Impact of a Statistical Bias Correction on the Projected Hydrological Changes Obtained from Three GCMs and Two Hydrology Models

STEFAN HAGEMANN, CUI CHEN, AND JAN O. HAERTER
Max Planck Institute for Meteorology, Hamburg, Germany

JENS HEINKE AND DIETER GERTEN
Potsdam Institute for Climate Impact Research, Potsdam, Germany

CLAUDIO PIANI
Abdus Salam International Centre for Theoretical Physics, Trieste, Italy

(Manuscript received 30 June 2010, in final form 4 February 2011)

ABSTRACT

Future climate model scenarios depend crucially on the models’ adequate representation of the hydrological cycle. Within the EU integrated project Water and Global Change (WATCH), special care is taken to use state-of-the-art climate model output for impacts assessments with a suite of hydrological models. This coupling is expected to lead to a better assessment of changes in the hydrological cycle. However, given the systematic errors of climate models, their output is often not directly applicable as input for hydrological models. Thus, the methodology of a statistical bias correction has been developed for correcting climate model output to produce long-term time series with a statistical intensity distribution close to that of the observations. As observations, global reanalyzed daily data of precipitation and temperature were used that were obtained in the WATCH project. Daily time series from three GCMs (GCMs) ECHAM5/Max Planck Institute Ocean Model (MPI-OM), Centre National de Recherches Météorologiques Coupled GCM, version 3 (CNRM-CM3), and the atmospheric component of the L’Institut Pierre-Simon Laplace Coupled Model, version 4 (IPSL CM4) coupled model (called LMDZ-4)—were bias corrected. After the validation of the bias-corrected data, the original and the bias-corrected GCM data were used to force two global hydrology models (GHMs): 1) the hydrological model of the Max Planck Institute for Meteorology (MPI-HM) consisting of the simplified land surface (SL) scheme and the hydrological discharge (HD) model, and 2) the dynamic global vegetation model called LPJmL. The impact of the bias correction on the projected simulated hydrological changes is analyzed, and the simulation results of the two GHMs are compared. Here, the projected changes in 2071–2100 are considered relative to 1961–90. It is shown for both GHMs that the usage of bias-corrected GCM data leads to an improved simulation of river runoff for most catchments. But it is also found that the bias correction has an impact on the climate change signal for specific locations and months, thereby identifying another level of uncertainty in the modeling chain from the GCM to the simulated changes calculated by the GHMs. This uncertainty may be of the same order of magnitude as uncertainty related to the choice of the GCM or GHM. Note that this uncertainty is primarily attached to the GCM and only becomes obvious by applying the statistical bias correction methodology.

1. Introduction

The climate of Earth is influenced by increasing greenhouse gas (GHG) concentrations, changing aerosol compositions and loads, and land surface changes. In climate research, a special emphasis is placed on the hydrological cycle, which is crucial to life on Earth. Its importance is highlighted by the Global Energy and Water Cycle Experiment (GEWEX; e.g., Sorooshian et al. 2005). The implications of changes in the hydrological cycle induced by climate change may affect society more than any other changes (e.g., with regard to flood risks and changes in water availability and water quality). Consequently, the
quantification of these implications is also a major objective of the EU project Water and Global Change (WATCH; http://www.eu-watch.org). Simulations of projected components of the hydrological cycle, under a range of GHG forcing scenarios (Gutowski et al. 2007; Boberg et al. 2007), are essential tools for strategic freshwater resource management, particularly in situations where the hydrological climate change signal is unclear (Mudelsee et al. 2003; Milly et al. 2002). Global climate models (GCMs) are used to investigate possible trends in the past and future global climate. To quantify details of projected changes in the hydrological cycle and their potential impacts on water resources, commonly used global hydrology models (GHMs) or land surface hydrology models (LSHMs) are forced with GCM output. These hydrological simulations largely depend on the accuracy of the GCM data, especially of precipitation.

An accurate representation of the exchange of water among the atmosphere, the ocean, the cryosphere, and the land surface is one of the biggest challenges in global climate modeling. Simulating these fluxes is extremely difficult because they depend on processes occurring on spatial scales that are generally several orders of magnitude smaller than the typical grid size in a GCM. The formation of precipitation, for example, is controlled by a multitude of processes such as cloud microphysics and particle growth, radiative transfer, atmospheric dynamics on a variety of space and time scales, and inhomogeneities of the earth’s surface, all of which have to be properly represented in a GCM. It is well known that unless the GCM output is corrected for biases, results from a forced hydrological simulation will be unrealistic and of little use (Sharma et al. 2007; Hansen et al. 2006). Such a bias correction should correct more than only one aspect (e.g., mean or variability) of a specific variable in order to capture future changes in the whole distribution. For example, most studies of hydrological change in the past have used the delta change approach (Hay et al. 2000), where the projected changes derived from climate modeling studies are added to observational data before these data are used to force hydrology models. However, this approach considers only the changes in the mean but not in the variability so that the representation of extremes from future climate scenarios effectively is filtered out in the transfer process (e.g., Graham et al. 2007), which is not desirable in studies of future changes in extreme events. Themeßl et al. (2011) compared several empirical–statistical downscaling and error correction methods applied to daily precipitation simulated by regional climate models over the Alps. These methods include indirect methods such as multiple linear regression (e.g., von Storch 1999; Hay and Clark 2003), analog methods (e.g., von Storch and Navarra 1999; Moron et al. 2008), and direct methods such as local intensity scaling (Widmann et al. 2003; Schmidli et al. 2006) and quantile mapping (Panofsky and Brier 1968). Themeßl et al. (2011) found that quantile mapping shows the best performance in reducing biases, particularly at high quantiles, which is advantageous for applications related to extreme precipitation events. Piani et al. (2010b) have developed a statistical bias correction methodology for global climate simulations that is based on quantile mapping. Their method is applicable to daily precipitation and mean, minimum, and maximum daily temperatures and was used in the present study.

