Hydrol. Earth Syst. Sci. Discuss., 8, 3469–3505, 2011
www.hydrol-earth-syst-sci-discuss.net/8/3469/2011/
doi:10.5194/hessd-8-3469-2011

Hydrology and Earth System Sciences Discussions

Skill assessment of a global hydrological model in reproducing flow extremes

N. Candogan Yossef[1,2], L. P. H. van Beek[1], J. C. J. Kwadijk[2], and M. F. P. Bierkens[1,2]

[1]Department of Physical Geography, Utrecht University, Utrecht, The Netherlands
[2]Deltares, Delft, The Netherlands

Received: 14 February 2011 – Accepted: 4 March 2011 – Published: 6 April 2011

Correspondence to: N. Candogan Yossef (naze.candoganyossef@deltares.nl)

Published by Copernicus Publications on behalf of the European Geosciences Union.

HESSD
8, 3469–3505, 2011

Abstract

As an initial step in assessing the prospect of using macro-scale hydrological models (MHMs) for hydrological forecasting, this study investigates the skill of the MHM PCR-GLOBWB in reproducing past discharge extremes on a global scale. Global terrestrial hydrology from 1958 until 2001 is simulated by forcing PCR-GLOBWB with daily meteorological data obtained by downscaling the CRU dataset to daily fields using the ERA-40 reanalysis. Simulated discharge values are compared with observed monthly streamflow records for a selection of 20 large river basins that represent all continents and a wide range of climatic zones.

We assess model skill in three ways. First, the general performance of the model in reproducing hydrographs is evaluated. Second, model skill in reproducing significantly higher and lower flows than the monthly normals is assessed in terms of skill scores used for forecasts of categorical events. Third, model skill in reproducing flood and drought events is assessed by constructing binary contingency tables for floods and droughts for each basin.

The results show that the model has skill in all three types of hindcasting. After bias correction the model skill in simulating hydrographs is improved considerably. For most basins it is much higher than that of the climatology. The skill in hindcasting monthly anomalies is high compared to that of an imaginary unskilled system. The model also performs better than an unskilled system in hindcasting floods and droughts, with a markedly higher skill in floods. We conclude that the prospect for using PCR-GLOBWB for monthly and seasonal hydrological forecasting is positive. Our results which we argue are representative for other similar MHMs, show that MHMs have sufficient skill for use in forecasting flow extremes.

HESSD

8, 3469–3505, 2011

1 Introduction

Macro-scale hydrological models (MHMs) simulate land surface dynamics of the hydrological cycle at a regional to global scale. These models have developed rapidly over the past couple of decades (Nijssen et al., 2001a). MHMs are comparable to land surface models (LSMs), such as H-TESSEL (Balsamo et al., 2009), ISBA-SGH (Decharme and Douville, 2006), MOSES (Gedney and Cox, 2003), NOAH (Ek et al., 2003) and SWAP (Gusev and Nasonova, 2003), which were introduced in general circulation models (GCMs) to resolve the land component and provide realistic lower boundary conditions on temperature and moisture (Decharm and Douville, 2007). Although largely similar to LSMs, MHMs focus more on modeling runoff and streamflow, as well as a more comprehensive representation of the terrestrial hydrological processes. Examples are VIC (Wood et al., 1992), WaterGap (Döll et al., 2003), LaD (Milly and Schmakin, 2002), WBM (Fekete et al., 2002), and Macro-PDM (Arnell, 1999). MHMs have been widely applied to estimate current and future continental runoff (Nijssen et al., 2001a; Fekete et al., 2002; Milly, et al., 2005), to investigate the hydrological response to global warming, (Arnell, 2004; Lehner et al., 2006; Nijssen et al., 2001b; Milly, et al., 2005) and to assess freshwater availability (Alcamo et al., 2003; Islam et al., 2007; Oki et al., 2001; Vörösmarty et al., 2000).

Given the capability of MHMs to quantify streamflow, their relevance for integrated water resources management of large river basins has been recognized (Refsgaard, 2001). Reliable and timely forecasts of extremes in streamflow can help mitigate flood and drought risks and optimize water allocations to different sectors and sub-regions. The application of MHMs could be particularly promising for developing regions of the world where no effective flood and drought early warning systems are in place. However, up to now MHMs have rarely been used for river flow forecasting, mainly because appropriate routing of river discharge is not included, and forecasting systems are limited to higher resolution national or regional domains (e.g., the European LISFLOOD system with a grid resolution of 5 × 5 km; De Roo et al., 2000).

In this paper we investigate the skill of the macro-scale hydrological model PCR-GLOBWB in reproducing past extremes in the discharges of 20 large rivers of the world that represent all continents and a wide range of climatic zones. The motivation for the paper is twofold. The first objective is to present our evaluation of PCR-GLOBWB as an initial step in assessing the prospect of using an MHM for forecasting hydrological extremes. The second one is to identify a methodology that can serve as a benchmark verification procedure for hydrological forecasting. This procedure uses methods and skill scores that were developed primarily for verification of meteorological forecasts.

Global terrestrial hydrology is simulated for a historical period from 1958 until 2001, by forcing PCR-GLOBWB with a meteorological data set produced by combining ERA-40 reanalysis (Uppala et al., 2005) and CRU data (New et al., 2000). The use of a historical meteorological dataset implies that the hydrological forecasts are not affected by forecasting uncertainty in the forcing and the propagation thereof with increasing lead times. In this sense, the results presented here are indicative of the maximum skill that can currently be achieved by this and similar MHMs given the associated errors in model structure, forcing and parameterization.

We assess the skill of PCR-GLOBWB in reproducing hydrological extremes in three ways. First, a general verification of simulated hydrographs is carried out. Second, model skill in reproducing significantly higher and lower flows than the monthly normals is assessed by constructing categorical contingency tables and applying skill scores used in meteorology for forecasts of ordinal categorical events. Third, model skill in reproducing flood and drought events is assessed by applying verification measures for forecasts of binary events, where floods and droughts are defined in terms of discharge values being higher or lower than discharges associated with a given return period.

We use discharge observations from the GRDC reference dataset which contains monthly discharges for most basins. Consequently, the forecasting skill that we assess in this study is indicative for the potential skill that could be achieved in monthly and seasonal forecasting, rather than medium-range forecasting.

HESSD

8, 3469–3505, 2011

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◂◂ ▸▸
◂ ▸
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Among other studies in which the discharge simulations of other MHMs and LSMs have been compared to discharge observations, the novelty of this work is to evaluate the ability of MHMs to reproduce anomalous flows and past flood and drought events with skill measures used in verification of meteorological forecasts.

The rest of this paper is set up as follows. Section 2 describes the MHM PCR-GLOBWB, the historical simulation, the meteorological forcing as well as the discharge data used for skill assessment. Section 3 describes the assessment of skill in reproducing hydrographs, anomalous flows and floods and droughts. Results are presented and discussed in Sect. 4, followed by conclusions in the last section.

2 Historical simulation

2.1 Hydrological model

PCR-GLOBWB (PCRaster Global Water Balance) is a macro-scale hydrological model that simulates the terrestrial part of the global water cycle (Van Beek and Bierkens, 2009; Bierkens and Van Beek, 2009). It is coded in the high-level computer language PCRaster for constructing environmental models (Wesseling et al., 1996). PCR-GLOBWB is fully distributed and operates on a regular grid with a cell size of $0.5 \times 0.5^{\circ}$ (ca. 55 km squared at the Equator). Meteorological forcing is applied on a daily time step and assumed to be constant over the grid cell. Sub-grid variability is taken into account in the representation of short and tall vegetation, open water, different soil types, saturated area, surface runoff, interflow and groundwater discharge.

