Hydrol. Earth Syst. Sci. Discuss., 8, C2766–C2775, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C2766/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Skill assessment of a global hydrological model in reproducing flow extremes" *by* N. Candogan Yossef et al.

N. Candogan Yossef et al.

ncandogan@hotmail.com

Received and published: 14 July 2011

Comment: I applaud the authors for their efforts and read some of the paper findings with great interest. However, I feel that the paper is premature and a more in depth analysis of the results is needed and in particular a clearer scientific justification for the novel contribution.

Reply: We thank Anonymous Referee 1 for his interest in our findings and for his useful comments. We have made changes and additions to our revised paper in response to some comments. With regard to some other comments, we would like to clarify our position and hopefully remove any remaining concern.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

We would like to clarify our scientific justification for the novel contribution of our study. There are several studies which compare discharge simulations of GHMs and LSMs to discharge observations. (The expression macro-scale hydrological model (MHM) is replaced in the revised manuscript by global hydrological model (GHM) after the comments of Anonymus Referee 2.) The novelty of this work is that it applies skill measures used in the verification of deterministic meteorological forecasts, to assess the ability of a GHM in reproducing anomalous flows and past flood and drought events, in the prospective context of operational hydrological forecasting. The explanation of the novelty of this work in the introduction has been modified to emphasize this prospective context.

Comment: I also believe that the current scientific methology is partially floored inthe way uncertainty is treated (not) and skill is evaluate. The reason for some of the results are unclear (E.g. bias vs non bias correction). The methodology is floored as the paper simply ignores any form of uncertainty be it in the hydrological model (e.g. application of Darcys law on a 55km grid??), the observations (discharge measurements have errors between 10-30% and more!), the forcings (partially done) or the post-processing. This is a major floor of the study andneeds to be rectified.

Reply: Here we would like to clarify our approach towards uncertainty. We acknowledge that there are several sources of uncertainty which reduce the accuracy of our simulations and thus affect our skill assessment. These sources include “errors in model structure, forcing and parametrization” as we mention in the introduction of the discussion paper, but of course also errors in the discharge measurements, which can be quite high as the referee points out. We have modified the introduction in the revised paper to include observation errors.

In this study we use the best available forcing and discharge observations, where all values are known a priori (in contrast to actual forecasts). We assess the maximum attainable skill that can currently be achieved by PCR-GLOBWB and similar GHMs or LSMs, given the inevitable errors from all sources. We do not attempt to quantify

HESSD

8, C2766–C2775, 2011

the contribution to the total uncertainty by different sources of error, but rather we quantify the forecasting skill. The ultimate purpose of forecasting is to decrease the uncertainty inherent in future events. The skill that we try to assess is the potential ability to decrease this uncertainty, i.e., compared to the data mean or climatology.

Since this paper aims to assess the prospect of using a GHM for operational seasonal forecasting, we mention that an additional source of uncertainty will be introduced when meteorological forecasts are used as forcing and this uncertainty will increase with increasing lead times. In that stage it will be necessary to perform a full uncertainty analysis. We intend to carry out this exercise in our upcoming paper where we use actual meteorological seasonal forecasts, but it is beyond the scope of the present paper.

Comment: p3473/section 2.1: I cannot see from this section how this model is different in any respects from many of the LSM models quoted in the introduction - please clarify that in the text.

Reply: In Section 2.1 we describe the global hydrological model PCR-GLOBWB and refer to Van Beek and Bierkens (2009) for a more extensive description but we indeed do not point out how it is different from other GHMs or LSMs. Actually two strong points of PCR-GLOBWB are its groundwater component and routing scheme. We argue in the discussion paper that our findings can be extended to other GHMs, given the similarities in model structure, parameterization and forcing dataset, as well as their performance in reproducing past hydrographs. We have modified the revised paper to extend our conclusions to LSMs as well. The referee rightly points out in Minor Points, that “just because of different focuses of the two approaches to hydrological modelling does not mean that results with respect to streamflow are different (in particular on a monthly scale)”. The performance of PCR-GLOBWB in reproducing streamflow is indeed comparable to other LSMs, provided that they incorporate a routing scheme, as shown by Sperna Weiland et al. (2011). We have updated the reference to this paper in the revised manuscript.

HESSD

8, C2766–C2775, 2011

Comment: Why is the model run on daily timesteps, when most of the analysis focuses on monthly data?

Reply: We use monthly data in our analysis and the resulting forecasting skill is for seasonal and monthly forecasts rather than medium-range. This is because our eventual goal is to develop a forecasting system on monthly/seasonal lead times. However we run the model with daily forcing because many hydrological processes are non-linear in input and state dependency, and they cannot be accurately represented on a monthly time-step without introducing additional errors.

Comment: p3476/3.1 definition of skill. I believe the authors make it too simple for proving that there is 'skill' - skill reference should be adequate (in particular in comparison on a monthly scale and the effort of running such a global model). Monthly discharge can be computed by a simple water balance model without routing (simply dumping the P-E of each catchment to the outlet including the bias correction). Please, see e.g. discussions by Schäfli and Gupta.