This study investigates the impact of the bias correction on the projected simulated hydrological changes and the associated uncertainty that it introduces. To consider the entire GCM–GHM modeling chain with its uncertainties, the bias correction was conducted for three GCMs and applied to two GHG scenario simulations for each of the GCMs. Then, the uncorrected and bias-corrected outputs from these GCM simulations were used to force two GHMs. Differences in the simulated hydrological changes were analyzed, and the results of the two hydrology models were compared focusing on projected changes for 2071–2100 relative to 1961–90.

The bias correction method, the models, and the simulations considered in this study are briefly described in section 2. Section 3 provides a short validation of the bias correction method using past and present-day climates. Projected climate change signals from uncorrected and bias-corrected simulations are compared in section 4. Particular uncertainties associated with the bias correction in specific regions are discussed in more detail in section 5, and the main findings are summarized in section 6.

2. Description of methods, models, and simulations

a. GCMs

Three coupled atmosphere–ocean GCMs are considered in the present study. For each GCM, a present-day control period of 1960–99 was used to derive the parameters necessary for the bias correction. Then, the bias correction was applied to the control period as well as to two scenario simulations (B1 and A2) from 2000 to 2100. These scenario simulations followed specific assumptions for the evolution of greenhouse gases and aerosols, which have been defined by the Intergovernmental Panel on Climate Change (IPCC; Houghton et al. 2001) and are described in the IPCC Special Report on Emission Scenarios (SRES; Nakicenovic et al. 2000). All GCM simulations used were conducted for the Fourth Assessment Report (AR4) of the IPCC.
The GCMs were forced by observed GHG and aerosol concentrations for the past climate until 2000, and by prescribed concentrations in accordance with the respective emission scenario afterward. Table 1 gives an overview on the three GCMs and their spatial resolutions, while additional information is provided below.

### 1) ECHAM5

The coupled atmosphere–ocean GCM ECHAM5–Max Planck Institute Ocean Model (MPI-OM) (denoted simply ECHAM5 henceforth; Roeckner et al. 2003; Jungclaus et al. 2006) of the Max Planck Institute for Meteorology (MPI-M) has been used to conduct an ensemble of climate simulations where, in the present study, the third initial condition ensemble member was always used. The GCM takes into account concentrations of CO$_2$, CH$_4$, N$_2$O, chlorofluorocarbons (CFCs), O$_3$ (tropospheric and stratospheric), and sulfate aerosols, thereby considering the direct and first indirect aerosol effect.

### 2) CNRM

The Centre National de Recherches Météorologiques Coupled GCM, version 3 (CNRM-CM3; henceforth simply CNRM) comprises the submodels Action de Recherche Petite Echelle Grande Echelle (ARPEGE)-Climat version 3 for the atmosphere (Déqué et al. 1994; Déqué and Piedelievre 1995; Royer et al. 2002), Océan Parallélisé (OPA) 8.1 for the ocean (Madec et al. 1998), and GELATO 2 for sea ice (Salas-Mélia 2002). The distributions of marine, desert, urban aerosols, and sulfate aerosols were specified, whereas for aerosols only the direct effect of anthropogenic sulfate aerosols was taken into account.

### 3) IPSL

The L’Institut Pierre-Simon Laplace Coupled Model, version 4 (IPSL CM4; hereafter simply IPSL) includes the submodels LMDZ-4 for the atmosphere (Hourdin et al. 2006), ORCA for the ocean (based on the OPA model; Madec et al. 1998), and LIM for sea ice (Fichefet and Morales Maqueda 1997; Goosse and Fichefet 1999). With regard to prescribed aerosols, the direct effect of sulfate aerosols was taken into account, as well as the first indirect effect.

---

### b. Observational data

As observations, temperatures (daily mean, min, and max) and precipitation (daily mean) from the newly available global WATCH dataset of hydrological forcing data (henceforth referred to as WFD; Weedon et al. 2011) were used. This dataset covers the period 1958–2001 and is based on the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40; Uppala et al. 2005). The ERA-40 data were interpolated to 0.5° and only considered over land points using the land–sea mask from the Climate Research Unit dataset TS2.1 (CRU; Mitchell and Jones 2005). A correction for elevation differences between ERA-40 and CRU was applied. For 2-m temperatures, a correction of the monthly means with CRU data was performed. For precipitation, a correction of the monthly means with data from the Global Precipitation Climatology Centre full dataset version 4 (GPCC; Fuchs et al. 2007) was conducted. In addition, a gauge-undercatch correction following Adam and Lettenmaier (2003) was used, which takes into account the systematic underestimation of precipitation measurements that have an error of up to 10%–50% (see, e.g., Rudolf and Rubel 2005). In this way, the WFD combine the daily statistics of ERA-40 with the monthly mean characteristics of CRU temperature and gauge undercatch-corrected GPCC precipitation amounts. A comparison to FLUXNET data (http://www.fluxnet.ornl.gov/fluxnet/) demonstrated a close correspondence between field measurements and the WFD for all variables (Weedon et al. 2011).

### c. Statistical bias correction

The bias correction methodology is described in detail in Piani et al. (2010b). Here, only a short summary is given.

