POINT-BY-POINT REPLY TO COMMENTS

1   RESPONSE TO THE EDITOR'S COMMENTS

The manuscript has now undergone 2 rounds of review. The first set of review comments were considered major and, therefore, the revised manuscript was sent out for review again after the author had made substantial changes to the manuscript.

In this round of reviews, Reviewer #1 notes that there has been significant improvement to the revised manuscript and recommends that the manuscript be accepted with minor revision. Reviewer #2 notes there are interesting points raised in this comment but felt that there was further major revision (and review) needed for the manuscript to be accepted.

In looking at both reviewers comments and my own review of the revised manuscript, the main points of this commentary still need clarification and be generalized enough to make the comment of interest to a wide audience. Reviewer #2 notes that this quite possible but that major revision is still needed.

To help the author clarify the contribution, the reviewers have noted their interpretations of the major points of this commentary and provide specific comments about how this can be done. I offer the same here to help with revisions: In my understanding of the contribution of this comment, I find of most interest the point that models are often called upon for applications that require information to make decisions of high societal value - such as flood forecasting - and that we have a responsibility as hydrologists to understand the limits and applicability of our models and the data that drive those models. As crowd-sourcing data (CSD) can now be used as a further tool to enhance hydrologic modeling efforts, out community needs to consider the use of CSD in our responsible assessment of the applicability of models to answer questions of such importance and relevance. I see the author using a recent paper on CSD, study area, and model to make this comment.

I believe that an additional revision effort, will continue to improve the general interest of the paper and elevate the impact of the comment. For this reason, I am recommending additional revision before the manuscript can be accepted.

I thank again the author and reviewers for the productive discussion and progress towards publication.

Again, I am grateful to the Editor and the Reviewers for their suggestions, which I found insightful and helpful.

I used the Editor's suggestions mainly to improve the Introduction of the Comment, highlighting the importance of debating and deeply understanding both models and data characteristics in order to provide useful and reliable tools to support the decision-making related to flood forecasting and management. Now the beginning of the Introduction reads (additions in red):

"Flood forecasting has a critical importance as it results in decisions of high societal value. It is essential to provide public authorities with the best combination of data and models in order to produce the most accurate flood predictions, and with a robust knowledge of the model behaviour in terms of reliability and uncertainty. Modellers thus have a responsibility to deeply assess the strengths and limitations of models, and to explore different kind of forcing data as well.

Within this general picture, the topic of crowdsourced data is gaining increasing attention among hydrologists. Indeed, the availability of hydrometric data, collected by active citizens in the course of severe flood events, offers a new, unexpected chance to improve real-time flood forecasts. However, the use of crowdsourced data poses severe challenges to modellers since their information content, reliability, arrival frequency, and location are a-priori unknown (Mazzoleni et al., 2015, 2017; van Meerveld et al., 2017). In addition, long time series of crowdsourced data are in fact unavailable.

In pioneering applications, crowdsourced data (CSD) collected in the upper part of a basin were assimilated into adaptive hydrological models..."

Moreover, I'm grateful to M. Mazzoleni as his suggestion helped in amending some imprecisions in the text and in better introducing the matter of the comment, and to the anonymous reviewer for giving a chance to make the manuscript significantly clearer and easily understandable.

2   RESPONSE TO THE COMMENTS OF REVIEWER #1

The author clearly improved the manuscript and I really appreciate his effort. The structure of the comment is now clear and the main objective well described. I do believe that author critical remarks will be helpful for future publications in the same area.

5   *Maurizio Mazzoleni*

I thank M. Mazzoleni (Reviewer #1) for his appreciation of the Comment and for his valuable comments and suggestions, which helped me to further improve my Comment.

1. *However, I still have a couple of comments/suggestions regarding section 3 (use of CSD in operation flood forecasting). I think this section needs more clarifications. In particular, from the abstract and introduction (page 2 lines 1-3) it seems*

10   *that the main objective of the comment is to discuss the effects of assimilation of unreliable synthetic CSD derived from a poorly calibrated model on the assimilation performances. Nonetheless, section 3 provides a qualitative analysis on the effect of model structural uncertainty in data assimilation, which is different than the original scope of this comment. The author need to clarify this issue or to better frame the scope of the comment.*

I agree. The two arguments mentioned by M. Mazzoleni are actually quite different and, although being mutually related

15   in the scope of the manuscript, they both deserve to be mentioned when introducing the main issues of the Comment. Accordingly, I modified the Abstract to read (additions in red):

"... In most real-world applications, hydrological models are calibrated using data from traditional sensors; CSD are typically collected at different locations, where (semi-)distributed models are not calibrated. As a result of either equifinality, poor model identifiability, and lacks in model structure, internal states of (semi-

20   )distributed models can hardly mimic the actual states of complex systems away from calibration points. ~~Indeed, in operational frameworks, the assimilation of real (rather than synthetic) CSD requires a careful assessment~~ Synthetic CSD generated by such models are unreliable and do not allow to assess the effects of model structural uncertainty; their use may lead to overestimating the performance of CSD assimilation with respect to real applications. Additional guidelines are given that are useful for the a-priori evaluation of

25   crowdsourced data for real-time flood forecasting and, hopefully, to plan apt design strategies for both model calibration and collection of crowdsourced data. ..."

In the Introduction, the following modifications were made:

"A practical verification of the results by Mazzoleni et al. (2017) is indeed necessary; furthermore, particular attention has to be paid to ~~additional~~ possible drawbacks inherent in the use of CSD for operational flood

30   forecasting and related to model structural uncertainty, which are not discussed in their proof-of-concept study."

