RESPONSE TO REVIEWER #2

General comment

This manuscript is a useful contribution to the field of hydrology and flood modelling. It also complements similar papers recently published which have used real crowdsourced/citizen science observations to support real hydrological applications. It is evident that the manuscript has been reworked considerably since the original submission and has addressed the reviewers' comments where necessary, which in turn has strengthened the quality of work. However, I still think it is confusing in places, particularly the methods section, and how the synthetic and real-time data were actually generated. I also think that a lot of assumptions are made (which are not always referenced or explained) without fully appreciating the complexity of engaging and involving real citizens. Furthermore, I do not think data quality is fully appreciated; for instance, physical and automatic sensors are subject to error throughout the data collection phase, and not just when level is converted to streamflow. Citizen-based observations are also subject to error in a number of places, particularly if relying on photograph submissions v's a quantitative value. The value of these particular scenarios are therefore limited. It is difficult to apply the findings to a broader scale and be used to influence operational activities. The manuscript reads well in places, but some sentences could do with being rephrased or reworked to improve fluency or clarity. The wider picture needs to be clear and reiterated throughout. The results and accompanying figures are well presented.

We appreciate the critical reviewer's comments and suggestions to strengthen our contribution. We recognize the need to make explicit the limitations and aspects for further research. We acknowledge that data quality of physical and automatic sensors is subjected to random (e.g. signal noise) and systematic errors (e.g. maintenance needs). In addition, we recognize that visual observations coming from either experts or citizen-based are subject to various errors which may be related to the observer and data collection method (expertise and photography) but also to the specific conditions at the reporting location (accessibility and visibility). We have added a section referring to the limitations of the scenarios analysed which includes recommendation for further research. We recognize that our approach cannot yet be used in operational activities. However, repeated data collection exercises will help to refine these scenarios. Description of the method used to generate synthetic data is improved. In addition, we have also restructured section 2.2 to clearly separate the general description of the sensors implemented used in the WSI project and the specific method and assumption on data accuracy used for this research. We believe that the manuscript has significantly improved and benefit from reviewer suggestions.

The updated text in the manuscript is marked using green colour. Below, you can find a point-to-point discussions of all the comments.

Specific comments (suggested additions in blue italics, text to delete in red italics):

Page 1, Title: *consider changing 'a model study based in…' to 'a modelling study based in...'*

Page 1, Title: *consider either adding '(Italy)' to the end of the title so the reader knows where Bacchiglione is, or make it more relevant to a wider audience by removing the place catchment name altogether.*

Based on reviewer comments we have change the title of the manuscript accordingly.

Page 1, Line 17: *make this clearer. 'less accurate' because non-professionals / the public are collecting valuable hydrological datasets? It needs something to describe what crowdsourcing actually is within the abstract.*

Crowdsourced water levels are recently being considered as complementary data to traditional sensors. These water levels are provided by interested (engaged) citizens with an estimated value and/or photography via smartphone application. These observations are less accurate due to the variable expertise of the observer, qualitative measurement and variable conditions at the reporting location.

Page 1, Line 19/20: *'the extreme flood event which occurred in'*

We agreed with the suggestions and have accordingly implemented them in the text.

Page 1, Line 21: *what do you mean by target point? Do you mean receptor or impact zone? It isn't a very common flood risk management term.*

Thanks to the reviewer' comment we realized it was not clear the term "target point". That is why we have replaced it with "prediction point" throughout the manuscript.

Page 1, Line 21: *'Ponte degli Angeli (Vicenza), at the outlet of the Bacchiglione catchment'*

The target point where the water levels are predicted is the outlet of the Bacchiglione. To make it explicit we now simply refer it as: By means of a semi-distributed hydrological model, flood water levels are predicted at the point of Ponte degli Angeli (Vicenza) which is the outlet of the Bacchiglione.

