Anonymous Referee #1:

This contribution deals with the question how crowdsourced observations could be utilized in hydrological modelling. The approach used here is to use such observations to update models used for forecasting runoff. The challenge is that the observations might come at irregular times and with varying accuracy. This is an interesting and timely issue and I was excited when I started reading the manuscript. In the end, however, I have to admit, I was not fully convinced, and feel that a major revision is needed.

RC: *My major is the selection of catchments/models and limited event data being used here. Two catchments were chosen (the selections seem a bit random, but ok) and two different models were used for the two catchments. The latter seems to make little sense, as it makes results hardly comparable. It should be noted that also issues like calibration largely varied: in the Brue case, calibration was based on one event only (p11381,l12), whereas in the Bacchiglione catchment ten years were used for calibration (p11383,l20). In the first case some form of effective precipitation must have been used, as only the so called direct runoff is simulated (but it is unclear how this was determined), whereas in the second case the entire runoff has been simulated.*

AC: Thank you for this comment. We realise we were not fully clear in the manuscript. The mentioned important points are now clarified in the revised manuscript. Ideally, a new method has to be tested on a variety of catchments and models, but in the presented study we were constrained by the available funds, the projects limitations and the data availability.

We have chosen these two case studies because we wanted to assess if the results we have obtained are valid for areas with different topographical and hydrometeorological features and represented by two different hydrological models.

First, the Bacchiglione basin is one the official cases studies of the WeSenseIt project which is funding this research. We had the model available from AAWA, the water authority (Ferri et al. (2012)). The developed methodology has been tested using synthetic observations, and the intention is to then apply it on the existing early warning system on the Bacchiglione Basin, previously designed and implemented by Ferri et al. (2012).

The second case, the Brue catchment, is an experimental catchment and is used in many studies, including us (Mazzoleni et al. (2015)). In addition, we specified in section 3.1 that, in case of the Brue catchment:

> “The choice of the model is based on previous studies performed on the Brue catchment in case of assimilation of streamflow observations from dynamic sensors (Mazzoleni et al., 2015)”

We agree the cases and models are indeed different, but the presented study demonstrated that the results obtained, Figure 11 and 13, are very similar in terms of model behaviour assimilating asynchronous observations. This does not allow for a universal generalisation (that the methods works in all cases), and we are not claiming this. In the present version we clearly state that:

> ...additional analyses on different case studies and larger time series of flood event should be carried out in order to provide more general conclusions.

These additional clarifications have been added as in the introduction:

> “The Brue catchment is considered because of the availability of precipitation and streamflow data, while the Bacchiglione river is one of the official case studies of the WeSenseIt Project (Huwald et al., 2013), which is funding this research”

and conclusion:

> “We agree the cases and models are indeed different, but the presented study demonstrated that the results obtained are very similar in terms of model behaviour assimilating asynchronous observations”

On the data sets used for calibration, we understand the reviewer concern regarding the short time series used to calibrate the lumped model in the Brue catchment. For this reason, an additional calibration is performed considering the period from 23-10-1994 to 17-03-1995, affected by a long series of flood events, to demonstrate that model parameters estimated in this new calibration do not differ from the ones previously estimated. In both calibration we used the correlation R as objective function of the optimization method. The result of the calibration performed in case of longer time series is showed in Figure 1. Table 1 shows the calibrated values of the parameter c_k and n obtained using two different time series

Table 1 Parameters values in case of two different time series used to calibrate the model

	n(-)	c (-)
Short time series	5	0.035
Longer time series	4	0.026

As it can be seen from Table 1, similar results are obtained in both cases for short and long time series. In particular, no difference is visible in the flood event 1 and 2 when applied the two different calibrated parameter sets. For this reason, the text in section 3.1 of the revised manuscript have been changed as:

> “The model calibration is performed maximizing the correlation between the simulated and observed value of discharge, at the outlet point of the Brue catchment, during the

flood events occurred from the 23-10-1994 to 17-03-1995. The results of such calibration provided a value of the parameters *n* and c_k equal to 4 and 0.026 respectively”