2. *Page 5, lines 25-30: Is the author referring to the assimilation of synthetic or real CSD? I do agree that synthetic CSD estimated at different points than the calibration ones may not be accurate when model is poorly calibrated. This concept was already reported in section 2.3. On the other hand, I do not see the point of this paragraph if the author is referring*

35   *to real CSD. Obviously, a traditional physical sensor will provide more reliable observations than CSD if located at the same point. However, the benefit of CSD is in their spatial distribution and availability in points where physical sensors are not available, as stated already in the original paper. I suggest the author to clarify which type of CSD (synthetic or real) are considered in section 3.*

I agree that the paragraph at hand was not properly structured. I changed the paragraph by moving the first sentence at the

40   end of the paragraph, so that the reason why I put this paragraph at the beginning of section 3 should now be clear. In other words, since CSD typically do not refer to calibration points, we must look carefully at the model behaviour/performance away from calibration points. Now the paragraph at hand reads (moved text in green):

"~~First, it must be observed that CSD typically do not refer to model calibration points, since their natural purpose is to enhance (rather than replace) data from traditional sensors.~~ In general, historical data recorded by traditional sensors are first used to calibrate a model; then, in real-time mode, the same sensors provide data both to force the model and to update the model states (e.g., Ercolani and Castelli, 2017); moreover, the reliability of data from traditional sensors outperforms that of CSD. Hence, from a practical point of view, CSD have limited usefulness at locations already equipped with traditional sensors. Since the natural purpose of CSD is to enhance (rather than replace) data from traditional sensors, CSD typically do not refer to model calibration points."

According to the reviewer's suggestion, I also added several references to "real" CSD in Section 3 (starting from the title of Section 3) in order to clarify what I was referring to.

3. *Page 6, lines 18-24: It is stated in the comment that data assimilation can be used only at calibration points in case of poorly calibrated model. I have some doubts about this statement (in particular page 6 lines 20-22). It is worth noting that even in case of not-properly calibrated model (which is the case of semi-distributed and distributed models) assimilation of reliable observations can help improving model performances. The author should include that a proper estimation of the model error, expressed by means of the error covariance matrix in linear DA or model spread in Ensemble method, is necessary to ensure an appropriate assimilation process in case of poor model calibration. There are many studies in which observations are assimilated at internal points different than the calibration points. This is the case of assimilation of remote sensing observations in case of distributed model calibrated only in few discharge locations (when available). I suggest the author to improve this part.*

I remark that the sentence referred to in this comment is not valid in general; as stated in the text, it specifically applies to cases in which "only internal states are updated". Anyway, I added a sentence to this paragraphs, which now reads (additions in red):

"... assimilation of CSD in operational flood forecasting can be helpful provided that the model is able to well represent the physics of the system at locations where CSD are collected. Of course, data assimilation can contribute, in many cases, to improve such a representation. However, when only internal states are updated (as in Mazzoleni et al., 2017), this condition is met if (and only if) the model is properly calibrated and verified at locations where CSD refer to. Otherwise, correcting internal states of a poorly calibrated model can even lead, in principle, to worse predictions at the outlet than performing no corrections at all (Crow and Van Loon, 2006). It is undoubtedly difficult to assess this issue when only synthetic CSD, generated by the same model, are available for testing the overall method."

4. *Page 7, lines 28-30: I do not agree with this statement. It is because of the random behaviour and involvement of citizens that CSD location cannot be determined a-priori. CSD can be provided at any point of the basin/river and not only at calibration ones. There has been many studies in which mobile applications are developed to correctly estimate river velocity or flow (e.g. Lüthi et al. 2014) by using river cross section (which can be assumed rectangular). Definitely, such tools will provide an uncertainty estimation of the flow characteristics of the river which may lead to a wring update of the model states at interior points. The hope is that more reliable tools for accurately measuring river flow will be developed in the next years.*

I agree, the sentence was formulated in a too much general terms. I then restricted the sentence to water level CSD, which actually need the existence of a rating curve to be converted into streamflow data.

"It must be observed that, while scarce control on the collection of CSD can be exerted during significant flood events, the locations at which citizens can collect CSD of water levels is always determined a-priori, since the availability of rating curves is a necessary condition in order to convert water levels into discharges. The amount of measured data needed to develop reliable rating curves can also be profitably used to calibrate the model at those sections as well."

5. *I have a final question to the author. Since CSD can be assimilated only at points when model is calibrated and physical sensors (more reliable than CSD) are already installed at these points, is the author implicitly suggesting that CSD from citizens should not be used for improving flood predictions even in case that system states are well represented?*

No, I'm not saying that. CSD can be assimilated at locations different from calibration points, but in this case attention must be paid to update not only model internal states, but also model parameters. Moreover, the suitability of model structure in correctly represent the physics of the real system at these points (when the model is properly forced, of course) must be a-priori checked.

6. *I hope these suggestions will help the author to further improve his valuable contribute to the use of CSD from citizens for improving flood prediction.*

I thank again M. Mazzoleni for his helpful suggestions.

3 RESPONSE TO THE COMMENTS OF REVIEWER #2

Thanks for the opportunity to review the revised version o this comment. I read both this version and the preceding discussion with interest. Honestly, I have somewhat split feelings about this comment. On the one hand some valuable points are made, but on the other hand the 'comment-aspects' are not fully clear.

I thank Reviewer #2 for his/her effort in reviewing the Comment and in providing useful suggestions, which are addressed in the following

1. *As I see it, the author makes two main comments: 1) the use of synthetic data for the crowdsourced observations and 2) the model choice and application. While the author is very critical about the first point, I would disagree to phrase this as general as done in the comment. Studies using synthetic data can actually be quite informative to investigate the question on how valuable such data potentially could be if they would be available. I would argue this is a suitable approach can actually provide guidance on how to collect crowed sourced data (see also van Meerveled et al, in review, HESS-D, as an example of this approach). The author needs to provide more convincing arguments why the approach in general is not suitable or on where exactly he sees short-comings of the particular implementation of this approach in the study by Mazzoleni et al.*

I admit I was quite puzzled in reading this comment. I carefully re-read the manuscript and I did not find anything, with reference to the use of synthetic data, saying that "the approach in general is not suitable". I searched for all the occurrences of the word "synthetic" throughout the text, and I found that synthetic data (and limitations related to their use) are always mentioned with reference to the use of (semi-)distributed (and overparametrized) hydrological models and, more specifically, with reference to the Bacchiglione case study presented by Mazzoleni et al. (2017). I agree with the reviewer that, in general, the use of synthetic data can actually be quite informative. As a matter of fact, I remark that Mazzoleni et al. (2017) used synthetic data in three additional case studies and, in my Comment, I was not critical at all about those applications. However, synthetic data must be reliable or, alternatively, their uncertainty/inaccuracy have to be (fairly) known. In my comment, I show that this is not the case when synthetic data are generated away from calibration points by a (semi-)distributed hydrological model. Specifically, if synthetic data are extracted from the best-fit scenario and than assimilated into the same model, they are obviously leading to better performance than real crowdsourced data. In other words, the question on how valuable such data potentially could be if they would be available can not be properly answered if synthetic data are surely better (but no one can say how much better) than real crowdsourced data.

