gradients discussed in Section 2.1), in order to face the dimensionality curse issue. Its suitability and potentials were demonstrated with regard to an urban outskirt area located near ~~to~~ Brescia (Lombardy, Northern Italy), which is affected by widespread and frequent floodings. In Figure 3 the possibility of expressing the spatiotemporal variability of exposed people by using time variable maps was illustrated. These data feature high spatial resolution (150 m) and short time step (15'), thus providing reliable assessments even for the smallest ~~analyzed~~analysed areas (about 4–5 ha) and a precise evaluation of the temporal dynamics. Indeed, daily density profiles can be derived according to this procedure. Then, these profiles can be clustered yielding groups of similar daily time patterns. Clustering results are definitively meaningful, since working days and weekends are acknowledged to show different temporal dynamics, when they belong to working months (from October to June). Conversely, daily dynamics in summer months (July, August and September, usually exploited for the longest hol~~y~~idays in Italy), must be regarded as different from the others. In addition, working days and weekends feature more similar daily density profiles during such months. As can be seen in Figure 6 and in Figure 7, the daily temporal variability of people exposed to floodings can be assessed with respect to the day cluster and the type of urban areas (residential or industrial–commercial), both in terms of expected value (the median) and uncertainty (confidence band), thus providing ~~a~~ comprehensive information to agencies and authorities devoted to the flood risk management.

The need to assess the entire population would theoretically require the gathering of a huge amount of datasets, from all the providers that operate in the area of interest. This issue would lead to a ~~relevant~~ remarkable increase in the data collection cost and would be difficult to overcome. Nevertheless, census data make it possible to infer the total population from the users of a single provider, by means of local estimates of its market share, as discussed in Section 2.2.

It is worth underlining that this statistical support, along with the high spatiotemporal resolution and the reliability of the raw data, makes the proposed procedure particularly appealing in order to decrease the errors of exposed people estimates. Such a support is not provided by crowdsourcing techniques, which are based on voluntary data supplies and commonly rely on very limited datasets with respect to the amount of the exposed people. A second advantage that must not be disregarded lies in the possibility of exploiting dynamic exposure maps, or alternatively the clustered daily density profiles, by directly implementing them in emergency plans, ~~off-line that is independently of~~regardless of the potential malfunctioning of the mobile phone connection during the flood episode. Conversely, crowdsourcing could be strongly compromised by the difficulties of connecting to the network during the emergency period. Indeed, dynamic exposure maps derived by mobile phone data have strong potentials to substantially improve emergency plans, so that real-time rescues, relief supplies and traffic management could be better addressed~~.~~.

Future developments of the ~~geo-statistic~~geostatistical approach proposed in this paper could be addressed towards multiple items which deserve further in-depth analyses. First, complete mobile phone data feature more additional information than those available for this study. Actually, datasets include the collection of matrices of OD-Origin-Destination vectors in different seasons, days of the week and hour of the day, which are ~~referred to~~constructed by using the SIM identification numbers. Hence, it would be possible to track users down. By knowing the density of origin-destination vectors ~~with origin and destination~~ of the paths around critical traffic nodes, it would be possible to forecast potential critical conditions for mobility and better manage traffic in a more precise manner.

Second, coupling traffic management decision support systems with real-time rainfall-runoff-flooding modelling is also a research perspective being considered. Presently, the exploitation of mobile phone data in real-time is problematic. Nevertheless, in the future, a more common use of 5G and GPS technologies in mobile devices will facilitate the real-time assessments of people spatial distribution. From a prevention perspective, this could make ~~it possible~~ the identification of preferential traffic flows possible, thus evidencing potential risks during inundation onsets or emergency situations. Alternative safe pathways could be identified and ~~enforced~~communicated to exposed people, in order to facilitate their evacuation.

Third, it would be possible to profile the SIM users, ~~even though keeping~~while remaining anonymous~~ness~~ and respecting their privacy. Users could be categorized, in order to isolate specific targets from the whole user set, and their ~~behaviors~~behaviours could be statistically ~~analyzed~~analysed separately from the others. Thus, a future development of the statistical matching procedure between mobile phone data and census data could use demographic and socio-economic information about the SC areas, for example the ISTAT ARCH.I.M.E.DE database (www.istat.it/it/archivio/190365). Since~~,~~ it is likely to assume heterogeneous ~~behaviors~~behaviours of individuals, this database will be beneficial in future works in order to ~~share~~organize individuals in classes in terms of their age, gender, income or their job. In fact, different mobile phone companies have different costs, and this may affect ~~differently~~ the choice of different classes of individuals.

Fourth, exposed people ~~behaviors~~behaviours and habits can significantly change after hydro-climatic alarms or during flood event~~ occurrence~~s, ~~as~~since people ~~enhance consciousness~~would be aware of flood hazard ~~-~~and limit their exposure in flood prone areas. Mobile phone data make the identification of anomalous exposed people mobility during alarms possible, while the ~~geo-statistic~~geostatistical approach herein proposed provides a tool to ~~analyze~~analyse whether, and how far, people ~~behaviors~~behaviours are different from those of common days. Actually, a sample far larger than two years would be beneficial to make such statistics more reliable, since alarms ~~interest~~involve a few days in a year. It is worth noting that in the ~~analyzed~~analysed area the risk perception towards the secondary network is almost absent, as well as a capillary local warning system. Flood risk perception is mainly related to the primary hydrographic network (i.e. Mella River). Therefore, ~~in the~~ regardings~~ of~~ the specific test case, in this research the possibility of drastic changes in human ~~behavior~~behaviour during heavy rainfall alarms are not expected. Indeed, people virtuous ~~behaviors~~behaviours are usually the result of extensive campaigns to raise public awareness against flood risk, coupled with trusted and effective warning systems. Hence, a final objective of the ongoing research will take into consideration the effectiveness of the non-structure practices that will be adopted to mitigate flood risk in the test watershed.

