Editor Decision:

Dear Authors,
I still find issues with the manuscript and agree with the comments of the Reviewer, with improvements needed in three areas: 1) editing for the English language, 2) clearer descriptions of the methods and modeling procedures, and 3) softening of the use of the crowd-source aspect of this study. I also ask that the reviews please carefully follow the HESS publication guidelines for equations and variables:
http://www.hydrology-and-earth-system-sciences.net/for_authors/manuscript_preparation.html.
I continue to believe, as does the Reviewer, that the manuscript has value and has the potential to be a worthwhile contribution. These needed revisions will alter framing of the manuscript substantially and, therefore, I am recommending again major revision to address the remaining reviewer and editor comments.
I ask the authors to now respond to these comments and revise the manuscript. The manuscript can still be rejected at this point if the comments are not accounted for the in the revised version.
I thank the authors for their continued revisions and submission to HESS and I wish them all the best on this next set of revisions.
Kind regards,
Stacey.

Dear Editor,

Thank you for all the comments and suggestions. We have corrected the manuscript based on the reviewer suggestions as much as we could, at this stage of reviews. The additional text in the updated version of the manuscript is reported using green color. We have had long discussions about these last comments of the reviewer. The three most important points we have summarized below:

1) *What data is assimilated*. We recognize in the previous version of the paper there is a lack of clarity on what data citizens would collect. We totally agree with the reviewer that to assume that citizens would collect streamflow is unrealistic, even by stating that these measurements are 10% to 30% accurate, as the 'crowd' would require having access to measurement equipment and knowledge. What we really meant (failed in describing it properly), is that citizens would collect water level data, for instance using low-cost sensors developed within the WeSenseIt project, such as the QRcode-based gauge boards (http://wesenseit.eu/?page_id=73). Then, the discharge values can be estimated using rating curves for each particular location in the rivers, with their associated uncertainty. As a consequence, to clarify these points, we have added a new section (2.2 Crowdsourced Data), where the aspects related to the differences between crowdsourced observations of discharge and water level are clearly explained. After section 2.2, we are now referring only to crowdsourced data and not observations of water levels in order to avoid any misunderstanding. The text related to the estimation of synthetic observations in Section 4 (Experimental setup) has been moved to section 2.2. In addition, we now specify that only discharge values are used in the calibration and evaluation phase and no water depth is considered. Overall, the paper has been restructured and now has 6 sections, where models, data assimilation and experimental setup have been merged into one new section (3. Methodology).

2) *Principle of crowdsourcing*. In the first round of reviews (29.01.2016) the referee No. 2 wrote that:

> *"This article presents an evaluation of methods for improving the accuracy of hydrologic models by incorporating crowdsource (social sensors) data. This is an interesting idea and the first paper on the topic that I have read. The opportunity to get the public to engage in extreme-events using technology they are already familiar with is exciting and will likely be a great success"*
> http://editor.copernicus.org/index.php/hessd-12-C6442-2016.pdf?_mdl=msover_md&_jrl=13&_lcm=oc108lcm109w&_acm=get_comm_file&_ms=32535&c=100473&salt=10035230031186881023

and this is quite encouraging for us. In terms of the Reviewer #1's opinion, we regret that the whole principle of using crowdsourced observations, which is fundamental idea of this paper, is questioned only now, after three rounds of reviews. This paper presents the results from and motivated by the EU-FP7 WeSenseIt (Citizen Observatory of Water) project, in which one of the fundamental research questions is how crowdsourcing observations can be used in water models. In our view, removing the reference to crowdsourcing will completely change the whole idea and motivation for the paper.
One can see more and more interest to crowdsource environmental data, but its use could be further extended by assimilating this data into models. Since such data is irregular in availability and variable in accuracy, it poses a number of interesting questions on how to update the existing techniques of data assimilation to make this possible. This is the main thrust of the paper and we would not like to change it. With the revisions in the manuscript we hope that our message is now clearly presented.

3) *"Overselling" the results*. First of all, the reviewer states that our study demonstrates that the new observations which are asynchronous in time and have variable accuracy can actually improve hydrological modelling. What he/she has a problem is the fact that we link this type of data to crowdsourcing. To some extent we understand the feeling that we are overselling our results just because the use of the term "crowdsourcing". This is still the main conclusion, and we still would like to link it to the fact that this type of data typically comes from crowdsourcing. At the same we have added the statement that crowdsourcing should be seen as complementary to traditional hydrometric networks.

We have also changed the title removing "streamflow" from it, rewritten the abstract, restructured the paper, and have brought a considerable number of editorial changes.

We would like to thank you and the referees again for all the efforts, and believe the manuscript has considerably benefitted from your comments.

Anonymous Referee #1:

I appreciate that the authors now included more cases. I still would have like to see more different events per catchments, but can accept the choice by the authors as a different approach.

1. *My main remaining concern is the general framing of the study using crowdsourced observations as motivation. It is assumed that citizen scientist can observe streamflow with an accuracy of 10-30%. I find this highly unrealistic. This accuracy would require the 'crowd' to have access to measurement equipment and knowledge and I just cannot see how this should work in practice.*

We thank the reviewer for the critical comment. Based on this, we realized the lack of clarity in the description of the method and in the main framework of this paper. We totally agree with the reviewer that to assume that citizens would collect streamflow is unrealistic, even by stating that this measurements are 10% to 30% accurate, as the 'crowd' would require to have access to measurement equipment and knowledge. What we really meant (failed in describing it properly), is that citizens would provide measures of hydrological variables which are easy to observe, such as precipitation or water level, for instance using low-cost sensors developed within the WeSenseIt project, such as the QRcode-based gauge boards (http://wesenseit.eu/?page_id=73).

