RESPONSE TO REVIEWER #1

Paper: Towards assimilation of crowdsourced observations for different levels of citizen engagement: the flood event of 2013 in the Bacchiglione catchment

M. Mazzoleni, V.J. Cortes Arevalo, U. Wehn, L. Alfonso, D. Norbiato, M. Monego, M. Ferri and D.P. Solomatine

General comment

The paper aims at assessing the usefulness of assimilating crowdsourced observations for improving model-based predictions of flood events, by distinguishing the contribution from static physical sensors from the one derived from either static or dynamic social sensors. Each family of sensors is characterized by a different level of reliability and time of availability. The application of the analysis using hypothetical data to the extreme flood event in the Bacchiglione catchment on May 2013 (when no real crowdsourced observations were available) show a good potential for including this novel type of information in flood control applications. The manuscript is well written and the topic is definitely interesting. However, I suggest the paper needs a major revision to clarify the specific points discussed below.

We appreciate reviewer's compliments and suggestions, which were extremely helpful to improve the manuscript. Additional analyses have been included to assess theoretical scenarios on the effect of smartphone penetration and citizen engagement in urban areas. In addition, some parts have been removed to improve the structure and highlight the most innovative aspects of the manuscript. In the updated version, more focus is given on the citizen engagement effects on DA. The number of scenarios in Experiment 2 are reduced from 6 to 3.

Specific comments

RC 1: *Literature Framework: despite the literature about crowdsourcing information and citizens science is relatively new, I believe that the authors should improve the manuscript's introduction to better frame their work within the existing approaches, which are currently only listed in section 2. In my opinion this is key for clarifying the novelty with respect to previous publications by the same authors, particularly the paper "Can assimilation of crowdsourced streamflow observations in hydrological modelling improve flood prediction?" which also has the same case study application but, more generally, with respect to the entire series Mazzoleni et al. (2015a; 2015b; 2016). In addition, such improved analysis of the literature allows reinforcing the value of the results (obtained with hypothetical data) with respect to the few existing applications run over real crowdsourced observations. Practically, I would suggest re-structuring sections 1 and 2 with the purpose of reviewing the existing approaches and of clarifying how the current paper represents a step-forward with respect to other works.*

AC 1: We thank the reviewer for this critical comment. Indeed, we realized that sections 1 and 2 should be re-structured, and in the revised version of the manuscript we have done that. On the one hand, section 1 (introduction) in the revised version focuses now on the problem definition, the description and limitations of previous studies that include crowdsourced observations in flood modelling and the novelty of the manuscript with respect to the previous one. On the other hand, section 2 now aims at describing theoretical considerations of crowdsourced observations and citizen engagement as basis for the engagement scenarios. We believe that the manuscript now reads better due to the reviewer's comments.

RC 2: *Given the focus on the use of crowdsourced observations, part of the results' discussion (e.g., the analysis on the lead time vs location) is relatively basic and would apply to any type of sensor available along the river. I'm not impressed by the fact that observations far from the outlet sections allow increasing the lead-time. I would hence suggest the authors to consider shortening this discussion in favour of a more extensive analysis of pros and cons of using/relying on crowdsourced data (see point 3).*

AC 2: We thank the reviewer for this comment, which have led to several changes in the revised manuscript, as follows. First, we have removed the experiments on assimilation of observations from physical static sensors (experiment 1 in the manuscript) to favour a wider and more detailed analysis on the effect of different engagement levels in the assimilation process (experiments 2 and 3). Second, the paper have been shortened by removing some results in experiments 2 and 3 related to the influence of social sensor location on the assimilation performances. Finally, we have included a discussion section where pros and cons of using crowdsourced observations are considered.

RC 3: *A major limitation of the analysis is the lack of real crowdsourced observations. To overcome this issue, I believe the results would need a more extensive discussion about some key aspects that may strongly impact the results in case real data were available: first of all the level of public engagement is crucial and I would recommend trying to justify the theoretical formulations with respect to some preliminary findings either from the WeSenseIt project or from similar studies in the literature. I'm quite skeptical about the assumption that the 41% mobile phone penetration can be considered a good proxy for estimating a ratio of active citizens equal to 41%. In addition, I would assume this may vary spatially (even though I don't know whether it could be higher in cities or in the rural areas). Moreover, the distinction of the different behaviours seems also quite theoretical and should be somewhat mapped to the specific case study. Finally, it is not clear how many observations are assumed to be available in each experiment. Given the fast dynamics of flood events, the whole process lasts few hours and indeed the maximum lead-time is one day. This temporal dynamics may however represent a strong constraint for collecting crowdsourced observations, because active citizens might not be there at the right time. I would hence recommend discussing the upper and lower limits in terms of number of observations needed to provide an accurate flood forecast.*

