

## Combined evaluation of MPI-ESM land surface water and energy fluxes

Stefan Hagemann,<sup>1</sup> Alexander Loew,<sup>1</sup> and A. Andersson<sup>2</sup>

Received 29 May 2012; revised 1 October 2012; accepted 27 November 2012; published 17 May 2013.

[1] To assess the robustness of projected changes of the hydrological cycle simulated by an Earth system model (ESM), it is fundamental to validate the ESM and to characterize its major deficits. As the hydrological cycle is closely coupled to the energy cycle, a common large-scale evaluation of these fundamental components of the Earth system is highly beneficial, even though this has been rarely done up to now. Consequently, the purpose of the present study is the combined evaluation of land surface water and energy fluxes from the newest ESM version of the Max Planck Institute for Meteorology (MPI-ESM), which was used to produce an ensemble of Coupled Model Intercomparison Project Phase 5 (CMIP5) simulations. With regard to energy fluxes, we especially make use of recent satellite data sets. Additionally, MPI-ESM results are compared with CMIP3 results from the predecessor of MPI-ESM, ECHAM5/MPIOM, as well as to results from the atmosphere/land part of MPI-ESM (ECHAM6/JSBACH) forced by observed sea surface temperature (SST). Analyses focus on regions where notable differences occur between the two ESM versions as well as between the fully coupled and the uncoupled SST-driven simulations. In general, our results show a considerable improvement of MPI-ESM in simulating surface shortwave radiation fluxes. The precipitation of the fully coupled simulations notably differs from those of the SST-forced simulations over a few river catchments. Over the Amazon catchment, the coupling to the ocean leads to a large negative precipitation bias, while for the Ganges/Brahmaputra, the coupling significantly improves the simulated precipitation.

**Citation:** Hagemann, S., A. Loew, and A. Andersson (2013), Combined evaluation of MPI-ESM land surface water and energy fluxes, *J. Adv. Model. Earth Syst.*, 5, 259–286, 10.1029/2012MS000173.

### 1. Introduction

[2] The climate of the Earth is influenced by increasing greenhouse gas concentrations, changing aerosol compositions and loads as well as 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.

[3] An accurate representation of the exchange of water between the atmosphere, the ocean, the cryosphere, and the land surface is one of the biggest challenges in

earth system 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 an Earth system model (ESM). 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 timescales, and inhomogeneities of the Earth's surface. All of these processes have to be properly represented in an ESM.

[4] The surface water and energy fluxes are closely related with each other as well as the terrestrial carbon fluxes. A strong coupling of the land surface dynamics with the atmosphere exists, especially in transitional climate regions [see, e.g., Koster *et al.*, 2004; Seneviratne *et al.*, 2010]. The feedback from the land surface can therefore have a strong effect on regional patterns of surface water and energy fluxes. Consequently, the focus of this study is on evaluating the skill of the current ESM version of the Max Planck Institute for Meteorology (MPI-ESM) to simulate land surface water and energy fluxes. MPI-ESM was used to produce an ensemble of Coupled Model Intercomparison Project Phase 5

<sup>1</sup>Max Planck Institute for Meteorology, Klima Campus, Hamburg, Germany.

<sup>2</sup>Deutscher Wetterdienst, CM-SAF, Offenbach, Germany.

(CMIP5) simulations for the forthcoming fifth Intergovernmental Panel on Climate Change (IPCC) assessment report. The evaluation mainly comprises the comparison of simulated and observed land surface climatologies of 2m temperature, precipitation, evaporation, and river runoff as well as surface solar radiation fluxes. In particular, we assess the dependence of biases in land surface water and energy fluxes on cases of fully or partially coupled MPI-ESM experiments to quantify differences due to the coupling with a dynamic ocean model. We also compare the results from the present MPI-ESM to results from its precedent version to quantify the differences between the two considerably different model versions of MPI-ESM (see section 2.1). Moreover, we quantify their accuracy compared with independent data sets. The analysis of other components of the hydrological cycle such as clouds, snow or soil moisture, is beyond the scope of this study and will therefore not be discussed.

[5] The model and the simulations considered in this study are briefly described in section 2. Simulated and observed components of the hydrological and energy cycle at the land surface are compared in section 3 on a large scale, and in section 4 for specific regions. Particular model biases in specific regions are discussed in more detail in section 5, and the main findings are summarized in section 6.

## 2. Description of MPI-ESM, Model Simulations Used and Observations

### 2.1. ESM of the Max Planck Institute for Meteorology

[6] The coupled MPI-ESM consists of the following model components: ECHAM6 in the atmosphere [Stevens *et al.*, 2012], MPIOM in the ocean [Jungclaus *et al.*, 2012], and JSBACH [Raddatz *et al.*, 2007; Brovkin *et al.*, 2009] for land surfaces. It takes into account observed concentrations of CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O, CFCs, O<sub>3</sub> (tropospheric and stratospheric), and sulphate aerosols, thereby considering the direct and first indirect aerosol effect. Major differences of MPI-ESM to its predecessor ECHAM5/MPIOM [Roeckner *et al.*, 2003; Jungclaus *et al.*, 2006] are a new radiative transfer scheme in the atmosphere, the use of a new aerosol climatology, and the incorporation of the carbon cycle including ocean biogeochemistry and an interactive and dynamic vegetation scheme at the land surface. Also the standard setup of MPI-ESM with 47 vertical atmospheric levels (LR) reaches to higher atmospheric layers than the previous one with 31 levels.

### 2.2. Model Simulations

[7] The experiments analyzed in this study were conducted following the CMIP5 protocol [Taylor *et al.*, 2012]. The following ESM ensemble simulations were considered in this study, comprising three members each. (1) An ensemble of fully coupled MPI-ESM “historical” (twentieth century) simulations conducted with the CMIP5 model setup, thereby focusing on the period 1971–2000 representing the current climate. This ensemble is referred as MPI-ESMh in the following. (2) An

ensemble of Atmospheric Model Intercomparison Project 2 (AMIP2) type simulations where the land/atmosphere component of MPI-ESM (ECHAM6/JSBACH) was forced with observed sea surface temperature (SST) and sea ice from 1979 onward. Here, we consider the period 1979–2000 and refer to this ensemble as MPI-ESMa in the following. (3) An ensemble of fully coupled ECHAM5/MPIOM historical twentieth century simulations that was conducted for CMIP3 [Program for Climate Model Diagnosis and Intercomparison (PCMDI), 2007] and used in fourth IPCC assessment report [Solomon *et al.*, 2007]. Comparisons focus on the period 1971–2000 and the ensemble is referred as ECHAM5h in the following. (4) In sections 4 and 5, we also show results from an AMIP2 simulation (1979–1999) using ECHAM5 with T63 horizontal resolution and 31 vertical atmospheric levels (ECHAM5a). This simulation has been evaluated by Hagemann *et al.* [2006], where the impact of different horizontal and vertical resolutions on the simulated hydrological cycle of ECHAM5 was considered.

[8] A notable difference of the setup of the land surface scheme JSBACH in MPI-ESM to those within the AMIP2 simulations is the use of a dynamical vegetation algorithm [Brovkin *et al.*, 2009] while in the latter a prescribed distribution of vegetation (i.e., Plant Functional Types (PFTs)) is used that is based on a 1 km global distribution of major ecosystem types [Loveland *et al.*, 2000]. Note also the difference of the land surface representation between ECHAM5 and MPI-ESM, where the first uses a static description of the land surface following Hagemann [2002], while the latter uses a dynamic model (see above) with interactive vegetation, e.g., comprising the phenological simulation of leaf area index and surface background albedo.

[9] We conducted climatological comparisons to observations and previous analogous ECHAM5 simulations. The time periods considered depend on the availability of model and observational data (see section 2.3). Note that for the coupled simulations MPI-ESMh and ECHAM5h we focus on the period 1971–2000 as this will be the common reference period for the quantification of climate change signals in studies of future global warming. With regard to the SST forced simulations, the available subset of this period will be used, i.e., 1979–2000 for MPI-ESMa. Results obtained with MPI-ESMh show that in general climatological differences between 1971–2000 and 1979–2000 are relatively small for the variables considered in this study, especially when compared with the differences to observations as reported in sections 3 and 4.

[10] Generally, simulated climatologies are directly taken from the model output of the respective variable. An exception is the surface albedo, which was calculated as the ratio of the monthly means of the upward and downward shortwave radiation fluxes to appropriately take into account changes in the fractions of direct and diffuse radiation during a month. This is necessary since the JSBACH land surface model in MPI-ESM and also ECHAM5 calculate the surface albedo as a function of the actual state of the vegetation, the

background albedo below the canopy and the amount of snow cover in a model grid cell [Roesch and Roeckner, 2006; Vamborg, 2011].

### 2.3. Observational Data

[11] For the evaluation of surface water and energy fluxes, various global observational data sets are used, which are summarized in Table 1 and introduced in the following. As the observations comprise some very recent data sets of surface solar irradiance (SSI), these are discussed in more detail.

#### 2.3.1. Hydrological Observations

[12] As observations, temperature and precipitation from the newly available global WATCH data set of hydrological forcing data (henceforth referred to as WFD [Weedon *et al.*, 2011]) were used. The WFD combine the daily statistics of the 40-years reanalysis of the European Centre for Medium-Range Weather Forecasts (ERA40 [Uppala *et al.*, 2005]) with the monthly mean observed characteristics of temperature from the Climate Research Unit data set TS2.1 [Mitchell and Jones, 2005] and precipitation from the Global Precipitation Climatology Centre full data set version 4 [Fuchs *et al.*, 2007]. For the latter, 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]. Note that the uncertainties in the precipitation data are still high in regions with a sparse station density and those with significant snowfall amounts, especially in mountainous areas [see, e.g., Rudolf and Rubel, 2005; Biemans *et al.*, 2009]. In addition, climatological observed discharge data were taken for major rivers from the Global Runoff Data Centre (GRDC; see, e.g., Dümenil Gates *et al.*, [2000]).

#### 2.3.2. Surface Albedo

[13] Ten years of Moderate Resolution Imaging Spectroradiometer (MODIS) surface albedo (MCD43C3, ver. 5) observations [Schaaf *et al.*, 2002] are used for comparison with MPI-ESM results. The albedo observations are filtered in accordance to the product quality flags to ensure that only best quality observations are considered for the reference data set. The data are then reprojected to the MPI-ESM model grid (T63 resolution, Gaussian grid). The mean surface albedo and its variance are calculated from the 10-year time series for each month and grid cell, respectively. With a record of just ten years of

observations, the estimation of robust climatic mean values of surface albedo is difficult. On the other hand, changes in vegetation cover might already change significantly the surface albedo on decadal timescales and therefore affect climate [Loew and Govaerts, 2010; Fensholt *et al.*, 2012]. Results shown in this paper are based on the full 10-year record of MODIS observations. However, we also analyzed the effect of sampling of the MODIS observations as well as the MPI-ESM simulations on shorter timescales (5 years) and found no significant differences to those shown in the present study.

#### 2.3.3. Surface Solar Irradiance

[14] Satellite based methods enable the estimation of the surface solar flux at high temporal and spatial resolutions at global scales. Satellite-based estimates of atmospheric radiation fluxes have been developed throughout the last 20 years. Different global products exist, which also allow to assess the interproduct variability of SSI. Four different surface solar radiation data sets are used for the present study. The Clouds and Earth Radiation Energy System (CERES) surface solar radiation fluxes are provided at global scale and are derived from measurements onboard of the Earth Observing System (EOS) Terra and Aqua satellites [Loeb *et al.*, 2012]. The CERES surface fluxes are obtained from the monthly TOA/Surface Averages (SRBAVG; edition 2) product for a limited time period (2000–2003). A new reprocessed and flux corrected Energy Balanced And Filled (EBAF) surface flux product, covering the period 2000–2010, became available after the analysis of the present paper had been finalized [D. R. Doelling, 2012, personal communication]. An analysis of the differences between climatologies derived from the two different data sets revealed no major differences and no impact on the results and conclusions of the present paper are therefore expected by using the time-limited CERES data for the present study.

[15] Although the CERES data record is limited to a few years, the recently released SSI data set of the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Satellite Application Facility on Climate Monitoring (CMSAF) covers the period from 1983 to 2005. The product is distributed on a regular latitude/longitude grid with a grid spacing of 0.03° in both directions and is limited to the spatial domain of Meteosat centered at 0°E (see Figure A1). Details on the retrieval scheme can be found in Müller *et al.* [2009, 2011]. A recent study investigated the

**Table 1.** Used Observational Data Sets

Variable	Data Set	Original Spatial Resolution	Original Temporal Resolution	References
Surface albedo	MODIS MCD43C3, ver. 5	0.05° × 0.05°	8 days	Schaaf <i>et al.</i> [2002]
Meteosat albedo	EUMETSAT	<5 × 5 km	10 days	Loew and Govaerts [2010]
SSI	CMSAF	0.03° × 0.03°	Hourly, monthly	Müller <i>et al.</i> [2009, 2011]
	CERES	1.0° × 1.0°	Monthly	Loeb <i>et al.</i> [2012]
	BSRN	Point	Hourly	Ohmura <i>et al.</i> [1998]
	SRB	1° × 1°	3 hourly, monthly	Cox <i>et al.</i> [2006]
	ISCCP	280 km	3 hourly, monthly	Zhang <i>et al.</i> [2004]
Temperature	WFD	0.5°	Daily, monthly	Weedon <i>et al.</i> [2011]
Precipitation	WFD	0.5°	Daily, monthly	Weedon <i>et al.</i> [2011]
Discharge	GRDC	Stations	Monthly	Dümenil Gates <i>et al.</i> [2000]

potential for the generation of a seamless surface solar radiation product from first and second generation of Meteosat satellites [Posselt *et al.*, 2011a].

[16] The accuracy of the CMSAF surface solar radiation product has been carefully investigated using reference measurements from the Baseline Surface Radiation Network (BSRN) [Ohmura *et al.*, 1998]. Ineichen *et al.* [2009] compared the CMSAF data product at hourly basis for eight stations in Europe and estimated accuracies of 80 to 100 W m<sup>-2</sup>. Posselt *et al.* [2011b] showed that the accuracy of the monthly mean product is better than 10 W m<sup>-2</sup> and better than other existing available surface radiation data products.

[17] The International Satellite Cloud Climatology project (ISCCP) Flux Data (FD) product provides information on estimated SSI at a spatial scale of 280 km and with a temporal resolution of 3 h to monthly values [Rossow and Zhang, 1995]. Further details on the estimation of surface fluxes for ISCCP are given in Zhang *et al.* [2004]. Data from 1989 to 2005 were used for the present analysis.

[18] The National Aeronautics and Space Administration (NASA)/GEWEX surface radiation budget (SRB) project aims at the production of long-term data sets of shortwave and longwave surface and top-of-atmosphere radiation fluxes. It provides 3 hourly to monthly flux estimates at global scale with a resolution of 1° × 1°. The fluxes are calculated based on cloud parameters obtained from ISCCP and meteorological fields from the NASA Global Modeling and Assimilation Office (GMAO) reanalysis data sets. This study uses monthly means of the SRB shortwave surface radiation flux product (version 3.0 [Cox *et al.*, 2006.]).

[19] All satellite data sets were reprojected to the MPI-ESM model grid at T63 resolution before further analysis using a conservative remapping procedure. The different SSI observational data sets exhibit systematic differences that may have to be considered in the comparisons to the model output. To illustrate these systematic differences, seasonal means of SSI differences are shown in Figure 1 using CERES data as a reference. Similar plots using CMSAF data are given in the Appendix. A comprehensive analysis of the accuracy of the different satellite flux estimates is beyond the scope of the present paper. A detailed analysis of the different data sets is conducted in the frame of the GEWEX radiation assessment (E. Raschke *et al.*, GEWEX radiative flux assessment (RFA), A project of the World Climate Research Programme Global Energy and Water Cycle Experiment (GEWEX) Radiation Panel, WCRP-Report, in preparation, 2013), which is also addressing temporal inconsistencies in the long-term data records caused by perturbations such as the Pinatubo eruption, El-Niño, as well as artifacts resulting from data processing. The significance of the difference between the seasonal means (Figure 1) was tested using a Student's *t* test.

[20] The ISCCP data set shows consistent differences to CERES as well as to the CMSAF data set. Largest differences are observed over the tropical Atlantic, the Sahara, as well as over Eurasia during the northern hemispheric summer and over the Antarctic ocean

during the southern hemispheric summer. These differences might be due to a different characterization of sea ice in the different data products.

[21] Although the SRB data set is based on ISCCP cloud properties as an input, differences in the surface radiation fluxes can be observed, which are due to different algorithms employed as well as different ancillary data used for flux calculations. While the flux differences to CERES and CMSAF data are rather similar to those for ISCCP in the Extratropics, SRB shows smaller surface solar radiation fluxes in the tropical areas than ISCCP. The difference plot in Figure 1 also shows artifacts caused by the spatial coverage of the different satellites used for the derivation of the data products. Over Africa, the spatial coverage of the Meteosat satellites is clearly visible in the difference images, which is likely to be an artifact of the data processing.

[22] However, it needs to be emphasized that the observed differences between the different data products are not statistically significant in most cases. Only for a very few grid boxes, statistically significant differences between the seasonal means can be detected, which is due to the large variability of SSI between the different years. Thus, mostly a comparison of simulated SSI to one data set, in this case CERES data, is sufficient for the evaluation. For better illustration, additional comparisons to CMSAF data are provided in the Appendix.