Page 1, Line 28/29: *upstream sub-catchment scenarios are very catchment specific. It depends on whether you have a community upstream (less likely) or nearer the outlet (more common). This will affect your results.*

We agree with the reviewer. Our study demonstrate the usefulness of integrating crowdsourced observations however the results are catchment specific and rely on the quality control procedures of crowdsourced observations. Therefore, we have accordingly mentioned it in the abstract. To specify that these results are case specific we have modified the text of the manuscript as follows:

The results demonstrate the usefulness of integrating crowdsourced observations. First, the assimilation of crowdsourced observations located at upstream points of the Bacchiglione catchment ensure high model performance for high lead time values, whereas observations at the outlet of the catchments provide good results for short lead times. Second, low quality crowdsourced observations can significantly affect model results. Third, for the theoretical

scenario of citizens motivated for their feeling of belonging to a "community of friends", flood prediction is improved when such small communities are located in the upstream portion of the Bacchiglione catchment. Growing participation of citizens motivated by personal interests, sharing hydrological observations can help improving model performance, and therefore effective communication between water authorities and citizens is encouraged. Finally, decreasing involvement over time lead to a reduction of the model performance and consequently inaccurate flood forecasts

Abstract: *would benefit from documenting more/clearer results in the abstract.*

Based on reviewer' recommendation, a few lines have been reformulated and added to clarify the results, matching the conclusions. To improve the clarity of the abstract, now results are numbered and aspects for further research are also stated.

Page 2, Line 1: *find a better term for 'proper'. What does this even mean?*

A new sentence, more related to the conclusions, has been added. The previous one was not clear and it was removed.

Page 2, Line 5: *'for example, to operate control river structures…'*

Page 2, Line 6: *'Reliable and accurate streamflow simulation…'*

Page 2, Line 7: *'inherently uncertain due to the: lack of reliable…'*

We agreed with the suggestions and have accordingly implemented them in the text.

Page 2, Lines 7-13: *embed the list of points into the sentence better. E.g. use 'for example..'*

We thank the reviewer for this suggestion. The introduction has been restructured, among others to address this comment.

Page 2, Lines 13/14: *'Data assimilation is a common method for updating model input.'*

Page 2, Line 23: *'citizen-based data (Shanley et al…'*

Page 2, Line 29: *'In both studies, the observation filtering process…'*

The previous three suggestions have been addressed in the update version of the manuscript

Introduction: *make it clearer what crowdsourcing actually is and perhaps link them to other similar terms e.g. citizen science ad VGI. Also consider that 'usefulness' doesn't just relate to flood forecasting and real time information. Crowdsourcing can also contribute to, or generate new, long-term datasets over time, and support other types of management activities.*

We appreciate reviewers' valuable suggestion to relate our study with the diverse terms used in the scientific literature. We have made the necessary additions in the introduction

and have further clarified the way we propose to use crowdsourced observations in our study. A new paragraph with the use of similar terms has been added, and some text has been moved. See below the text that has been added.

"In parallel, the availability of recent technological advances to the public has motivated the idea of involving people in data collection. Various terms have been used for this concept in different areas (Wehn and Evers, 2015). In Natural Science this idea is known as 'citizen science' (e.g, Silvertown, 2009); in Geography, 'volunteer geographic information, VGI' (Goodchild, 2007) and 'crowdsourcing geospatial data' (Heipke, 2010), and in Computer Science 'people-centric sensing' (Campbell et al., 2006) and 'participatory sensing' (Höller et al., 2014). Other terms explicitly emphasise the involvement of the public, for instance the 'value of information and public participation' (Alfonso, 2010), 'public computing' (Anderson, 2003) and 'community data collection' (Aanensen et al., 2000)."

Page 3, Line 4: *avoid repeating the same word in the sentence ('mentioned')*

We have addressed this comment in the manuscript.

