Assessment of JSBACHv4.30 as land component of ICON-ESM-V1 in comparison to its predecessor JSBACHv3.2 of MPI-ESM1.2

Rainer Schneck[1], Veronika Gayler[1], Julia E.M.S. Nabel[1,2], Thomas Raddatz[1], Christian H. Reick[1], and Reiner Schnur[1]

[1]Max Planck Institute for Meteorology, Hamburg, Germany
[2]Max Planck Institute for Biogeochemistry, Jena, Germany

Correspondence: Rainer Schneck (rainer.schneck@mpimet.mpg.de)

Abstract. We assess the land surface model JSBACHv4, which was recently developed at the Max Planck Institute for Meteorology as part of the effort to build the new Earth System model ICON-ESM. We assess JSBACHv4 in simulations coupled with ICON-A, the atmosphere model of ICON-ESM, hosting JSBACHv4 as land component to provide the surface boundary conditions. The assessment is based on a comparison of simulated albedo, Land Surface Temperature (LST), Leaf Area Index (LAI), Terrestrial Water Storage (TWS), Fraction of Absorbed Photosynthetic Active Radiation (FAPAR), Net Primary Production (NPP), and Water-Use-Efficiency (WUE) with corresponding observational data. JSBACHv4 is the successor of JSBACHv3, therefore, another purpose of this study is to document how this step in model development has changed model biases. This is achieved by also assessing in parallel results of coupled land-atmosphere simulations with the preceding model ECHAM6 hosting JSBACHv3.

Large albedo biases appear in both models over ice sheets and in central Asia. The temperate to boreal warm bias observed in simulations with JSBACHv3 largely remained in JSBACHv4, despite the very good agreement with observed LST in the global mean. For the assessment of changes in land water storage, a novel procedure is suggested to compare the gravitational data from the GRACE satellites to simulated TWS. It turns out that the agreement of changes in the seasonal cycle of TWS is sensitive to the representation of precipitation in the atmosphere model. The LAI is generally too high which is partly caused by too high soil moisture but also by the parameterization of the phenology itself. The pattern of WUE is for both models largely as observed. In India WUE is too high probably because JSBACH does not incorporate irrigation in our simulations. WUE differences between the two models can be traced back to differences in precipitation patterns in the two coupled land-atmosphere simulations. For both models, most NPP biases can be associated with biases in water stress, LAI and FAPAR. Particularly the NPP bias of the Eurasian steppes has switched from positive in JSBACHv3 to negative in JSBACHv4. This difference is mainly caused by weaker precipitation and lower FAPAR of ICON-A+JSBACHv4 in July, which is most probably caused by a feedback loop between too little soil moisture, evaporation and clouds. While the size and patterns of biases in albedo and LST are largely similar between the two model versions, they are less well correlated for precipitation and vegetation related variables like FAPAR. Overall, the biases found in the different assessment variables are either already known from the previous implementation in MPI-ESM, or have changed because of the coupling with the new atmospheric component ICON-

A. As discussed, there is a good perspective to mitigate these biases by an improved processes representation. Accordingly, this study demonstrates the technically successful completion of the re-implementation of JSBACH into ICON-ESM-V1.

1 Introduction

With the massively increasing parallelism in high performance computing during the last two decades (TOP500 project, 2021) unstructured grids became a favorable choice for general circulation models of the atmosphere and the ocean. Their usage offers good scalability, prevention of grid singularities at the poles, exact conservation of mass, and flexibility in resolution (e.g. for local refinements). Accordingly, the utilization of unstructured grids was the primary motivation to build the icosahedral nonhydrostatic (ICON) modeling framework for a unified next-generation numerical weather prediction and climate modeling system, a joint development of the German Weather Service (DWD) and the Max Planck Institute for Meteorology (MPI-M).

ICON was successfully introduced into the operational forecast system of the DWD in 2015 (Zängl et al., 2015; DWD, 2014). The atmosphere model ICON-A (Giorgetta et al., 2018) as well as the ocean model ICON-O (Korn, 2017; Korn and Linardakis, 2018) were established for climate simulations. Furthermore, a complex Earth System model named ICON-ESM is currently assembled based on ICON and a coupled version of ICON-A and ICON-O (Jungclaus et al., 2022).

As part of this major effort, the ICON-Land framework was developed at MPI-M to facilitate the implementation of complex land surface models not only in ICON-ESM, but also into other modeling environments for global simulations. ICON-Land has a code structure following object-oriented programming concepts. It offers management of processes and a hierarchical handling of tiles that represent different types of land cover within a grid box. As a first application of this framework, it now hosts JSBACHv4. As a consequence JSBACHv4 is the land component of ICON-A and thus also integral part of the first version of ICON-ESM. Technically, JSBACHv4 is a subroutine of ICON-A and part of its scheme for the implicit solution of vertical diffusion. Because of this intimate connection, the assessment of JSBACHv4 performed in the present study is done by means of *coupled* atmosphere-land simulations.

The base version of ICON-ESM – of which JSBACHv4 is part – is described and evaluated in a recent study by Jungclaus et al. (2022). Their evaluation focuses on the characterization of the main features of the new ICON-ESM as well as its performance in simulating the Earth system in its historical development. Our paper is a companion study to amend this evaluation by a more detailed investigation of the ability of its land component JSBACHv4 to represent land surface processes for climate modeling as well as for Earth System Modeling. For this purpose we focus on variables that are on the one hand meaningful for the physical and biological aspects of the overall model result and on the other hand are available as an observation-based dataset for comparison. For these reasons we selected albedo, Land Surface Temperature (LST), Terrestrial Water Storage (TWS), Leaf Area Index (LAI), Fraction of Absorbed Photosynthetic Active Radiation (FAPAR), Net Primary Production (NPP) and Water-Use-Efficiency (WUE) for our assessment. These variables represent processes that are fast compared to the cycling of carbon between land storage pools or climate induced biogeographical changes in landcover. Modules of JSBACHv4 representing the latter two are switched off in our simulations and are thus not part of the assessment.

JSBACHv4 is the successor of JSBACHv3 (Brovkin et al., 2013; Reick et al., 2013; Schneck et al., 2013; Reick et al., 2021), the land component of the Max Planck Earth System Model MPI-ESM 1.2 (Mauritsen et al., 2019) that participated in the Coupled Model Intercomparison Project (CMIP) phase 5 (Giorgetta et al., 2013) and phase 6 (Tebaldi et al., 2021). Most of the parametrizations of JSBACHv4 are re-implementations from JSBACHv3. One aim of the present study is to check if the new ICON-Land framework encompassing the implementation of JSBACHv4 is free from defects. At the same time we want to document the changes in the quality of simulation results when stepping from ECHAM6+JSBACHv3 to ICON-A+JSBACHv4. This is achieved by performing the same assessment as done for JSBACHv4 also for JSBACHv3 so that simulation biases can be compared. Hence, the present study is in fact a parallel assessment of the two JSBACH versions although our main interest lies in JSBACHv4.

We base our assessment on simulations with atmosphere coupling performed in AMIP configuration (Atmospheric Model Intercomparison Project; Gates 1992; Taylor et al. 2000). For JSBACHv4 we perform AMIP simulations of the last three decades with ICON-A that hosts JSBACHv4. For JSBACHv3 we use an existing AMIP simulation that was performed for the same period with the atmosphere model ECHAM6.3 that hosts JSBACHv3 as land component. The key characteristic of AMIP simulations is that the observed historical development of monthly sea surface temperature and sea-ice cover is prescribed. Thereby most of the simulated interannual variability (e.g. related to the El Niño Southern Oscillation or the Pacific Decadal Oscillation) is in phase with the real climate. Moreover, such a model configuration prevents biases that arise in full ESM simulations from internal variability and biases in the ocean model; such biases would impede our discussion of land model performance. Besides, driving alternatively the considered land surface models by observed climate data in stand-alone mode would not permit their assessment as proper ESM components: such simulations would lack the interaction of the land surface with the atmosphere and thereby presumably lead to smaller biases than those that have to be coped with in ESM simulations for which JSBACHv4 and JSBACHv3 were designed. For JSBACHv3, such biases arising from the coupling to atmosphere and ocean has been discussed for some variables in Dalmonech et al. (2015).

The structure of the paper is as follows: First we describe in section 2 our methodology by introducing the considered models and the simulation set-ups, then we describe each assessment variable, how it is calculated in JSBACH, by which observational data it is assessed, and how observational data are pre-processed for this purpose. The subsequent section 3 contains our main results. There we compare for each assessment variable simulation results from the two JSBACH versions with observational data and investigate differences in simulation results from the two models. These results are summarised in section 4 with emphasis on the potential reasons for the emergence of simulation biases and their potential mitigation.

2 Methods

2.1 The models ICON-A+JSBACHv4 and ECHAM6+JSBACHv3

To document the advancement from JSBACHv3 to JSBACHv4, we compare ICON-A+JSBACHv4 with ECHAM6+JSBACHv3 simulation results. As ICON-A (Giorgetta et al., 2018) is the atmospheric component of the Icosahedral Nonhydrostatic Earth System Model (ICON-ESM), so is ECHAM6 (Stevens et al., 2013; Giorgetta et al., 2013) the atmospheric component of the

Max Planck Institute Earth System Model (MPI-ESM), while JSBACHv4 and JSBACHv3 are the respective land components. ICON-A is the successor of ECHAM6 in the sense that it is based on the ECHAM6 physics package with only slight modifications (see Giorgetta et al., 2018). Besides technical infrastructure the major difference concerns their dynamical cores: while ICON-A employs a nonhydrostatic solver for the primitive equations on an icosahedral grid (Zängl et al., 2015), ECHAM6

95 invokes a spectral solver for the hydrostatic approximation on a latitude-longitude grid (Stevens et al., 2013). In both models, the dynamical core solves for atmospheric motion, temperature, density, and concentrations of water in forms of vapor, clouds and cloud ice. The parameterizations of the physical processes, like radiation and cloud condensation, alter these dynamical variables. Crueger et al. (2018) found that, compared to ECHAM6.3, the representation of climate has slightly improved in ICON-A.

100 JSBACHv4 (Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg) is the land component of ICON-A. It was developed at MPI-M as successor of JSBACHv3 (Reick et al., 2013, 2021), with a completely renewed technical infrastructure. However, it applies the same parametrizations as JSBACHv3 (documented in Reick et al., 2021) and includes the additional feature of frozen soil water and a five-layer snow scheme (Ekici et al., 2014; de Vrese et al., 2021). It comprises processes that are important for a land surface scheme: the surface energy balance, terrestrial water budget and runoff, surface exchange

105 fluxes (moisture, heat, carbon), phenology, surface albedo and roughness, radiation in the canopy, plant productivity (photosynthesis, gross and net primary productivity), anthropogenic land cover change and land cover disturbances by wind and fire. The radiation fluxes in the canopy that are needed in the photosynthesis routines to determine the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) are calculated by employing a two-stream approximation (Sellers, 1985). The hydrological soil scheme uses five soil layers ranging 9.8 meter below the surface with increasing thickness towards the bottom.

110 The coupling to the atmosphere is identical in ICON-A+JSBACHv4 and ECHAM6+JSBACHv3, happening via the calculation of albedo, surface roughness, and the surface fluxes of sensible and latent heat, where the latter are obtained as part of the implicit numerical scheme solving for the vertical diffusion of dry static energy and humidity in the atmosphere (Reick et al., 2021).

2.2 Simulation setup

115 For our assessment we perform simulations with ICON-A+JSBACHv4 and compare the results with data from an existing simulation with ECHAM6+JSBACHv3. For the present study we ran ICON-A+JSBACHv4 in R2B4 resolution (grid spacing ∼160 km) using 47 layers for the atmosphere. The simulation employ the 11 plant functional types (PFTs) Tropical broadleaf evergreen, Tropical broadleaf deciduous, Extra-tropical evergreen, Extra-tropical deciduous, Raingreen shrubs, Deciduous shrubs, C3 grass, C4 grass, C3 pasture, C4 pasture and C3 crops. Land cover change is prescribed by annual maps for the

120 fractional cover of the PFTs within each grid cell. These maps are derived as described in Mauritsen et al. (2019) from the Land Use Harmonization (LUH) project (Hurtt et al., 2011) version LUH2v2h (Hurtt et al., 2016, 2020) dataset.

Both simulations were performed for the years 1979 to 2014 in a setup according to the AMIP II protocol (Taylor et al., 2000, 2012), prescribing monthly sea surface temperatures (SSTs), sea ice concentrations (Durack and Taylor, 2017), observed

solar irradiance, and historical greenhouse gas concentrations (Meinshausen et al., 2017). The AMIP simulations were started from the state of an existing historical simulation with the respective full ESM (see below).

For ICON-A+JSBACHv4 the particular model versions used are ICON-A of ICON-ESM-V1 and JSBACHv4.30. For ECHAM6+JSBACHv3 we used published CMIP6 simulations performed with the atmospheric component ECHAM-6.3.05 of MPI-ESM1.2-HR that hosts JSBACHv3.20 as land model (Mauritsen et al., 2019; Müller et al., 2018). These latter data are published as Max Planck Institute for Meteorology (2020). We used ensemble member r1i1p1f1. The ECHAM6+JSBACHv3 AMIP setup is similar to our ICON-A+JSBACHv4 run except for the model grid (T127 $\sim$ 100 km and 95 vertical levels), integer lake mask, no frozen soil, one snow layer, and the additional use of a C4 crop PFT (Table 1). In particular, it applies the same land cover change scheme and input maps as our ICON-A experiment.

In case of ICON-A+JSBACHv4, data from year 1980 were taken for initialization from the respective historical simulation (Jungclaus et al., 2022), while the initialization of ECHAM6+JSBACHv3 is based on the state of the year 1979 of the associated historical CMIP6 simulation (Max Planck Institute for Meteorology, 2020). After model start the atmosphere equilibrates within days. Because land carbon and biogeographical components are not active in our AMIP simulations, the slowest land variable in this setup is soil moisture. By the initialization from a historical simulation, the soil water reservoirs are upon start already filled to a realistic level. Soil water memory is typically a few months, only in desert regions it lasts up to a year (Hagemann and Stacke, 2015). But soil water memory stems (in our model) mostly from below the root zone (Hagemann and Stacke, 2015) so that it is largely decoupled from the active water cycle at the monthly scale that we consider for our assessment.

Table 1. Simulation differences between JSBACHv4 and JS-BACHv3.

Parametrization	JSBACHv4	JSBACHv3
Plant Functional Types	11	11 + C4 crops
Frozen soil	yes	no
Number of snow layers	5	1
Fractional lakes	yes	no

2.3 Assessed variables and their representation in JSBACH

Here we introduce the variables that we selected for our model assessment. Only variables describing fast processes (seconds to years) are considered. Slow processes (decades and longer, e.g. climate induced changes in biogeography and wood or soil carbon turnover) are as well implemented in JSBACHv4 but are not activated in our simulations and are thus not subject to assessment here. To shed some more light on the origin of biases in the selected assessment variables, we consider also some auxiliary variables, these are introduced below jointly with the respective assessment variable. As process descriptions are largely similar in JSBACHv4 and JSBACHv3, in the following we distinguish between those two versions only when necessary.

2.3.1 Albedo

Land surface albedo controls the balance of shortwave radiation at the surface and thus the land energy uptake. It plays a role for the land biosphere as well as for the climate of the lower atmosphere. Its changes are shown to have positive feedback effects with climate (e.g. Claussen, 1997). Changes are driven by natural seasonal and diurnal alterations, as well as human interventions like vegetation changes due to land use (Forster et al., 2007). Surface albedo depends on the canopy properties (e.g. LAI), soil color, the color of litter covering the soil, as well as on snow cover and changes in snow color when aging. As all these properties and the surface albedo itself are typically calculated by the land surface scheme (land model), the albedo is one of the most common variables for its evaluation. This is especially true in consideration that albedo results vary strongly among land models. For example, Levine and Boos (2017) show for northern hemisphere summer that intermodel albedo variance in CMIP5 (Coupled Model Intercomparison Project Phase5) simulations is large compared to interannual albedo variance. Wang et al. (2016) also show a strong intermodel albedo variability in the CMIP5 simulations but for northern hemisphere winter.

The albedo scheme of JSBACH computes temporal and spatial albedo changes. The scheme differentiates between near infrared (NIR) and visible range (VIS) albedo, only for lakes this differentiation is not done. On lakes, the fraction of lake ice and snow on lake ice is taken into account. The albedos for PFTs and bare soil as well as for snow are computed separately. Snow albedo decreases with increasing snow age and surface temperature. The overall albedo on land is then calculated from the fractions of soil, canopy and the corresponding overlaying snow, considering the influence of the incoming solar radiation angle on their fractions. Thereby the coverage of the soil by stems and branches within forest is considered. JSBACHv4 uses the same albedo scheme as JSBACHv3 which is described in detail in Otto et al. (2011).

As a benchmark for our assessment of the JSBACH albedo results we use the MODIS MCD43C3 CMG Collection 6 Albedo Product (Schaaf and Wang, 2015) which is suitable for climate model comparisons (Cescatti et al., 2012; Schaaf et al., 2002). For our comparison we exclude data flagged for minor quality from inversion (quality flag 4 and 5).

The contributions from snow cover make a large contribution to the albedo values in the extra tropics. Therefore, as part of the discussion of albedo results, we also discuss the simulated distribution of snow cover and compare it with observed snow cover from the same MODIS dataset.

2.3.2 Land Surface Temperatures (LST)

Virtually all land processes depend directly or indirectly on LST. It is a key variable in the surface energy balance and thus takes part in the control of thermal, radiative, and hydrological exchange fluxes at the interface between land and atmosphere that shape local climate and the state of the lower atmosphere. It plays a central role for cryospheric processes (amount and duration of snow cover, formation of soil ice) and determines local living conditions for fauna and flora.

Atmospheric temperature is one of the most regarded prognostic variables in climate models. Its calculation depends over land on LST as lower boundary condition that is provided by the respective land component. In JSBACH land surface temperature is obtained from the surface energy balance equation, whose solution is embedded in the vertical diffusion scheme for

heat and moisture fluxes in the atmosphere (implicit coupling). LST is also used as upper boundary condition to determine the vertical temperature profile within the five-layer soil model assuming vanishing heat fluxes at the bottom (10-m depth).

For comparison of simulated LST with observations, we use the MOD11C1 Moderate Resolution Imaging Spectroradiometer (MODIS) Terra Land Surface Temperature/Emissivity V006 dataset (Wan et al., 2015). We excluded data points where the quality flags indicate no retrieval because of clouds.

As a part of the LST analysis we regard the modelled Total Cloud Cover (TCC) as compared to observed TCC from the Collection 6.1 EOS-TERRA MODIS Atmosphere Level-3 Daily Product (Platnick, 2017).