PCR-GLOBWB is a "leaky-bucket" type of model that calculates the water balance for every grid cell by tracking the transfer of water between the atmosphere and the cell, through stores within each cell, and laterally, as discharge, from one cell to the next. The model calculates the storages and fluxes of water, simulates the generation of runoff and its propagation as discharge through the river network. Precipitation falls either as snow or rain depending on atmospheric temperature. It can be intercepted

HESSD
8, 3469–3505, 2011

by vegetation and added to the finite canopy storage, which is subject to open water evaporation. Snow is accumulated when the temperature is lower than 0 °C and melts when it is higher. Snow melt is added to rain and throughfall; it is stored in the available pore space in the snow cover, or reaches the top soil layer. Part of this water is transformed in surface runoff and the remainder infiltrates into the soil through two vertically stacked soil layers and an underlying groundwater layer. Water is exchanged between these layers following Darcy's law and the resulting soil moisture is subject to evapotranspiration. The remaining water contributes to lateral drainage as interflow from the soil layers or baseflow from the groundwater reservoir. The total drainage which consists of surface runoff, interflow and baseflow is routed through the drainage network of rivers, lakes and wetlands, based on DDM30 (Döll and Lehner, 2002), using the kinematic wave approach. An extensive description of PCR-GLOBWB can be found in Van Beek and Bierkens (2009).

2.2 Meteorological data set

The meteorological variables required to force PCR-GLOBWB are daily values of precipitation, evapotranspiration and temperature. In the absence of direct estimates of actual evapotranspiration, the model can be forced with values of potential evapotranspiration calculated from temperature, radiation, cloud cover, vapour pressure and wind speed.

In order to force PCR-GLOBWB with daily meteorological data at 0.5° resolution, the monthly fields of the CRU TS 2.1 data set (New et al., 2000) have been downscaled to daily fields using ERA-40 reanalysis (Uppala et al., 2005). Precipitation fields are downscaled multiplicatively while an additive correction is used for temperature. Reference potential evapotranspiration is calculated first on a monthly basis, based on monthly cloud cover and vapour pressure deficit from CRU TS 2.1 as well as radiation and wind speed from CRU CLIM 1.0 (New et al., 2002). Reference evapotranspiration is converted to crop-specific potential evapotranspiration using crop factors derived following FAO guidelines. Finally, potential evapotranspiration is downscaled multiplicatively to

HESSD
8, 3469–3505, 2011

daily values using ERA-40 temperature fields. The methodology used to calculate potential evaporation for the different land surfaces in PCR-GLOBWB and the downscaling of the meteorological data is described in detail by Van Beek (2008). The resulting meteorological data set is limited to the period from 1958 to 2001 for which ERA-40 data are available.

2.3 Simulated and observed discharge time series

The simulated discharge time series represent non-regulated, unmodified, natural flow. Twenty large river basins are selected for comparison of simulated and observed time series on the basis of two criteria. The first one is to represent all the continents, a wide range of climate zones and latitudes as well as a variety of precipitation regimes. The second criterion is the availability of observed monthly streamflow records for at least part of the period 1958–2001. Selected basins can be seen in Fig. 1 (Sperna Weiland et al., 2010a). Basin characteristics and record length are presented in Table 1.

The discharge data for most of the selected basins are obtained from the Global Runoff Data Center (GRDC, 2007). When GRDC data are not available, records from the Global River Discharge Database, RivDis 1.1 (Vörösmarty et al., 1998) are used. The period of record for the discharge values reported in the GRDC and RivDis databases varies widely from basin to basin (Table 1). Simulated daily discharges for the model grid cells corresponding to gauging stations are aggregated into monthly values, since this is the temporal resolution at which observed discharge data are available for validation. The simulated and observed discharge time series are used in the assessment of skill as described in the following section.

HESSD
8, 3469–3505, 2011

3 Skill assessment methodology

3.1 Measuring the skill in reproducing hydrographs

Model bias due to errors in the input data, model parameters, or simplifying assumptions, can highly degrade the quality of the output of a hydrological model (Hashino et al., 2007). Since our intention is to predict high and low values of river discharge at the correct time, it makes sense to bias-correct model results. In this study a simple method of a posteriori correction is carried out. For the correction of each monthly discharge, the mean bias is calculated using hindcasts and observations of the same month of other years. This bias is then removed from the hindcast monthly discharge.

The general performance of the model in hydrograph simulation is assessed in terms of verification measures used in hindcasting of continuous variables, without applying thresholds. For the purpose of general verification, the most commonly applied statistical measure, mean squared error (MSE) is calculated for each river basin. In order to judge the predictive skill, the raw MSE scores are transferred into MSE Skill Scores, (MSESS). The MSESS provide a relative measure of the quality of the simulation compared to the mean climatology as a low skill alternative hindcasting method. Here climatology refers to the long term mean of the available monthly discharge records for each of the 12 months of the year. The MSESS is defined as:

$$\mathrm{MSESS} = 1 - \frac{\mathrm{MSE}}{\mathrm{MSE}_{\mathrm{climatology}}} \tag{1}$$

The range of values that MSESS can take is $[-\infty, 1]$; with the maximum value of 1 indicating perfect skill; a value of 0 indicating a model skill equivalent to the climatology; and a negative value implying that the model performs worse than the climatology.

Additionally we used the coefficient of determination (R^2) and Nash and Sutcliffe's coefficient of efficiency (NS), which are often employed in the validation of hydrological models. These coefficients provide a measure of the model skill relative to the

HESSD
8, 3469–3505, 2011

long-term mean, and independent of the climatology. *NS* takes on the values [$-\infty$, 1] and R^2 [0, 1], with higher values indicating higher skill.

3.2 Measuring the skill in reproducing anomalous flows

In order to analyze whether the model is capable of reproducing higher or lower flows than usual for a given month, the discharge time series are transformed into categorical events defined in terms of three categories of high, normal and low flow. High flow is defined as discharge values above the 75th percentile for the month in question; normal flow between the 75th and the 25th percentile; and low flow below the 25th percentile. Thresholds are identified separately for simulated and observed discharge. This approach eliminates any systematic under or overestimation in the simulations and allows us to use the simulations without bias correction. The skill in simulating these three classes is assessed by constructing categorical contingency tables and applying skill scores used in meteorology for ordinal categorical events.

Here we use Gerrity Scores (GS) (Gerrity, 1992) which is a subset of the Gandin and Murphy (GM) family of equitable scores for deterministic categorical forecasts (Gandin and Murphy, 1992). 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 (Murphy and Daan, 1985). GM scores use a scoring matrix which represents the reward or penalty accorded to each pair of simulation and observation on the contingency table. In contrast to other equitable scores such as the Heidke skill score and Peirce skill score, the Gandin and Murphy (GM) family considers differences in relative sample probabilities of categories when according a reward or penalty (Livezey, 2003). A correct forecast of a low probability category is rewarded more than that of a high probability category. Likewise failure to forecast a rare event receives a lighter penalty than a common event.

GS and LEPSCAT scores (Potts et al., 1996) are the two subsets of the GM family, that are appropriate for the specific case of ordinal categories defined as ranges of a continuous variable such as discharge. In this study, GS are preferred since they are

HESSD
8, 3469–3505, 2011

recommended by Livezey (2003) for ordinal categorical events, on the practical basis of being more convenient to use compared to LEPSCAT. GS provide higher penalties as the discrepancy between simulated and observed classes increase. For example a hindcast of low flow receives a heavier penalty when the observed flow is high, and a lighter one when the observed flow is normal.

