Explicit Representation of Spatial Subgrid-Scale Heterogeneity in an ESM

PHILIPP DE VRESE AND STEFAN HAGEMANN

Max Planck Institute for Meteorology, Hamburg, Germany

(Manuscript received 11 May 2015, in final form 15 February 2016)

ABSTRACT

In present-day Earth system models, the coupling of land surface and atmosphere is based on simplistic assumptions. Often the heterogeneous land surface is represented by a set of effective parameters valid for an entire model grid box. Other models assume that the surface fluxes become horizontally homogeneous at the lowest atmospheric model level. For heterogeneity above a certain horizontal length scale this is not the case, resulting in spatial subgrid-scale variability in the fluxes and in the state of the atmosphere. The Max Planck Institute for Meteorology's Earth System Model is used with three different coupling schemes to assess the importance of the representation of spatial heterogeneity at the land surface as well as within the atmosphere. Simulations show that the land surface–atmosphere coupling distinctly influences the simulated near-surface processes with respect to different land-cover types. The representation of heterogeneity also has a distinct impact on the simulated gridbox mean state and fluxes in a large fraction of land surface.

1. Introduction

In a simplistic representation, the link between surface and atmosphere comprises the surface heat, radiation, mass, and momentum fluxes. These fluxes have a nonlinear dependency on the state of the surface and the atmosphere. Land surface characteristics such as topography, the properties of the soil and hydrological characteristics, land use, vegetation, etc. vary on scales ranging from millimeters to hundreds of kilometers. Consequently, the state of the land surface and many spatially heterogeneous processes vary likewise (Sellers 1991). This wide range of scales, together with the nonlinear nature of processes involved, poses one of the fundamental difficulties in accurately describing the interaction of the land surface and the atmosphere, making a realistic representation a key challenge in Earth system modeling.

Present-day Earth system models (ESMs) employ different strategies to integrate the spatial subgrid-scale (SSGS) information of the land surface in the model's physical parameterizations and to aggregate the information to match the grid of the atmosphere. These strategies require simplifying assumptions that, because of the nonlinear nature of the processes involved, result in a distinct and possibly inaccurate representation of the land surface–atmosphere interaction. One basic distinction between different aggregation methods can be made between the parameter-aggregating and flux-aggregating methods. Both methods assume that a grid box can be sectioned into discrete subdivisions, the so-called ''tiles'' or ''patches,'' which themselves exhibit homogeneous characteristics. In the parameter-aggregating methods SSGS heterogeneity is not explicitly accounted for; instead, soil and vegetation parameters of the tiles are aggregated to one effective value representing the entire grid box. This is usually done by averaging all SSGS parameter values weighted by the respective cover fraction. The fluxes connecting surface and atmosphere are calculated based on these effective grid parameters (Giorgi and Avissar 1997). In this study, this will be referred to as the ''parameter aggregation.''

In the flux-aggregating methods, SSGS heterogeneity is explicitly represented and fluxes are calculated for each tile in a grid box individually based on the tile-specific characteristics. One of these methods is the ''mixture approach'' (Koster and Suarez 1992). This approach assumes that heterogeneity at the surface consists of numerous clusters, that is, homogeneous subareas, which cover only small areas and are evenly distributed across the grid box. The surface fluxes that originate from individual tiles are assumed to have completely blended horizontally below the lowest atmospheric model level.

Corresponding author address: Philipp de Vrese, Max Planck Institute for Meteorology, Bundesstrasse 53, 20146 Hamburg, Germany. E-mail: philipp.de-vrese@mpimet.mpg.de

DOI: 10.1175/JHM-D-15-0080.1

Consequently, the atmosphere interacts only with the mixed flux and remains horizontally homogeneous. The term "simple flux aggregation" used in this work refers to the mixture approach.

Many mostly local studies have compared techniques that apply an aggregation of parameters to those that aggregate fluxes (Avissar and Pielke 1989; Polcher et al. 1996; van den Hurk and Beljaars 1996; Cooper et al. 1997; Molod and Salmun 2002; Essery et al. 2003; Heinemann and Kerschgens 2005; Ament and Simmer 2006). These studies found pronounced differences between simulations performed with the two coupling techniques, and they agree that by employing an aggregation of fluxes the representation of processes is clearly improved. Many of the studies additionally found an improvement of the simulated climate.

It is evident that the assumption of atmospheric spatial homogeneity may be valid only in specific circumstances. Numerous studies, both modeling and observational, showed that it indeed becomes erroneous when the scale of surface heterogeneity increases beyond the microscale (Mason 1988; Claussen 1995; Raupach and Finnigan 1995; Avissar and Schmidt 1998; Mahrt 2000; Bou-Zeid et al. 2004; Patton et al. 2005; Ma et al. 2008). In this case, the signal associated with a specific surface feature may be detectable far above the surface layer, resulting in a spatially heterogeneous state of the atmosphere at heights above the lowest model level of many ESMs, which is often located below a height of 50 m (Arola 1999).

The present study investigates the importance of an explicit representation of SSGS heterogeneity at the surface and in the lowest parts of the atmosphere, using simulations performed with the Max Planck Institute Earth System Model (MPI-ESM), more specifically, with the land surface model JSBACH (Raddatz et al. 2007; Brovkin et al. 2009; Ekici et al. 2014) coupled to the general circulation model ECHAM6 (Stevens et al. 2013), which will in the following be referenced as ECHAM–JSBACH. Section 2 gives a brief overview over the model and the different coupling schemes used in the study. In addition, the simulations will be described in more detail. In section 3, the investigation is focused on determining whether different land-cover types display systematic differences in the near-surface processes with respect to the different coupling schemes. Section 4 focuses on the impact of surface and atmospheric SSGS heterogeneity on the simulated global climate. The main results are summarized and shortly discussed in section 5.

2. Methods

In the operational setup, ECHAM–JSBACH uses a parameter aggregation scheme to couple land surface and atmosphere. Here, the determination of an effective gridbox mean albedo is described in Otto et al. (2011); the aggregation of the surface roughness length of different tiles follows Mason (1988), Claussen (1991), and Claussen et al. (1994); and the aggregation of hydrological and soil parameters is done as a weighted (by the respective cover fractions) average (Kabat et al. 1997; Feddes et al. 1998). Based on the effective parameters, the surface fluxes are calculated using a bulk-exchange formulation (Giorgetta et al. 2013). For the present study, ECHAM–JSBACH was altered to enable a surface–atmosphere coupling using a simple flux aggregation scheme and an improved coupling method that represents the turbulent mixing process more realistically, that is, the Vertical Tile Extension (VERTEX) scheme (de Vrese et al. 2016). In both schemes, the homogeneous subareas in a grid box are represented by individual tiles that only interact via the vertical turbulent fluxes. In the soil, a horizontal transport of water and heat is not modeled and the soil moisture and temperature of a given tile is independent of the other tiles.

The simple flux aggregation scheme implemented into ECHAM–JSBACH, follows the general method for an implicit surface–atmosphere coupling proposed by Polcher et al. (1998) and Best et al. (2004). Similar to the simple flux aggregation scheme, the VERTEX scheme explicitly represents surface SSGS heterogeneity by individual tiles. Additionally, the scheme explicitly represents SSGS heterogeneity within the three lowest layers of the atmosphere, that is, up to a height of roughly 350 m. In the atmosphere, the vertical fluxes within the individual tiles are modeled by a modified version of the turbulent kinetic energy scheme described in Brinkop and Roeckner (1995). In the VERTEX scheme, the fluxes within the individual tiles are not treated independently of each other but are assumed to blend horizontally to a certain extent. Thus, the vertical flux from a given tile can influence the states of all the tiles on the level above. The extent to which the vertical fluxes blend horizontally is determined based on the ratio of the height of a model level and the blending height. The latter can be estimated as a function of friction velocity, the horizontal wind speed, and the characteristic length scale of the respective surface heterogeneity (Mahrt 2000). To approximate the characteristic length scale of a given land-cover type within a grid box, the horizontal extents of all clusters (homogeneous subareas) of the respective land-cover type within the grid box need to be determined. These are derived from the GlobCover dataset (Arino et al. 2012). The resolution of 300 m $\times$ 300 m of GlobCover allows us to resolve a T63 model grid box by approximately 390 000 pixels. From the 22 land-cover classes employed

in the dataset, 10 are selected and matched to the tiles used in JSBACH. For each pixel within a grid box, the distance in the north–south, east–west, northwest–southeast, and northeast–southwest direction to the nearest pixel with a different land-cover class is determined. These distances provide a rough estimate for the dimensions (in the given directions) of the pixel's superordinate cluster. For each cover type within a given grid box, a characteristic length scale can be derived from the average horizontal dimensions of the homogeneous subareas belonging to the respective cover type.

Because of the many degrees of freedom incorporated in global-scale simulations, that is, in this study the atmospheric model coupled to a land surface model with prescribed sea surface temperature and sea ice extent (AMIP-style; Gates et al. 1999), these can be difficult to interpret. Single-column simulations, on the other hand, facilitate the understanding of relevant processes but the findings are often limited to the specific location the model was applied to. Offline simulations, that is, the land surface model forced by observations, have the disadvantage that a feedback from the surface onto the atmosphere is being omitted so that findings may not be applicable to the fully coupled land surface–atmosphere system. However, they serve as a compromise between AMIP-style and single-column simulations in that they facilitate process understanding on the global scale. In section 3, a slightly modified version of this type of experiment was conducted as a quasi-offline study.

For a 5-yr AMIP-style simulation, ECHAM–JSBACH was adapted in such a way that all calculations pertaining to surface processes, including the calculation of the state and the surface fluxes, were duplicated twice, effectively creating three different land surfaces within JSBACH. Each surface developed an individual state, as they were all coupled to the same atmospheric conditions but via a different coupling scheme. These are coupled with parameter aggregation ($\mathrm{PARAM_{SRF}}$), a simple flux aggregation ($\mathrm{SIMPLE_{SRF}}$), and the VERTEX scheme ($\mathrm{VERTEX_{SRF}}$). Differences between the above surfaces will be referenced by ΔSRF indexed with two letters indicating the two respective surfaces. For example, $\Delta\mathrm{SRF}_{S-P}$ refers to the differences between $\mathrm{SIMPLE_{SRF}}$ and $\mathrm{PARAM_{SRF}}$.