I then concluded that maybe the Abstract and the Introduction were not well formulated, and thus formed a sort of wrong perception of what I was commenting on in the following sections of the manuscript.

Accordingly, I revised the Abstract to read (additions in bold):

"In their recent contribution, Mazzoleni et al. (2017) investigated the integration of crowdsourced data (CSD) in hydrological models to improve the accuracy of real-time flood forecasts. The Authors used synthetic CSD (i.e., not actually measured), because real crowdsourced data were not available at the moment of the study. In their work, which is actually a proof-of-concept study, Mazzoleni et al. (2017) showed that assimilation of CSD improves the overall model performance; the impact of irregular frequency of available CSD, and that of data uncertainty, were also deeply assessed. However, ~~it has to be remarked that, in their work, the Authors used synthetic (i.e., not actually measured) crowdsourced data, because actual crowdsourced data were not available at the moment of the study. For this reason, the work by Mazzoleni et al. (2017) is actually a proof-of-concept study.~~ the use of synthetic CSD in conjunction with a semi-distributed hydrological model deserves further discussion. In most real-world applications, hydrological models are calibrated using data from traditional sensors; CSD are typically collected at different locations, where (semi-)distributed models are not calibrated. As a result of either equifinality, poor model identifiability, and lacks in model structure, internal states of (semi-)distributed models can hardly mimic the actual states of complex systems away from calibration points. ~~Indeed, in operational frameworks, the assimilation of real (rather than synthetic) CSD requires a careful assessment~~ Synthetic CSD generated by such models are unreliable and do not allow to assess the effects of model structural uncertainty; their use may lead to overestimating the performance of CSD assimilation with respect to real applications. Additional guidelines are given that are useful for the a-priori evaluation of crowdsourced data for real-time flood forecasting and, hopefully, to plan apt design strategies for both model calibration and collection of crowdsourced data."

I also found an overstatement in the Introduction, and I amended the text to read (additions in red):

"A practical verification of the results by Mazzoleni et al. (2017) is indeed necessary; furthermore, particular attention has to be paid to ~~additional~~ possible drawbacks inherent in the use of CSD for operational flood forecasting and related to model structural uncertainty, which are not discussed in their proof-of-concept study."

Finally, once clarified that in my Comment it was never stated that the approach is in general not suitable, I remark that Section 2.3 is completely devoted to explain where exactly I see shortcomings in the particular implementation of this approach in the study by Mazzoleni et al. (2017), in a way that M. Mazzoleni found this comment reasonable and useful. Nonetheless, to further clarify this key point, I substantially enhanced Section 2.3 (see my answer to your next comment).

2. *The second point I actually find more interesting. The author nicely provides reasons on why a so called physical model is not as physical as one might think. I find this discussion really helpful, although it could be a bit more to the point. I can clearly sympathize with the argument that the use of such a 'physical but still not so physical' model has implementations for the result in a study which heavily relies on model calibration, and probably the use of a simpler model by Mazzoleni et al. could have been appropriate. However, it is not clear from the comment in which way the author thinks this could have influenced the results.*

I realized that, in the previous version of the manuscript, this part was not clear enough. Therefore, I substantially enhanced Section 2.3, which is entirely devoted to explain this point. Now it reads (additions in red, moved text in green):

"In the Bacchiglione case study, Mazzoleni et al. (2017) calibrated the model using measured rainfall data to well reproduce the streamflow hydrograph at the closing section (call this post-event simulation "scenario 1"). Then they forced the model with predicted rainfall fields that were completely different from the actual storm event ("scenario 2"); in this case, the discharge simulated using forecasted input was very different from that obtained using recorded rainfall, with a significant time shift and errors in predicted discharge ranging between 25 and 50% at the flood peak (and up to 90% if considering synchronous data). ~~In Mazzoleni et al. (2017), synthetic CSD of streamflow are results of the model itself.~~ In the "scenario 3", similarly to

5

the "observing system simulation experiment" (OSSE) approach, synthetic streamflow CSD extracted from the "scenario 1" were assimilated into a new run using the same forcing as in the "scenario 2". ~~synthetic CSD were calculated by forcing the hydrological model with measured precipitation recorded during the considered flood events (post-event simulation).~~ Not surprisingly, the model performance in the "scenario 3" was significantly better than in the "scenario 2".

The Authors claimed that ~~these~~ synthetic CSD they used are realistic. ~~; however,~~ For this condition to be met, given that these CSD are results of the model itself, the model must represent well the physics of the real system (i.e., it must be calibrated or, at least, verified) at locations where CSD are first generated and then assimilated~~, which~~; this is a fundamental hypothesis behind the OSSE approach. As a matter of fact, the synthetic CSD used in Mazzoleni et al. (2017) for the Bacchiglione case study are representative of the model internal states of the best-fit scenario. ~~But~~ However, recalling that such CSD do not refer to model control points, nothing can actually be said about the model performance at locations where CSD are generated and, as a consequence, about their accuracy. Real CSD are then expected to be farther from the best-fit scenario than the synthetic CSD generated by the model; that is, real CSD are likely biased with respect to the synthetic CSD actually used. Therefore, assimilation of real crowdsourced data can not be as effective as that performed in Mazzoleni et al. (2017).

From one point of view, such an inconsistency could have led Mazzoleni et al. (2017) to overrate the importance of CSD ~~in Mazzoleni et al. (2017), who~~, as they considered issues related to CSD precision, but not accuracy (Mazzoleni et al., 2016). ~~In other words, real CSD are likely biased with respect to the synthetic CSD they used but, contrarily to Mazzoleni et al. (2016), this aspect was not accounted for in Mazzoleni et al. (2017).~~ From a more general point of view, additional care must be taken in operational flood forecasting when assimilating CSD into (semi-)distributed hydrological models at locations other than model control points. This last point is further discussed in the next section."

3. *From the comment it is clear that certain aspects of Mazzoleni et al. could have done differently or have been described clearer. However, to be really useful a comment has to be specific and raise issues of general importance. As argued above, I do not agree with the fundamental critic against the use of synthetic data. While point 2 still could be of general interest, it is not yet formulated in such a way. In the current form it mainly describes the details but misses to frame this in a more general discussion on which model to use when, including the consequences of using a too complex model. To summarize, while the comment raises an important point which could be of general interest, some significant work is needed to make the comment as useful as it could be.*

I agree with the reviewer that, to be really useful, a comment has to respond to two different (and substantially opposite) needs: it has to analyse and debate specific aspects, and has to raise issues of general importance as well.