Reply: We absolutely agree that skill should be adequately defined in relation to an appropriate reference. For some modelling studies this reference could be a benchmark model, or as the referee points out an estimation such as simply dumping the P-E of each catchment to the outlet.

The performance of PCR-GLOBWB was compared to other methods of runoff generation by Sperna-Weiland et al. (2011). According to this study, LSMs with a comparable level of complexity could be as useful as large-scale hydrological models, provided they are tuned to reproduce realistic water partitioning at the grid scale and a routing scheme is added. For large continental basins, methods which lack discharge routing result in too high peak flows. Routing and temporal storage in a groundwater reservoir introduce a necessary delay, realistic travel times, more constant baseflows and reduced extremes.

We want to use the results of the present study as benchmark in our forthcoming

HESSD

8, C2766–C2775, 2011

studies where we assess the model performance when it is driven by meteorological forecasts. In the present study skill is not defined relative to a benchmark model. In contrast we use skill scores which define skill relative to a zero-skill forecasting system. We believe this is a more appropriate reference for our case because our purpose in skill assessment is to evaluate the prospect of forecasting extemes. We want to see which types of forecasting are more promising. We conclude that there is skill because the skill scores that we use are fair verification measures and they indicate the presence of skill.

Comment: P3477/3.2 Your choice of binary scores suffers from the large number of correct rejections - hence i find it very laudable that you publish the contingency tables. However discuss your results with respect to that issue (follow discussion of Finley affair, Murphy1996 and Stephenson, 200) - you actually note that effect in your results, but given not enough discussion in context.

Reply: There is indeed a large number of correct rejections in the binary contingency tables as expected for rare events. However, Peirce's skill score (PSS) which we use for the verification of binary forecasts, is an equitable score. The criterion of equitability is based on the principle that random forecasts or constant forecasts of the same single category receive a no-skill score. Thus the unjust claim of skill (such as in Finley's tornado forecasts) is taken care of through an equitable score. With these scores, forecasts of rare events can be verified fairly and the large number of correct rejections is not a problem anymore.

We discuss the criterion of equitability in section 3.2 where we explain our choice of Gerrity Scores (GS) for the verification of categorical forecasts. So we do not repeat the discussion in section 3.3 for binary forecasts. However, in the revised manuscript, we have added a reference to the discussion of equitability also in Section 3.3 where we explain why we chose PSS for the verification of binary forecasts.

Comment: p3479/3.3 Despite your claims, I cannot see how your evaluation is realis-

HESSD

8, C2766–C2775, 2011

tically flood and drought related. Please substantiate your claim that the 5year return period "provides and acceptable common measure". I have fundamentally no problem with such an evaluation, however, your are completely overstating the significance of your results w.r.t to floods and droughts. Please properly explain how you have derived the return period and also indicate what the results are with other return periods. In addition, I am not sure I understand your numbers, e.g. in a 20 year data set of the Amazon, you seem to have a total of 336 events (table C1). There are only 240 months and you state p3480/L6 "we limit ourselves to monthly..." - please clarify as you presumable used daily data here.

Reply: We thank the referee for his comment. It made us realize that our expression “provides an acceptable common measure” is quite misleading and we have consequently removed it in the revised paper. These words give the wrong impression that we claim that a 5-years return period is a definition for floods and droughts which is commonly accepted by scientists. However this is not at all what we meant. Rather, we chose a 5-years return period because it was “acceptable” to us as a “common measure” for our selected basins and for both flood and drought events. As we explain in section 3.3, our decision to use 5-years return period was a compromise between the rareness of the events on one hand and the limited record of discharge observations on the other.

The referee correctly points out that we should explain how we have derived the return period. It has been added to Section 3.3 that for calculating the 5-year flood and drought discharges, the Annual Maximum Series method has been used.

Since the referee requests that we indicate the results for other return periods, we have included an appendix where we present the skill for 10-year return periods for the two basins with the longest records, the Danube and the Mississippi.

The referee mentions that the Amazon data set is 20 years and hence 240 months in contrast to 336 events in Table C1. This comment alarmed us that we had made a mis-

HESSD

8, C2766–C2775, 2011

take, but to our relief we saw that the record length given in Table 1 for the Amazon is 28 years, which adds up to 336 months. So, all the numbers in Table C1 of the discussion paper correspond to monthly events. No daily data were used. The contingency tables are presented in the revised manuscript as Table 3, 4, etc.

Comment: p3480+Appendix A. I do not see the evidence for the assumption that daily maxima necessarily transform to monthly maxima. I do not understand your justification for the extrapolation from the Rhine - and I am not sure why it is necessary (rather than just stating this limitation).

