Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-201-AC2, 2020

Natural Hazards and Earth System Sciences
Discussions
Open Access
EGU

Interactive comment on "Dynamic maps of people exposure to floods based on mobile phone data" *by* Matteo Balistrocchi et al.

Matteo Balistrocchi et al.

m.balis@unimore.it

Received and published: 21 October 2020

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 at maximizing 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,

Printer-friendly version

Discussion paper

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 155-157). Some details on this aspect and the following reference have been added to the manuscript: 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 chosen Bouveyron and Come (2015)'s model-based functional data clustering method because its better flexibility compared to alternative methods. As already reported in lines 167âĂŤ171 (first submission) 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 171âĂŤ173, 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 174-181).

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 225-229.

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 increases, 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.

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?

A4: 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.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-201, 2020.

Printer-friendly version

Discussion paper