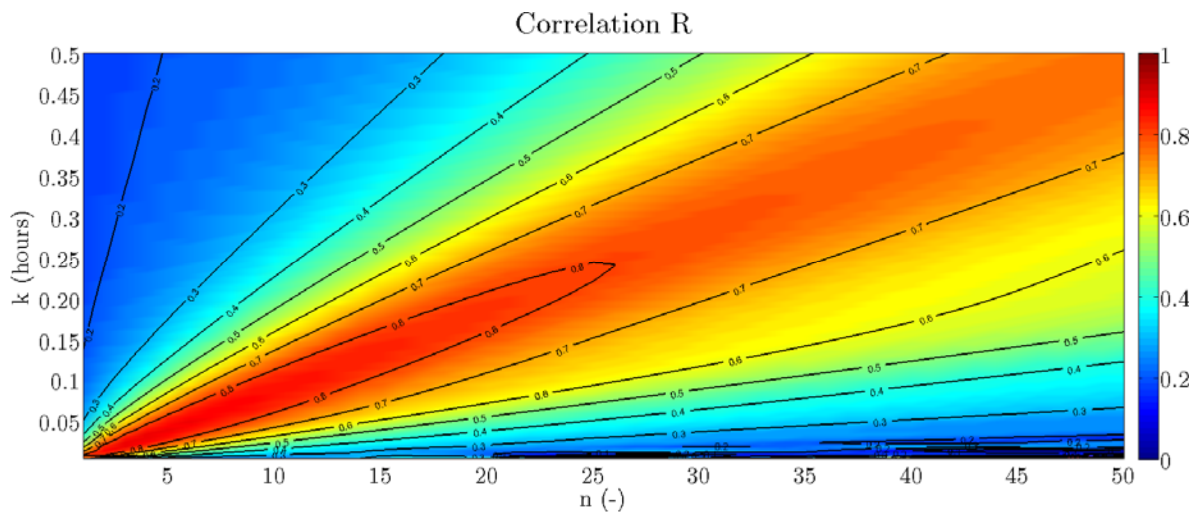

Figure 1. Calibration of Brue model in case of longer time series.

In case of the Bacchiglione catchment, as already stated, the model was already calibrated by AAWA in order to be used in the early warning system for the Bacchiglione River.

In order to clarify the input and output of the conceptual model used in the Brue catchment, the following text in the revised manuscript has been added:

> “where *I* is the model forcing (in this case direct runoff), *n* (number of storage elements) and *k* (storage capacity) are the two parameters of the model and *Q* is the model output (streamflow).”

Direct runoff is estimated using the approach used in Mazzoleni et al (2015) as referred in the text. On the other hand, in the Bacchiglione model the input is the time series of precipitation.

RC: *The discussion of the models, especially the second one, largely ignores recent findings on runoff generation processes. For instance, the statement on residence time (P11383,l10) should be reformulated with the recent paper of Beven and McDonnell in mind. In the end, for this study the physical correctness of the models is probably less important, but I still find the uncritical description of the models with their partly unrealistic assumptions a bit troublesome.*

AC: Indeed, descriptions and limitations of the models used had to be made more explicit. Following this reviewer comment we revised the model description related to the runoff generation process. The following text has been added to the revised version of the manuscript (sec. 3.2):

> “In particular, in case of Q_{sur} the value of the parameter *k*, which is a function of the residence time in the catchment slopes, is estimated relating the slopes velocity of the surface runoff to the average slopes length L. However, one of difficulties involved is the proper estimation of the surface velocity, which should be calculated for each flood event (Rinaldo and Rodriguez-Iturbe, 1996). According to Rodríguez-Iturbe et al. (1982), such velocity is a function of the effective rainfall intensity and event duration. In this study, the estimate of the surface velocity is performed using the relation between velocity and intensity of rainfall excess proposed in Kumar et al. (2002). In this way it is possible to estimate the average time travel and the consequent parameter *k*, which assumes a different value for each time step as the rainfall changes. Even though the method of this study is based on geomorphological theory of hydrological response, a distributed implementation of such method, for which the distributions of residence times in the possible paths within the basin are related instant unit hydrograph of the basin, should be implemented to better represent runoff generation processes (Bhaskar et al., 1997; Rodríguez-Iturbe and Valdés, 1979; Rosso, 1984). In addition, the correct estimation of the residence time should be derived considering the latest findings reported in McDonnell and Beven (2014). In case of Q_{sub} and Q_g the value of *k* is calibrated comparing the observed and simulated discharge at Vicenza as previously described.”