The statistical bias correction is designed to adjust all moments of the probability distribution function (PDF) of intensity for a specific variable (In this respect, intensity refers to the value of the variable that is valid for the time step considered in the PDF.). Once bias corrected, the modeled variable should have the same intensity PDF as the observed one. In the bias correction methodology the corrected variable is a function of the modeled one. This function is referred to as the “transfer

---

### Table 1. Three IPCC AR4 GCMs and their spatial resolution.

<table>
<thead>
<tr>
<th>Center</th>
<th>GCM</th>
<th>Horizontal resolution</th>
<th>Vertical resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPI-M</td>
<td>ECHAM5/MPI-OM</td>
<td>T63 – 1.9° – 200 km</td>
<td>L31</td>
</tr>
<tr>
<td>CNRM</td>
<td>CNRM-CM3</td>
<td>T42 – 2.8° – 300 km</td>
<td>L45</td>
</tr>
<tr>
<td>IPSL</td>
<td>LMDZ-4</td>
<td>3.75° × 2.5° – 300 km</td>
<td>L19</td>
</tr>
</tbody>
</table>

(Solomon et al. 2007).
function.” To obtain the transfer function, modeled and corresponding observed time series of the same length are sorted according to their magnitude, from smallest to largest, and plotted one versus the other (Fig. 1). Plotting observed versus modeled data (e.g., daily precipitation) yields a perfect transform function in that the corrected data would yield an intensity PDF identical to the observed (Fig. 1, dotted curve). This perfect transfer function is not valid for data outside the observed range and has too many degrees of freedom to be stable in time (see Piani et al. 2010b for details). Hence, an idealized transfer function, defined by a greatly reduced number of parameters, is fitted (Fig. 1, solid line). This transfer function can then be used to find the corresponding (corrected) value to each model value. Following the arrows in Fig. 1 gives an example of how a corrected value can be obtained from a model value using the transfer function.

In practice, a two-parameter fit to the daily precipitation transform function was used as a good approximation for most regions (as shown in Fig. 1). For some specific regions, three- or four-parameter transfer functions produced better results [as described in Figs. 2 and 3 in Piani et al. (2010b)]. Note that using a larger number of parameters may not be adequate as the correction needs to be time independent on climatological time scales (The more parameters are needed to define the transfer function, the longer the observational period has to be in order to evaluate them adequately). A similar procedure has been followed for the correction of daily temperatures where a two-parameter fit of a linear transfer function was generally sufficient. In both cases, monthly transfer functions were used, with smooth transitions between months for temperature. The latter removed jumps that would otherwise occur at the transitions.

The bias correction transfer functions were derived for the period 1960–99 and applied to 1960–2100 for both scenarios. Here, total precipitation was corrected using transfer functions, and snowfall was corrected accordingly using the snowfall fraction taken from the GCM (snowfall was not used in the present study as both GHMs compute their own snowfall fraction). In addition, mean \( T_{\text{mean}} \), minimum \( T_{\text{min}} \), and maximum \( T_{\text{max}} \) daily temperatures were also corrected. But \( T_{\text{min}} \) and \( T_{\text{max}} \) were not corrected directly as it turned out that correcting the diurnal range \( \Delta T = T_{\text{max}} - T_{\text{min}} \) and the skewness \( \sigma = (T_{\text{mean}} - T_{\text{min}})/\Delta T \) yielded a correction with smaller relative errors (Piani et al. 2010b).

---

Fig. 1. Example for correcting original model data using a transfer function obtained from cumulative distribution functions of observed and modeled intensities.
It should be pointed out that the bias correction also includes statistical downscaling to 0.5°, which is due to the higher resolution of the WFD of 0.5° as compared to the GCM resolutions (see Table 1). This means that the GCM data were interpolated to 0.5° resolution, and then they were bias corrected at 0.5° with the WFD. Especially in regions with large orographic gradients, the higher resolution contains features not present in the low-resolution GCM data, such as orographic precipitation. Note that the accuracy of the bias correction is always limited by the quality of the observational data used. Consequently, bias-corrected GCM data may be less reliable in data-sparse regions. Furthermore, measurement of small precipitation amounts is generally more difficult than that of large amounts. Therefore, a lower measurement cutoff of 1 mm day\(^{-1}\) for precipitation was used and only values larger than this cutoff were employed in the construction of the transfer function as shown in Fig. 1. In addition, a prior removal of outliers was conducted using a criterion common to all grid points and models. Including outliers in the derivation of transfer functions would increase the error when the bias correction is done on different time periods, since these are the most likely to change from one decade to another. For the application of the transfer functions, all daily GCM data were corrected, thereby leading to small deviations between WFD and GCM statistics in the respective areas. Also, the bias correction cannot correct temporal errors of major circulation systems in the GCM data (e.g., the onset of the monsoon). With regard to future climate, it is assumed that the bias behavior of the model does not change with time; that is, the transfer functions are time independent and thus applicable in the future. Note that the bias correction method used here is not restricted to GCM output but can be also applied to regional climate model (RCM) output, given availability of observational data at the high resolution of the RCM, such as done by Piani et al. (2010a), who applied the same concept to RCM precipitation over Europe.

d. GHMs

The original and the bias corrected GCM data were used to force two GHMs: 1) the hydrological model of the Max Planck Institute for Meteorology (MPI-HM), and 2) the dynamic global vegetation model LPJmL operated by the Potsdam Institute for Climate Impact Research. Both models have a daily time step and use a spatial resolution of 0.5°, and their main characteristics are given in Table 2.

