AGU100 ADVANCING EARTH AND SPACE SCIENCE

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1002/2018MS001284

Evaluating the Interplay Between Biophysical Processes and Leaf Area Changes in Land Surface Models

Giovanni Forzieri[1], Gregory Duveiller[1], Goran Georgievski[2], Wei Li[3], Eddy Robertson[4], Markus Kautz[5], Peter Lawrence[6], Lorea Garcia San Martin[1], Peter Anthoni[5], Philippe Ciais[3], Julia Pongratz[2,7], Stephen Sitch[8], Andy Wiltshire[4], Almut Arneth[5], and Alessandro Cescatti[1]

[1]European Commission, Joint Research Centre, Ispra, Italy, [2]MPI, Hamburg, Germany, [3]LSCE IPSL, Gif sur Yvette, France, [4]UK Met Office, Exeter, UK, [5]KIT, Garmisch-Partenkirchen, Germany, [6]NCAR, Boulder, Colorado, USA, [7]Now at Ludwig-Maximilians-Universität München, Munich, Germany, [8]University of Exeter, Exeter, UK

Abstract Land Surface Models (LSMs) are essential to reproduce biophysical processes modulated by vegetation and to predict the future evolution of the land-climate system. To assess the performance of an ensemble of LSMs (JSBACH, JULES, ORCHIDEE, CLM, and LPJ-GUESS) a consistent set of land surface energy fluxes and leaf area index (LAI) has been generated. Relationships of interannual variations of modeled surface fluxes and LAI changes have been analyzed at global scale across climatological gradients and compared with those obtained from satellite-based products. Model-specific strengths and deficiencies were diagnosed for tree and grass biomes. Results show that the responses of grasses are generally well represented in models with respect to the observed interplay between turbulent fluxes and LAI, increasing the confidence on how the LAI-dependent partition of net radiation into latent and sensible heat are simulated. On the contrary, modeled forest responses are characterized by systematic bias in the relation between the year-to-year variability in LAI and net radiation in cold and temperate climates, ultimately affecting the amount of absorbed radiation due to LAI-related effects on surface albedo. In addition, for tree biomes, the relationships between LAI and turbulent fluxes appear to contradict the experimental evidences. The dominance of the transpiration-driven over the observed albedo-driven effects might suggest that LSMs have the incorrect balance of these two processes. Such mismatches shed light on the limitations of our current understanding and process representation of the vegetation control on the surface energy balance and help to identify critical areas for model improvement.

Key Points:
- The covariability of land biophysics and changes in vegetation density predicted by Land Surface Models is compared with satellite data
- Biophysical properties of vegetation are explored across climatological gradients of temperature and precipitation
- Model-specific and systematic strengths and deficiencies are diagnosed separately for tree and grass biomes

Supporting Information:
- Supporting Information S1

Correspondence to:
G. Forzieri,
giovanni.forzieri@ec.europa.eu

Citation:
Forzieri, G., Duveiller, G., Georgievski, G., Li, W., Robertson, E., Kautz, M., et al. (2018). Evaluating the interplay between biophysical processes and leaf area changes in land surface models. *Journal of Advances in Modeling Earth Systems*, *10*, 1102–1126. https://doi.org/10.1002/2018MS001284

Received 30 JUN 2017
Accepted 30 MAR 2018
Accepted article online 3 APR 2018
Published online 6 MAY 2018

1. Introduction

Variations in the structure and density of vegetation influence land surface fluxes by affecting the water and energy exchange between the Earth surface and the atmosphere (Arora & Montenegro, 2011; McPherson, 2007). Leaf area in particular plays a key role in determining resistances to heat, moisture, and momentum transfer (Bright et al., 2015). Its representation has been included as a key prognostic variable in the last generation of Land Surface Models (LSMs) and its interactions with biophysics, hydrology, and biogeochemistry are represented through equations of different complexity (Bonan, 2008; Piao et al., 2013). Being key components of Earth System Models, LSMs are now widely applied as supporting tools for policy relevant scientific assessment of climate change and its impact on terrestrial ecosystems (Arora et al., 2011; IPCC, 2013; Quéré et al., 2018). However, there is still limited knowledge of the LSMs performance in representing the net effect of changes in vegetation density on the climate system (de Noblet-Ducoudré et al., 2012; Zhang et al., 2016). This is particularly relevant in light of the observed global greening (Zhu et al., 2016) that may impact on regional climates through changes in leaf area index (LAI) (Forzieri et al., 2017; Zeng et al., 2017). Evaluating LSMs for the interplay between the surface energy budget and vegetation density (leaf area) is therefore essential to reduce their uncertainty and improve the representation of biophysical effects.