Page 3, Line 7: *how is your study real-time? It is not clear. There are reasons why real-time has not been focussed upon e.g. citizens submit their observations at a later date when they have phone signal, wifi or data to submit them.*

Our study is real time as a synthetic dataset of CS observations is assimilated at the time periods in which the flood event in the Bacchiglione occurred. However, we agreed with the reviewer that the assimilation of real time observations bring forth challenges that go beyond the availability of observers at a given location. Real time observations require to ensure safety conditions, internet connection and trusted observers by water authorities. We further elaborate these limitations in the discussion. In the introduction, we refer to our study simply assessing to the usefulness of CS observations while taking into account its variable distribution, intermittency and potentially lower quality at the time period of the flood event.

Page 3, Lines 4-8: *this text outlines the main research gap for your work (which is good) but are not clearly reflected in your abstract.*

We have rephrased the abstract to more explicitly mention these aspects. A new sentence has been added to the abstract:

Unfortunately, just few studies have investigate the benefit of assimilating crowdsourced data in hydrological and hydraulic model.

Page 3, Lines 10-13: *'To that end, wWe analyse athe flood event which occurred in May 2013 in the Bacchiglione basin (Italy) derived from a distributed network of StPh, StSc and DySc sensors. Synthetic CS observations of water level are assimilated in a cascade of hydrological*

and hydraulic models since real CS measurement are not yet available for this particular study site.'

We have incorporated suggestions accordingly and further introduced the acronyms as they are being referred for first time in the introduction.

Page 3, Line 15: *Useful to include a final sentence to say how your papers aim/objectives have a broader relevance.*

We thank the reviewer for this good idea. Introduction concludes now with a sentence that makes broader our contribution while acknowledging the aspects for further research that are also relevant for other fields:

"The achievement of the paper's objective is a step forward in understanding the effect of public involvement on the possible improvement of physical models, with methods that can be replicated in other fields"

Page 3, Line 18: *Useful to include a link to the WeSenseIt project. Don't assume the reader knows what this is.*

We have include the link to the project as suggested

Page 3, Lines 23-25*: The project set up a pilot – this is confusing. Makes it sound like citizens were actually involved. If they weren't, who was?*

As suggested, we have rearranged the description of the pilot setup for clarity. We further refer to the section 2.3 that describes the progress and limitations of the citizen involvement in the Bacchiglione catchment.

Page 3, Line 25: *'usefulness of assimilating CS WL observations or WL to improve the model performance and consequently flood prediction'.*

Page 3, Line 26: *'Northern East' should be written as 'North Eastern'. But I would move this to Line 18, when the catchment is first introduced 'The Bacchiglione catchment (North East Italy)'.*

Page 3, Line 27/28: *'river length of about 50km'. Use approximately instead of about.*

Page 3, Line 28: *change left side / right side to east? West?*

The four previous suggestions have been included in the updated version of the manuscript

Page 4, Line 1: *Forecasted and measured precipitation time series – were these subject to quality assurance and control checks? Are they of a high quality?*

We thank the reviewer for pointing out this misunderstanding. Forecasted and measured precipitation time series are available for a flood event that occurred in May 2013. In

particular, the forecasted precipitation time series is provided by the Cosmo-LAMI model, a regional model that provides numerical prediction over the national territory at 7 km resolution and three-day time interval. Currently, AAWA is performing quality control on the forecasted data before using them in the Bacchiglione flood early warning system. On the other hand, the measured precipitations are supplied and validated by Veneto Regional Agency of Environmental Prevention and Protection (ARPAV). We have included this text in the updated manuscript.

Section 2.1: *Do you need to refer to a location map within this section (i.e. Fig 1)?*

We have include the reference to Figure 1.

Page 4, Line 6: *'Three types of sensors used to measure WL, static physical (StPh), static social (StSc) and'*

The text has been modified.

Page 4, Line 9/10: *Any quality control checks for the StPh traditional sensors? Sensors are still subject to error. Why assume? Why not prove this?*

The measured river water levels at Ponte Angeli are provided and validated by ARPAV. The levels are measured with an ultrasonic transducer that compensates the echo received basing on the air temperature measured by the integrated thermometer. Certainly the measured levels, even if validated, are associated with an instrumental error. Typically the instrument precision is about 0.01 m.