#### 2.3.4. BSRN Data

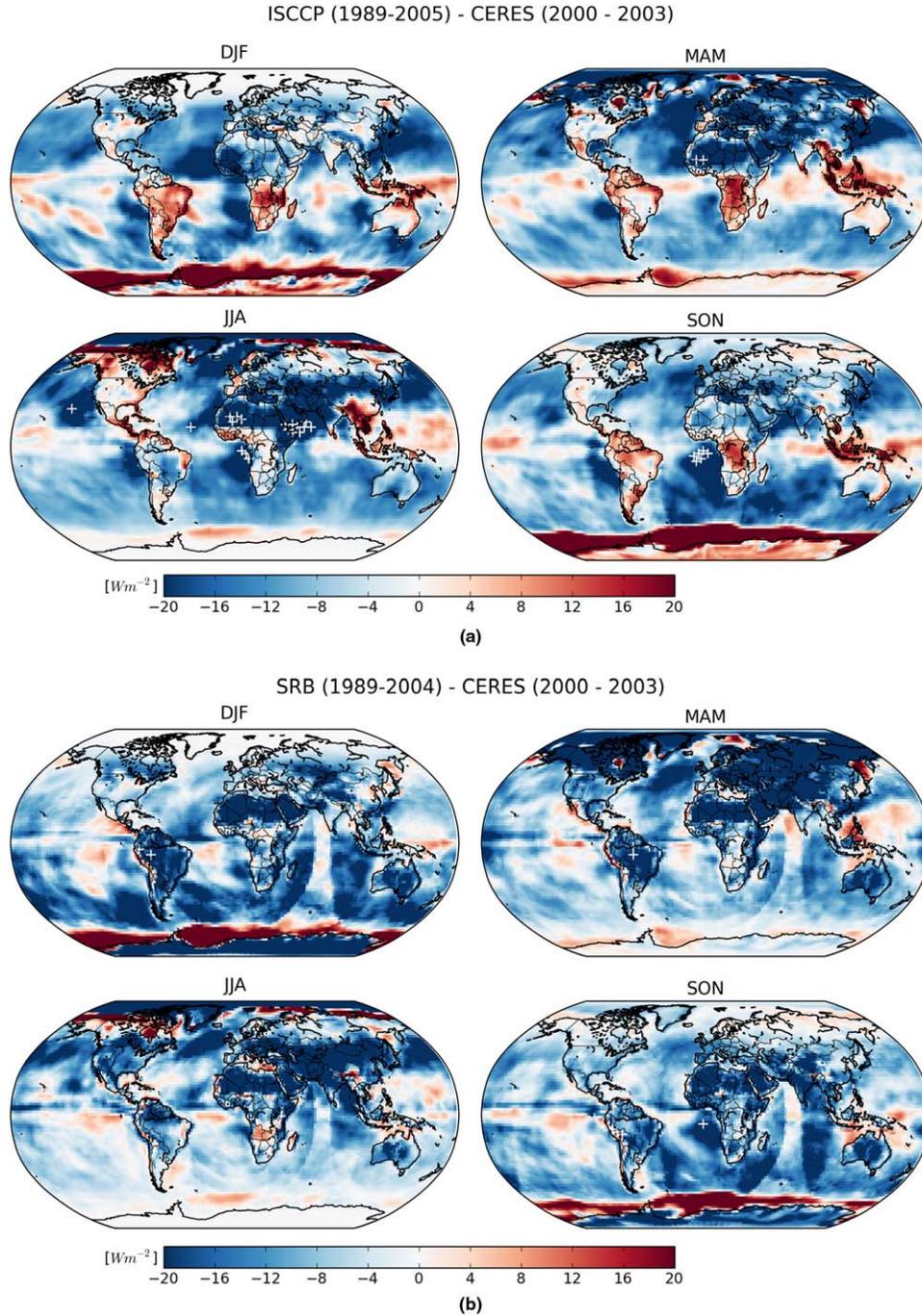
[23] The BSRN provides high quality in situ measurements of the surface solar radiation flux at the local scale. A total of 61 stations collect surface radiation data on a regular basis [Ohmura *et al.*, 1998]. BSRN station measurements are used for an intercomparison with the model simulations on a climatological timescale. While the in situ data correspond to point like observations, the radiation simulated in the model corresponds to an average flux over an entire model grid box. However, characteristic seasonal patterns of the radiation flux can be extracted from both, model simulations and in situ measurements and be compared. From the available 61 BSRN stations, only those were used where at least 60 continuous months of measurements were available to compile a monthly climatology of SSI. This resulted in a total of 39 BSRN stations worldwide (see Figure 2). As the CMSAF SSI data cover only a limited part of the globe, only 16 BSRN stations can be used for comparison. The data were obtained from the BSRN archive (<http://www.bsrn.awi.de/>). Climatological mean SSI and its variability were then calculated for each station from the entire available record of a BSRN station.

### 3. Results of Model Evaluation

[24] Results of the MPI-ESM evaluation are analysed in the following, focusing on global difference maps and zonal mean statistics for the different variables analysed in this study.

#### 3.1. Precipitation

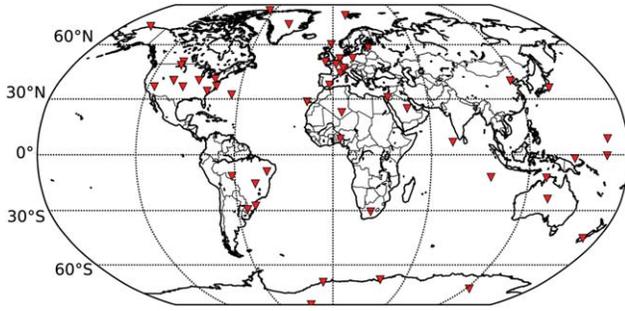
[25] Considering the zonal distribution of precipitation over land (Figure 3), the general shape of the



**Figure 1.** Differences of seasonal cycle of SSI between different observational data sets: ISCCP compared to (top) CERES, SRB compared to (bottom) CERES. Statistically significant differences ( $p > 0.05$ ) between the satellite products are indicated by stippled areas.

distribution is captured well by all models for all seasons. Common deviations from the WFD denote a pronounced dry bias in the Tropics north of the equator, which also extends to south of the equator during the boreal spring and summer, a dry bias in the low precipitation region in the southern Subtropics accompanied by a wet bias around  $50^{\circ}\text{S}$ , and a wet bias in the northern high latitudes during boreal spring and summer.

The most notable difference between the models is that MPI-ESMh and MPI-ESMa generally have an improved simulation of peak rainfall in the tropics compared with ECHAM5h. In the boreal summer, MPI-ESMh even simulates a better peak than MPI-ESMa while this peak is better captured by MPI-ESMa during the boreal winter and spring. Over the high northern latitudes in the boreal spring and summer, the slightly



**Figure 2.** Distribution of 39 BSRN stations with long-term SSI measurements.

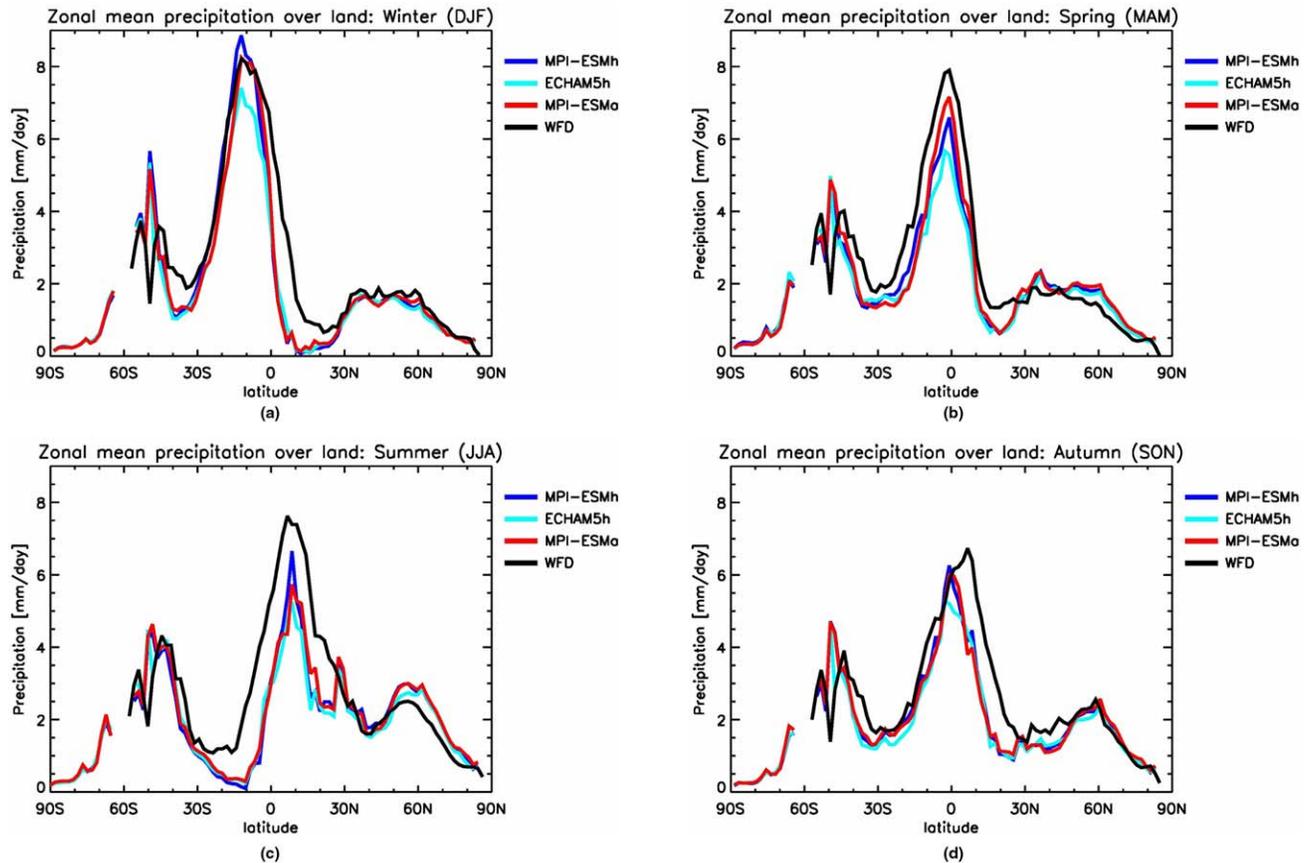
lower precipitation in ECHAM5h is somewhat closer to WFD than for the other two models.

[26] Considering the spatial differences to WFD (Figure 4), further common deficiencies are the overestimation of precipitation along steep mountain slopes (Andes, Himalayas, Rocky Mountains), distinct dry biases in the northern part of South America and around the Sahel zone that extend further south during the boreal summer. In the boreal winter, a large wet bias over eastern Asia can be noted. The pronounced dry bias in ECHAM5h over central and southern

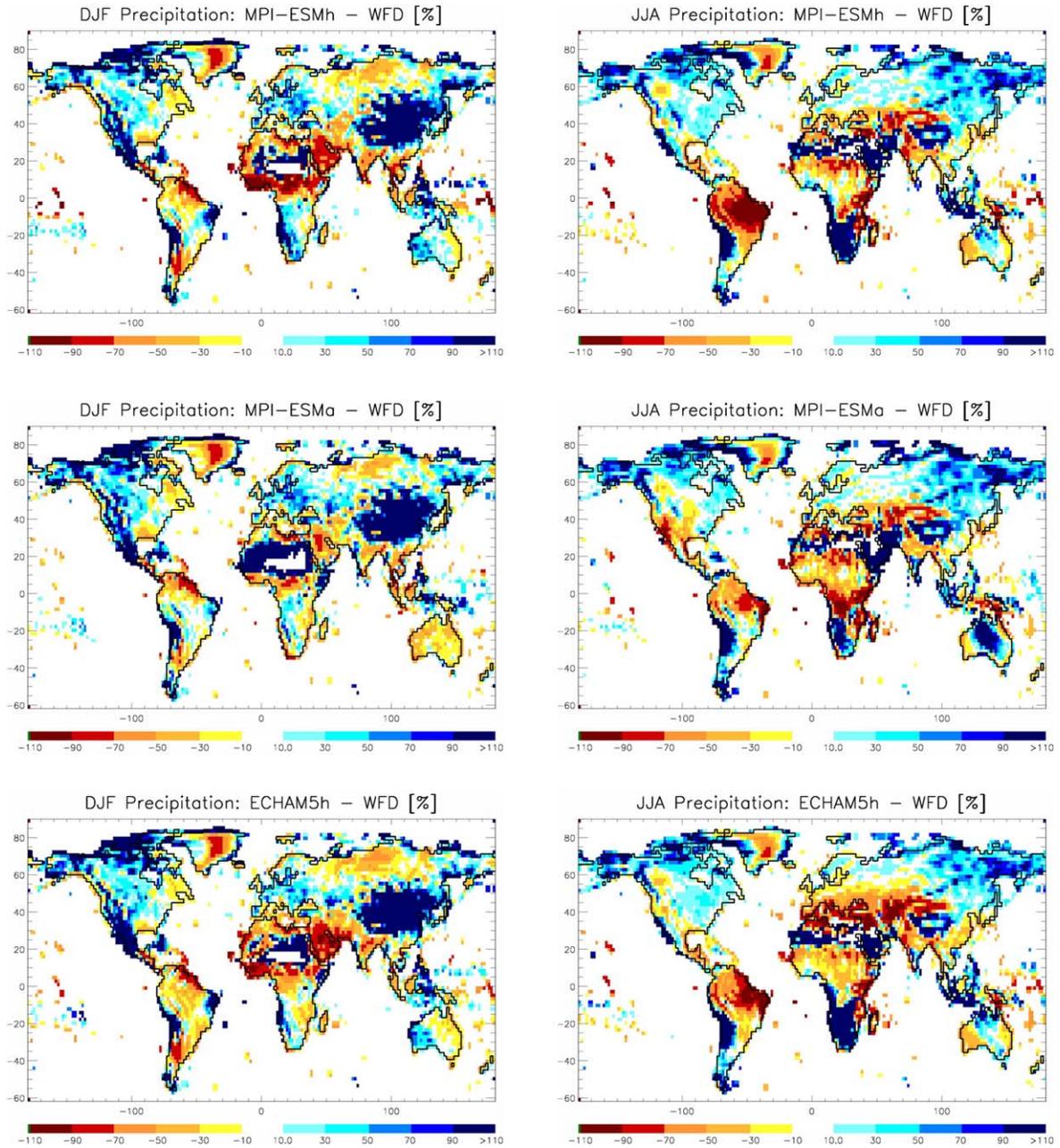
Europe during the boreal summer is largely reduced in both MPI-ESM simulations.

### 3.2. Temperature

[27] Figure 5 shows the zonal distribution of the 2m temperature difference to the WFD over land averaged over the four seasons. All models are close (within  $\pm 1\text{K}$ ) to the WFD in the tropics in all seasons, but show pronounced warm biases in the northern mid- and high latitudes in the boreal winter and spring, which is replaced in the northern hemisphere by a cold bias in the boreal summer. Further, a warm bias around  $30^\circ\text{S}$  can be noted that might be related to the dry bias in the southern Subtropics (see above). In the boreal summer, this extends to the southern mid-latitudes while in the boreal autumn and winter, a cold bias becomes apparent. MPI-ESMh tends to be colder than the other two models south of  $30^\circ\text{N}$  and shows the smallest bias compared with the WFD in the tropics and southern subtropics in the boreal summer and autumn (reduced or even no warm bias around  $30^\circ\text{S}$ ). In the northern high latitudes, ECHAM5h is colder than MPI-ESMh and MPI-ESMa, the latter being the warmest in this region. Thus, ECHAM5h has the smallest warm bias in winter and spring, while MPI-ESMa has a strongly reduced cold bias around  $60^\circ\text{N}$  in the boreal summer.



**Figure 3.** Zonal distribution of precipitation over land in the boreal (a) winter, (b) spring, (c) summer, and (d) autumn. Unit:  $\text{mm d}^{-1}$ .



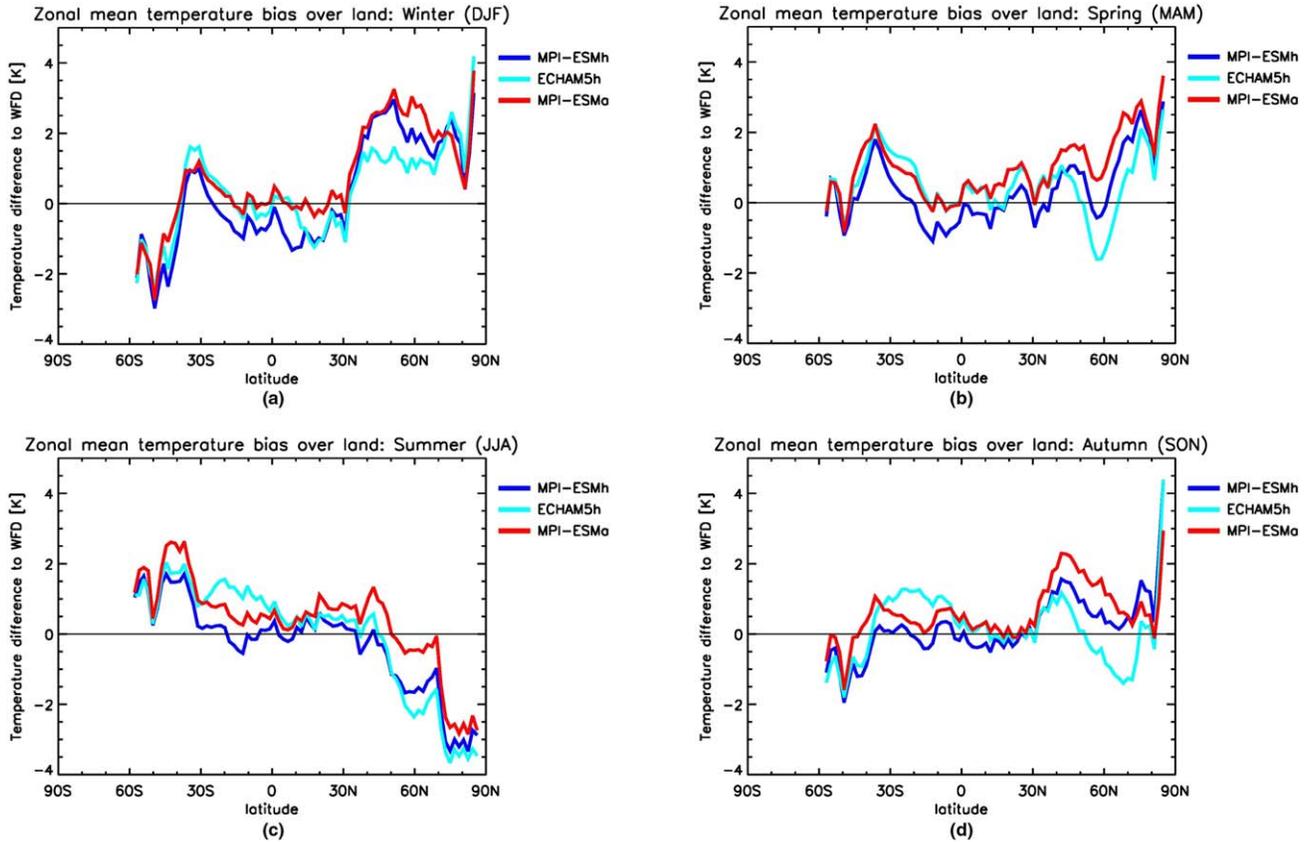
**Figure 4.** Relative precipitation difference to the WFD over land in the boreal winter (left column) and summer (right column) for MPI-ESMh (upper row), MPI-ESMa (middle row), and ECHAM5h (lower row). Unit: %.