6 Acknowledgements

This research was developed at University of Brescia in the course of the FLOod RIsk MAnagement Policies (FLORIMAP) Project which received funding by the Cariplo Foundation under grant number 2017/0708. The content of this article does not reflect the official opinion of the Cariplo Foundation. Responsibility for the information and views expressed in this paper lies entirely with the authors. The research will continue~~s~~ after additional funding by the Lombardy Region (Mo.So.Re. project on sustainable and resilient mobility).

References

Agrawal, R., Gehrke, J., Gunopulos, D., and Raghavan, P.: Automatic subspace clustering of high dimensional data for data mining applications, in: Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data, ACM Press, 94–105, doi:10.1145/276304.276314, 1998.

Arabie, P., and De Soete, G.: Clustering and classification, World Scientific, doi:10.1142/1930, 1996.

Arrighi, C., Pregnolato, M., Dawson, R. J., and Castelli, F.: Preparedness against mobility disruption by floods. Sci. Total Environ., 654, 1010–1022. doi:10.1016/j.scitotenv.2018.11.191, 2019.

Balistrocchi, M., Grossi, G., and Bacchi, B.: Deriving a practical analytical-probabilistic method to size flood routing reservoirs, Adv. Water Resour., 62, 37–46, doi:10.1016/j.advwatres.2013.09.018, 2013.

Balistrocchi, M., Moretti, G., Orlandini, S., and Ranzi, R.: Copula-based modelling of earthen levee breach due to overtopping, Adv. Water Resour., 134, 103443, doi:10.1016/j.advwatres.2019, 2019.

Barredo, J. I.: Normalised flood losses in Europe: 1970-2006, Nat. Hazard. Earth Sys., 9/1, 97–104, doi:10.5194/nhess-9-97-2009, 2009.

Becker, R. A., Caceres, R., Hanson, K., Loh, J. M., Urbanek, S., Varshavsky, A., and Volinsky, C.: Route classification using cellular handoff patterns, in: Proceedings of the 13th international conference on Ubiquitous computing, 123–132, ACM press, 2011.

Blöschl, G. et al.: Changing climate both increases and decreases European river floods, Nature, 573, 108–111, doi:10.1038/s41586-019-1495-6, 2019.

Bouveyron, C., Girard, S., and Schmid, C.: High-dimensional data clustering, Comput. Stat. Data An., 52(1), 502–519, doi:10.1016/j.csda.2007.02.009, 2007.

Bouveyron, C., Come, E., and Jacques, J.: The discriminative functional mixture model for a comparative analysis of bike sharing systems, Ann. Appl. Stat., 9(4), 1726–1760, doi:10.1214/15-AOAS861, 2015.

Bouwer, L. M., Bubeck, P., and Aerts, J. C. J. H.: Changes in future flood risk due to climate and development in a Dutch polder area, Global Environ. Chang., 20/3, 463–471, doi:10.1016/j.gloenvcha.2010.04.002, 2010.

Calabrese, F., Ferrari, L., and Blondel, V. D.: Urban sensing using mobile phone network data: A survey of research, ACM Comput. Surv., 47(2), 2655691, doi:10.1145/2655691, 2015.

Carpita M.: The mobile phone big data tell the story of Christo's The Floating Piers impact on the Lake Iseo. In: Carpita M., Fabbris L. (Eds.) Book of Short Papers of the ASA 2019 Conference "Statistics for Health and Well-being", CLEUP, ISBN: 978-88-5495-135-8, 2019.

Carpita, M., and Fabbris, L.: The mobile phone big data tell the story of the impact of Christo's The Floating Piers on the Lake Iseo, in: ASA Conference 2019 Statistics for Health and Well-being Book of Short Papers, 2019.

Carpita, M., and Simonetto, A.: Big data to monitor big social events: Analysing the mobile phone signals in the Brescia smart city, Electron. J. Appl. Stat. Anal., 5/1, 31–41, doi:10.1285/i2037-3627v5n1p31, 2014.

CENSIS: 53° Rapporto sulla situazione sociale del ~~P~~paese ~~– 2019~~, Franco Angeli, ~~–~~https://www.censis.it/rapporto-annuale/sintesi-del-53%C2%B0-rapporto-censis, 2019. (*in Italian*)

Dawson, R. J., Peppe, R., and Wang, M.: An agent-based model for risk-based flood incident management. Natural Hazards, 59(1), 167-189, doi:10.1007/s11069-011-9745-4, 2011.

Dalal, N., and Triggs, B.: Histograms of oriented gradients for human detection, in: Proceedings of the International Conference on Computer Vision & Pattern Recognition (CVPR '05), doi:10.1109/CVPR.2005.177, 2005.