I hope that, with the enhancements made in the revised version of the paper (as described above), the part referring to specific aspects is now sufficiently clear and complete. I am also convinced that the issue concerning the critic against the use of synthetic data (not really present in the manuscript) is now clarified.

On the other hand, I point out that a "general discussion on which model to use when" goes far beyond the scope of the present manuscript that, being in fact a specific commentary on the paper by Mazzoleni et al. (2017), mainly deals with possible shortcomings in the assimilation of crowdsourced data into (semi-)distributed hydrological models for real-time flood forecasting.

I believe that the comment, at least for people operating in the field of flood forecasting and interested in the use of crowd-sourced data, already presents issues of quite general interest. Nevertheless, in Section 3, which is mainly devoted to generalize the main specific issues of the Comment, I added some sentences in order to make the reasoning more clear. Specifically, in the revised version of the manuscript, it is explicitly stated that CSD are typically spatially distributed. Accordingly, spatially explicit models are needed in order to take advantage from this kind of data. Unfortunately, physically-based, (semi-)distributed models suffer from equifinality, poor identifiability of model parameters, and structural deficiencies, leading to possible shortcomings related to the assimilation of CSD referring to locations different

from the calibration point of the model. Finally, guidelines and possible solutions (along with associated limitations) are discussed.

References

Lüthi, B., Philippe, T. and Peña-Haro, S.: Mobile device app for small open-channel flow measurement, International Environmental Modelling and Software Society (iEMSs) 7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA, 2014.

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

van Meerveld, I., Vis, M., and Seibert, J.: Information content of stream level class data for hydrological model calibration, Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-72, in review, 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.

5

Comment on "Can assimilation of crowdsourced data in hydrological modelling improve flood prediction?" by Mazzoleni et al. (2017)

Daniele P. Viero[1]

[1]Department of Civil, Environmental, and Architectural Engineering, University of Padova, via Loredan 20, 35131, Padova (Italy).

Correspondence to: Daniele P. Viero (daniele.viero@unipd.it)

Abstract. In their recent contribution, Mazzoleni et al. (2017) investigated the integration of crowdsourced data (CSD) in hydrological models to improve the accuracy of real-time flood forecasts. The Authors used synthetic CSD (i.e., not actually measured), because real crowdsourced data were not available at the moment of the study. In their work, which is actually a proof-of-concept study, Mazzoleni et al. (2017) showed that assimilation of CSD improves the overall model performance; the impact of irregular frequency of available CSD, and that of data uncertainty, were also deeply assessed. However, ~~it has to be remarked that, in their work, the Authors used synthetic (i.e., not actually measured) crowdsourced data, because actual crowdsourced data were not available at the moment of the study. For this reason, the work by Mazzoleni et al. (2017) is actually a proof-of-concept study.~~ the use of synthetic CSD in conjunction with a semi-distributed hydrological model deserves further discussion. In most real-world applications, hydrological models are calibrated using data from traditional sensors; CSD are typically collected at different locations, where (semi-)distributed models are not calibrated. As a result of either equifinality, poor model identifiability, and lacks in model structure, internal states of (semi-)distributed models can hardly mimic the actual states of complex systems away from calibration points. ~~Indeed, in operational frameworks, the assimilation of real (rather than synthetic) CSD requires a careful assessment~~ Synthetic CSD generated by such models are unreliable and do not allow to assess the effects of model structural uncertainty; their use may lead to overestimating the performance of CSD assimilation with respect to real applications. Additional guidelines are given that are useful for the a-priori evaluation of crowdsourced data for real-time flood forecasting and, hopefully, to plan apt design strategies for both model calibration and collection of crowdsourced data.

1 Introduction

Flood forecasting has a critical importance as it results in decisions of high societal value. It is essential to provide public authorities with the best combination of data and models in order to produce the most accurate flood predictions, and with a robust knowledge of the model behaviour in terms of reliability and uncertainty. Modellers thus have a responsibility to deeply assess the strengths and limitations of models, and to explore different kind of forcing data as well.

1

Within this general picture, the topic of crowdsourced data is gaining increasing attention among hydrologists. Indeed, the availability of hydrometric data, collected by active citizens in the course of severe flood events, offers a new, unexpected chance to improve real-time flood forecasts. However, the use of crowdsourced data poses severe challenges to modellers since their information content, reliability, arrival frequency, and location are a-priori unknown (Mazzoleni et al., 2015, 2017; van Meerveld et al., 2017). In addition, long time series of crowdsourced data are in fact unavailable.

In pioneering applications, crowdsourced data (CSD) collected in the upper part of a basin were assimilated into adaptive hydrological models to reduce uncertainty in forecasting flood hydrographs at downstream sections (Mazzoleni et al., 2015). In a recent work, Mazzoleni et al. (2017) paid particular attention to the issues of uncertainty and irregular arrival frequency of CSD. Their results showed that assimilation of CSD improves the overall model performance. They also showed that the accuracy of CSD is, in general, more important than their arrival frequency.

~~However,~~ In their work, the Authors used synthetic (i.e., not actually measured) CSD, because real streamflow CSD were not available at the moment of the study. Commenting on this aspect, the Authors wrote "*the developed methodology is not tested with data coming from actual social sensors. Therefore, the conclusions need to be confirmed using real crowdsourced observations of water level*". A practical verification of the results by Mazzoleni et al. (2017) is indeed necessary; furthermore, particular attention has to be paid to ~~additional~~ possible drawbacks inherent in the use of CSD for operational flood forecasting and related to model structural uncertainty, which are not discussed in their proof-of-concept study.

The Comment is outlined as follows. Section 2 presents a deep assessment of the Bacchiglione River case study (i.e., the fourth case study presented in Mazzoleni et al., 2017), in order to highlight the actual gap between a proof-of-concept study and a real application for operational flood forecasting. Given the complexity of the basin and the relatively paucity of available data, it is shown that the semi-distributed model used in Mazzoleni et al. (2017) is unable to properly represent the physics of the whole hydrological and hydraulic system, with adverse effects on the assimilation of real CSD. Based on the key features delineated in Sect. 2, a more general assessment of CSD assimilation in (semi-)distributed hydrological models is given in Sect. 3. A brief summary closes the Comment.

