Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2017-102-AC1, 2017

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

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

D. P. Viero

daniele.viero@unipd.it

Received and published: 20 May 2017

I'm grateful to the reviewers for their suggestions, which are actually helpful. I'm confident that the revision of the paper will lead to a more focused comment, allowing to better order its structure, and possibly enriching the paper in terms of contents.

1 RESPONSE TO THE COMMENTS OF REVIEWER #1

The comment on "Can assimilation of crowdsourced data in hydrological modelling improve flood prediction?" addresses the subtle drawback hidden behind the practice

Printer-friendly version

Discussion paper

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 **?** *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.*

Maurizio Mazzoleni

I thank Maurizio Mazzoleni (Reviewer #1) for his appreciation of the Comment and for his valuable comments and suggestions, which are addressed in the following.

1. *Overall, I found that the main message of the author have been stretched and repeated many times throughout the Comment.*

   I agree. This is due, in part, to the brevity of the Comment; the main message is repeated at least in the Abstract, in the Introduction, and in the Summary. In the revised version of the Comment, I'll try to better organize the text in order to avoid (or, at least, to limit) useless repetitions.

2. *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.

The work by Mazzoleni et al (2017) is actually a proof-of-concept, which analyze major aspects of the assimilation of crowdsource data in order to improve the forecasting of hydrological models. My Comment essentially focuses on what should follow a proof-of-concept, i.e., on the use of actual crowdsourced data in real, operational flood forecasting. Indeed, it is the passage from a proof-of-concept study to a real-world application (i.e., from synthetic CSD to actually measured CSD) that entails the additional significant drawbacks related to equifinality, overparameterization, and deficiency in model structure, which are not discussed in Mazzoleni et al. (2017).

Accordingly, my Comment is not specifically aimed at proposing a different, better way to generate synthetic CSD when measurements from traditional sensors are available only at different locations from the ones of CSD, as this would mainly pertain to the proof-of-concept study.

I agree that the solution proposed in my Comment opens many other questions, but a problem do exists with when assimilating CSD referring to locations in which the model states can substantially differ from real states. The existence of this problem cannot be ignored; rather, being aware of it is per se important (this is one of the most important reason behind my comment).

Thanks to the suggestion by Reviewer #2, in the revised version of the Comment I'm going to suggest an additional possible solution, which can potentially solve (or aid solving) other questions raised by M. Mazzoleni (e.g., quality of CSD that can be different in calibration and in operational use, etc.). This solution is the use of ensemble-based data assimilation methods to update jointly model states and parameters (and not only model states). Again, like the rest of the Comment, this solution refer to operational use of hydrological models with real (i.e., not synthetic) CSD.

3. *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.*

   I'm aware that it is actually difficult to properly balance comments that must be specific (in that they refer to particular aspects of a given work) and, at the same time, they should be significant in a wider sense. Consider that Reviewer #2 criticized the Comment (in particular Section 2.1) as too specific. I'll try to find a better equilibrium between specificity and generality in the revised version of the Comment.

4. *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.

I thank M. Mazzoleni for this comment, which help me to clarify the focus and the structure of my Comment. In the revised version of the Comment, I'm going to better separate comments referring to the reliability of synthetic CSD due to equifinality issues, from comments on the use of actual CSD in operational forecasting. In this way, I'm confident that misunderstandings, such that those here underlined, could be removed.

5. *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.*

   The issue raised by M. Mazzoleni is undoubtedly interesting; although being not the primary objective of my Comment, it deserve further discussion. Nevertheless, it seems to me that this specific criticism descends from the fact that one of the main goals of Mazzoleni et al. (2017) was how to generate realistic synthetic CSD (this issue pertains to the proof-of-concept scope), whereas I am mainly interested in the differences that modelers may find when moving from proofs-of-concept to real-world applications. I remain convinced that it is impossible to determine if synthetic CSD are realistic or not, when these CDS are generated using a hydrological model at locations where it is impossible to calibrate/verify this model.

In the revised version of the Comments, I'll try to better explain this issue, which is actually intricate. Measured CSD are obviously affected by uncertainty (this issue is assessed in Mazzoleni et al., 2017), and by bias defined as systematic deviation from the actual values (this issue is assessed, e.g., in Dee, 2005 and in Liu et al., 2012, but also in Crow and Van Loon, 2006, who stated that *"inappropriate model error assumptions can lead to circumstances in which the assimilation of surface soil moisture observations actually degrades the performance of a land surface model"*).