The goal of this study is to assess the effect of assimilating crowdsourcing observations coming at irregular time steps within an existing early warning system to then apply such methodology in real-life. For this reason, we wanted to keep the Bacchiglione model used in this study as close as possible to the one implemented in the AMICO platform used by AAWA.

RC: *Most importantly, however, I find the small number of tested events problematic (2 in Brue, 1 in Bacchiglione). Obviously the results depend largely on the characteristics of the event and the quality of the precipitation data. I am afraid that this extremely small number of events makes results rather ‘random’. Honestly, I find it therefore difficult to see what this study contributes beyond that the additional information improves simulation somewhat (which one would have expected anyway).*

AC: We agree with the reviewer that the small number of flood events might be problematic to draw general conclusions. We have revised the conclusions and added the following (a limitation of this study):

> “...additional analyses on different case studies and the longer time series of flood events should be carried out in order to draw more general conclusions about assimilation of the crowdsourced observations and their value in different types of catchments and model setups”.

and formulated the corresponding recommendation in the end of sec. 7.

At the same time we would like to stress that we found similar trends in the dependency of Nash index on the number of observations - in both flood events in Brue (Figure 2, Figure 3 and Figure 4) and Bacchiglione. It is now stated in Sec. 6.1 clearer. In fact, after a threshold number of observations, NSE asymptotically approaches a certain value, in both flood events, meaning that no improvement is achieved with additional observations (Figure 2). However, the only difference is that threshold number of observations and asymptotic NSE values are different because model performances can change according to the considered flood events. Example of these considerations are showed in the following two figures

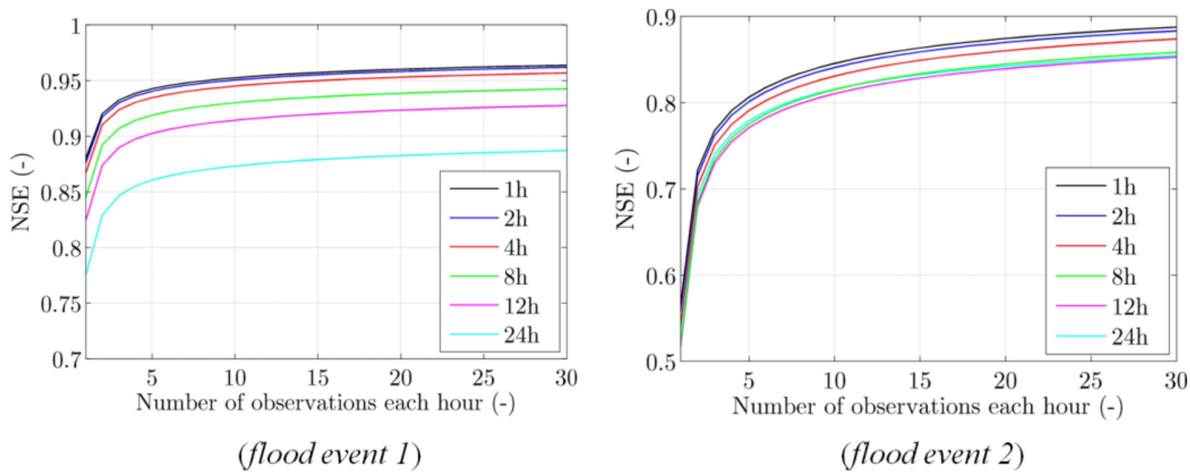

Figure 2. Model improvement during flood event 1 and 2, in case of different lead times, assimilating streamflow observations according to scenario 1