2.3.3   Terrestrial Water Storage (TWS)

TWS influences many surface properties connected to climate. Especially its particular representation in land surface schemes has a major impact on calculated evaporation, transpiration (root water uptake), NPP and LAI (water limitation). As a result, the latent heat flux into the atmosphere heavily depends on TWS. On a global scale, its seasonality peaks in about April and is lowest around September (Swenson and Milly, 2006). Due to the local seasonality of precipitation and evapotranspiration, TWS has a strongly site dependent seasonality (Feng et al., 2012; Hickel and Zhang, 2006; Settin et al., 2007). Even when TWS has a strong dependency on the atmosphere through precipitation and air temperature it is still determined by surface properties like the soil type and vegetation. Therefore, its depiction in climate models depends largely on the particular land model. Koster et al. (2009) drove a number of land surface models with the same atmospheric forcing and concluded that soil moisture is highly dependent on the particular land surface model.

TWS encompasses water in snow, vegetation, soil, runoff and aquifers. In JSBACH, TWS is the sum of water stored as snow and water on the surface and the canopy, as soil water and soil ice, and as runoff. Aquifer water impacts climate only when it emerges as open water (runoff, lake or ocean). Therefore, aquifer water is not explicitly represented in JSBACH (as in most land models). JSBACH also has no explicit store for water in vegetation. Formally, this water is part of the soil water pool because for transpiration the water is taken directly from the soil water pool. The water budget in JSBACH is calculated from processes above ground, the soil hydrology and the river runoff. Above ground processes are the snow and water that fall on the surface or that are intercepted by the canopy. The water that reaches the surface (as rain, snow melt and dew) can infiltrate the soil. From that only a part is taken up by the soil, the remaining part enters the surface runoff. Vertical movements of water in the soil are the result of its vertical diffusion and its gravitational percolation. Water in- and output occur at the surface in form of rain, evaporation and snow melt (on the surface and the canopies). Water transpired by the vegetation is extracted from the soil water reservoir. The size of transpiration depends on primary production, root depth, and specific humidity of the lowest atmospheric layer. At the bottom of the soil water is lost as drainage which is added to the overall runoff. Because TWS could not be calculated from the ECHAM6+JSBACHv3 simulation output we only assess JSBACHv4.

The Gravity Recovery And Climate Experiment (GRACE) is a project to accurately determine the Earth's gravitational field with detailed measurements from satellites. For our assessment of simulated TWS we use the ITSA_Grace2018s unconstrained monthly TWS Dataset (Kvas et al., 2019; Mayer-Gürr et al., 2018). It is derived from the GRACE satellite sensing of changes in the gravitational field of the Earth. These changes are reprocessed to reduce the effects of mass trends originating from

glacial isostatic adjustment, postseismic deformation after large earthquakes and atmospheric mass variability. Therefore, the dataset essentially senses water mass anomalies independently of their surface exposure and thus integrates all water mass changes in snow cover, vegetation, soil, runoff and aquifers.

Here we are mainly interested in the question how JSBACHv4 performs in the context of climate and Earth System modelling. For the fluxes to the atmosphere a correct reproduction of the seasonal cycle in TWS is essential. Thus, only changes of TWS in the course of the year are analysed and not long-term trends. Our comparison is based on the assumption that the additional signal from the hydrology below 10 m that is present in the observational data but not in JSBACH does only negligible contribute to monthly TWS changes. We assume that below 10 m depth the hydrological processes are already so slow that they don't add to the phasing of the overall seasonal signal. This pertains in particular to potential signals from aquifers, whose recharging times are much larger than the monthly time scale considered in our comparison (typically decades to millennia).

As model results and GRACE data are derived by totally different methods we expect that the size of monthly TWS change will systematically differ. For example, both data sets are expected to have different standard deviations. Before comparison we therefore normalize the observed and simulated data according to their own size of variation. For the normalization we use averaged absolute values instead of the standard deviation. Using the standard deviation would put more weight on extreme values, which makes sense to lower the impact of statistical outliers. However, here statistical outliers are unlikely as we regard multi year averages. Therefore we use for each particular grid cell:

$$\overline{\Delta x} = \frac{1}{12} \sum_{m=1}^{12} |\overline{x}_m|, \tag{1}$$

where $\overline{x}_m$ is the average of observed TWS change for month m calculated over the period of available observation data (years 2003 to 2014). From this we calculate the normalized month-to-month differences as

$$\Delta \text{TWS}_m := \frac{\overline{x}_{m+1} - \overline{x}_m}{\overline{\Delta x}}, \tag{2}$$

with m = 1, 2,...11. Denoting simulated values by TWS' and applying to them the same calculations, we measure the average mismatch between simulated and observed TWS in month-to-month changes across a year by

$$Q_{\text{TWS}} := \sum_{m=1}^{11} |\Delta \text{TWS}'_m - \Delta \text{TWS}_m|. \tag{3}$$

This value can reach a maximum of 22 and the closer to zero it is, the more synchronous are observed and simulated month-to-month changes.

Only the average seasonal changes in TWS are evaluated by this method. The amplitude of TWS variations as well as long-term trends in TWS are not considered. Therefore, a low value of Q_{TWS} does not necessarily mean that variations in TWS are simulated realistically in a quantitative sense, but it shows that the seasonal phasing of TWS is captured by the model.

2.3.4 Leaf Area Index (LAI)

The LAI strongly affects the exchange of energy and matter with the atmosphere through its impact on albedo and the fluxes of water and carbon (Chase et al., 1996; Betts et al., 1997; Piao et al., 2007). A high LAI typically enhances transpiration of water

and in general also primary productivity. The LAI depends primarily on the climatic conditions for a given biome. With the onset and eventual loss of leaves at the beginning and end of the growing season the LAI reveals a strong seasonality. However, in contrast to soil moisture, it remains unaffected by short precipitation spells. In climate models, the LAI is time-dependent and either prescribed from observations or calculated by the corresponding land surface scheme. How it is calculated strongly depends on the particular land model (see e.g. Wang et al., 2016).

JSBACHv3 provides different schemes to compute the LAI (Reick et al., 2021). The default scheme, which is also implemented in JSBACHv4, is the LoGro-P model (Logistic Growth Phenology). In LoGro-P the LAI is only coupled to climate and not to a specific carbon allocation to leaves. The advantage of this simplified approach is that due to the missing feedback of the carbon pools to the LAI it is possible to run the carbon and vegetation dynamics submodels in different configurations without the phenology model and with the same NPP. Each PFT is assigned to one of the following phenology types: evergreen, summergreen, raingreen, grasses, tropical and extra-tropical crops. Each phenology type has a maximum LAI, which represents its physiological limit. Changes in the summergreen phenology are based on three phases: a growth (in spring), vegetative (in summer) and a rest phase (in autumn and winter). The raingreen, grasses and tropical crop phenology types only have a growth phase; they grow whenever the environmental conditions (soil moisture, temperature, NPP) are favorable. The evergreen and extratropical crop phenologies have no vegetative phase. Dependent on the phenology type, different environmental conditions determine the advancement from one phase to the next. For example, the evergreen, summergreen and extratropical crop phenologies use accumulated temperatures (heat sum) for entering the growth phase when it reaches a PFT-specific threshold value. Overall, the LAI increases when the soil is wet and temperatures and NPP are high. A detailed description of the LoGro-P model is given in Böttcher et al. (2016).

For the assessment, we compare our simulation results for LAI with the MODIS MOD15A2Hv006 LAI Product (Myneni et al., 2015). We use only data where the primary or secondary quality flag reveal more than a 50% fraction of measurements with best or good quality (primary/secondary quality flag of 51-100 and 251-300).

2.3.5 Fraction of Absorbed Photosynthetic Active Radiation (FAPAR)

The Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) quantifies the fraction of the Photosynthetically Active Radiation (PAR, solar radiation in the 400 -700 nm spectral domain) absorbed by leaves for photosynthesis. FAPAR depends on the LAI, the optical properties of the leafs and their orientation, atmospheric conditions, the angle of incoming radiation, and the albedo of the underlying soil. It determines Gross Primary Productivity (GPP) and plays a key role for plant respiration and transpiration.

To solve the complicated radiation problem JSBACH uses a so called two stream approximation approach. This approximation requires that the leaves are distributed homogeneously in the canopy and the radiation distribution within the canopy is horizontally invariant. Hence for closed canopies it is sufficient within the canopy to consider vertically up- and downward radiation fluxes. However, the concrete implementation of this approach into JSBACH is rather complex and we refer the reader to Loew et al. (2014) for details.

280 For our assessment of FAPAR we regard the differences between our model and the MODIS MOD15A2Hv006 Product. We took only the data where the primary or secondary quality flag have more than a 50% fraction with best or good quality measurements.

2.3.6 Net Primary Production (NPP)

The exchanges of carbon between terrestrial ecosystems and the atmosphere are dominated by GPP and plant respiration
285 (Houghton, 2007). NPP, as the net carbon flux from these two, is therefore very important for the carbon balance between atmosphere and land. It depends on LAI, TWS and WUE as well as on temperature and radiation. It varies seasonally and annually, depending on the climatic conditions (precipitation, temperature and radiation). Globally it peaks in northern hemisphere summer and has its low point in winter. As shown in Cramer et al. (1999) its representation strongly depends on the land model.

290 In JSBACH GPP is calculated from FAPAR under consideration of water stress. Water stress is quantified by calculating the relative amount of water found in the root zone taking into account the wilting point below which plants cannot extract any more water from the soils and above which transpiration is largely unhindered. In JSBACH carbon is respired on the one hand as maintenance respiration (to cover the basic plant functions like photosynthesis, nutrient and water transport, repairs, defence) and on the other hand as growth respiration (to cover plant growth). Note that autotrophic respiration is calculated
295 as a prescribed fraction of GPP instead from carbon allocated in the different plant tissues. Accordingly, NPP belongs as the other variables considered here to the 'fast' variables of JSBACH that can be assessed without consideration of the cycling of carbon between different storage pools. In JSBACH the net balance of GPP and respired carbon is called potential NPP. In times when more carbon is respired than gained from the assimilation potential NPP is negative. In this case as much carbon as possible is taken from the reserve pool (representing carbon in plants stored in sugars and starches) to obtain the actual NPP.
300 If potential NPP is positive and can be allocated to the plant carbon pools or through root exudates to the soil carbon, potential NPP equals actual NPP. However, if potential NPP can not be allocated because of structural limits of the plants (plants can not grow infinitely large) as much carbon as possible is allocated (actual NPP is smaller than the potential NPP) and the remaining carbon is added as root exudates to the soil carbon.

 We use the MODIS-C006_MOD17A2Hv006 dataset as benchmark for our actual NPP assessment (Running and Zhao,
305 2019). We use only data with confidence flag having the values 0 (very best possible) and 1 (good, very usable, but not the best).

2.3.7 Water-Use-Efficiency (WUE)

Continuous water uptake is a prerequisite for plant growth. But how much they need to photosynthesize a unit of organic carbon is different between different plants and, at a larger scale, different between ecosystems. This efficiency by which
310 plants use water is closely related to the way plants operate their leaf stomata: By closing them less leaf water evaporates into the surrounding air which reduces the risk of desiccation, while their opening facilitate access to atmospheric CO_2 and thereby assimilation and growth. In dry areas growth of plants is typically limited by lack of water and they have developed means

to use water to assimilate carbon more efficiently than plants in wet regions (e.g. C_4 instead of C_3 carbon fixation). But even under the same climate conditions Water-Use-Efficiency (WUE) can differ depending on the species (Niu et al., 2011), stand age and ecosystem structure. Overall, WUE strongly defines carbon and water exchanges with the atmosphere (Sun et al., 2016).

WUE can be characterized in different ways (Beer et al., 2009). In the context of land surface models with their large grid cells only a definition at biome or ecosystem level is useful. Here we define WUE as the ratio between annual GPP and annual evapotranspiration (ET) integrated over all PFTs in a grid cell. This characterizes WUE at ecosystem level as it employs evapotranspiration that not only includes transpiration but also evaporation from ground. Moreover, this definition has the advantage that for GPP and ET global observational data sets exist from which this WUE can be calculated.

In JSBACH, GPP is derived in a two step approach. First, *potential* GPP is calculated assuming absence of water stress so that the plants photosynthesize at maximum rate. Then the soil hydrology model computes the corresponding potential water loss from transpiration. In a second step *actual* GPP is calculated accounting for a possible lack of soil water. JSBACH differentiates between C3 and C4-photosynthesis. For C3-plants an implementation of the Farquhar et al. (1980) model and for C4-plants an implementartion of the Collatz et al. (1992) model is used. Both models take into account that the carboxylation and electron transport rate depend on temperature.

We calculate the JSBACH WUE for each grid cell as total actual GPP from all PFTs divided by total evapotranspiration in that grid cell and average these quotients over the years 2001 to 2014. We compare the JSBACH results with the WUE calculated from MODIS data. For that we divide the MODIS-C006_MOD17A2Hv006 GPP (Running and Zhao, 2019) with MODIS MOD16A3GF.061 evapotranspiration (Running and Zhao, 2019). For evapotranspiration we excluded data flagged with a minor quality above 80% (bad days/total days). We compare the JSBACH results also with the global distribution of WUE published by Sun et al. (2016) calculated partially from different observation based data. To ease visual comparison, we plotted our figures such that contour levels and colors are as theirs.

2.4 Data Preparation

Table 2 summarizes the variables that are compared with corresponding observational data. The spatial resolution of the MODIS data is 0.05° x 0.05° (about 5.6 km at the equator), of the GRACE data 2° x 2° (about 220 km at the equator), and of the GPCP data 2.5° x 2.5° (about 280 km at the equator). The observational data as well as our model results are interpolated with a first order conservative remapping to a Gaussian 96 x 192 lonlat grid (T63; about 1.88°/210 km at the equator). In the MODIS data quality flags are included which indicate issues from elicitation arising from atmospheric scattering and absorption, anisotropy, inadequate temporal, spatial and spectral sampling, and narrowband to broadband conversions. For the MODIS data the interpolation is done after the data points with minor quality are excluded (see subsections above for details). Except of TWS, we average our model results, the MODIS data, and the GPCP data over the years 2001 to 2014. The TWS and the corresponding model data are averaged over the years 2003 to 2014 because the GRACE dataset misses 2001 and parts of 2002. For our precipitation results we use only grid cells where the differences between JSBACH and GPCP are lager than the GPCP error bar. For the assessment of the seasonality we selected January and July to represent deep winter and high summer

in the two hemispheres. The respective maps were obtained by averaging all Januaries and Julys in the considered time range 2001 to 2014. All presented spatial correlations are weighted with the grid cell area.

Table 2. Variables and the observational data used for their assessment. Assessment variables are plotted bold.

Variable	Corresponding observational data	Reference
Albedo and snow cover	MODIS MCD43C3 CMG Collection 6 Albedo Product	Schaaf and Wang (2015)
LST	MODIS MOD11C1 Terra Land Surface Temperature/Emissivity V006	Wan et al. (2015)
TCC	Collection 6.1 EOS-TERRA MODIS Atmosphere Level-3 Daily Product	Platnick (2017)
Precipitation	GPCP monthly precipitation dataset from 1979-2021	Adler et al. (2003)
TWS	ITSA_Grace2018s unconstrained monthly TWS Dataset	Mayer-Gürr et al. (2018)
LAI	MODIS MOD15A2Hv006 LAI Product	Myneni et al. (2015)
FAPAR	MODIS MOD15A2Hv006	Myneni et al. (2015)
NPP	MODIS-C006_MOD17A2Hv006	Running and Zhao (2019)
WUE	see GPP and Evapotranspiration; Sun et al. (2016)	Sun et al. (2016)
GPP	MODIS-C006_MOD17A2Hv006	Running and Zhao (2019)
Evapotranspiration	MODIS MOD16A3GF.061	Running et al. (2021)

Abbreviations: Land Surface Temperature (LST), Total Cloud Cover (TCC), Terrestrial Water Storage (TWS), Leaf Area Index (LAI), Fraction of Absorbed Photosynthetic Active Radiation (FAPAR), Net Primary Production (NPP), Water-Use-Efficiency (WUE), MODerate resolution Imaging Spectroradiometer (MODIS), Gravity Recovery And Climate Experiment (GRACE), Global Precipitation Climatology Project (GPCP).

To give a general overview of the agreement between the model results and the observation data we use the Normalized
350 Mean Error (NME) of Kelley et al. (2013). It is calculated as

$$NME = \frac{\sum_{i=1}^{n} |X_{model,i} - X_{obs,i}|}{\sum_{i=1}^{n} |X_{obs,i} - \overline{X_{obs}}|} \tag{4}$$

where n is the number of land grid cells, X is the actual value of the assessment variable for the model or the observations and $\overline{X_{obs}}$ is the mean value of the observations over the grid cells. The higher the value of NME the stronger the model bias as compared to the spatial variability of the observations. Zero means perfect agreement between model and observations. One
355 means the model bias is as strong as the spatial variability of the observations.

3 Results

Here we present our results from ICON-A+JSBACHv4 and ECHAM6+JSBACHv3 simulations. One general tendency is that simulation results from the two models are quite similar, so that often the discussion of the results applies to both models – in such cases we simply talk of 'JSBACH' without version number. In all other cases, when we discuss specifics of results from
360 a particular model version, the full model name will be used. Because biospheric variables strongly depend on environmental conditions, we discuss physical variables first.
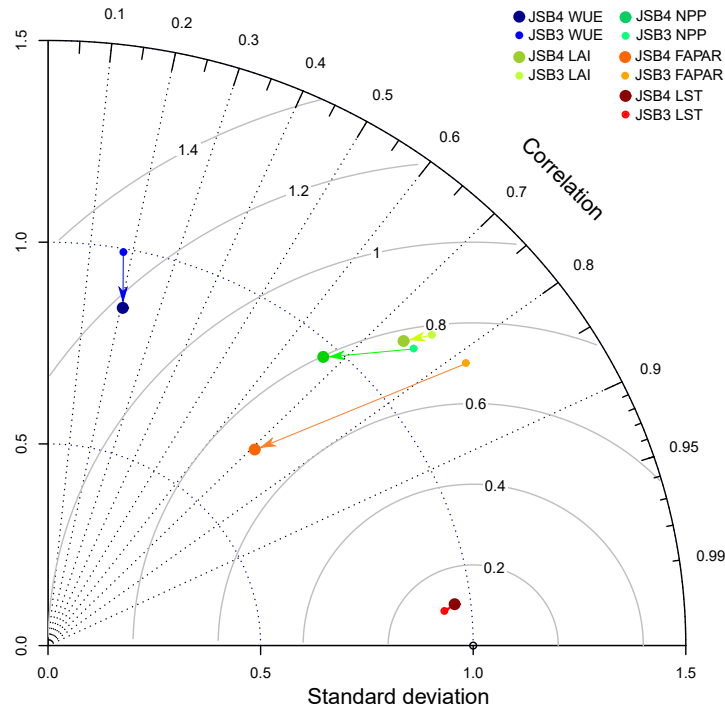
Figure 1. Taylor Diagram (Taylor, 2001) of normalized pattern statistics for annual means of our main evaluation variables LST, LAI, FAPAR, NPP and WUE. The diagram contains no values for albedo because during polar winters observations are missing. JSBACHv3 is shown in small dots and JSBACHv4 in big darker dots. Arrows indicate the change from JSBACHv3 to JSBACHv4. The centered pattern root-mean-square error (RMSE) and standard deviations have been normalized by the observed standard deviation of each field before plotting. Correlations are depicted by lines from the origin representing angles relative to the horizontal base line. Standard deviations are represented as arcs around the origin. Normalized centered pattern RMSE values are seen as circular arcs around the value 1 of the baseline (normalization on standard deviation 1.0).