1) MPI-HM

MPI-HM consists of the simplified land surface (SL) scheme (Hagemann and Düménil Gates 2003), which computes vertical water fluxes, and the hydrological discharge (HD) model (Hagemann and Düménil 1998) that globally simulates the lateral freshwater fluxes at the land surface. The latter is a state-of-the-art discharge model that is applied and validated on the global scale, and it is also part of the coupled atmosphere–ocean GCM ECHAM5/MPI-OM. The SL scheme incorporates the main components of the hydrological cycle at the land surface and primarily uses relations that are functions of temperature and precipitation. The soil is represented by a single soil layer, and major process representations account for the separation of throughfall into surface runoff and infiltration according to the improved Arno scheme (Hagemann and Düménil Gates 2003), the separation of precipitation into rain and snow according to Wigmosta et al. (1994), snowmelt using a daily degree formula according to Bergström (1992), and potential evapotranspiration using the Thornthwaite formula (Chebotarev 1977), which is purely based on temperature. For the current study, slight modifications compared to Hagemann and Düménil Gates (2003) were implemented. Land sea mask, glacier mask, and total (field capacity) and plant-available soil water capacity were taken from the LSP2 dataset (Hagemann 2002). Lake and wetland fractions were obtained from the global lake and wetlands database (Lehner and Döll 2004), and the lake and wetland evaporation at the potential rate was modified in both components of MPI-HM (T. Stacke 2010, personal communication).

2) LPJmL

The global ecohydrological model LPJmL (Bondeau et al. 2007; Rost et al. 2008) simulates at 0.5° resolution
the growth, production, and phenology of natural and agricultural vegetation in direct coupling with the carbon and water cycling. Atmospheric CO2 concentration is simulated to affect plant transpiration and biomass production via both physiological and structural plant responses (Gerten et al. 2004). The establishment and dynamic distribution of natural vegetation and the seasonal phenology of natural and agricultural vegetation are simulated based on long-term average climate. The model distinguishes two soil layers with fixed thickness (upper, 50 cm; lower, 100 cm). Soil moisture of each layer is updated daily, according to the balance between the amount of water infiltrating into the soil (throughfall minus surface runoff) and that removed from the soil layers through subsurface runoff, percolation, soil evaporation, and plant transpiration. Evapotranspiration is calculated from radiation and temperature using the Priestley–Taylor formulation (with a modification for plant transpiration to mimic boundary layer effects). Runoff is generated when field capacity of the upper and/or lower soil layer is surpassed. Snowmelt is modeled following a degree-day approach. For details on the hydrological scheme see Gerten et al. (2004) and Rost et al. (2008); for the most recent description of the model version and the land use input datasets used herein, see Gerten et al. (2011).

3. Validation

a. Global scale

In this subsection, the month April is considered as an example since other months show similar behavior. With regard to the climatological monthly mean precipitation from 1960 to 1999, Fig. 2 shows that the three GCMs have partially large precipitation biases compared to the WFD (left column). These biases were largely reduced by the bias correction (Fig. 2, right column). While the mean precipitation could also easily be corrected by simpler methods such as the “delta change approach,” this would not be the case for the standard deviation of precipitation, which is considered over the control period for the month of April in Fig. 3. Note the regional details for the WFD shown in Fig. 3a, especially over mountainous terrain. Figure 3b presents the original and the corrected ECHAM5 data over the same period, respectively. The comparison to the WFD confirms the utility of the correction. The analogous results for IPSL (Fig. 3c) and CNRM (Fig. 3d) show that the agreement of the bias-corrected data with the WFD is good in all cases. Similar results were obtained for the mean, standard deviation, and daily range of temperature (not shown). Note that the bias correction did not lead to a perfect agreement of the monthly means of bias-corrected GCM data and WFD. For precipitation (Fig. 2), this is mainly related to the prior outlier removal (see section 2c), especially for CNRM data (Fig. 2d), while for temperature (not shown), this originates from the smooth transitions of the transfer function between the months.

b. Catchment scale

To evaluate the present-day GHM simulations, several catchments were selected (Fig. 4), for which river discharge data have been compiled by Dümenil Gates et al. (2000). The catchments comprise the following regions representing different climate regimes: the Amazon, Arctic Ocean represented by its six largest rivers (Jenisei, Kolyma, Lena, Mackenzie, North Dvina, and Ob), Baltic Sea catchment (land only), Congo, Danube, Ganges/Brahmaputra, Mississippi, Murray, Nile, Parana (La Plata), and Yangtze Kiang.

Climatological annual mean biases in precipitation $P$ and temperature $T$ are not considered because the bias correction methodology results in biases of their corrected values that are close to zero (see above) for the control period. Biases of annual mean runoff $R$ and evapotranspiration $E$ are shown in Figs. 5 and 6, respectively. Note that because of the lack of observational data evapotranspiration has been diagnosed as $E = P - R$ by assuming that the long-term storage of soil water is negligible. The observational values used to calculate the biases are given in Table 3.

For most of the rivers, the bias correction led to a reduced bias in the simulated runoff coefficient of both GHMs. The large positive biases shown for the rivers Murray and Nile (station before the Aswan dam) mainly have the following reasons. First, the water of both rivers is heavily used for irrigation, which is not accounted for in the models (neither in the GCMs nor in the GHMs). Here, it can be noted that for most of the rivers, human interference (especially reservoirs) has larger impacts on the seasonal variations than on the annual means (Biemsans et al. 2010), except for those rivers where large amounts of water are taken out of the river system for irrigation, such as for the Nile and Murray rivers as stated above. Second, large evaporation losses from swamps and ephemeral water bodies are strongly underestimated in these and other macroscale hydrological models (see, e.g., Gerten et al. 2004). And third, in the long-term mean, the runoff of mostly arid regions like these two catchments is the relatively small residual of two large values of $P$ and $E$ (cf. Table 3). Thus, small biases in $P$ or $E$ may lead to large relative biases in runoff. It can also be noted that the runoff biases of MPI-HM (Fig. 5a) seem to be
generally somewhat smaller than for LPJmL (Fig. 5b). The same is also the case for bias in evapotranspiration (Fig. 6). Apart from general model differences, this might be partially related to the fact that LPJmL uses two GCM variables in addition to $P$ and $T$ as forcing (see Table 2), which are required to calculate evapotranspiration (see section 2d). The two radiation fluxes are not bias corrected and are thus not fully consistent with bias-corrected $P$ and $T$. Therefore, biases in these radiation fluxes are still present and may cause biases in LPJmL evapotranspiration. The majorly reduced biases due to the bias correction are also seen for evapotranspiration although the reduction seems less pronounced than for runoff.
4. Projected climate change signals