AC 3: We agree with the reviewer that the model used to represent citizen engagement levels is rather theoretical, and we have clearly stated this. In the framework of WeSenseIt project we have carried out an exercise with (52) interested (engaged) citizens who were providing water level observations via the smartphone app to initiate the citizen observatory. However, no formal assessment of citizens' engagement under preparedness and emergency situations has been carried out for this case study. In relation to the Bacchiglione case study, the WeSenseIt Project focused of the analysis of the role of personal weather stations in sharing data via online amateur weather networks, but this data was not neither. Under the consideration that the Civil Protection of Vicenza was involved within the WeSenseIt project, and in reality there is a high chance that it will be happening in the future, however, for this paper it was assumed that only volunteers and/or trained volunteers will participate in providing water level observations. A further assumption is that the mobile application available for the project is easy-to-use and accessible for all participants (this in fact may increase the assumed level of engagement).

The engagement scenarios adopted in our study are assumed to represent the hypothetical situations where citizens will be fully engaged by the Alto Adriatico Water Authority within the Citizen Observatory project. CS observations are assumed to be collected at a particular location and time upon request from the Alto Adriatico Water Authority. The distinction between favourable attitudes are treated from a theoretical point of view, based on Batson et al. (2002): 1) own personal purposes (usefulness of the collected data for personal interest or direct flood risk management impact); 2) shared or community interests belonging to a community of peers with shared interested; and 3) altruism (beneficing society at large).

In order to clarify this aspect, in the revised manuscript we have included an additional sub-section where citizen engagement level in collecting water level data within the Bacchiglione catchment is described.

Regarding the assumption of 41% for the mobile phone penetration, we have considered this percentage based on Statistica (2016). This value depends on the geographic area and on the characteristic of the population. In fact, we assumed that not everyone will be willing to use smartphone to collect and share water level data due to for example their age and habits. More exhaustive analysis should be performed in order to better define the percentage of active citizens. We agree with the reviewer on the fact that the percentage of active citizen can change spatially. That is why we have conducted additional analysis considering higher percentage of smartphone users (80%) in areas with higher population density, i.e. the municipality of Vicenza. The results of such analysis are reported below (figure 1), and included in the updated version of the manuscript.

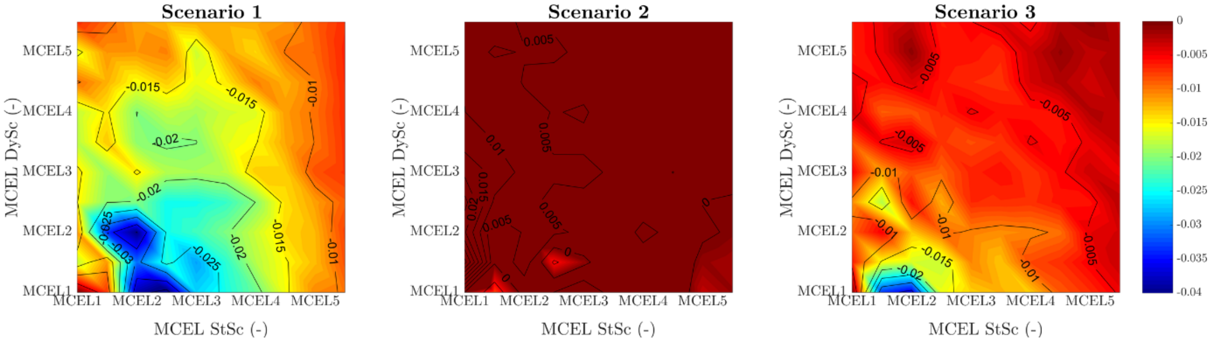

Figure 1. Difference between μ(NSE) values obtained considering standard and higher active citizen percentage in the municipality of Vicenza for different engagement levels from StSc and DySc

In figure 1, the difference between the $\mu(N_{SE})$ values obtained assuming smartphone penetration of 41% and 80% are represented. From the previous graphs it can be seen that increasing the smartphone penetration in Vicenza does not affect model results in case of scenario 2. In fact, for this scenario, no engagement is assumed in densely urbanized areas such as the municipality of Vicenza. On the other hand, higher number of smartphones in Vicenza affects (partially) only scenarios 1 and 3. In these scenarios we can observe an expected increment in the model performance due to the higher engagement in Vicenza. However, small increments in the NSE values are reported in figure 1, with a maximum difference of 0.04 between normal and higher smartphone penetration. Even in case of the 41% of smartphone usage, high engagement level is achieved for small number of active citizen (see figure 3 of the manuscript). This means that in the proposed theoretical engagement model more active citizens (more mobiles available) won't significantly improve engagement and affect the model performance. In addition, regardless of the number of CS observations if the quality is not good enough the model performance will not improve further improve as described in Mazzoleni et al. (2017). In the revised version, we have included these considerations in the discussions of the results.