3.1 General Performance

Taylor (2001) proposed a graphical way to depict the overall performance of a model. The corresponding Taylor Diagram for JSBACHv4 and JSBACHv3 is shown in Fig. 1. The LST achieved the by far best agreement of the statistical metrics with observations. These metrics remained nearly unchanged from JSBACHv3 to JSBACHv4. Except for the LST all standard deviations are reduced in JSBACHv4 (especially FAPAR). Overall some statistical metrics of JSBACHv4 improved while others worsened as compared to JSBACHv3. Thus the overall performance remained more or less the same. This is also visible in the NME in Table A1 and A2, where the NME magnitude is very similar between the assessment variables of both models.

365
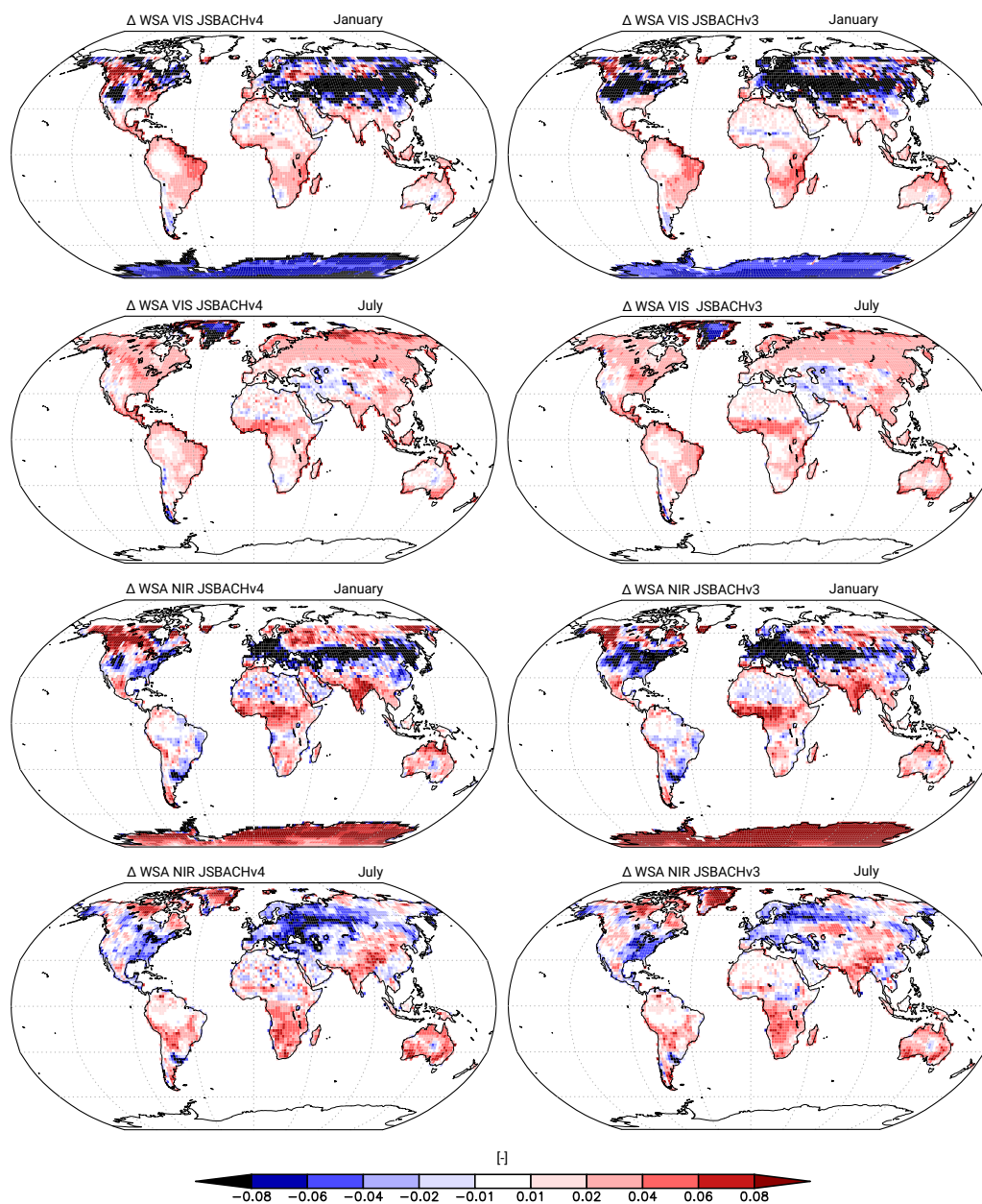
Figure 2. Bias in White Sky Albedo (WSA) for JSBACHv4 (left) and JSBACHv3 (right) compared to MODIS data averaged over the years 2001 to 2014. Shown are biases in the near infrared (NIR) and visible (VIS) range for January and July. Because of polar night no observation data are available for Arctic and Antarctic winter. Note that significance is not shown because nearly all differences are significant (p=0.05) according to an independent two-sample t-test.

Table 3. Spatial Spearman rank correlations (rho) between the biases of JSBACHv4 against those of JSBACHv3 for all grid cells.

	Albedo VIS	Albedo NIR	LST	LAI	FAPAR	NPP
Annual	-	-	0.854	0.658	0.336	0.735
January	0.754	0.756	0.856	0.735	0.333	0.769
July	0.844	0.829	0.701	0.599	0.340	0.551

3.2 Albedo

370 Fig. 2 shows for both JSBACH versions the albedo biases, separately for the visible (VIS) and near infra red (NIR) range. The bias pattern is very similar for JSBACHv3 and JSBACHv4 with exceptions only found in central North America and some regions in North-Eastern Europe (e.g. in the January values of the VIS albedo). The similarity of bias patterns between the two model versions is visible in Table 3. It is also visible from the scatter plots in Fig. B2: The correlations (r and rho) are about 0.7 and higher except of r∼0.6 for VIS albedo in January when a larger scatter arises from polar regions.

375 For the biases seen in Fig. 2 three main causes can be identified. First, the too low VIS albedo and the too high NIR albedo over glaciers (Antarctica, Greenland) is likely a direct result of the fixed minimum and maximum albedo values used in the JSBACH calculations of glacier albedo. Second, the strong negative VIS and NIR albedo bias seen for January in western N-America, eastern Europe and large parts of central Asia is most probably a result of the bias in the dynamically calculated snow cover fraction (compare Fig. 3). And third, the biases in all other regions that are not covered by glacier or snow are at

380 least partly caused by old canopy and soil albedo maps used in JSBACH. These biases will be explained in the next paragraph.
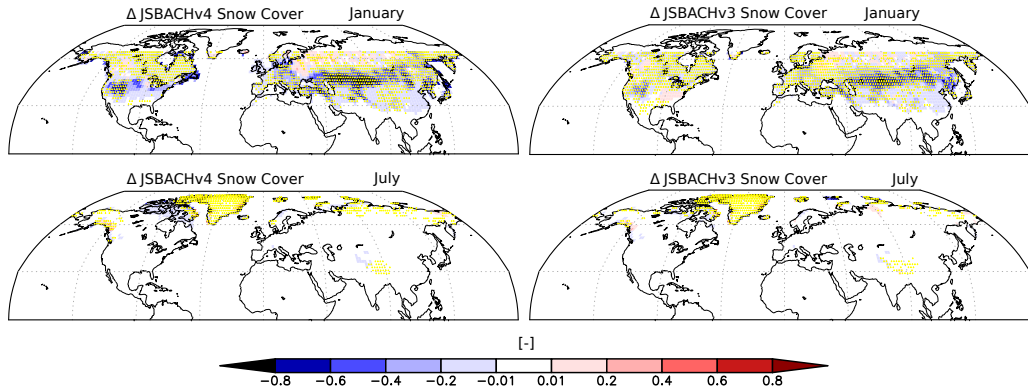
Figure 3. Bias in snow cover for JSBACHv4 (left) and JSBACHv3 (right) compared to MODIS data. Shown are the averages of January (top row) and July (bottom row) averaged over the years 2001 to 2014. The yellow dots indicate grid boxes where the bias is significant (p=0.05) according to an independent two-sample t-test.

Our analyses reveals that the reason for the too low NIR albedo obtained for northern midlatitudes summer is the static map used for soil albedo. The origin of the mostly too large albedo values in other regions (S-America, Africa, India, Australia)
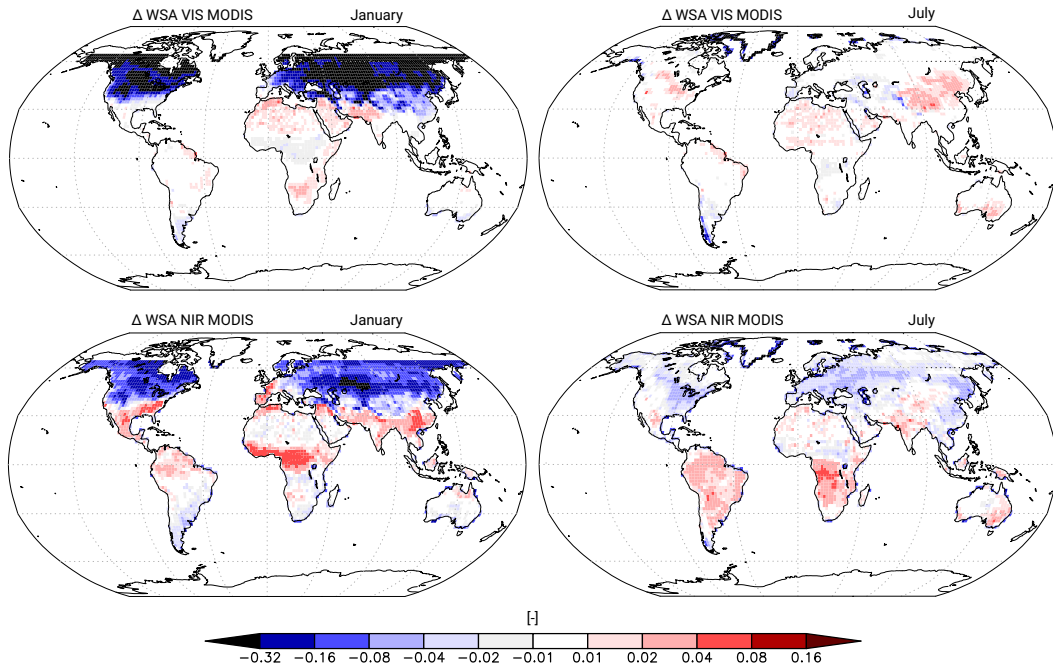
Figure 4. Difference in White Sky Albedo (WSA) between the two versions of MODIS data averaged over the years 2001 to 2014: The older one, from which the soil albedo maps used in JSBACH were derived, and the newer one used for our bias analysis in Fig. 2. Shown are NIR and VIS. As JSBACH does not account for seasonality in the soil albedo maps, the old version is a yearly mean. From this version we subtracted the January and July values of the newer version. The differences shown here can be interpreted as the 'bias' of JSBACH compared to the observations for January and July.

are not traceable to single causes but must be a geographically complex result of the calculated superposition of the assumed values for canopy albedo (which depend on vegetation type) and the static map for soil albedo. A closer inspection (not shown) indicates that in the NIR range JSBACH has a too high canopy and a too low soil albedo while for VIS albedo the converse must be suspected. JSBACH uses the same canopy and soil albedo maps for the whole year and does not incorporate seasonality. The canopy and soil albedo maps used in JSBACH are derived from a fit (see discussion for details) to the 2001 to 2004 mean of an older MODIS Product (MODIS MOD43C1 CMG Collection 4 Albedo Product, Strahler et al. (2021), Gao et al. 2005). As the version used internal in JSBACH is a yearly mean one would expect strong differences to observations for January and July. In Fig. 4 we show the difference between the version used internal in JSBACH and the MODIS observations for January and July we use here for our bias analysis. These differences explain a considerable part of the albedo biases seen in Fig. 2: In Africa for example for both months considered and both spectral ranges analyzed large parts of the JSBACH biases are already visible as bias of the older compared to the newer MODIS product. The same holds for South-East Asia, and for the NIR albedo in India. Accordingly, a considerable part of the biases in albedo arise from the maps for canopy and soil albedo. Note that the rather large differences between the two albedo products seen in Fig. 4 at high latitudes in northern winter in the

visible range are irrelevant for our analysis here, because in these regions at that time of the year the values from the static maps of soil and canopy albedo do not enter the calculated albedo value in JSBACH because of snow there.

3.3 Land Surface Temperature (LST)

Global mean LST of 9.6° C simulated by JSBACHv4 matches well with observational data (see Fig. 5 top). The lower northern
400   hemisphere winter values found for JSBACHv4 (Fig. 5 bottom) are the main reason for the overall about one degree colder climate compared to JSBACHv3 and the better fit with the MODIS data. For JSBACHv4 the values for January are up to two degrees lower than for JSBACHv3, otherwise the seasonality is very similar for the two model versions and the phasing is consistent with MODIS observations.
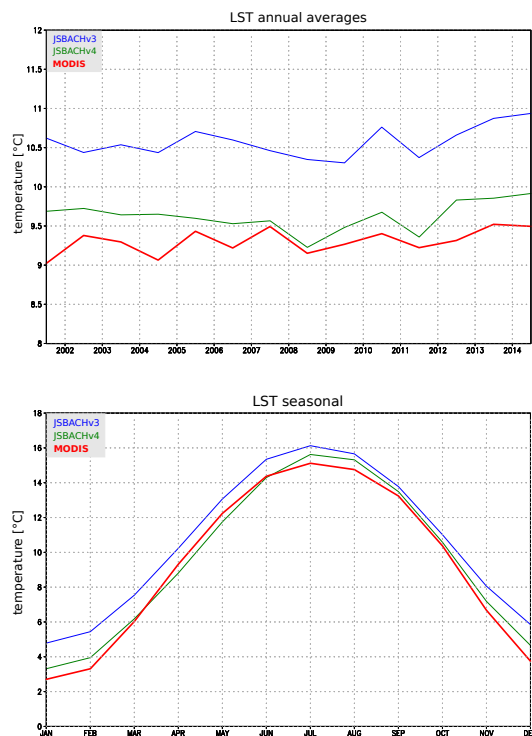
Figure 5. Global mean LST for JSBACHv3, JSBACHv4, and observation values from MODIS. Top: Annual averages from 2001 to 2014. The vertical dashed lines indicate January 15th of the respective year. Bottom: Monthly mean values over the time period form 2001 to 2014. The vertical dashed lines indicate the 15th of the respective month.

In Fig. 6 we show the geographic distribution of temperature biases with respect to MODIS; to the left we show the biases
405   for JSBACHv4, and to the right those for JSBACHv3. For both models the strong regional biases cancel each other in the global mean leading to the relatively weak biases in Fig. 5 (top). At annual mean (Fig. 6 top row plots), the geographic distribution of temperature biases is very similar for the two models, and their difference is typically smaller than the bias. The zonal plots to the left and right reveal that the higher northern latitudes warm bias is as well significant when considered as a large scale phenomenon (for regions north of 45°N the uncertainty range seen in the zonal plots lies off the line of zero bias). This
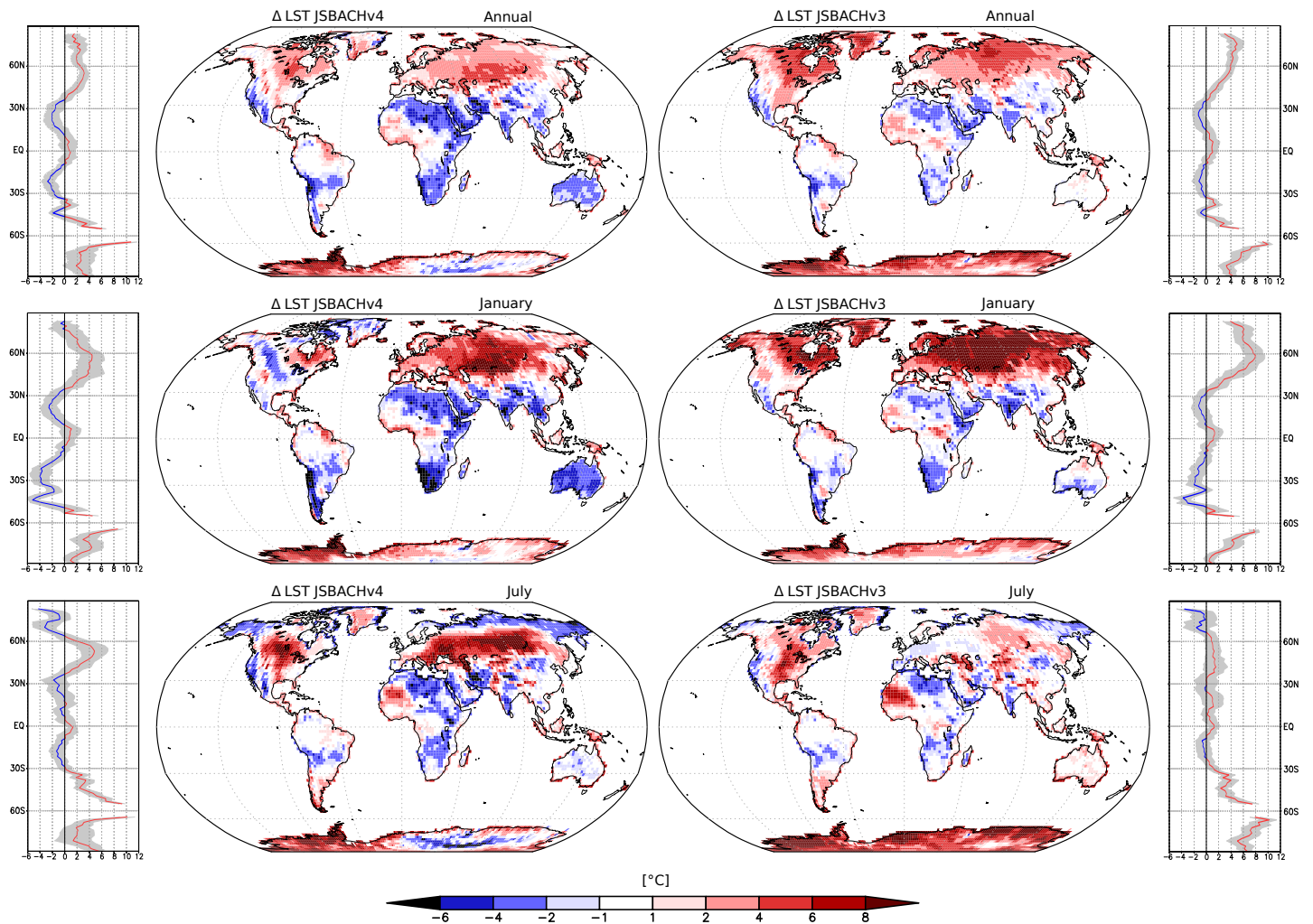
Figure 6. Bias in Land Surface Temperature (LST) compared to MODIS data averaged over the years 2001 to 2014. Shown are the annual averages (top row), and averages of Januaries (mid row), and Julys (bottom row). The zonal and spatial plots in the left two columns refer to JSBACHv4, while those in the right two columns refer to JSBACHv3. The grey range in the zonal plots indicates the 95% uncertainty range calculated from a Student t-distribution so that zonal values are significantly different when the zero bias line is found outside this grey range. Note that the significance for the grid cells of an independent two-sample t-test is not shown because nearly all biases are worldwide significant (p=0.05).