This score takes on the maximum value of 1 for perfect skill, and the value of 0 for no-skill. The value of GS for a categorical forecast with K number of categories is given by Eq. (2):

$$\mathrm{GS} = \sum_{i=1}^{K}\sum_{j=1}^{K} p_{ij} s_{ij} \tag{2}$$

where the relative sample frequency p_{ij} of each outcome on the $K \times K$ contingency table is multiplied by the corresponding scoring factor s_{ij} $(i, j = 1, \ldots, K)$ from a scoring matrix **S** with relative levels of rewards and penalties; and summing the values. The elements s_{ij} of the scoring matrix **S** is given by Eq. (3):

$$\mathbf{S} = \begin{pmatrix} s_{ii} & s_{ij} & \cdots & s_{iK} \\ s_{ji} & s_{jj} & \cdots & s_{jK} \\ \vdots & \vdots & \ddots & \vdots \\ s_{Ki} & s_{KK} & \cdots & s_{KK} \end{pmatrix} \tag{3}$$

$$s_{ii} = b\left(\sum_{r=1}^{i-1} a_r^{-1} + \sum_{r=i}^{K-1} a_r\right)$$

$$s_{ij} = b\left(\sum_{r=1}^{i-1} a_r^{-1} - (j-i) + \sum_{r=j}^{K-1} a_r\right); (1 \le i \le j \le K)$$

$$s_{ji} = s_{ij}$$

$$a_i = \frac{1 - \sum_{r=1}^{i} p_r}{\sum_{r=1}^{i} p_r}$$

$$p_r = \sum_{j=1}^{K} p_{rj}$$

$$b = \frac{1}{K-1}$$

3.3 Measuring the skill in reproducing floods and droughts

Floods and droughts are regarded as simple binary events defined as exceedences of threshold discharges. Decision thresholds for a basin may be defined using various hydrological and economical criteria. A comprehensive approach with verification over the full range of possible thresholds for each basin is beyond the scope of this study. Therefore, a single set of decision thresholds for floods and droughts common for all river basins is selected, that can reasonably distinguish between the usual and extreme states of each basin.

The flood and drought thresholds used in this study are calculated as discharges corresponding to 5-yr return periods for each river. The choice of 5-yr return periods for floods as well as droughts is made on the basis of two considerations. On one hand, events with return periods of a few years do not reflect the long-term variability, and do not represent unusually extreme states of a river. On the other hand the limited availability of discharge observations does not allow the estimation of rare events beyond a fraction of the record length. 5 years in this case appears to be a reasonable

HESSD
8, 3469–3505, 2011

return period for the assessment of model skill in reproducing both types of hydrological extremes observed in 20 basins, the record lengths for which are given in Table 1. These discharges do not represent the actual most critical decision thresholds for the selected basins; nevertheless they provide an acceptable common measure.

It should be noted that for most rivers a monthly time scale may be too coarse to correctly predict flood sizes. However, when we limit ourselves to forecasting monthly flows in terms of binary events, these will certainly be indicative for increased probability of floods for large rivers. It can be seen in Appendix A that at gauging station Lobith on the Rhine, throughout the years with available records during the period from 1815 to 2008, extreme daily discharges almost always coincide with large monthly discharges. When the annual maxima of daily discharge are plotted against the monthly mean discharge of the month in which this daily maximum occurred, resulting points cluster along a straight line (see Fig. A1), with daily maxima higher than monthly mean values as would be expected. Moreover, Fig. A2 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. In many of the other years, it is either the previous or next month. Since the Rhine is the smallest of the 20 global rivers in this study, and given the fact that it has a rather complex regime, one can safely conclude that the same assumption holds for other larger basins as well.

Similar to the approach used for the verification of categorical hindcasts described in Sect. 3.2, for the verification of binary hindcasts the thresholds for observations and simulations are identified separately, in order to decrease the effect of any systematic under or overestimation. The skill in simulating flood and drought events is assessed by constructing 2×2 contingency tables and applying binary skill scores. Binary contingency tables present the 2×2 possible combinations of hindcast and observed event outcomes: hit, false alarm, miss and correct rejection.

Equitable skill scores used in the verification of binary forecasts are Heidke skill score (HSS) (Heidke, 1926), Peirce's skill score (PSS) (Haansen and Kuipers, 1965), Gilbert's skill score (GSS) (Schaefer, 1990) and odds ratio skill score (ORSS)

HESSD

8, 3469–3505, 2011

(Stephenson, 2000). Two of these four equitable scores, namely HSS and GSS, are markedly dependent on sample climate. Sample climate, defined as the sample estimate of the unconditional probability of occurrence of an event is purely a characteristic of the observations with no direct relevance to skill assessment (Mason, 2003). Since dependence on sample climate makes a skill score unjustifiably sensitive to variations in observed climate and therefore unreliable, HSS and GSS are excluded in this study. The remaining two equitable scores PSS and ORSS are independent of the sample climate and recommended by several studies (McBride and Ebert, 2000; Stephenson, 2000; Göber et al., 2004). ORSS is also excluded because the value of zero in any cell of the contingency table suggests that this skill score is no longer appropriate (Livezey, 2003). PSS is preferred to other scores in this study on the basis of these considerations.

The possible values of PSS are within the range [−1, 1] and its true zero-skill value is 0. Negative values imply less skill than a random prediction. The PSS for floods and droughts for each basin are calculated in terms of cell counts of the relevant contingency tables according to the formula:

$$\mathrm{PSS} = \frac{a}{a+c} - \frac{b}{b+d} \tag{4}$$

where *a, b, c* and *d* represent the cell counts for each of the possible outcomes of hit, false alarm, miss and correct rejection respectively.

4 Results and discussion

4.1 Skill in reproducing hydrographs

The results of the historical simulation and observed discharge time series for the selected rivers are presented in Fig. 2 for visual inspection. The simulation by PCR-GLOBWB is in reasonable agreement with the streamflow records for most river basins.

HESSD
8, 3469–3505, 2011

Three groups of rivers present large discrepancy between the simulations and observations. The first group is the Arctic rivers such as the Lena and McKenzie and snow and glacier dominated rivers such as the Indus. Undercatch in the CRU snowfall amounts reported by Fiedler and Döll (2007) results in a large underestimation of the spring discharge after the start of snowmelt. The second group consists of those basins with heavy regulation and large amounts of withdrawal for irrigation and consumption, such as the Murray, Zambezi and Parana. The routing scheme in the current version of PCR-GLOBWB simulates natural discharge and does not include reservoir operations and withdrawals. Therefore the simulated natural flow on these heavily regulated rivers is in disagreement with the measured discharge. Although it is one of the most heavily regulated rivers, the Nile does not show this discrepancy since measurements of natural flow upstream of the High Dam is available for comparison. The last group consists of rivers in the tropics, which show either overestimation as in Africa or underestimation as in the Amazon. This is mostly attributable to the low station coverage over the tropics in the CRU dataset and to a lesser extent poor precipitation forecasts in ERA-40 (Troccoli and Kalberg, 2004).

The improvement in predictive skill due to the correction of bias can be seen on the discharge time series before and after the bias correction (Figs. 2 and 3), as well as the reliability diagrams (Fig. 4). It can be observed from these figures that bias correction highly improves the results. This improvement is documented quantitatively in Table 2, which shows the MSE skill scores for the selected basins, both before and after the bias correction. Table 2 shows that without a bias correction, the MSESS for the majority of basins are negative. The improvement in the MSESS due to the correction varies widely, but is quite high in general, yielding a skill higher than the climatology for most basins. The three basins where the highest skill is observed are the Yangtze, the Rhine and the Mississippi, with MSESS above 0.70. The model performs worse than the climatology in four basins. It is interesting to note that the three basins with the worst performance, namely the Niger, the Nile, and the Congo are all African rivers. The fourth basin with negative skill is the Amazon. The relatively low skill in the Amazon and

HESSD
8, 3469–3505, 2011
Skill assessment of a global hydrological model
N. Candogan Yossef et al.

other monsoon dominated basins such as the Indus and the Mekong can be explained to a certain degree by the fact that for such basins the climatology is already a good estimate of the expected discharge, so that it is difficult to perform better than that. The relatively high values of R^2 and *NS* for these basins, which are also presented in Table 2, indicate that the model performance is not poor in monsoon dominated basins provided that it is evaluated using measures independent of the climatology.

4.2 Skill in reproducing anomalous flows

A complete summary of the joint distribution of categorical hindcasts and observations for the selected basins is presented in the 3×3 contingency tables in Appendix B. These tables provide the basis for the calculation of the Gerrity Scores for each basin. As can be seen in Table 3, all the resulting values of GS are positive, indicating that the skill obtained by categorical hindcasts is high compared to that of an imaginary unskilled forecasting system.

The same three rivers with the highest skill in hindcasting exact discharges, namely the Yangtze, the Rhine and the Mississippi, have again the highest scores for categorical hindcasts (>0.60). The model performance in categorical hindcasting for the African rivers the Niger, the Nile, and the Congo is much better than in reproducing hydrographs. The lowest skill among all the basins is observed for another African river, the Zambezi, though still above the climatology. For the Amazon, where the skill in reproducing hydrographs is less than that of the climatology, we observe that the skill in reproducing anomalous flows is rather high compared to other basins. This shows that even in cases where the model simulations are biased and do not outperform the climatology in reproducing hydrographs, the skill in reproducing anomalous flows can be relatively high.