The number of observations used in each experiment varies based on a particular engagement level scenario used. In fact, considering a 48h flood event and an hourly model time step, an engagement equal to 1 corresponds to 48 available observations. An engagement of 0.5 implies 24 observations (randomly distributed in time and space) to assimilate. As the reviewer correctly pointed out, active citizens might not be always available during a flood event. A limitation of our study is that we did not consider the fact that citizen may not provide observations for instance during night hours, as done in an earlier paper by Mazzoleni et al. (2015). This limitation will be mentioned in the concluding section of the updated manuscript.

Finally, it is difficult to recommend a defined number of observations needed to achieve an accurate flood forecast. In fact, as mentioned in Mazzoleni et al. (2017), it is not possible to define a-priori the upper and lower limits of crowdsourced data needed to improve a model because it depends on the uncertainty level of CS observations (which depends on the training level and experience of citizens).

RC 4: *The last point regards the need of discussing two additional key aspects of crowdsourcing (or more in general citizen science) experiments: how to stimulate citizens engagement and how to keep them engaged in the long term. I understand the authors are assuming a kind of self-motivated behaviour differentiated according to the level of engagement. However, in the final discussion, I would suggest the authors to comment about possible techniques for motivating citizens in participating to this data collection experiment and increasing their engagement level (e.g., gamification techniques). In addition, it would be nice discussing also about the potential evolution in time of such engagement as many studies observed decreasing levels of engagement in time. How this would affect the overall flood forecasting system? Assuming it is possible to have a good level of engagement in a critical event, how many citizens are expected to remain active until the next flood? Given the case study analysed in the paper where floods are not frequent, in my opinion this point is critical as I see a high probability of having a lot of people potentially involved just after a catastrophic flood event who will lose interest in time and may not be active anymore at the next flood event.*

AC 4: We agree with the reviewer's comment on the importance of stimulating citizens' engagement for a long period of time. Many studies reported a temporal pattern of citizen participation driven by self-directed motivations and person's interests. Engagement during flood event tends to disappear if no other event will occur in a short time. In fact, depending on the memory of the community, the awareness of flood risk decays with time, and, therefore, the tendency to be engaged in data collection will also tend to reduce or even disappear. As the reviewer stated correctly, many citizens may be potentially involved just after a catastrophic flood event but they might not be active anymore at the next flood event because they lost interest in time. A possible solution for collecting water level data over time could be the involvement of staff of the civil protection agency which act as (trained) volunteers. This approach is currently being used in the Bacchiglione catchment by the Alto Adriatico Water Authority, which requests the water level data at particular location and time from the Civil Protection to validate model results in near real-time.

As our study is based on theoretical considerations, we simply acknowledged in the discussion the importance to keep citizens engaged. We suggest, for example, gaming approaches or periodic meetings/seminars with interested participants. A more critical and detailed analysis of citizen engagement motivations is reported in Geoghegan et al. (2016).

As the reviewer mentioned, in this study the citizen engagement is assumed constant in time because only one flood event is considered. However, when multiple flood events are simulated, some model has to be used to represent the possible decay of engagement level on time. A possible way to represent such decay is to use a logistic curve and to vary the value of growth rate r over time. Figure 2 (which will be included in the revised version of the manuscript) presents sensitivity analysis of model results with respect to the varying values of the coefficient r to representing the varying engagement levels over time (see in Eq.15). Only engagement scenario 3, for three different values of w, is considered. The results demonstrated that decreasing engagement over time (low values of r) will lead to a reduction of the model performance, and consequently influence flood prediction. This is somehow an expected result that, once more, demonstrates the importance of keeping citizens engaged not only for a short period of time but on the long run. However, such reduction of model performances is significant only for values of *r* lower than 0.3, leading to the conclusion that model performances can still be high even if engagement reduces over time up to a given threshold value. Additional relevant literature about engagement strategies are included in the updated version of the manuscript. The ongoing H2020 project GroundTruth2.0 is trying to answer these questions and further research will take those outcomes into account.