410 temperate to boreal warm bias is for both model versions visible during northern winter (see the mid row plots for January), but less strong for JSBACHv4. In July the bias is much more pronounced in JSBACHv4 (mainly around 50°N, also visible in Fig. B3) and in contrast to JSBACHv3 significant in the zonal mean (see Fig. 6). In January the above named snow-albedo feedback surely contributes to the warm bias in JSBACH. In July the negatively biased JSBACHv4 precipitation (Fig. 8)
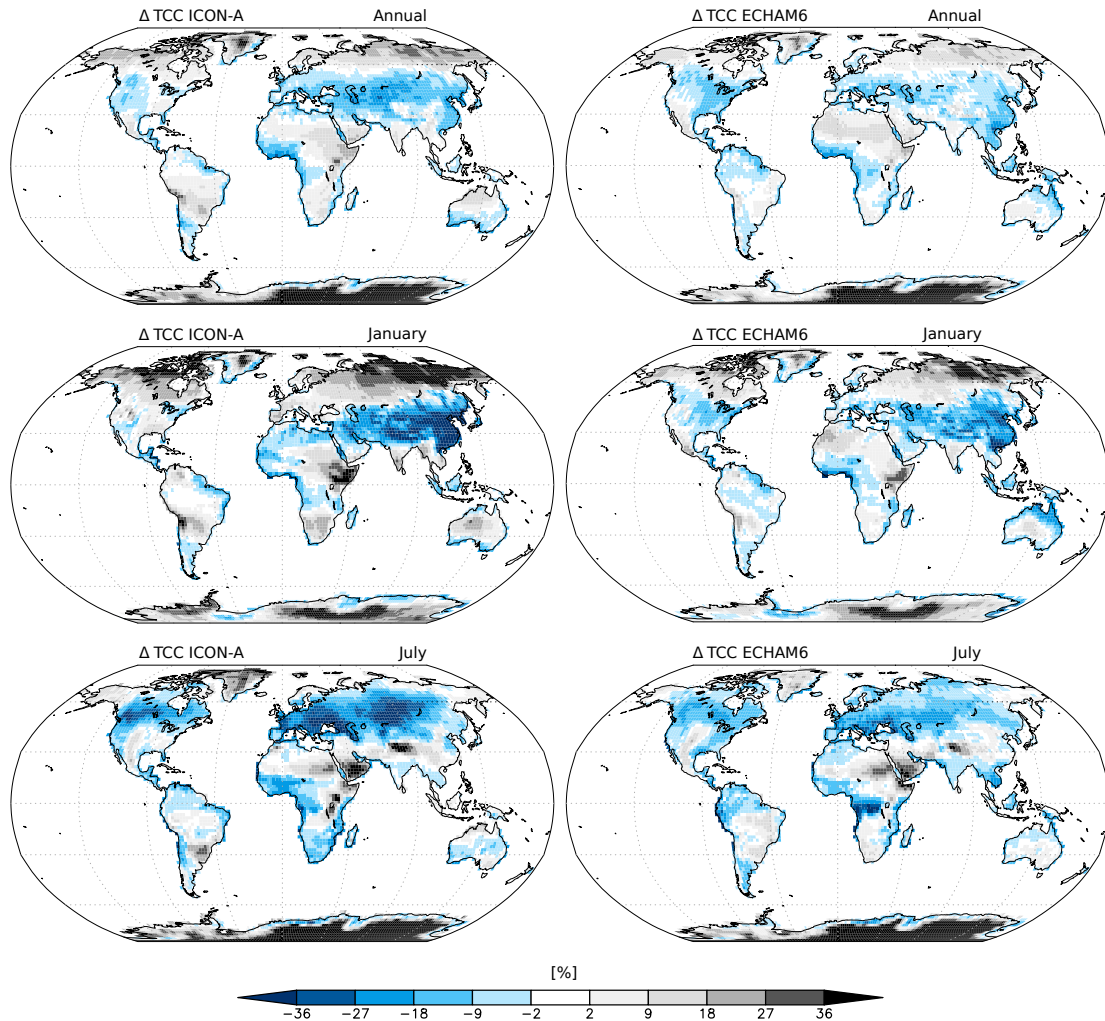
Figure 7. Bias in Total Cloud Cover (TCC) for ICON-A (JSBACHv4) and ECHAM6 (JSBACHv3) compared to MODIS data. Shown are the annual, January and July means over the years 2001 to 2014.

obviously supports the warm bias in Eurasia through the associated evaporation reduction (Table A3 shows a correlation of
415 ∼-0.6). Table 4 also shows a high correlation (∼-0.8) between the JSBACHv4 July LST bias in Eurasia and a weaker Total
Cloud Cover (TCC, see Fig. 7). It seem obvious that the JSBACHv4 biases in this area are the result of a feedback loop: too
little clouds and precipitation result in too low soil moisture (SM) and evaporation which feeds back to precipitation and is
associated with the too high LST. In the follwing this will be called the precipitation-SM feedback. It is probably caused by the
frozen soil and 5-layer snow scheme implemented in JSBACHv4 (see discussion). In contrast, in JSBACHv3 the precipitation-
420 SM feedback seems to have no importance in this region. Instead the January warming effect of the TCC is more pronounced
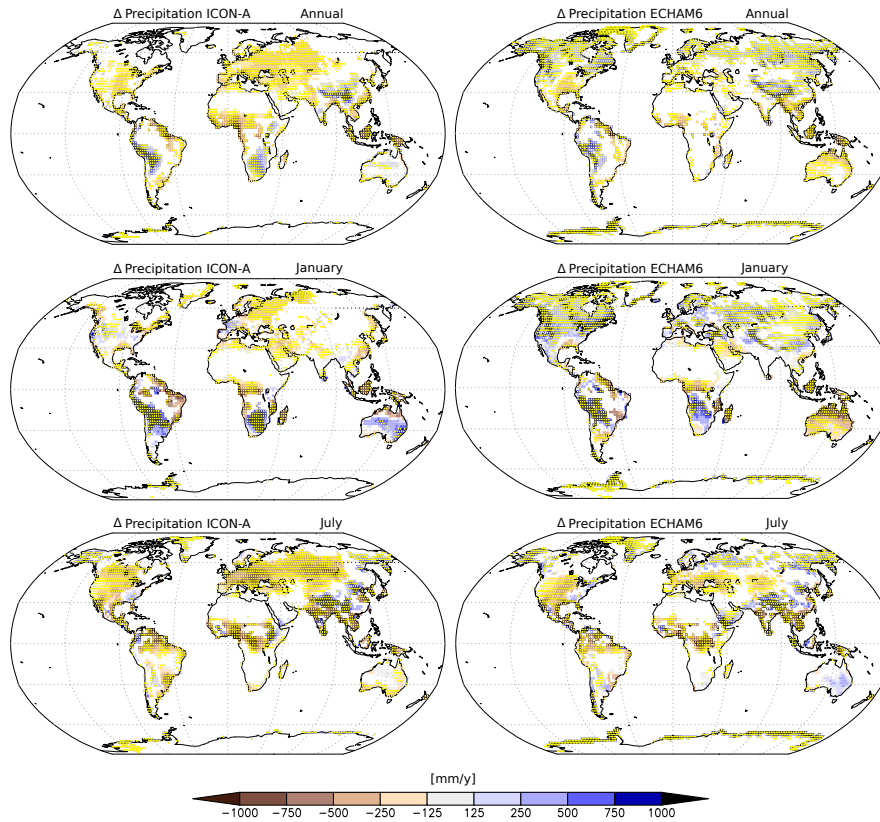
Figure 8. Bias in precipitation for JSBACHv4 (left) and JSBACHv3 (right) compared to GPCP precipitation. Shown are the annual, January and July means over the years 2001 to 2014. Note, land areas where the bias is smaller than the GPCP error bar are not plotted. The yellow dots indicate grid boxes where the bias is significant (p=0.05) according to an independent two-sample t-test.

in JSBACHv3 (see Table 4). However, for both models a higher TCC has a warming effect in January - as expected from the general rule that a cloud cover tends to warm the surface in winter (Chen et al., 2000).

The Antarctic and Greenland ice sheets tend to be warm biased. Only for JSBACHv4 in January Greenland shows no uniform pattern. Cold biases are seen in northern and southern subtropics, and are more expressed for JSBACHv4. These two
425 observations together – lower high latitudes warm bias, stronger subtropical cold bias – are together the reason for the globally lower temperature seen for JSBACHv4 in the plot of Fig. 5 (top). This is somewhat more pronounced for January which is visible in the zonal averages and in Fig. 5 (bottom).

The differences in bias patterns between the two model versions are also visible in the scatter plots of Fig. B3: The low July correlation (r=0.653) is caused by JSBACHv4 being warmer in the 'temperate zone' and colder in the 'polar zone' (for the
430 definition of the zones see Fig. B1). For January and also for the whole year, LSTs are much stronger correlated (r>0.84). This is also true when we regard the corresponding rank correlations of Spearman (Table 3).

Overall, for both models the bias pattern is very zonal, showing a warm bias at northern and southern higher latitudes, a cold bias in the northern and southern subtropics, and an insignificant bias in the equatorial regions (Fig. 6). The north-south symmetry of this zonal bias pattern hints to an atmospheric origin in the simulated climate since for biases caused by land processes one would not expect a north-south symmetry because of the very different continental distribution in the two hemispheres.

In summary we find that: (1) The global mean LSTs of JSBACH only fit the MODIS data because regional biases cancel each other. The global mean LSTs of JSBACHv4 fit better with the MODIS data because of its lower January values in the subtropics. (2) Because of the overall zonal bias pattern of both models a strong atmospheric contribution has to be assumed. (3) Both models exhibit a higher northern latitudes warm bias. In January, especially in JSBACHv3, the TCC partly causes this warming and both models show a strong amplification by the snow-albedo feedback. In July, but only in JSBACHv4, a feedback between precipitation and SM causes a warm bias. (4) Except of Greenland in January for JSBACHv4 the Antarctic and Greenland ice sheets tend to be warm biased.

Table 4. Spatial Spearman rank correlations (rho) between the JSBACHv4 and JSBACHv3 biases of TCC against those of LST and FAPAR.

	JSBACHv4		JSBACHv3	
	TCC vs. LST (Eurasia)	TCC vs. FAPAR	TCC vs. LST (Eurasia)	TCC vs. FAPAR
Annual	0.027	0.217	0.596	-0.016
January	0.141	0.218	0.593	0.134
July	-0.797	0.138	-0.329	-0.060

Eurasia refers here to the square between longitude 0° to 184° and latitude 35° to 90°.

3.4 Terrestrial Water Storage (TWS)

Fig. 9 shows the agreement between JSBACHv4 and GRACE in the alterations of TWS changes in the course of the year. The strongest differences occur in dry regions with small TWS changes. In these regions precipitation is very sensitive to the exact atmosphere dynamics and thus small shortcomings result in strong effects in Fig. 9. At the same time in regions where precipitation is not very sensitive to exact atmospheric conditions the model is in better agreement. For example, the inner tropics with strong deep convection (central Africa, central S-America, northern Australia), regions that are strongly affected by the monsoon (India, Africa) and regions that are dominated by the westerlies (Europe, and eastern Europe). Also north-eastern Asia reveals relatively good agreements despite the fact that the amount of TWS change is not high there. In contrast to this general rule, quite the opposite applies for the Indonesian islands. This is probably due to the resolution of the model which can not depict precipitation very exactly for small land areas surrounded by ocean. We find the same problem also for all extratropical islands.
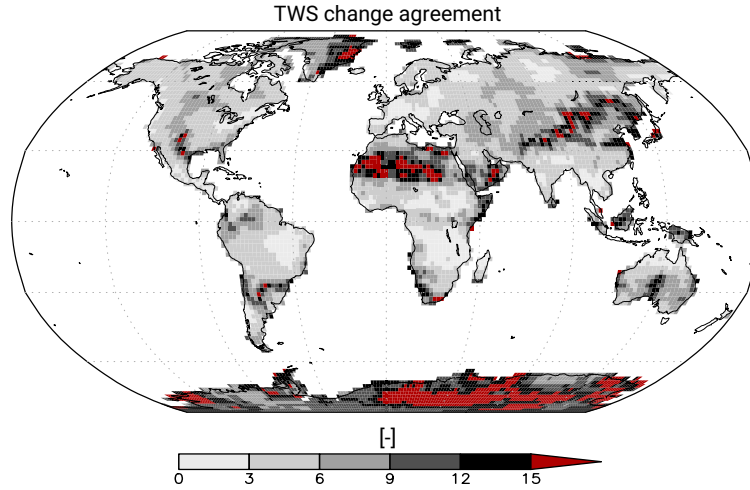
Figure 9. Normalized TWS change differences in the course of the year between the JSBACHv4 and GRACE data. Note that, due to the statistics, only the relative agreement between different regions can be rated but not an amount of water mass change agreement in absolute terms. However, the higher the value the stronger is the difference in the course of the year between JSBACHv4 and GRACE.

3.5 Leaf Area Index (LAI)

In Fig. 10 we show the bias in LAI for JSBACHv4 and JSBACHv3 compared to MODIS LAI data. The bias patterns between both models are strongly correlated (see Table 3). In the tropics the bias pattern is quite similar. Differences between the models are mainly seen in the northern extratropics, where in July the LAI is too large for JSBACHv4 across the high northern latitudes (Canada, Alaska, Norway) and for JSBACHv3 too large along the Eurasian steppe belt. For the latter region this may be related to the slightly longer growth period in JSBACHv3 because of the earlier rise of northern hemisphere spring temperatures (see Fig. 5 bottom). The wide spread and very similar significance pattern between the JSBACH versions reveals that JSBACH is systematically biased independent of the atmosphere model.

Overall, the bias is small in regions that are only sparsely covered with vegetation (e.g. northern Africa, Middle East, north-eastern Asia and central North-America) and in regions with strong bias the LAI is generally too large, as is also seen in the scatter plots of Fig. B5, where the center data cloud is at an LAI bias of about 1. Clearly visible, particularly for JSBACHv3, is a positive LAI bias in tropical and subtropical regions with high relative soil water levels (compare Fig. 11), which is insofar plausible as only for good growing conditions the LAI may overshoot. In other regions a spatial correlation between LAI bias and relative soil water level is not as clear. The general higher Relative Soil Moisture (RSM) in JSBACHv3 as compared to JSBACHv4 is obviously a consequence of its higher precipitation (Fig. 8).

For a more detailed analysis of the bias patterns one must note that JSBACH uses different parameterizations for its vegetation types to calculate the seasonal and multi-annual dynamics of the LAI, i.e. to describe their phenology. These parameter-
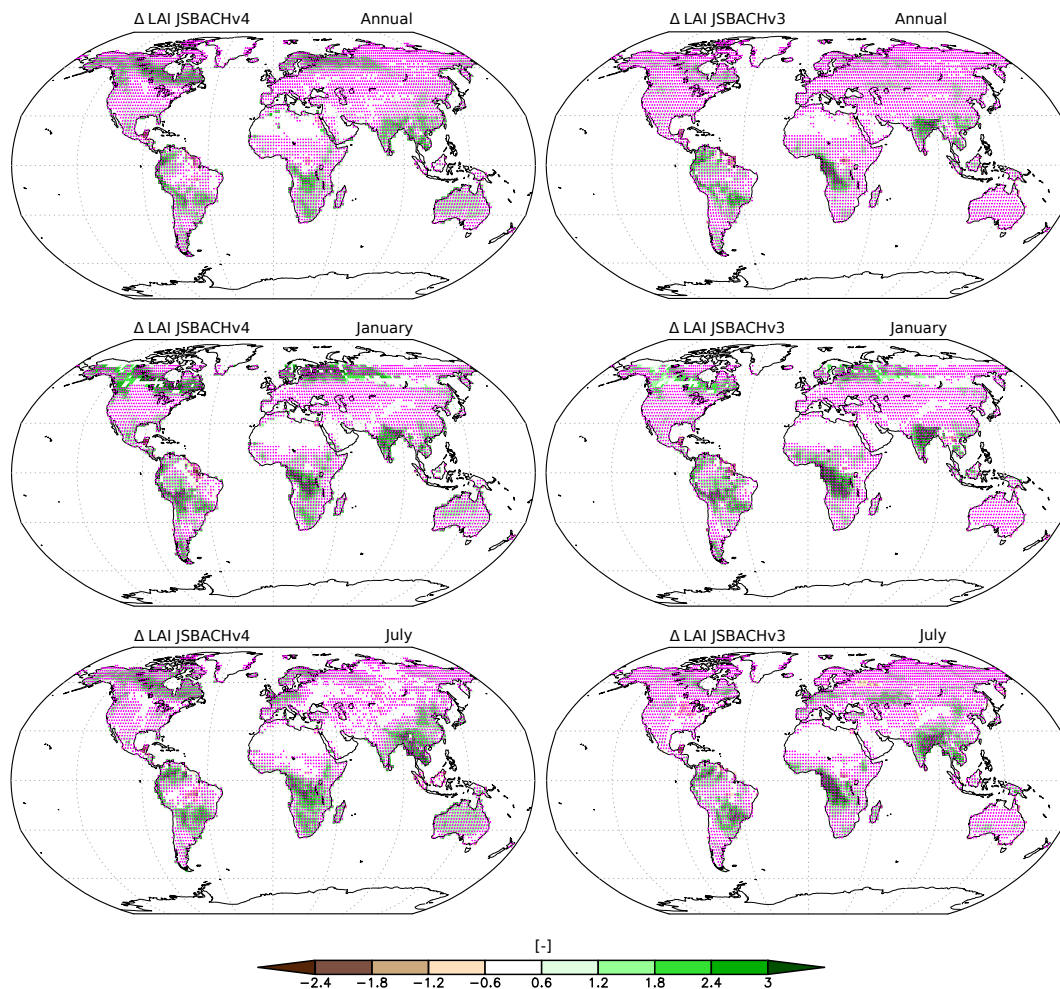
Figure 10. Bias in Leaf Area Index (LAI) compared to MODIS data for JSBACHv4 (left) and JSBACHv3 (right) averaged over the years 2001 to 2014. Shown are the annually averaged bias and those for January and July. Pink dots indicate grid boxes where the bias is significant (p=0.05) according to an independent two-sample t-test.