Donoho, D. L.: High-dimensional data analysis: The curses and blessings of dimensionality, AMS Math Challenges Lecture, 1–32, 2000.

ha formattato: Italiano (Italia)

Codice campo modificato

ha formattato: Italiano (Italia)

ha formattato: Italiano (Italia)

ha formattato: Inglese (Regno Unito)

ha formattato: Tipo di carattere: Corsivo, Inglese (Regno Unito)

ha formattato: Inglese (Regno Unito)

Di Baldassarre, G., Viglione, A., Carr, G., Kuil, L., Salinas, J. L., and Blöschl, G.: Socio-hydrology: Conceptualising human-flood interactions. *Hydrology and Earth System Sciences, 17*(8), 3295-3303, doi:10.5194/hess-17-3295-2013, 2013.

European Union: Directive 2007/60/EC on the assessment and management of flood risks, Official Journal of European Union, 227, 27–34, 2007.

Erlang. A.K.: The theory of probabilities and telephone conversations. Nyt Tidsskrift Matematika, B.20, 33-39, 1909.

Febrero–Bande, M., Galeano, P., and Gonzalez–Manteiga, W.: Outlier detection in functional data by depth measures, with application to identify abnormal NO_x levels, Environmetrics, 19(4), doi:10.1002/env.878, 331–345, 2008.

Feyen, L., and Dankers, R.: Impact of global warming on streamflow drought in Europe, J. Geophys. Res.-Atmos., 114/D17, D17116, doi:10.1029/2008JD011438, 2009.

Fuchs, S., Keiler, M., and Zischg, A.: A spatiotemporal multi-hazard exposure assessment based on property data, Natural Hazards and Earth System Sciences, 15, 2127-2142, doi:10.5194/nhess-15-2127-2015, 2015.

Gabrys, R., and Kokoszka, P.: Portmanteau test of independence for functional observations, J. Am. Stat. Assoc., 102(480), 1338–1348, doi:10.1198/016214507000001111, 2007.

Guntha, R., Rao, S., Benndorf, M., and Haenselmann. T.: A comprehensive crowd-sourcing approach to urban flood management, in: Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering (LNICST), 218, 13–24, Springer Verlag, 2018.

Goodchild, M. F.: Citizens as sensors: The world of volunteered geography, Geojournal, 69/4, 211–221, doi:10.1002/9780470979587.ch48, 2007.

Goodchild, M. F., Aubrecht, C., and Bhaduri. B.: New questions and a changing focus in advanced VGI research, T. GIS, 21/2, 189–190, doi: 10.1111/tgis.12242, 2017.

Haer, T., Botzen, W. J. W., and Aerts, J. C. J. H.: The effectiveness of flood risk communication strategies and the influence of social networks-insights from an agent-based model. Environmental Science and Policy, 60, 44-52, doi:10.1016/j.envsci.2016.03.006, 2016.

Han, S., and Ciravegna, F.: Rumour detection on social media for crisis management. Paper presented at the Proceedings of the International ISCRAM Conference, 660–673, 2019.

Hartigan, J. A., and Wong, M. A.: Algorithm AS 136: A k-means clustering algorithm, J. R. Stat. Soc., Series C (Applied Statistics), 28(1), 100–108, doi:10.2307/2346830, 1979.

Hartmann, D. L., et al.: Observations: Atmosphere and surface, In: Stocker, T. F., et al. “Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change”, Cambridge University Press, Cambridge, UK, 2013.

Hirata, E., Giannotti, M. A., Larocca, A. P. C., and Quintanilha, J. A.: Flooding and inundation collaborative mapping - use of the Crowdmap/Ushahidi platform in the city of Sao Paulo, Brazil, J. Flood Risk Manag., 11 (Supplement 1), S98–S109, doi:10.1111/jfr3.12181, 2018.

ha formattato: Tipo di carattere: (Predefinito) Times New Roman

Hung, K.-C., Kalantari, M., and Rajabifard., A.: Methods for assessing the credibility of volunteered geographic information in flood response: a case study in Brisbane, Australia, Appl. Geogr., 68, 37–47, doi:10.1016/j.apgeog.2016.01.005, 2016.

Kellens, W., Terpstra, T., and De Maeyer, P.: Perception and communication of flood risks: A systematic review of empirical research, Risk Anal., 33/1, 24–49, doi:10.1111/j.1539-6924.2012.01844.x, 2013.

Keogh, E., and Mueen, A.: Curse of dimensionality, in: Sammut, C., and Webb, G. I. (eds.), Encyclopedia of Machine Learning and Data Mining, 314–315, Springer, doi:10.1007/978-1-4899-7687-1_192, 2017.

Johnson, C. L., and Priest. S. J.: Flood risk management in England: A changing landscape of risk responsibility?, Water Resour. Dev., 24/4, 513–525, doi:10.1080/07900620801923146, 2008.

~~Kreibich, H. et al.: Adaptation to flood risk: Results of international paired flood event studies, Earth's Future, 5/10, 953–965. doi:10.1002/2017EF000606, 2017.~~

Kreibich, H., Thaler, T., Glade, T., and Molinari, D.: Preface: Damage of natural hazards: assessment and mitigation, Nat. Hazards Earth Syst. Sci., 19, 551–554, https://doi.org/10.5194/nhess-19-551-2019, 2019.

Jovi, A., Brki, K., and Bogunovi, N.: A review of feature selection methods with applications, in: Proceedings of the 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 1200–1205, doi:10.1109/MIPRO.2015.7160458, 2015.

