Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2018-119-RC2, 2018

Natural Hazards and Earth System Sciences
Discussions

NHESSD

Interactive comment

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.

Anonymous Referee #2

Received and published: 3 August 2018

Overview

The manuscript presents an application of an event-based runoff generation module driven by forecasts provided at high horizontal resolution by two new numerical weather prediction tools (a deterministic one and a probabilistic one). The meteo-hydrological model coupling is tested for a small-size catchment (and its sub-catchment) in a mountainous area of Switzerland. The manuscript addresses relevant and interesting methods (in particular, the proposed hydrological module does not need a calibration task) to improve flood forecasting. The description of the experiments, calculations and re-

Printer-friendly version

Discussion paper

sults is clear and accurate. However, the results are based on a very limited dataset (just one summer season for the meteorological analysis and about twenty events for the hydrological analysis), that may result as not sufficient to support the interpretations and the conclusions.

General comments

(1) It is not so clear the main goal of the manuscript, with respect to the companion paper. On the one hand, if the focus of the paper is on the meteorological input (as stated by authors), therefore the dataset may results as quite limited to test the improvements of the new meteorological forecasting tools (why have authors not considered a larger period of data availability of COSMO-1 and COSMO-E? For instance, the years 2016-2018?). Actually, if the focus of the paper is the evaluation of the performance provided by the new meteorological chains, then the computation of the statistical scores could be carried out over a longer period, without the need to perform a comparison with the older forecasting chains (model benchmarking), given that the statistical scores are able to give an objective evaluation of performance for the tested meteorological forecasting tools. On the other hand, if the focus of the research is on the performance of the new meteo-hydrological chain, therefore the dataset seems to be not so significance in terms of flood events, in particular with respect to the operational aims of civil protection authorities (just one event with a 2-yr return period in the investigated dataset). The limited amount of data here used for the statistical analysis may not justify a separate manuscript with respect to the companion paper. Maybe, the investigations and results shown in the present manuscript could be synthesized (as done, for instance, in Section 6.6) and added to the companion paper in order to enlarge the statistics for the proposed coupling of RGM-PRO approach to COSMO-1 and COSMO-E.

(2) The meteo-hydrological model coupling is used as a verification tool for the new meteorological model chains at higher resolution. But, by the hydrological perspective, it seems that there are not enough high-impact events in the investigated period. More-

over, it could be questionable that the false alarms are "realistically" evaluated (Pag.8, lines 7-9), given that the investigated period is the summer season. I mean, which is the soil saturation of the study area in summer? Is summer a dry season for the study area or the soil saturation is quite high in summer so that a light/moderate rain event could trigger a flood event? Or are there floods in summer only due to extreme rainfall events which cause rapid surface runoff without rainfall infiltration? Authors should add "hydrological" details about the occurrence of flood events in summer for the selected catchments.

(3) The introduction (i.e., Section 1) could be shortened (for instance, lines 14-33 at Pag.2; lines 13-34 at Pag.3; lines 1-31 at Pag.4). Some issues are repeatedly discussed and too much detailed descriptions of past studies are provided, even though not strictly related to the contents and methodologies proposed in the manuscript. Thus, a synthesis may result advantageous. Moreover, the contents may appear as dispersive (too general) with respect to the context of the study area and the proposed forecasting methodologies. The contents of Section 1 should focus on contents which show similar features to the present study. The citation of past studies should highlight the feasibility of those approaches with respect to spatial and temporal characteristics of phenomena (for instance, catchment dimension, return time of the basin, forecast lead time), focusing on the similarities with the present manuscript. In the current form, this section seems as a general review of the flood forecasting subject.

(4) I guess that the hourly runoff climatology of the period May-August 2016 was used as reference climatology to carry out the statistical analyses (for instance, to compute the quantiles of Figures 6-8). Is the May-August 2016 runoff climatology statistically meaningful with respect to a longer climatology (for instance, some decades) for the selected study area? Section 5 provides a very detailed analysis of the performance for the tested forecasting chains. Nevertheless, it is not so evident that these performances are significantly different. Even, some scores provide outcomes in contrast to the companion paper (for instance, the process-based forecasts are not better for the

nested sub-catchment). The limited dataset may hamper a solid comparison.

(5) Sections from 6.3 to 6.6 go into a detailed analysis about comparisons of the proposed new forecasting chains with previous studies. However, models, data input, study areas and investigated period are not always the same. Therefore, the content of these sections may result too long, redundant and not so interesting with respect to evaluation of the proposed forecasting chains. The discussion recalls results and trends which are general and well known in the past specific literature (in particular Section 6.4). Authors should highlight the original contribution of the proposed forecasting chains and summarize the comments on the comparisons (for instance, authors could move each specific comments of Section 6.5 in a position within the manuscript where that issue has already been discussed, rather than devote a specific section to comment all the issues of the comparisons).

Specific comments

Pag.1, Lines 20-22: The comment on the performance of the proposed model chains should stress the feasibility of these chains with respect to the dimension of the study area. This point of view for the discussion of results could be an added value for the present manuscript.