4.3 Skill in reproducing floods and droughts

The 2×2 contingency tables for flood and drought events for the selected basins can be seen in Appendix C. The PSS calculated on the basis of these tables are presented in Table 4. The resulting PSS show that the skill obtained by binary forecasts of 5-year floods and droughts is also higher than an unskilled forecasting system. The system has a markedly higher skill in forecasting floods compared to droughts.

There are no basins where the model has a negative skill in reproducing either floods or droughts; but for 7 basins, the PSS indicates no skill in drought hindcasting. This is because the PSS takes on the value of 0 when the contingency table shows no hits. For some basins the model demonstrates perfect skill in reproducing floods. This is a shortcoming of the skill score that is used. The score takes on the value of 1 in cases where there are either no misses or no false alarms. Whereas to be able to assign perfect skill, one would expect the number of both misses and false alarms to be zero.

The skill assessment in reproducing 5-yr events is not applicable to the Zambezi which has an available length of discharge records of only 4 years (see Table 1). For this basin PSS is undefined due to the absence of any observed event. Similarly in the Brahmaputra and the Ganges with discharge record lengths of 5 years 10 months and 9 years respectively, the short length of the observed discharge records affects the assessment of skill negatively, because the number of available data points is low. Notwithstanding the problems related to limited observation lengths, reasonable to high skill is achieved for floods in most basins. However, the skill is significantly lower for droughts.

HESSD

8, 3469–3505, 2011

5 Conclusions and recommendations

As an initial step in assessing the prospect of global hydrological forecasting, we tested the ability of a macro-scale hydrological model PCR-GLOBWB in reproducing past extremes in the discharges of 20 large rivers of the world. We assessed the model skill in three ways: first in simulating hydrographs, second in reproducing monthly anomalies and third in reproducing flood and drought events. The advantage of such a procedure is that it provides a more detailed assessment of forecasting skill and an insight into which types of forecasting are more promising.

For most basins, the model skill in simulating hydrographs is reasonable and improves significantly by bias correction. Bias corrected hindcasts show higher skill than the observed climatology for most basins. The skill obtained in hindcasting monthly anomalies is high compared to that of an imaginary unskilled system. The model also performs better than an unskilled system in hindcasting floods and droughts. The skill in reproducing floods is markedly higher than droughts.

The results show that although simulated hydrographs may be biased and do not always outperform the observed climatology even after bias correction, acceptable to high skills can be attained in forecasting monthly anomalies as well as floods. The prospects for forecasting of hydrological extremes are thus positive. Given the similarity of PCR-GLOBWB to other MHMs in model structure, parameterization and forcing data set, as well as its performance in reproducing past hydrographs being comparable to those of other MHMs (Sperna Weiland et al., 2010b), we argue that this conclusion is valid for most other MHMs as well.

This assessment in hindcast is a preliminary one; and it shows a potential skill given the current MHM, with a meteorological forcing based on observations. The true skill should be assessed in forecasting mode using meteorological forecasts subject to uncertainty from numerical weather prediction (NWP) models.

HESSD

8, 3469–3505, 2011

References

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., and Siebert, S.: Development and testing of the WaterGAP 2 global model of water use and availability, Hydrol. Sci. J., 48(3), 317–337, 2003.

Arnell, N.: A simple water balance model for the simulation of streamflow over a large geographic domain, J. Hydrol., 27, 314–335, 1999.

Balsamo, G., Viterbo, P., Beljaars, A., Van den Hurk, B., Hirschi, M., Betts, A. K., and Scipal, K.: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system, J. Hydrometeorol., 10, 623–643, 2009.

Bierkens, M. F. P. and Van Beek, L. P. H.: Seasonal predictability of European discharge: NAO and hydrological response time, J. Hydrometeorol., 10, 953–968, 2009.

De Roo, A. P. J., Wesseling, C. G., and Van Deursen, W. P. A.: Physically based river basin modeling within a GIS: The LISFLOOD model, Hydrolog. Process., 14, 1981–1992, 2000.

Decharme, B. and Douville, H.: Introduction of a sub-grid hydrology in the ISBA land surface model, Clim. Dyn., 26, 65–78, 2006.

Decharme, B. and Douville, H.: Global validation of the ISBA sub-grid hydrology, Clim. Dyn., 29, 21–37, 2007.

Döll, P. and Lehner, B.: Validation of a new global 30-minute drainage direction map, J. Hydrol., 258, 214–231, 2002.

Döll, P., Kaspar, F., and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, J. Hydrol., 270, 105–134, 2003.

Ek, M. B., Mitchell, K. E., Lin, Y., Grunmann, P., Rogers, E., Gayno, G., Koren, V., and Tarpley, J. D.: Implementation of the upgraded Noah land-surface model in the NCEP operational mesoscale Eta model, J. Geophys. Res., 108, 8851, doi:10.1029/2002JD003296, 2003.

Fekete, B. M., Vörösmarty, C. J., and Grabs, W.: High-resolution fields of global runoff combining observed river discharge and simulated water balances, Global Biogeochem. Cy., 16(3), 1042, doi:10.1029/1999GB001254, 2002.

Fiedler, K. and Döll, P.: Global modeling of continental water storage changes – sensitivity to different climate data sets, Adv. Geosci., 11, 63–68, 2007, http://www.adv-geosci.net/11/63/2007/.

Gandin, L. S. and Murphy, A. H.: Equitable scores for categorical forecasts, Mon. Weather Rev., 120, 361–370, 1992.

Gedney, N. and Cox, P. M.: The sensitivity of global climate model simulations to the representation of soil moisture heterogeneity, J. Hydrometeorol., 4, 1265–1275, 2003.
Gerrity, J. P. Jr.: A note on Gandin and Murphy's equitable score, Mon. Weather Rev., 120, 2707–2712, 1992.
Göber, M., Wilson, C. A., Milton, S. F., and Stephenson, D. B.: Fairplay in the verification of operational quantitative precipitation forecasts, J. Hydrol., 288, 225–236, 2004.
GRDC.: Major River Basins of the World, Global Runoff Data Centre, Federal Institute of Hydrology, D 56002, Koblenz, Germany, 2007.
Gusev, Y. M. and Nasonova, O. N.: The simulation of heat and water exchange in the boreal spruce forest by the landsurface model SWAP, J. Hydrol., 280, 162–191, 2003.
Hanssen, A. W. and Kuipers, W. J. A.: On the relationship between the frequency of rain and various meteorological parameters. Koninklijk Nederlands Meteorologisch Institut, Mededelingen en Verhandelingen, 81, 2–15, 1965.
Hashino, T., Bradley, A. A., and Schwartz, S. S.: Evaluation of bias-correction methods for ensemble streamflow volume forecasts, Hydrol. Earth Syst. Sci., 11, 939–950, doi:10.5194/hess-11-939-2007, 2007.
Heidke, P.: Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst, Geografiska Annaler Stockholm, 8, 301–349, 1926.
Islam, M. S., Oki, T., Kanae, S., Hanasaki, N., Agata, Y., and Yoshimura, K.: A grid-based assessment of global water scarcity including virtual water trading, Water Resour. Manage., 21(1), 19–33, 2007.
Livezey, R. E.: Categorical events, in: Forecast Verification: A Practitioner's Guide in Atmospheric Science, edited by: Jollife, I. T. and Stephenson, D. B., Wiley, West Sussex, United Kingdom, 77–96, 2003.
Mason, I. B.: Binary events, in: Forecast Verification: A Practitioner's Guide in Atmospheric Science, edited by: Jollife, I. T., Stephenson, D. B., Wiley, West Sussex, United Kingdom, 37–73, 2003.
McBride, J. L. and Ebert, E. E.: Verification of quantitative precipitation forecasts from operational numerical weather prediction models over Australia, Weather Forecast., 15, 103–121, 2000.
Milly, P. C. D. and Schmakin, A. B.: Global modelling of land water and energy balances, Part I: the land dynamics (LaD) model, J. Hydrometeorol., 3(3), 283–299, 2002.