Reply: We agree that daily maxima do not necessarily transform to monthly maxima and we accept this limitation by stating in Section 3.3 that "for some rivers a monthly time scale may seem to be too coarse to correctly predict flood sizes". However we claim that for large rivers threshold exceedances of monthly discharge will be "indicative for increased probability" of floods. We support this claim for the Rhine by the results shown in Figs. A1 and A2. Fig A1 shows that "extreme daily discharges almost always coincide with large monthly discharges". Fig. 2 shows that "for most of the years, the month in which the annual maximum daily discharge occurred is also the month of maximum monthly flow". We used the Rhine as an example because it is the smallest of the 20 global rivers in this study and it has a complex regime. We assume that if it can be proven that this relation holds for the Rhine, it would hold for other larger basins as well. Unfortunately the same exercise can not be repeated for all the basins because of the lack of long-term daily records.

Comment: Section 4.1: most of your skill improvement comes through the bias correction. At this stage I am unconvinced that there is any value in running a complex hydrological model.

Reply: We would like to try to convince the referee about the value of running a complex hydrological model for forecasting purposes. It is true that our simulations are biased and some of the possible reasons are discussed in Section 4.1. However, for

HESSD

8, C2766–C2775, 2011

forecasting the extremes, what is important is the ability to predict high and low values of river discharge at the correct time. It is for this reason that even for basins where simulated hydrographs are biased and do not outperform the climatology, acceptable to high skills are attained in forecasting monthly anomalies and extreme events.

It should be emphasized that we use uncorrected simulations in reproducing monthly anomalies and extreme events. For the categorical and binary contingency tables, thresholds for observations and simulations are calculated separately; so systematic errors are dismissed and we use the simulation results without bias correction.

Comment: Please plot time series for a station where there is no improvement and one where there is a large improvement. Please also plot the climatology you are using. Currently, it is difficult to make heads and tails of these results and in particular understand why the bias correction brings such an improvement.

Reply: The time series for all 20 basins are plotted in Figs. 2 and 3, uncorrected and bias corrected respectively. Also Fig. 4 shows the reliability diagrams where the improvement through bias correction can be seen clearly. One of the basins with the highest improvement is Lena. It serves as a good example for our argument that if the model has the ability to predict high and low values of river discharge at the correct time, a simple post-processing can take care of the systematic bias and the model is useful for forecasting. In the reliability diagram for Lena, we see that most bias corrected (X) as well as uncorrected (O) values cluster around the 1:1 line, whereas for the two months of the year with highest discharge model simulations are strongly overestimated (O) and bias corrected (X) values are brought towards the 1:1 line.

The climatology that we are using is the long term mean of the available monthly discharge records for each of the 12 months of the year. This is explained in Section 3.1.

Comment: p3483/4.2 I liked this section and think a more in depth analysis could actually be very interesting.

HESSD

8, C2766–C2775, 2011

Reply: We thank the referee for his comment. A more detailed analysis of what we discuss in this section can be found in the paper by Van Beek, Wada and Bierkens (2011).

Minor points:

Comment: I believe that the paper is not set properly into context. There have been multiple publications of the LSMs mentioned in the introduction with respect to their capability in reproducing streamflow. None of these studies is mentioned although they are highly relevant. Just because of different focuses of the two approaches to hydrological modelling does not mean that results with respect to streamflow are different (in particular on a monthly scale). The paper needs to mention these studies and explain the added value to the scientific community (please see a very limited number of references of recent papers below Gong, Papenberger, Yamazki (all this year - amongst many others!)). In addition, there are multiple papers using meteorological skill scores within a hydrological setting (Laio and Tamea, 2007 amongst many others - also see specialissues of HEPEX) thus I cannot see how the contribution is novel in this context. Having said that the way those measures are applied are very interesting and could be the focus of this paper.

Reply: In the revised paper, we have included the reference to the most relevant of recent publications on LSMs suggested by the referee. Among other studies in which the discharge simulations of other GHMs and LSMs have been compared to discharge observations, the novelty of this work is to evaluate the ability of a GHMs in reproducing anomalous flows and past flood and drought events with skill measures used in verification of deterministic meteorological forecasts. Here we would like to thank the referee for appreciating the way these measures are applied. We think the focus of this paper is indeed this application which is a preliminary assessment to evaluate the prospect of operational forecasting.

Concerning the prior use of meteorological skill scores in hydrological applications, this

HESSD

8, C2766–C2775, 2011

mainly concerns the verification of ensemble hydrological forecasts in operational forecasting mode. The studies within HEPEX Hydrological Ensemble Prediction mostly use the Brier Skill Score or other scores developed for ensemble forecasts. Since we conduct a preliminary skill assessment with observed meteorological input, we use scores intended for deterministic categorical and binary forecasts. In our forthcoming studies however we wish to conduct retroactive and operational forecasts using ensemble monthly/seasonal meteorological forecasts with different lead times. To that end we will use ensemble skill scores.

Comment: p3490 Table1: Where do these basin data come from? Please quote reference.

Reply: The reference for the data on Table 1 has been added to Section 2.3 in the revised manuscript.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 3469, 2011.