a. GCM temperature and precipitation

Figure 7 shows the changes of temperature and precipitation projected under the A2 scenario by the end of the twenty-first century over selected catchments. For most catchments, the temperature signal shifts induced by the bias correction are smaller than the signal differences between the GCMs (Fig. 7a). But a few exceptions can be noted where a strong impact of the bias...
correction on the signal occurs. This is pronounced for ECHAM5 and CNRM in the Amazon and CNRM in the Mississippi, as well as IPSL in Congo, Nile, and Yangtze catchments. Some impact is also seen for ECHAM5 and CNRM in the Congo, ECHAM5 and IPSL in the Ganges/Brahmaputra, IPSL in the Baltic Sea, and CNRM in the Arctic and Murray catchments. Similar results were obtained for precipitation (Fig. 7b) where the bias correction had a strong impact on the signal for IPSL and CNRM in the Ganges/Brahmaputra, CNRM in the Arctic, ECHAM5 in Congo and Nile catchments, and some impact for CNRM in Amur and Parana and for IPSL in Amazon, Arctic and Murray catchments. In general it can be noted that the bias-corrected signals of the three GCMs are not closer together than the original data, which suggests that the bias correction does not correct for uncertainties in the signal related to the choice of the GCM. This result was expected, as the projected signals are strongly associated with the different climate sensitivities of the GCMs. Generally, the projected A2 warming is stronger than for the B1 scenario (not shown), but the bias correction had a similar relative impact for both scenarios and all three GCMs over most of the catchments. Here, it can be noted that CNRM projects the strongest warming of the three GCMs over the Nile catchment (0.3°–0.5°C larger) and a similar warming over the Murray catchment, but it has the weakest B1 signal (about 1°C lower for Murray, 0.3°–0.5°C lower for Nile), pointing to some feedback processes that are differently handled in CNRM compared to the other GCMs. For precipitation, the projected B1 changes are similar in magnitude or somewhat lower than for the A2 scenario. The same applies to changes in the climate change signal imposed by the bias correction.

b. GHM runoff and evapotranspiration

Figure 8 shows the projected mean annual runoff changes simulated by the two GHMs. For clarity of the figure, the plot is restricted to ECHAM5 and CNRM data but included also the B1 changes to give an idea of the variations between the scenarios.

For most of the catchments, the obvious differences between the runoff signals of the original and bias-corrected GCM data are directly related to the differences that the bias correction imparted to the projected precipitation changes. Exceptions are ECHAM5 in the Amazon and Ganges/Brahmaputra for MPI-HM, where runoff increases due to the bias correction are related to decreases in evapotranspiration (Fig. 9). Noticeable impacts occurring for ECHAM5 in the Arctic and CNRM in Amazon (MPI-HM only) and the Baltic Sea catchments show that even smaller changes in the precipitation signal may lead to noticeable changes in the runoff signal. In some cases, precipitation increases due to the bias correction also led to increases in evapotranspiration (Fig. 9), such as for CNRM in Parana (MPI-HM only; LPJmL shows increases in runoff instead), ECHAM5 in Nile (LPJmL only), CNRM in Amur (LPJmL), and CNRM in Ganges/Brahmaputra (MPI-HM only) catchments. In general the relative projected climate changes are lower for evapotranspiration than for runoff, as is also the case for the relative impact of the bias correction on the projected climate change signal of both variables. The projected B1 changes are mostly similar or lower in magnitude than the A2 changes, which also applies to changes in the climate change signal imposed by the bias correction.

In addition, while for some catchments (Amur, Arctic rivers, Danube), both GHMs agree reasonably well in
terms of their projected discharge changes, at least in the direction of change, there are notable differences even in the direction of change for most of the catchments, especially for those located in tropical areas (Amazon, Congo, Nile, Parana). These differences are likely directly related to differences in projected changes in evapotranspiration (Fig. 9) and thus dependent on the parameterizations used to calculate evapotranspiration in the respective GHM. This issue will be taken up in section 4c.

A further reason causing the differences in the evapotranspiration signal might be related to the different treatment of soil moisture in both GHMs. MPI-HM uses only one soil layer, the root zone, where all water above the wilting point is in principle accessible.
for transpiration by plants. LPJmL has a more sophisticated, two-layer soil moisture budget, in which water uptake for transpiration is regulated by the root mass present in each layer (with relatively little uptake from the lower soil layer), and it assumes a maximum transpiration rate limited by plant hydraulic traits (Gerten et al. 2004). Thus, the soil water content simulated by LPJmL tends to be higher than that of MPI-HM, and especially it shows little variation over the Amazon basin.

For some catchments (especially the Amur, Arctic, Congo, and Ganges/Brahmaputra), the impact of the bias correction on the discharge signal is of the same order as differences between the two GHMs, while for evapotranspiration its impact is generally smaller than the respective GCM differences. This is likely due to the fact that discharge is directly affected by precipitation, and hence by its correction, while evapotranspiration is only indirectly affected by precipitation via wet and dry
TABLE 3. Observed values for WFD precipitation (1961–90), evaporation (WFD precipitation minus climatological discharge), and runoff (climatological); unit is mm yr⁻¹.