Il Sole24ore, Telefoni cellulari: è corsa al ribasso per le tariffe, 23-05-2017. URL: https://www.ilsole24ore.com/art/telefoni-cellulari-e-corsa-ribasso-le-tariffe-AESqTnQB

ISPRA: Soil consumption in Italy – 2019 Edition, Report SNPA, 08/19, ISBN 978-88-448-0964-5, 2019. *(in Italian)*

ISTAT: Basi territoriali e variabili censuarie, 2017. (Last accessed: 17.07.2019)

Liao, K.-H.: A theory on urban resilience to floods – A basis for alternative planning practices, Ecol. Soc., 17/4, doi:10.5751/ES-05231-170448, 2012.

Maaskant, B., Jonkman, S. N., and Bouwer, L. M.: Future risk of flooding: an analysis of changes in potential loss of life in South Holland (the Netherlands), Environ. Sci. Policy, 12/2, 157–169, doi:10.1016/j.envsci.2008.11.004, 2009.

Manfredini, F., Pucci, P., Secchi, P., Tagliolato, P., Vantini, S., and Vitelli, V.: Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the Milan urban region, in: Paganoni, A. M. and Secchi, P. (eds.), Advances in Complex Data Modeling and Computational Methods in Statistics, Contributions to Statistics, Springer, Cham, 133–147, doi:10.1007/978-3-319-11149-0__9, 2015.

Mazumdar, S., Wrigley, S., and Ciravegna, F.: Citizen science and crowdsourcing for earth observations: An analysis of stakeholder opinions on the present and future, Remote Sens., 9/1, doi:10.3390/rs9010087, 2017.

Mazzoleni, M., et al.: Exploring the influence of citizen involvement on the assimilation of crowdsourced observations: A modelling study based on the 2013 flood event in the Bacchiglione catchment (Italy), Hydrol. Earth Syst. Sc., 22/1, 391–416, doi:10.5194/hess-22-391-2018, 2018.

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A.: Assessment of economic flood damage, Nat. Hazard. Earth Sys., 10/8, 1697–1724, doi:10.5194/nhess-10-1697-2010, 2010.

ha formattato: Italiano (Italia)
ha formattato: Inglese (Regno Unito)
ha formattato: Inglese (Regno Unito)
ha formattato: Inglese (Regno Unito)
ha formattato: Inglese (Regno Unito)
ha formattato: Inglese (Regno Unito)
ha formattato: Inglese (Regno Unito)
ha formattato: Italiano (Italia)
ha formattato: Tipo di carattere: Corsivo, Italiano (Italia)
ha formattato: Italiano (Italia)

Metulini, R., Carpita, M.: The HOG-FDA Approach with Mobile Phone Data to Modeling the Dynamic of Peoples Presences in the City, in: Bini, M., Amenta, P., D'Ambra, A. and Camminatiello, I. (Eds.), IES 2019 Innovation & Society - Statistical Evaluation Systems at 360: Techniques, Technologies and new Frontiers Book of Abstracts, Cuzzolin Editing, 2019a.

Metulini, R., Carpita, M.: A strategy for the matching of mobile phone signals with census data, in: Arbia, G., Peluso, S., Pini, A., and Rivellini, G. (Eds.), SIS 2019 Smart Statistics for Smart Applications Book of Short Papers, 427–434, Pearson Publishing, 2019b.

Metulini, R., Carpita, M.: A Spatio-Temporal Indicator for City Users based on Mobile Phone Signals and Administrative Data, Social Indicator Research, doi: 10.1007/s11205-020-02355-2, 2020.

~~Munich RE: Risks from floods, storm surges and flash floods, underestimated natural hazards, https://www.munichre.com/en/risks/natural-disasters-losses-are-trending-upwards/floods-and-flash-floods-underestimated-natural-hazards.html, accessed in June 2020.~~

Poser, K., and Dransch, D.: Volunteered geographical information for disasters management with application to rapid flood damage estimation, Geomatica, 64/1, 89–98, doi:10.5623/geomat-2010-0008, 2010.

Rojas, R., Feyen, L., and Watkiss, P.: Climate change and river floods in the European Union: socio-economic consequences and the costs and benefits of adaptation, Global Environ. Change 23/6, 1737–1751, doi:10.1016/j.gloenvcha.2013.08.006, 2013.

Rosser, J. F., Leibovici, D. G., and Jackson, M. J.: Rapid flood inundation mapping using social media, remote sensing and topographic data, Nat. Hazards, 87/1, 103–120, doi:10.1007/s11069-017-2755-0, 2017.

Salhi, A. I., Kardouchi, M., and Belacel, N.: Histograms of fuzzy oriented gradients for face recognition. In: 2013 International conference on computer applications technology (ICCAT), 1–5, IEEE, 2013.

Secchi, P., Vantini, S., and Vitelli, V.: Analysis of spatio-temporal mobile phone data: A case study in the metropolitan area of Milan, Stat. Method. Appl., 24(2), 279–300, doi:10.1007/s10260-014-0294-3, 2015.

Solomon, S., et al.: AR4 Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK, 2007.

Schumann, A.: Flood safety versus remaining risks - options and limitations of probabilistic concepts in flood management, Water Resour. Manag., 31(10), 3131–3145, doi:10.1007/s11269-017-1700-z, 2017.