[28] Maps of the 2m temperature difference to the WFD (Figure 6) reveal that the winter warm bias is spread throughout the whole range of northern latitudes except for Greenland and the west coasts of North America and Europe, and it is most pronounced in Eastern Siberia. The reduced warm bias in ECHAM5h stems from generally colder simulation of temperatures over northern Eurasia. In the boreal summer, the reduced cold bias of MPI-ESMa in the high northern latitudes is not only related to a strongly improved simulation of temperatures over Eurasia but also to the compensating effect of a warm bias over Central United States that

does not appear for the other two models. The general cold bias over Greenland is likely related to the too simple treatment of the glacier ice sheets in the models.

### 3.3. Surface Solar Radiation

[29] The differences in SSI between the different models and experiments are illustrated in Figure 7. The MPI-ESM simulations are very similar for the global mean field with a global average of  $162.5 \pm 94.2 \text{ W m}^{-2}$  and  $162.2 \pm 94.2 \text{ W m}^{-2}$  for MPI-ESMa and MPI-ESMh experiments, respectively. Observed seasonal differences are not significant except for small regions in



**Figure 5.** Zonal distribution of 2m temperature difference to the WFD over land in the boreal (a) winter, (b) spring, (c) summer, and (d) autumn. Unit: K.

the inner tropics during boreal winter as well as over the Benguela current during June–July–August/September–October–November (JJA/SON). Spatial differences in SSI are mainly related to changes in the cloud pattern that also reflect in changes in the simulated precipitation patterns (A. Andersson et al., Evaluation of MPI-ESM ocean surface fluxes, *J. Adv. Model. Earth Syst.*, in preparation, 2013, hereinafter referred to as Andersson et al., in preparation, 2013).

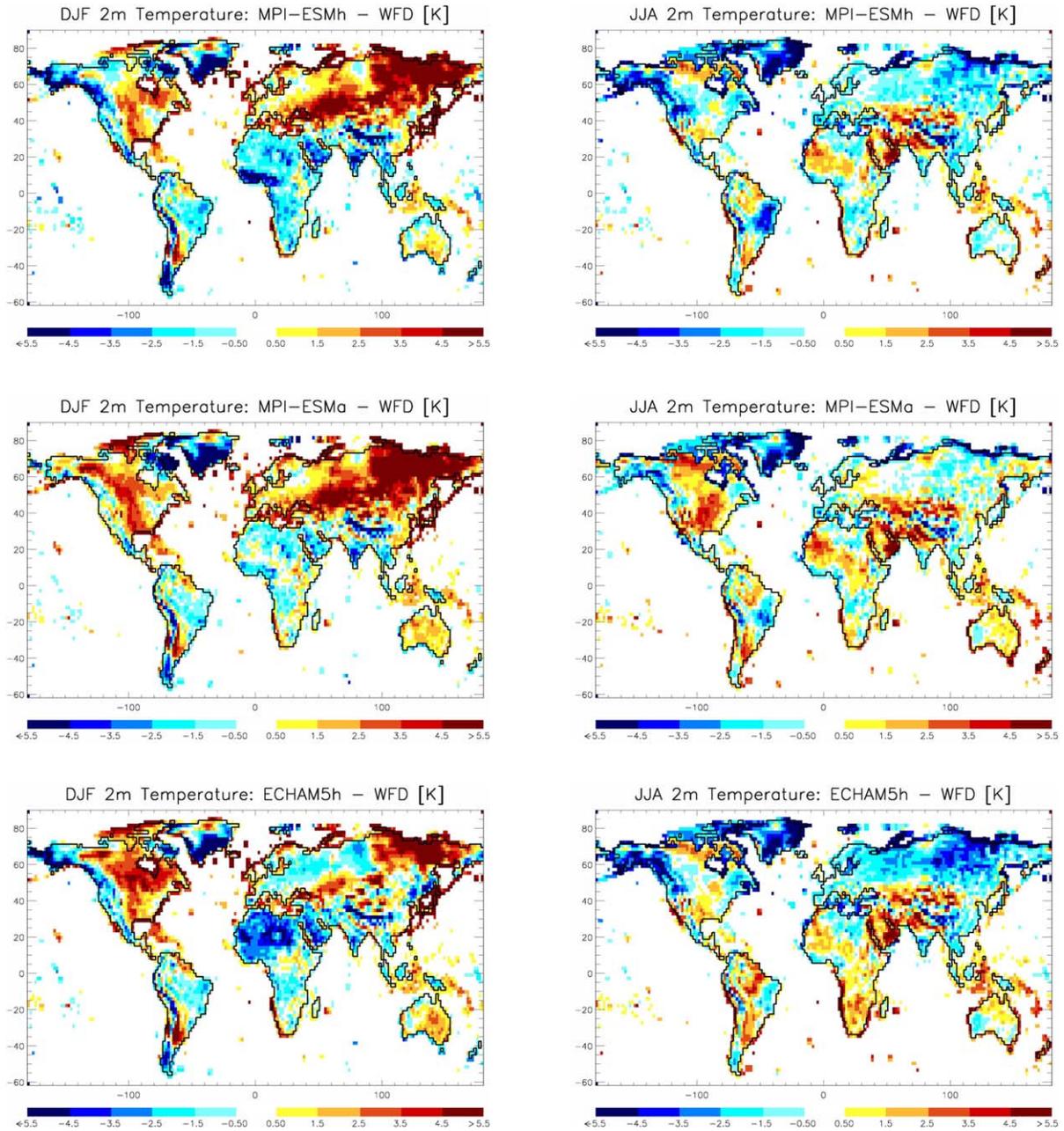
[30] Figure 7b shows the SSI differences between MPI-ESMh and ECHAM5h. The global mean of ECHAM5h is lower ( $156.1 \pm 95.5 \text{ W m}^{-2}$ ) than the corresponding simulation results from MPI-ESM. Largest differences are observed over Eurasia and the tropical Africa with deviations  $>20 \text{ W m}^{-2}$ . However, the estimated differences are only statistically significant in central Africa as well as in the Eastern Pacific. The observed differences between the model versions are due to different cloud coverage as well as aerosol optical depth.

### 3.3.1. Evaluation Using BSRN Station Data

[31] Comparison of simulated SSI fields with BSRN station data shows high correlations for all stations and models. Figure 8 shows the distribution of obtained Pearson correlation coefficients and root-mean-square error (RMSE;  $\text{W m}^{-2}$ ) for a total of 39 stations the global radiation products and 16 stations for CMSAF.

Overall, the CMSAF SSI product shows highest correlations and smallest RMSE for this climatological comparison, which is also the case if RMSE and correlation of CERES, SRB, and ISCCP are calculated only for the 16 stations covered by the CMSAF data. This result is consistent with validation results of the CMSAF data product (see also section 2.3), which showed that this satellite-based product has an accuracy better than  $10 \text{ W m}^{-2}$  and shows smaller errors compared with other existing solar irradiance products [Posselt et al., 2011b]. The RMSE obtained in this study is slightly larger, which is explained by the much larger scale discrepancy between the BSRN stations and the model grid scale. The other satellite data sets show slightly higher RMSE compared with the CMSAF data. While ISCCP and CERES show an almost similar spread in the RMSE as the CMSAF data ( $\sigma < 10 \text{ W m}^{-2}$  inner quartile range), the SRB inner quartile range is  $14 \text{ W m}^{-2}$ .

[32] The MPI-ESM model simulations have a median RMSE of  $14$  ( $12$ )  $\text{W m}^{-2}$  for MPI-ESMa (MPI-ESMh) with a similar variability. Contrary, ECHAM5h has a higher RMSE of  $18 \text{ W m}^{-2}$  with a much larger spread. This indicates that the improvements in the new radiation scheme applied in MPI-ESM as well as the new aerosol climatology have improved the skill of the model to simulate SSI. In the following, we will compare the model results against the satellite-based estimates of SSI.

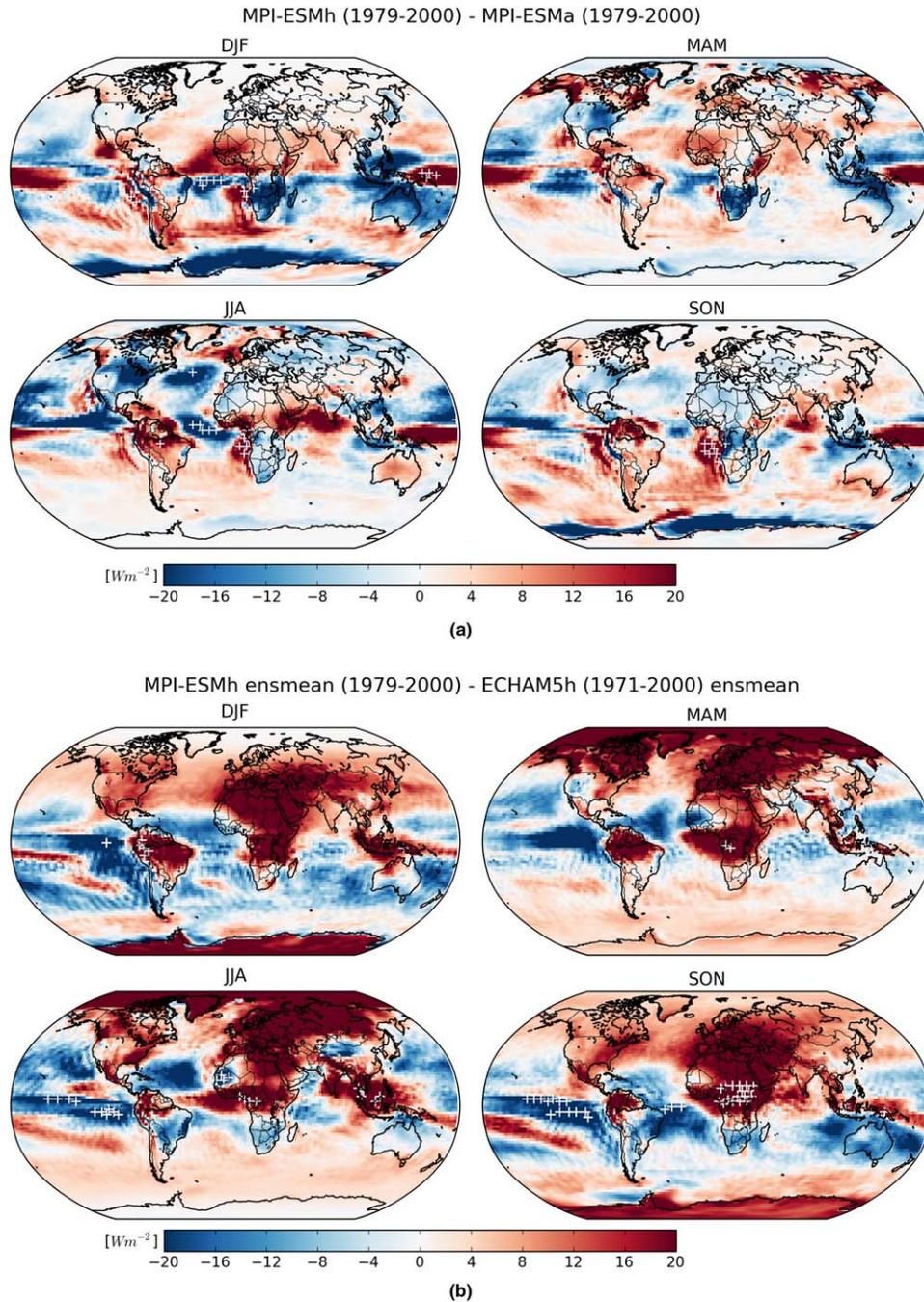


**Figure 6.** The 2m temperature difference to the WFD over land in the boreal winter (left column) and summer (right column) for MPI-ESMh (upper row), MPI-ESMa (middle row), and ECHAM5h (lower row). Unit: K.

### 3.3.2. Results of Spatial SSI Comparison

[33] The differences between the various satellite-based surface radiation products were discussed in section 2.3. Table 2 summarizes the global mean annual and seasonal SSI for the various model simulations and observations, whereas Figure 9 shows corresponding zonal means of SSI over land. In general, there is a good agreement between the different data sets and the models. Largest differences occur during the summer season of each hemisphere. The estimated global mean surface solar radiation flux is  $162 \text{ W m}^{-2}$  for MPI-ESM, for both the coupled and uncoupled simulations

and therefore well within the range of uncertainties obtained from the analysis of reanalysis data [Trenberth *et al.*, 2009, 2011]. The satellite-derived global mean SSI flux is  $169.3$ ,  $164.7$ , and  $160.8 \text{ W m}^{-2}$  for CERES, ISCCP, and SRB, respectively. ECHAM5h has a much lower global mean of  $156.1 \text{ W m}^{-2}$ . Overall, the differences between ECHAM5h and the new MPI-ESM are larger than the differences between the coupled and uncoupled experiments of MPI-ESM. Statistically different seasonal means occur between MPI-ESM, ECHAM5h, and the various observational data sets. The same regions with significant differences are identified in both MPI-ESM

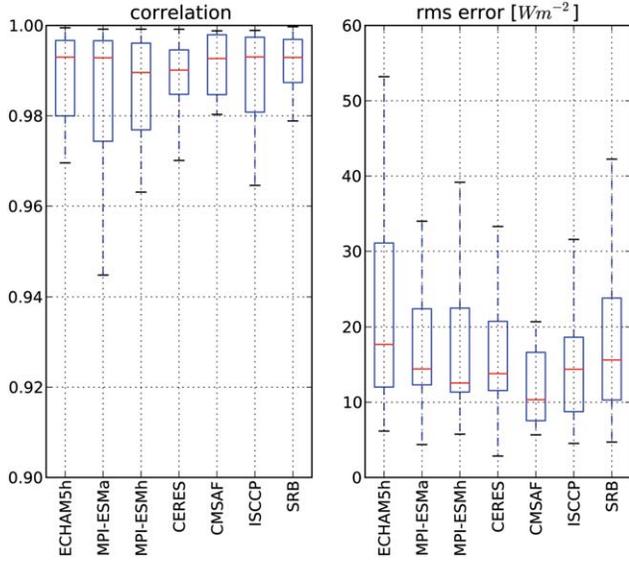


**Figure 7.** Differences in seasonal SSI ( $\text{W m}^{-2}$ ) between (a) MPI-ESMh and MPI-ESMa and (b) between MPI-ESMh and ECHAM5h. Statistical significant differences ( $p > 0.05$ ) are indicated by stippled areas.

experiments. We therefore only show results for MPI-ESMh in Figure 10 (and Figure A2). Consistent differences of MPI-ESM occur with all observational data sets. Statistically significant differences of SSI are mainly evident during December-January-February (DJF) and JJA when comparing against CERES data. For DJF, SSI is underestimated for the southern hemispherical tropical ocean and land surfaces. In JJA, a statistically significant underestimation of SSI is shown over large parts of the

northern hemisphere Atlantic and Pacific oceans. Significant positive differences occur in the western Pacific throughout all seasons. In general, similar patterns are simulated by ECHAM5h.

[34] While the differences between simulated and observed SSI reveal generally consistent patterns and are very useful to identify regions with significant differences, an overall assessment of the accuracy of the different model experiments is difficult. We therefore calculate a



**Figure 8.** BSRN error statistics: correlation and RMSE of climatological mean seasonal cycle of SSI. Boxes correspond to lower and upper quartiles, whiskers correspond to the data range, and red lines signify the median value.

single scalar normalized difference ( $e^2$ ) between the model results ( $SSI_m$ ) and a reference data set ( $SSI_{ref}$ ), following Reichler and Kim [2008] as

$$e^2 = \sum_n \left( w_n (SSI_{m,n} - SSI_{ref,n})^2 / \sigma_{ref,n}^2 \right),$$

whereas  $w_n$  are proper weights for changes in grid cell area,  $\sigma_{ref}^2$  corresponds to the interannual variance of the reference data set, and  $n$  is an index over all grid boxes investigated. The normalized error provides a relative ranking of different data sets. The CERES as well as CMSAF data sets were used as references for the calculation of  $e^2$ . Figure 11 shows results of this ranking with the dimensionless  $e^2$  on the  $x$  axis, stratified by season. The differences in  $e^2$  between CMSAF and CERES data are due to the different spatial domains covered by the two data sets. In general, the observational data sets (ISCCP, SRB) show smaller errors than the model simulations, indicating a general higher agreement between the different observational data sets, as could be expected. SRB tends to have slightly larger values for  $e^2$  for both CERES and CMSAF data used as a reference.