<table>
<thead>
<tr>
<th>Catchment</th>
<th>Precipitation</th>
<th>Evaporation</th>
<th>Runoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>2231</td>
<td>1178</td>
<td>1053</td>
</tr>
<tr>
<td>Amur</td>
<td>541</td>
<td>363</td>
<td>178</td>
</tr>
<tr>
<td>Six largest Arctic rivers</td>
<td>457</td>
<td>266</td>
<td>191</td>
</tr>
<tr>
<td>Baltic Sea catchment</td>
<td>686</td>
<td>406</td>
<td>279</td>
</tr>
<tr>
<td>Congo</td>
<td>1635</td>
<td>1270</td>
<td>365</td>
</tr>
<tr>
<td>Danube</td>
<td>808</td>
<td>557</td>
<td>251</td>
</tr>
<tr>
<td>Ganges/Brahmaputra</td>
<td>1392</td>
<td>720</td>
<td>672</td>
</tr>
<tr>
<td>Mississippi</td>
<td>869</td>
<td>684</td>
<td>185</td>
</tr>
<tr>
<td>Murray</td>
<td>491</td>
<td>482</td>
<td>8</td>
</tr>
<tr>
<td>Nile</td>
<td>660</td>
<td>611</td>
<td>49</td>
</tr>
<tr>
<td>Parana</td>
<td>1274</td>
<td>1006</td>
<td>268</td>
</tr>
<tr>
<td>Yangtze Kiang</td>
<td>1072</td>
<td>535</td>
<td>537</td>
</tr>
</tbody>
</table>

soil moisture conditions, and is also indirectly influenced by temperature via potential evaporation.

c. Changes in the annual cycle for specific catchments

To analyze more thoroughly why the bias correction has such a large impact on the climate change signal for a specific GCM–GHM combination in some regions, annual cycles of discharge changes for selected rivers are considered in Fig. 10.

The absolute value of the runoff change signal (Fig. 8) and the size of precipitation bias do not appear to be decisive for whether the bias correction has a strong impact on the climate change signal or not. This can be seen for the Danube, where Hagemann et al. (2009) found a prominent dry bias in the ECHAM5 simulation that is especially large during the summer. Figure 10 (upper panels) shows that the effect of the bias correction on the Danube discharge change signal is relatively small. Here, it is only noticeable in a few months, especially in winter, and it is fairly similar for all GCMs and both GHMs.

For the Arctic rivers, the monthly discharge signal (Fig. 10, second row) is largely affected by the bias correction of precipitation (Fig. 11a), the projected changes of which were increased throughout the year except for the summer. The maximum projected discharge increase in spring is shifted toward the winter for the bias-corrected GHM simulations, especially when using CNRM and IPSL data. It seems that the increase in precipitation due to the bias correction led to generally wetter soil conditions, which in turn enhanced subsurface runoff in the winter. The latter is the main source of winter discharge in the Arctic catchment, which is also comparatively low so that small absolute changes lead to larger relative ones. The impact of the temperature bias correction was rather small as it caused mainly some smaller future increases in the winter when the temperatures were far below the freezing point (not shown). Only for CNRM, projected temperatures were increased by 1°–1.5°C for most parts of the year, which was causing somewhat more rain/less snow and enhanced snowmelt in the transitional seasons.

For the Ganges/Brahmaputra, the bias correction of CNRM data notably increased the two peaks in the climate change signal for spring and autumn discharge (Fig. 10, third row), which is due to a corrected increase in precipitation around the beginning and the end of the monsoon season (Fig. 11b). This feature will be considered in more detail in section 5a.

With regard to the Amazon it can be noticed that the strong impact of the bias correction on MPI-HM discharge (Fig. 10, lower panels) is not related to the precipitation correction that has only a weak impact on the precipitation signal (Fig. 12a). Instead it is related to shifts in the MPI-HM evapotranspiration signal (Fig. 12c), which was directly influenced by the bias correction of temperature (Fig. 12b). Here, an especially strong impact of the temperature bias correction of ECHAM5 occurs in the dry season in the second half of the year. The impact of the bias correction on the temperature signal will be investigated more thoroughly in section 5a.

This distinct sensitivity of the evapotranspiration signal to the bias correction of the ECHAM5 temperature data cannot be seen for LPJmL (Fig. 12d). This points directly to a formulation of potential evaporation in MPI-HM that is purely based on temperature. It seems that especially in tropical and subtropical areas, this formulation reacts quite sensitively to future increases in temperature, thereby causing overly large increases in total evapotranspiration under global warming conditions (cf. Fig. 9). This is supported by Fig. 13, where the projected annual evapotranspiration changes of both GHMs using original and corrected ECHAM5 data are compared with changes directly simulated by the GCM. Here, the latter is relatively close to LPJmL but strongly differs from MPI-HM over the Amazon, Congo, Ganges/Brahmaputra, Nile, and Parana catchments. The GHM intercomparison study of Haddeland et al. (2011) also indicated some deviating behavior of MPI-HM for present-day summer evapotranspiration using WFD forcing over the Ganges catchment.

5. Discussion of particular uncertainties

a. Temperature

To investigate the effect of the bias correction on the temperature signal in more detail, daily temperatures (original and bias-corrected ECHAM5) are shown in Fig. 14 for September–December 1991 over the Amazon catchment. It can be seen that higher temperatures were
reduced much further by the bias correction than lower temperatures during September and October. In November (days 305–334), lower temperatures were even raised. The total daily temperature variation in these three months is about 3°C to 6°C. As the mean projected ECHAM5 A2 increase (∼5°C–7°C; see Fig. 12b) is generally larger than the variation in these months, the whole daily temperature band in these months was shifted toward a higher temperature mean state, which was strongly decreased by the bias correction. Consequently, this largely reduced the projected temperature increase.