However, in hydrological modelling it is easier to assimilate streamflow rather than water level. For this reason, crowdsourced observations of water level are used to assess crowdsourced data of streamflow by means of a rating curve. That is why, we should clearly state the difference between observations (provided by citizens) and data (calculated from the previous ones and then used inside hydrological model). In this study we assimilated only crowdsourced data and not observations. For this reason, we change the term "crowdsourced observations" in "crowdsourced data" when referred to streamflow values assimilated within the hydrological model. In order to clarify this aspect, we included a new figure (figure 2) with the representation of the ideal approach to follow to assimilate crowdsourced observations from citizens. However, due to the fact that these observations are not available at the time of this study, a synthetic approach using synthetic streamflow observations is followed. It is during the process from observations to data that the uncertainty comes from. In fact, unless flood extreme conditions or obvious measurement mistakes, water level measures have low uncertainty due to the fact that the staff gauge can assist citizens in correctly retrieve the information. Therefore, an uncertain estimation of the rating curve at the staff gauge location can reflects in high errors in the crowdsourced data assessment. That is why we assigned uncertainty three times higher than the one of physical sensors. However, as stated in the conclusions, advancing methods for a more accurate assessment of the data quality and accuracy of streamflow data derived from social sensors need to be considered.

As a consequence, to clarify these points, we have added a new section (2.2 Crowdsourced Data), where the aspects related to the differences between crowdsourced observations of discharge and water level are clearly explained. After section 2.2, we are now referring only to crowdsourced data and not observations of water levels in order to avoid any misunderstanding. The text related to the estimation of synthetic observations in Section 4 (Experimental setup) has been moved to section 2.2. In addition, we now specify that only discharge values are used in the calibration and evaluation phase and no water depth is considered. Overall, the paper has been restructured and now has 6 sections, where models, data assimilation and experimental setup have been merged in one new section (3. Methodology).

2. *The study has its value regardless of this framing, but I see a great risk that the catchy title and intro on crowdsourced observations will lead to wrong conclusions from this study. The water levels mentioned in the conclusions are obviously not sufficient unless there is rating curve (and if there is a rating curve measuring water levels would be the easy part ….).*

We appreciate reviewer comment. With the water level in the conclusions we referred to crowdsourced observations which can be used to derive crowdsourced data by means of a rating curve, as underlined by the reviewer. We went through the paper and we modified the term crowdsourced observations with crowdsourced data (CSD) in order to underline that the information that are assimilated is discharge, which is derived from citizen-observations. However, because we did not have either crowdsourced observations or data, we used synthetic streamflow data.

3. *Although the framing is likely to catch attention, I therefore would strongly recommend toning down the crowd-parts to avoid that people draw wrong conclusions from this study; I do not agree that this study "demonstrates that crowdsourced observations, asynchronous in time and with variable accuracy, can improve flood prediction if integrated in hydrological models.". If crowdsourced is deleted, I'd agree, with it I feel the authors are overselling their results.*

We thank the reviewer for his comment. In our view, removing the references to crowdsourcing will weaken the paper, as Citizen Science projects are currently expanding fast everywhere and a clear link to them in modelling is urgently needed. Therefore, the reference and link to crowdsourcing should be maintained. One can see more and more interest to crowdsource environmental data, but its use could be further extended by assimilating this data into models. Since such data is irregular in availability and variable in accuracy, it poses a number of interesting questions on how to update the existing techniques of data assimilation to make this possible. This is the main thrust of the paper and we would not like to change it. With the revisions in the manuscript we hope that our message is now clearly presented.
However, in order to toning down and not oversell the crowdsourced part we have included few sentences in the conclusion regarding the importance of a dense network of physical sensors. In fact, this study demonstrates that crowdsourced data can improve flood prediction if integrated in hydrological models. However, social-sensors should not replace existing network of physical sensors but integrate them to compensate lack of observations. Obviously, in case of a dense network of physical sensors, the additional information from social sensors might not be necessary because of the high accuracy of the hydrological observation derided by physical sensors. These considerations have been included in the updated version of the manuscript

4. *Only after quite some time I realized that the authors used (as far as I understand) simulated discharge as for calibration/evaluation. This is ok as it takes away one source of error by assuming that the model is a perfect representation of reality, but this needs to be described much clearer to avoid confusion (and a misinterpretation of the very good simulation results, where a perfect fit actually would have been possible, which is never the case in reality).*

Thanks to the reviewer comment we have clarified this aspect since the abstract and introduction section. In fact, water level it is mentioned as feasible crowdsourced observations that citizen can provide, while in hydrological modelling only discharge is used. For this reason, all calibration and evaluation parts are performed comparing observed and simulated discharges. In addition, the description of the synthetic observations is reported since the second section of the paper and not in the experiment description as it was done before.

5. *The English needs some improvement. This applies especially to some of the new (red) parts (Did the senior co-authors actually look at the revised text? Honestly I doubt this).*

Based on the reviewer comment we improved the English

6. *Equations: avoid the use of variable names such as ET or Sw. Mathematically, these mean E times T or S tmes w!*

Following reviewer and editor suggestion, we have included the HESS publication guidelines for equations and variables