The surface–atmosphere feedback is taken from $\mathrm{PARAM_{SRF}}$, whereas the potential feedback from the other land surfaces is calculated for each time step but does not affect the state of the atmosphere in the subsequent time step.

The VERTEX scheme requires the knowledge of atmospheric SSGS variations of temperature and specific humidity at the beginning of each time step. To preserve the SSGS variability in the atmosphere, a refinement was applied to the state variables at the lowest atmospheric model levels. At the end of a 20-min time step, the tile-specific SSGS temperature and humidity deviations from the grid mean were stored and added to the prescribed grid mean values at the beginning of the following time step.

The simulation was performed for the period of 1979–84 and the multiannual mean of the years 1980–84 will be analyzed for the global scale with a focus on the relation between soil moisture, surface temperature, and the turbulent heat fluxes. The simulation was performed at a standard vertical resolution of 47 levels, of which the lowest is located on a height of approximately 30 m and a horizontal resolution of T63, that is, a grid spacing of 140 km × 210 km at midlatitudes. Because surface–atmosphere feedbacks are not taken into account, the analysis in section 3 only facilitates a qualitative understanding of the effects on near-surface processes.

To estimate the impact of a given coupling scheme on the simulated global climate, 3 × 5 AMIP-style simulations were performed, which will be discussed in section 4. These are an ensemble of five simulations using a parameter aggregation scheme ($\mathrm{PARAM_{ENS}}$), an ensemble of five simulations performed with the simple flux aggregation scheme ($\mathrm{SIMPLE_{ENS}}$), and an ensemble of five simulations employing the VERTEX scheme ($\mathrm{VERTEX_{ENS}}$). The latter two may also be referred to as the two $\mathrm{FLUX_{ENS}}$ as opposed to $\mathrm{PARAM_{ENS}}$. The comparison of two of the above ensembles will be referenced by ΔENS indexed with two letters indicating the two respective ensembles. For example $\Delta\mathrm{ENS}_{S-P}$ refers to the differences between $\mathrm{SIMPLE_{ENS}}$ and $\mathrm{PARAM_{ENS}}$.

Each simulation was performed for the same 21-yr period (1979–99; 1979 was required for the model spinup and is omitted from the analysis) but initialized with slightly different conditions. All the above simulations were performed at a standard vertical resolution of 47 levels and a horizontal resolution of T63.

3. Impact of spatial heterogeneity on the near-surface processes in individual tiles

The major conceptual difference between parameter and flux aggregation as realized in JSBACH is the existence of SSGS variations in the state of the land surface and in the surface fluxes. In the operational setup (parameter aggregation), technically SSGS heterogeneity exists in the form of the different surface tiles. However, the majority of physical surface and soil processes, including the calculation of surface fluxes, surface temperature, and soil moisture content, are modeled based on grid mean values, and thus de facto heterogeneity is not being considered. With the introduction of the flux aggregation, not only was the land surface–atmosphere

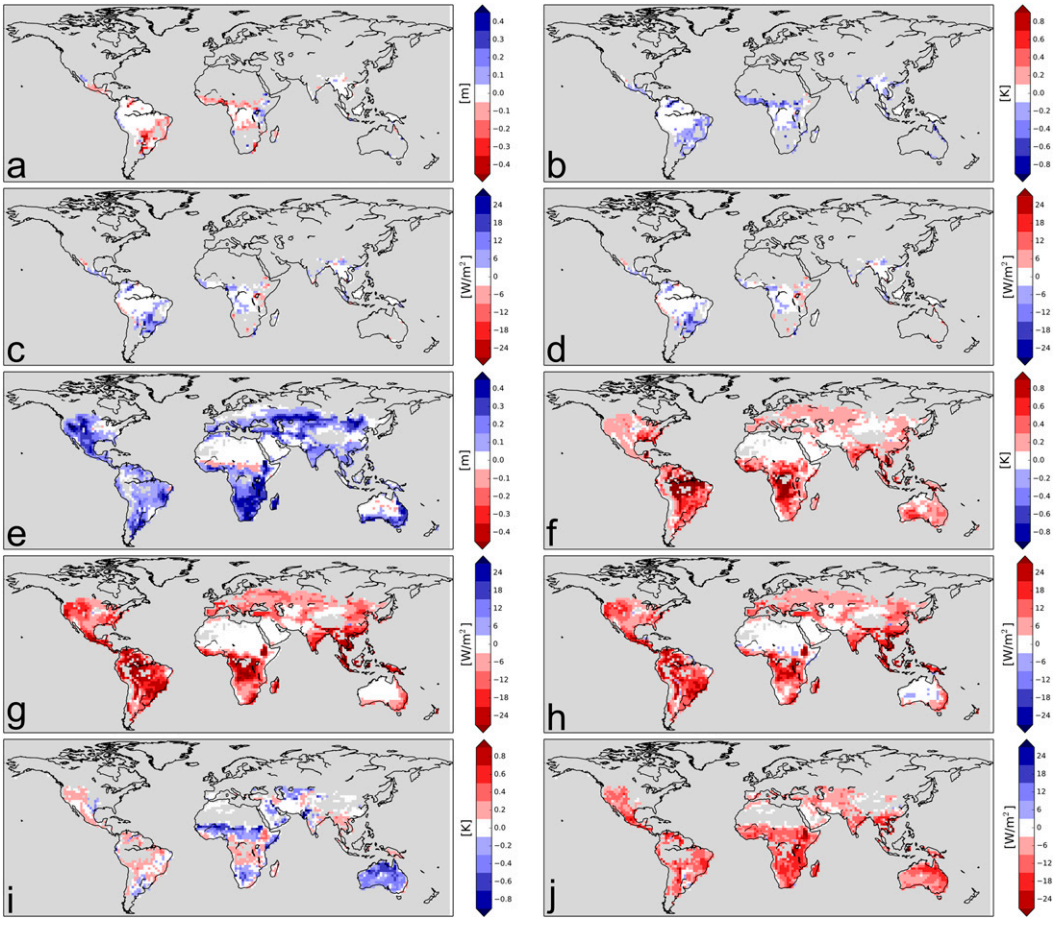

FIG. 1. The 5-yr mean difference between $\text{SIMPLE}_{\text{SRF}}$ and $\text{PARAM}_{\text{SRF}}$ for tropical evergreen forest (a) soil moisture, (b) surface temperature, (c) latent heat flux, and (d) sensible heat flux; for C_4 pasture (e) soil moisture, (f) surface temperature, (g) latent heat flux, and (h) sensible heat flux; and for raingreen shrubs (i) surface–atmosphere temperature difference and (j) sensible heat flux. Gray shading indicates absence of tile.

coupling changed, but SSGS variations are considered in the majority of the surface and soil processes.

In JSBACH, tiles represent different plant functional types (PFTs; with the exception of glaciers), which in turn each represent different species aggregated to groups based on their functional properties. For the following considerations, one of the most important properties of a PFT is its ability to transpire, which is closely related to the photosynthetic activity. In JSBACH this is reflected by a PFT's stomatal conductivity and leaf area. The impact of differences in albedo was found to be considerably smaller and has been omitted in the analysis by calculating the absorbed radiation at the surface based on the gridbox mean albedo.

a. Tile-based comparison of $SIMPLE_{SRF}$ and $PARAM_{SRF}$

Because of the PFT-specific properties determining the stomatal conductivity and the leaf area, some PFTs transpire at relatively high rates. In the present model setup these are tropical and extratropical trees, C_3 grass, and C_3 pasture (Figs. 1a–d). When comparing $\text{SIMPLE}_{\text{SRF}}$ to $\text{PARAM}_{\text{SRF}}$, an immediate effect within the respective tiles is a predominant increase in transpiration. As the excess in transpiration is not balanced by a decrease in evaporation, it leads to an increase in evapotranspiration and in the latent heat flux. Because more energy is expended in evapotranspiration, a new energetic equilibrium is reached at the surface with lower surface temperatures and a reduced sensible heat flux. Precipitation is calculated based on $\text{PARAM}_{\text{SRF}}$ and is unaffected by $\text{SIMPLE}_{\text{SRF}}$; thus, the shift toward increased evapotranspiration in $\text{SIMPLE}_{\text{SRF}}$ does not increase precipitation and therefore mostly leads to lower soil moisture values in these tiles. In turn, tiles in which the respective PFTs transpire at relatively low rates such as shrubs, C_4 grass, and C_4 pasture show an opposite behavior (Figs. 1e–h). Here the shift toward less transpiration leads to wetter soils,

a shift from latent to sensible heat flux, and higher surface temperatures.

In many regions there are no distinct differences between $SIMPLE_{SRF}$ and $PARAM_{SRF}$ for the dominant tile, for example, for tropical evergreen trees in the Amazon basin (Figs. 1a–d). This is related to the high cover fraction of the tile in these regions, that is, often more than 90% (not shown). In case of large cover fractions, the effective parameters and consequently the gridbox mean states and fluxes are largely determined by the dominant tile. As a result, $PARAM_{SRF}$ and the dominant tile in $SIMPLE_{SRF}$ and $VERTEX_{SRF}$ are very similar.