2   Specific comments

2.1   The Bacchiglione catchment closed at Ponte degli Angeli (Vicenza)

The catchment of the upper Bacchiglione River, closed at Ponte degli Angeli in the historical centre of Vicenza (Fig. 1), is located in the north of the Veneto Region, a plain that is fringed by the Alpine barrier at a distance of less than 100 km to the north of the Adriatic Sea (Barbi et al., 2012).

With regard to the precipitation climatology, the southern part of this plain is the drier, with approximately 700–1000 mm of mean annual rainfall, whereas more than 2000 mm are measured close to the pre-alpine chain due to the interaction of the southerly warm and humid currents coming from the Mediterranean Sea with the mountain barrier (Smith, 1979). A significant portion of the annual rainfall often concentrates into very short periods of time in the form of what often turns out to be an extreme event with deep convection playing a central role (Barbi et al., 2012; Rysman et al., 2016). As a consequence, severe
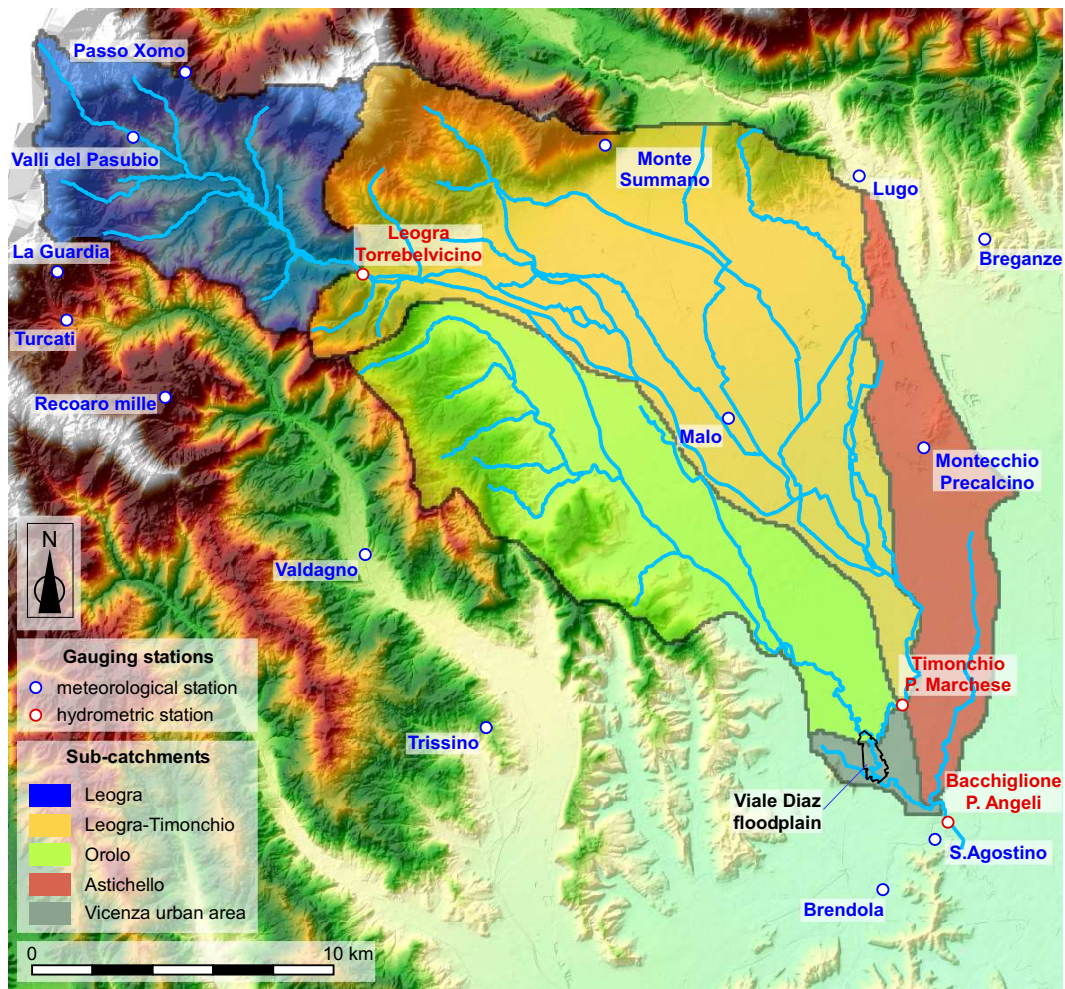
Figure 1. The catchment of the Bacchiglione River closed at Ponte degli Angeli, Vicenza (Italy).

flooding event have threatened agricultural and urban areas in the recent years (e.g. Viero et al., 2013; Scorzini and Frank, 2015).

Due to the spatial and temporal variability of the rainfall fields meteorological models are often unable to provide accurate and reliable quantitative precipitation estimates (QPE) for the upper Bacchiglione catchment. An example of this inadequacy

5 is given, for instance, by Fig. 13 in Mazzoleni et al. (2017). ~~The discharge simulated using forecasted input is very different~~ ~~from that obtained using recorded rainfall, with a significant time shift and errors in predicted discharge ranging between 25~~ ~~and 50% at the flood peak (and up to 90% if considering synchronous data).~~

The upper Veneto plain is a highly populated and urbanized area, with extremely complex drainage and irrigation networks that significantly affect both runoff production and propagation (Viero and Valipour, 2017). Within this plain, the Bacchiglione

10 River and its tributaries are provided with relatively high levees (Viero et al., 2013), which prevent the exchange of water from

3

inside to outside the riverbed (and vice-versa) when the inner water levels are relatively high. As a consequence, the minor channel networks are not always allowed to deliver their drainage water towards the nearest tributary, i.e., the inflow points along the main river reaches change during a flood event depending on the instantaneous water level within the river. This occurrence modifies the network connectedness which, in turn, leads to different mechanisms of hydrologic response in the overall catchment.

Just upstream of the City of Vicenza, an area of up to 1 km^2 of the "Viale Diaz" floodplain (Fig. 1) is flooded when the Bacchiglione flow rate exceeds ~ 160 m^3/s. Since about $2 \cdot 10^6$ m^3 of water can be temporarily stored in this area, a significant flood attenuation can be produced, particularly in case of floods with a steep rising limb (which is often the case due to the climatic regime and the catchment characteristics).