izations differ in particular by their dependence on growth conditions (temperature, relative soil water and NPP). To analyze the origin of the biases it is therefore necessary to consider each region with its prevailing vegetation types separately. Such analysis (not shown) reveals only weak correlations between growth conditions and the diagnosed LAI biases (except for the above named RSM). It seems unlikely that this is caused by several environmental conditions affecting the LAI dynamics simultaneously in a way that correlations cancel. More likely the LAI biases are caused by shortcomings in the parameterizations of the phenology, for example, by an unlucky choice for the combination of the many parameter values like spring growth rate for LAI and maximum length of growth season, or by more structural deficiencies like the missing coupling between LAI and leaf biomass in this type of phenology model.
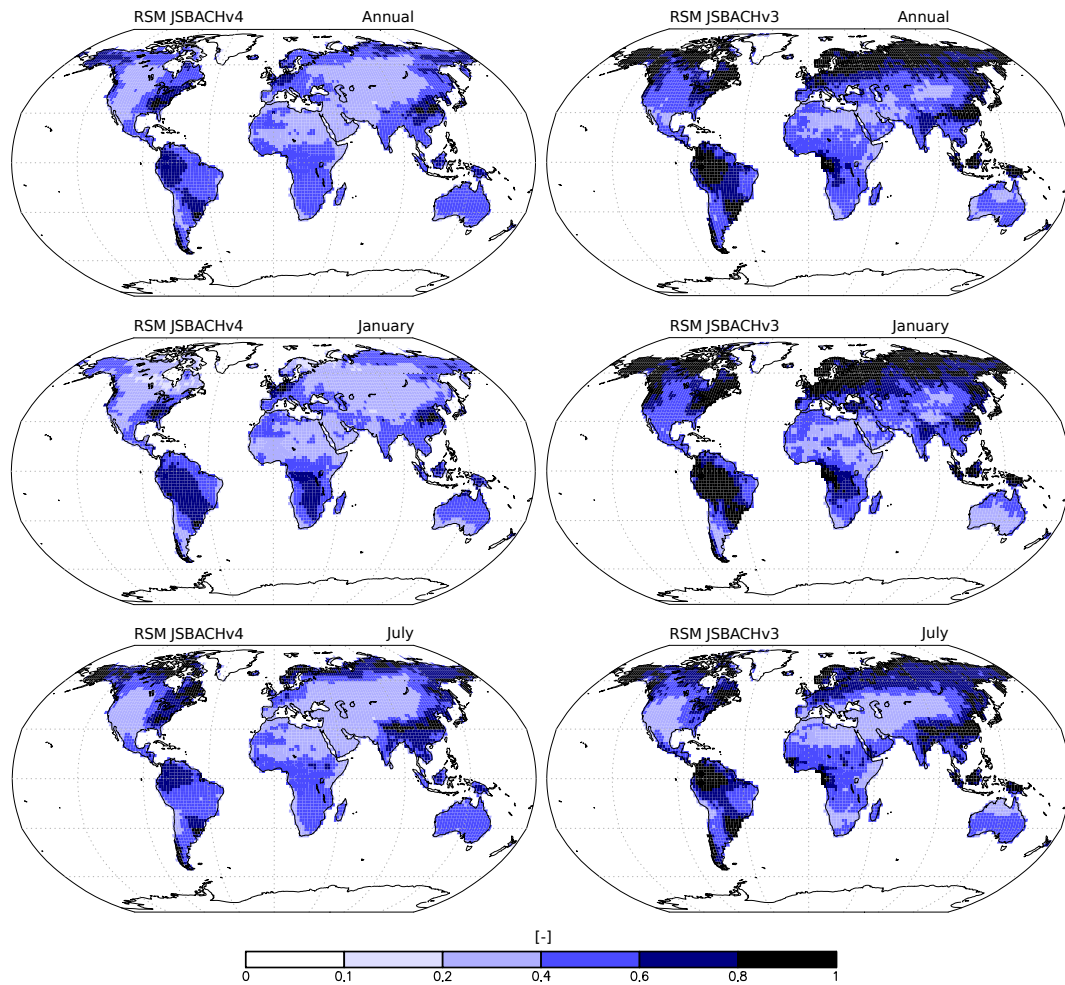
Figure 11. Relative Soil Moisture (RSM) in JSBACHv4 (left) and JSBACHv3 (right). Shown are the annual, January and July means over the years 2001 to 2014. RSM is calculated as the ratio between amount of soil water above wilting point and the amount of water between field capacity and wilting point. RSM thus characterizes the soil conditions concerning water stress of plants (the plant usable field capacity).

480   However, for the LAI bias of JSBACHv4 in Australia we can indeed identify environmental causes. Australia's canopy is especially in the areas with too high LAI dominated by shrubs and C4 pasture whose phenology is calculated by the parametrizations 'raingreen phenology' and 'grass phenology', respectively. The LAI of both phenology parametrizations depends on the relative soil water level and NPP of the model vegetation. The grass phenology additionally depends on air temperature of the lowest atmospheric level. As the surface temperature in this area is too low throughout the year (Fig. 6) this is most likely not

485   the cause for the positive LAI bias. In terms of simulated relative soil moisture (Fig. 11) Australia is not particularly dry. There is sufficient soil water available for plant growth in the annual average as well as for January and July (values range between 0.4 and 0.5). Particularly in July the dry season of central and northern Australia is not at all represented in the simulations.

As the model bias in NPP is rather weak for July (see Fig. 13) the overly wet soils should be the main reason for the too high LAI. For January, however, in reality the soil water does not limit plant growth as in July. Therefore, even though our models exhibits too much soil moisture in January, this might not cause the too high LAI. Hence, in this case the positive NPP bias could be the cause for the LAI bias and not vice versa.

In summary we find that: (1) The LAI bias pattern are strongly correlated between JSBACHv4 and JSBACHv3. (2) The LAI is in general high biased which occurs in regions with strong vegetation cover. (3) In Australia for JSBACHv4 the main reasons for the high biases might be the overly wet soils in July and the too high NPP in January. (4) Except for Australia, the LAI biases are probably caused by shortcomings in the parameterizations of the phenology themself and not by the growth conditions.

3.6 Fraction of Absorbed Photosynthetic Active Radiation (FAPAR)

Table 5. Spatial Spearman rank correlations (rho) between the JSBACHv4 and JSBACHv3 biases of LAI against those of FAPAR. In brackets the corresponding Pearson coefficients.

	JSBACHv4		JSBACHv3	
	LAI vs. FAPAR (Kazakhstan)	LAI vs. FAPAR	LAI vs. FAPAR (Kazakhstan)	LAI vs. FAPAR
Annual	0.784 (0.766)	0.640 (0.683)	0.831 (0.835)	0.784 (0.798)
January	0.324 (0.285)	0.505 (0.579)	0.413 (0.329)	0.780 (0.818)
July	0.925 (0.939)	0.732 (0.725)	0.879 (0.888)	0.539 (0.632)

Kazakhstan refers here to the square between longitude 46° to 87° and latitude 40° to 54°N.

In general FAPAR biases are strongly correlated with the LAI biases (see Tab 5). For JSBACHv3 the FAPAR bias pattern follows in the tropics and in regions of austral and boreal summer largely the bias pattern of LAI (compare Figs. 10 and 12), as can be expected from FAPAR being typically larger for deeper canopies. For JSBACHv4 this correlation is weaker but still strong in the yearly average (see Tab 5). Other possible source of FAPAR deviations are clouds because they might alter the incoming PAR or the fraction between its direct and diffuse parts. However, the comparison with the bias in Total Cloud Cover (TCC) (see Fig. 7 and Table 4) reveals no general hints for that. FAPAR deviations could also be caused by the VIS soil albedo which is used in the JSBACH canopy radiation model as lower boundary condition. (The biases in the JSBACH soil albedo maps were already discussed above.)

The deviations in FAPAR of the two JSBACH versions (Fig. 12) clearly differ from each other, as also reflected in the medium correlation values around 0.4 (see Fig. B6). The main differences occur over Kazakhstan, India, south China and central Africa. The difference over Kazakhstan (around 50°N) is obviously caused by the above described precipitation-SM feedback in July. Differences in India, south China and central Africa occur because the high-biased LAI in JSBACHv4 does not cause the same FAPAR high-bias in these regions as in JSBACHv3. As the VIS soil albedo is the same in both model
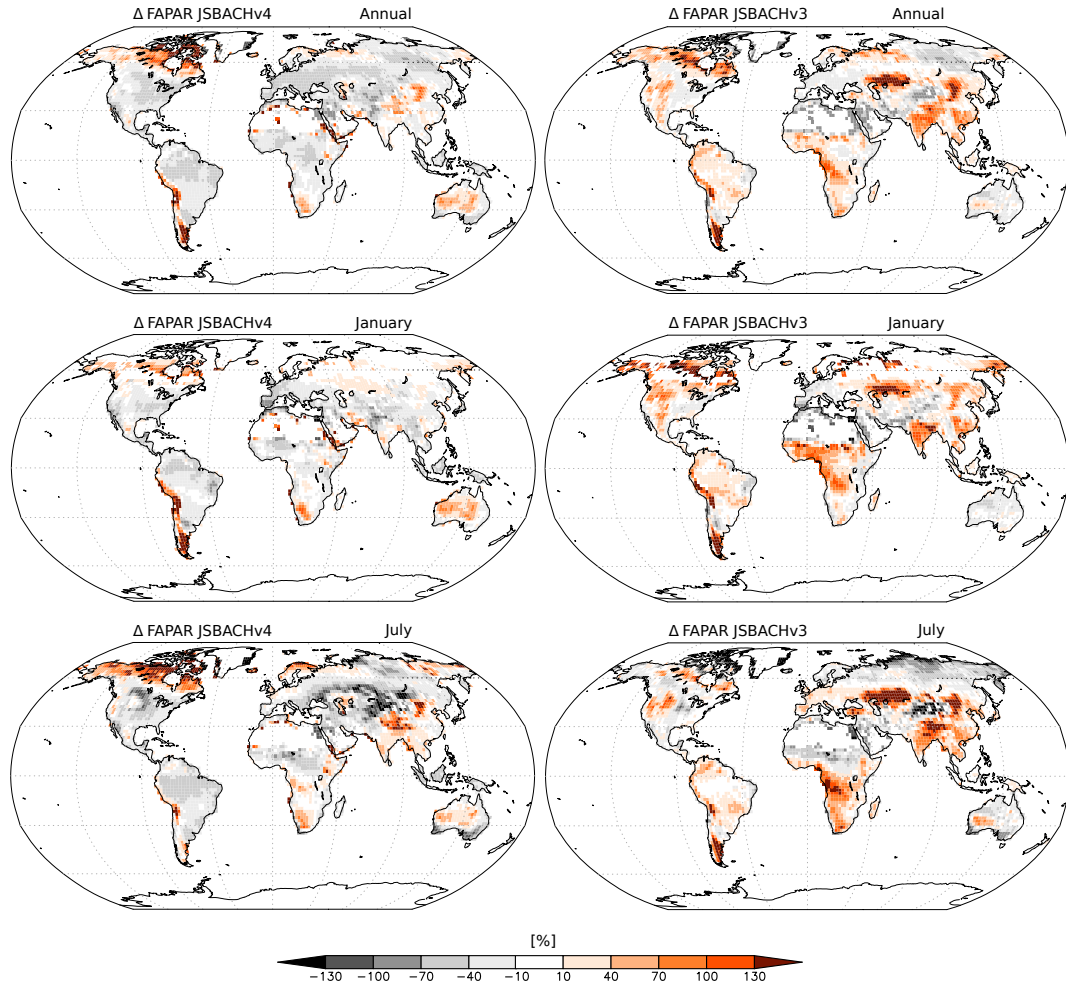
Figure 12. Bias in the Fraction of Absorbed Photosynthetic Active Radiation (FAPAR) of JSBACHv4 (left) and JSBACHv3 (right) compared to MODIS FAPAR. Shown are the annual, January and July means over the years 2001 to 2014 in percent to MODIS FAPAR.

versions this can not be the explanation. Unfortunately our simulation outputs do not provide the necessary information about incoming PAR to verify an atmospheric origin.

3.7 Net Primary Production (NPP)

Except for the northern mid latitudes in northern hemisphere summer, the patterns in NPP biases are very similar between the two JSBACH versions (Fig. 13 and Fig. B7). This is also demonstrated by the rather high correlations in Table 3 for Januaries (∼0.8) and throughout the year (∼0.75) while there are lower correlation values for the Julys (∼0.5). In general NPP shows strong correlations with Relative Soil Moisture (RSM) (Table 6). Only for JSBACHv3 in January they are nearly uncorrelated

515

Table 6. Spatial Spearman rank correlatations (rho) between the JSBACHv4 and JSBACHv3 biases of NPP against those of LAI, FAPAR and RSM.

	JSBACHv4			JSBACHv3		
	NPP vs. LAI	NPP vs. FAPAR	NPP vs. RSM	NPP vs. LAI	NPP vs. FAPAR	NPP vs. RSM
Annual	0.197	0.270	0.823	0.397	0.550	0.739
January	0.279	0.348	0.573	0.506	0.553	0.026
July	0.400	0.362	0.789	0.435	0.366	0.793
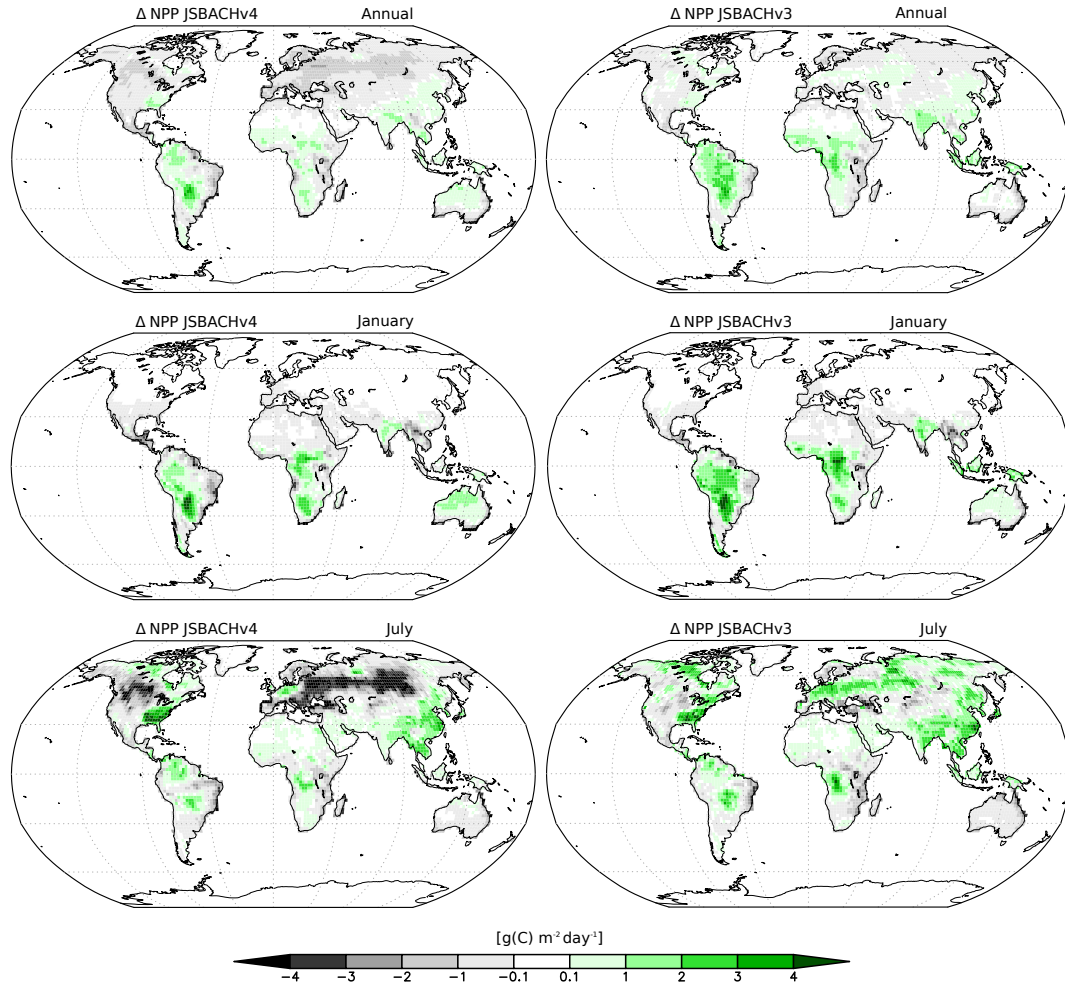
Figure 13. Bias in Net Primary Production (NPP) for JSBACHv4 (left) and JSBACHv3 (right) compared to MODIS NPP. Shown are the annual, January and July means over the years 2001 to 2014. Note that significance is not shown because except for the Sahara region all biases are worldwide significant (p=0.05) according to an independent two-sample t-test.

which is due to the fact that north of 60° there is no NPP at that time but JSBACHv3 has a lot of RSM. However, in general the bias correlation between NPP and LAI is much weaker.

The shift in NPP bias from JSBACHv3 to JSBACHv4 seen in northern mid latitudes in July has several reasons. For JS-BACHv4 it is obvious that too low precipitation and FAPAR lead to the strong negative NPP bias (compare with Fig. 8 and Fig. 12). This bias is a side effect of the precipitation-SM feedback already explained above (similar to the LST bias for this region and time). For JSBACHv3 the bias is caused by a too high LAI leading to a too high FAPAR (compare with Fig. 10 and Fig. 12).

That the NPP bias is vanishing in winter for regions with seasonal vegetation is trivial, so that only the patterns in summer need explanation. For both JSBACHv4 and JSBACHv3 in the tropics and in the regions of boreal and austral summer the bias pattern in NPP largely follows that of LAI (compare Figs. 10 and 13), which is plausible in view of the already diagnosed positive LAI bias. The correlations between the NPP and LAI biases are not high (Table 6) as there are no correlations for the respective winter part of the globe (compare Figs. 10 and 13). The same accounts for the correlations between the NPP and FAPAR biases although primary production directly depends on FAPAR.

In summary we find that: (1) The NPP bias pattern highly agree between JSBACHv4 and JSBACHv3 and the strongest mismatch is caused by the July precipitation-SM feedback of JSBACHv4. (2) The NPP deviations are strongly correlated with RSM.