Another important aspect is the vegetation roughness length associated with the PFTs. Under given wind conditions, it determines the generation of shear-driven turbulence and accordingly has a strong impact on the turbulent fluxes. This characteristic becomes especially relevant for regions in which the vegetation comprises PFTs with distinctly differing heights such as savanna. Here, the vegetation consists of trees or shrubs and very low vegetation such as grasses or pasture. In tiles with a surface roughness larger than the grid mean, the consideration of heterogeneity leads to a closer link between surface and air temperatures, as more turbulence is created near the surface under given wind conditions. In turn, in the tiles with low vegetation, less turbulence is generated, resulting in a looser link between surface and air temperature. At the southern border of the Sahel zone and northern Australia, raingreen shrubs constitute vegetation with a comparable large roughness length, and the link between surface and atmosphere is relatively close. Therefore, in the respective regions the sensible heat flux increases distinctly despite the occurrence of a cooling of the surface and a reduction of the surface–atmosphere temperature gradient (Figs. 1i–j). Essery et al. (2003) found comparable effects for a comparison of trees and bare soil areas.

b. *Tile-based comparison of $VERTEX_{SRF}$ and $SIMPLE_{SRF}$*

When comparing the VERTEX and the simple flux scheme, the most relevant conceptual change is the reduction in horizontal blending of the surface fluxes in the VERTEX scheme, which results in SSGS variability in the state of the lowest layers of the atmosphere.

In comparison to the mean state, the tile-specific state of the atmosphere above PFTs with pronounced transpirational abilities is moister and cooler for $VERTEX_{SRF}$ as less dry and warm air from other tiles is being mixed into the local atmosphere. As the surface is usually warmer and the specific humidity higher than at the lowest atmospheric model level, this primarily leads to a reduced surface–atmosphere moisture gradient and an increase in the surface–atmosphere temperature gradient within the respective tiles (Figs. 2a,b). Given similar wind conditions and roughness lengths at the surface, differences in the surface–atmosphere gradients determine the differences in the surface fluxes between $VERTEX_{SRF}$ and $SIMPLE_{SRF}$. With a reduced humidity gradient, the latent heat flux decreases and, because of the increased surface–atmosphere temperature gradient, the sensible heat flux increases (Figs. 2c,d). Note that the cooling of the atmosphere due to the reduced horizontal blending is so pronounced that the surface–atmosphere temperature gradient increases, even though surface temperatures mostly decrease (not shown). Furthermore, because of the decrease in evapotranspiration, the soil moisture in the tiles with pronounced transpirational abilities is also higher for $VERTEX_{SRF}$ than for $SIMPLE_{SRF}$ (not shown). Note that the absence of pronounced differences in regions in which the dominant tile has a large cover fraction, for example, tropical evergreen trees in the Amazon basin, is related to the argument given in the section above. Here, the cover fraction of the dominant tile is so large that the aggregated surface flux is mainly determined by the flux from this tile. Therefore, the state in the atmosphere above this tile is also very similar to the mean state without refinement, and both schemes yield similar results.

For PFTs that transpire at relatively low rates, the refined atmosphere for the VERTEX scheme is drier and warmer. This increases the atmospheric moisture demand and reduces the surface–atmosphere temperature gradient, which results in a shift from sensible to latent heat flux (Figs. 2e–h). In the respective tiles, the increase in atmospheric temperature, due to the reduced horizontal blending of the surface fluxes, is often so pronounced that the sensible heat flux decreases despite increasing surface temperatures (not shown).

c. *Effect on the gridbox mean state*

Because of the nonlinear dependency of the surface fluxes on the state of the surface, especially on the plant available water, $SIMPLE_{SRF}$ and $VERTEX_{SRF}$ differ distinctly from $PARAM_{SRF}$ also with respect to the gridbox mean. Precipitation is not resolved with respect to the tiles. This leads to an aggregation of water within the tiles with poor transpirational abilities and a decline in soil moisture within tiles with pronounced transpirational abilities, when resolving heterogeneity at the surface. The increase in transpiration in tiles with pronounced transpirational abilities is often limited by the availability of water, whereas transpiration in tiles with less pronounced transpirational abilities does not necessarily increase with increasing soil moisture. Therefore,

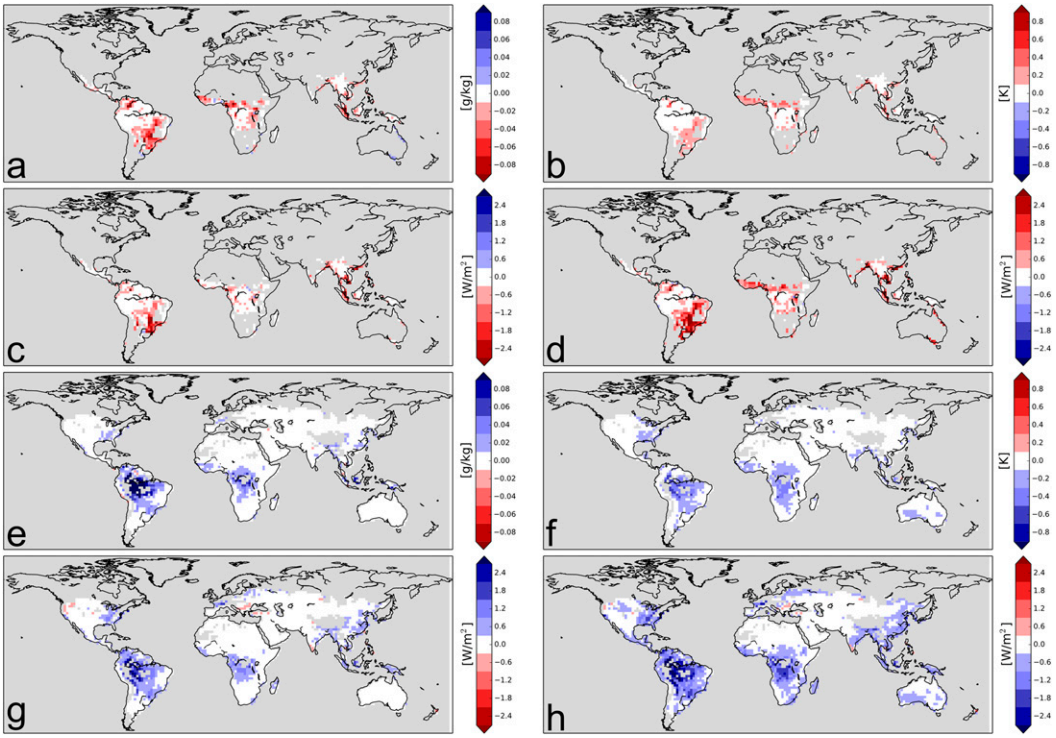

FIG. 2. The 5-yr mean difference between $\text{VERTEX}_{\text{SRF}}$ and $\text{SIMPLE}_{\text{SRF}}$ for tropical evergreen forest (a) surface–atmosphere specific humidity difference, (b) surface–atmosphere temperature difference, (c) latent heat flux, and (d) sensible heat flux; and for C_4 pasture (e) surface–atmosphere specific humidity difference, (f) surface–atmosphere temperature difference, (g) latent heat flux, and (h) sensible heat flux. Gray shading indicates absence of tile.

the effect of reduced transpiration in certain tiles clearly dominates the effect of the increased transpiration in other tiles when comparing the surfaces coupled with the flux aggregation schemes to that coupled with the parameter aggregation scheme. On the grid scale, the effect of introducing the simple flux aggregation is a distinct reduction in transpiration in a large part of the land surface that causes a decrease in evapotranspiration and an increase in soil moisture (Fig. 3a). Accordingly, a shift from surface latent to sensible heat flux occurs, and there is a corresponding increase in the surface temperature (Figs. 3b–d). This agrees well with a conceptual study performed by Koster and Suarez (1992), who found that when aggregating SSGS parameters to effective grid values (as in $\text{PARAM}_{\text{SRF}}$), these are dominated by the tiles that are less effected by water stress. This results in increased evapotranspiration and lower surface temperatures in comparison to using a simple flux aggregation (as in $\text{SIMPLE}_{\text{SRF}}$).

In certain tiles the effect of resolving atmospheric SSGS heterogeneity (ΔSRF_{V-S}) is of the same order of magnitude as that of resolving heterogeneity at the surface (ΔSRF_{S-P}). For example, with respect to the average surface temperature for tiles representing evergreen (deciduous) tropical trees, ΔSRF_{V-S} amounts to 75% (80%) of ΔSRF_{S-P}. For shrubs, evergreens, and deciduous extratropical trees, these figures still range between 30% and 50%. However, the differences in soil moisture and in the surface heat fluxes are mostly an order of magnitude smaller in the case of ΔSRF_{V-S} than for ΔSRF_{S-P}. This discrepancy is also present in the comparison of the gridbox mean values. This implies that, given an identical mean state in the atmosphere, the surface is comparably insensitive to the representation of SSGS heterogeneity in the atmosphere. This is in good agreement with studies that indicate that atmospheric SSGS heterogeneity only has a minor impact on the magnitude of the surface fluxes (Mahrt and Sun 1995; Ament and Simmer 2006; Schomburg et al. 2012).

This may only be valid when omitting the surface–atmosphere feedbacks that, in the given experimental setup, are indicated by the differences in atmospheric temperature and specific humidity calculated at the end of each time step (Figs. 3e–h). The omitted feedbacks are also the reason why the quasi-offline experiment does not allow any conclusions on the absolute magnitude of impacts or their importance in a coupled land surface–atmosphere system. In some grid boxes, the 5-yr