On the other hand, the conversion from rating curves is subject to significant uncertainties due to the poor reliability of the steady state hypothesis, for the variability of the roughness of the river bed, for periodic variations of the geometry of the river cross section and for extrapolation-induced errors (may be of the order of 20%).

We have reported the quality control check and instrumental precision in the manuscript. Considerations regarding high uncertainty in the rating curve estimation have also been included.

Page 4, Line 13/14: *Is the mobile app used to submit photos, videos and/or quantitative values? Are date, time and location also submitted (i.e. metadata)? Any data quality checks anticipated/required? The app and use of QR codes is very specific and difficult to synthetically generate.*

Concerning the QR code function, the mobile app does not allow to send photos and videos but only the quantitative value of the river level observed at a specific staff gauge. This value is submitted in association with date/time. The location is incorporated in the QR code. The WSI mobile app is equipped with a filter that automatically discards the river level measurements that fall outside the range associate to the staff gauge. We have also restructured section 2.2 to clearly separate the general description of the sensors

implemented used in the WSI project and the specific method and assumption on data accuracy used for this research.

Page 4, Line 23: *'We assume a direct relationship…'.*

Page 4, Line 24: *'i.e. the probability of receiving a CS observations.'*

We have included these changes in the text

Page 4, Lines 16-18 / Page 9, Line 13: *estimating velocity and runoff induces significant uncertainty and defeats the object of involving citizens in a cost-effective and simple way. Is it worth the effort if additional data is required or has to be derived? It is unlikely that rating curves would be available in reality. Some studies are extracting velocities, levels and discharge from videos and photographs automatically using image analysis techniques.*

We agreed with the reviewer on the increasing uncertainty of estimating velocity and runoff with dynamic sensors. In addition, it is extremely difficult to have rating curve information at any random location of the DySc sensors. That is why, a Manning equation for rectangular channel and given channel roughness can be used to derive river flow. However, this approach will introduce significant uncertainty. A possible solution is the use of mobile apps able to retrieve flow observations as described by Luthi et al. (2014). We believe that this type of mobile app will increasingly become available (at reasonable low costs) to citizen in order to easily measure river flow. We have included these considerations in the updated version of the manuscript. We extended the description of Michelsen et al (2016) as example from the literature to overcome these limitations.

Page 4, Lines 31/32: *CS activities are not yet operational but this page describes these activities. This makes it confusing to follow. It is not clear how/if synthetic data is used.*

We agree with the reviewer that this part may look confusing. Currently, only StPh sensors are used by AAWA to provide daily flood forecast in the Bacchiglione catchment. Despites that CS observations were not operational nor available in the case study for the flood event of 2013, we analysed the (potential) characteristics of each sensor to generate the synthetic data used in this study. We have added this clarification at the beginning of section 2.2. We have also removed the reference to the synthetic observations.

Page 5, Table 1: *How are the photos used? Who extracts the information? Social observations can come in a variety of formats, and is often one of the biggest challenges/barriers when involving citizens. How would this be managed in practice?*

Social observations can come in a variety of formats, and it is often one of the biggest challenges when involving citizens. Photos can be automatically analysed using image recognition methods as proposed by van Overloop and Vierstra (2015) and Le Boursicaud et al. (2015), in which a reference gauge must be available. For the WSI project, Quick

Response codes (QR) have been added to gauges with automatic water level sensors, which are read in combination with a dedicated mobile application. No pattern recognition for water level photos is implemented in the WSI mobile app. We have included these considerations in the updated version of the manuscript.

Page 5, Table 1: *Why is StPh regarded as a CS method here? It is automatic and generates the data for you.*

Table 1 refers to all type of observations and not only CS. We agreed with the reviewer and have explicitly described StPh as one of the sensor types, which is not a CS type due to its automatic method for data collection. To clarify this point we have also modified the table caption.

