Hydrol. Earth Syst. Sci. Discuss., 5, S490–S493, 2008
www.hydrol-earth-syst-sci-discuss.net/5/S490/2008/


Hydrology and Earth System Sciences Discussions



# *Interactive comment on* "The European Flood Alert System EFAS – Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts" *by* J. C. Bartholmes et al.

**Anonymous Referee #2**

Received and published: 11 June 2008

The paper presents the statistical skill assessment of the probabilistic and deterministic forecasts computed by the EFAS. The system provides flood warnings for large river basins across Europe with lead times from 3-10 days based on the deterministic and probabilistic forecasts provided by the ECMWF (along with medium range deterministic forecasts provided by the DWD). Part 2 of a series of 2 papers deals with the qualitative assessment of the forecasts provided by EFAS using different meteorological data sets as input data.

General comments:

It is the merit of this paper to apply a large range of skill tests that were originally

developed for assessing the performance of probabilistic forecasts in meteorology to evaluate the skill of the EFAS hydrological forecasts. The paper is well written and the findings are very relevant for the scientific community. In particular, the authors provide a valuable discussion on the use of persistence criteria and on the number of EPS that need to give discharge forecasts above certain thresholds before issuing reliable flood alerts. Also they discuss the relation between the skill that is achieved by EFAS using different meteorological input data. Although it can be argued that the selected performance criteria were already used in similar skill assessment studies and that the innovative aspect of the paper might therefore be rather limited, I believe that due to the scale of the study area and the discussion of the impact of various input data sets and various constraints applied to these data sets, this paper is an important contribution in the field of medium range hydrological forecasting. There is no doubt that the pragmatic approach that the authors chose to assess the skill of their forecasting system can be criticized from a purely statistics point of view. I found that the approach is appropriate here, because it was the goal of the paper to demonstrate the skill of different products derived from EFAS. Since many of these EFAS products are based on some sort of thresholding (alert yes or no) it seems unavoidable to reduce the continuous variables to binary events to process the analysis of these products.

Since the aim of this study was to critically assess EFAS, I was missing in this paper some sort of benchmarking that would allow evaluating the performances of EFAS compared to other flood forecasting systems. Since comparable results have been published for other regional forecasting system it might be interesting to compare the skills obtained by EFAS with those obtained by other systems (cf. list of studies given in the introduction). Moreover, I feel that some of the terminology that stems from meteorology needs to be better explained in order to make the reading a little bit more straightforward for the hydrological community. A large amount of skill tests are presented but since the hydrological community might not be so familiar with these methods adopted from meteorology, it could be difficult for many of them to assess (and appreciate) the performance that was achieved by EFAS over the 2 years. The authors



HESSD

5, S490–S493, 2008

argued that the Nash Sutcliffe criterion is not a suitable criterion for evaluating probabilistic forecast performances. They are of course right but due to the wide spread of this particular performance measure it would have been nonetheless interesting to present the Nash values that were obtained – for example – by the ensemble mean. As the authors mention it themselves, there is a lot of “number-crunching” in this paper. At times this makes the reading quite difficult and it’s easy to get lost in the sheer amount of setups and constraints that the authors evaluated. Sometimes less is more and the paper would benefit from being a bit more concise. Is it really necessary to present all skill tests given that many of them provide very similar results? If you want to keep all of them, please specify the added value of each one.

Overall, I found this an interesting paper to read and I recommend it for publication in HESS subject to some minor revisions.

Specific comments:

Introduction: I found the introduction very much focussed on different performance criteria. Since it is the aim of the authors to assess the skill of the EFAS with different input data rather than to provide a new indicator for assessing the performance of a forecast, it would be more interesting to report the skills of forecasting systems presented in literature. This would help to put the skills that they computed for EFAS in a more general context. I suggest to specify the results of skill studies that the authors mention on p. 291 (l. 3-8). These could serve as a kind of benchmark for EFAS to which the skills found for EFAS could be compared. This would enable a true assessment of the performances achieved by EFAS. In the introduction I would also give some more details on the meteorological input data that were used. What are the skills of these products with respect to predicted rainfall amounts? I think it is necessary to better evaluate the quality of the input data before using them in hydrological forecasting. This is important in order to evaluate the rankings that they establish for the hydrological forecasts using different input data sets.



HESSD

5, S490–S493, 2008

p. 293 l. 21 please clarify what the contribution of this paper is compared to the EFAS skill study of Bartholmes et al. 2006

p.293 l. 27: using the same kind of meteorological input data any regional or national service could provide medium-range flood forecasts. Hence, provided the same data sets are available (and normally they are), one could argue that national or regional hydro-meteorological services could also provide better medium-range forecasts than EFAS (since they should have more complete data sets for model calibration).

p. 296 l.7 I feel that more explanations are needed with respect to ECMWF’s EPS products.

p. 298 l. 4 please explain in more detail what you mean by “climatology”

p. 304 in Fig. 3 you give the absolute numbers of the contingency table. But could you also briefly mention how many HAL and SAL per pixel were observed in average.

p. 305 l. 5 I found it surprising that the skill using DWD data is in general smaller than the one obtained with ECMWF data. I was especially surprised by the explanation given by the authors. They claim that the ECMWF resolution is more similar to the JRC-MARS and that due to this, the skill of EFAS forecasts with ECMWF data might be better. Didn’t you do any resampling of the DWD raster to make it compatible with the Lisflood grid? Was there no averaging of the rainfall predictions over each Lisflood cell? Why should the lower resolution product necessarily provide better skills?

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 5, 289, 2008.