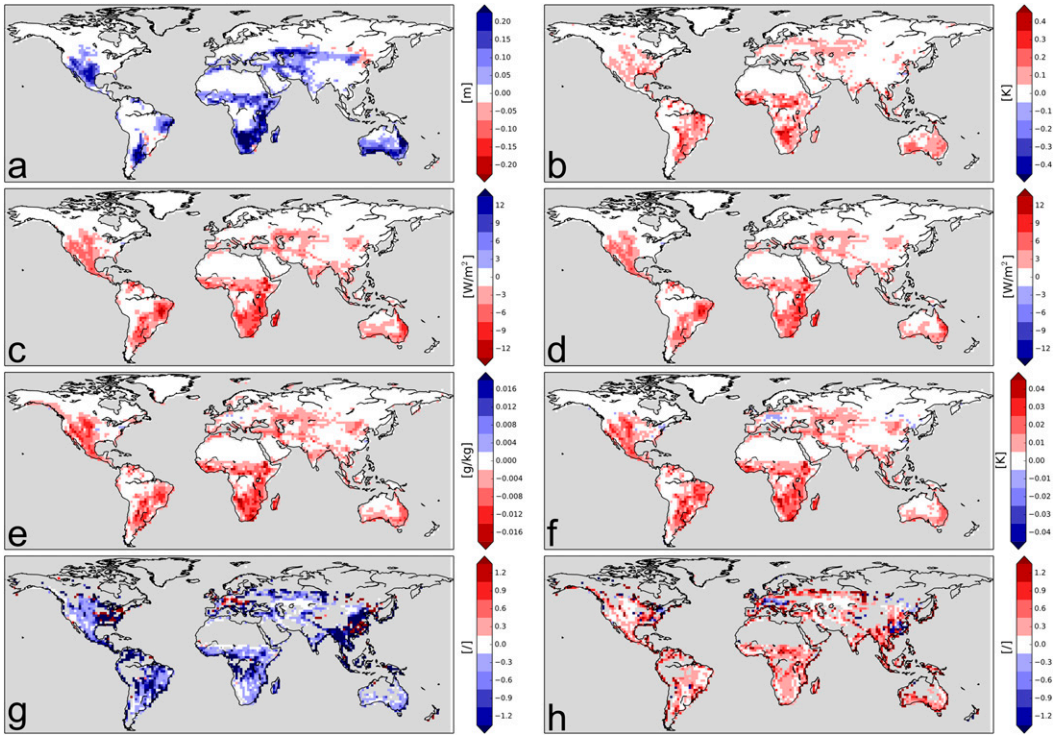

FIG. 3. The 5-yr mean difference between $\text{SIMPLE}_{\text{SRF}}$ and $\text{PARAM}_{\text{SRF}}$ for gridbox average (a) soil moisture, (b) surface temperature, (c) latent heat flux, (d) sensible heat flux, (e) specific humidity on the lowest atmospheric level, and (f) temperature on the lowest atmospheric level; and the 5-yr mean difference between $\text{VERTEX}_{\text{SRF}}$ and $\text{SIMPLE}_{\text{SRF}}$ relative to $\text{SIMPLE}_{\text{SRF}}$ and $\text{PARAM}_{\text{SRF}}$ for gridbox average (g) specific humidity on the lowest atmospheric level and (h) temperature on the lowest atmospheric level. Panels (g) and (h) are masked as to not include grid boxes in which the absolute value of the differences between $\text{SIMPLE}_{\text{SRF}}$ and $\text{PARAM}_{\text{SRF}}$ or between $\text{VERTEX}_{\text{SRF}}$ and $\text{SIMPLE}_{\text{SRF}}$ are larger than $0.003\,\text{g}\,\text{kg}^{-1}$ per time step and 0.005 K per time step.

mean temperature (specific humidity) differences at the end of each 20-min time step may be as large as 0.05 K ($0.025\,\text{g}\,\text{kg}^{-1}$). Hence, by omitting the surface–atmosphere feedback, a constant heating/cooling (humidifying/drying) at a rate of $0.15\,\text{K}\,\text{h}^{-1}$ ($0.075\,\text{g}\,\text{kg}^{-1}\,\text{h}^{-1}$) is introduced. When considering that even for ΔSRF_{S-P} the gridbox mean differences in surface temperature are mostly below 0.5 K, it has to be assumed that artificially cooling the atmosphere by $0.1\,\text{K}\,\text{h}^{-1}$ substantially affects the state of the surface and the surface fluxes.

For the gridbox mean atmospheric state at the end of each time step, ΔSRF_{V-S} is comparable to ΔSRF_{S-P}. With respect to the average temperature on the lowest atmospheric model level, ΔSRF_{V-S} amounts to roughly two-thirds of ΔSRF_{S-P}. For specific humidity ΔSRF_{V-S} is even larger (by about 10%) than ΔSRF_{S-P}, which indicates that the state of the entire atmospheric column is distinctly modified by accounting for SSGS heterogeneity at the lowest atmospheric model levels. This is also suggested by single-column studies performed with the VERTEX scheme (de Vrese et al. 2016) and by other modeling studies (Seth et al. 1994; Arola 1999; Giorgi et al. 2003; Molod et al. 2004; Salmun et al. 2007; Dimri 2009; Manrique-Suñén et al. 2013). To further explore this possibility, in the next section a model setup is used in which the surface–atmosphere feedback is accounted for.

4. Impact of spatial heterogeneity on the simulated global climate

With the atmospheric feedback taken into account, there is a good agreement between the results obtained with ECHAM–JSBACH and similar modeling studies, which gives some confidence that these findings have a more general validity. For example, the seasonally averaged surface temperature differences (averaging period is 20 years) between individual ensemble members range within roughly ±2.5 K, which agrees with the findings of Essery et al. (2003), who compared a flux aggregation to a parameter aggregation version of the Met Office Surface Exchange Scheme in combination with the Met Office GCM. They found that differences in surface temperature (averaging period 10 years) are

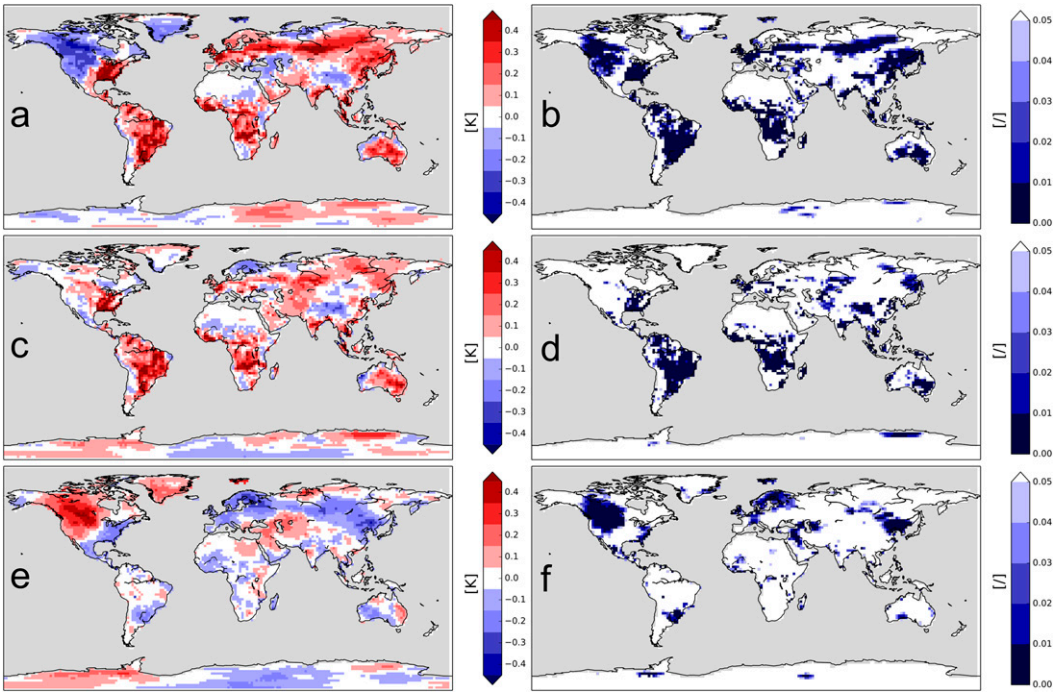

FIG. 4. (left) The 20-yr mean difference in surface temperature between (a) $\text{SIMPLE}_{\text{ENS}}$ and $\text{PARAM}_{\text{ENS}}$, (c) $\text{VERTEX}_{\text{ENS}}$ and $\text{PARAM}_{\text{ENS}}$, and (e) $\text{VERTEX}_{\text{ENS}}$ and $\text{SIMPLE}_{\text{ENS}}$. (right) The p value of statistical significance of 20-yr mean difference in surface temperature between (b) $\text{SIMPLE}_{\text{ENS}}$ and $\text{PARAM}_{\text{ENS}}$, (d) $\text{VERTEX}_{\text{ENS}}$ and $\text{PARAM}_{\text{ENS}}$, and (f) $\text{VERTEX}_{\text{ENS}}$ and $\text{SIMPLE}_{\text{ENS}}$.

mostly smaller than ±3 K. In the following, we focus exclusively on differences between the annual means, as these resemble the seasonal averages and no important effect is neglected because of the higher temporal aggregation.

The differences in the 20-yr annual mean surface temperature between the ensembles are much smaller, rarely exceeding 1 K (Fig. 4), and in the majority of the land surface ΔENS_{S-P}, ΔENS_{V-P}, and ΔENS_{V-S} are below 0.125 K. The differences in the surface energy fluxes are predominantly below 1 W m^{-2} and with respect to precipitation they mostly do not surpass 0.05 mm day^{-1} (Tables 1, 2). For most of the variables ΔENS_{S-P}, ΔENS_{V-P}, and ΔENS_{V-S} are comparable in magnitude. This is contrary to expectations based on the quasi-offline experiment, which suggest that ΔENS_{V-S} should be an order of magnitude smaller than ΔENS_{S-P} and ΔENS_{V-P}. It can be seen that ΔENS_{V-S} is consistently less pronounced, but generally by a factor ranging between 1.5 and 3 and not by an order of magnitude. In some cases, such as the relative precipitation differences of more than 10%, ΔENS_{V-S} is actually larger than ΔENS_{S-P} and ΔENS_{V-P}.

A two-sample, two-sided Student's t test is used to determine the statistical significance of the 20-yr mean differences between two ensembles. Here, the ensembles are treated as one series of 100 independent annual means (Table 3; Fig. 4). The Student's t test shows that,

TABLE 1. Share of land surface (%) exhibiting differences larger than x (y).