[36] The ECHAM5h simulations show the highest errors and a considerable improvement of the simulated SSI is found for both MPI-ESM experiments, proving the increased capability of MPI-ESM to simulate SSI using its refined atmospheric radiative transfer scheme and new aerosol climatology. The relative ranking of the different data sets remains consistent throughout the seasons and between the two reference data sets. Major changes in  $e^2$  are observed between the seasons with highest errors in DJF and JJA respectively.

### 3.4. Surface Albedo

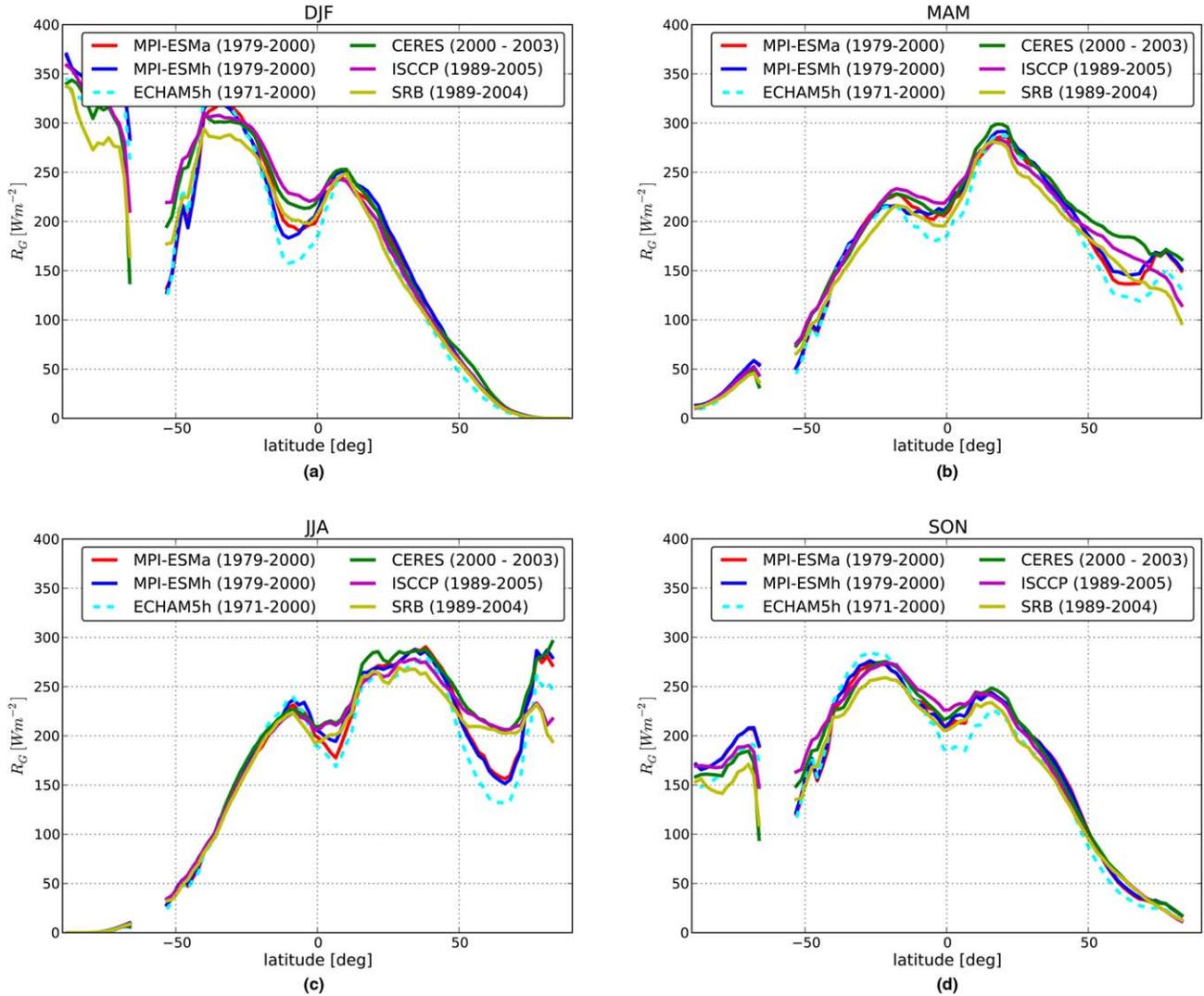
[37] The surface albedo of the MPI-ESM simulations is very similar. Seasonal values of zonal means of surface albedo are shown in Figure 12. On average, the MPI-ESM simulations show good agreement with the MODIS observations, except for high latitudes where larger differences occur due to snow. However, these larger absolute deviations have a small effect on the total surface radiative fluxes, as the SSI is small during boreal winter. Zonal means of ECHAM5h show a similar performance as the MPI-ESM simulations compared with the MODIS observations. To better assess the importance of the albedo differences for the surface radiation fluxes, the differences between model simulations and MODIS observations are expressed in terms of differences of the upward shortwave radiation flux which is obtained by scaling the surface albedo by the model SSI. As mentioned in section 2.2, all results are presented for the ensemble mean of three ensemble members of the model experiments. However, a sensitivity analysis (not shown) revealed that the differences caused by the internal variability of the model are much smaller than the differences yielded between the model simulations and the satellite observations. Figure 13 shows difference maps of upward shortwave flux between the ESM simulations and MODIS data.

[38] In general, the land surface albedo (upward flux) is slightly overestimated by the model almost everywhere during summertime, except for the Sahel region and some desert regions in Asia and Australia. A significance test of the differences between simulated and observed surface shortwave upward flux showed that the observed differences are significant ( $p < 0.05$ ) nearly everywhere on the globe. Differences between model and observations are most pronounced during boreal winter (DJF) and spring (March–April–May (MAM)) season with snow cover in high latitudes. In this period, the surface albedo dynamics are affected by snow cover and its masking by tree cover [e.g., Essery *et al.*, 2009]. In the northwestern part of North America (British

**Table 2.** Global Means of Seasonal SSI for the Simulations and Observational Data Sets<sup>a</sup>

Data Set	DJF	MAM	JJA	SON	Mean
MPI-ESMa (1979–2000)	177.7 (111.9)	157.4 (82.2)	152.1 (96.5)	162.9 (81.0)	162.5 (94.2)
MPI-ESMh (1979–2000)	176.7 (111.5)	157.9 (82.7)	151.8 (96.1)	162.5 (81.5)	162.2 (94.2)
ECHAM5h (1971–2000)	172.6 (111.6)	150.4 (85.6)	144.3 (94.7)	157.3 (85.3)	156.1 (95.5)
CERES (2000–2003)	179.8 (110.0)	165.7 (83.8)	167.8 (101.9)	163.8 (82.4)	169.3 (95.5)
ISCCP (1989–2004)	179.1 (111.4)	158.3 (82.1)	159.3 (97.5)	162.2 (80.5)	164.7 (94.1)
SRB (1989–2004)	172.2 (105.1)	155.5 (81.6)	158.7 (96.0)	156.8 (79.6)	160.8 (91.4)

<sup>a</sup>Values in brackets correspond to spatial standard deviations. All values are in  $W m^{-2}$ .



**Figure 9.** Zonal distribution of SSI over land in the boreal (a) winter, (b) spring, (c) summer, and (d) autumn. Unit:  $\text{W m}^{-2}$ .

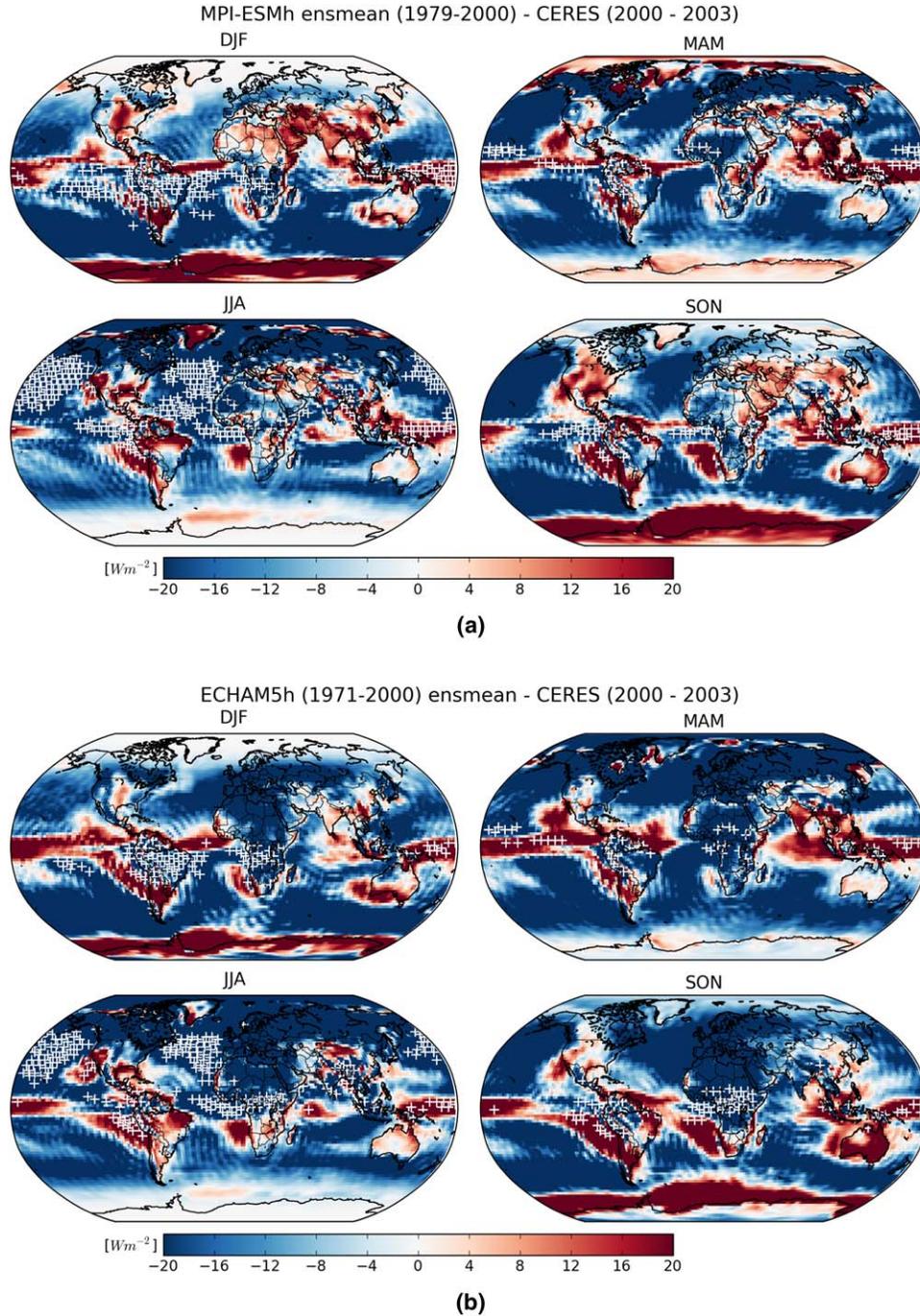
Columbia), the model significantly overestimates the surface albedo (upward flux) throughout the year. *Brovkin et al.* [2012] concluded that this positive bias is mainly due to an underestimation of the tree coverage by the model in this area. It also can be noted that the Sahara is brighter in ECHAM5h compared with MPI-ESMh, which is closer to the observations in this area. This can be attributed to the bare soil correction of albedo with Meteosat data in the LSP2 data set [*Hagemann, 2002*] that was used in ECHAM5h. Here, these data seem to be more adequate than the soil albedo data in MPI-ESMh that were derived from MODIS data by *Rechid et al.* [2009].

[39] The overall performance of the three models to simulate the surface albedo was estimated by calculating the intermodel performance index of *Reichler and Kim* [2008], similar to the analysis performed for the SSI, which provides a relative ranking of the individual experiments compared with the multimodel mean. Both

experiments with MPI-ESM clearly outperform its predecessor ECHAM5. The experiment with forced SST (MPI-ESMa) is by 39% better than the average, while the MPI-ESMh is still 15% better. Contrary, ECHAM5h is by 54% worse than the multimodel mean. It should be emphasized that the *Reichler and Kim* [2008] skill score provides only a relative ranking of the different models and experiments and is not an absolute measure of model performance. Nevertheless, it clearly demonstrates the improvement of surface albedo simulations by MPI-ESM using a dynamic surface albedo scheme, compared to its predecessor.

#### 4. Regional Validation Over Large Catchments

[40] Results from a regional analysis for selected major catchments are shown in the following. The distribution of catchments selected for the model validation is shown in Figure 14. To represent closed

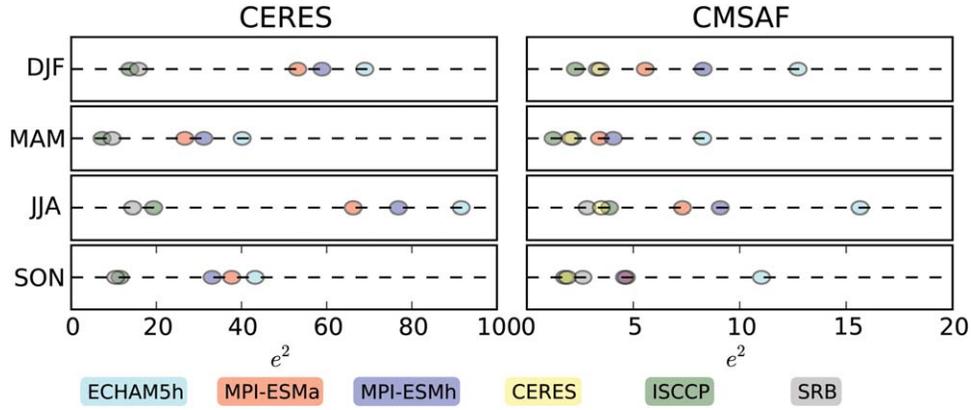


**Figure 10.** SSI differences of (top) MPI-ESMh and (bottom) ECHAM5h to CERES.

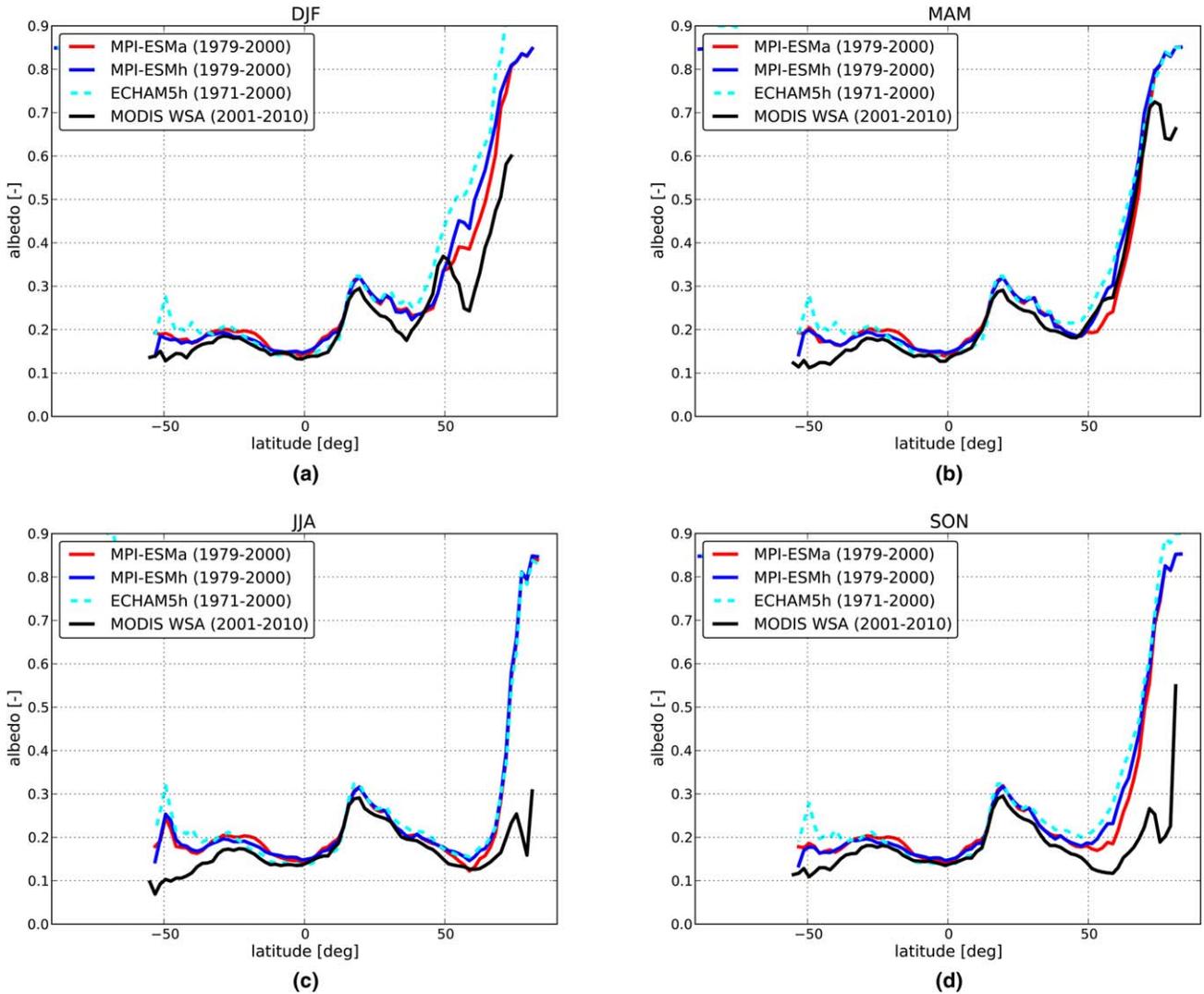
hydrological units over the different continents, the largest rivers on Earth are included as well as a few smaller ones in Europe (Baltic Sea, Danube) and Australia (Murray). Biases of annual mean 2m temperature, precipitation ( $P$ ), evapotranspiration ( $E$ ), and runoff ( $R$ ) are shown in Figures 15 and 16. As the accuracy of global observational evapotranspiration data sets [e.g., Jiménez *et al.*, 2011; Mueller *et al.*, 2011] is highly uncertain, evapotranspiration has been diagnosed as  $E = P - R$  by assuming that the long-term storage of soil water and snow is negligible. The observational values

used to calculate the biases are given in Table 3. In addition to MPI-ESMh, MPI-ESMa, and ECHAM5h, results from an AMIP2 simulation (1979–1999) using ECHAM5 are also shown (ECHAM5a, see section 2.2).