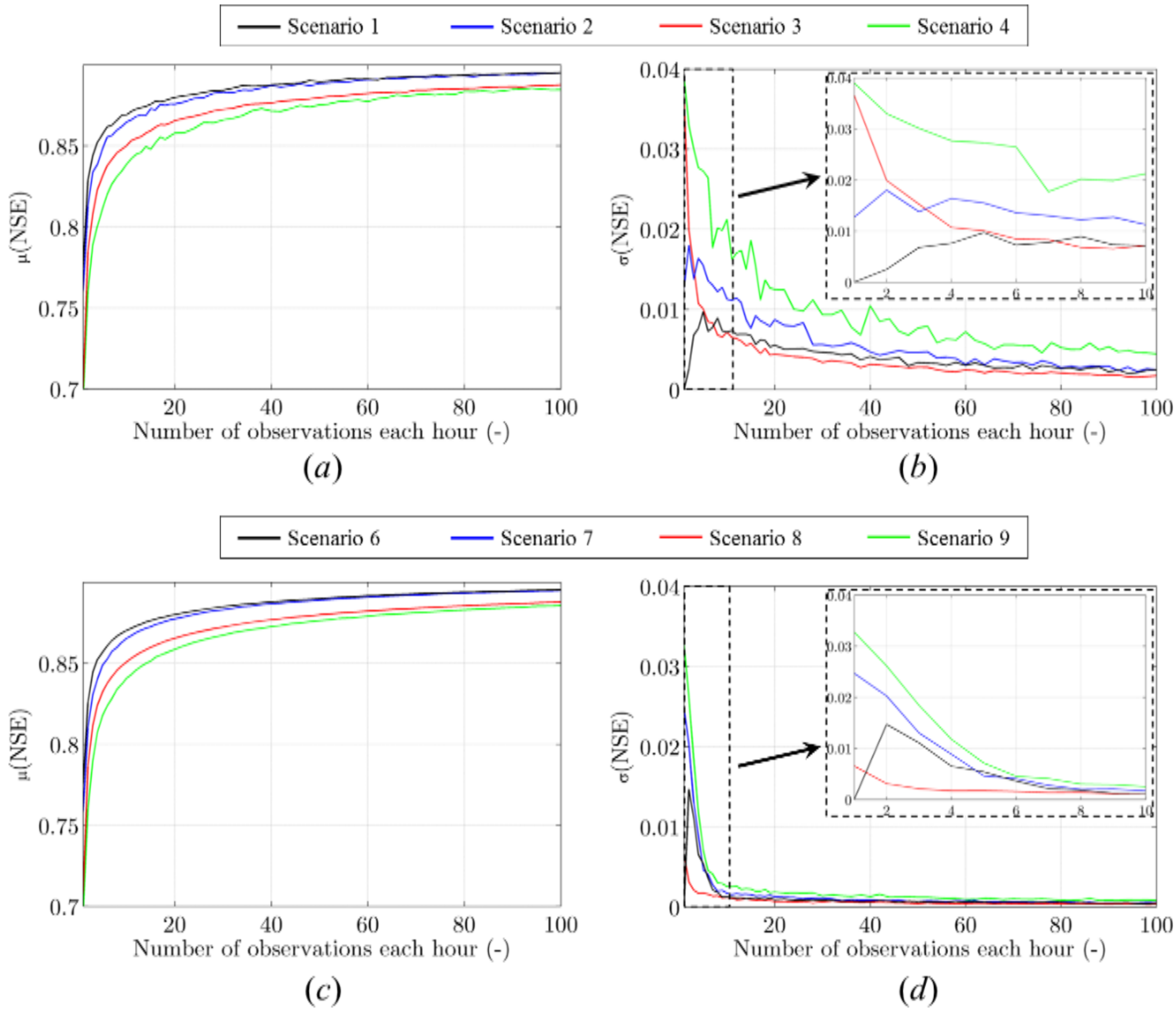

Figure 3. Dependency of μ(NSE) and σ(NSE) on the number of observations, for the scenarios 2, 3, 4, 5, 6, 7, 8 and 9 in case of flood event 1 .