	ΔENS_{S-P}	ΔENS_{V-P}	ΔENS_{V-S}
Temp at surface: 0.125 (0.4) K	44.8 (6.4)	31.8 (4.3)	25.8 (0.7)
Temp at 2 m: 0.125 (0.4) K	40.6 (2.8)	27.4 (1.6)	22.1 (0.4)
Surface net shortwave radiation: 1.0 (2.0) W m^{-2}	18.7 (1.0)	13.6 (1.0)	9.1 (0.5)
Surface net longwave radiation: 1.0 (2.0) W m^{-2}	28.2 (6.0)	21.1 (4.3)	6.1 (0.1)
Sensible heat flux: 1.0 (2.0) W m^{-2}	30.3 (10.4)	29.1 (11.7)	18.8 (5.3)
Latent heat flux: 1.0 (2.0) W m^{-2}	35.7 (17.3)	34.1 (16.8)	20.4 (6.6)
Precipitation: 0.05 (0.2) mm day^{-1}	35.4 (1.8)	31.0 (3.3)	28.9 (1.3)
Vertically integrated water vapor: 0.1 (0.25) kg m^{-2}	41.8 (12.2)	35.1 (8.5)	30.3 (4.4)
Soil moisture: 0.01 (0.025) m	27.2 (10.2)	28.0 (9.4)	10.1 (2.0)

TABLE 2. Share of land surface (%) exhibiting relative differences >5% (10%).

	ΔENS_{S-P}	ΔENS_{V-P}	ΔENS_{V-S}
Temp at surface	—	—	—
Temp at 2 m	—	—	—
Surface net shortwave radiation	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Surface net longwave radiation	1.6 (0.0)	0.9 (0.0)	0.0 (0.0)
Sensible heat flux	23.9 (10.9)	24.6 (11.6)	15.2 (5.8)
Latent heat flux	23.2 (8.4)	22.5 (9.5)	19.4 (10.1)
Precipitation	20.9 (5.9)	17.2 (6.3)	18.5 (7.3)
Vertically integrated water vapor	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Soil moisture	26.2 (15.4)	27.1 (14.9)	5.9 (3.5)

TABLE 3. Share of land surface grid boxes (%) exhibiting significant differences ($p < 0.05$).

	ΔENS_{S-P}	ΔENS_{V-P}	ΔENS_{V-S}
Temp at surface	38.6	27.3	17.9
Temp at 2 m	36.6	24.6	16.6
Surface net shortwave radiation	28.0	20.0	12.4
Surface net longwave radiation	42.5	34.3	16.9
Sensible heat flux	30.6	30.3	21.6
Latent heat flux	35.8	35.9	21.5
Precipitation	21.6	18.3	16.4
Vertically integrated water vapor	19.8	15.6	9.5
Soil moisture	47.5	48.0	23.6

at a significance level of 0.05, between 9.5% and 48.0% of the land surface exhibits a significant impact due to changes in the surface–atmosphere coupling, depending on the variables and schemes considered. The largest impact on simulated climate can be found due to the explicit representation of surface heterogeneity. For ΔENS_{S-P} the share of the land surface exhibiting significant impacts mostly ranges between 30% and 40%, whereas the corresponding values for ΔENS_{V-S} generally lie in a range between 15% and 20%. For ΔENS_{V-P}, they predominantly range between 15% and 35%. Thus, the Student's t test supports the above results, that is, with respect to statistical significance, the impact of representing atmospheric SSGS heterogeneity primarily ranges between one-third and two-thirds of the impact of an explicit representation of surface SSGS heterogeneity. Furthermore, it confirms that the VERTEX scheme gives results that are more similar to the parameter aggregation scheme than the simple flux aggregation scheme.

In the following, correlations between the quasi-offline and the AMIP-style experiment are investigated in order to estimate whether the differences between two ensembles can be related to the mechanisms discussed in section 3. Here, the Pearson correlation coefficient r is used to test for a linear correlation. Weak ($|r|$ ranging between 0.30 and 0.52) yet statistically significant ($p \ll 0.01$) positive correlations can be found between ΔENS_{S-P} and ΔSRF_{S-P} for surface temperature, soil moisture, and the surface heat fluxes. This indicates that the impact of an explicit representation of surface heterogeneity on the 20-yr mean state of the surface and the surface fluxes, which were discussed in this section, can be partly related to the mechanisms discussed in section 3. Consequently, the impact of an explicit representation of SSGS heterogeneity at the surface is also similar in the coupled simulation and in the quasi-offline simulation. In condensed and simplified form, between 50°N and 50°S, an explicit representation of SSGS heterogeneity at the surface (ΔENS_{S-P}) results in generally higher temperatures and higher soil moisture values, less precipitation, larger sensible heat fluxes, and reduced evapotranspiration and thus lower latent heat fluxes and a smaller amount of vertically integrated water vapor (Table 4).

For the surface variables, no such correlations can be found between ΔENS_{V-S} and ΔSRF_{V-S}. This is not surprising as, in the quasi-offline experiment, $\text{VERTEX}_{\text{SRF}}$ and $\text{SIMPLE}_{\text{SRF}}$ do not exhibit large differences in the gridbox mean state and surface fluxes, whereas $\text{VERTEX}_{\text{ENS}}$ and $\text{SIMPLE}_{\text{ENS}}$ do. It was hypothesized that this is because of the neglect of the surface–atmosphere feedback in the quasi-offline experiment. To further investigate this possibility, the surface variables in ΔENS_{V-S} are being tested for a correlation with the differences in the state on the lowest atmospheric model level between $\text{VERTEX}_{\text{SRF}}$ and $\text{SIMPLE}_{\text{SRF}}$, calculated at the end of each time step in the quasi-offline experiments. For the sensible heat flux and surface temperature, ΔENS_{V-S} shows a weak yet statistically significant ($p \ll 0.01$) anticorrelation with the differences in atmospheric temperature and specific humidity for ΔSRF_{V-S}. A positive correlation was found for the latent heat flux, soil moisture, and precipitation. Here, the correlations with respect to the surface state variables are weaker ($|r|$ ranging between 0.16 and 0.20) and the correlations with respect to the surface fluxes and precipitation are stronger ($|r|$ ranging between 0.30 and 0.43). This indicates that strong differences between $\text{VERTEX}_{\text{ENS}}$ and $\text{SIMPLE}_{\text{ENS}}$ can be found whenever the differences in the surface processes have a strong feedback on the atmospheric processes.

In the following, the relation between surface and atmospheric processes will be investigated in more detail for two regions that exhibit opposing statistically significant surface temperature differences between $\text{VERTEX}_{\text{ENS}}$ and $\text{SIMPLE}_{\text{ENS}}$ (Fig. 5a). The first region,

TABLE 4. AWAD and global land surface mean differences.

	ΔENS_{S-P}	ΔENS_{V-P}	ΔENS_{V-S}
Temp at surface (K)	0.15 (0.09)	0.11 (0.08)	0.09 (−0.01)
Temp at 2 m (K)	0.13 (0.07)	0.10 (0.07)	0.08 (−0.00)
Surface net shortwave radiation (W m^{-2})	0.57 (0.24)	0.51 (0.25)	0.42 (0.01)
Surface net longwave radiation (W m^{-2})	0.75 (−0.54)	0.64 (−0.47)	0.39 (0.07)
Sensible heat flux (W m^{-2})	0.83 (0.39)	0.89 (0.31)	0.65 (−0.08)
Latent heat flux (W m^{-2})	1.02 (−0.68)	1.07 (−0.52)	0.67 (0.15)
Precipitation (mm day^{-1})	0.05 (−0.02)	0.05 (−0.01)	0.04 (0.01)
Vertically integrated water vapor (kg m^{-2})	0.11 (−0.07)	0.09 (−0.05)	0.08 (0.02)
Soil moisture (m)	0.009 (0.006)	0.009 (0.006)	0.004 (0.000)

which is located in western North America (42.0°–60.6°N, 99.4°–127.5°W), exhibits predominantly higher surface temperatures, larger sensible heat fluxes, and a reduction in soil moisture, evapotranspiration, and precipitation. Here, the land cover consists mainly of extratropical evergreen trees and C_3 grass, which constitute PFTs with pronounced transpirational abilities (about 74%), as well as crops and C_4 grass (about 26%; Fig. 5b). The second region is located in East Asia roughly between 36.4° and 53.2°N and 101.3° and 129.4°E. Here, mainly C_3 grass, extratropical deciduous trees, and extratropical evergreen trees constitute PFTs that transpire at high rates (about 62%), while mainly rainfed crops and C_4 grass constitute PFTs that exhibit low transpiration rates (about 38%; Fig. 5c). In the second region, the land surface exhibits predominantly lower temperatures, smaller sensible heat fluxes, and an increase in soil moisture, evapotranspiration, and precipitation for the $\text{VERTEX}_{\text{ENS}}$.

In the first region, the explicit representation of SSGS heterogeneity in the atmosphere results in relatively more humid and colder conditions in the air above the tiles with pronounced transpirational abilities (Figs. 5d,f). This reduces the atmospheric moisture demand relative to the dry tile and consequently leads to a relative reduction of the latent heat flux in the tiles with pronounced transpirational abilities (Fig. 5j). Consequently, more water remains within the soil, relative to the tiles with less pronounced transpirational abilities (Fig. 5l).

However, in region 1, both groups experience a decrease in latent heat flux and soil moisture. This is related to the relatively large eddy diffusivity within the tiles with pronounced transpirational abilities (Fig. 5h). The eddy diffusivity denotes the intensity with which properties are being mixed vertically within the atmosphere. The co-occurrence of a higher specific humidity and a larger eddy diffusivity results in a relatively intense vertical moisture transport and thus in a larger specific humidity at greater heights in the atmosphere while the atmosphere is drier closer to the surface (Fig. 6b). Additionally, the atmosphere is warmer, especially close to the surface (Fig. 6c), which has a negative effect on cloud formation (Fig. 6d). The decrease in cloud cover causes a reduction of precipitation of around $0.04\,\text{mm day}^{-1}$ and an increase in incoming shortwave radiation of about $0.8\,\text{W m}^{-2}$. This again has a negative feedback on soil moisture and evapotranspiration while it contributes to the increases in surface temperature and sensible heat flux that in turn feeds back on the temperature increase within the atmosphere.

