Ensemble flood forecasting considering dominant runoff processes: I. Setup and application to nested basins (Emme, Switzerland)

Authors replies to RC1:

We want to thank the reviewer for his/her assessment of our manuscript. We are glad that the topics of our paper have been found to be interesting. In the following we give our answers to the comments and recommendations that have been raised. Reviewer comments RC are **bold,** our reply AR is in *italic*. Insertions in the revised manuscript MI are underlined.

RC: I find that the presentation requires significant improvement. Particularly, in the current version the manuscript is somewhat disorganized and does not stand by itself (the authors send the reader to external references in an excessive number of times).

AR: The reviewer is right in his statement, that the current paper might not stand by itself in its current organization. As this manuscript is part I of 2 companion papers, we tried to avoid redundancies between the two manuscripts and also to avoid duplicating too large portions of the papers by Antonetti et al. (2016, 2017 and 2018). Reviewer RC2 suggests to include a diagram showing the logic of our past and recent papers in order clarify where relevant information on the adopted data. The diagram will be uploaded in our comment to RC2. This diagram will be also provided as supplementary material to the revised manuscript.

In the revised paper we will also replace some of the "external" references to the other manuscripts with short descriptions of the previously referenced topic.

MAJOR COMMENTS

RC: 1) Introduction: The organization of the first section requires some sharpening and reorganization. Section 1.1 gives too much detail on some of the approaches and very little detail about some others.

AR: In order to streamline section 1.1 and correct the imbalance with the other parts of the introduction we removed a full paragraph (P3 L 25-33 of the original manuscript). Further adjustments have been taken in accordance with the comments of reviewer RC2.

RC: 2) The presentation of the datasets and methods used in the study should be improved. The authors could focus on the following points:

a. **Description of the datasets should include their origin and resolutions. In particular,a brief description of the analyzed COSMO configurations is missing.**

b. Calibration of RGM-PRO and RGM-TRD. Given that one of the main differences between the two systems is in the way they are calibrated, the authors should provide a detailed description of the calibration process.

c. How were the uncalibrated and calibrated versions of PREVAH set up? Was the TRD-UC configuration calibrated using the observations of a single event? Did the authors use uniform parameters in the Emme catchment?

d. I find Section 2 too long. The authors could consider splitting g it into "Target area and datasets", and "Models and methods."

AR: We will accommodate the reviewer request and split the section as suggested (d). When preparing the two manuscripts we decided to put the focus of this first paper on the DRP model and on the description of COMBIPRECIP and present the details on COSMO in the companion paper. Furthermore COSMO data are also introduced as follows:

MI: "As future rainfall input, quantitative precipitation forecasts were used from NWP by MeteoSwiss, namely COSMO-E and COSMO-1, and were processed as in Addor et al. (2011). COSMO-1 has a grid spacing of 1.1 km and runs as deterministic model with initialisations every three hours. Lead time is 33 hours except for the 03 UTC run, where a 45 hour forecast is 10 available. COSMO-E is an ensemble prediction system with 2.2 km grid spacing, two initialisations each day and a lead time of 120 hours. Both COSMO-E and COSMO-1 are available for only one season and there is no prior experience in applying these models in a forecasting chain."

(b,c) A detailed description of how the two models have been configured (RGM-PRO) and calibrated (RGM-TRD), would basically duplicate the papers Antonetti et al. (2017) and Antonetti et al. (2018, published in the meantime).The reviewer is of the opinion of dealing with a "calibrated" and "uncalibrated" version of our "classic model PREVAH, while as described, RGM-PRO is a new module, configured a priori as presented in Antonetti et al. (2017), where the concepts of runoff generation are strongly oriented to the dominant runoff process approach with 5 separated storage according to the runoff types, while RGM-TRD is identical with the runoff generation module integrated in PREVAH, where one concepts fits all processes (as shown in Figure 2). As far as the calibration of the TRD (one set for the whole basin) version is concerned, yes, only one event is used. We choose this approach in order to have a setup with minimum requirements of observed discharge. This should show the potential of the TRD approach, when a single measurement campaign is available, as discussed for example in Pool et al., (2017).

We will adapt the sections on RGM-PRO and RGM-TRD to include these argumentations and, more important, we will better declare RGM-PRO and RGM-TRD as standalone modules.