Sun, Y., and Genton, M. G.: Functional boxplots, J. Comput. Graph. Stat., 20(2), 316–334, doi:10.1198/jcgs.2011.09224, 2011.

Sun, Y., and Genton, M. G.: Adjusted functional boxplots for spatiotemporal data visualization and outlier detection, Environmetrics, 23(1), 54–64, doi:10.1002/env.1136, 2012.

Székely, G. J., and Rizzo, M. L.: The distance correlation t-test of independence in high dimension, J. Multivariate Anal., 117, 193–213, doi:10.1016/j.jmva.2013.02.012, 2013.

Te Linde, A. H., Bubeck, P., Dekkers, J. E. C., De Moel, H., and Aerts, J. C. J. H.: Future flood risk estimates along the river Rhine, Nat. Hazard. Earth Sys., 11/2, 459–473, doi:10.5194/nhess-11-459-2011, 2011.

Tomasi, C.: Histograms of oriented gradients, Computer Vision Sampler, 1–6, 2012.

UN DESA Population Division: World urbanization prospects: The 2011 revision, United Nations Department of Social Affairs (UN DESA) Population Division, New York, USA, 2012.

UN ISDR: 2009 UNISDR terminology on disaster risk reduction, United Nations International Strategy for Disaster Reduction (UN ISDR), Geneva, CH, 2009.

Viglione, A., Di Baldassarre, G., Brandimarte, L., Kuil, L., Carr, G., Salinas, J. L., Scolobig, A., and ~~. . .~~ Blöschl, G.: ~~. (2014).~~ Insights from socio-hydrology modelling on dealing with flood risk - roles of collective memory, risk-taking attitude and trust. *Journal of Hydrology*, *518*(PA), 71-82, doi:10.1016/j.jhydrol.2014.01.018, 2014. ~~. doi:10.1016/j.jhydrol.2014.01.018~~

Vorogushyn, S., Merz, B., and Apel, H.: Development of dike fragility curves for piping and micro-instability breach mechanisms, Nat. Hazard Earth Syst. Sci., 9/4, 1383–1401, https://doi.org/10.5194/nhess-9-1383-2009, 2009.

Vu, T. T., and Ranzi R.: Flood risk assessment and coping capacity of floods in central Vietnam. Journal of Hydro-Environment Research, 14, 44-60. doi:10.1016/j.jher.2016.06.001, 2017.

Zanini, P., Shen, H., and Truong, Y.: Understanding resident mobility in Milan through independent component analysis of Telecom Italia mobile usage data, Ann. Appl. Stat., 10(2), 812–833, doi:10.1214/16-AOAS913, 2016.

ha formattato: Inglese (Regno Unito)

ha formattato: Inglese (Regno Unito)

ha formattato: Tipo di carattere: (Predefinito) Times New Roman

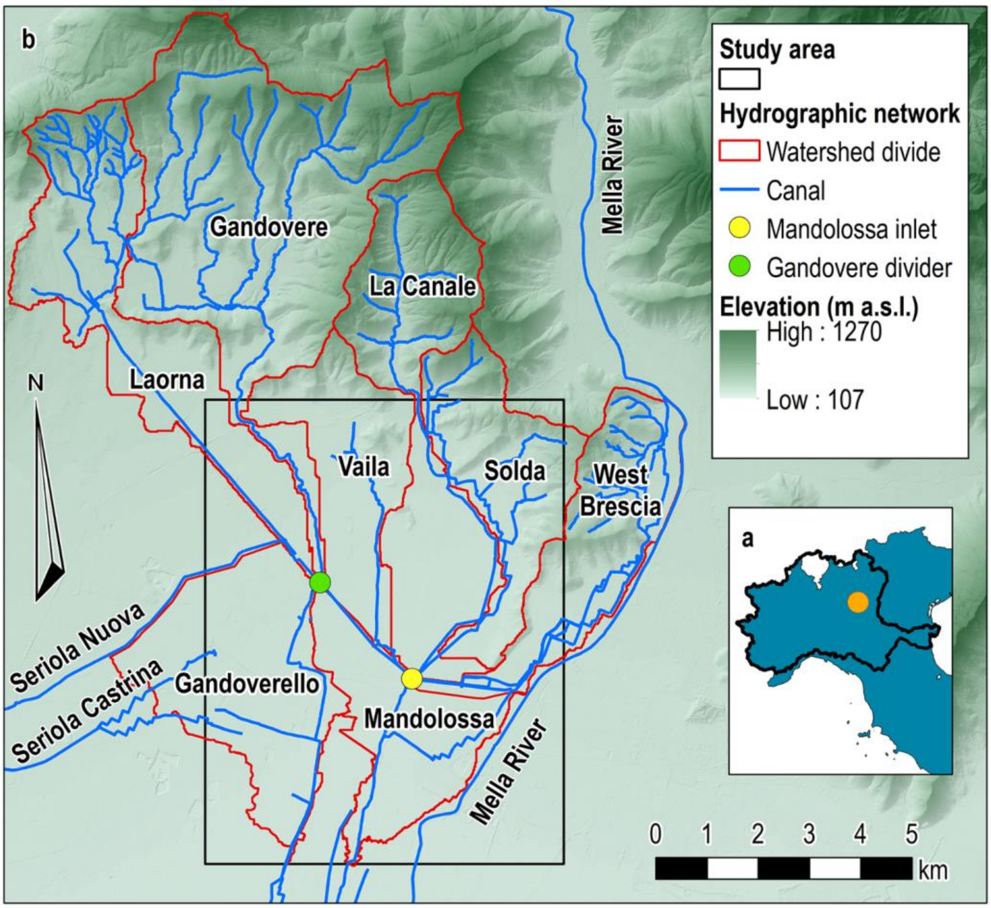

Figure 1. Location of the study area with respect to the Po River Valley in northern Italy *(a)*, and main hydrographic features of the foothill area west of Brescia town *(b)*; base map 5 m Digital Elevation Model provided by Lombardy Region (www.geoportale.regione.lombardia.it).