Milly, P. C. D., Dunne, K. A. and Vecchia, A. V.: Global pattern of trends in streamflow and water availability in a changing climate, Nature, 438(17), 347–350, doi:10.1038/nature04312, 2005.
Murphy, A. H. and Daan, H.: Forecast evaluation, in: Probability, Statistics and Decision Making in the Atmospheric Sciences, edited by: Murphy, A. H., Katz, R. W., Westview Press, Boulder, Colorado, USA, 379–437, 1985.
New, M., Hulme, M., and Jones, P.: Representing twentieth-century space-time climate variability, Part 1: Development of a 1961-90 mean monthly terrestrial climatology, J. Climate, 12, 829–856, 2000.
New, M., Lister, D., Hulme, M. and Makin, I.: A high-resolution data set of surface climate over global land areas, Climate Res., 21, 1–25, 2002.
Nijssen, B., O'Donnell, G. M., Lettenmaier, D. P., Lohmann, D., and Wood, E. F.: Predicting the discharge of global rivers, J. Clim., 14(15), 3307–3323, 2001a.
Nijssen, B., O'Donnell, G. M., Hamlet, A. F., and Lettenmaier, D. P.: Hydrologic sensitivity of global rivers to climate change, Clim. Change, 50(1–2), 143–175, 2001b.
Oki, T., Agata, Y., Kanae, S., Saruhashi, T., Yang, D., and Musiake, K.: Global assessment of current water resources using total runoff integrating pathways, Hydrolog. Sci. J., 46, 983–995, 2001.
Potts, J. M., Folland, C. K., Jolliffe, I. T., and Sexton, D.: Revised LEPS scores for assessing climate model simulations and long-range forecasts, J. Climate, 9, 34–53, 1996.
Refsgaard, J. C.: Discussion of model validation in relation to the regional and global scale, in: Model Validation: Perspectives in Hydrological Science, edited by: Anderson, M. G. and Bates P. D., Wiley, West Sussex, United Kingdom, 461–483, 2001.
Schaefer, J. T.: The critical success index as an indicator of forecasting skill, Weather Forecast., 5, 570–575, 1990.
Sperna Weiland, F. C., Van Beek, L. P. H., Kwadijk, J. C. J., and Bierkens M. F. P.: The ability of a GCM-forced hydrological model to reproduce global discharge variability, Hydrol. Earth Syst. Sci., 14, 1595–1621, doi:10.5194/hess-14-1595-2010, 2010a.
Sperna Weiland, F. C., Van Beek, L. P. H., Kwadijk, J. C. J., and Bierkens M. F. P.: Hydrological value of GCM runoff, J. Hydrometeorol., submitted, 2010b.
Stephenson, D. B.: Use of the "Odds Ratio" for diagnosing forecast skill, Weather Forecast., 15, 221–232, 2000.

Troccoli, A. and Kallberg, P.: Precipitation correction in the ERA-40 reanalysis, ERA-40 Project Rep. Series 13, 6. ECMWF, Reading, United Kingdom, 2004.
Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Van De Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fischer, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J. -F., Morcrette, J.J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, P., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Q. J. Roy. Meteor. Soc., 131, 2961–3012, 2005.
Van Beek, L. P. H.: Forcing PCR-GLOBWB with CRU meteorological data, available at: http://vanbeek.geo.uu.nl/suppinfo/vanbeek2008.pdf, 2008.
Van Beek, L. P. H. and Bierkens, M. F. P.: The global hydrological model PCR-GLOBWB: Conceptualization, parametrization and verification, available at: http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf, 2009.
Vörösmarty, C. J., Fekete, B., and Tucker, B. A.: River Discharge Database, Version 1.1. Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham NH, USA, 1998.
Vörosmarty, C. J., Green, P., Salisbury, J., and Lammers, R. B.: Global Water Resources: Vulnerability from Climate Change and Population Growth, Science, 289, 284–288, 2000.
Wesseling, C. G., Karssenberg, D., Van Deursen, W. P. A., and Burrough, P. A.: Integrating dynamic environmental models in GIS: the development of a Dynamic Modeling language, Transactions in GIS. 1, 40–48, 1996.
Wood, E. F., Lettenmaier, D. P., and Zartarian, V. G.: A land-surface hydrology parameterization with subgrid variability for general circulation models, J. Geophys. Res., 97, 2717–2728, 1992.

HESSD
8, 3469–3505, 2011

Table 1. Basins data.

Basin	Area (km^2)	Q avg ($m^3 s^{-1}$)	Length of records
Amazon	6 915 000	190 000	28 years
Congo	3 680 000	41 800	26 years
Mississippi	2 981 076	12 743	40 years 9 months
Nile	3 400 000	2830	40 years 7 months
Lena	2 500 000	17 000	24 years
Parana	2 582 672	18 000	33 years
Yangtze	1 800 000	31 900	31 years
MacKenzie	1 805 000	10 700	16 years 4 months
Volga	1 380 000	8060	24 years
Niger	2 117 700	6000	21 years 10 months
Murray	1 061 469	767	16 years
Orange River	973 000	365	20 years 3 months
Ganges	907 000	12 015	9 years
Indus	1 165 000	6600	10 years 6 months
Danube	817 000	6400	42 years 10 months
Yellow River	752 000	2571	30 years
Brahmaputra	930 000	48 160	5 years 10 months
Rhine	65 638	2200	29 years
Zambezi	1 390 000	3400	4 years
Mekong	795 000	16 000	29 years 5 months

HESSD
8, 3469–3505, 2011

Table 2. Skill scores for reproducing hydrographs.

Basin	uncorrected			bias corrected		
	MSESS	R^2	*NS*	MSESS	R^2	*NS*
Amazon	−4.92	0.55	−0.13	−0.29	0.79	0.75
Congo	−3.83	0.27	−0.87	−0.35	0.64	0.48
Mississippi	0.40	0.77	0.68	0.72	0.85	0.85
Nile	−31.51	0.59	−4.35	−4.38	0.57	0.11
Lena	−7.81	0.62	0.52	0.40	0.97	0.97
Parana	−2.10	0.48	−1.70	0.48	0.65	0.54
Yangtze	−0.89	0.89	0.64	0.75	0.95	0.95
Mackenzie	−10.51	0.62	0.11	0.33	0.95	0.95
Volga	−0.81	0.58	0.51	0.50	0.86	0.86
Niger	−81.30	0.11	−18.62	−6.75	0.32	−0.85
Murray	−0.70	0.37	−0.45	0.32	0.48	0.42
Orange River	0.11	0.22	0.20	0.17	0.26	0.25
Ganges	0.33	0.90	0.90	0.47	0.92	0.92
Indus	−1.63	0.12	0.12	0.08	0.69	0.69
Danube	−0.04	0.68	0.38	0.50	0.76	0.70
Yellow River	−1.98	0.77	−0.49	0.57	0.79	0.78
Brahmaputra	−1.40	0.88	0.71	0.32	0.92	0.92
Rhine	0.57	0.72	0.65	0.74	0.79	0.79
Zambezi	−1.49	0.16	−1.13	0.24	0.38	0.35
Mekong	−0.61	0.85	0.82	0.13	0.90	0.90

HESSD
8, 3469–3505, 2011

Table 3. . Gerrity skill scores for anomalous flows.

Basin	GS	Basin	GS
Amazon	0.47	Murray	0.33
Congo	0.40	Orange River	0.34
Mississippi	0.63	Ganges	0.47
Nile	0.32	Indus	0.21
Lena	0.35	Danube	0.60
Parana	0.58	Yellow River	0.39
Yangtze	0.67	Brahmaputra	0.25
Mackenzie	0.29	Rhine	0.61
Volga	0.53	Zambezi	0.07
Niger	0.15	Mekong	0.39

HESSD
8, 3469–3505, 2011

Table 4. Peirce's skill scores for floods and droughts.