In the second region, the latent heat flux in all tiles within the grid box is slightly larger for $\text{VERTEX}_{\text{ENS}}$ than for $\text{SIMPLE}_{\text{ENS}}$ (Fig. 5k). However, because the atmosphere is slightly more humid within the tiles with pronounced transpirational abilities (Fig. 5e) and there is a smaller surface–atmosphere moisture gradient (not shown), the average increase in latent heat flux is smaller in these tiles. This leads to a more even soil moisture distribution within the grid box, in which PFTs with pronounced transpirational abilities are less limited by water stress (Fig. 5m). Thus, averaged over the grid box and over the 20-yr period, the latent heat flux increases. As more energy is expended in the latent heat flux, less energy is available at the surface and the sensible heat flux and surface temperatures decrease (not shown).

The average eddy diffusivity is slightly lower for the PFTs with pronounced transpirational abilities (Fig. 5i). As the atmosphere within these tiles is relatively more humid, more moisture remains lower within the atmosphere for $\text{VERTEX}_{\text{ENS}}$ (Fig. 6b). In contrast the lower atmosphere in tiles with a larger eddy diffusivity is predominantly drier and warmer (Figs. 5e,g), so that dry static energy is being mixed relatively faster throughout the atmosphere. Even though the SSGS variability of the eddy diffusivity, specific humidity, and temperature is small, it results in colder and more humid conditions low within the atmosphere (Figs. 6b,c). These promote cloud cover (Fig. 6d), which in turn leads to an increase in annual mean precipitation of about $0.07\,\text{mm day}^{-1}$ and a decrease in incoming shortwave radiation of about $1.1\,\text{W m}^{-2}$. The increase in precipitation has a positive

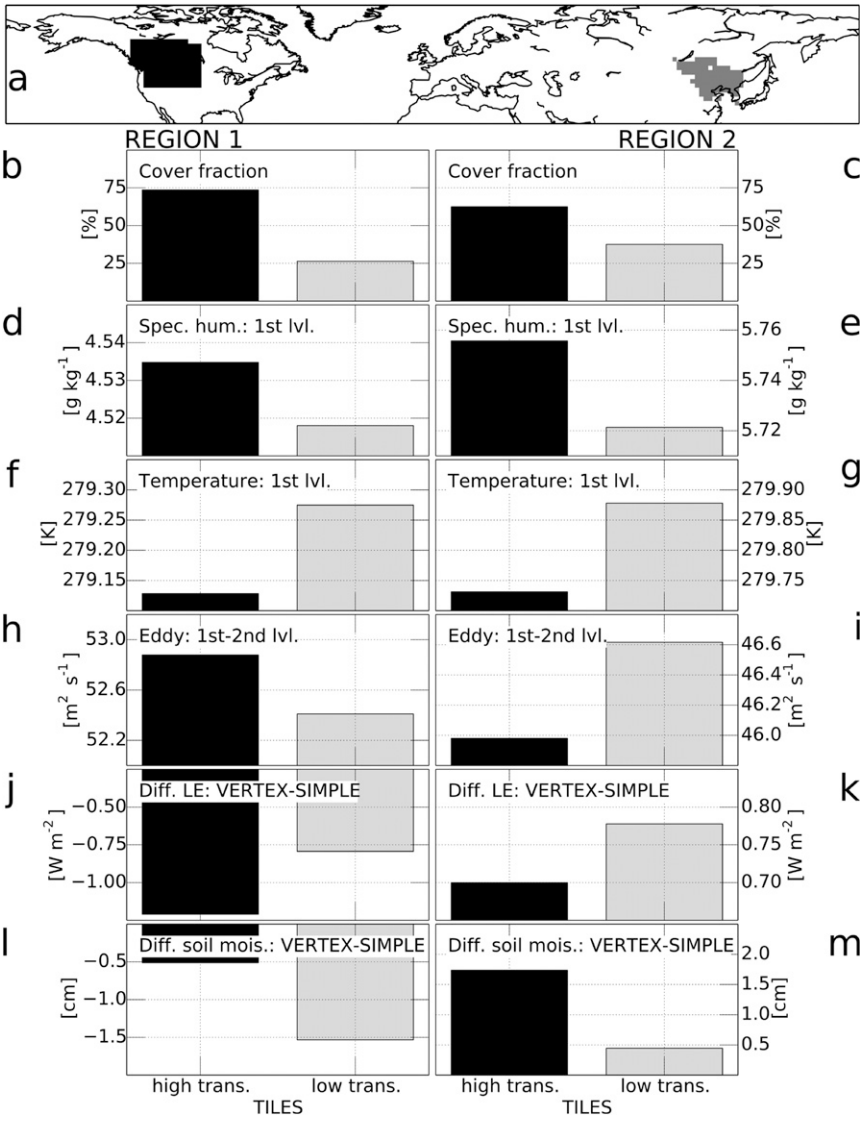

FIG. 5. (a) Focus regions 1 (black area: 42.0°–60.6°N, 99.4°–127.5°W) and 2 (gray area: 36.4°–53.2°N, 101.3°–129.4°E). (b),(c) Cover fraction; (d),(e) $\mathrm{VERTEX}_{\mathrm{ENS}}$ 20-yr mean specific humidity on the lowest atmospheric model level; (f),(g) $\mathrm{VERTEX}_{\mathrm{ENS}}$ 20-yr mean temperature on the lowest atmospheric model level; (h),(i) $\mathrm{VERTEX}_{\mathrm{ENS}}$ 20-yr mean eddy diffusivity in the lowest atmospheric layer; (j),(k) 20-yr mean $\Delta\mathrm{ENS}_{V-S}$ latent heat flux; and (l),(m) soil moisture for (left) region 1 and (right) region 2. Black bars indicate PFTs with pronounced transpirational abilities; gray bars denote PFTs with poor abilities to transpire.

feedback on soil moisture, and thus on evapotranspiration, while the reduction in incoming solar radiation reduces the energy available at the surface. This contributes to the reduction of the sensible heat flux and the surface temperature.

In the two regions, resolving the vertical turbulent transport in the atmosphere with respect to the tiles results in strongly diverging impacts on the state of the surface and the surface fluxes. Consequently, the corresponding impact on the global mean climate is much smaller than the respective impact of an explicit representation of SSGS at the surface (Table 4). When taking into account that the direction of an impact may vary regionally, the area-weighted absolute difference between two ensembles ($\mathrm{AWAD}_{\mathrm{ENS1-ENS2}}$) can be used as a measure for the similarity of two ensembles. The $\mathrm{AWAD}_{\mathrm{ENS1-ENS2}}$ is defined as the sum of the absolute difference ($|x^{\mathrm{GB}_j}_{\mathrm{ENS1}} - x^{\mathrm{GB}_j}_{\mathrm{ENS2}}|$) over all land surface grid boxes GB_j, weighted by the relative size of grid box ($A^{\mathrm{GB}_j}/\sum_{k=1}^{n} A^{\mathrm{GB}_k}$):

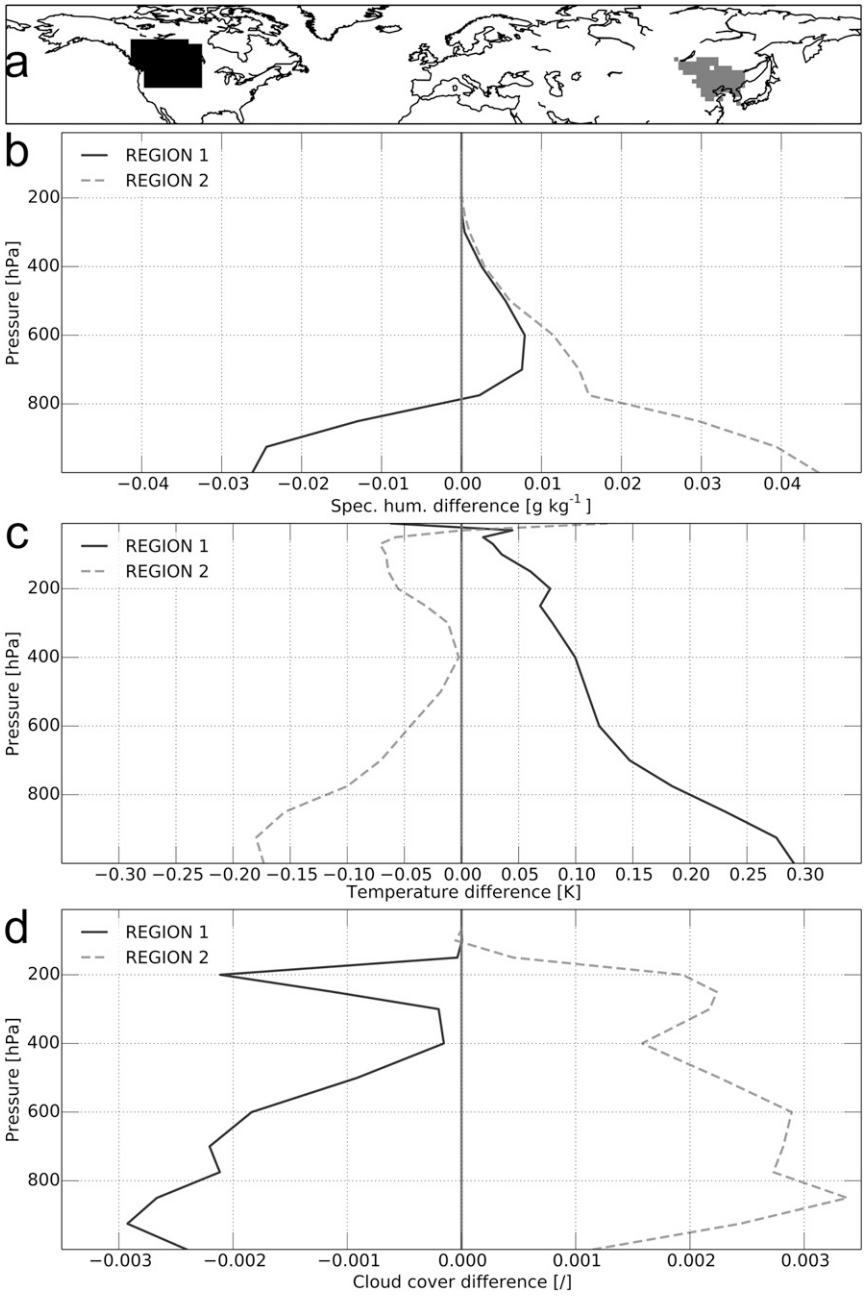

FIG. 6. (a) Focus regions 1 (black area: 42.0°–60.6°N, 99.4°–127.5°W) and 2 (gray area: 36.4°–53.2°N, 101.3°–129.4°E) and 20-yr mean $\Delta\mathrm{ENS}_{V-S}$ in (b) atmospheric specific humidity, (c) temperature, and (d) cloud cover.