Page 5, Table 1: *Do you have any references to add to the observational error column? Examples do exist in the literature and data quality is important.*

We appreciated reviewer's comment and have accordingly add a column with supporting examples from the literature.

Page 5, Line 9: *'from a wide range but limited number of' – this is not clear.*

We have improved this sentence removing "a wide range but"

Page 5, Line 10/11: *due to the limited number of participants – isn't that the point? Recruitment and low participation is a huge barrier.*

The exercise that was carried out within the framework of the WeSenseIt project had the purpose of testing the infrastructure and set up for CS observations. Despites the limited number of participants, the duration and set up of the exercise did not intend to involve other citizens than the volunteers of Civil Protection. We further clarify that in this study we do not refer to the engagement process (how to get citizens involved) but rather to the probability of receiving a CS observation based on the citizen's own interest or intention in collecting water levels. We agreed that engagement and involvement level are related and represent a huge barrier to collect CS observations therefore we have accordingly referred it in the discussion.

Page 6, Line 8: *'In the case of the main river channel,'*

The sentence has been improved

Page 6, Figure 1: *it would be useful to mark on the map where the urban area of Vicenza or 'target point' is.*

The prediction location of Ponte degli Angeli (PA) corresponds to the StPh-3 sensor. We have included this information in the caption of figure 1.

Page 6, Line 14: *'Figure 1. Spatial distribution of the sub--catchments, river reaches, and StPh and StSc sensors implemented in the catchment by AAWA'*

Page 6, Line 18: *'relate to the model equation here as a detailed description is available in Ferri et al. (2012)…'*

The text has been modified including the previous suggestions

Page 6, Line 18/19: *Precipitation time series – can/have the citizens observe this too? Many examples in the literature where they have.*

Previous studies (de Vos et al., 2017; Starkey et al., 2017) have demonstrated useful CS reports from 24-hours rainfall measurements. However, these measurements are not within the scope of this study as the input precipitation for the hydrological model are hourly instead of 24-hours time series. In the current version of the WSI mobile app citizen may report observed precipitation values. However, this is out of the scope of our paper (as specified in section 2).

Page 7, Line 3: *'The tTemperature is used for the estimation'*

Text has been corrected.

Page 7, Line 7: *Information on the quality/success of the calibrated model would be useful. Do you have any statistics to validate its performance?*

The hydrological model was calibrated and validated by AAWA. Calibration results are briefly reported in Ferri et al. (2012). We have included this reference in section 3.1.1.

Page 10, Line 4: *I do not agree that rating curves are the only source of error/uncertainty. Especially when physical sensors often measure water level indirectly using temperature and pressure.*

We agreed with the reviewer that other sources of uncertainty exist in physical sensors (i.e. ultrasonic) such as temperature fluctuations and turbulence (Irrigation Training and Research Center, 1998, p. 58). However, for simplification purposes we assume that the main source of uncertainty is the rating curve. That is why we have changed "only" by "main".

Page 10, Line 14: *WL can be easily measured by citizens using a staff – this depends! Some studies have found that their ability to manually observe level using a staff can vary greatly. It*

can also depend on when it is installed, how turbulent the flow is etc. I feel as though any error associated with the citizens is completely bypassed here. It cannot be assumed that error is the same spatially, temporally and for each participant.

We agreed with the reviewer on that random error and bias or systematic errors (Bird et al., 2014) for WL observations vary temporally, spatially and for each type of sensor (physical or social). Moreover, we acknowledge that the distribution of accuracy levels for CS observations is a simplified and first approximation that is aspect for further research

Page 10, Table 2: *It would be useful to include a citation for the coefficients used in your study, within the table itself or within the table caption so it is clear when they have each come from.*

We agreed with the reviewer that the assumptions behind the distribution of accuracy levels should be clarified and it is now listed in bullet points above the Table 2.

Page 11, Lines 15-20: *What NSE value do you regard as being 'good' or 'acceptable'?*

Based on Moriasi et al. (2007), NSE values between 0.0 and 1.0 are generally considered as acceptable levels of model performance.