Pag.2, Lines 1-2: This statement should be based on a larger dataset.

Pag.2, Line 11: The meteorological perspective is not so deeply investigated in the manuscript.

Pag.3, Line 2: Authors should specify the reasons for the suitability just for catchments with areas up to 1000-2000 km2.

Pag.3, Line 18: Authors should specify the country of FOEN.

Pag.3, Line 29: Authors should specify the year of the event.

Pag.6, Line 10: Authors should specify the meaning of the acronym “WSL”.

Pag.6, Line 10: Authors should specify that PREVAH is a semi-distributed hydrological model (as done in the abstract).

Pag.7, Line 9: Should "and Avalanche Research SLF." be "and Avalanche Research (SLF)."?

Pag.7, Lines 15-22: Is the configuration of the COSMO model changed with the increase of horizontal resolution (from 2.2 to 1.1 km)? Authors should add details and references about this issue.

Pag.7, Lines 23-27: Is the configuration of the COSMO-based ensemble changed with the increase of horizontal resolution (from 7 km of COSMO-LEPS to 2.2 km of COSMO-E)? Authors should add details and references about this issue.

Pag.8, Lines 1-3: This sentence is not clear.

Pag.8, Lines 6-7: Does a threshold exist to identify major flood events (namely, flood events which are of interest for the authority in charge of the public safety)? How many major flood events occurred in summer 2016 for the study area?

Pag.8, Lines 23-24: The comparison may results as not fully proper, given that the observed rainfall input is different for the two chains. Why has the same input not been used for both the chains?

Pag.9, Line 13: Is the hourly runoff climatology of the period May-August 2016 statistically meaningful with respect to a longer climatology (for instance, some decades) for the selected study area? It could be useful to show a comparison between the reference climatology of this study and a historical ones for the selected catchment.

Pag.9, Lines 14-16: Are the statistical scores computed at each hourly time step of the simulation event? I mean, is the threshold exceedance evaluated each hour and the corresponding score computed at an hourly time step, then averaged over each lead time window? Or is the threshold exceedance evaluated just one time within the whole lead time window? Please clarify.

Pag.10, Lines 10-13: The description of the Brier Score decomposition could be omitted, given that it is not discussed in the main manuscript.

Pag.11, Lines 20-21: Which is the spatial domain (Switzerland? Verzasca catchment?) over which the scores shown in Fig.3 were computed? Please specify. The scores for POD and FAR are not so satisfying, especially for the higher thresholds (namely, rainfall events which likely trigger flood peaks). The FAR scores may results quite high with respect to the usefulness for operational decisions of civil protection authorities. Could authors add a comment to justify this results? Which is the rainfall threshold that trigger major flood events for the study catchments?

Pag.12, Lines 1-2: Why are the BSS values not shown in Fig.4 (or discussed) for the thresholds higher than 10 mm (as done in Fig.3)?

Pag.12, Lines 14-15: Please specify that the cited scores refer to runoff data and the quantiles refer to the hourly runoff climatology of summer 2016 (in case of my interpretation is right).

Pag.13, Line 14: The panels "b" and "c" of Fig.8 are very friendly to convey the best performing method, but this visualization does not allow to evaluate if the difference of performance is significant in term of ROCa.

Pag.13, Lines 16-19: Authors should try to justify this result. May the reason for this result be ascribed to the larger dimension of the Verzasca catchment (with respect to its sub-catchment), which can allow to "average" spatial errors in the localization of the rainfall field?

Pag.13, Lines 21-22: The magnitude of the event (i.e., 2-yr return period) should also be cited here (as done at Pag.14, Line 4). Authors could add a comment about the frequency and amount of the flood peaks corresponding to more intense events for the selected catchments.

Pag.13, Lines 24-30: The significance of the comparison may result as weak, given that it is discussed for two forecasting chains that differ for the model and input used. Authors should add a comment in order to justify this result with respect to model characteristics and data input.

Pag.15, Line 6: With this statement authors recognize that the discussed results and conclusions may be invalidated by the limited dataset that has been used to test the proposed new forecasting chain. Actually, the usefulness and added value of the new forecasting chain are questionable due to the limited test period. The use of a different dataset could provide different (and opposing) results.

Pag.16, Lines 1-19: The detailed comment on the comparison results does not lead to a clear conclusion about which chain should be preferable. Some scores provide opposing outcomes (for instance, in contrast to the companion paper, the process-based forecasts are not better in the sub-catchment), then this analysis may result as inconclusive.

Pag.19, Line 25: Authors should specify the magnitude of these 22 events with respect to the catchment climatology.

Pag.20, Lines 1-4: This result is quite general and recalls several past studies. Here, it appears as a local application based on a limited dataset.

Pag.20, Lines 5-12: Authors recognize the major drawback of the present study. They would evaluate a new model-based approach to flood forecasting which does not require the calibration task for the hydrological module, but the available dataset is not appropriate (due to the very limited size) to prove the added value of the proposed approach.

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