Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2017-102-RC1, 2017

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Comment on "Can assimilation of crowdsourced data in hydrological modelling improve flood prediction?" by Mazzoleni et al. (2017)" *by* Daniele P. Viero

M. Mazzoleni (Referee)

m.mazzoleni@unesco-ihe.org

Received and published: 27 March 2017

The comment on "Can assimilation of crowdsourced data in hydrological modelling improve flood prediction?" addresses the subtle drawback hidden behind the practice of using traditional and crowdsourced data, recorded at different locations, disjointly. The former are used to calibrate semi-distributed models and to force them in real-time, the latter only to update the model states in operational forecasting.

In Mazzoleni et al. (2017), synthetic CSD were generated as model results using observed precipitation, while simulated results were obtained using forecasted precipitation. Since the semi-distributed hydrological model used in Mazzoleni et al. (2017) was calibrated at only one location, Viero (2017) underlined that synthetic CSD at interior

points (different from the calibration one) cannot be considered reliable due to equifinality issues. In fact, semi-distributed hydrologic models are commonly over-parametrized and may provide accurate predictions only where the model is calibrated, and it can fail to represent the relative contribution of upstream tributaries. I read the comment with interest and I really appreciate all the author's efforts. However, I have many doubts and considerations that I would like to share with him.

Overall, I found that the main message of the author have been stretched and repeated many times throughout the comment. It is not clear to me what would the author propose to generate synthetic CSD when only measurements from traditional sensors, located at points different from the ones of CSD, are available. In the summary section, only a pragmatic solution is suggested in case of availability of distributed flow data (not the case in Mazzoleni et al., 2017). This solution involves the collection of CSD for a suitable test period, to verify the model ability in describing the system states correctly at the locations in which CSD are collected. However, this solution will open many other types of questions. For example, how would the author assess the quality of the CSD? Which category of citizen the author would engage in order to collect CSD? For how long will this data collection take place? How can it be insured that CSD quality during data collection will be the same as the CSD quality during real-time modelling updating (no control)? Citizens accuracy is different and data quality assessment is still a burning topic in citizen sciences. In addition, CSD in calibration may be different from the ones in real-time model updating.

Moreover, I do not understand to which extent the comments of the Author are referred to the paper of Mazzoleni et al. (2017) or to a generic issue on the use of CSD in hydrological modelling.

The Author mentioned that "Indeed, for synthetic streamflow CSD to be realistic, two specific requirements have to be met: i) a reliable rating curve must be available for the cross sections where hydrometric CSD are recorded, and ii) the model has to be calibrated at these locations". I agree with the author in case of CSD provided by static

sensors, like in case of Mazzoleni et al. (2017). However, in a real scenario where CSD are provided by citizens at random moments and locations within the catchment, by means of dynamic sensors, I do not agree with the second point of the comment for two reasons. Firstly, assuming the author is right, it would be extremely difficult to calibrate the model with observed data at unknown locations in which synthetic CSD will be assimilated. Secondly, it is not clear to me why synthetic CSD based on model results should be generated if observed data are already available at the CSD/calibration points. Obviously, such observed data should be directly used to generate synthetic scenarios of CSD, like in case of the first three case studies in Mazzoleni et al. (2017), without using any model result.

Another extremely important point is the assimilation of CSD observations. From Viero's comment, it is not clear how erroneous synthetic observations can affect assimilation performances. The author briefly mentions this issue referring to Dee (2005) and Liu et al. (2012). Honestly, since the main objective of Mazzoleni et al. (2017) was the assimilation of realistic synthetic CSD, I was expecting a more comprehensive analysis on the effect of assimilating biased/uncertain observations within hydrological model.

In addition, Viero stated, "In a context of equifinality and of poor identifiability of model parameters, the model internal states can hardly mimic the actual system states away from calibration points, thus reducing the chances of success in assimilating real (i.e. not synthetic) CSD." Why the chances of success in assimilating real CSD is reduced if the model is not calibrated at CSD location? Does this mean that in case of assimilation of distributed soil moisture observations from remote sensing, within a distributed hydrological model, we would need to calibrate the model in each grid cell? I disagree with the author. The main purpose of data assimilation is to use real-time (noisy) observations to update the wrong estimate of the states of a dynamic model (not able to mimic the actual system states away from calibration point). Assimilation of observations at internal points of the catchment is very useful when model states are

less accurate that real-time observations. If a model is able to correctly predict actual system states away from calibration points, why should someone bother to add complexity and uncertainty assimilating CSD observations? The literature provides many studies (e.g. Rakovec et al., 2012) in which hydrological models are updated using measurements at internal points, even if such observations are not used during model calibration.

I am puzzled with the sentence "Therefore, beside the key points identified by Mazzoleni et al. (2017), not only data, but also the model has to match specific requirements for data assimilation to be successful". What are these specific requirements that model has to match?" Is the Author referring to the reliability of synthetic data at calibration points and to the capability of the model to represent truth states?

Maurizio Mazzoleni

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-102, 2017.