In my Comment, I want to point out that even accurate and unbiased measured data can be "seen" as biased data by a model. This can occur when the model is not properly calibrated at sections where data refer to (and model parameters are not update along with model states), or when the model is unable to reproduce the actual dynamic of the system at that location due to intrinsic limitations of the model structure. This issue is better explained with practical examples in my answer to the following point #6.

6. *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 inter-*

nal points of the catchment is very useful when model states are less accurate than 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.*

Thank you for this comment. I realize that I was not precise enough in that part of my Comment, which should be improved.

To answer the key question in this Reviewer's comment (which I highlighted above), I remark that a model can predict wrong system states away from calibration points for different reasons (e.g., wrong/insufficient input data and/or poor calibration and/or structural model deficiencies). Assimilation of observations at internal points of the catchment can be extremely useful when model states are less accurate than real-time observations, but not when this lack of accuracy of model states is due to problems with model structure (or due to poor calibration of model parameters if such parameters are not updated through data assimilation along with the model states).

Therefore, I stress that the statements by M. Mazzoleni are reasonable, but they implicitly assume that the model structure (and the set of model parameter as well, since they were not updated through data assimilation in his work) is (are) able to correctly estimate both the internal states and the model outputs. Although this is desirable for physical-based models (see also the comment #3 of Reviewer #2), one must admit that it is not true in general and, reasonably, it is not true for the model application of the Bacchiglione River presented in Mazzoleni et al. (2017).

I try to clarify the question using first a hypothetical example. Consider a hydrological model, not calibrated at internal points, which provides the right prediction at the closing section as the result of wrong predictions at some internal points. The updating of model states at this internal points based on real data (i.e., different to the internal states needed to provide the 'correct' prediction at the outlet) will likely cause this model to produce worse predictions at the closing section with respect to no assimilation at all. This possible occurrence cannot be detected, nor assessed, if data to be assimilated are extracted from the model itself, because in this case the synthetic data represents the wrong internal states (with respect to the real data at these points).

The problem of assimilating data not coherent with internal model states (when this is due to poor estimation/identifiability of model parameters) could be limited by updating the model parameters along with the internal states of the model (as suggested by Reviewer #2), but this strategy could not be sufficient if the model has structural deficiencies.

As a practical example, consider the "Viale Diaz" floodplain, described in my Comment, which acts as a sort of in-line natural flood control reservoir on flood propagation. Since the attenuation of flood wave exerted by this floodplain can not be properly accounted for by the routing model used in Mazzoleni et al. (2017), the (hypothetical) assimilation of a correct flood hydrograph upstream of the Viale Diaz floodplain leads to incorrect predictions at Ponte degli Angeli, downstream of the "Viale Diaz" floodplain.

7. *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?*

The need to assimilate suitable crowdsourced data was assessed in Mazzoleni et al. (2017). With respect to the specific requirements that the model has to match, its ability of well representing the physics of the hydrological system (i.e., of correctly representing true internal states when forced by correct input data) is actually the key aspect. I'll try to make this point clearer in the revised version of the Comment.

2 RESPONSE TO THE COMMENTS OF REVIEWER #2

The author makes some significant critical remarks on the work of Mazzoleni et al. (2017) that are worth to be considered for publication.

I thank Reviewer #2 for his/her appreciation of my Comment and for his/her very useful and precise suggestions, which are addressed in the following

1. *However, I would first advise to mellow the tone of the narrative.*

   Thanks for the suggestion. I'll revise the text of the paper, trying to smooth the English (and to fix typos).

2. *In addition, I invite the author to make sure that the comments are more general and less focused on the upper Bacchiglione river catchment presented by Mazzoleni et al. (2017). In doing so, Section 2.1 should be reduced considerably, as most of the information and comments seem too specific, and might not be supported for the other test sites.*

   I thank the reviewer for his suggestion. I'll try to shorten Section 2.1 in the revised version of the Comment. While I agree that Sect. 2.1 is very specific, I do believe that most part of this specificity is not meaningless for other test sites. Indeed, I remain convinced that much can be learned from in-depth analyses of specific cases. The opposite risk is the (often unperceived) oversimplification of real systems and processes in our schematic representations (i.e., models) of the reality.

Besides its evident specificity, one of the goals of Sect. 2.1 is to highlight that real-world case studies are often far more complex than what can emerge from most of the applications reported in the literature (this is undoubtedly due to actual limits in papers' length). I am convinced that hydrologists can easily find similarities with other case studies.

Finally, the Comment is indeed a comment to a specific paper, and only one of the four model applications reported by Mazzoleni et al (2017) is here commented, since the contents of the Comment only apply to semi-distributed (and over-parameterized) models and to the assimilation of CSD data in location where the model cannot be calibrated. In the other test cases presented in Mazzoleni et al (2017), the Authors used a lumped model and assimilated CSD only at the calibration sections.

