Hydrol. Earth Syst. Sci. Discuss., 8, C2787–C2797, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C2787/2011/

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: This paper investigates the ability of a global hydrological model PCR-GLOBWB to reproduce flow extremes using the example of large river basins located all over the World. The model performance is evaluated using different efficiency measures and skill scores. The paper is well written and features a clear structure. However, the methods and terminology applied are not common in the community addressed by the title which makes the paper initially difficult to access. According to the introduction, the main objective is to evaluate the potential skill of the model to forecast flow extremes. Although stated by the authors that the prospects for forecasting hydrological extremes are positive, the results lag behind expectations. The main strength

of the contribution lies in the application of skill scores for model validation (or testing), which are rarely used in hydrological modelling yet are an interesting way of judging our models' performances.

Reply: We thank Anonymous Referee 3 for referring to our paper as well written and featuring a clear structure, and for pointing out the main strength of the contribution which is the application of the chosen skill scores for skill assessment. We also thank him for his useful comments that helped us improve our paper. Here, we address his comments point by point, hoping to resolve all the issues that he raises.

Comments chapter 1

Comment: Global hydrological and land surface models are introduced in chapter 1. Here, a discussion on studies focussing on hydrological extremes is missing, e.g. Lehner et al. 2006, Hirabayashi et al. 2008. (Hirabayashi, Y., Kanae, S., Emori, S., Oki, T. & M. Kimoto (2008): Global projections of changing risks of floods and droughts in a changing climate. Hydrological Sciences Journal, 53(4), 754-772.)

Reply: In the revised manuscript, we have included the study by Hirabayashi et al. (2008) in the introduction. The study by Lehner et al. (2006) was already mentioned but missing in the reference list. Now it has been added.

Comment: p. 3472, line 14-16: The sentence stated by the authors is rather subjective and the results obtained in this study, however, cannot be generalised.

Reply: We acknowledge that the skill scores obtained in this study are only representative for the GHM used, PCR-GLOBWB (According to the suggestion of Referee 2, the term macro-scale hydrological model, MHM has been replaced with global hydrological model, GHM.). Also, we understand that the term "maximum skill" may be confusing. We do not suggest that a different model may not perform better. Rather, we mean that forcing by an a priori known dataset without forecasting uncertainty will return skill scores that denote the upper limit to the performance of the present model. Notwith-

standing, we are of the opinion that given the overall similarity in parameterization and model structure, the results of this study are amenable to other forecasting systems that employ GHMs. To stress this point, we refer to the comparison of PCR-GLOBWB with other model results in Van Beek et al. (2011) and, more generally, to the findings of the WATCH intercomparison study.

Comments chapter 2

Chapter 2.1.

Comment: PCR-GLOBWB operates on a 0.5 by 0.5 degrees resolution but sub-grid variability is taken into account. Does this mean that soil type, land cover and other input data are considered for model calculation on a higher resolution? Is runoff generation calculated on sub-grid level? On the other side the river routing is based on DDM30.

Reply: To represent the influence of sub-grid variability on the hydrological response, PCR-GLOBWB divides the area fractionally into different land cover types. In the present study, these are the open water surface, comprising rivers or lakes, and short and tall vegetation. These vegetation classes are different in their below- and above-ground characteristics and have been distinguished given their different behaviour with respect to interception and evapotranspiration. For these surfaces, the water balance is evaluated over unit area. Underlying this water balance are effective values, aggregated in different ways, that were derived from global datasets on soil (FAO) and land cover (GLCC), as well as elevation (Hydro1k). All relevant variables are described for these surface areas (e.g., soil moisture) but represented at scale of the overarching cell size (0.5 deg) as the weighed averages of the respective fractions. Subsequently, the composite runoff is passed on to the drainage network where it is routed at the general resolution (DDM30 at 0.5 deg) using process descriptions pertaining either to channels, lakes or reservoirs (not included here). Additional information on the model concepts, process descriptions and parameterization of PCR-GLOBWB can be found

in Van Beek & Bierkens (2009).

Chapter 2.3.

Comment: The simulated daily river discharge is summed-up to monthly values for any further analysis. The question which then arises is why this advanced approach of temporal downscaling of climate input was chosen. At least for some of the GRDC stations observed daily discharges are available. Here a comparison between daily and monthly results would be highly interesting.

Reply: We run the model with daily forcing because many hydrological processes are non-linear in input and state dependency, and they can not be accurately represented on a monthly time-step without introducing additional errors. On the other hand, no GHM has been tested on daily data, to our knowledge, without basin-specific calibration. For this reason, although our GHM runs on a daily time-step, and even for the basins for which daily records are available, we restrict our analysis to monthly data.

Concerning hydrological extremes, for forecasting high or low flows, we believe a monthly time-step is appropriate since these extremes are determined to a large extent by persistent characteristics. A daily time-step seems to be more suitable for floods, whereas a monthly time-step is certainly more appropriate for droughts. It can be argued that a monthly time step is too coarse to correctly predict flood sizes. However we demonstrate in Appendix 1 that monthly high flows will certainly be indicative for increased probability of floods for large rivers. The skill we want to assess is the skill in forecasting increased probabilities of flow extremes rather than exact discharges, and we want to make these forecasts on monthly/seasonal lead times. For this reason we believe a monthly time-step is more appropriate and more promising.