Basin	PS-f	PS-d	Basin	PS-f	PS-d
Amazon	0.44	0.40	Murray	0.36	0.07
Congo	0.33	0.00	Orange River	0.50	0.24
Mississippi	0.70	0.39	Ganges	0.33	0.00
Nile	0.45	0.02	Indus	0.33	0.00
Lena	0.33	0.00	Danube	0.50	0.36
Parana	0.65	0.00	Yellow River	0.50	0.25
Yangtze	1.00	0.50	Brahmaputra	1.00	0.33
Mackenzie	1.00	0.00	Rhine	0.60	0.36
Volga	0.67	0.25	Zambezi	n.a.	n.a.
Niger	0.14	0.00	Mekong	0.40	0.25

HESSD

8, 3469–3505, 2011

Skill assessment of a global hydrological model

N. Candogan Yossef et al.

Table B1. Categorical contingency tables: o: observed, s: simulated, L: low flow, N: normal flow, H: high flow.

Amazon

o \ s	L	N	H
L	53	27	4
N	35	96	37
H	1	32	51

Parana

o \ s	L	N	H
L	73	23	0
N	37	140	27
H	2	34	60

Murray

o \ s	L	N	H
L	30	14	4
N	29	46	21
H	4	18	26

Yellow River

o \ s	L	N	H
L	34	45	4
N	37	116	40
H	2	25	57

Congo

o \ s	L	N	H
L	24	40	8
N	16	101	51
H	1	14	57

Yangtze

o \ s	L	N	H
L	76	20	0
N	21	141	19
H	0	29	66

Orange River

o \ s	L	N	H
L	32	26	1
N	38	76	10
H	5	28	26

Brahmaputra

o \ s	L	N	H
L	6	6	0
N	9	29	7
H	2	7	4

Mississippi

o \ s	L	N	H
L	83	37	0
N	34	181	34
H	2	27	91

McKenzie

o \ s	L	N	H
L	24	28	0
N	19	73	10
H	3	32	17

Ganges

o \ s	L	N	H
L	18	4	2
N	18	31	11
H	2	8	14

Rhine

o \ s	L	N	H
L	59	24	0
N	25	131	25
H	1	24	59

Nile

o \ s	L	N	H
L	61	49	10
N	57	133	57
H	11	48	61

Volga

o \ s	L	N	H
L	51	19	2
N	38	93	14
H	2	26	43

Indus

o \ s	L	N	H
L	12	11	4
N	25	32	14
H	2	11	15

Zambezi

o \ s	L	N	H
L	0	9	3
N	1	14	9
H	1	5	6

Lena

o \ s	L	N	H
L	26	39	6
N	14	103	28
H	2	29	41

Niger

o \ s	L	N	H
L	11	40	15
N	6	72	52
H	2	25	39

Danube

o \ s	L	N	H
L	92	35	3
N	34	182	38
H	2	38	90

Mekong

o \ s	L	N	H
L	41	36	7
N	24	119	43
H	7	27	49

HESSD
8, 3469–3505, 2011

Table C1. Binary contingency tables for floods and droughts: o: observed, s: simulated.

Amazon

Flood: o \ s	yes	no	Drought: o \ s	yes	no
yes	4	5	yes	4	6
no	5	322	no	3	323

Parana

Flood: o \ s	yes	no	Drought: o \ s	yes	no
yes	11	6	yes	0	17
no	7	372	no	13	366

Congo

Flood: o \ s	yes	no	Drought: o \ s	yes	no
yes	3	6	yes	0	5
no	3	300	no	10	297

Yangtze

Flood: o \ s	yes	no	Drought: o \ s	yes	no
yes	4	0	yes	5	5
no	2	366	no	2	360

Mississippi

Flood: o \ s	yes	no	Drought: o \ s	yes	no
yes	7	3	yes	7	11
no	3	476	no	11	460

McKenzie

Flood: o \ s	yes	no	Drought: o \ s	yes	no
yes	1	0	yes	0	4
no	3	202	no	7	195

Nile

Flood: o \ s	yes	no	Drought: o \ s	yes	no
yes	5	6	yes	1	49
no	8	468	no	11	426

Volga

Flood: o \ s	yes	no	Drought: o \ s	yes	no
yes	2	1	yes	2	6
no	3	282	no	4	276

Lena

Flood: o \ s	yes	no	Drought: o \ s	yes	no
yes	2	4	yes	0	1
no	3	279	no	5	282

Niger

Flood: o \ s	yes	no	Drought: o \ s	yes	no
yes	1	6	yes	0	31
no	3	252	no	6	225

HESSD
8, 3469–3505, 2011

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀◀ | ▶▶
◀ | ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Table C1. Continued.

Amazon

Flood

o \ s	yes	no
yes	4	5
no	5	322

Drought

o \ s	yes	no
yes	4	6
no	3	323

Parana

Flood

o \ s	yes	no
yes	11	6
no	7	372

Drought

o \ s	yes	no
yes	0	17
no	13	366

Congo

Flood

o \ s	yes	no
yes	3	6
no	3	300

Drought

o \ s	yes	no
yes	0	5
no	10	297

Yangtze

Flood

o \ s	yes	no
yes	4	0
no	2	366

Drought

o \ s	yes	no
yes	5	5
no	2	360

Mississippi

Flood

o \ s	yes	no
yes	7	3
no	3	476

Drought

o \ s	yes	no
yes	7	11
no	11	460

McKenzie

Flood

o \ s	yes	no
yes	1	0
no	3	202

Drought

o \ s	yes	no
yes	0	4
no	7	195

Nile

Flood

o \ s	yes	no
yes	5	6
no	8	468

Drought

o \ s	yes	no
yes	1	49
no	11	426

Volga

Flood

o \ s	yes	no
yes	2	1
no	3	282

Drought

o \ s	yes	no
yes	2	6
no	4	276

Lena

Flood

o \ s	yes	no
yes	2	4
no	3	279

Drought

o \ s	yes	no
yes	0	1
no	5	282

Niger

Flood

o \ s	yes	no
yes	1	6
no	3	252

Drought

o \ s	yes	no
yes	0	31
no	6	225

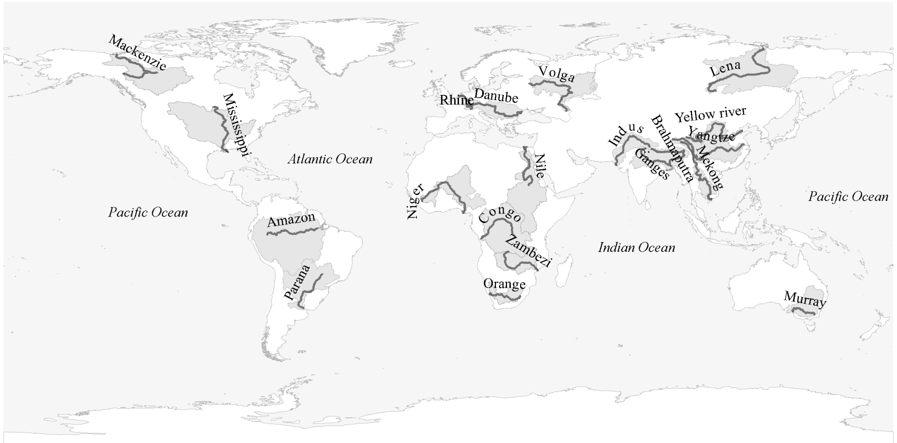

Fig. 1. Selected catchments.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