3. *The paper of Mazzoleni et al. (2017) aimed at investigating the value of information retained by crowdsourced data (CSD) when assimilated in surface flow models for flood prediction. Their work is admittedly a proof-of-concept study and the synthetic feature of CSD is quite clear, rather than "briefly mentioned". Their conclusions are correct so long as one assumes the model well represents the physics of the hydrological system, which is a fundamental hypothesis behind observation simulation system experiments. On the other hand, I agree that there seems to be an inherent tendency in Mazzoleni et al. (2017) to present results in a way that somehow overstates the importance of CSD.*

   I agree with the reviewer. The fact that a model well represents the physics of the hydrological system is a fundamental hypothesis for physically-based models, and is tacitly assumed in Mazzoleni et al. (2017). However, it must be stressed that this requirement is not necessarily matched when semi-distributed, physically based models are actually used as lumped models, i.e., they are calibrated only at the closing sections. Given the complexity of the Bacchiglione catchment, the relatively paucity of measured data, and the structure of the model used (see my answer to comment #6 of Reviewer #1 for further details), reasonably it is not true for the model application of the Bacchiglione River presented in Mazzoleni et al. (2017).

4. *There are, in my view, some major points that need to be highlighted: the method chosen for calibrating a model should be consistent regardless of the type of data used. For non-linear models, ensemble based data assimilation methods (e.g the EnKF or the PF) are attractive in that they can be used to update jointly model states and parameters and provide a direct measure of uncertainty. Note that these models cope directly with problems of over parameterization and equifinality since parameter posterior distributions are represented by ensembles. CSD can be instrumental to reduce model uncertainty. Indeed, one can assimilate these data together with traditional hydrologic observations, thereby reducing parameter uncertainty even in those regions where the original reliability of the model is inadequate. In general, the value of information of these data is strictly dependent on their quantity, quality, spatial-temporal distribution. Note that typical data assimilation algorithms are in principle able to screen out noisy data automatically, but need to be modified to tackle possible data bias, which otherwise leads to poorly calibrated models. Thus, it is important, regardless of the nature of the data, to verify if such bias exists before any data assimilation is applied.*

   I thank the Reviewer for these interesting considerations. Ensemble based data assimilation methods are indeed powerful tools. On one hand, their use to jointly update model states and parameters can effectively circumvent the problem of uncertainty in model internal states at crowdsourced data points; on the other hand, such methods can help diagnosing poor identifiability of model parameters.

   However, sophisticated tools to update jointly model parameters and states may fail if assimilating data in locations where the model is unable to correctly repro-

Printer-friendly version

Discussion paper

duce the actual physics of the system. While this possible occurrence can be a-priori conjectured through a close inspection of both the physical system and the model characteristic/capabilities, it can be proved (and quantified) only by comparing model results with measured data (i.e., model validation). The “blind” use of CSD (i.e., its assimilation at locations where the model is neither calibrated nor verified) is at least questionable (see, e.g., the examples reported in the answer to comment #6 of Reviewer #1).

Finally, in the Reviewer’s comment it is stressed the importance of detecting bias in data to be assimilated. This observation pertains also to the object of my Comment, since real (i.e., not synthetic) data referring to locations where the model is unable to reproduce the physic of the system are equivalent to biased data for an model providing accurate estimates at these sections.

I’m going to add this considerations in the revised version of the Comment.

References

Crow, W. T. and Van Loon, E.: Impact of Incorrect Model Error Assumptions on the Sequential Assimilation of Remotely Sensed Surface Soil Moisture, J. Hydrometeorol., 7, 421–432, doi:10.1175/JHM499.1, 2006.

Dee, D.: Bias and data assimilation., Q. J. R. Meteorol. Soc., 131, 3323-3343, doi:10.1256/qj.05.137, 2005.

Liu, Y., Weerts, A., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A., van Velzen, N., He, M., Lee, H., Noh, S., Rakovec, O., and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, Hydrol. Earth Syst. Sc., 16, 3863–3887, doi:10.5194/hess-16-3863-2012, 2012.

Mazzoleni, M., Verlaan, M., Alfonso, L., Monego, M., Norbiato, D., Ferri, M., and Solomatine, D.: Can assimilation of crowdsourced data in hydrological modelling improve flood prediction?, Hydrol. Earth Syst. Sc., 21, 839–861, doi:10.5194/hess-21-839-2017, 2017

Viero, D.P: Comment on “Can assimilation of crowdsourced data in hydrological modelling improve flood prediction?” by Mazzoleni et al. (2017), Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-102, in review, 2017.

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