Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2018-119-RC1, 2018

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Ensemble flood forecasting considering dominant runoff processes: II. Benchmark against a state-of-the-art model-chain (Verzasca, Switzerland)" *by* Christoph Horat et al.

E. Gaume (Referee)

eric.gaume@ifsttar.fr

Received and published: 20 July 2018

The manuscript submitted by Horat and al. compares a newly developed flash flood forecasting chain based on a so-called "process-based" rainfall-runoff model to a chain that is running operationally in Switzerland. Some additional elements of comparison of the respective qualities of deterministic and probabilistic forecasts are proposed (i.e. this question was the main issue of previous publications of the same authors). The paper builds on the large experience of the authors in the field of flood and especially flash flood forecasting. The methods are clear and well-established except for the use

of the Brier score to compare deterministic and probabilistic forecasts (see hereafter). The connection with real-world operational methods and application is a very positive aspect of the presented work. To my knowledge, particularly advanced methods are implemented operationally in Switzerland if compared to the rest of the world. Nevertheless, the proposed comparison is conducted for one single watershed (and one of its sub-watersheds but the results are not presented in the manuscript) and for a period of time of 3 months only (May to August 2016). This is by far insufficient to draw real convincing conclusions. The implementation of a rainfall-runoff model without calibration can only be evaluated if conducted on a significant number of watersheds – typically some tens. Likewise, a three months period seems far to short for a faithful evaluation of forecasting models and does not correspond to the general standards of the scientific publications. The results may be too dependent on some few if not a single flood event with no possibility of generalization. The authors themselves acknowledge in the discussion part of their manuscript that sparse data may be problematic (P 18 L6). They also mention that the obtained results and skills and their variations with lead times or considered threshold quantiles are not consistent with the ones obtained in previous studies. There is, to my opinion, a high probability that part of these observed inconsistencies may be explained by the limited size of the test data set. At least, the authors have to demonstrate the robustness of their results and interpretations. The objectives and methods presented in the paper are correct, but the manuscript can hardly be published to my opinion unless a much larger data set in time and space (larger period of time and larger number of watershed is considered). The team seems to have access to reach data sets in Switzerland ; It is time consuming of course, but I do not see any reason why they could not conduct a large and necessary test and validation study based on the approach presented in the manuscript. Apart from this major issue, some other less important comments can be done on the presented manuscript:

1) The authors mostly refer to their own works. Indeed, interesting and innovative methods are implemented in Switzerland to forecast flash floods. But it would important also to cite works conducted in other countries and by other teams on the

same issue at least in the introduction of the manuscript to show the originality of the proposed approach. Flash flood forecasting has been an active field of research in the recent years. 2) The manuscript refers in many places to a companion paper and to supplementary materials. This is frustrating for the readers since some important information is not provided in the manuscript such as the implementation of the “process based” model (what are the input variables, how are the values of the parameters of the model fixed) or the results obtained for the Pincascia sub-watershed. Supplementary material is interesting but a manuscript must be to a certain extent self-sufficient and contain at least the basic information needed for the interpretations and the results that are commented and interpreted. 3) Brier scores are used to compare deterministic and probabilistic forecasts. I know that some other papers did the same, but this comparison is not appropriate. Indeed, a Brier score can be computed in both cases, but do not measure exactly the same things and can therefore not be directly compared. Forecasts must be combined with a utility function and evaluated in a decision making context for a proper and rigorous comparison. An annotated manuscript is attached to this review.

Please also note the supplement to this comment:
https://www.nat-hazards-earth-syst-sci-discuss.net/nhess-2018-119/nhess-2018-119-RC1-supplement.pdf

---

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2018-119, 2018.