$$\mathrm{AWAD}_{\mathrm{ENS1-ENS2}} = \sum_{j=1}^{n} |x_{\mathrm{ENS1}}^{\mathrm{GB}_j} - x_{\mathrm{ENS2}}^{\mathrm{GB}_j}| \left(A^{\mathrm{GB}_j} \Big/ \sum_{k=1}^{n} A^{\mathrm{GB}_k} \right). \quad (1)$$

A low AWAD indicates small differences between two ensembles, that is, a high degree of similarity. A global comparison of the different AWADs confirms that the two $\mathrm{FLUX}_{\mathrm{ENS}}$ are generally more similar than one of the $\mathrm{FLUX}_{\mathrm{ENS}}$ and the $\mathrm{PARAM}_{\mathrm{ENS}}$ (Table 4). However, the AWAD_{V-S} ranges between 60% and 80% of the AWAD_{S-P}, showing that the $\mathrm{FLUX}_{\mathrm{ENS}}$ are much more different than indicated by the comparison of the simulated global mean climate.

To evaluate whether the alterations to the standard model improve the simulated climate, the ensembles were compared to the Water and Global Change

programme's forcing data (WFD; Weedon et al. 2011). Compared to the uncertainty inherent to such a dataset, the impact of an explicit representation of SSGS heterogeneity is only of subordinate importance. In certain regions precipitation datasets often exhibit differences of several hundreds of millimeters per year (Fekete et al. 2004), whereas the differences between the ensembles rarely surpass $50\,\mathrm{mm\,yr^{-1}}$. Furthermore, the inherent model biases are an order of magnitude larger than the differences between the ensembles. For the standard operational model setup ($\mathrm{PARAM_{ENS}}$), the bias in 2-m temperature is on the scale of several kelvins (Hagemann et al. 2013), whereas the difference in 2-m temperature between any of the ensembles rarely exceeds 0.5 K. Therefore, a clear improvement of the simulated climate could not be found for any of the coupling schemes.

5. Conclusions and discussion

The representation of the link between the horizontally heterogeneous land surface and the atmosphere remains one of the key challenges in Earth system modeling. It has been argued that a realistic representation requires the consideration of SSGS variability at the surface but also within the atmosphere. To investigate the importance of SSGS heterogeneity, two surface–atmosphere coupling schemes were implemented into the MPI-ESM, in which a parameter aggregation scheme is currently being used: 1) a simple flux aggregation scheme, based on the assumption that even at the lowest model level the state of the atmosphere is horizontally homogeneous, and 2) the VERTEX scheme, in which the turbulent fluxes in the lowest atmospheric model layers are resolved with respect to the surface tiles. In a first step, the sensitivity of the near-surface processes to the representation of SSGS heterogeneity was investigated. To preclude large-scale atmospheric effects, a modified version of the offline experiment, a quasi-offline experiment, was used in which the atmospheric state is prescribed. In a second step, ensembles of global AMIP-style simulations were used to determine whether the choice of coupling scheme also has a robust impact on simulated climate.

With the quasi-offline experiments, it could be shown that the near-surface processes are systematically affected by the differences between two coupling schemes. When surface heterogeneity is explicitly represented (simple flux aggregation scheme vs parameter aggregation scheme), tiles representing plants with a high stomatal conductance and a large leaf area exhibit an increase in transpiration and a decrease in sensible heat flux, surface temperature, and soil moisture. Plants with a lower stomatal conductance and leaf area show the opposite effect. By additionally accounting for SSGS heterogeneity in the atmosphere (VERTEX scheme vs simple flux aggregation scheme), local surface–atmosphere moisture gradients are reduced in tiles with pronounced transpirational abilities while the temperature gradients increase. This results in a decrease in latent heat flux and a consequent increase in sensible heat flux. Tiles with less pronounced transpirational abilities display the opposite reaction.

The importance of the surface–atmosphere coupling was confirmed by the analysis of global AMIP-style simulations. Depending on the variable and schemes considered, between 10% and 48% of the land surface is significantly affected by the differences in the aggregation technique. Here, the largest impact on simulated climate is related to the explicit representation of SSGS heterogeneity at the surface. Simulations with a simple flux aggregation scheme generally result in higher soil moisture values, surface temperatures, and sensible heat fluxes, whereas precipitation and evapotranspiration are reduced, in comparison to simulations with a parameter aggregation scheme. But the explicit representation of SSGS heterogeneity in the atmosphere also has a distinct impact on the state of the surface and the surface fluxes. Depending on the variables and measures considered, the impact is roughly half as large as the impact of explicitly accounting for SSGS heterogeneity at the surface. Here, the impact is less unidirectional and strongly differs between regions. Even though the explicit representation of SSGS heterogeneity results in a more accurate physical model, especially for regions with strongly contrasting surface characteristics, it did not result in a clear improvement of the simulated global climate.

Finally, studies suggest that atmospheric SSGS variability may substantially affect processes such as convection and cloud formation and precipitation (Koster 2004; Rieck et al. 2014; Guillod et al. 2015). In the present study, the simulated SSGS variability on the lowest atmospheric levels is often large in comparison to the respective surface values (Fig. 7). This information on atmospheric SSGS variability could potentially be used to improve the representation of many atmospheric processes. Furthermore, the present study considered heterogeneity only for the vertical turbulent transport and on the lowest two atmospheric model levels in order to limit the level of complexity. Also, the representation of heterogeneity was limited to temperature and specific humidity whereas wind speeds were assumed to be horizontally homogeneous on and above the lowest atmospheric model level. Hence, it is plausible that atmospheric SSGS variability plays an even more important role than has been argued in this study.

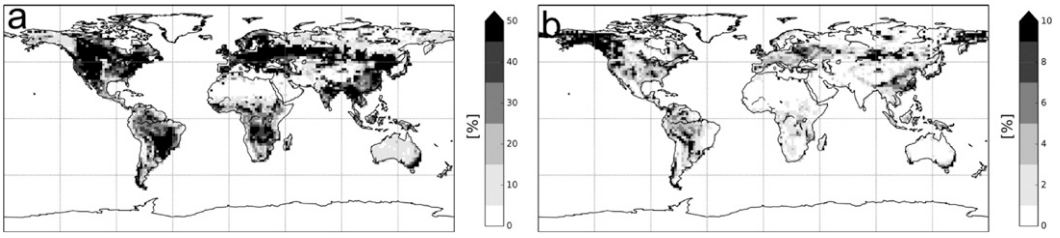

FIG. 7. Standard deviation of tile-specific temperature relative to standard deviation of tile-specific surface temperature for (a) lowest atmospheric level and (b) second-lowest atmospheric level.

However, additional model development is required to be able to include more variables and levels into the scheme. At this point the high computational demand of solving a matrix for a number of atmospheric levels and for a number of variables limits the scheme's applicability. Using the scheme for a model setup with 14 tiles (land surface and atmospheric model) increases the computational demand by almost 40% (compared to the simple flux aggregation scheme). Here, an optimization with respect to high performance machines is a prerequisite for the scheme's operational use. Additionally, the computational demand depends nonlinearly on the number of tiles used in the simulation, and for a simulation with four tiles, the two flux aggregation schemes have almost identical computational costs. Therefore, certain structural changes such as a preaggregation of tiles based on the associated blending heights may also reduce the computational demand drastically. Possible optimization strategies are currently being investigated.

Acknowledgments. The present work was supported by the International Max Planck Research School on Earth System Modeling and funding from the European Commission's 7th Framework Programme, under Grant Agreement 282672, within the EMBRACE project. Moreover, we thank Tobias Stacke, Thomas Raddatz, and Andreas Chlond from the MPI-M for very helpful discussions and advice.

REFERENCES

Ament, F., and C. Simmer, 2006: Improved representation of land-surface heterogeneity in a non-hydrostatic numerical weather prediction model. *Bound.-Layer Meteor.*, **121**, 153–174, doi:10.1007/s10546-006-9066-4.

Arino, O., J. J. Ramos Perez, V. Kalogirou, S. Bontemps, P. Defourny, and E. Van Bogaert, 2012: Global Land Cover Map for 2009 (GlobCover 2009). Pangaea, accessed 15 March 2016, doi:10.1594/PANGAEA.787668.

Arola, A., 1999: Parameterization of turbulent and mesoscale fluxes for heterogeneous surfaces. *J. Atmos. Sci.*, **56**, 584–598, doi:10.1175/1520-0469(1999)056<0584:POTAMF>2.0.CO;2.

Avissar, R., and R. A. Pielke, 1989: A parameterization of heterogeneous land surfaces for atmospheric numerical models and its impact on regional meteorology. *Mon. Wea. Rev.*, **117**, 2113–2136, doi:10.1175/1520-0493(1989)117<2113:APOHLS>2.0.CO;2.

——, and T. Schmidt, 1998: An evaluation of the scale at which ground-surface heat flux patchiness affects the convective boundary layer using large-eddy simulations. *J. Atmos. Sci.*, **55**, 2666–2689, doi:10.1175/1520-0469(1998)055<2666:AEOTSA>2.0.CO;2.

Best, M. J., A. Beljaars, J. Polcher, and P. Viterbo, 2004: A proposed structure for coupling tiled surfaces with the planetary boundary layer. *J. Hydrometeor.*, **5**, 1271–1278, doi:10.1175/JHM-382.1.

Bou-Zeid, E., C. Meneveau, and M. B. Parlange, 2004: Large-eddy simulation of neutral atmospheric boundary layer flow over heterogeneous surfaces: Blending height and effective surface roughness. *Water Resour. Res.*, **40**, W02505, doi:10.1029/2003WR002475.