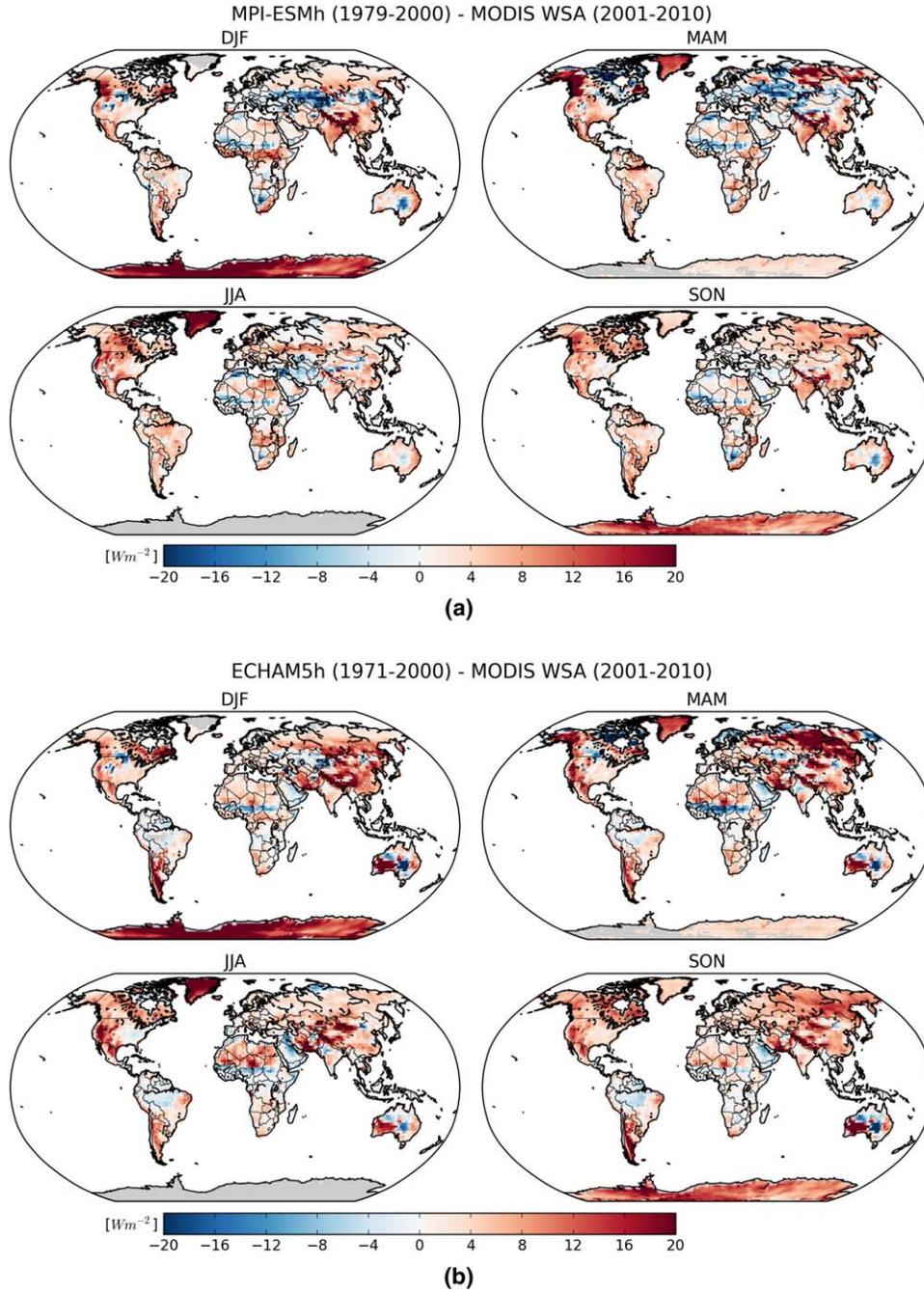
[41] For the Eurasian high and mid-latitude catchments (Arctic rivers, Amur, Baltic Sea, Danube), the MPI-ESM simulations are warmer than their ECHAM5 counterparts (Figure 15a; see also section 3.2), which for the Danube shows up in a warm bias of about 1 K that does not occur in both ECHAM5 simulations. For most of the catchments, MPI-ESMh has the smallest temperature



**Figure 11.** Relative normalized SSI error ( $e^2$ ) (–) for different model experiments as well as observations, relative to (left) CERES and (right) CMSAF observations. Note differences in scale due to the smaller spatial coverage of the CMSAF field.



**Figure 12.** Zonal distribution of surface albedo over land in the boreal (a) winter, (b) spring, (c) summer, and (d) autumn.



**Figure 13.** Differences in upward shortwave flux ( $\text{W m}^{-2}$ ) compared to MODIS data scaled with model incoming radiation for (top) MPI-ESMh and (bottom) ECHAM5h.

bias, or is at least close to the smallest bias. Here, apart from the Danube (see above) only the Congo and Yangtze Kiang stick out where MPI-ESMh shows a much stronger cold bias than the other three models.

[42] With regard to precipitation (Figure 15b), differences between the two historical simulations seem to be of minor importance. For most catchments, MPI-ESMh precipitation is somewhat larger than for ECHAM5h, thereby leading to some enhancement of the wet biases over the

Arctic rivers, Amur, Baltic Sea, and Yangtze Kiang, and to an apparent reduction of dry biases over the Danube, Congo and Parana. For the Ganges/Brahmaputra, the wet bias is almost halved in MPI-ESMa compared with ECHAM5a. Further difference between the two AMIP2 simulations can be noted for the Arctic rivers (increased wet bias in MPI-ESMa), Yangtze Kiang (decreased wet bias), Murray (removed dry bias), and Nile (10% dry bias instead of a similar wet bias).

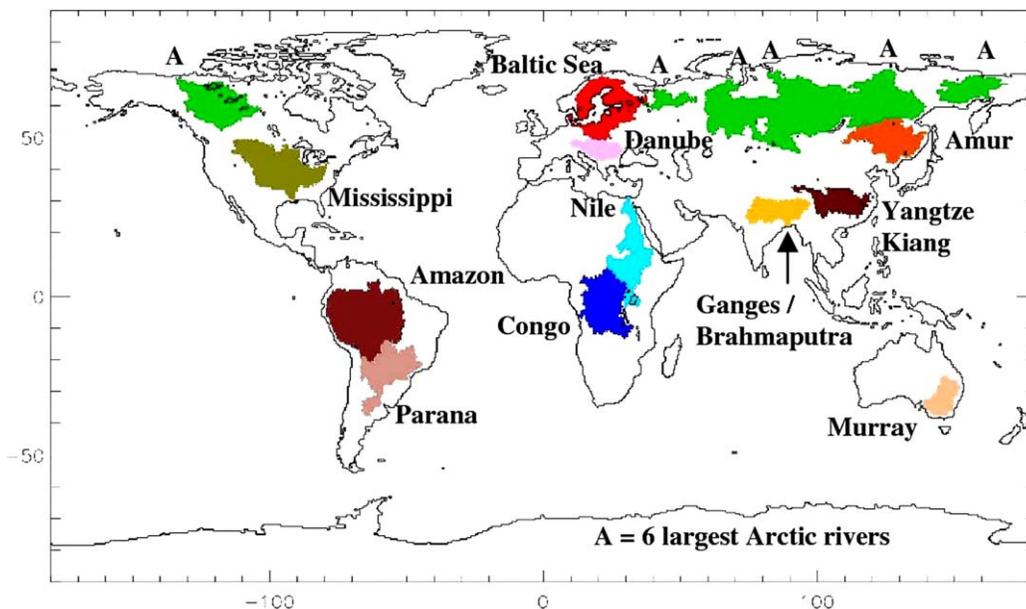


Figure 14. Selected large catchments of the globe at 0.5° resolution.

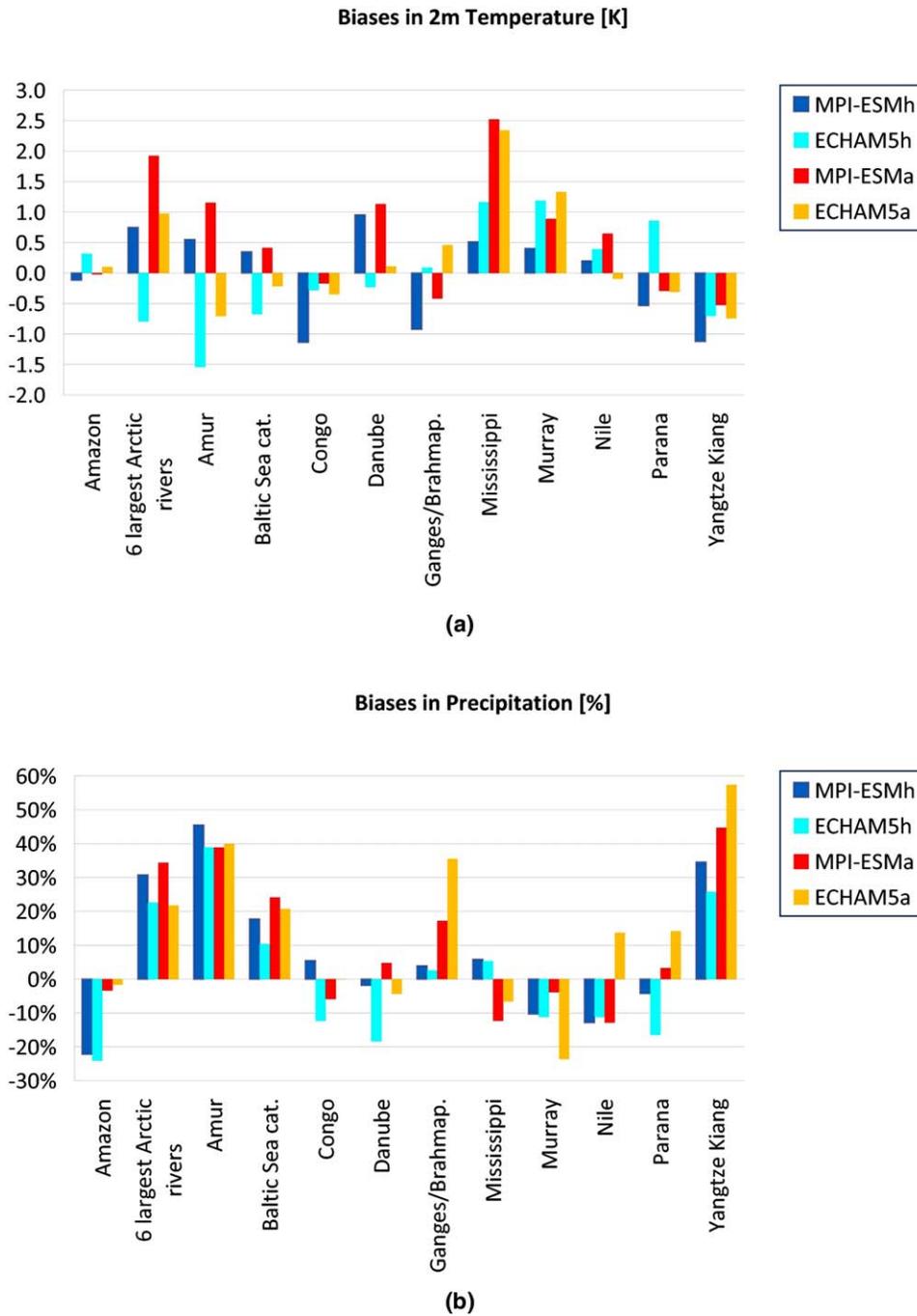
[43] Biases and differences between the models in evapotranspiration (Figure 16a) generally follow those in precipitation, e.g., the positive biases over Arctic Rivers, Amur, Baltic Sea and Yangtze Kiang of all models as well as the reduced negative bias for the Murray and the enhanced negative biases for the Nile of MPI-ESMa compared with ECHAM5a. It seems that all models produce a too enhanced evapotranspiration over the Amazon, which becomes obvious by the positive biases for the AMIP2 simulations where the precipitation shows no bias (Figure 15b), but also for the historical simulations that show no evapotranspiration bias even though precipitation is largely underestimated. Similarly over the Danube, evapotranspiration is overestimated by MPI-ESMh after the removal of the dry precipitation bias, while the dry precipitation bias in ECHAM5h also leads to a negative evapotranspiration bias. For the Ganges/Brahmaputra, evapotranspiration generally seems to be underestimated by the models as even in the AMIP2 simulations the surplus of water due to the wet precipitation bias only leads to a removal of the negative bias shown for the historical simulations.

[44] Positive or negative biases in precipitation often exceed the biased amounts of evapotranspiration so that also wet (Arctic rivers, Amur, Baltic Sea, Yangtze Kiang) and dry (Congo for ECHAM5h and MPI-ESMa, Mississippi for AMIP2 simulations) biases in runoff (Figure 16b) occur, respectively. Differences between the models in precipitation and evapotranspiration mostly seem to be of the same order so that they often compensate each other, which causes model differences in the runoff and the associated bias to be rather similar with a few exceptions. The too enhanced evapotranspiration over the Amazon leads to a dry runoff bias for all models, which is more severe for both historical simulations due to their too low precipitation. Overestimated evapotranspiration also leads to dry runoff biases over the

Danube catchment for all models except ECHAM5h where the dry bias in precipitation dominates the dry runoff bias. Underestimated evapotranspiration causes wet runoff biases over the Ganges/Brahmaputra and Murray catchments for all models. For the first, these are largely enhanced by the wet precipitation biases for the two AMIP2 simulations. As for precipitation (see above), the wet runoff bias in MPI-ESMa is only about half the bias of ECHAM5a. For the Murray, wet runoff bias of ECHAM5a is less severe due to the larger dry bias in precipitation. For the Nile, negative biases in precipitation and evapotranspiration compensate each other except for ECHAM5a where the overestimation of precipitation causes a large overestimation of runoff. Note that for the Nile climatological discharge, observations are used before the Aswan dam was built as this is not considered in the models either. For the Parana, the reduction in the dry precipitation bias of MPI-ESMh compared with ECHAM5h leads also to an analogue removal of the dry runoff bias. For both AMIP2 simulations, large positive runoff biases occur that are related to a wet precipitation bias for ECHAM5a, while for MPI-ESMa it seems to be caused by an adding up of smaller positive and negative biases in precipitation and evapotranspiration, respectively.

## 5. Annual Cycles and Differences Between Model Families

[45] In this section, we are considering noticeable differences of the different model families (historical/AMIP2 and MPI-ESM/ECHAM5) in more detail, thereby considering the mean annual cycles of precipitation, 2m temperature, surface albedo, and SSI over the selected major catchments (Figure 14). Here, we also investigated why some of these differences occur and how they relate to the model setup.

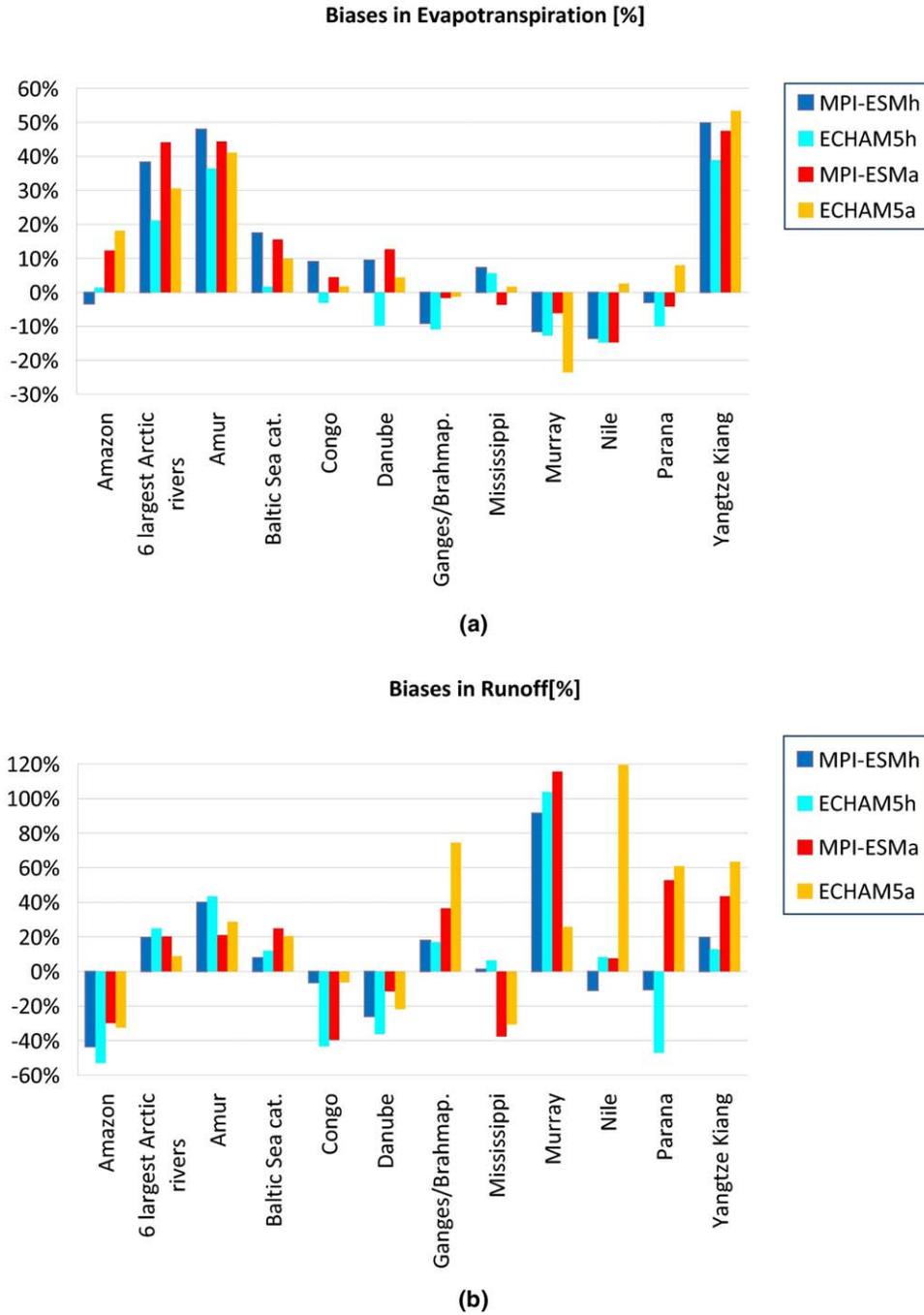


**Figure 15.** Annual mean biases in simulated (a) 2m temperature and (b) precipitation over several catchments. The bias were calculated from the differences of the simulations to the WFD data.