Comments chapter 3

Chapter 3.1.

Comment: In large-scale modelling the usage of correction factors is common. How-

ever, does it make sense to bias-correct station discharges? With the method applied the water balance is not closed, i.e. discharges do not correspond to any other state variable in the upstream area. Thus the bias-corrected discharges are only valid at the given location and have no significance for the rest of the river basin.

Reply: We agree with the referee that bias-correcting station discharges does not correspond to any other state variable upstream and has no significance for other locations. However we think that in our case bias-correcting the simulation results makes sense and we would like to try to convince the referee that this a posteriori correction is justifiable. First of all, we would like to emphasize that we use uncorrected data for the skill assessment in reproducing monthly anomalies and extreme events. For the verification of categorical and binary hindcasts, thresholds for observations and simulations are calculated separately so systematic errors are eliminated and we can use the simulation results without bias correction.

For the skill assessment in reproducing hydrographs, we use uncorrected results as well as bias-corrected results, and we apply the verification methods on both sets of results. This serves two purposes. Verification with uncorrected data gives the reader the opportunity to compare our simulations with the results of other studies which use uncorrected data. Verification with bias-corrected data, on the other hand, provides an indication of maximum skill that can be achieved when the systematic bias is eliminated (whether it is due to model errors or forcing). It is relevant for the assessment of forecasting skill, which is the ultimate purpose of this study. It is also necessary for the sake of consistency, since the two following types of skill assessment eliminate systematic errors inherently. We have further elaborated the reasoning for the use of bias correction in the revised paper.

Comment: p. 3476, line 16: the term climatology is confusing as it refers to monthly mean discharge.

Reply: We prefer this term since it is common in the context of forecasting and is

inherent to the definition of the skill scores that we apply.

Chapter 3.2

Comment: In this chapter GS is introduced but the description should be more detailed.

Reply: We have included the most important properties of the GS, how it is calculated and why we have chosen this score. We think that this amount of information is sufficient given the scope of this study. However if there is any other specific information on GS that the referee thinks should be included we will certainly do so.

Chapter 3.3

Comment: The authors should describe how they derived the 5-yr return levels for floods and droughts.

Reply: It has been added to Section 3.3 of the revised manuscript that for calculating the 5-year flood and drought discharges, the Annual Maximum Series method has been used.

Comment: p. 3480, line 27 to p. 3481, line 12: This paragraph should be shortened and only concentrate on PSS and why it was chosen.

Reply: In explaining why we chose PSS, we prefer to include a discussion of other available scores as well, because in this way we can justify our choice, especially since all scores including PSS have shortcomings.

Comments chapter 4

Chapter 4.1

Comment: p. 3481, line 24ff: I don't observe reasonable agreement in the graphs shown in figure 2 using uncorrected results which is also confirmed by the efficiency criteria listed in table 2. Additionally, it becomes visible in figure 4 as well.

Reply: We have removed the assertion that the non bias-corrected simulations are in

reasonable agreement with the streamflow records for most river basins, because this is somehow subjective.

Comment: p. 3482, line 19-20: A better agreement after correction should be the purpose of this exercise.

Reply: The main purpose of this exercise is to test the maximum skill that can be achieved when the systematic bias is eliminated, so that we have an indication of the potential skill which we may expect in actual forecasting, after a simple post-processing.

Chapter 4.2

Comment: The thresholds of Q25 and Q75 set for anomalous flows are rather low as 50% of all values are actually considered "anomalous". Furthermore, how do the observed flow anomalies relate to precipitation anomalies of the given month (and the previous month)? I get the impression that precipitation anomaly compared to mean climatology is a strong predictor for the occurrence of anomalous flows as defined here which would also true for any forecast.

Reply: In the revised paper, we have repeated the analysis of anomalous flows for the 10th and 90th percentiles for all basins. Concerning the relation between flow anomalies and precipitation anomalies, we think it would certainly be interesting to test and see to what extent a precipitation anomaly is a predictor for a flow anomaly. We would like to mention here that the added value of running an GHM has been tested by Sperna-Weiland et al. (2011) in their upcoming paper under review at the Journal of Hydrometeorology. In this study they compare the performance of PCR-GLOBWB to other methods of runoff generation and they conclude that GHMs and LSMs (with a comparable level of complexity, and a routing scheme) produce more realistic results for large continental basins than methods which lack discharge routing (resulting 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.

Chapter 4.3

Comment: The outcomes of the PSS reflect again, that the model is able to reproduce floods better than droughts. The methodology applied to assess the skill in reproducing floods and droughts (PSS) is reasonable but comprises some weakness which should be discussed as well. Here, the weakness of PSS is hidden in the second term of the equation (eq. 4) in which the wrong simulation results are considered. Due to the fact that the number of events not simulated and not observed (i.e. no-no combination in Table C1) is rather high, this second term will be always very small. Consequently, this term shows almost no effect on the PSS results which means that the wrong simulation results are not reflected. For example, Table 4 indicates a perfect score of 1 for floods for the Mackenzie but Table C1 also shows that three more flood events were simulated. Thus, the PSS might not be the right choice here.