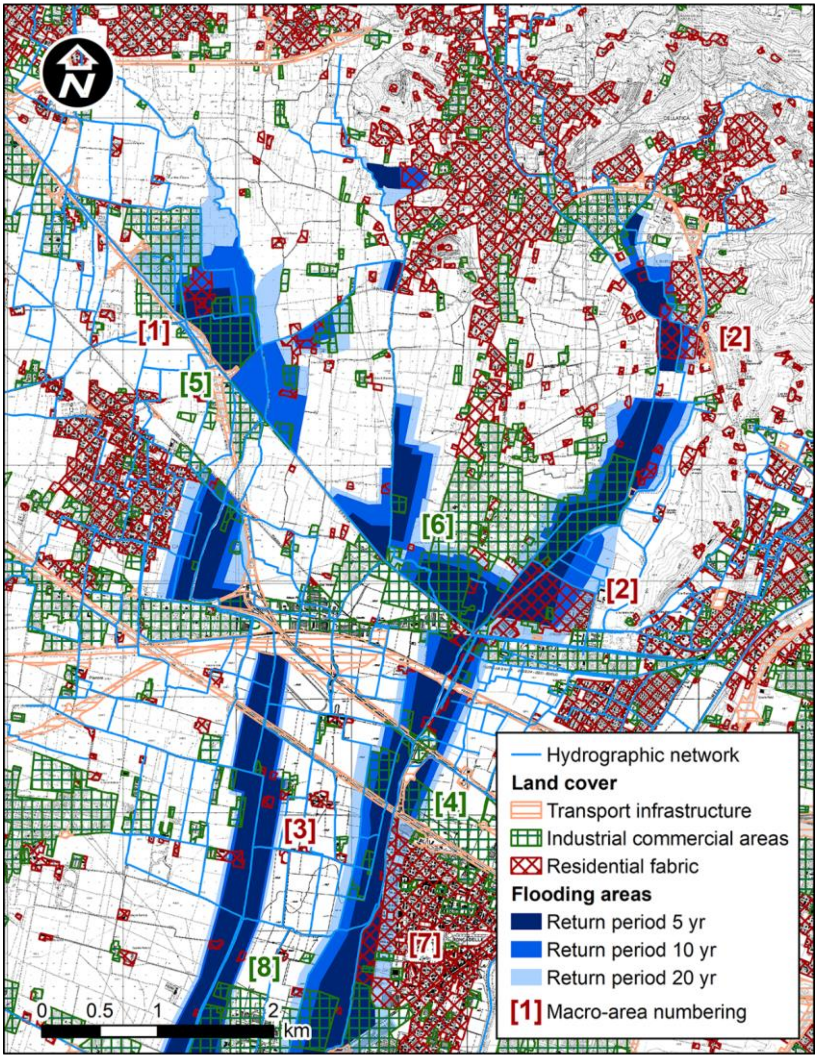

Figure 2. Flooding hazard map of the study area comparing present land cover and flooding area extensions referred to return periods varying between 5 years and 20 years; exposed residential areas are: [1] Laorna and Gandovere streams, [2] La Canale and Solda streams, [3] Southern Gandovere canal, [7] Mandolossa canal, and exposed industrial and commercial settlements are: [4] Mandolossa canal, [5] Laorna and Gandovere streams, [6] La Canale and Solda streams, [8] Southern Gandovere canal; base map Lombardy Regional Technical Map CTR 1:5000 provided by Lombardy Region (www.geoportale.regione.lombardia.it).