**5.1. Historical Versus AMIP2 Simulations**

[46] The fully coupled historical and the uncoupled AMIP2 SST driven simulations do not show any significance differences for surface albedo. With regard to SSI (Figure 11), the SST driven simulation (MPI-ESMa) generally shows a relatively better performance than the coupled model simulation (MPI-ESMh), except for SON, indicating a general increase of SSI uncertainties due to the coupling. However, the difference between the

MPI-ESM experiments is in general much smaller than the difference to ECHAM5h. For 2m temperature (Figure 15a), it can be noted that the historical simulations are consistently colder over Eurasian high and mid-latitude catchments (Arctic rivers, Amur, Baltic Sea, Danube) than the corresponding AMIP2 simulations. The latter is also the case for the Mississippi catchment where both historical simulations have a reduced warm bias. The most apparent differences between the historical



**Figure 16.** Annual mean biases in simulated (a) evapotranspiration and (b) runoff over several catchments. The observed evaporation was calculated from the difference of WFD precipitation and observed climatological discharge [Dümenil Gates *et al.*, 2000]. The runoff bias was calculated from the difference of the simulated runoff and the observed climatological discharge.

and the SST driven simulations occur for the land surface water fluxes, especially for precipitation (Figure 15b). For the Amazon, Ganges/Brahmaputra, and the Mississippi, both historical simulations behave strongly differently than the AMIP2 simulations. For the Amazon, a dry precipitation bias of about 20% occurs that is not present in the AMIP2 simulations, while for the Ganges/Brahmaputra, the wet precipitation bias of the AMIP2

simulations is reduced almost to zero. Consequently, we consider these three catchments in more detail in the following.

[47] For the Amazon, the dry biases in the hydrological cycle are partially caused by the too enhanced evapotranspiration (see section 4), which points to deficits in the representation of associated land surface processes. These deficits seem to affect the simulated precipitation of

**Table 3.** Observed Values for WFD 2m Temperature (1971–2000; °C), WFD Precipitation (1971–2000), Evaporation (WFD Precipitation Minus Climatological Discharge), and Runoff (Climatological)<sup>a</sup>

Catchment	Temperature	Precipitation	Evaporation	Runoff
Amazon	24.6	2242	1189	1053
Amur	−1.3	540	367	173
Six largest Arctic Rivers	−5.0	455	264	191
Baltic Sea catchment	4.4	692	413	279
Congo	24.1	1576	1211	365
Danube	9.1	797	545	251
Ganges/Brahmaputra	17.7	1397	725	672
Mississippi	10.2	893	697	196
Murray	17.8	510	502	8
Nile	25.5	646	597	49
Parana	21.5	1308	1085	223
Yangtze Kiang	12.0	1069	532	537

<sup>a</sup>Unit: mm a<sup>−1</sup>.

all models during the boreal summer, where precipitation is significantly lower than the WFD (Figure 17). But the large underestimation of precipitation in the historical simulations throughout the first nine months of the year is not caused by these deficits as no dry bias occurs in the SST driven AMIP2 simulations during winter and spring. During most parts of this period, the moisture available for precipitation is mainly transported from the Tropical Atlantic north of the equator, i.e., passing over the northeast (NE) coast of South America [see e.g., *Trenberth, 1998*]. Andersson et al. (in preparation, 2013) showed that both coupled simulations have annual mean cold SST bias at the NE coast of South America that is not present in MPI-ESMa. This bias is associated with lower evaporation rates and a subsequent dry bias in the integrated water vapor. The latter is spatially enhanced by a large-scale low bias in 10 m wind speed in the northern Tropical Atlantic. These biases lead to a reduced moisture transport into the Amazon catchment that is causing the dry precipitation bias in the coupled simulations. These biases along the NE coast of South America are not present in MPI-ESMa, thereby leading to a more realistic moisture transport and associated precipitation.

[48] For the Ganges/Brahmaputra, the improved simulation of precipitation induced by the coupling occurs during the South Asian summer monsoon season, where AMIP2 simulations, especially ECHAM5a, overestimate the precipitation compared with WFD (Figure 17). We speculate that in the AMIP2 simulations, the ocean act as a too enhanced heat and moisture source as there is no counteracting feedback by ocean SST. This is leading to a too strong moisture transport from the Arabian Sea to northern India. As the models do not represent the large irrigation ongoing over northern India and Pakistan, there is too little moisture supply from the associated areas [see, e.g., *Saeed et al., 2009*], which inhibits the formation of precipitation indicated by the dry bias over the northern Indian plains (Figure 4). Therefore, the moisture is transported further inland toward the Himalayas, where it is causing a surplus of precipitation over the Ganges/Brahmaputra catchment and a subsequent too enhanced hydrological cycle over the region. In the coupled simulations, the large moisture fluxes over the ocean are reduced by the response of the ocean SST so

that the associated precipitation over the Ganges/Brahmaputra catchment agrees well with the WFD. This is supported by Andersson et al. (in preparation, 2013), whose results indicate a reasonable simulation of SST over the Arabian Sea in the coupled simulations compared with Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data (HOAPS) [*Andersson et al., 2010*] being slightly colder than the AMIP SST, while the wind speed in the AMIP simulation exhibits a positive bias for this region. The latter and the missing coupling leads to strongly overestimated evaporation fluxes in MPI-ESMa over the Arabian Sea that are not present in the coupled simulations.

[49] Noticeable differences between the coupled and the AMIP2 simulations can be also seen in summer and autumn over the Mississippi catchment, where a warm bias of up 2–3°C in the AMIP2 simulations is almost completely eliminated in both coupled simulations (Figure 18). Figure 19 indicates that the coupling leads to a lower simulation of SSI, which is closer to CERES data and, hence, causes the reduction of the warm bias. Between April and September, the coupling also leads to an increased precipitation (Figure 17). While this causes a wet bias until July, the prominent dry bias of the AMIP2 simulations from August to October is reduced. These results suggest that the coupling causes more cloud cover and enhanced precipitation during the summer half year, thereby overcompensating the dry bias of the AMIP2 simulations in the annual mean (Figure 15b). This enhanced precipitation subsequently leads to a removal of the negative bias in runoff (Figure 16b). During the summer, the moisture is mainly transported from the northern subtropical Atlantic via Gulf of Mexico into the catchment [see, e.g., *Trenberth, 1998*]. Here, both coupled models simulated a warm SST bias that is inducing larger evaporation fluxes than in MPI-ESMa (Andersson et al., in preparation, 2013). These fluxes lead to a wet bias in the integrated water vapor, and, very likely to the enhanced moisture flux into the Mississippi catchment that is compensating the dry precipitation bias which is present in the AMIP simulations. Part of the moisture flux into the Mississippi catchment during the hurricane season (June–November) is related to tropical cyclone activity that may contribute up to

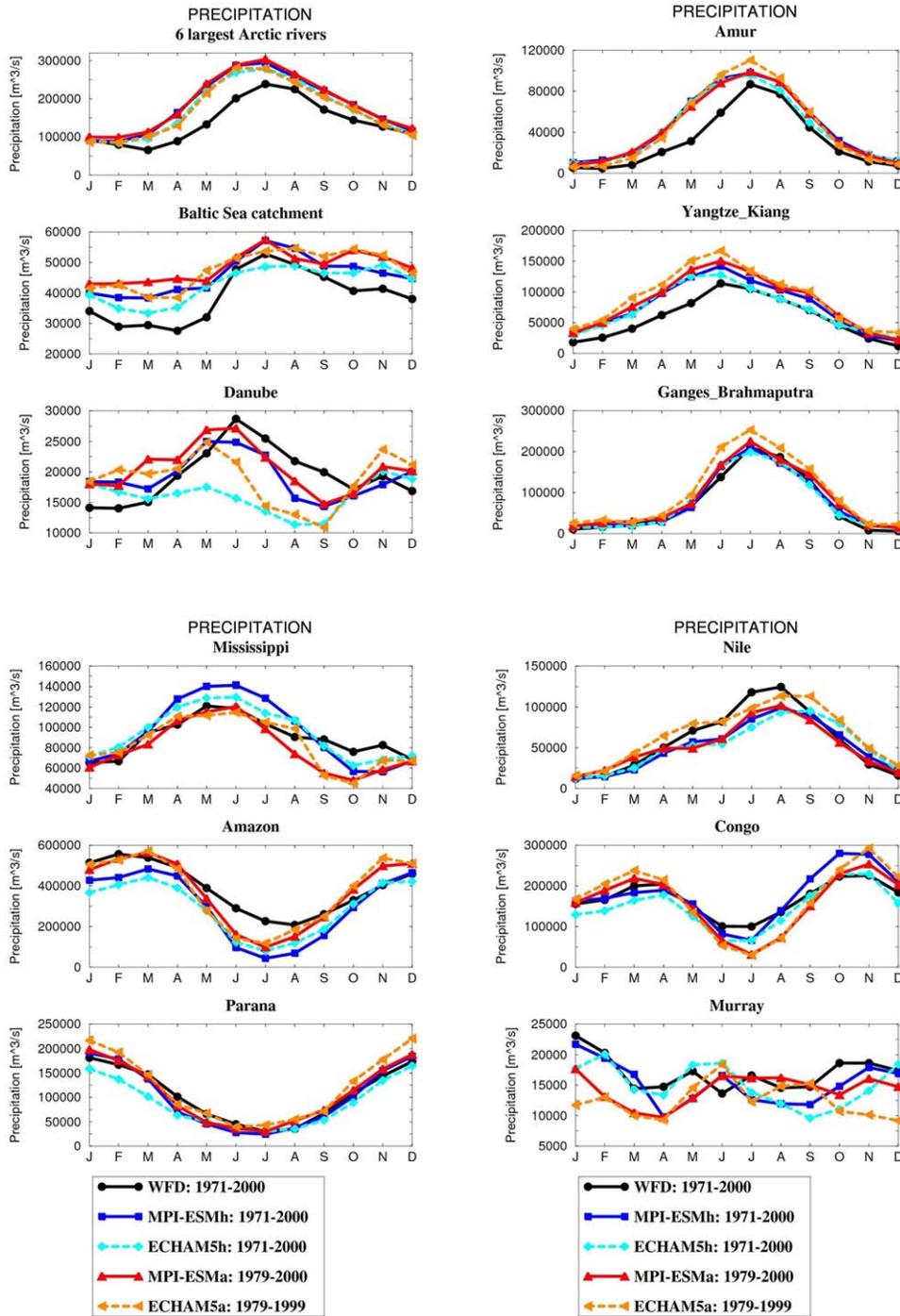


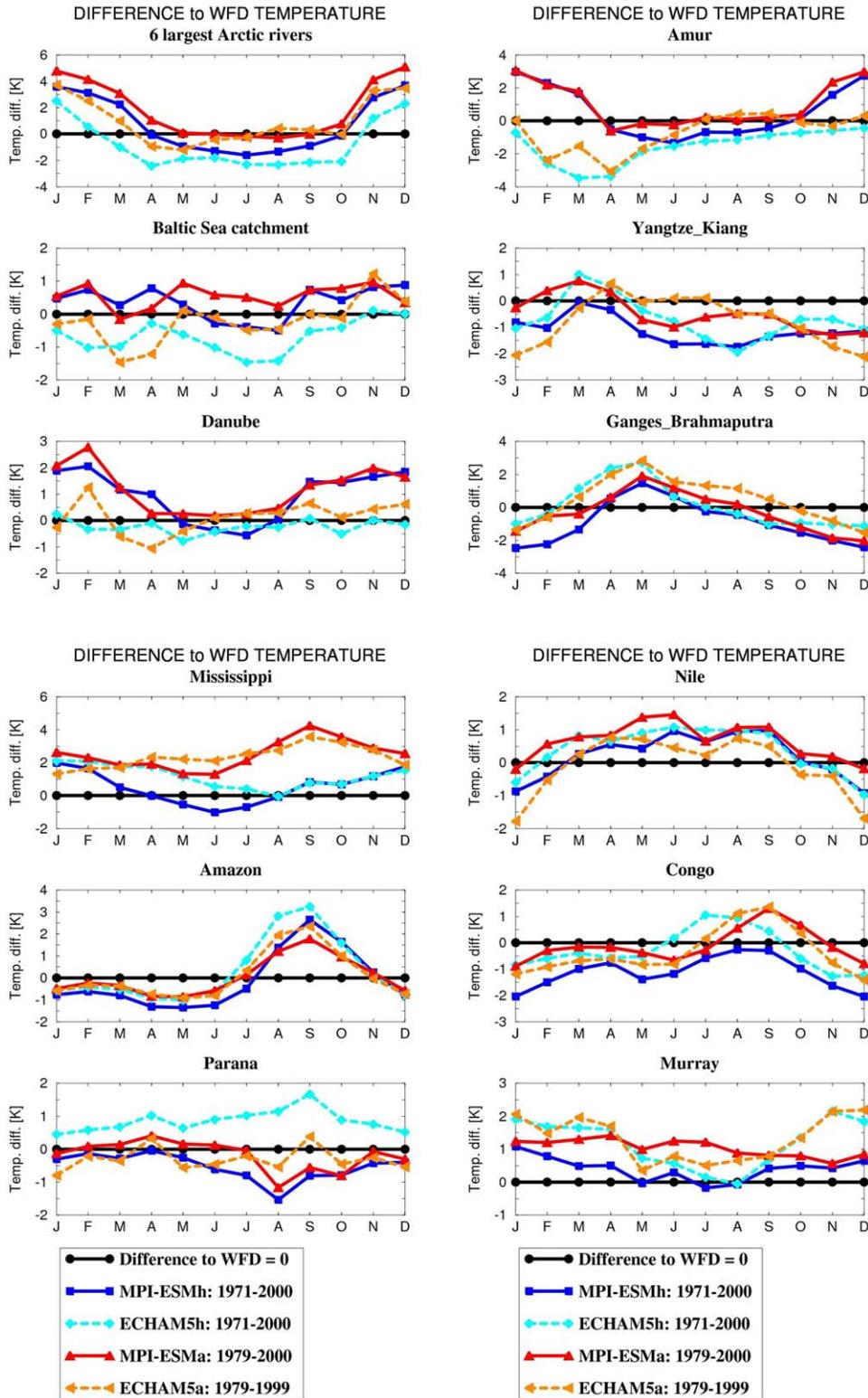
Figure 17. Simulated and observed precipitation ( $\text{m}^3 \text{s}^{-1}$ ) over large catchments.

10%–15% of the seasonal rainfall in the southern parts of the catchment [Larson et al., 2005]. We speculate that the summer/autumn dry bias is partially caused by the fact that tropical cyclones cannot be adequately represented at the used model resolution of about 200 km grid size.

### 5.2. MPI-ESM Versus ECHAM5

[50] As pointed out in section 3.2, larger differences in the simulated 2m temperature occur between MPI-ESM and ECHAM5, especially over northern Eurasia that are

most pronounced in the boreal winter half year. Here, the MPI-ESM simulations are considerably warmer than the ECHAM5 simulations and the WFD data, such as shown for the Arctic Rivers, Danube, Amur, and Baltic Sea in Figure 18. This behavior seems to be mainly imposed by a lower simulation of surface albedo by MPI-ESM that is closer to MODIS data than those simulated by ECHAM5 (Figure 20). In summer time, cold biases of ECHAM5h over the Arctic and Baltic Sea catchments seem to be caused by a large negative SSI bias (Figure 19)



**Figure 18.** Differences to WFD 2m temperature (K) over large catchments. Differences were calculated over the respective time period given for the ESM simulations.

that is reduced but still prominent in the MPI-ESM simulations. A corresponding behavior of precipitation biases to biases in SSI or 2m temperature is not obvious (cf. Figure 17).