For a more analytical explanation for this impact of the bias correction on the projected temperature signal, the linear transfer function \( a + bT \) of the temperature bias correction is now considered. Figure 15 shows the annual mean slope \( b \) of the monthly transfer functions.
for all three GCMs. The three plots show that the slope is often not close to unity. This means that the GCM has considerably stronger (less than unity) or considerably weaker (greater than unity) fluctuations than the WFD. The bias correction would hence remove the effect by simply narrowing the fluctuations. Close inspection of the three plots leads to the conclusion that there are regions where all models consistently have biases in temperature fluctuations (such as the Amazon, Greenland, western North America, and central Siberia) but in other regions there is little overlap between the different models. In a future climate as projected by the GCM
simulations the globe may experience a general increase of temperatures. This increase would be interpreted by the bias correction as a fluctuation with respect to the control-period mean and hence an increase would be corrected subject to the slope given in the plots. For a change signal of $\Delta T$ the corrected signal would become $b\Delta T$. Certainly, such impacts on future projections caused by the bias correction have to be taken very seriously and fundamental questions regarding the interpretation of model results for future climate can be raised (to be discussed elsewhere). At this point it should be mentioned that—especially in regions where all models consistently over- or underestimate the natural fluctuations—the meaningful interpretation of future climate projections becomes rather problematic. To check the goodness-of-fit of the temperature transfer functions, the $R^2$ values of the fit were calculated for ECHAM5, where values close to 1 indicate a good fit. In most regions of the globe the value is greater than 0.95. The only region where the linear fit cannot be

Fig. 9. Projected annual mean evapotranspiration changes in 2071–2100 relative to 1961–90 for (a) MPI-HM and (b) LPJmL.
FIG. 10. Monthly mean discharge changes (2071–2100 compared to 1961–90) for (top row) the Danube, (second row) the six largest Arctic rivers, (third row) the Ganges/Brahmaputra, and (bottom row) the Amazon as simulated by (left) MPI-HM and (right) LPJmL.
considered a very good statistical model is in the Amazon region (not shown), where $R^2$ values less than 0.95 occur. Figure 16 provides two examples of fits for selected grid boxes in the Amazon catchment where observed versus model temperatures are shown for January. For most areas of the globe the fit is rather good, as is the case in Fig. 16b. However, for several grid boxes (see Fig. 16a) in the Amazon region the transfer function is not well approximated by a linear fit. This is due to a convex behavior of the transfer function in that region. Technically, this could be improved by allowing a larger array of possible transfer functions, possibly using a quadratic term. However, in the current situation, a more dramatic challenge is posed by the transfer function itself, even if
the fit were perfect: How can the discrepancy in model versus observed behavior of fluctuations in temperature be interpreted? Can it be assumed that fluctuations that are too pronounced in the current-day model simulation imply that the temperature change caused by a shifting climate would also be overrated? If the model’s reaction to radiation changes is responsible for such fluctuations, and the radiation change in a future climate is similar to that of day-to-day changes, then this assumption might be justified. This may be different in regions where the temperature bias is strongly related to specific weather patterns. If the distribution of these patterns will not change in the future despite a warming in the mean, the bias behavior for low and high temperatures relative to the mean is not expected to change either. In that case, an alternative approach may be better suited for the correction of projected temperatures, such as the shift of the whole transfer function by the amount of the projected mean change $\Delta T$ [i.e., using $a + b(T - \Delta T)$ instead of $a + bT$]. However, answering the questions raised above goes beyond the scope of the current paper; these shall be discussed in future research.

b. Precipitation

As for temperature, a similar impact of the bias correction can also be seen for precipitation in some regions and seasons. Here, for example, pronounced shifts in the projected A2 signal can be noted for CNRM data in the Ganges/Brahmaputra catchment. These shifts mainly took place around the beginning and end of the Indian summer monsoon season. The daily precipitation of CNRM for January–April 1991 (Fig. 17) shows that in April (days 91–120), lower precipitation values were further reduced while higher values are increased by the bias correction. CNRM projects an increase of $\Delta P$ annual mean precipitation by more than 25% in April (Fig. 11b), which shifts larger parts of the April precipitation into the increasing range of the bias correction, thereby leading to the strong changes in the signal. Note that the

![Fig. 13. Projected annual mean A2 ECHAM5-based evapotranspiration changes in 2071–2100 relative to 1961–90 for MPI-HM, LPJmL, and changes directly simulated by ECHAM5.](image1)

![Fig. 14. Original and bias-corrected daily temperatures of ECHAM5 for September–December 1991 over the Amazon catchment.](image2)
absolute changes are less pronounced than the relative ones due to the lower precipitation amounts in April than in the summer monsoon period (original CNRM $\Delta P = 1$ mm day$^{-1}$, corrected $\Delta P = 1.7$ mm day$^{-1}$).

In most arid regions models and observations disagree strongly and the slope of the transfer function may become extreme. Extreme slopes are defined—somewhat arbitrarily—by a slope larger than 5 or less than $\frac{1}{5}$. In those cases, the bias correction does not produce a fit to the slope but simply corrects the mean by an additive correction; hence, in the curve $a + bx$, $b$ is taken as unity but $a$ is adjusted (Piani et al. 2010b). In Fig. 18a, regions are shown where such a correction was used in April. The arid regions stand out from the remaining regions.
because, even if slightly different definitions are used, the slope values are always far from unity in those regions. As an example, in the Ganges catchment in April the slope is extremely small, meaning that model precipitation is generally much larger than that of WFD. From this region, Fig. 18b provides a typical example of the model to observations mapping. As model precipitation is much larger, the offset \( a \) had to be chosen to be negative. All model precipitation values \( P \) less than the \( x \) intercept \( x_0 = -a \) were set to zero and values greater than \( x_0 \) were shifted to \( P - x_0 \). In the case of a future increase in projected precipitation (more values greater than \( x_0 \)), this type of correction then led to an amplification of the climate change signal (Fig. 11b).