Reply: We entirely agree with the referee about the shortcomings of PSS, and we discuss these weaknesses in section 4.3. Some of the issues with PSS are mainly due to short discharge records which result in no hits, misses or false alarms in the contingency tables. This is most evident in the cases where the value for either misses or false alarms is zero. The case of Mackenzie pointed out by the referee, is such a case where there are three false alarms but no misses. Here the PSS wrongly indicates perfect skill. We mention this shortcoming in our discussion. If there was even one miss in the contingency table for Mackenzie, the PSS would have been less than one. With longer records and increasing data points, this shortcoming tends to disappear.

As we explain in our reply to the referee comment on Section 3.3, we pay special attention to the justification of our choice of PSS. For this reason we consider alternatives and eliminate them, leaving us with PSS as the best available score in spite of its weaknesses.

Comment: p. 3484, line 5: What is meant by an unskilled forecasting system? This comparison is useless as a skilled system should be always better than an unskilled

one.

Reply: We use the term unskilled system, since it is common in the context of forecast verification and is inherent in the definition of the skill scores that we apply. It refers to a purely hypothetical system with absolute zero-skill (e.g., a system which forecasts randomly, or constantly the same value, or always the climatology etc.). In the revised paper we have removed the expressions which may be confusing.

Comment: p. 3484, line 20: reasonable to high skill is achieved within the framework of the analysis used (PSS). A high skill in being able to reproduce floods is somewhat overstated.

Reply: Here we have removed these words from the text since we agree with the referee that this is somewhat overstating the skill in reproducing floods, given the shortcoming of the scores used for some basins.

Comments chapter 5

Comment: The conclusions given in the paper are rather subjective and limited to the analysis and scoring systems applied. Simulated discharges of four out of 20 river basins are in reasonable agreement with the observations which means that 80% of the hydrographs look quite bad. This is completely neglected by the authors. However, it is recommended to improve the processes in the model instead of using a bias-correction that affects the water balance afterwards.

Reply: We have modified the wording in our conclusions to remove the expressions which come across as subjective. We agree with the recommendation of the referee to improve model processes, which is one of the main goals of our research group. However we believe that the simple a posteriori correction of bias that we apply has its advantages as well, in being transparent, straight-forward and most importantly since our eventual purpose is seasonal forecasting.

Comment: The skill to reproduce anomalies should be discussed in the light of using

Q25 and Q75 whereas it is necessary to clarify the definition of “anomalies”. Because an anomaly is an unusual or unique occurrence, it is questionable whether the chosen Q25 and Q75 are representative. Instead, the authors should go for Q10 and Q90 (or Q5 and Q95) which would better fit the terminology.

Reply: We have repeated the analysis of anomalous flows for the 10th and 90th percentiles for all basins in the revised paper.

Comment: The authors should avoid the comparison with an unskilled system that does not exist.

Reply: We have modified the text in order to avoid confusion.

Comment: The conclusion derived to apply the model for forecasting hydrological extremes seems to be overstated as this cannot be observed from the results given in the text and thus should be reconsidered. The ability to reproduce past hydrographs and extremes by other global hydrological models is well-known and published. Several studies on future projections of hydrological extremes are existent as well.

Reply: We acknowledge that simulations with PCR-GLOBWB have a certain bias and do not always outperform the observed climatology even after bias correction. Nevertheless, the results show that there is skill in reproducing monthly extremes. We believe that it is absolutely worthwhile to exploit this skill for forecasting purposes (after further testing in forecasting mode), while at the same time continuing with our efforts to improve the skill and increase the prospect for forecasting.

Comments on References

Comment: Lehner et al. 2006 is missing Please correct spelling for Vörösmarty et al. 2000

Reply: Lehner et al. 2006 has been added to the reference list. The spelling for Vörösmarty has been corrected.

Comments on Tables and Figures

Comment: Table C1 and C1 (continued) contain the same contingency tables.

Reply: We will make sure this mistake is not repeated in the revised manuscript.

Comment: Figure 2: very small and scarcely be legible. Figure 3: same as for Figure 2. Other colours should be used for better distinction. Figure 4: why are different colours used? Not explained in the text or figure. Figure A1 is redundant as it should be clear that maximum daily discharges exceed maximum monthly discharges.

Reply: We agree that Figures 2 and 3 are very small, but if they are enlarged they would cover too much page space. It has been clarified in the caption of Fig. 4 that the different colours in the reliability diagrams represent different months of the year. We also agree that Figure A2 conveys the most important information, but prefer to keep Figure A1 as well because it shows the relation between daily and monthly extremes in a more straightforward way.

Comment: Please use a consistent spelling, either “McKenzie” or “Mackenzie”, in the figures and throughout the text.

Reply: Mackenzie has been used throughout the text.

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