8, 3469–3505, 2011

Skill assessment of a global hydrological model

N. Candogan Yossef et al.

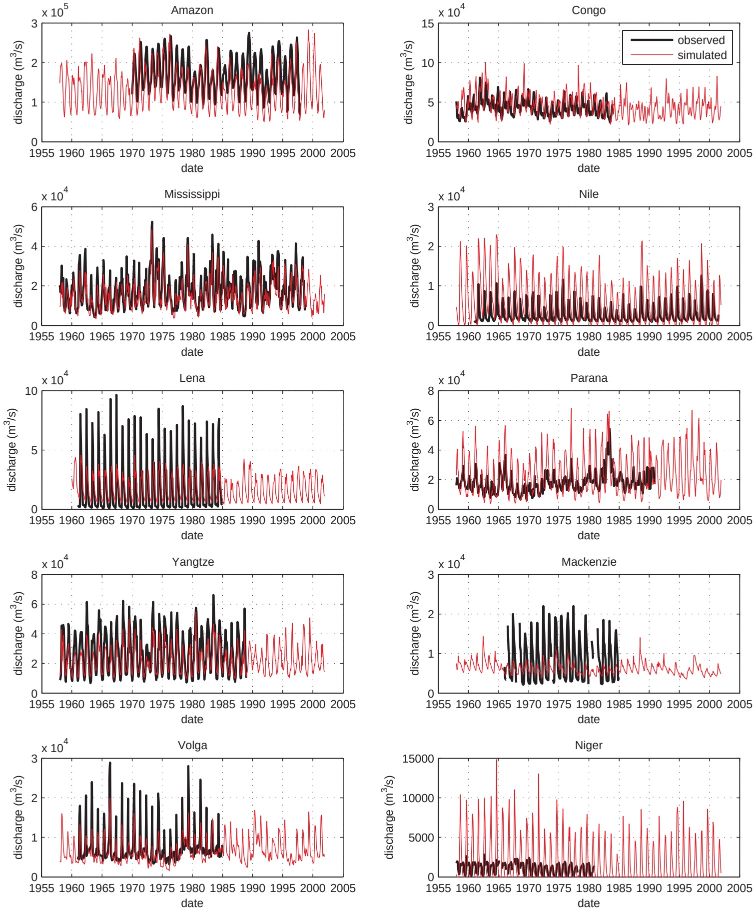

Fig. 2. Discharge time series.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

8, 3469–3505, 2011

Skill assessment of a global hydrological model

N. Candogan Yossef et al.

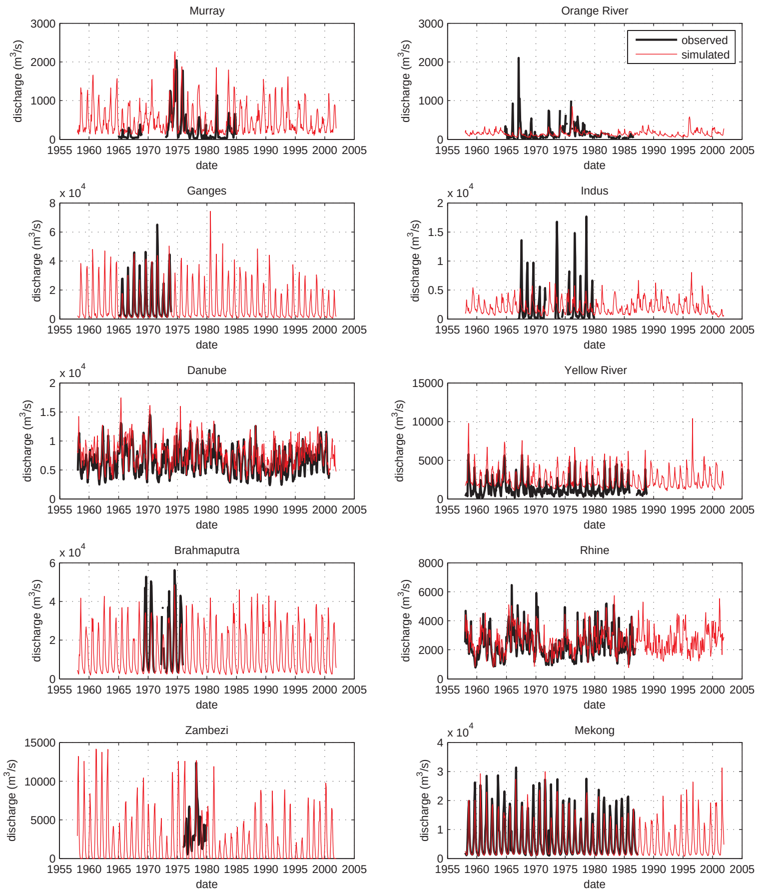

Fig. 2. Continued.

HESSD

8, 3469–3505, 2011

Skill assessment of a global hydrological model

N. Candogan Yossef et al.

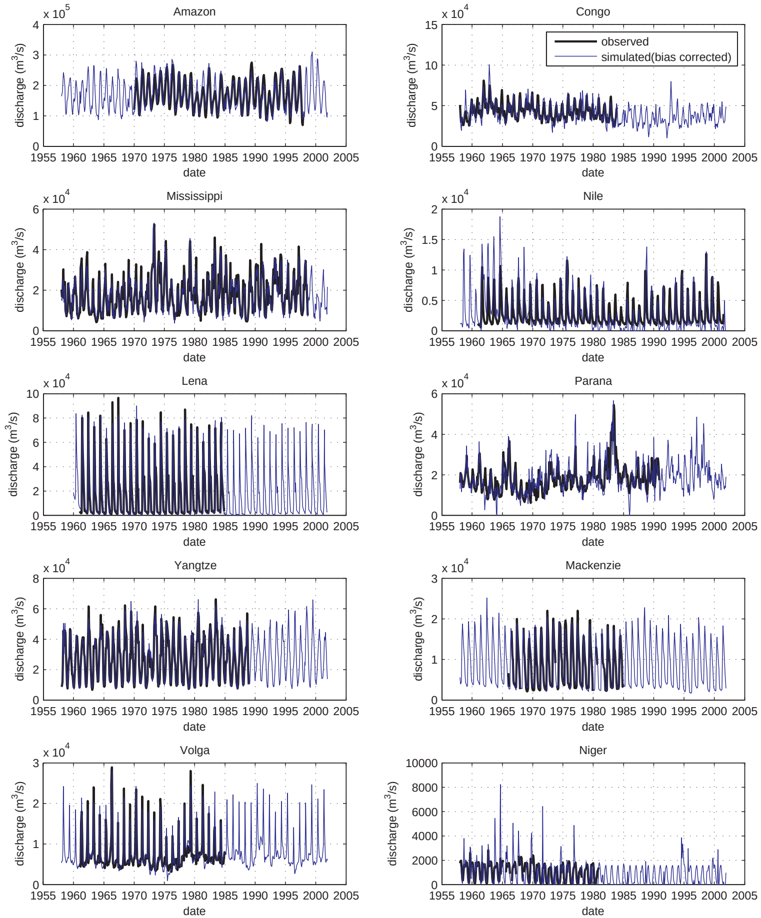

Fig. 3. Bias-corrected discharge time series.

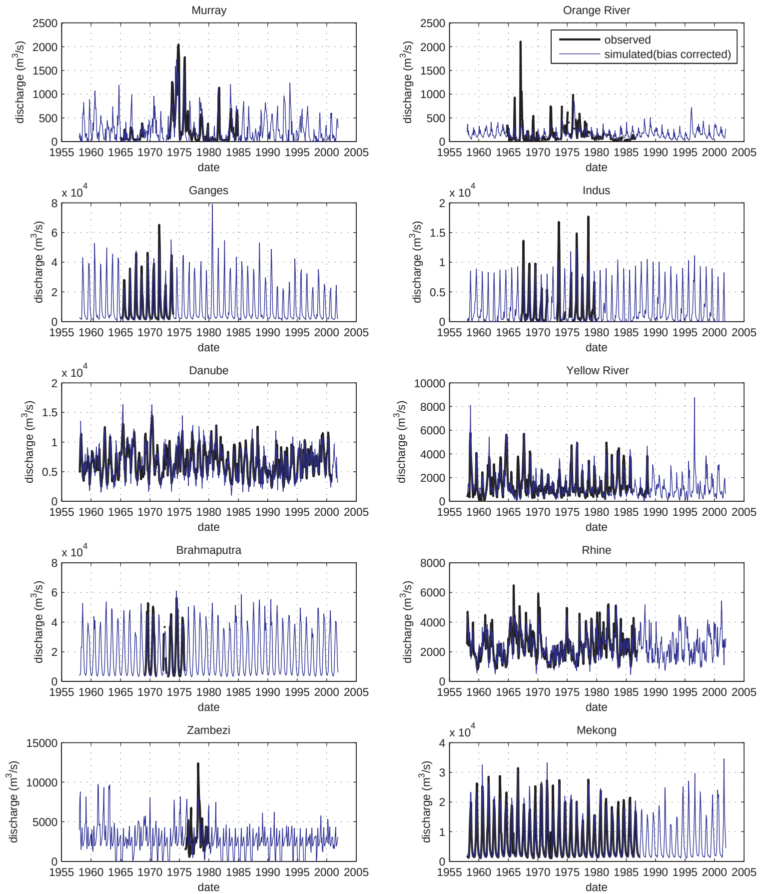

Fig. 3. Continued.