Moreover, the lower part of the Bacchiglione basin, North of Vicenza, includes a vast groundwater resurgence zone, in which it's difficult to assess both the actual contribution of resurgence to the Bacchiglione streamflow (up to ~ 30 m^3/s) and the time-variable behaviour of soil moisture.

Clearly, such a system is highly non-linear. Nonetheless, significant parts of the Bacchiglione catchments are poorly monitored, and the remaining parts are completely unmonitored. The Leogra subcatchment (blue shaded area in Fig. 1) is provided with a pressure-transducer for the measure of water level at Torrebelvicino (Fig. 1). A rating curve derived from theoretical considerations is available for this cross-section. However, the absence of instrumental measures of flow discharge limits its reliability. The Leogra-Timonchio subcatchment (orange shaded area in Fig. 1) is monitored by an ultrasonic stage sensor located at Ponte Marchese, just upstream of the confluence with the Orolo River. Flow rate measurements at Ponte Marchese refers only to low hydraulic regimes, and show great variability due to the operations of a hydroelectric power plant located just downstream of Ponte Marchese. The Orolo River (green shaded area in Fig. 1), with a discharge capacity of more than one third of the Bacchiglione at Ponte degli Angeli, is one of its major tributaries. Unfortunately, not only the Orolo subcatchment is completely uncovered by meteorological gauging stations, but also no hydrometric gauging stations are present along its reach. Similarly to the Orolo, the Astichello catchment (red shaded area in Fig. 1) is unmonitored and, due to backwater effects, significant areas adjacent to the Astichello are flooded when water levels in the Bacchiglione are relatively high. Hence, the discharge that effectively flows from the Astichello into the Bacchiglione River may significantly reduce depending on the water stage within the main course of the Bacchiglione River.

Attention must be paid to the fact that the three major tributaries (Orolo, Timonchio, and Astichello) meet just upstream of the closing section of Ponte degli Angeli (Fig. 1), making it difficult to correctly estimate the actual contribution of each single tributary to the total streamflow. By looking at the tree-like structure of the drainage network in an electrical analogy (Rodríguez-Iturbe and Rinaldo, 2001), the major tributaries of the Bacchiglione are in fact "conductors in parallel".

Certainly, given the irregular topography of the catchments, the heterogeneity of the landscape, and the complexity of the hydraulic network, it can be stated that the Bacchiglione catchment is poorly monitored.

2.2 The semi-distributed model of the Bacchiglione catchment

In catchments like that of the Bacchiglione River, for all the reasons reported in the previous section, the accurate prediction of flood hydrographs by performing continuous time simulations is unquestionably a hard task (Anquetin et al., 2010).

Mazzoleni et al. (2017) used an available semi-distributed hydrological model coupled with a Muskingum–Cunge scheme for flood propagation within the main river network, which was originally set up to forecast flood hydrographs at the closing section of Ponte degli Angeli (Vicenza). Sensibly, the model was calibrated by minimizing the root mean square error between observed and simulated values of water discharge only at Ponte degli Angeli, which is the only hydrometric station provided with a reliable rating curve. The semi-distributed model, although explicitly representing the hydrological processes within the main subcatchments, has to be intended as a lumped model from a practical standpoint, since the discharge in Ponte degli Angeli is its only control point.

Therefore, no matter the accuracy of the streamflow predictions in Ponte degli Angeli, little can be said about the accuracy of the model in describing the internal states of the system, such as the streamflow along the upstream tributaries. This limitation has to be ascribed to uncertainty in precipitation fields, to the paucity of (reliable) flow rate data upstream of Vicenza, and to inherent limitations of the model itself.

Indeed, it has to be remarked that the Muskingum–Cunge model for flood propagation used in Mazzoleni et al. (2017) considers rectangular river cross-sections for the estimation of hydraulic radius, wave celerities, and other hydraulic variables (Todini, 2007). Accordingly, the effects exerted by the "Viale Diaz" floodplain, which acts as a sort of in-line natural flood control reservoir on flood propagation, can not be properly accounted for. This means that, if the flood hydrograph is correctly modelled at Ponte degli Angeli, it can not be correctly modelled upstream of the Viale Diaz floodplain (and vice-versa).

2.3 The use of synthetic CSD in the Bacchiglione case study

In the Bacchiglione case study, Mazzoleni et al. (2017) calibrated the model using measured rainfall data to well reproduce the streamflow hydrograph at the closing section (call this post-event simulation "scenario 1"). Then they forced the model with predicted rainfall fields that were completely different from the actual storm event ("scenario 2"); in this case, the discharge simulated using forecasted input was very different from that obtained using recorded rainfall, with a significant time shift and errors in predicted discharge ranging between 25 and 50% at the flood peak (and up to 90% if considering synchronous data). In Mazzoleni et al. (2017), synthetic CSD of streamflow are results of the model itself. In the "scenario 3", similarly to the "observing system simulation experiment" (OSSE) approach, synthetic streamflow CSD extracted from the "scenario 1" were assimilated into a new run using the same forcing as in the "scenario 2". synthetic CSD were calculated by forcing the hydrological model with measured precipitation recorded during the considered flood events (post-event simulation). Not surprisingly, the model performance in the "scenario 3" was significantly better than in the "scenario 2".

The Authors claimed that these synthetic CSD they used are realistic. ; however, For this condition to be met, given that these CSD are results of the model itself, the model must represent well the physics of the real system (i.e., it must be calibrated or, at least, verified) at locations where CSD are first generated and then assimilated, which; this is a fundamental hypothesis

behind the OSSE approach. As a matter of fact, the synthetic CSD used in Mazzoleni et al. (2017) for the Bacchiglione case study are representative of the model internal states of the best-fit scenario. ~~But~~ However, recalling that such CSD do not refer to model control points, nothing can actually be said about the model performance at locations where CSD are generated and, as a consequence, about their accuracy. Real CSD are then expected to be farther from the best-fit scenario than the synthetic

5 CSD generated by the model; that is, real CSD are likely biased with respect to the synthetic CSD actually used. Therefore, assimilation of real crowdsourced data can not be as effective as that performed in Mazzoleni et al. (2017).