In the last decade several international initiatives have been developed to define new experiments and tools to objectively assess Land Surface Models. Along this line, the Carbon-Land Model Intercomparison Project (C-LAMP; Randerson et al., 2009), the International Land-Atmosphere Model Benchmarking project (ILAMB; Luo et al., 2012), the LandFlux-EVAL project (Jiménez et al., 2011; Mueller et al., 2011), and the Land

Surface Model Benchmarking Evaluation Project (PLUMBER; Best et al., 2015) have contributed significantly to lay the basis in this research field. Recently, the Protocol for the Analysis of Land Surface Models (PALS)—a free online Land Surface Model application—has been developed to assist in LSM evaluation (Abramowitz, 2012).

Direct observations available from field experiments (Sellers et al., 1997) and flux tower data (Baldocchi et al., 2001) are widely utilized for model performance evaluation (e.g., Best et al., 2015; Blyth et al., 2009; Williams et al., 2009). However, surface observations, while generally considered the most reliable reference, are limited to few selected locations and can only partially characterize the spatiotemporal variability of the global processes reproduced in LSMs. Global gridded data sets derived by up-scaling ground observations via statistical techniques (e.g., Jung et al., 2009; Tramontana et al., 2016) may solve the above-mentioned issue. However, such products have shown limitations in capturing the spatial and temporal variability of biophysical processes, thus hampering the analysis at large scales and over multiple years (Anav et al., 2015; Marcolla et al., 2017; Schimel et al., 2015). Model performance has been also evaluated against standardized experiments of well-accepted models (Dai et al., 2003), previous versions of the models or the ensemble mean of multiple models (Chen et al., 1997). However, such evaluations implicitly neglect the possibility that the considered reference model or experiment might be biased (Best et al., 2015). Other approaches compare model outputs with statistically based model results (Abramowitz, 2005). These approaches offer valuable advantages, such as the possibility to quantify the effective use of the available information used in models (Blyth et al., 2011), yet they have been applied only for specific environmental conditions. The increasing availability of extensive satellite Earth observations represents an important additional source of information to retrieve biophysical land and climate parameters from multiple observations, or from the combination of observations with relatively simple or empirically derived model formulations. Such data-driven products represent valuable tools to derive observation-based diagnostics that may serve as reference across different spatial and temporal scales (Kelley et al., 2013; Maignan et al., 2011; Randerson et al., 2009).

Existing evaluation systems of vegetation processes have mostly analyzed the model capability in reproducing single targeted processes or variables. They typically focused on assessing model performance on phenology and dynamics of vegetation greenness (Forkel et al., 2014; MacBean et al., 2015; Murray-Tortarolo et al., 2013). As example, an intermodel comparison study showed an overall tendency of LSMs to overestimate annual average LAI, seasonal amplitude and length of the growing season (Murray-Tortarolo et al., 2013). While these approaches are useful to quantify the model accuracy in the magnitude of the simulated vegetation metrics, they do not support a broader model evaluation where multiple processes and variables may covary. Such limitations inevitably hamper a comprehensive assessment of the modeled interplay between vegetation and biophysics. To this respect, evaluating Land Surface Models on functional relationships between variables and across climatological gradients may be particularly effective, especially in light of the key role of background climate conditions in modulating water and energy exchanges (Forzieri et al., 2017; Pitman et al., 2011).