HESSD
8, 3469–3505, 2011

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀I | ▶I
◀ | ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

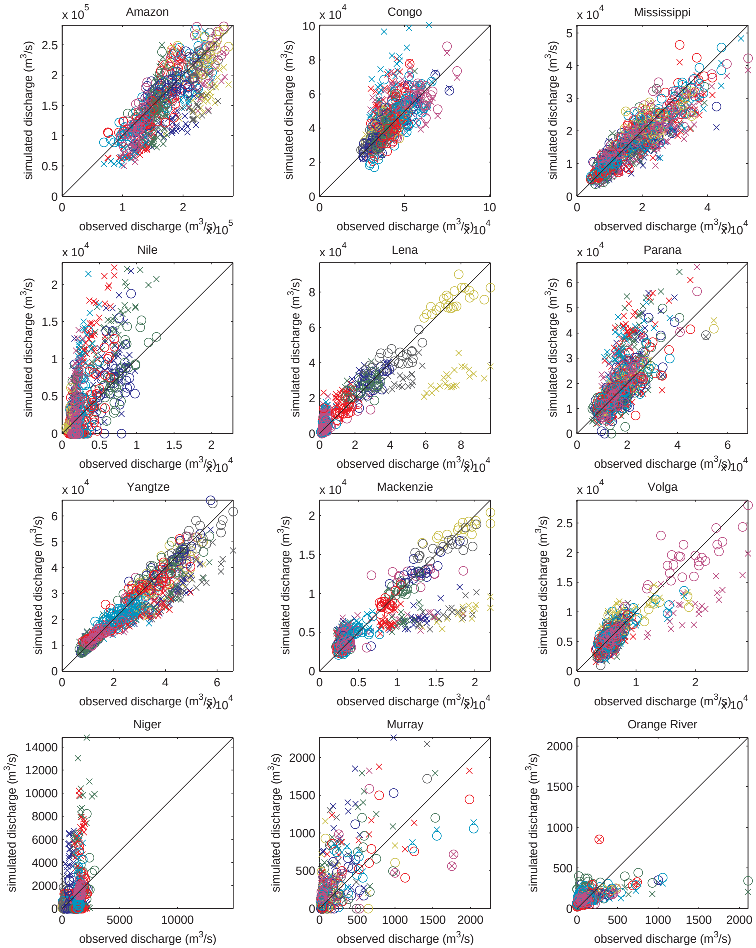

Fig. 4. Reliability diagrams.

HESSD
8, 3469–3505, 2011

Skill assessment of a global hydrological model

N. Candogan Yossef et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

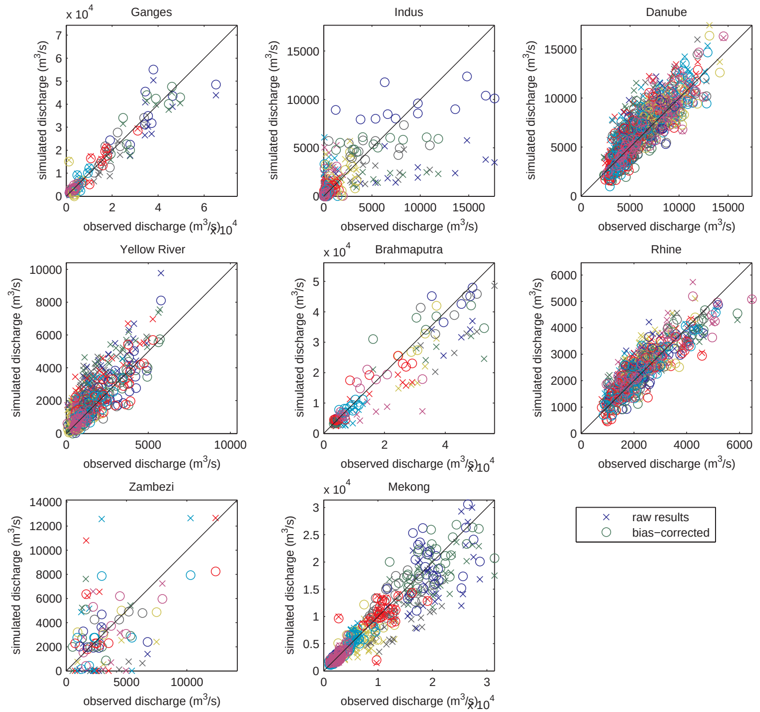

Fig. 4. Continued.

HESSD

8, 3469–3505, 2011

Skill assessment of a global hydrological model

N. Candogan Yossef et al.

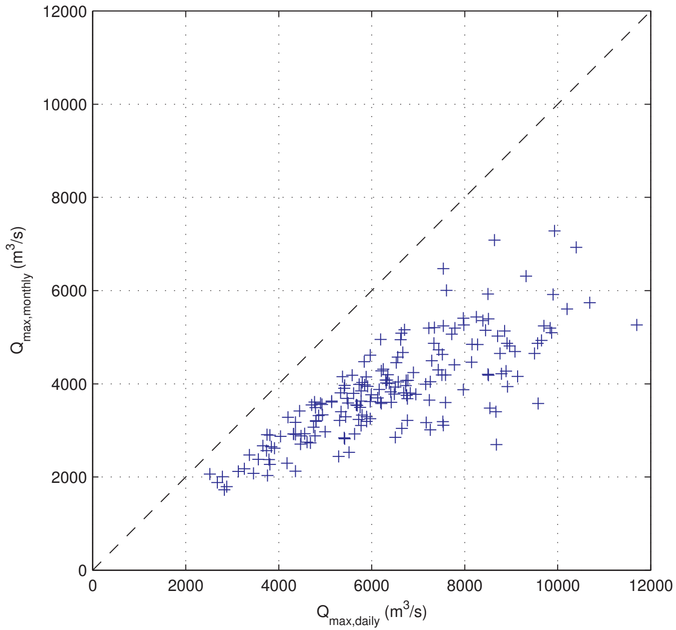

Fig. A1. Annual maxima of daily discharge vs. corresponding monthly mean flows at gauging station Lobith on the Rhine.

HESSD
8, 3469–3505, 2011

Skill assessment of a global hydrological model

N. Candogan Yossef et al.

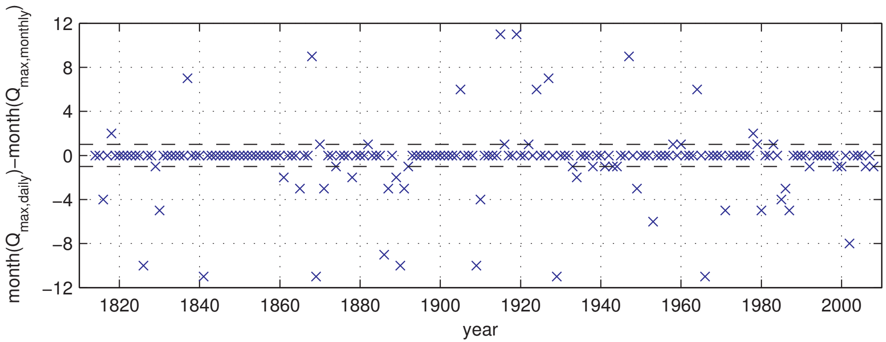

Fig. A2. The difference between the month in which the annual maximum daily discharge occurred and the month of maximum monthly flow at gauging station Lobith on the Rhine.

HESSD
8, 3469–3505, 2011