From one point of view, such an inconsistency could have led Mazzoleni et al. (2017) to overrate the importance of CSD ~~in Mazzoleni et al. (2017), who~~, as they considered issues related to CSD precision, but not accuracy (Mazzoleni et al., 2016). ~~In other words, real CSD are likely biased with respect to the synthetic CSD they used but, contrarily to Mazzoleni et al. (2016),~~

10 ~~this aspect was not accounted for in Mazzoleni et al. (2017).~~ From a more general point of view, additional care must be taken in operational flood forecasting when assimilating CSD into (semi-)distributed hydrological models at locations other than model control points. This last point is further discussed in the next section.

3   The use of real CSD in operational flood forecasting

As remarked by Mazzoleni et al. (2017), the success of assimilating real CSD in hydrological modelling strictly depends on

15 their accuracy, quantity, and spatial-temporal distribution. However, attention must be paid not only to CSD, but also to the model.

~~First, it must be observed that CSD typically do not refer to model calibration points, since their natural purpose is to enhance (rather than replace) data from traditional sensors.~~ In general, historical data recorded by traditional sensors are first used to calibrate a model; then, in real-time mode, the same sensors provide data both to force the model and to update the model states

20 (e.g., Ercolani and Castelli, 2017); moreover, the reliability of data from traditional sensors outperforms that of CSD. Hence, from a practical point of view, CSD have limited usefulness at locations already equipped with traditional sensors. Since their natural purpose is to enhance (rather than replace) data from traditional sensors, and considering that they can be collected at locations not known a priori, CSD typically do not refer to model calibration points.

Given the spatially distributed nature of CSD, spatially explicit hydrological models can take the major advantage from

25 CSD. On the other hand, ~~Accordingly,~~ particular care has to be taken when dealing with physically-based, (semi-)distributed models, which are known to suffer from equifinality and poor identifiability of model parameters (Beven, 2006).

After the critical work by Beven (1989), detailed investigations were carried out about the model complexity needed to simulate rainfall-runoff process. Several studies indicated that the information content in a rainfall-runoff record is sufficient to support models of only very limited complexity (Jakeman and Hornberger, 1993; Refsgaard, 1997). This implies that dis-

30 tributed, or semi-distributed, hydrological models are seldom calibrated; rather, they are commonly over-parametrized, since calibration rarely involves their internal states (Sebben et al., 2012; Viero et al., 2014).

In addition, flood routing processes are typically oversimplified in operational models meant to real-time flood forecasting (Mejia and Reed, 2011). For instance, significant effects related to either compound sections, large floodplains connected to the main channel, or confluences causing backwater effects, are seldom accounted for.

As a consequence, (semi-)distributed rainfall-runoff models may provide accurate predictions of outflow discharge at the closing section and, at the same time, poor predictions of internal states of the system (e.g., the soil moisture content, or the relative contribution of upstream tributaries); in other words, one can likely get the correct answer for the wrong reason (Loague et al., 2010). Therefore, (semi-)distributed models can be said calibrated only at calibration (or control) points, and verified only at locations in which model results are shown to compare favourably with enough (and accurate enough) measured data.

This caveat particularly applies to assimilation of CSD in hydrological modelling for operational, real-time flood forecasting. Indeed, while CSD typically refer to model internal states, they are assimilated in order to improve the accuracy of the main outputs of the model, such as streamflow hydrographs at closing sections (model internal states are relatively less important in this context).

Recalling that model input, states, parameters, and outputs (or a subset of them) can be updated using different data assimilation techniques (Refsgaard, 1997), assimilation of CSD in operational flood forecasting can be helpful provided that the model is able to well represent the physics of the system at locations where CSD are collected. Of course, data assimilation can contribute, in many cases, to improve such a representation. However, when only internal states are updated (as in Mazzoleni et al., 2017), this condition is met if (and only if) the model is properly calibrated and verified at locations where CSD refer to. Otherwise, correcting internal states of a poorly calibrated model can even lead, in principle, to worse predictions at the outlet than performing no corrections at all (Crow and Van Loon, 2006). It is undoubtedly difficult to assess this issue when only synthetic CSD, generated by the same model, are available for testing the overall method.

As a valid alternative for operational forecasting, ensemble based data assimilation methods (e.g., the Ensemble Kalman Filter or the Particle Filter) can be used to update jointly model states and parameters and to provide a direct measure of uncertainty. In this way, models cope directly with equifinality and problems of over-parametrization, since parameter posterior distributions are represented by ensembles. 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.

Nonetheless, also such sophisticated tools may fail if the model has structural deficiencies that make it unable to represent true system states at given locations. As a representative example, consider the Bacchiglione River (Fig. 1) and, specifically, the "Viale Diaz" floodplain described in Sec. 2. The role played by such an in-line flood control reservoir on flood routing can not be accounted for using a basic Muskingum–Cunge model that considers rectangular cross-sections. It follows that the assimilation of accurate streamflow data referring to a section located just upstream of the Viale Diaz floodplain (e.g., Ponte Marchese, see Fig. 1) can likely deteriorate the model predictions in Ponte degli Angeli, downstream of the floodplain.

Shortcomings similar to the one described above, which can be found in many different case studies, can be a-priori conjectured through a close inspection of both the physical system and the model characteristic. Their quantitative assessment needs an extensive comparison with measured data; of course, a "blind" use of CSD (i.e., their assimilation at locations where the model is neither calibrated nor verified) is at least questionable.

4 Summary

The approach proposed and investigated by Mazzoleni et al. (2017), based on the assimilation of crowdsourced data (CSD), can be generally valuable to improve real-time flood forecasts using non-traditional information now available thanks to active citizens and new technologies.

However, it has to be remarked that physically based modelling of rainfall-runoff and flow routing processes has to face actual limitations ascribed to the paucity of measured data, to the complexity of real environments, and to lacks in model structure and parametrization. As a consequence, (semi-)distributed rainfall-runoff models used for operational flood forecasting can provide reliable predictions at locations where calibration is performed (i.e., control points) and, at the same time, incorrectly represent system states elsewhere (e.g., discharges in upstream, ungauged tributaries).