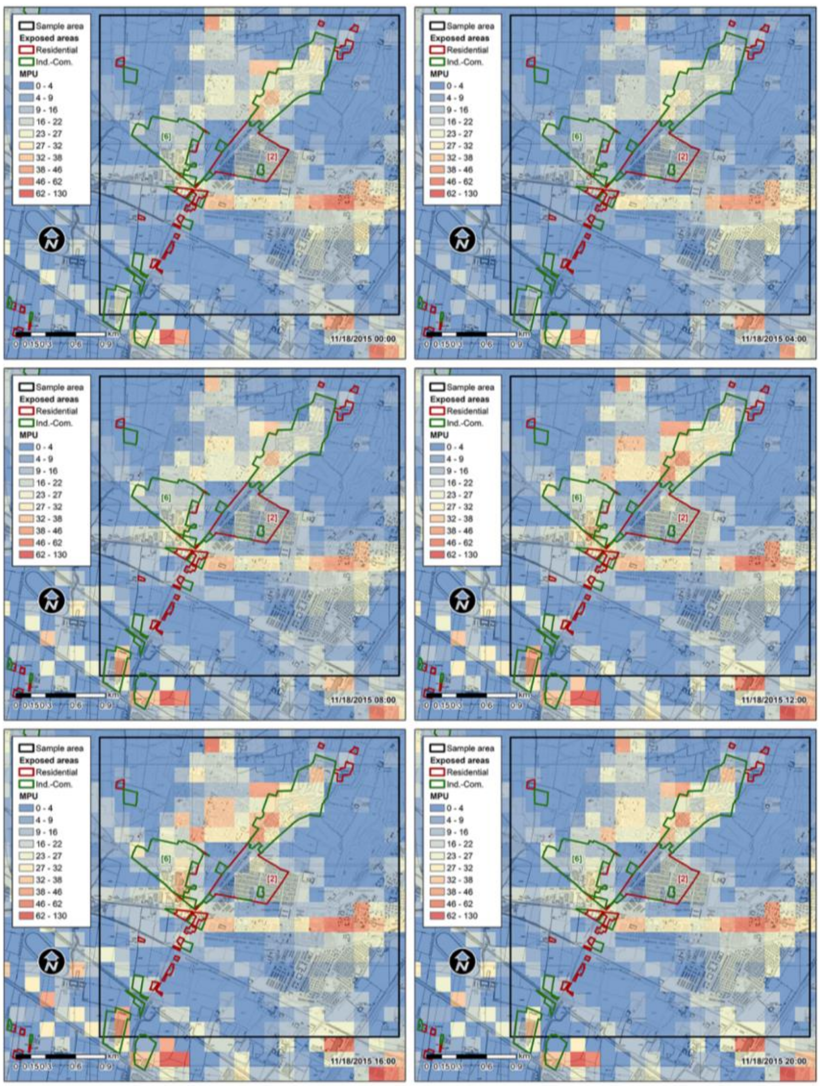

Figure 3. Snapshots of a dynamic map showing the spatiotemporal distribution of mobile phone users (MPU) occurred on 18/11/2015 (Wednesday) in urban areas exposed to 10 year return period floodings; base map Lombardy Regional Technical Map CTR 1:5000 provided by Lombardy Region (www.geoportale.regione.lombardia.it).

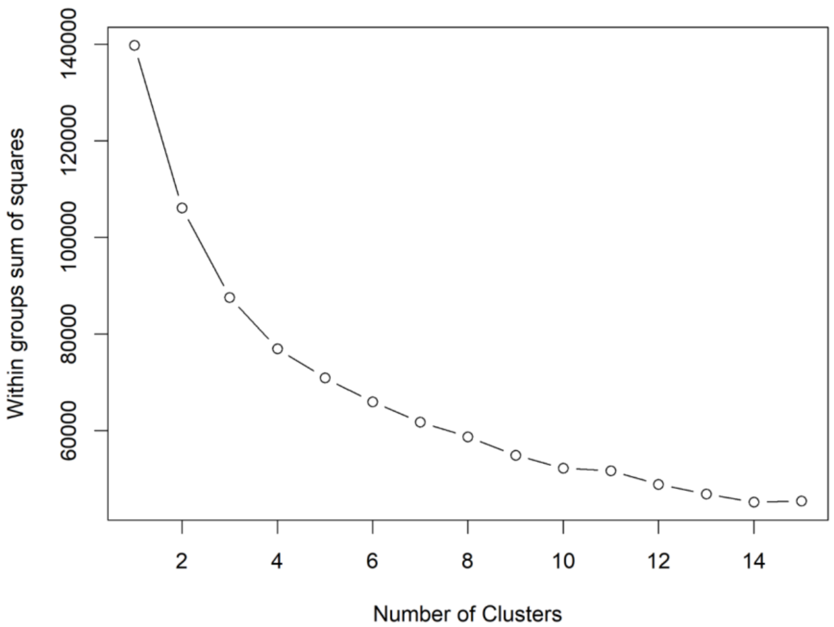

Figure 4. Diagnostic for the choice of the number of first-step clusters based on the within groups sum of squares.

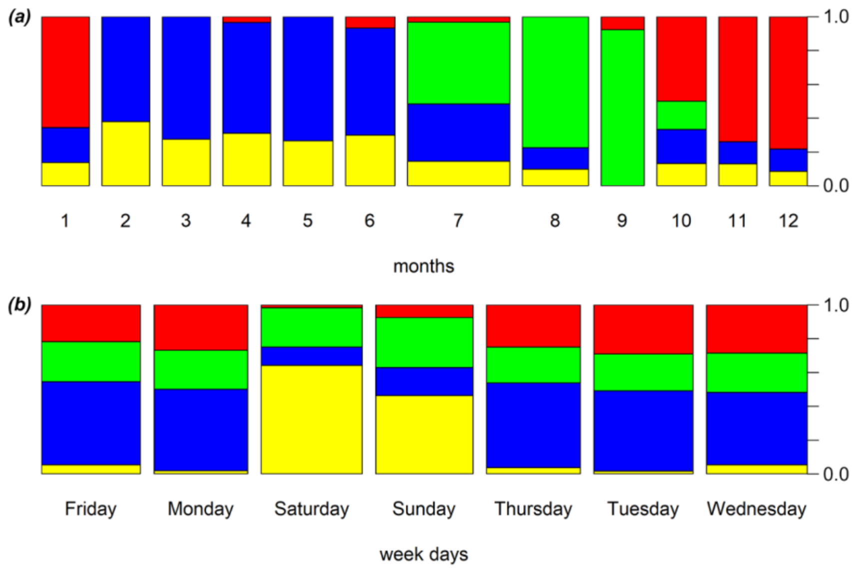

Figure 5. Spine-plots representing the first-step clustering of days along *(a)* months ~~days of the week~~ and *(b)* days of the week ~~months~~ (green: all days ~~from~~ mostly occurring in July, August and ~~to~~ September; blue: working days mostly occurring from February to June; red: working days mostly occurring from October to January; yellow: weekends mostly occurring from October to June).