Page 13, Line 17: *why have you used 500m and 1000m? Why are they assumed? Citizens may travel or walk elsewhere.*

In this study, a spatial discretization of 1000m is used in order to guarantee the numerical stability of the MC model scheme (as added at the end of section 3.1). On the other hand, the value of 500m is selected based on a subjective judgment, assuming that citizens located further than 500m from the river are not contributing to the collection of CS observations. Obviously, Different extents of the buffer will lead to different coverages of the active area, with significant effects on the simulated number of involved citizens. In order to consider more complex citizen behaviours we suggested (in the conclusion section) the use of Agent Based Model as next research step.

Page 13, Line 29: *41% still seems very vague/generic in the context of data submission.*

We recognise this limitation. According to Statistica (2016), the mobile phone penetration in Italy in 2013, the year of the flood event analysed in this study, was about 41%, which means that about 41% of the population was potentially able to submit data. In view of the lack of a better source, we assume that this proportion is valid also for the regional scope. The text in the manuscript has been modified in order to explicitly include this limitation.

Page 15, Line 7: *Batson et al 2002 seems an old reference to use for such an evolving topic which is heavily dictated and driven by technology.*

These citizen involvement scenarios are based on Batson et al., (2002), whose aggregated categories of citizen's motivations are still in agreement with more comprehensive and detailed analysis such the ones recently reported in Geoghegan et al. (2016) and (Gharesifard & Wehn, 2016).

Page 17, Line 8: *Why have you used 80%?*

We have used 80% as a double percentage of active citizens is assumed in this specific analysis. Thanks to reviewer' comment we have clarified this aspect in the text

Page 18, Line 6: *'and river reaches (hydraulic model) for a 1-hour lead time.'*

Text has been corrected accordingly

Page 18, Figure 4: *would be useful to include 'NSE' on or next to the colour ramp key. And repeat for all later figures.*

As requested by the reviewer, we have included NSE on the colour bar of figures 4, 5, 7, 8, 10 and 11.

Page 27, Line 11: *'so for the assimilation of CS observations it is also important to consider also this'*

Page 28, Line 10: *'This section aims to summarise at summarizing the main findings of our study and…'*

We have corrected the text based on reviewer suggestions.

Page 28, Discussion: *there is scope to relate your findings to the literature in more detail, including those which have used real crowdsourced observations.*

We agree with the reviewer. As referred in the introduction, there have been different studies in which crowdsourced observations were used to improve model prediction. In the discussion section we underlined how the findings of our study are in accordance with recent research carried out using real crowdsourced observations. We believe that the comment of the reviewer have strengthen the discussion section and improve the quality of the manuscript.

Page 29, Line 23: *'awareness of flood risk decreases over time' – do you have a reference to back this up?*

We have included Raaijmakers et al. (2008) as reference in the paper:

Raaijmakers, R., J. Krywkow and A. van der Veen (2008). "Flood risk perceptions and spatial multi-criteria analysis: an exploratory research for hazard mitigation." Natural Hazards 46(3): 307-322

Page 29, line 28: *'Gharesifard and Wehn (2016) are and Rutten et al. (2017) and being studied in detail in the H2020 GroundTruth..'*

Page 29, Line 32: *'This study demonstrates that high performance models value of model performance can still be achieved even…'*

The text of the manuscript has been corrected accordingly

Page 30, Lines 7-9: *This text is not reflected in the abstract, despite its importance.*

We thank the reviewer for this suggestion. We have restructured abstract in order to account for these aspects.

Page 30, Line 13: *Why discuss experiment 2 here and not experiment 1?*

Thanks to the reviewer' comment we have included an additional sentence summarizing the scope of experiment 1 with respect to experiment 2

Page 30, Discussion: *what do your results/conclusions mean for the wider picture? Ensure readers can relate to your study.*

The reviewer is right. The sentence at the beginning of the conclusions has been expanded to include a reflection for the wider picture. We reckon it was needed.