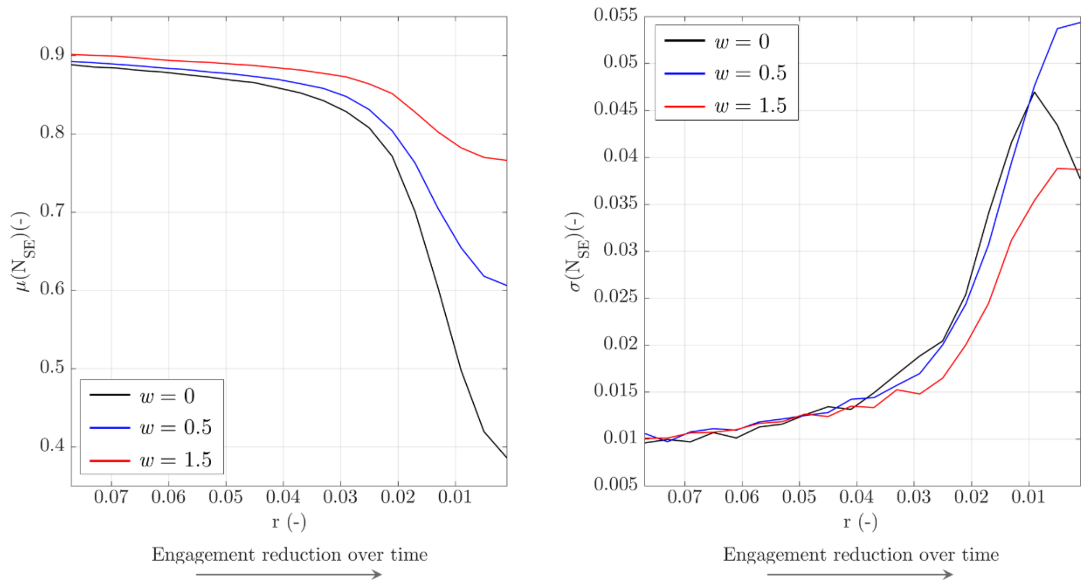

Figure 12. μ(NSE) and σ(NSE) values obtained considering varying values of the coefficient r for scenarios 1 and 3 with three different values of w

Minor points

- *There is a quite intensive use of acronyms. I would suggest - if possible - to reduce it and to add a table of acronyms to help the readers*

As suggested by the reviewer, we have included a table with all the acronyms used in this study.

AAWA	Alto Adriatico Water Authority
CEL	Citizen Engagement Level
CS	Crowdsourced
DySc	Dynamic Social
KF	Kalman filter
MCEL	Maximum Citizen Engagement Level
PA	Ponte degli Angeli
StPh	Static Physical
StSc	Static Social
WL	Water level
WSI	WeSenseIt

- *Page 3, lines 12-13: soil moisture (from AMSR-E) is repeated duplicated*

We have reduced the introductory section and removed the duplicated term

- *Page 3, lines 25-27 / Page 4, lines 14-15: the classification of behaviours from Bonney et al. is duplicated*

We thank the reviewer for spotting this mistake. We have removed the duplicated term

- *Page 9, lines 2-3: why the model does not depend on temperature? how evapotranspiration is estimated?*

In order to shorten the manuscript, we did not provide many details about the hydrological model. For this reason, we referred to Ferri et al (2012) and Mazzoleni et al. (2107) for the interested readers. The temperature is used for the estimation of the real evapotranspiration, which is calculated using the formulation of Hargreaves and Samani (1985)

Reference: Hargreaves, G.H., and Z.A. Samani. 1985. Reference crop evapotranspiration from temperature. Applied Engrg. in Agric. 1:96-99.

- *Page 23, line19: I assume there is an extra N in "allows to achieve higher N N_{SE}"*

Thank you for the comment. We have removed the additional N

- *Page 25, line 19: sigma(NSE) is never defined. I assume it is the standard deviation across the 100 experiments, but this must be explicitly stated*.

We clearly defined sigma(NSE) in the updated version of the manuscript

Reference:

Geoghegan, H., Dyke, A., Pateman, R., West, S. & Everett, G. (2016) Understanding motivations for citizen science. Final report on behalf of UKEOF, University of Reading, Stockholm Environment Institute (University of York) and University of the West of England.

Gharesifard, Mohammad, Uta Wehn, and Pieter van der Zaag
2017 Towards Benchmarking Citizen Observatories: Features and Functioning of Online Amateur Weather Networks. Journal of Environmental Management 193: 381–393.