In a context of equifinality and simplified representation of real physical processes, the accurate prediction of outflow hydrographs can be achieved even though model internal states don't match the true system states. In such cases, the assimilation of real CSD can lead to a substantially lower performance than the use of synthetic CSD would suggest, as it corresponds, in fact, to update a model using biased data (e.g., Dee, 2005; Liu et al., 2012). When only internal states (and not model parameters) are updated, or when the model suffers structural deficiencies, the assimilation of real (i.e., not synthetic) streamflow data at internal points can lead, in principle, to even worse model prediction at the outlet than no assimilation at all (Crow and Van Loon, 2006). The problem can arise due to the disjoint use of traditional and crowdsourced data, with the former used to calibrate (semi-)distributed models at control points, and the latter used only in real-time to update model states at different locations.

A possible solution is the use of ensemble based data assimilation methods to update jointly model states and parameters. An additional pragmatic recommendation is the collection of accurate measured data for a suitable period, for at least two reasons: i) to develop reliable rating curves at locations where water level CSD are planned to be collected, and ii) to calibrate and verify the model ability in describing the system states correctly at the locations in which CSD are collected.

It must be observed that, while scarce control on the collection of CSD can be exerted during significant flood events, the locations at which citizens can collect CSD of water levels is always determined a-priori, since the availability of rating curves is a necessary condition in order to convert water levels into discharges. The amount of measured data needed to develop reliable rating curves can also be profitably used to calibrate the model at those sections as well.

As a final remark, both modellers and environmental agencies should comprehensively account for the characteristics of the physical system, for model structure and parametrization, for the design of the sensors network, and for data to be used both in calibration and in operational mode.

References

Anquetin, S., Braud, I., Vannier, O., Viallet, P., Boudevillain, B., Creutin, J.-D., and Manus, C.: Sensitivity of the hydrological response to the variability of rainfall fields and soils for the Gard 2002 flash-flood event., J. Hydrol., 394, 134–147, doi:10.1016/j.jhydrol.2010.07.002, 2010.

Barbi, A., Monai, A. M., Racca, R., and Rossa, A. M.: Recurring features of extreme autumnal rainfall events on the Veneto coastal area, Nat. Hazard. Earth Sys. Sci., 12, 2463–2477, doi:10.5194/nhess-12-2463-2012, 2012.

Beven, K.: Changing ideas in hydrology: the case of physically based model., J. Hydrol., 105, 157–172, doi:10.1016/0022-1694(89)90101-7, 1989.

Beven, K.: A manifesto for the equifinality thesis., J. Hydrol., 320, 18–36, doi:10.1016/j.jhydrol.2005.07.007, 2006.

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

Ercolani, G. and Castelli, F.: Variational assimilation of streamflow data in distributed flood forecasting, Water Resour. Res., 53, doi:10.1002/2016WR019208, 2017.

Jakeman, A. J. and Hornberger, G. M.: How much complexity is warranted in a rainfall-runoff model?, Water Resour. Res., 29, 2637–2649, doi:10.1029/93WR00877, 1993.

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

Loague, K., Heppner, C. S., Ebel, B. A., and VanderKwaak, J. E.: The quixotic search for a comprehensive understanding of hydrologic response at the surface: Horton, Dunne, Dunton, and the role of concept-development simulation., Hydrol. Process., 24, 2499–2505, doi:10.1002/hyp.7834, 2010.

Mazzoleni, M., Alfonso, L., Chacon-Hurtado, J., and Solomatine, D.: Assimilating uncertain, dynamic and intermittent streamflow observations in hydrological models., Adv. Water Resour., 83, 323–339, doi:10.1016/j.advwatres.2015.07.004, 2015.

Mazzoleni, M., Alfonso, L., and Solomatine, D.: Influence of spatial distribution of sensors and observation accuracy on the assimilation of distributed streamflow data in hydrological modelling., Hydrolog. Sci. J., doi:10.1080/02626667.2016.1247211, 2016.

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

Mejia, A. I. and Reed, S. M.: Evaluating the effects of parameterized cross section shapes and simplified routing with a coupled distributed hydrologic and hydraulic model., J. Hydrol., 409, 512–524, doi:10.1016/j.jhydrol.2011.08.050, 2011.

Refsgaard, J. C.: Parametrisation, calibration and validation of distributed hydrological models., J. Hydrol., 198, 69–97, doi:10.1016/S0022-1694(96)03329-X, 1997.

Rodríguez-Iturbe, I. and Rinaldo, A.: Fractal river basins: Chance and self-organization, Cambridge University Press, Cambridge, UK, 2001.

Rysman, J.-F., Lemaître, Y., and Moreau, E.: Spatial and temporal variability of rainfall in the Alps–Mediterranean Euroregion., J. Appl. Meteorol. Clim., 55, 655–671, doi:10.1175/JAMC-D-15-0095.1, 2016.

Scorzini, A. R. and Frank, E.: Flood damage curves: new insights from the 2010 flood in Veneto, Italy., J. Flood Risk Manag., –, 1–12, doi:10.1111/jfr3.12163, 2015.

Sebben, M. L., Werner, A. D., Liggett, J. E., Partington, D., and Simmons, C. T.: On the testing of fully integrated surface-subsurface hydrological models., Hydrol. Process., doi:10.1002/hyp.9630, 2012.

5 Smith, R.: The influence of mountains on the atmosphere., Adv. Geophys., 21, 87–230, doi:10.1016/S0065-2687(08)60262-9, 1979.

Todini, E.: A mass conservative and water storage consistent variable parameter Muskingum-Cunge approach., Hydrol. Earth Syst. Sci., 11, 1645–1659, doi:10.5194/hess-11-1645-2007, 2007.

van Meerveld, I., Vis, M., and Seibert, J.: Information content of stream level class data for hydrological model calibration, Hydrol. Earth Syst. Sci. Discuss., in review, doi:10.5194/hess-2017-72, 2017.

10 Viero, D. P. and Valipour, M.: Modeling anisotropy in free-surface overland and shallow inundation flows., Adv. Water Resour., 104, 1–14, doi:10.1016/j.advwatres.2017.03.007, 2017.

Viero, D. P., D'Alpaos, A., Carniello, L., and Defina, A.: Mathematical modeling of flooding due to river bank failure., Adv. Water Resour., 59, 82–94, doi:10.1016/j.advwatres.2013.05.011, 2013.

Viero, D. P., Peruzzo, P., Carniello, L., and Defina, A.: Integrated mathematical modeling of hydrological and hydrodynamic response to
15 rainfall events in rural lowland catchments., Water Resour. Res., 50, 5941–5957, doi:10.1002/2013WR014293, 2014.