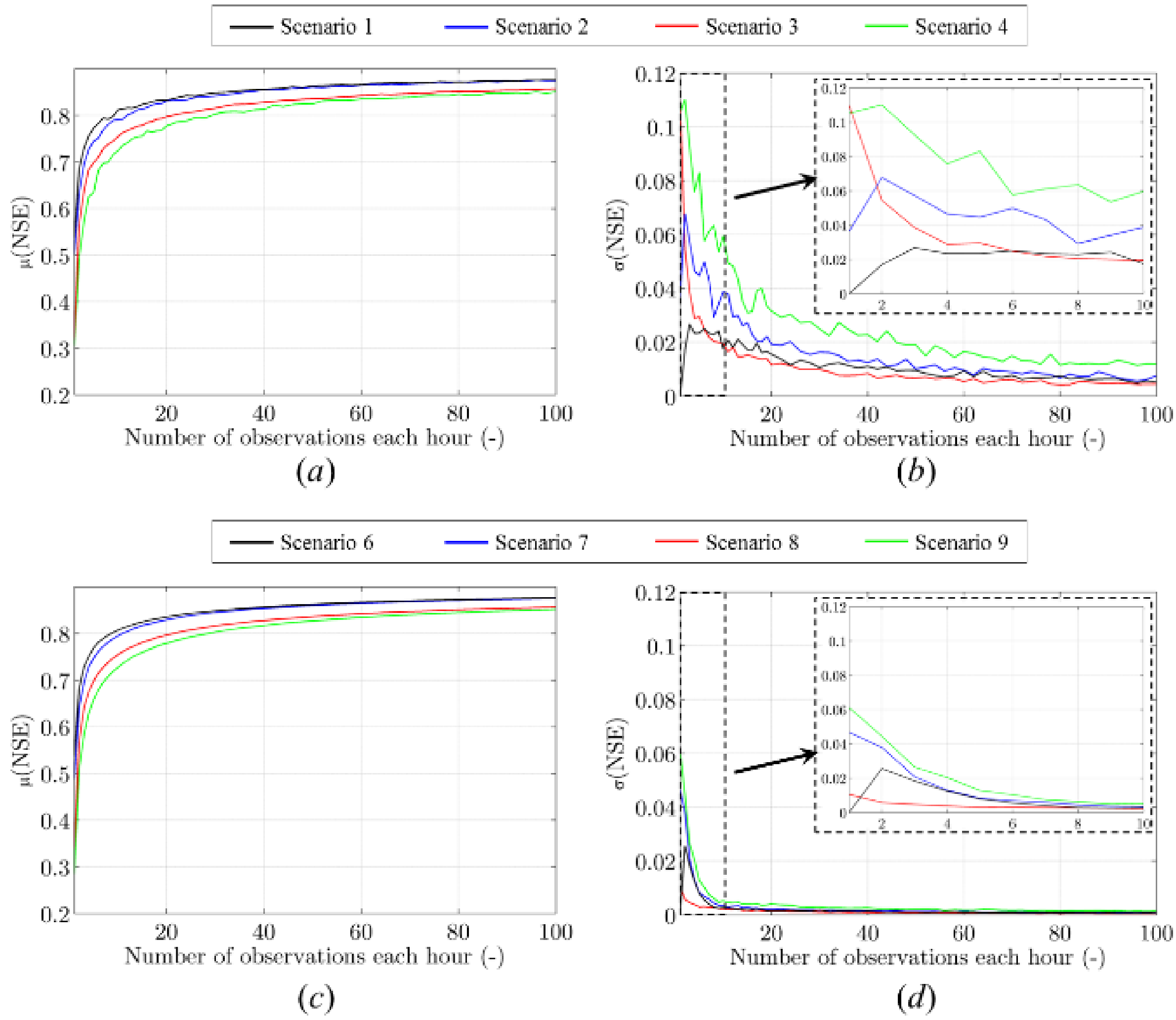

Figure 4. Dependency of μ(NSE) and σ(NSE) on the number of observations, for the scenarios 2, 3, 4, 5, 6, 7, 8 and 9 in case of flood event 2.

This is one of the first studies that address the issue of including crowdsourcing observations coming at asynchronous instant within hydrological modelling. We believe that this study demonstrated that assimilation of crowdsourcing observations can improve hydrological modelling requesting a limited amount of observations, depending on the flood event, even if such observations have variable uncertainty and are coming at random moment.

Our ongoing research is actually looking at a higher number of flood events. After discussions between the authors, we concluded that the current manuscript is already too long to include these new analyses, which will require a number of extra figures.

RC: *The more interesting questions of how big the improvement is, how many observations are needed, at which accuracy, ... all are too heavily influenced by the choice of the one or two event(s) to be of a more general value.*

AC: We agree with reviewer comment on the importance of knowing the amount of needed observations and their level of accuracy. However, model performances vary according to the type of flood event as showed in this study. Consequently, improvement should be related to the particular flood event. It is not possible to define a priori number of observations needed to improve model. In fact, considering figure 1 reported in this reviewer response, in case of flood event 1, we needed only 5 observations to reach NSE equal to 0.9, while during flood event 2, we needed 15 observations to reach the same NSE value. That is why the number of observations necessary to achieve a good model performance might vary according to the flood event. A possible improvement of our method could be a pre-filtering module aimed to select only observations having good accuracy while discarding the ones with low accuracy. However, in this study we aim to assimilate all type of observations without giving recommendations on the minimum acceptable accuracy value of crowdsourced observations. In this way, we demonstrated that high number of observations can compensate with the varying accuracy of them. All these considerations are included in the revised manuscript.