Brinkop, S., and E. Roeckner, 1995: Sensitivity of a general circulation model to parameterizations of cloud–turbulence interactions in the atmospheric boundary layer. *Tellus*, **47A**, 197–220, doi:10.1034/j.1600-0870.1995.t01-1-00004.x.

Brovkin, V., T. Raddatz, C. H. Reick, M. Claussen, and V. Gayler, 2009: Global biogeophysical interactions between forest and climate. *Geophys. Res. Lett.*, **36**, L07405, doi:10.1029/2009GL037543.

Claussen, M., 1991: Estimation of areally-averaged surface fluxes. *Bound.-Layer Meteor.*, **54**, 387–410, doi:10.1007/BF00118868.

——, 1995: Flux aggregation at large scales: On the limits of validity of the concept of blending height. *J. Hydrol.*, **166**, 371–382, doi:10.1016/0022-1694(94)05098-I.

——, U. Lohmann, E. Roeckner, and U. Schulzweida, 1994: A global dataset of land surface parameters. MPI Rep. 135, Max-Planck-Institute for Meteorology, 35 pp.

Cooper, H. J., E. A. Smith, J. Gu, and S. Shewchuk, 1997: Modeling the impact of averaging on aggregation of surface fluxes over BOREAS. *J. Geophys. Res.*, **102**, 29 235–29 253, doi:10.1029/97JD00456.

de Vrese, P., J.-P. Schulz, and S. Hagemann, 2016: On the representation of heterogeneity in land-surface–atmosphere coupling. *Bound.-Layer Meteor.*, doi:10.1007/s10546-016-0133-1, in press.

Dimri, A., 2009: Impact of subgrid scale scheme on topography and landuse for better regional scale simulation of meteorological variables over the western Himalayas. *Climate Dyn.*, **32**, 565–574, doi:10.1007/s00382-008-0453-z.

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

Essery, R., M. J. Best, R. A. Betts, and C. M. Taylor, 2003: Explicit representation of subgrid heterogeneity in a GCM land

surface scheme. *J. Hydrometeor.*, **4**, 530–543, doi:10.1175/1525-7541(2003)004<0530:EROSHI>2.0.CO;2.

Feddes, R. A., P. Kabat, A. J. Doman, R. W. A. Hutjes, and M. J. Waterloo, 1998: Large-scale field experiments to improve land surface parameterisations. *Proc. Second Int. Conf. on Climate and Water*, Espoo, Finland, WMO/UNESCO, 619–647.

Fekete, B. M., C. J. Vörösmarty, J. O. Roads, and C. J. Willmott, 2004: Uncertainties in precipitation and their impacts on runoff estimates. *J. Climate*, **17**, 294–304, doi:10.1175/1520-0442(2004)017<0294:UIPATI>2.0.CO;2.

Gates, W. L., and Coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.*, **80**, 29–55, doi:10.1175/1520-0477(1999)080<0029:AOOTRO>2.0.CO;2.

Giorgetta, M. A., and Coauthors, 2013: The atmospheric general circulation model ECHAM6: Model description. MPI Rep. on Earth System Science 135, Max-Planck-Institute for Meteorology, 172 pp. [Available online at https://www.mpimet.mpg.de/fileadmin/publikationen/Reports/WEB_BzE_135.pdf.]

Giorgi, F., and R. Avissar, 1997: Representation of heterogeneity effects in earth system modeling: Experience from land surface modeling. *Rev. Geophys.*, **35**, 413–437, doi:10.1029/97RG01754.

——, R. Francisco, and J. Pal, 2003: Effects of a subgrid-scale topography and land use scheme on the simulation of surface climate and hydrology. Part I: Effects of temperature and water vapor disaggregation. *J. Hydrometeor.*, **4**, 317–333, doi:10.1175/1525-7541(2003)4<317:EOASTA>2.0.CO;2.

Guillod, B. P., B. Orlowsky, D. G. Miralles, A. J. Teuling, and S. I. Seneviratne, 2015: Reconciling spatial and temporal soil moisture effects on afternoon rainfall. *Nat. Commun.*, **6**, 6443, doi:10.1038/ncomms7443.

Hagemann, S., A. Loew, and A. Andersson, 2013: Combined evaluation of MPI-ESM land surface water and energy fluxes. *J. Adv. Model. Earth Syst.*, **5**, 259–286, doi:10.1029/2012MS20008.

Heinemann, G., and M. Kerschgens, 2005: Comparison of methods for area-averaging surface energy fluxes over heterogeneous land surfaces using high-resolution non-hydrostatic simulations. *Int. J. Climatol.*, **25**, 379–403, doi:10.1002/joc.1123.

Kabat, P., R. Hutjes, and R. Feddes, 1997: The scaling characteristics of soil parameters: From plot scale heterogeneity to subgrid parameterization. *J. Hydrol.*, **190**, 363–396, doi:10.1016/S0022-1694(96)03134-4.

Koster, R. D., 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138–1140, doi:10.1126/science.1100217.

——, and M. J. Suarez, 1992: A comparative analysis of two land surface heterogeneity representations. *J. Climate*, **5**, 1379–1390, doi:10.1175/1520-0442(1992)005<1379:ACAOTL>2.0.CO;2.

Ma, Y., M. Menenti, R. Feddes, and J. Wang, 2008: Analysis of the land surface heterogeneity and its impact on atmospheric variables and the aerodynamic and thermodynamic roughness lengths. *J. Geophys. Res.*, **113**, D08113, doi:10.1029/2007JD009124.

Mahrt, L., 2000: Surface heterogeneity and vertical structure of the boundary layer. *Bound.-Layer Meteor.*, **96**, 33–62, doi:10.1023/A:1002482332477.

——, and J. Sun, 1995: Dependence of surface exchange coefficients on averaging scale and grid size. *Quart. J. Roy. Meteor. Soc.*, **121**, 1835–1852, doi:10.1002/qj.49712152803.

Manrique-Suñén, A., A. Nordbo, G. Balsamo, A. Beljaars, and I. Mammarella, 2013: Representing land surface heterogeneity: Offline analysis of the tiling method. *J. Hydrometeor.*, **14**, 850–867, doi:10.1175/JHM-D-12-0108.1.

Mason, P. J., 1988: The formation of areally-averaged roughness lengths. *Quart. J. Roy. Meteor. Soc.*, **114**, 399–420, doi:10.1002/qj.49711448007.

Molod, A., and H. Salmun, 2002: A global assessment of the mosaic approach to modeling land surface heterogeneity. *J. Geophys. Res.*, **107**, 4217, doi:10.1029/2001JD000588.

——, ——, and D. W. Waugh, 2004: The impact on a GCM climate of an extended mosaic technique for the land–atmosphere coupling. *J. Climate*, **17**, 3877–3891, doi:10.1175/1520-0442(2004)017<3877:TIOAGC>2.0.CO;2.

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

Patton, E. G., P. P. Sullivan, and C.-H. Moeng, 2005: The influence of idealized heterogeneity on wet and dry planetary boundary layers coupled to the land surface. *J. Atmos. Sci.*, **62**, 2078–2097, doi:10.1175/JAS3465.1.

Polcher, J., K. Laval, L. Dümenil, J. Lean, and P. Rowntree, 1996: Comparing three land surface schemes used in general circulation models. *J. Hydrol.*, **180**, 373–394, doi:10.1016/0022-1694(95)02886-2.

——, and Coauthors, 1998: A proposal for a general interface between land surface schemes and general circulation models. *Global Planet. Change*, **19**, 261–276, doi:10.1016/S0921-8181(98)00052-6.

Raddatz, T., and Coauthors, 2007: Will the tropical land biosphere dominate the climate–carbon cycle feedback during the twenty-first century? *Climate Dyn.*, **29**, 565–574, doi:10.1007/s00382-007-0247-8.

Raupach, M., and J. Finnigan, 1995: Scale issues in boundary-layer meteorology: Surface energy balances in heterogeneous terrain. *Hydrol. Processes*, **9**, 589–612, doi:10.1002/hyp.3360090509.

Rieck, M., C. Hohenegger, and C. C. van Heerwaarden, 2014: The influence of land surface heterogeneities on cloud size development. *Mon. Wea. Rev.*, **142**, 3830–3846, doi:10.1175/MWR-D-13-00354.1.

Salmun, H., A. Molod, and A. Ira, 2007: Observational validation of an extended mosaic technique for capturing subgrid scale heterogeneity in a GCM. *Tellus*, **59B**, 625–632, doi:10.1111/j.1600-0889.2007.00257.x.

Schomburg, A., V. Venema, F. Ament, and C. Simmer, 2012: Disaggregation of screen-level variables in a numerical weather prediction model with an explicit simulation of subgrid-scale land-surface heterogeneity. *Meteor. Atmos. Phys.*, **116**, 81–94, doi:10.1007/s00703-012-0183-y.

Sellers, P., 1991: Modeling and observing land-surface–atmosphere interactions on large scales. *Surv. Geophys.*, **12**, 85–114, doi:10.1007/BF01903413.

Seth, A., F. Giorgi, and R. E. Dickinson, 1994: Simulating fluxes from heterogeneous land surfaces: Explicit subgrid method employing the Biosphere–Atmosphere Transfer Scheme (BATS). *J. Geophys. Res.*, **99**, 18 651–18 667, doi:10.1029/94JD01330.

Stevens, B., and Coauthors, 2013: Atmospheric component of the MPI-M Earth System Model: ECHAM6. *J. Adv. Model. Earth Syst.*, **5**, 146–172, doi:10.1002/jame.20015.

van den Hurk, B., and A. Beljaars, 1996: Impact of some simplifying assumptions in the new ECMWF surface scheme. *J. Appl. Meteor.*, **35**, 1333–1343, doi:10.1175/1520-0450(1996)035<1333:IOSSAI>2.0.CO;2.

Weedon, G. P., and Coauthors, 2011: Creation of the WATCH Forcing Data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *J. Hydrometeor.*, **12**, 823–848, doi:10.1175/2011JHM1369.1.