3.8 Water-Use-Efficiency (WUE)

To discuss WUE as simulated by JSBACHv3 and JSBACHv4, we show simulation results in two ways: as geographic distribution (Fig. 14), and in climate space (Fig. 15). In both plots colors are chosen to be consistent with the color coding of the respective Figs. 1 and 2 of Sun et al. (2016) showing WUE obtained from observation data and also from simulations with other land models.

Looking first at the geographical distribution (Fig. 14), WUE turns out to be worldwide slightly weaker in JSBACHv4 than in JSBACHv3, but otherwise the pattern is similar. A major difference between the model versions is the rather low WUE of JSBACHv4 found in the agricultural belt along the Eurasian steppe regions, where in the models vegetation consists of crops. (Note, there is no correlation with the additional C4 crops PFT of JSBACHv3.) Causal for this is the lower GPP of JSBACHv4 in this region (not shown, but also visible somewhat less pronounced in NPP Fig. 13).

The large scale pattern is best characterized by noting that JSBACH reveals low values for WUE in western North America, north-eastern Canada and in the broad belt ranging from the Sahara deep into inner Eurasia; elsewhere WUE is considerably larger. The reason for this pattern is obvious from Fig. 11, showing the relative soil water content above wilting point. RSM is therefore a measure of the water stress experienced by the vegetation in JSBACH: regions of high water stress broadly match the regions of low WUE. Accordingly, it is the low plant productivity that determines the regions of low WUE. Nevertheless, in the hot deserts – that make up large parts of the regions of low WUE – high potential evaporations surely adds to WUE being small there.
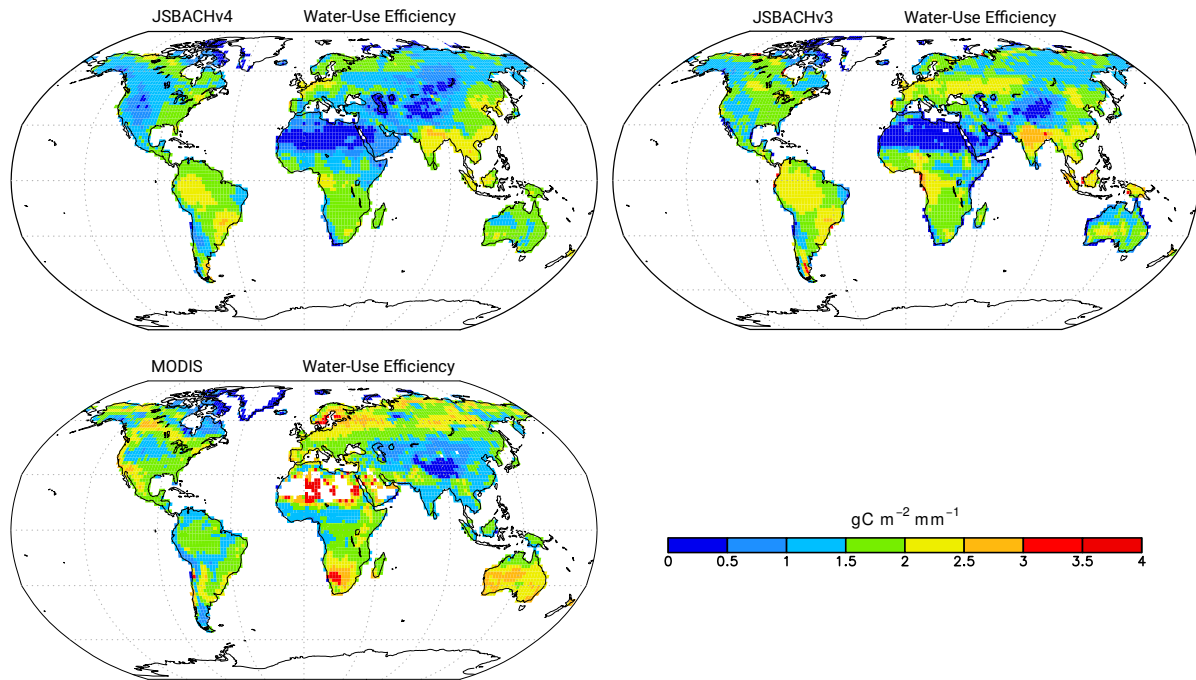
Figure 14. Water-Use-Efficiency (WUE). JSBACHv4 (top left), JSBACHv3 (top right) and MODIS (bottom left). All values are averaged over the years 2001 to 2014. The color scale is chosen to match the one used in Figs. 1 and 2 of the study by Sun et al. (2016).

This large scale pattern is also broadly consistent with the MODIS WUE (Fig. 14 bottom left). In contrast to JSBACH MODIS exhibits the highest global values over the dry areas in northern and southern Africa and in Australia. Sun et al. (2016) calculated the WUE from MODIS GPP and JUNG GPP (their Figs. 1a and 1b, the invoked evapotranspiration data are the same). The WUE calculated from their MODIS dataset is unsurprisingly very similar to our MODIS WUE and shows the same

555 high values in these dry areas. However, the JUNG data (Sun et al. (2016) Figs. 1b) show rather low values for these regions and are therefore in much better agreement with JSBACH. A general disagreement between all observation based datasets and JSBACH occurs in India: The too high WUE simulated for India is caused by the PFT 'C3 crops': In this region C3 types ('C3 grass', 'C3 pasture' and 'C3 crops') are the most productive PFTs and among these 'C3 crops' are in terms of area the most dominant PFT.

560 Regarding the distribution of our WUE in 'climate space' spanned by temperature and precipitation (Fig. 15) the results of JSBACHv3 are very similar to JSBACHv4. ECHAM6 (JSBACHv3) shows much higher precipitation in the temperature range around -15°C and 15°C. As a consequence JSBACHv4 fits much better in this range than JSBACHv3 when compared to our MODIS data or to Sun et al. (2016) Fig. 2 (a) and (b). The above mentioned much higher global values over dry areas in MODIS as compared to JSBACH are also visible here. MODIS has its highest values in the high temperature low precipitation
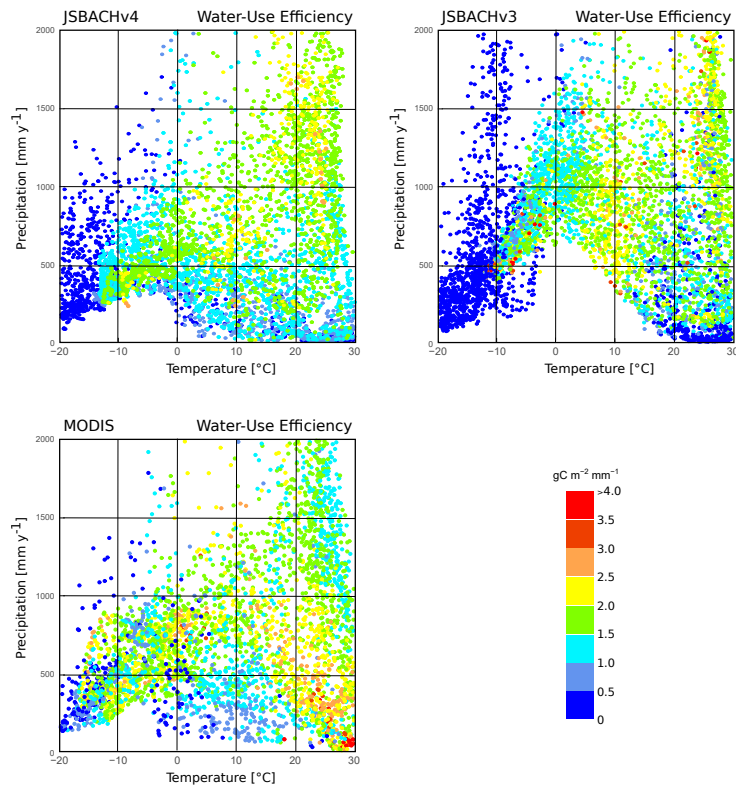
Figure 15. Water-Use-Efficiency (WUE). JSBACHv4 (top left), JSBACHv3 (top right) and MODIS (bottom left) for each grid cell plotted against annual 2m-temperature and precipitation. All values are averaged over the years 2001 to 2014. Color scale as in Fig. 14.

565 corner which is the opposite in both JSBACH versions. The JSBACH Indian high bias is visible at 25°C and precipitation above 1000 mm/y.

4   Summary and conclusions

General Performance

The Taylor diagram shown in Fig. 1 and the comparison of the NME for our assessment variables (see appendix A1 and A2)
570 reveal that the overall performance of JSBACHv4 remains more or less as in JSBACHv3. In the context of its totally new infrastructure and the new ICON-A atmosphere this is a good achievement and lays the groundwork for a further development of JSBACH.

Albedo

575 As compared to the JSBACHv3 results JSBACHv4 shows very similar biases (and significances). Overall the representation of the albedo in JSBACHv4 is as good as for JSBACHv3.

Nearly all albedo deviations of JSBACH are significant. In Antarctica and Greenland the albedo deviations are uniformly too high for NIR and uniformly too low for VIS. Therefore, these deviations can be mitigated without side effects within the JSBACH calculations of the glacier albedo. For glaciers albedo values are used that vary as function of temperature continuously

580 between fixed minimum and maximum values. Adapting these minimum and maximum values alone should already reduce the bias and tuning the temperature dependency could further improve the results. In the northern mid latitudes the January albedo of JSBACH is low biased and is largely caused by a too small snow cover. Most probably the already existing initial warm bias of ICON-A in these areas (see LST) is enhanced by the snow-albedo feedback. Tuning the snow scheme in JSBACH towards more snow might compensate the warm bias of ICON-A. However, there could be undesired side effects, such as a cold bias

585 in eastern North-America, a delay of the spring point or too little soil water.

The JSBACH albedo bias in other regions are at least partly caused by canopy or soil albedo biases. Another source could be that the scheme merging them to the overall albedo causes additional deviations. For example, wrong fractions of bare ground below vegetation that is "visible" for the sun or wrong fraction of soil below the vegetation which is covered by stems but not by the LAI. However, we found that the JSBACH albedo bias in these regions can be largely explained by the canopy and soil

590 albedo of the model. In the present model configuration the values for canopy are calculated from PFT specific leaf albedo values, which themselves do not account for seasonality. The NIR and VIS soil albedos in the present model configuration are taken from prescribed maps, which also do not account for seasonality. The albedo values for the leaves and for the soil are derived by linear regression of the FAPAR on total surface albedo (described in Otto et al., 2011). Thus, both the canopy and soil albedo values within JSBACH are based on the 2001 to 2004 average of the MODIS MOD43C1 CMG Collection 4

595 Albedo Product. We showed that the differences between this product and the January/July values of the MODIS product used for our assessment already explain large parts of the JSBACHv4 albedo deviations. Therefore, as a first mitigation step the MODIS database for JSBACH should be updated and a seasonality for leaves and soil should be introduced in the JSBACH albedo scheme. Only after that it would make sense to investigate improvements of the scheme that merges snow, canopy and soil albedo to the overall albedo.

600

Land Surface Temperatures

Overall, JSBACHv3 and JSBACHv4 show very similar temperature biases. The global mean LST simulated by JSBACH match well with observational data, especially for JSBACHv4 (see Fig. 5 top). However, this is only because strong regional biases mutually cancel each other. The overall zonal bias pattern of both models hint to a strong atmospheric bias contribution. We

605 find for both models a large scale warm bias throughout the year over central Asia which is also significant in the zonal averages. One reason for this is the warming effect of the TCC in January, which is enforced by a too small snow cover and the snow-albedo feedback in this region. For JSBACHv4 an additional feedback between low precipitation and low RSM causes a strong warm bias in July. It is almost certain that this feedback is caused by the combination of the 5-layer snow and soil

water freezing schemes implemented in JSBACHv4. The effect of this scheme on the JSBACH results is analysed in detail by
Hagemann et al. (2016) and is in very good agreement with our results. For the predecessor model ECHAM5 similar northern
hemisphere winter and summer warm biases as in our results are already shown in Ozturk et al. (2012) while the January and
July biases shown in Piani et al. (2010) agree only partly (Kazakhstan and northern China). We also find too high temperatures
over the Antarctic and Greenland ice sheets and cold-biased regions in S-America, northern- and southern Africa, Australia,
India and China (all significant). These biases are at least partly caused by albedo deviations. As the central Asian, Antarctic
and Greenland warm biases are directly connected to albedo deviations they might be mitigated by the model adaptations proposed above for the albedo.

Terrestrial Water Storage

We analysis the agreement of alterations of TWS changes in the course of the year between JSBACHv4 and GRACE. We
found that the agreement is best for regions where precipitation is not very sensitive to the exact representation of atmospheric
conditions in the model (i.e. regions where precipitation underlies dominant atmospheric processes like deep convection). An
implication of this result is that the representation of TWS changes in the course of the year are less dependent on soil properties like field capacity or porosity than on the atmospheric conditions. Thus, from this point of view the enhancement of soil
properties is not a main focus for the development of JSBACHv4.

Leaf Area Index

The patterns in LAI biases are very similar between both model versions. In general the LAI is significantly too high in JS-
BACH. This is not unique for our model (see Seiler et al., 2022). For JSBACHv4 in Australia the LAI deviations for July is
probably caused by too much soil moisture and for January by too high NPP. Therefore, for July, when our model does not
represent a dry season in Australia, the LAI bias would be fixed by improving relative soil moisture. For January the LAI bias
might be reduced with a mitigation of the NPP bias. Except for JSBACHv4 in Australia and the general too high RSM the LAI
biases and the corresponding growth conditions are correlated only weakly. However, as our growth conditions (temperature,
relative soil water and NPP) exhibit strong shortcomings a first step is to enhance them. As a second step one could reduce the
LAI by adapting the parameters for the phenology growth and death rates. As the LAI has strong influence on other processes
(sensible and latent heatflux, GPP and dark respiration, canopy albedo) it would be at the same time necessary to adapt these
processes to the lower LAI. Nabel et al. (2020) show that the introduction of forest age classes in JSBACHv4 also reduces the
LAI bias. However, this comes with additional computational costs for the model.

Fraction of Absorbed Photosynthetic Active Radiation and Net Primary Production

We find for both model versions that FAPAR deviation are mainly caused by LAI biases. Overall all NPP deviations in JS-
BACHv4 can be associated with water stress problems, FAPAR and LAI. The high-biased NPP would be bettered nearly
everywhere with the above named reduction of the LAI. With this reduction also the FAPAR biases would be reduced and contribute to an improved NPP (e.g. in southern Africa and India). Our NPP biases are strongly correlated with RSM respectively

with water stress. Higher water stress in S- and eastern N-America and central Africa would further reduce our NPP deviations.
However, to avoid the strong JSBACH4 low-bias in July for the mid latitudes it would be necessary to enhance the ICON-A precipitation in this area which is a consequence of the above named precipitation-SM feedback.

Water-Use-Efficiency

Overall the WUE of our models looks very reasonable. In the Sahel Zone JSBACH WUE seems to be too low as compared to MODIS data, but it is possible that MODIS is too high here, especially as neither the JUNG data nor the models used in Sun et al. (2016) show the same high bias. In India our WUE is too high. One reason for this is the strong assimilation from C3 crops. This is also visible in the NPP high-bias for India. However, a high-biased productivity of C3 crops can not be found somewhere else. Another reason might be that JSBACH does not incorporate irrigation in our simulations, which is substantial in India and enhances evapotranspiration. As a result WUE in JSBACH should be higher than observed. For JSBACHv3 an irrigation scheme was already implemented (de Vrese and Stacke, 2020) and could be transfered to JSBACHv4. JSBACHv4 WUE is low-biased in central Asia which is probably caused by our low-biased FAPAR. WUE in JSBACHv4 is generally somewhat weaker than in JSBACHv3 but the geographical patterns are similar. The overall temperature and precipitation range of ICON-A+JSBACHv4 and ECHAM6+JSBACHv3 is similar to observational data, but especially ECHAM6+JSBACHv3 reveals too high precipitation in the temperature range -15° to 15°C. Therefore, JSBACHv4 performs better than JSBACHv3 when compared to observations. With a general reduction of the LAI as proposed above and the associated reduction of GPP the WUE of JSBACH might get low biased.

Perspectives of JSBACHv4

Our comparison with JSBACHv3 shows that the performance of JSBACHv4 is overall quite similar to JSBACHv3. Differences can partly be attributed to the different atmospheric conditions generated from and in interaction with the atmospheric host models ICON-A and ECHAM6, e.g. differences in WUE could be traced back to differences in precipitation. In terms of process descriptions, JSBACHv3 and JSBACHv4 are almost identical. The only notable difference in simulation results arising from such structural differences concerns regional differences in surface temperature via modified feedbacks with the atmosphere that we attribute to the implementation of a multi-layer snow model and freezing soil water in JSBACHv4. A few options to improve JSBACHv4 have been outlined in this paper. It should be noted that the versions of ICON-A and ICON-Land/JSBACHv4 used for this study constitute the first milestone in the development of the new ICON Earth System Model ICON-ESM. The flexibility of the ICON-Land infrastructure provides the basis for further developments and improvements as the land component of ICON-ESM. The model configuration of the AMIP experiments described in this paper only use the physical and biogeophysical components of JSBACHv4. The carbon cycle and dynamic vegetation processes of JSBACHv3 have also already been ported to JSBACHv4 and can already be used to model the fully closed carbon cycle within ICON-ESM together with the ocean biogeochemistry module. The evaluation of these processes will be addressed in future work.

Code and data availability. The source code, scripts, and necessary forcing and boundary data of ICON-ESM-V1.0 used in this study for AMIP simulations with ICON-A is downloadable from https://doi.org/10.35089/WDCC/RUBY-0_ICON-_ESM_V1.0_Model (click on "Find data"). Note that by downloading the included license agreement (see https://mpimet.mpg.de/en/science/modeling-with-icon/code-availability) is accepted. The primary data and scripts used to produce the figures are downloadable from https://cera-www.dkrz.de/WDCC/ui/cerasearch/entry?acronym=DKRZ_LTA_060_ds00008. The data from the MPI-ESM1.2-HR AMIP simulations are available at the CMIP6 repository of the Earth System Grid Federation (Lorenz et al., 2021) and can be accessed via https://doi.org/10.22033/ESGF/CMIP6.6463.

680

Appendix A

685 This Appendix provides additional tables with statistical information.

Table A1. Normalized Mean Error (NME) of JSBACHv4 relative to observations.

	Albedo VIS	Albedo NIR	LST	LAI	FAPAR	NPP	WUE
Annual	-	-	0.138	1.294	0.796	0.783	1.341
January	0.267	0.410	0.149	1.266	0.604	0.569	-
July	0.411	0.508	0.193	1.102	0.757	0.777	-

Table A2. Normalized Mean Error (NME) of JSBACHv3 relative to observations.