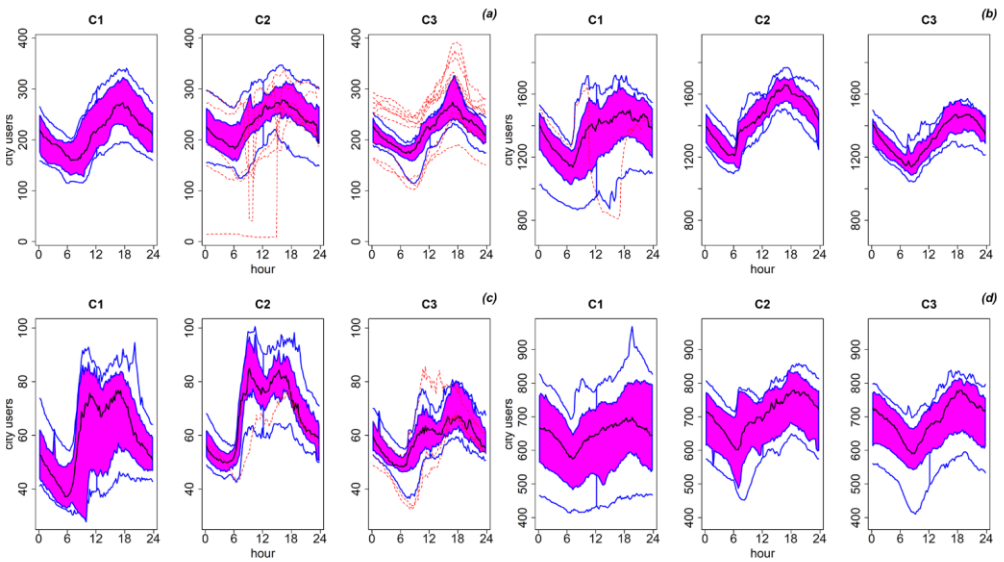

Figure 6. Functional box plots of exposed people ("city users") inside residential areas: *(a)* Moie di Sotto (area 1 in Figure 2), *(b)* Villaggio Badia and Fantasina (area 2 in Figure 2), *(c)* southern Gandovere canal (area 3 in Figure 2), *(d)* Roncadelle (area 7 in Figure 2). Cluster 1 (July, August, September, C1), Cluster 2 (working-days from October to June, C2), Cluster 3 (~~week-ends~~weekends from October to June, C3).

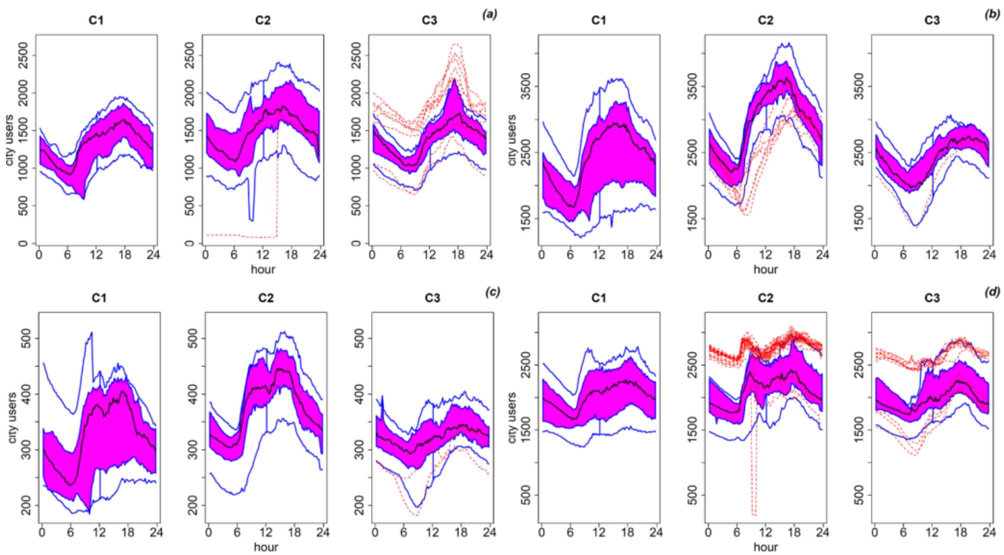

Figure 7. Functional box plots of exposed people (“city users”) inside industrial-commercial settlements: *(a)* Moie di Sotto (area 5 in Figure 2), *(b)* Villaggio Badia and Fantasina (area 6 in Figure 2), *(c)* southern Gandovere canal (area 8 in Figure 2), *(d)* Roncadelle (area 4 in Figure 2). Cluster 1 (July, August, September, C1), Cluster 2 (working-days from October to June, C2), Cluster 3 (~~week-ends~~weekends from October to June, C3).