6. Summary and concluding remarks

A statistical bias correction has been applied to daily temperature and precipitation data from three GCMs and two emission scenarios. Both, original GCM and bias-corrected data were used to force two global hydrology models. The application of the bias correction has shown that it effectively improves both the mean and the variance of the precipitation and temperature fields in all but a few regions of the globe. For precipitation, it can also correct higher moments of the distribution. Consequently, both GHMs have reduced biases in simulated discharge for most catchments when using bias-corrected GCM data instead of the original ones. Moreover for LPJmL, the simulated vegetation patterns (not shown), and hence the related land surface processes, were largely improved by using the bias-corrected data. It was shown that the bias correction can alter the climate change signal for specific locations and months. The bias correction will lead to changes in the analyzed signal (a) if low precipitation amounts (or temperatures) are differently corrected as high amounts (due to different model biases leading to transfer functions with a slope notably deviating from one), and (b) if the distribution between low and high amounts changes in a future climate. For some regions, the impact of the bias correction on the climate change signal may be larger than the signal itself, thereby identifying another level of uncertainty that is comparable in magnitude to the uncertainty related to the choice of the GCM or GHM.

Even though only two GHMs were used in the present study, the results indicate that the uncertainty related to the choice of the GHM may be as large as the choice of the GCM with regard to the projected change signal. In
In this respect it was found that the calculation of potential evapotranspiration based only on temperature seems to lead to large overestimations of projected evapotranspiration, especially in tropical areas. Thus, such temperature-based formulations cannot be recommended for climate change impact studies in those areas.

As mentioned before, it is rather difficult to judge whether the impact of the bias correction on the climate change signal leads to a more realistic signal or not. In this respect, Giorgi and Coppola (2010) analyzed 18 AR4 GCM projections and found that projected regional precipitation changes are significantly correlated with the respective regional biases for about 30% of the seasonal/regional cases investigated. For temperature, only a negligible effect of the regional bias on the projected change was noticed. Even though the specific relation between present-day model bias and projected climate change signals is not clarified yet, these results suggest that, at least for precipitation, an impact of the bias correction on the climate change signal may be reasonable. For temperature, the assumption of relating current-day model and observed variability differences to future scenario sensitivity to changes in GHG boundary conditions has yet to be verified. For precipitation, this is a similar problem but the situation becomes even more complicated because additional current-day model and observational variabilities affect the signal, and deriving the transfer functions (producing adequate fits that are robust for present-day climate) is not trivial. This means that for temperature, the issues lie in the assumptions on how to relate current day-to-day variability to future climate warming (due to possibly unrelated causes), but for precipitation the current-day fluctuations are already much greater and the fits by themselves produce some additional noise. Moreover, precipitation and temperature are corrected independently. Several studies, such as that of Berg et al. (2009), have shown that daily precipitation shows some scaling with temperature so that future improvements of the bias correction method may be achieved with multivariate approaches that take these dependencies into account.

GCMs cannot be used for projection impact studies without some form of bias correction. When only the climate change signal is taken from simulations, instead of the raw GCM output, this is tantamount to applying a bias correction of the sole mean. That said, it is a matter of scientific debate whether the bias correction is adding or uncovering another level of uncertainty that is related to the uncertainty induced by the choice of the GCM. The latter seems to apply in cases where biases lead to positive regional feedbacks to the climate change signal. This, for example, may be the case in regions with strong land–atmosphere coupling where the coupling strength will change under future climate conditions. For Europe, Seneviratne et al. (2006) pointed out that land–atmosphere coupling is significantly affected by global warming and is itself a key player for climate change, thereby highlighting the importance of soil moisture–temperature feedbacks (in addition to soil moisture–precipitation feedbacks) for regional future climate changes. Van den Hurk et al. (2005) stated that if models overemphasize the positive land–atmosphere feedback that leads to a dry soil, strong evaporation stress, and reduced precipitation, this poses severe problems in the interpretation of hydrological aspects of climate change in future GHG emission scenarios. In this respect, the ECHAM5 results over the Amazon catchment (see section 4) may fall in this category so that potentially the bias correction might lead to an improved temperature signal. But in order to investigate this more thoroughly, sensitivity tests with prescribed soil moisture and/or atmospheric conditions are necessary, which is subject to future studies.
Some larger uncertainties occurred over several dry regions and seasons, especially for precipitation. Here, any projected changes in discharge and their subsequent impact on water resources have to be carefully considered, with as well as without using bias-corrected GCM data. The bias correction applied in the present study has only identified but not necessarily caused this extra level of uncertainty within the GCM-hydrology model (or any other impact model) modeling chain. How to handle and possibly reduce this uncertainty is an important question in climate change impact research. Thus, it will be subject to future investigations whose outcomes have to be communicated to the impact research communities.

Despite the recent progress in the development of GCMs, they still exhibit a number of significant systematic biases in their ability to simulate key features of the observed climate system (Randall et al. 2007). The further development of GCMs will certainly reduce biases in some areas (e.g., improved cloud microphysics may lead to a better representation of orography, enhanced simulation of blocking situations, etc.). On the other hand, new processes will be implemented into the GCMs (dynamic vegetation, biogeochemistry, aerosol chemistry, etc.), which will lead to more degrees of freedom that may also increase some of the biases. Consequently, the issue of climate model bias correction will be of interest within the next years, even though it is desirable that this will no longer be necessary in the long-term perspective.

Acknowledgments. This study was supported by funding from the European Union within the WATCH project (Contract 036946). The authors thank Tobias Stacke for implementing several modifications into MPI-HM. The GCM data were obtained from the CERA database at the German Climate Computing Center (DKRZ) in Hamburg. Additional data were provided by Nathalie Bertrand from IPSL.

REFERENCES
——, H. Götzel, D. Jacob, P. Lorenz, and E. Roeckner, 2009: Improved regional scale processes reflected in projected