In sec. 6.1 the following text has been added:

> "In both flood events we found similar trends in the dependency of Nash index on the number of observations. However, it is not possible to define a priori number of observations needed to improve model. In fact, after a threshold number of observations (five for flood event 1 and fifteen for flood event 2), NSE asymptotically approaches to a certain value meaning that no improvement is achieved with additional observations."

Minor issues:

RC: *The language could be improved, there are several small language mistakes, which make reading more difficult.*

AC: Language has been improved in the revised version of the manuscript.

RC: *The graphs could be improved, they are in general quite hard to read (and not be too 'nice' to be honest)*

AC: Indeed, legends on some Figures were not clear. We have improved the quality of Figures 2, 3, 4, 6 and 10. Special attention was given to revising Figure 3 with the scheme of the hydrological model used in this study.

RC: *Use mathematically correct terms in your equations! ET (Eq.2), for instance, is not correct as it strictly mathematically means E times T (note that you actually use this in the directly following equation, where CS(t) actually means C times S(t))*

AC: We modified the equations of the manuscript according to reviewer comment

RC: *Please separate results and discussion; this would make reading the text so much easier.*

AC: We have seriously considered this suggestion. Indeed, it is a possibility – but still we would like to suggest that we would follow the initial logic of having results and discussion integrated in one section. This is in a way subjective decision but we checked with a colleague and all came to a conclusion that in this paper the results are presented clearer if each of them is immediately followed by interpretation and corresponding discussion.

RC: *Be careful with your references, some are missing in the reference list (e.g. Krouse) other are misspelled (e.g. Bergström)*

AC: We revised the reference list and corrected the misspelled and added the missing references.

RC: *Reference to recent work could be improved. The work could be better linked to recent work on the value of (limited) data in hydrological modeling. Also, the recent review on citizen science in hydrology by Buytaert et al. (2014) should be referred to.*

AC: We thank the reviewer for the useful comment. This paper deals only with citizen data used for hydrological modelling, and we do not study the value of information in the traditional sensor data. We know of attempts in France and Italy to use the videos of the flood events to estimate the water velocity and to use this information to re-calibrate the river models, however could not find the published papers on this subject. The only reference we could find is EGU abstract by Candela et al. (2013) where citizen data (video) was used to validate the urban flood model in a post-event analysis comparing water depth or flow velocity. To the best of our knowledge, there are no studies that would quantitatively analyse the usefulness of using citizens' data for (real-time) DA in flood modelling.

We do not want to extend the scope of the paper and to deal only with assimilation of crowdsourced observations. We have also followed an advice of reviewer #2 to shorten the Introduction. The reference to Buytaert et al (2014) is very useful indeed (however it does not consider the examples of DA of citizen data for hydrology). We added this references, and the reference to the CitiSense project (this project belongs to the pool of five Citizen Observatories EU projects, together with WeSenseIt which funds this study).

Candela, A., Naso, S. And Aronica, G.: On the use of innovative post-event data for reducing uncertainty in calibrating flood propagation models, EGU General Assembly, Vienna, Austria, 7-12 April, 2013

References:

- *McDonnell, J. J., and K. Beven (2014), Debates—The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph, Water Resour. Res., 50, 5342–5350, doi:10.1002/2013WR015141*
- *Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Bastiaensen, J., Bhusal, J., Clark, J., Dewulf, A., Foggin, M., Hannah, D.M., Hergarten, C., Isaeva, A., Pandey, B., Paudel, D., Sharma, K., Steenhuis, T., Tilahun, S., Van Hecken, G., Zhumanova, M., 2014. Citizen science in hydrology and water resources:opportunities for knowledge generation , ecosystem service management , and sustainable development. Front. Earth Sci. doi:10.3389/feart.2014.00026*