MINOR COMMENTS

RC: 1) Page 4, line 13: "which requires a high model resolution" I guess that the text refers to high NWP model resolution, but it would be worth making it explicit.

AR: Will be accommodated

RC: 2) Page 5, lines 26-29: "For the Trueb catchment, measurements from the Bau-, Verkehr- und Energiedirektion of the Canton of Berne were available. For the evaluation of hindcasts, only four events are investigated as runoff data is not available from 2005 to 2010." It is not fully clear that the last sentence refers to the Trueb catchment.

AR: Will be accommodated

RC: 3) Page 6, lines 10-15: How was CombiPrecip applied in the study? Was the operational CombiPrecip product the one applied here? Were the rain gauge measurements of the Napf station blended in CombiPrecip?

AR: A reanalysis of COMBIPRECIP is available for the period 2005 to 2013. Since 2013 we receive the operational COMBIPRECIP product, which includes the station of Napf in the blending procedure.

RC: 4) Page 7, lines 1-9: It is difficult to follow the retrieval of the RTs maps. Could you provide some details about the Magreth map of SoilCom GmbH, DEM used for the Müller map, resolution. . .?

AR: The RT maps used coincides with the products developed in the Antonetti and Zappa (2018) study. As detailed in Antonetti et al. (2016) the "SoilCom" map requires more input information than the Müller Map and uses a DTM in 2 m resolution. The Müller map is obtained from a DTM of 25 m resolution. The information required by the reviewer will be added in the revised manuscript.

RC: 5) Page 7, line 31. Why is RGM-PRO event-based? Is this the only alternative?

AR: We did not want to develop a new fully fledged model, but develop something that we can setup ad hoc in any area and feed with precipitation information only, while prescribing initial soil moisture deficit, in this case from a nation wide simulation of the "donor" model PREVAH. This tool should be only active when thunderstorms are to be expected and provide information to anticipate flash-floods in fast-reacting area small areas, while for larger catchments the "established" PREVAH versions are adequate. We will integrate some of these thoughts when we first introduce RGM-PRO.

RC: 6) Page 8, lines 4-7: It is surprising that the calibration of the RGM-TRD was done using a single event. How could this affect the performance of the system?

AR: We replied to the second question above, i.e. to mimic nearly ungauged conditions. For sure calibration over several events would have led to more robust parameters.

7) Figure 4 - caption: In the figure "Uncalibrated PREVAH" is referred to as "NC".

AR: corrected, thanks

8) Figure 7 – caption: Please, specify the duration of the analysis. Is this for the event of 12 May? Over what period?

AR: It refers to all events considered events between May and July. Caption will be modified accordingly.

AR: We assume the reviewer refers to Figure 6. The analyses cover the

9) Please, refer to the accepted version of the work of Kienzler and Naef (2008): Kienzler, P. M. and Naef, F.: Temporal variability of subsurface stormflow formation, Hydrol. Earth Syst. Sci., 12, 257-265, https://doi.org/10.5194/hess-12-257-2008, 2008.

AR: corrected, thanks

References:

Antonetti, M., Buss, R., Scherrer, S., Margreth, M., and Zappa, M.: Mapping dominant runoff processes: an evaluation of different approaches using similarity measures and synthetic runoff simulations, Hydrol. Earth Syst. Sci., 20, 2929-2945, https://doi.org/10.5194/hess-20-2929-2016, 2016.

Antonetti, M., Scherrer, S., Kienzler, P. M., Margreth, M., & Zappa, M. (2017). Process-based hydrological modelling: the potential of a bottom-up approach for runoff predictions in ungauged catchments. *Hydrological Processes*, *31*(16), 2902-2920. https://doi.org/10.1002/hyp.11232

Antonetti, M., & Zappa, M. (2018). How can expert knowledge increase the realism of conceptual hydrological models? A case study based on the concept of dominant runoff process in the Swiss Pre-Alps. *Hydrology and Earth System Sciences*, *22*(8), 4425-4447. https://doi.org/10.5194/hess-22-4425-2018

Pool, Sandra; Viviroli, Daniel; Seibert, Jan (2017). Prediction of hydrographs and flow-duration curves in almost ungauged catchments: Which runoff measurements are most informative for model calibration? Journal of Hydrology, 554:613-622.