	Albedo VIS	Albedo NIR	LST	LAI	FAPAR	NPP	WUE
Annual	-	-	0.147	1.430	0.669	0.658	1.406
January	0.232	0.415	0.164	1.304	0.687	0.576	-
July	0.400	0.452	0.176	1.310	0.688	0.653	-

Table A3. Spatial Spearman rank correlations (rho) between the JSBACHv4 and JSBACHv3 biases of precipitation against those of LST.

	JSBACHv4		JSBACHv3	
	Preci. vs. LST (Eurasia)	Preci. vs. LST	Preci. vs. LST (Eurasia)	Preci. vs. LST
Annual	-0.416	-0.543	0.283	0.262
January	-0.090	-0.403	0.278	0.312
July	-0.638	-0.413	-0.169	-0.031

Eurasia refers here to the square between longitude 0° to 184° and latitude 35° to 90°.

Appendix B

For direct comparison of the simulation bias of the two model versions we plot in this appendix the difference to observations of the various assessment variables for each grid cell of one model against that of the other (see e.g. the scatter plot B3). The statistical values noted in these plots are the intercept and slope of the linear regression, the Pearson correlation r, its square

690   R^2 (also called 'coefficient of determination'), and the Spearman rank correlation rho. In addition, to obtain a rough idea how differences in simulation bias varies with region, we colorize the data points according to their broad geographical origin. For this we distinguish four regions (Fig. B1):

Polar Zone (black): 90° to 60° and -90° to -60°

Temperate Zone (red): 59,9° to 37.9° and -59,9° to -37.9°

695   *Subtropical Zone* (yellow): 38° to 15° and -38° to -15°

Tropical Zone (green): 14,9° to -14.9°

Statistics Δ Albedo

● Polar Zone ● Temperate Zone ● Subtropical Zone ● Tropical Zone

VIS   **January**

Intercept = -0.005
Gradient = 0.563
R^2 = 0.475
r   = 0.690
rho   = 0.754

NIR   **January**

Intercept = 0.007
Gradient = 0.708
R^2 = 0.481
r   = 0.694
rho   = 0.756

VIS   **July**

Intercept = 0.009
Gradient = 0.618
R^2 = 0.370
r   = 0.608
rho   = 0.844

NIR   **July**

Intercept = 0.007
Gradient = 0.714
R^2 = 0.623
r   = 0.789
rho   = 0.829
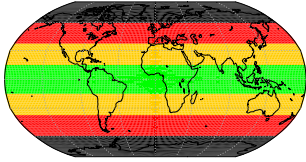
Figure B1. Zones of the scatterplots.
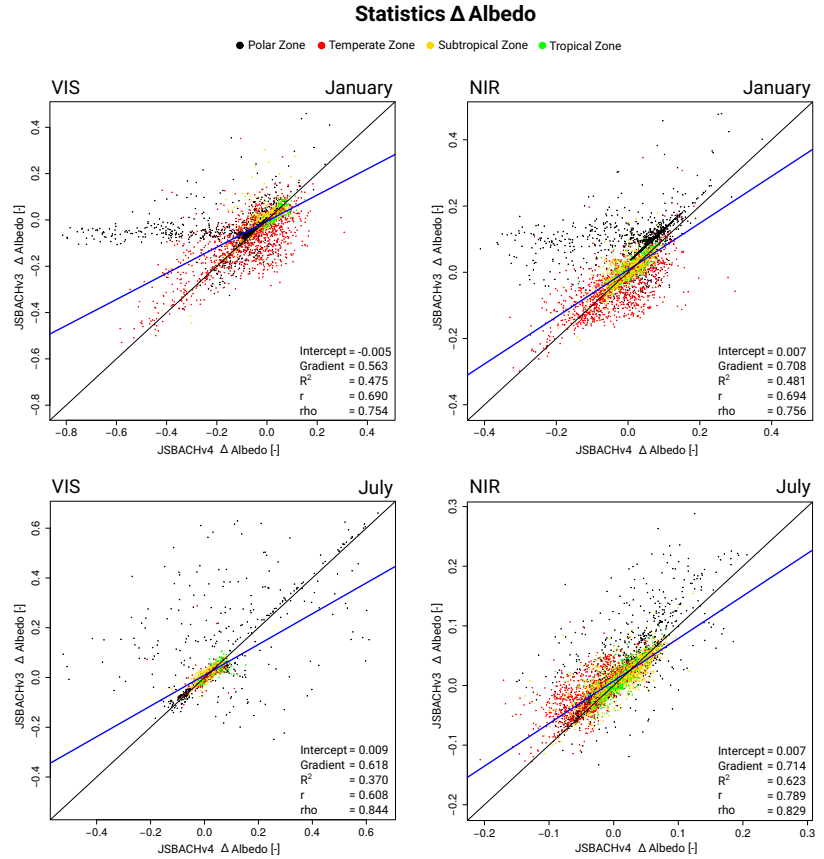
Figure B2. Comparison of simulated albedo bias (Δ Albedo) between the two JSBACH versions. Plotted is for each grid cell the albedo bias of JSBACHv3 against that of JSBACHv4. The bias data are the same as in Fig. 2. Shown are the averages of Januaries (top), and Julys (bottom) for VIS (left) and NIR (right) albedo. For more details see section 2.4.
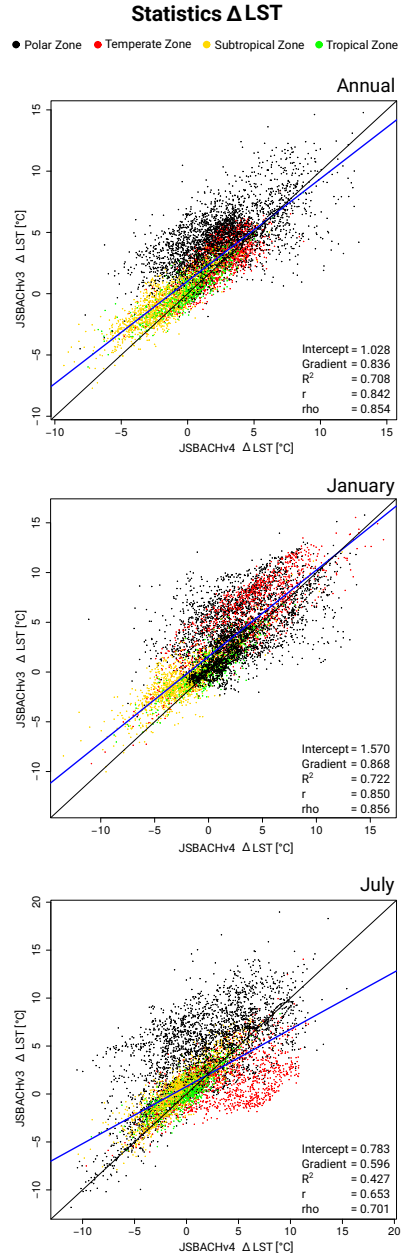
Figure B3. Comparison of simulated LST bias (ΔLST) between the two JSBACH versions. Plotted is for each grid cell the LST bias of JSBACHv3 against that of JSBACHv4. The bias data are the same as in Fig. 6. Shown are the annual averages (top), average of Januaries (mid), and Julys (bottom). For more details see section 2.4.
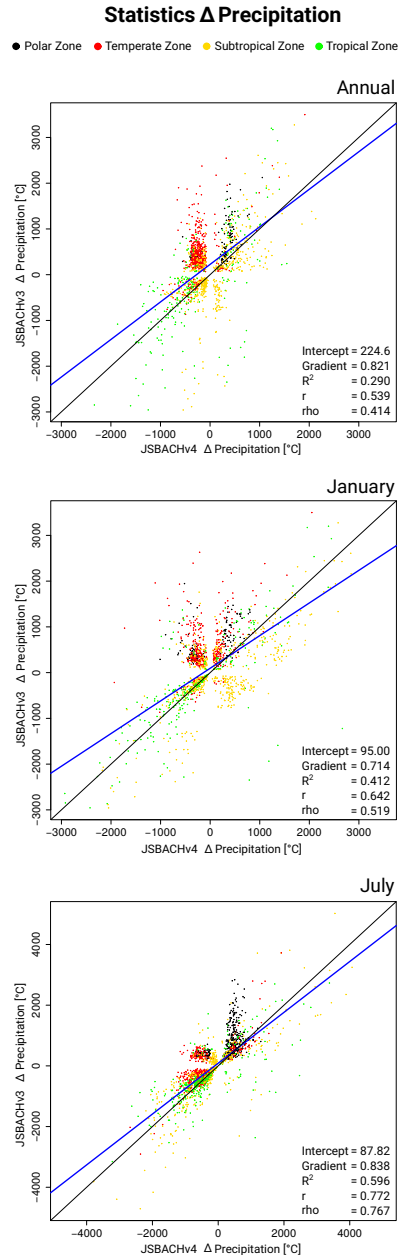
Figure B4. Comparison of simulated precipitation bias (Δ Precipitation) between the two JSBACH versions. For each grid cell the precipitation bias of JSBACHv3 is plotted against that of JSBACHv4. The bias data are the same as in Fig. 8. Shown are the annual averages (top), averages of Januaries (mid), and Julys (bottom). For more details see section 2.4.
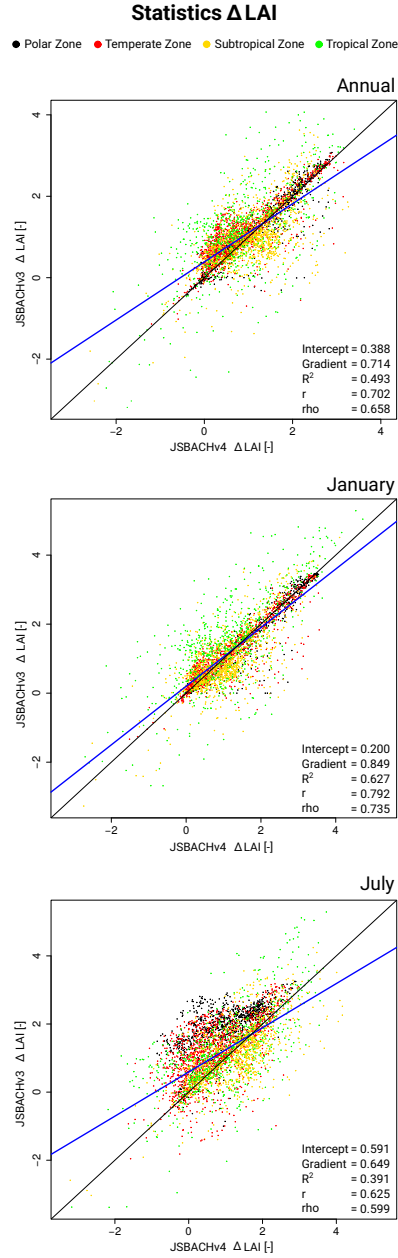
Figure B5. Comparison of simulated LAI bias (Δ LAI) between the two JSBACH versions. For each grid cell the LAI bias of JSBACHv3 is plotted against that of JSBACHv4. The bias data are the same as in Fig. 10. Shown are the annual averages (top), averages of Januaries (mid), and Julys (bottom). For more details see section 2.4.
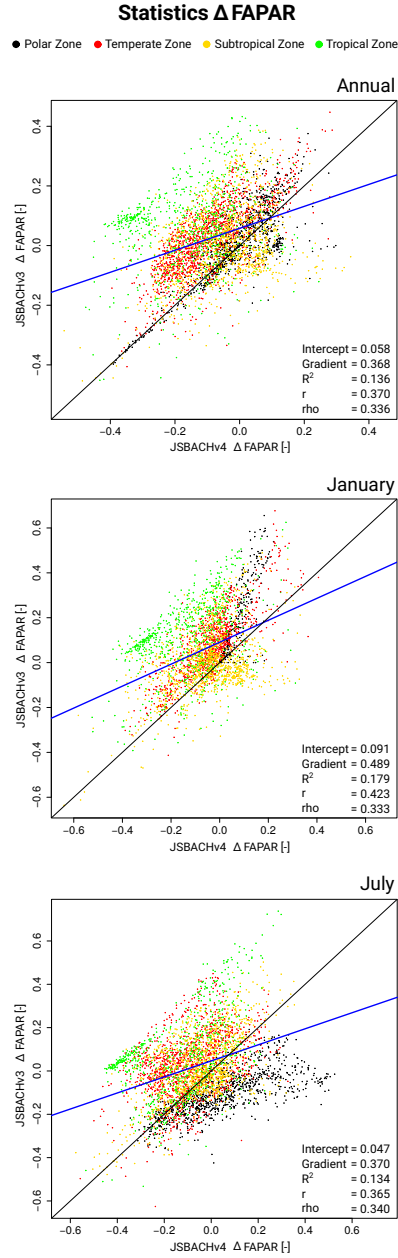
Figure B6. Comparison of simulated FAPAR bias (Δ FAPAR) between the two JSBACH versions. For each grid cell the FAPAR bias of JSBACHv3 is plotted against that of JSBACHv4. The bias data are the same as in Fig. 12. Shown are the annual averages (top), averages of Januaries (mid), and Julys (bottom). For more details see section 2.4.
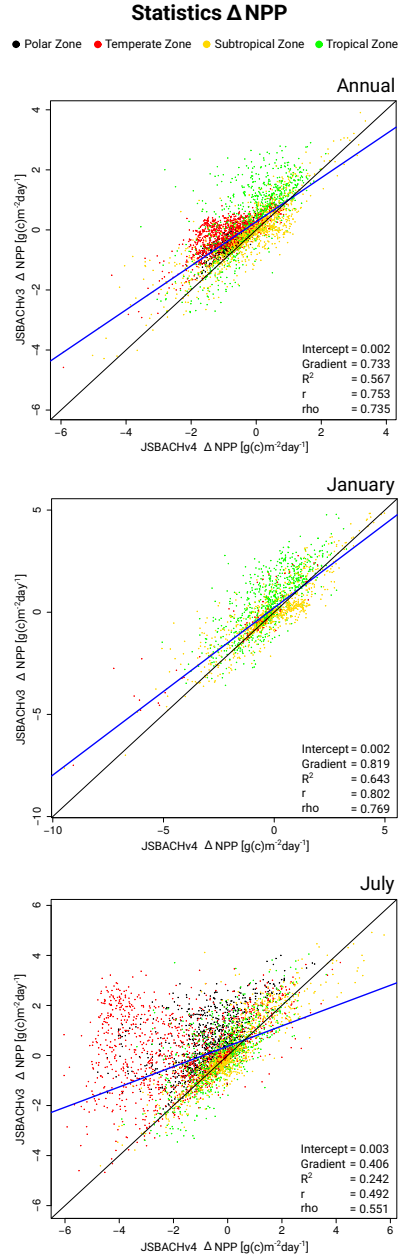
Figure B7. Comparison of simulated NPP bias (Δ NPP) between the two JSBACH versions. For each grid cell the NPP bias of JSBACHv3 is plotted against that of JSBACHv4. The bias data are the same as in Fig. 13. Shown are the annual averages (top), averages of Januaries (mid), and Julys (bottom). For more details see section 2.4.

Author contributions. R. Schneck performed the simulations, and prepared the data and plots. All authors contributed to the design, data interpretation and text of the study.

Competing interests. There are no competing interests.

700 *Acknowledgements.* We thank Dr. Stefan Kern for his advice concerning the handling of satellite remote sensing data and the German Climate Computation Center (DKRZ) for providing IT resources and support. Furthermore, we like to thank Dr. Jürgen Bader for his internal review of our manuscript and his constructive comments.

References

Adler, R., Huffman, G., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., and Arkin, P.: The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present), J. Hydrometeor, 4, 1147–1167, 2003.

Beer, C., Ciais, P., Reichstein, M., Baldocchi, D., Law, B., Papale, D., Soussana, J.-F., Ammann, C., Buchmann, N., Frank, D., et al.: Temporal and among-site variability of inherent water use efficiency at the ecosystem level, Global Biogeochemical Cycles, 23, 1–13, https://doi.org/10.1029/2008GB003233, 2009.

Betts, R., Cox, P., Lee, S., and Woodward, F.: Contrasting physiological and structural vegetation feedback in climate change simulations, Nature, 387, 796–799, 1997.

Böttcher, K., Markkanen, T., Thum, T., Aalto, T., Aurela, M., Reick, C. H., Kolari, P., Arslan, A. N., and Pulliainen, J.: Evaluating Biosphere Model Estimates of the Start of the Vegetation Active Season in Boreal Forests by Satellite Observations, Remote Sensing, 8, https://doi.org/10.3390/rs8070580, 2016.

Brovkin, V., Boysen, Raddatz, T., Gayler, V., Loew, A., and Claussen, M.: Evaluation of vegetation cover and land-surface albedo in MPI-ESM CMIP5 simulations, J. Adv. Model. Earth Syst., 5, 48–57, https://doi.org/10.1029/2012MS000169, 2013.

Cescatti, A., Marcolla, B., Vannan, S. K. S., Pan, J. Y., Román, M. O., Yang, X., Ciais, P., Cook, R. B., Law, B. E., Matteucci, G., et al.: Intercomparison of MODIS albedo retrievals and in situ measurements across the global FLUXNET network, Remote sensing of environment, 121, 323–334, https://doi.org/10.1016/j.rse.2012.02.019, 2012.

Chase, T., Pielke, R., Kittel, T., Nemani, R., and Running, S.: Sensitivity of a general circulation model to global changes in leaf area index, J. Geophys. Res., 101, 7393–7408, 1996.

Chen, T., Rossow, W. B., and Zhang, Y.: Radiative Effects of Cloud-Type Variations, Journal of Climate, 13, 264 – 286, https://doi.org/10.1175/1520-0442(2000)013<0264:REOCTV>2.0.CO;2, 2000.

Claussen, M.: Modeling bio-geophysical feedbacks in the African and Indian monsoon region, Clim. Dyn., 13, 247–257, 1997.

Collatz, G. J., Ribas-Carbo, M., and Berry, J.: Coupled photosynthesis-stomatal conductance model for leaves of C4 plants, Functional Plant Biology, 19, 519–538, 1992.

Cramer, W., Kicklighter, D., Bondeau, A., Iii, B., Churkina, G., Nemry, B., Ruimy, A., Schloss, A., and Intercomparison, T.: Comparing global models of terrestrial net primary productivity (NPP): overview and key results, Global Change Biology, 5, 1–15, https://doi.org/10.1046/j.1365-2486.1999.00009.x, 1999.

Crueger, T., Giorgetta, M., Brokopf, R., Esch, M., Fiedler, S., Hohenegger, C., Kornblueh, L., Mauritsen, T., Nam, C., Naumann, A. K., Peters, K., Rast, S., Roeckner, E., Sakradzija, M., Schmidt, H., Vial, J., Vogel, R., and Stevens, B.: ICON-A, The Atmosphere Component of the ICON Earth System Model: II. Model Evaluation, Journal of Advances in Modeling Earth Systems, 10, 1638–1662, https://doi.org/10.1029/2017MS001233, 2018.