[51] Over the Danube catchment, both ECHAM5 simulations suffer from the so-called summer drying problem that was already highlighted for ECHAM5h by Hagemann et al. [2009]. The problem still exists for MPI-

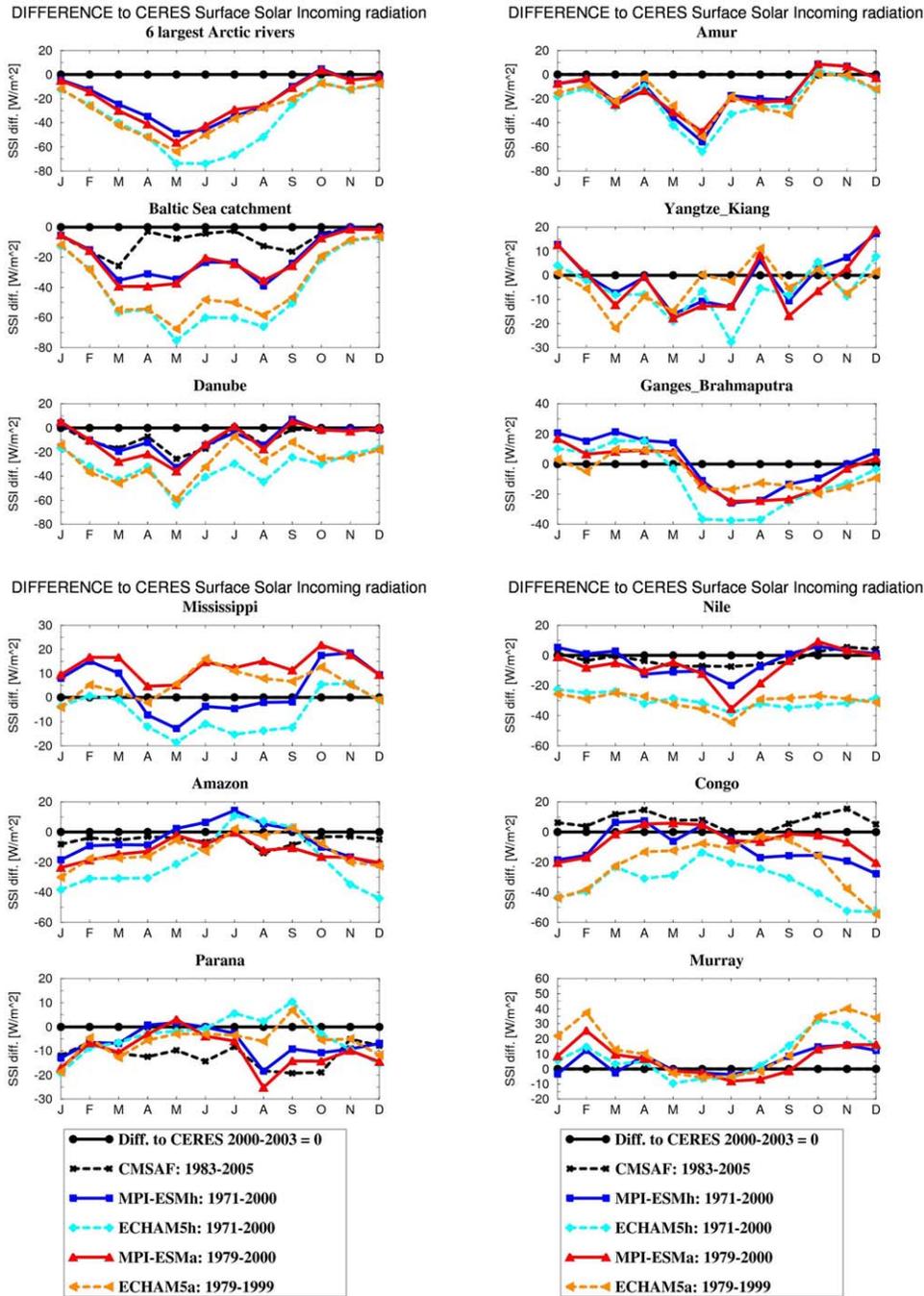


Figure 19. Differences to CERES SSI ( $\text{W m}^{-2}$ ) over large catchments.

ESM, but it is largely reduced compared with ECHAM5. The representation of land surface hydrology is very similar in both model versions, as is the surface albedo during summer time (Figure 20), which is close to MODIS data. This indicates that the treatment of land surface processes is not responsible for the reduction of the summer drying problem in MPI-ESM. This is consistent with results from a regional climate modeling study of Hagemann *et al.* [2004], who pointed out that for two regional climate models using ECHAM4 physics (HIRHAM, REMO), systematic errors in the atmospheric dynamics appear to be causing the summer drying prob-

lem over the Danube catchment. These are likely also leading to the positive evapotranspiration bias in the MPI-ESM simulations over this region (Figure 16a, see also section 4). The strong improvement in the simulation of SSI (Figure 19) suggests that this reduction is mainly caused by changes in the atmospheric component of MPI-ESM, ECHAM6 [Stevens *et al.*, 2012].

[52] In section 3.3, the better performance of MPI-ESM in simulating SSI was already pointed out. This cannot only be seen for the Danube, but also over many other catchments (Arctic rivers, Baltic Sea, Amazon, Nile, Congo, and Murray) where the simulated SSI

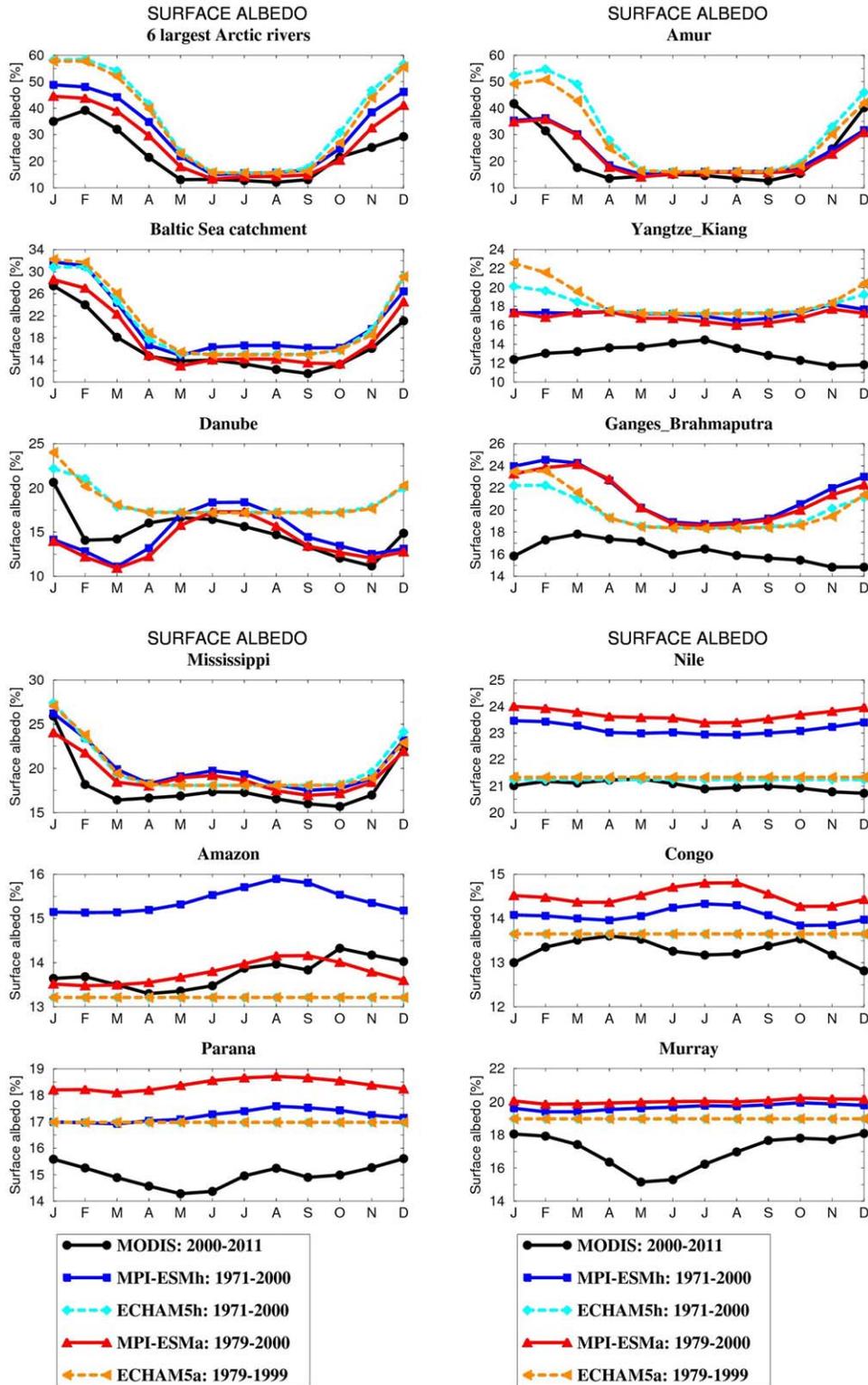


Figure 20. Simulated and observed surface albedo (%) over large catchments.

(Figure 19) of MPI-ESM is closer to the satellite observations of CERES and CMSAF than the ECHAM5 SSI. For the Parana, ECHAM5 SSI is closer to CERES data than MPI-ESM SSI (CMSAF data does not fully cover this catchment). However, a systematic impact on

the simulated temperatures (Figure 18) is only visible for the Parana in the southern winter and the Murray in the southern summer half year. For the latter, the overestimated SSI in the ECHAM5 simulations leads to a related warm bias which is largely reduced in

MPI-ESM where also the simulated SSI is closer to the CERES data. For the Parana, the deviations of simulated SSI to CERES data look rather similar to the temperatures biases compared with WFD during July to October.

[53] Despite the fact that MPI-ESMh was run with a dynamic vegetation scheme while MPI-ESMa uses a prescribed PFT distribution, the surface albedo over snow free areas does not differ significantly between both simulations (cf. section 3.4). Some noticeable differences occur over the Amazon (Parana) catchment, where MPI-ESMa surface albedo is systematically lower (higher) by about 2% (1%) than in MPI-ESMh (Figure 20). For the Amazon, MPI-ESMa agrees quite well with MODIS data while for the Parana MPI-ESMh has lower positive bias than MPI-ESMa compared with MODIS. This implies that the dynamic vegetation scheme simulates a too low tree cover over the Amazon catchment as a direct consequence of the dry bias in MPI-ESMh. Over the Parana instead, the dynamic vegetation scheme seems to reduce a bias in the PFT distribution that is present in MPI-ESMa. A direct effect on the simulated temperature cannot be concluded from Figure 18, even though it seems that over the Amazon, the overestimated surface albedo of MPI-ESMh might be associated with a slightly increased cold bias compared with MPI-ESMa in the first half of the year. In the second half of the year, this effect is exceeded by the warm bias that is likely related to the dry bias in the boreal summer (Figure 17).

## 6. Summary and Concluding Remarks

[54] In the present study, we have jointly evaluated land surface water and energy fluxes from the very recent simulations that have been conducted with the MPI-ESM for CMIP5 exercise. These simulations comprise three-member ensembles of AMIP2 SST forced simulations of the land/atmosphere component of MPI-ESM and of the fully coupled ESM. MPI-ESM model output was compared with various observational data sets as well as to simulations by its predecessor ECHAM5. Apart from a general evaluation of the fluxes, we focused on differences between the two ESM versions as well as on differences between the fully coupled simulation and the SST-driven simulations.

[55] The study has proven that the simulated surface shortwave radiation fluxes and land surface albedo have considerably improved in MPI-ESM compared with its predecessor ECHAM5. This has led to subsequent differences in simulated 2m temperature between MPI-ESM and ECHAM5. To a large extent, these are caused by the improved simulation of SSI in MPI-ESM. Over the high northern latitudes in the winter, these differences mainly originate from an improved simulation of surface albedo related to the associated snow cover. But compared with WFD data, the MPI-ESM simulated 2m temperature did not necessarily improve in the same way as SSI and surface albedo. The latter is the case for the reduction of the boreal summer cold bias in the high northern latitudes that can be attributed to the improved SSI in the MPI-ESM simulations. During the

boreal winter, the cold bias over Europe is removed by MPI-ESM, but the northern Asian warm bias is largely extended in the MPI-ESM simulations, which is mainly limited over Eastern Siberia in ECHAM5h.

[56] For the hydrological cycle, large-scale bias patterns are rather similar between the different models over many regions. For precipitation, common deviations from the WFD denote a pronounced dry bias in the Tropics north of the equator, which also extends to south of the equator during the boreal spring and summer, a dry bias in the low precipitation region in the southern Subtropics accompanied by a wet bias around 50°S, a wet bias in the northern high latitudes during boreal spring and summer and an overestimation of precipitation along steep mountain slopes. The most notable difference between the models is that MPI-ESMh and MPI-ESMa generally have an improved simulation of peak rainfall in the Tropics compared with ECHAM5h. Also, the summer drying problem over southern and eastern Europe (especially over the Danube catchment) is largely reduced in the MPI-ESM simulations.

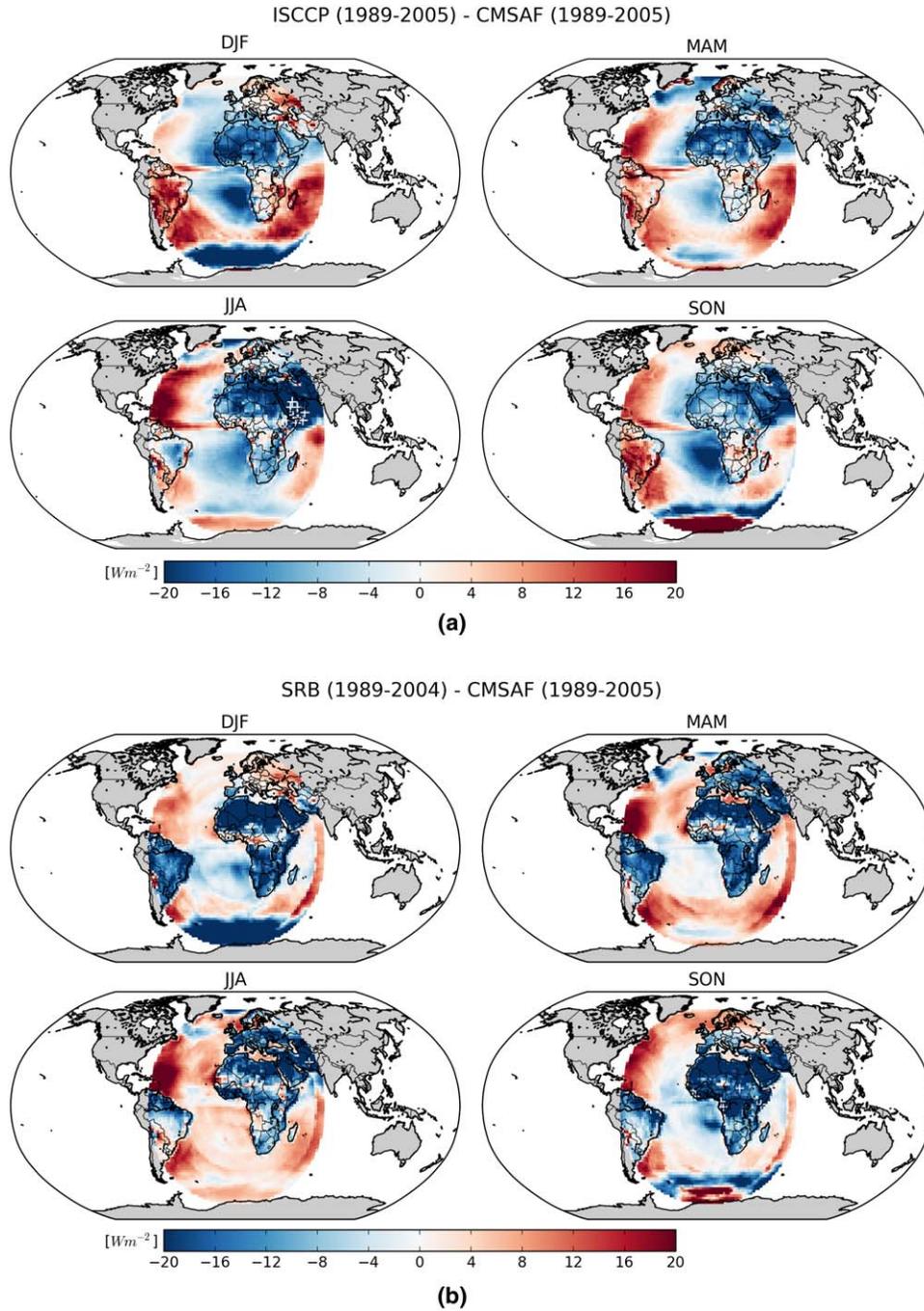
[57] For many land areas, the coupling to an ocean model does not lead to significant differences in land surface water and energy fluxes compared to the simulations forced with observed SST. But three areas can be highlighted where the coupling causes noticeable effects. On one hand, the coupling induces a dry bias over the Amazon catchment, while on the other hand it leads to an improved precipitation over the Ganges/Brahmaputra and Mississippi catchments. For the Amazon, this deficit of the coupled simulations cannot be attributed to land surface processes, but instead it is primarily induced by biases in simulated SST patterns and associated moisture transport. Here, it can be noted that an insufficient representation of land surface processes is probably only causing the dry bias during the boreal summer that is persistent in all models. For the Mississippi, warm biases in SST of the coupled simulations and associated wet biases in the associated moisture transport lead to increased summer precipitation that is compensating a dry bias caused by other model deficits that are likely associated with the coarse spatial resolution of the ESMs. For the Ganges/Brahmaputra, the missing interaction of the ocean with the atmosphere over the Arabian Sea has been identified as the main cause of the too enhanced precipitation in the SST-forced simulations. This conclusion is supported by results of *Wang et al.* [2005], who found that state-of-the-art atmospheric GCMs, when forced by observed SST, are unable to simulate properly Asian-Pacific summer monsoon rainfall.

[58] In summary, the combined evaluation of land surface water and energy fluxes has shown that MPI-ESM has generally improved compared with ECHAM5, especially with regard to SSI and surface albedo. Bias pattern for precipitation and 2m temperature are similar for both ESM versions, and improvements slightly outweigh worsenings. As ECHAM5h was already one of the best performing models in the CMIP3 exercise [*Reichler and Kim*, 2008], it can be concluded that MPI-ESM is well suited for climate change studies focusing on the water and energy cycle at the land surface.