Dalmonech, D., Zaehle, S., Schürmann, G.J., Brovkin, V., Reick, C., and Schnur, R.: Separation of the effects of land and climate model errors on simulated contemporary land carbon cycle trends in the MPI Earth System Model version 1, Journal of Climate, 28, 272–291, https://doi.org/10.1175/JCLI-D-13-00593.1, 2015.

de Vrese, P. and Stacke, T.: Irrigation and hydrometeorological extremes, Climate Dynamics, 55, 1521–1537, https://doi.org/10.1007/s00382-020-05337-9, 2020.

de Vrese, P., Stacke, T., Kleinen, T., and Brovkin, V.: Diverging responses of high-latitude CO 2 and CH 4 emissions in idealized climate

740   change scenarios, The Cryosphere, 15, 1097–1130, https://doi.org/10.5194/tc-15-1097-2021, 2021.

Durack, P. J. and Taylor, K. E.: PCMDI AMIP SST and sea-ice boundary conditions version 1.1.2, https://doi.org/10.22033/ESGF/input4MIPs.1161, 2017.

DWD: Operationelle Einführung des neuen nichthydrostatischen Globalmodells ICON, Tech. rep., Deutscher Wetterdienst,   www.dwd.de/DE/fachnutzer/forschung_lehre/numerische_wettervorhersage/nwv_aenderungen/_functions/DownloadBox_

745   modellaenderungen/icon/pdf_2015/pdf_icon_17_12_2014.pdf?__blob=publicationFile&v=6, accessed 10 February 2021, 2014.

Ekici, A., Beer, C., Hagemann, S., Boike, J., Langer, M., and Hauck, C.: Simulating high-latitude permafrost regions by the JSBACH terrestrial ecosystem model, Geoscientific Model Development, 7, 631–647, https://doi.org/10.5194/gmd-7-631-2014, 2014.

Farquhar, G. D., von Caemmerer, S. v., and Berry, J. A.: A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species, Planta, 149, 78–90, 1980.

750 Feng, X., Vico, G., and Porporato, A.: On the effects of seasonality on soil water balance and plant growth, Water Resour. Res., 48, W05 543, https://doi.org/10.1029/2011WR011263, 2012.

Forster, P. et al.: Changes in atmospheric constituents and in radiative forcing, in: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., K. B. Avery and, M. T., and Miller, H. L., p. 131–234, Cambridge University Press,

755   Cambridge, United Kingdom and New York, NY, USA, 2007.

Gao, F., Schaaf, C. B., Strahler, A. H., Roesch, A., Lucht, W., and Dickinson, R.: MODIS bidirectional reflectance distribution function and albedo Climate Modeling Grid products and the variability of albedo for major global vegetation types, J. Geophys. Res., 110, D01 104, https://doi.org/10.1029/2004JD005190, 2005.

Gates, W. L.: AMIP: The Atmospheric Model Intercomparison Project, Bulletin of the American Meteorological Society, 73, 1962–1970,

760   https://doi.org/10.1175/1520-0477(1992)073<1962:ATAMIP>2.0.CO;2, 1992.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., et al.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, Journal of Advances in Modeling Earth Systems, 5, 572–597, 2013.

Giorgetta, M. A., Brokopf, R., Crueger, T., Esch, M., Fiedler, S., Helmert, J., Hohenegger, C., Kornblueh, L., Köhler, M., Manzini, E.,

765   Mauritsen, T., Nam, C., Raddatz, T., Rast, S., Reinert, D., Sakradzija, M., Schmidt, H., Schneck, R., Schnur, R., Silvers, L., Wan, H., Zängl, G., and Stevens, B.: ICON-A, the Atmosphere Component of the ICON Earth System Model: I. Model Description, Journal of Advances in Modeling Earth Systems, 10, 1613–1637, https://doi.org/10.1029/2017MS001242, 2018.

Hagemann, S. and Stacke, T.: Impact of the soil hydrology scheme on simulated soil moisture memory, Climate Dynamics, 44, 1731–750, 2015.

770 Hagemann, S., Blome, T., Ekici, A., and Beer, C.: Soil-frost-enabled soil-moisture–precipitation feedback over northern high latitudes, Earth Syst. Dynam., 7, 611–625, https://doi.org/10.5194/esd-7-611-2016, 2016.

Hickel, K. and Zhang, L.: Estimating the impact of rainfall seasonality on mean annual water balance using a top-down approach, J. Hydrol., 331, 409–424, 2006.

Houghton, R.: Balancing the global carbon budget, Annual Review of Earth and Planetary Sciences, 35, 313–347, 2007.

775 Hurtt, G., Chini, L., Frolking, S., and Sahajpal, R.: http://gsweb1vh2.umd.edu/LUH2/LUH2_v2h_README.pdf, for a short description of the LUH2v2h dataset see, 2016.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., Fisk, J. P., Hibbard, K., Houghton, R. A., Janetos, A., Jones, C. D., Kindermann, G., Kinoshita, T., Klein Goldewijk, K., Riahi, K., Shevliakova, E., Smith, S., Stehfest, E., Thomson, A., Thornton, P., van Vuuren, D. P., and Wang, Y. P.: Harmonization of land-use scenarios for the period 1500-2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary land, Climatic Change, 109, 117–161, https://doi.org/10.1007/s10584-011-0153-2, 2011.

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Klein Goldewijk, K., et al.: Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6, Geoscientific Model Development, 13, 5425–5464, https://doi.org/10.5194/gmd-2019-360, 2020.

Jungclaus, J., Lorenz, S., Schmidt, H., Brovkin, V., Brüggemann, N., Chegini, F., T. Crüger1, P. D.-V., Gayler, V., Giorgetta, M., Gutjahr, O., Haak, H., Hagemann, S., Hanke, M., Ilyina, T., Korn, P., Kröger, J., Linardakis, L., Mehlmann, C., Mikolajewicz, U., Müller, W., Nabel, J., Notz, D., Pohlmann, H., Putrasahan, D., Raddatz, T., Ramme, L., Redler, R., Reick, C., Riddick, T., Sam, T., Schneck, R., Schnur, R., Schupfner, M., von Storch, J.-S., Wachsmann, F., Wieners, K.-H., Ziemen, F., Stevens, B., Marotzke, J., and Claussen, M.: The ICON Earth System Model Version 1.0, Journal of Advances in Modeling Earth Systems, p. e2021MS002813, https://doi.org/10.1029/2021MS002813, 2022.

Kelley, D., Prentice, I., Harrison, S., Wang, H., Simard, M., Fisher, J., and Willis, K.: A comprehensive benchmarking system for evaluating global vegetation models, Biogeosciences, 10, 3313–3340, https://doi.org/https://doi.org/10.5194/bg-10-3313-2013, 2013.

Korn, P.: Formulation of an unstructured grid model for global ocean dynamics, J. Comp. Phys., 339, 525–552, https://doi.org/10.1016/j.jcp.2017.03.009, 2017.

Korn, P. and Linardakis, L.: A conservative discretizations of the shallow-water equations on triangular grids, J. Comp. Phys., 375, 871–900, 2018.

Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K., and Puma, M. J.: On the Nature of Soil Moisture in Land Surface Models, J. Climate, 22, 4322–4335, https://doi.org/10.1175/2009JCLI2832.1, 2009.

Kvas, A., Behzadpour, S., Ellmer, M., Klinger, B., Strasser, S., Zehentner, N., and Mayer-Gürr, T.: ITSG-Grace2018: Overview and evaluation of a new GRACE-only gravity field time series, Journal of Geophysical Research: Solid Earth, 124, 9332–9344, https://doi.org/10.1029/2019JB017415, 2019.

Levine, X. J. and Boos, W. R.: Land surface albedo bias in climate models and its association with tropical rainfall, Geophys. Res. Lett., 44, https://doi.org/10.1002/2017GL072510, 2017.

Loew, A., van Bodegom, P. M., Widlowski, J.-L., Otto, J., Quaife, T., Pinty, B., and Raddatz, T.: Do we (need to) care about canopy radiation schemes in DGVMs? Caveats and potential impacts, Biogeosciences, 11, 1873–1897, https://doi.org/10.5194/bg-11-1873-2014, 2014.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fläschner, D., Gayler, V., Giorgetta, M., Goll, D., Haak, H., Hagemann, S., Hedemann, C., et al.: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2, Journal of Advances in Modeling Earth Systems, 11, 998–1038, https://doi.org/10.1029/2018MS001400, 2019.

Max Planck Institute for Meteorology: CRP CMIP6: Max Planck Institute for Meteorology (MPI-M) MPI-ESM1-2-HR model output for the "amip" experiment. Centre for Environmental Data Analysis, 2021/02/13, https://catalogue.ceda.ac.uk/uuid/348138bb004242238bfc6979e8b8e55f, 2020.

Mayer-Gürr, T., Behzadpur, S., Ellmer, M., Kvas, A., Klinger, B., Strasser, S., and Zehentner, N.: ITSG-Grace2018 - Monthly, Daily and Static Gravity Field Solutions from GRACE, GFZ Data Services, https://doi.org/10.5880/ICGEM.2018.003, 2018.

815   Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., Fraser, P. J., Montzka, S. A., Rayner, P. J., Trudinger, C. M., Krummel, P. B., Beyerle, U., Canadell, J. G., Daniel, J. S., Enting, I. G., Law, R. M., Lunder, C. R., O'Doherty, S., Prinn, R. G., Reimann, S., Rubino, M., Velders, G. J. M., Vollmer, M. K., Wang, R. H. J., and Weiss, R.: Historical greenhouse gas concentrations for climate modelling (CMIP6), Geoscientific Model Development, 10, 2057–2116, https://doi.org/10.5194/gmd-10-2057-2017, 2017.

820   Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., et al.: A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1. 2-HR), Journal of Advances in Modeling Earth Systems, 10, 1383–1413, https://doi.org/10.1029/2017MS001217, 2018.

Myneni, R., Knyazikhin, Y., and Park, T.: MOD15A2H MODIS/Terra Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V006 dataset, distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at the U.S. Geological Survey (USGS) Earth Re-

825   sources Observation and Science (EROS) Center (lpdaac.usgs.gov);, https://doi.org/10.5067/MODIS/MOD15A2H.006, last access date: January 28, 2020, 2015.

Nabel, J. E. M. S., Naudts, K., and Pongratz, J.: Accounting for forest age in the tile-based dynamic global vegetation model JS-BACH4 (4.20p7; git feature/forests) - a land surface model for the ICON-ESM, Geoscientific Model Development, 13, 185–200, https://doi.org/10.5194/gmd-13-185-2020, 2020.

830   Niu, S., Xing, X., Zhang, Z., Xia, J., Zhou, X., Song, B., Li, L., and Wan, S.: Water-use efficiency in response to climate change: from leaf to ecosystem in a temperate steppe, Global Change Biology, 17, 1073–1082, 2011.

Otto, J., Raddatz, T., and Claussen, M.: Strength of forest-albedo feedback in mid-Holocene climate simulations, Climate of the Past, 7, 1027–1039, https://doi.org/10.5194/cp-7-1027-2011, 2011.

Ozturk, T., Hamza, A., Murat, T., and Levent, K. M.: Simulation of temperature and precipitation climatology for the Central Asia CORDEX

835   domain using RegCM 4.0, Clim Res, 52, 63–76, https://doi.org/10.3354/cr01082, 2012.

Piani, C., Weedon, G., Best, M., Gomes, S., Viterbo, P., Hagemann, S., and Haerter, J.: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, Journal of Hydrology, 395, 199–215, https://doi.org/10.1016/j.jhydrol.2010.10.024, 2010.

Piao, S., Friedlingstein, P., Ciais, P., Viovy, N., and Demarty, J.: Growing season extension and its impact on terrestrial carbon cycle in the

840   Northern Hemisphere over the past 2 decades, Glob. Biogeochem. Cycles, 21, GB3018, 2007.

Platnick, S. e. a.: MODIS Atmosphere L3 Daily Product. NASA MODIS Adaptive Processing System, Tech. rep., Goddard Space Flight Center, USA, https://doi.org/http://dx.doi.org/10.5067/MODIS/MOD08_D3.061, 2017.

Reick, C., Raddatz, T., Brovkin, V., and Gayler, V.: Representation of natural and anthropogenic land cover change in MPI-ESM, Journal of Advances in Modeling Earth Systems, 5, 459–482, https://doi.org/doi.org/10.1002/jame.20022, 2013.

845   Reick, C. H., Gayler, V., Goll, D., Hagemann, S., Heidkamp, M., Nabel, J. E. M. S., Raddatz, T., Roeckner, E., Schnur, R., and Wilkenskjeld, S.: JSBACH 3 - The land component of the MPI Earth System Model: Documentation of version 3.2, Berichte zur Erdsystemforschung, 240, https://doi.org/10.17617/2.3279802, 2021.

Running, S. and Zhao, M. S.: MOD17A2H MODIS/Terra Gross Primary Productivity 8-Day L4 Global 500m SIN Grid V006 dataset, Distributed as globally gridded MODIS GPP and net photosynthesis data at 0.5 resolution in netCDF format by the Integrated Climate Data

850   Center (ICDC, icdc.cen.uni-hamburg.de) University of Hamburg, Hamburg, Germany, https://doi.org/10.5067/MODIS/MOD17A2H.006, last accessed August 27, 2019, 2019.

Running, S., Mu, Q., Zhao, M., and Moreno, A.: MODIS/Terra Net Evapotranspiration Gap-Filled Yearly L4 Global 500m SIN Grid V061, https://lpdaacsvc.cr.usgs.gov/, https://doi.org/10.5067/MODIS/MOD16A3GF.061, accessed January 29, 2022, 2021.

Schaaf, C. and Wang, Z.: MCD43C3 MODIS/Terra+Aqua BRDF/Albedo Albedo Daily L3 Global 0.05 Deg CMG V006 Dataset, www.umb.edu/spectralmass/terra_aqua_modis/v006/mcd43c3, https:\icdc.cen.uni-hamburg.de/en/modis-landalbedo.html#c2294, https://doi.org/10.5067/MODIS/MCD43C3.006, 2015.

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N. C., Zhang, X., Jin, Y., Muller, J.-P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entremont, R. P., Hu, B., Liang, S., Privette, J. L., , and Roy, D. P.: First operational BRDF, albedo nadir reflectance products from MODIS, Tech. rep., NASA Publications 38, 2002.

Schneck, R., Reick, C., and Raddatz, T.: Land contribution to natural CO_2 variability on time scales of centuries, J. Adv. Model. Earth Syst., 5, 354–365, https://doi.org/10.1002/jame.20029, 2013.

Seiler, C., Melton, J. R., Arora, V. K., Sitch, S., Friedlingstein, P., Anthoni, P., Goll, D., Jain, A. K., Joetzjer, E., Lienert, S., Lombardozzi, D., Luyssaert, S., Nabel, J. E. M. S., Tian, H., Vuichard, N., Walker, A. P., Yuan, W., and Zaehle, S.: Are Terrestrial Biosphere Models Fit for Simulating the Global Land Carbon Sink?, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 946, https://doi.org/https://doi.org/10.1029/2021MS002946, e2021MS002946 2021MS002946, 2022.

Sellers, P.: Canopy reflectance, photosynthesis and transpiration, Int. J. Remote Sensing, 6, 1335–1372, 1985.

Settin, T., Botter, G., Rodriguez-Iturbe, I., and Rinaldo, A.: Numerical studies on soil moisture distributions in heterogeneous catchments, Water Resour. Res., 43, W05 425, https://doi.org/10.1029/2006WR005737, 2007.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., et al.: Atmospheric component of the MPI-M Earth system model: ECHAM6, Journal of Advances in Modeling Earth Systems, 5, 146–172, 2013.

Strahler, A., Schaaf, C., Gao, F., Li, X., and Lucht, W.: MODIS MOD43 Product User Guide V004, Tech. rep., Boston University, accessed 20 October 2021, 2021.

Sun, Y., Piao, S., Huang, M., Ciais, P., Zeng, Z., Cheng, L., Li, X., Zhang, X., Mao, J., Peng, S., Poulter, B., Shi, X., Wang, X., Wang, Y.-P., and Zeng, H.: Global patterns and climate drivers of water-use efficiency in terrestrial ecosystems deduced from satellite-based datasets and carbon cycle models, Global Ecology and Biogeography, 25, 311–323, 2016.

Swenson, S. C. and Milly, P. C. D.: Climate model biases in seasonality of continental water storage revealed by satellite gravimetry, Water Resour. Res., 42, W03 201, https://doi.org/10.1029/2005WR004628, 2006.

Taylor, K.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research, 106, 2001.

Taylor, K., Williamson, D., and Zwiers, F.: The sea surface temperature and sea- ice concentration boundary conditions for AMIP II simulations, PCMDI Report, 60, 2000.

Taylor, K., Stouffer, R., and Meehl, G.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93(4), 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2012.

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., van Vuuren, D., Riahi, K., Meinshausen, M., Nicholls, Z., Tokarska, K. B., Hurtt, G., Kriegler, E., Lamarque, J.-F., Meehl, G., Moss, R., Bauer, S. E., Boucher, O., Brovkin, V., Byun, Y.-H., Dix, M., Gualdi, S., Guo, H., John, J. G., Kharin, S., Kim, Y., Koshiro, T., Ma, L., Olivié, D., Panickal, S., Qiao, F., Rong, X., Rosenbloom, N., Schupfner, M., Séférian, R., Sellar, A., Semmler, T., Shi, X., Song, Z., Steger, C., Stouffer, R., Swart, N., Tachiiri, K., Tang, Q., Tatebe, H., Voldoire, A., Volodin, E., Wyser, K., Xin, X., Yang, S., Yu, Y., and Ziehn, T.: Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6, Earth System Dynamics, 12, 253–293, https://doi.org/10.5194/esd-12-253-2021, 2021.

890   TOP500 project: Performance Development, https://www.top500.org/statistics/, 2021.

Wan, Z., Hook, S., and Hulley, G.: MOD11C1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 0.05Deg CMG V006 data set, distributed in netCDF format by the Integrated Climate Data Center (ICDC, icdc.cen.uni-hamburg.de) University of Hamburg, Hamburg, Germany, /pool/data/ICDC/land/modis_terra_landsurfacetemperature/DATA/MONTHLY/ MODIS-C06__MOD11C3__MONTHLY__LandSurfaceTemperature__0.05deg__UHAM-ICDC_, https:\icdc-test.cen.uni-hamburg.
895   de/en/modis-landsurfacetemperature.html, last access date: January 6 2020, accessed 2021-04-16, 2015.

Wang, L., Cole, J. N. S., Bartlett, P., Verseghy, D., Derksen, C., Brown, R., and von Salzen, K.: Investigating the spread in surface albedo for snow-covered forests in CMIP5 models, J. Geophys. Res. Atmos., 121, 1104–1119, https://doi.org/10.1002/2015JD023824, 2016.

Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M.: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core, Q.J.R. Meteorol. Soc., 141, 563–579, https://doi.org/10.1002/qj.2378, 2015.