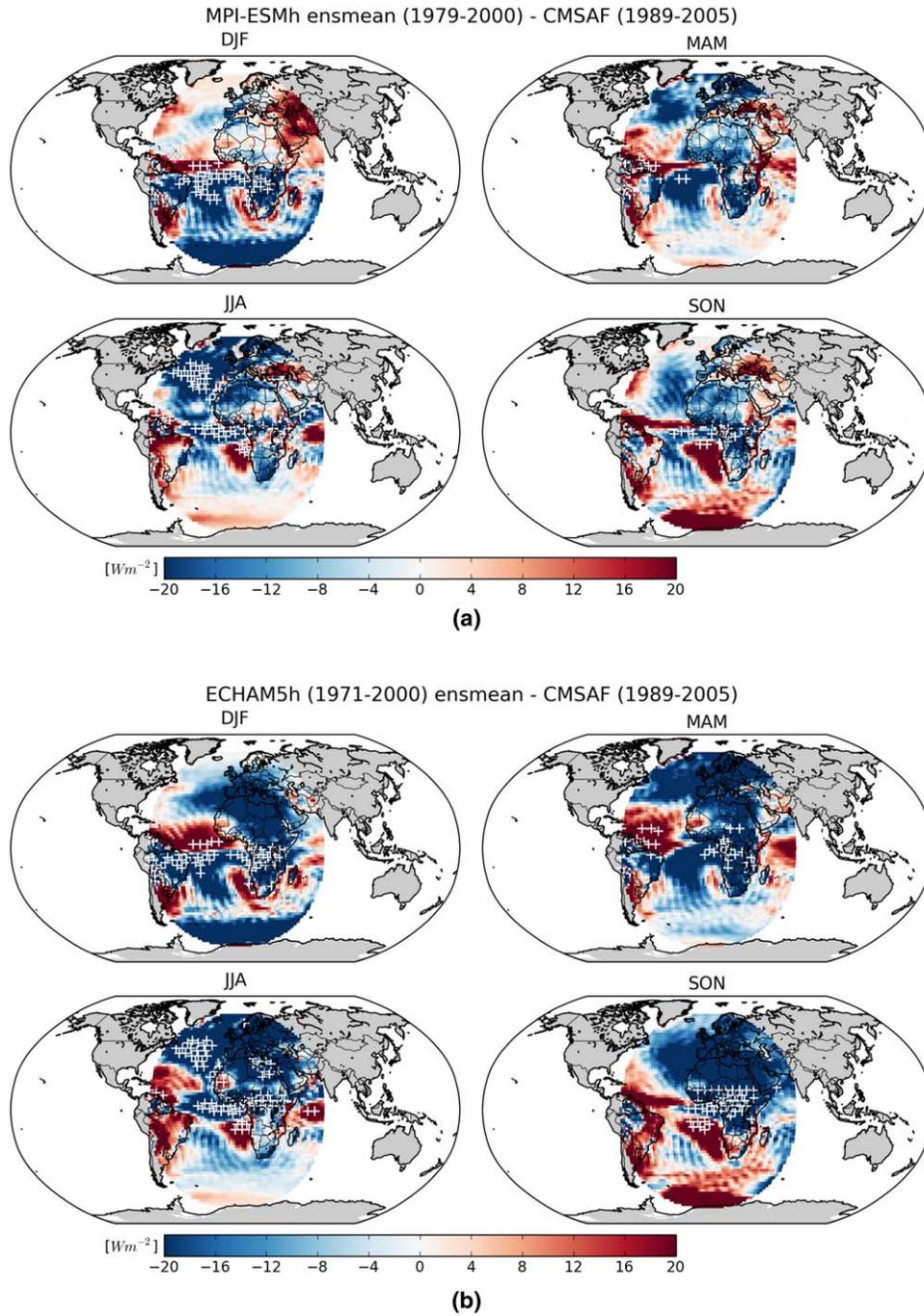
Appendix A

[59] In order to provide complementary information for the usage of CERES as reference SSI in the spatial maps of the manuscript, we added comparisons to CMSAF SSI in this appendix. CMSAF data are available for a longer period (1989–2005) than CERES data, but they cover only a limited part of the globe. Figure

A1 compares the seasonal SSI cycles of ISCCP and SRB data to CMSAF data, and is, thus, complementary to Figure 1. Figure A2 shows an analogue comparison of MPI-ESMh and ECHAM5h to CMSAF data that is complementary to Figure 10.



**Figure A1.** Differences of seasonal cycle of SSI between different observational datasets: ISCCP compared to CMSAF (top), SRB compared to CMSAF (bottom). Statistically significant differences ( $p > 0.05$ ) between the satellite products are indicated by stippled areas.



**Figure A2.** SSI differences of MPI-ESMh (top) and ECHAM5h (bottom) to CMSAF.

[60] **Acknowledgments.** This work was partly supported by funding from the European Union within the EMBRACE project (grant 282672) and through the Cluster of Excellence “CliSAP” (EXC177), KlimaCampus, University of Hamburg, funded by the German Science Foundation (DFG). The GCM data were obtained from the CERA database at the German Climate Computing Center (DKRZ) in Hamburg. CERES data were obtained from the NASA atmospheric science data center. The Meteosat surface radiation data were provided by the EUMETSAT CMSAF, which is gratefully acknowledged. BSRN surface radiation data were obtained from the WRMC-BSRN hosted at AWI, Bremerhaven.

## References

- Adam, J. C., and D. P. Lettenmaier (2003), Adjustment of global gridded precipitation for systematic bias, *J. Geophys. Res.*, *108*(D9), 4257, doi:10.1029/2002JD002499.
- Andersson, A., K. Fennig, C. Klepp, S. Bakan, H. Graßl, and J. Schulz (2010), The Hamburg Ocean atmosphere parameters and fluxes from satellite data—HOAPS-3, *Earth Syst. Sci. Data*, *2*, 215–234, doi:10.5194/essd-2-215-2010.
- Biemans, H., R. W. A. Hutjes, P. Kabat, B. Strengers, D. Gerten, and S. Rost (2009), Effects of precipitation uncertainty on discharge

- calculations for main river basins, *J. Hydrometeorol.*, *10*, 1011–1025, doi:10.1175/2008JHM1067.1.
- Brovkin, V., T. Raddatz, C. H. Reick, M. Claussen, and V. Gayler (2009), Global biogeophysical interactions between forest and climate, *Geophys. Res. Lett.*, *36*, L07405, doi:10.1029/2009GL037543.
- Brovkin, V., L. Boysen, T. Raddatz, V. Gayler, A. Loew, and M. Claussen (2012), Evaluation of vegetation cover and land-surface albedo in MPI-ESM CMIP5 simulations, *J. Adv. Model. Earth Syst.*, doi:10.1029/2012MS000169.
- Cox, S. J., P. W. Stackhouse Jr., S. K. Gupta, J. C. Mikovitz, T. Zhang, L. M. Hinkelman, M. Wild, and A. Ohmura (2006), The NASA/GEWEX surface radiation budget project: Overview and analysis, Preprints, in *Proceedings of the 12th Conference on Atmospheric Radiation*, Madison, Wis., *Amer. Meteor. Soc.*, 10.1 [Available at <http://ams.confex.com/ams/pdfpapers/112990.pdf>].
- Dümenil Gates, L., S. Hagemann, and C. Golz (2000), Observed historical discharge data from major rivers for climate model validation, *Max Planck Institute for Meteor. Rep.*, *307*, MPI for Meteorology, Hamburg, Germany.
- Essery, R., N. Rutter, J. Pomeroy, R. Baxter, M. Stahli, D. Gustafsson, A. Barr, P. Bartlett, and K. Elder (2009), An evaluation of forest snow process simulations, *Bull. Am. Meteorol. Soc.*, *90*, 1120–1135, doi:10.1175/2009bams2629.1.
- Fensholt, R., et al. (2012), Greenness in semi-arid areas across the globe 1981–2007—An Earth Observing Satellite based analysis of trends and drivers, *Remote Sens. Environ.*, *121*, 144–158, doi:10.1016/j.rse.2012.01.017.
- Fuchs, T., U. Schneider, and B. Rudolf (2007), Global precipitation analysis products of the GPCC, Global Precipitation Climatology Centre (GPCC), Deutscher Wetterdienst, Offenbach, Germany.
- Hagemann, S. (2002), An improved land surface parameter dataset for global and regional climate models, *Max Planck Institute for Meteor. Rep.*, *336*, Hamburg, Germany [Available at <http://www.mpimet.mpg.de/en/wissenschaft/publikationen.html>].
- Hagemann, S., B. Machenhauer, R. Jones, O. B. Christensen, M. Déqué, D. Jacob, and P. L. Vidale (2004), Evaluation of water and energy budgets in regional climate models applied over Europe, *Clim. Dyn.*, *23*, 547–567.
- Hagemann, S., K. Arpe, and E. Roeckner (2006), Evaluation of the hydrological cycle in the ECHAM5 model, *J. Clim.*, *19*, 3810–3827.
- Hagemann, S., H. Göttel, D. Jacob, P. Lorenz, and E. Roeckner (2009), Improved regional scale processes reflected in projected hydrological changes over large European catchments, *Clim. Dyn.*, *32*, 767–781.
- Ineichen, P., C. Barroso, B. Geiger, R. Hollmann, A. Marsouin, and R. Müller (2009), Satellite application facilities irradiance products: Hourly time step comparison and validation over Europe, *Int. J. Remote Sens.*, *30*(10), 5549–5571.
- Jiménez, C., et al. (2011), Global inter-comparison of 12 land surface heat flux estimates, *J. Geophys. Res.*, *116*, D02102, doi:10.1029/2010JD014545.
- Jungclaus, J. H., M. Botzet, H. Haak, N. Keenlyside, J.-J. Luo, M. Latif, J. Marotzke, U. Mikolajewicz, and E. Roeckner (2006), Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM, *J. Clim.*, *19*, 3952–3972.
- Jungclaus, J. H., N. Fischer, H. Haak, K. Lohmann, J. Marotzke, D. Matei, U. Mikolajewicz, D. Notz, and J. S. von Storch (2012), Characteristics of the ocean simulations in MPIOM, the ocean component of the MPI-Earth System Model, *J. Adv. Model. Earth Syst.*, doi:10.1002/jame.20023.
- Koster, R. D., et al. (2004), Regions of strong coupling between soil moisture and precipitation, *Science*, *305*(5687), 1138–1140.
- Larson, J., Y. Zhou, and R. W. Higgins (2005), Characteristics of land-falling tropical cyclones in the United States and Mexico: Climatology and interannual variability, *J. Clim.*, *18*, 1247–1262.
- Loeb, N. G., S. Kato, W. Su, T. Wong, F. G. Rose, D. R. Doelling, and J. Norris (2012), Advances in understanding top-of-atmosphere radiation variability from satellite observations, *Surv. Geophys.*, *33*, 359–385, doi:10.1007/s10712-012-9175-1.
- Loew, A. and Y. Govaerts (2010), Towards multidecadal consistent Meteosat surface albedo time series, *Remote Sens.* *2*(4), 957–967.
- Loveland, T. R., B. C. Reed, J. F. Brown, D. O. Ohlen, J. Zhu, L. Yang, and J. W. Merchant (2000), Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR data, *Int. J. Remote Sens.* *21*, 1303–1330.
- Mitchell, T. D., and P. D. Jones (2005), An improved method of constructing a database of monthly climate observations and associated high-resolution grids, *Int. J. Climatol.*, *25*, 693–712.
- Mueller, B., et al. (2011), Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, *Geophys. Res. Lett.*, *38*, L06402, doi:10.1029/2010GL046230.
- Müller, R. W., C. Matsoukas, A. Gratzki, H. D. Behr, and R. Hollmann (2009), The CMSAF operational scheme for the satellite based retrieval of solar surface irradiance—A LUT based eigenvector hybrid approach, *Remote Sens. Environ.*, *113*(5), 1012–1024.
- Müller, R., J. Trentmann, R. Stöckli, and R. Posselt (2011), Algorithm theoretical baseline document: Meteosat (MVIRI) solar irradiance and effective cloud albedo climate datasets—The MAGIC SOL method, *Tech. Rep.*, EUMETSAT Satellite Application Facility on Climate Monitoring, SAF/CM/DWD/ATBD/MVIRI HEL v1.1.
- Ohmura, A., et al. (1998), Baseline surface radiation network (BSRN/WCRP): New precision radiometry for climate research, *Bull. Am. Meteorol. Soc.*, *79*(10), 2115–2136.
- Program for Climate Model Diagnosis and Intercomparison (PCMDI) (2007), IPCC model output [Available at [www.pcmdi.llnl.gov/ipcc/about\\_ipcc.php](http://www.pcmdi.llnl.gov/ipcc/about_ipcc.php)].
- Posselt, R., R. Müller, R. Stöckli, and J. Trentmann (2011a), Spatial and temporal homogeneity of solar surface irradiance across satellite generations, *Remote Sens.*, *3*, 1029–1046.
- Posselt, R., R. Müller, J. Trentmann, and R. Stöckli (2011b), CMSAF validation report: Meteosat (MVIRI) climate data sets of SIS, SID and CAL, *Tech. Rep.*, Satellite Application Facility on Climate Monitoring, SAF/CM/DWD/VAL/MVIRI HEL, v.1.1.
- Raddatz, T. J., C. Reick, W. Knorr, J. Kattge, E. Roeckner, R. Schnur, K.-G. Schnitzler, P. Wetzell, and J. H. Jungclaus (2007), Will the tropical land biosphere dominate the climate-carbon cycle feedback during the twenty-first century?, *Clim. Dyn.*, *29*, 565–574, doi:10.1007/s00382-007-0247-8.
- Rechid D., T. J. Raddatz, and D. Jacob (2009), Parameterization of snow-free land surface albedo as a function of vegetation phenology based on MODIS data and applied in climate modelling, *Theor. Appl. Climatol.*, *95*, 245–255, doi:10.1007/s00704-008-0003-y.
- Reichler, T., and J. Kim (2008), How well do coupled models simulate today's climate? *Bull. Am. Meteorol. Soc.*, *89*(3), 303–311, doi:10.1175/BAMS-89-3-303.
- Roeckner, E., et al. (2003), The atmospheric general circulation model ECHAM5. Part I: Model description, *Max Planck Institute for Meteor. Rep.*, *349*, 127 pp., MPI for Meteorology, Hamburg, Germany.
- Roesch, A., and E. Roeckner (2006), Assessment of snow cover and surface albedo in the ECHAM5 general circulation model, *J. Clim.*, *19*, 3828–3843.
- Rossow, W. B., and Y.-C. Zhang (1995), Calculation of surface and top-of-atmosphere radiative fluxes from physical quantities based on ISCCP datasets, Part II: Validation and first results, *J. Geophys. Res.*, *100*, 1167–1197.
- Rudolf, B., and F. Rubel (2005), Global precipitation, in *Observed Global Climate, Landolt-Boernstein: Numerical Data And Functional Relationships in Science and Technology—New Series, Group 5: Geophysics*, vol. 6, edited by M. Hantel, chap. 11, p. 567, Springer, Berlin.
- Saeed, F., S. Hagemann, and D. Jacob (2009), Impact of irrigation on the South Asian Summer Monsoon, *Geophys. Res. Lett.*, *36*, L20711, doi:10.1029/2009GL040625.
- Schaaf, C. B., et al. (2002), First operational BRDF, albedo and nadir reflectance products from MODIS, *Remote Sens. Environ.*, *83*, 135–148.
- Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling (2010), Investigating soil moisture-climate interactions in a changing climate: A review, *Earth Sci. Rev.*, *99*, 125–161, doi:10.1016/j.earscirev.2010.02.004.
- Solomon, S., D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. B. Tignor, H. L. Miller Jr., and Z. Chen, Eds. (2007), *Climate Change 2007: The Physical Science Basis*, Cambridge Univ. Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp.
- Sorooshian, S., J. Roads, J. Polcher, R. Schiffer, R. Lawford, P. Try, W. Rossow, and G. Sommeria (2005), Water and energy cycles: Investigating the links, *WMO Bull.*, *54*(2), 58–64.

- Stevens, B., et al. (2012), The atmospheric component of the MPI-M Earth system model: ECHAM6, *J. Adv. Model. Earth Syst.*, doi:10.1002/jame.20015.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bull. Am. Meteor. Soc.*, 93(4), 485–498.
- Trenberth, K. E. (1998), Atmospheric moisture recycling: Role of advection and local evaporation, *J. Clim.*, 12, 1368–1381.
- Trenberth, K. E., J. T. Fasullo, and J. Kiehl (2009), Earth's global energy budget, *Bull. Am. Meteorol. Soc.*, 90, 311–323.
- Trenberth, K. E., J. T. Fasullo, and J. Mackaro (2011), Atmospheric moisture transports from ocean to land and global energy flows in reanalyses, *J. Clim.*, 24, 4907–4924, doi:http://dx.doi.org/10.1175/2011JCLI4171.1.
- Uppala, S. M., et al. (2005), The ERA-40 Re-analysis, *Q. J. R. Meteorol. Soc.*, Oct. 2005 Part B.
- Vamborg, F. S. E. (2011), Vegetation-albedo-precipitation interactions in North Africa during the Holocene, Rep. Earth System Science 94, Max Planck Institute for Meteorology, Hamburg, Germany [Available at <http://www.mpimet.mpg.de/wissenschaft/publikationen/berichte/erdsystemforschung.html>].
- Wang, B., Q. Ding, X. Fu, I.-S. Kang, K. Jin, J. Shukla, and F. Doblas-Reyes (2005), Fundamental challenge in simulation and prediction of summer monsoon rainfall, *Geophys. Res. Lett.*, 32, L15711, doi:10.1029/2005GL022734.
- Weedon, G. P., S. Gomes, P. Viterbo, W. J. Shuttleworth, E. Blyth, H. Österle, J. C. Adam, N. Bellouin, O. Boucher, and M. Best (2011), Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century, *J. Hydrometeorol.*, 12, 823–848, doi:10.1175/2011JHM1369.1.
- Zhang, Y.-C., W. B. Rossow, A. A. Lacis, V. Oinas, and M. I. Mishchenko (2004), Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data, *J. Geophys. Res.*, 109, D19105, doi:10.1029/2003JD004457.

---

Corresponding author: S. Hagemann, Max Planck Institute for Meteorology, Bundesstr. 53, 20146 Hamburg, Germany. (stefan.hagemann@zmaw.de)