Disentangling vegetation-mediated biophysical processes in LSMs is not straightforward, especially when they are coupled with climate models whose signal (and biases) may largely determine the response of the land surface (Blyth et al., 2011; Winckler et al., 2016). Furthermore, differences in reference plant functional types (PFTs) and climate forcing across LSMs may hinder the development of a consistent comparison. In order to reduce potential methodological biases, evaluating multiple LSMs on their capacity to reproduce vegetation-mediated land-atmosphere interactions requires the use of a common modeling framework.

In this study, we present a novel evaluation framework to assess model capacity in reproducing land surface-atmosphere energy exchanges modulated by vegetation density. Through a collaborative effort of different modeling groups developed within the LUC4C project (Land use change: assessing the net climate forcing, and options for climate change mitigation and adaptation, http://luc4c.eu/), a consistent set of land surface energy fluxes and LAI dynamics has been generated from multiple LSMs. Relationships of interannual covariations of modeled surface fluxes and LAI changes have been analyzed at global scale across climatological gradients and compared with satellite-based products. A set of scoring metrics is used to assess the overall models' performance, and a detailed analysis in the climate space is provided to diagnose model strengths and deficiencies associated to background climate conditions. Results may help to guide future developments for improved representation of biophysical processes in land-atmosphere models.

2. Materials and Methods

2.1. Observation-Based Products

Multiple remote sensing products were employed to derive observation-based information for the interplay between vegetation and biophysics. The chosen LAI and energy fluxes data sets are independent of any LSM considered (not utilized for model's calibration or development), have global (or quasi global) coverage, span over multiple years, have spatial and temporal resolutions suitable for the purposes of the study with documented high accuracies.

Net radiation (RN) was retrieved from the Clouds and the Earth's Radiant Energy System (CERES SYN1deg, http://ceres.larc.nasa.gov/) (Wielicki et al., 1996) onboard Terra, Aqua, and the Tropical Rainfall Measuring Mission (TRMM). Latent heat (LE) was derived from observational-based evapotranspiration data generated from the Global Land Evaporation Amsterdam Model Version 2b (GLEAM v2B, http://www.gleam.eu/) (Martens et al., 2016; Miralles et al., 2011) and accounting for latent heat of vaporization (Dingman, 2015). The sum of sensible and ground heat fluxes (H + G) was obtained from the closure of the energy balance by subtracting LE from RN. Daily means of energy fluxes were computed after a preliminary gap-filling processing applied via temporal linear interpolation (Miralles et al., 2011).

In order to characterize the vegetation density, we used the LAI product derived from the Global Inventory Modeling and Mapping Studies Normalized Difference Vegetation Index (GIMMS3g, http://cliveg.bu.edu/modismisr/lai3g-fpar3g.html) (Zhu et al., 2013) that was based on the combination of the Advanced Very High Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS) sensors. This product was preferred with respect to alternative LAI products (e.g., GLASS LAI) for its improved harmonization on multisensor records (Forzieri et al., 2017).

Climate dynamics over the last decades were derived from global gridded precipitation (P) and air temperature (T) data obtained from the CRU-NCEP product based on the interpolation of climate station records and reanalysis data (version 6, https://crudata.uea.ac.uk/cru/data/ncep/).

In addition, we used data sets of PFTs derived from the 2010 land cover map (2008–2012 reference period) of the European Space Agency's Climate Change Initiative (ESA-CCI, https://www.esa-landcover-cci.org/). The procedure to pass from categorical land cover classes to continuous PFT maps involved the cross-walking procedure described in Poulter et al. (2015), with the exception that we used a simplified table to generate seven PFTs including: tree broadleaf evergreen, tree broadleaf deciduous, tree needleleaf (evergreen and deciduous), shrub, natural grass, managed grass, and bare soil (supporting information Figure S1). We derived the respective cover fraction for each PFT. A further aggregation to two broad PFT classes based on physiognomy alone was used here to synthesize results: grass (including natural and managed grasses) and trees (including broadleaf and needleleaf trees) (Figure 1). Finally, we used the Global Map of Irrigation Areas, GMIA (http://www.fao.org/nr/water/aquastat/irrigationmap/index.stm) to identify areas

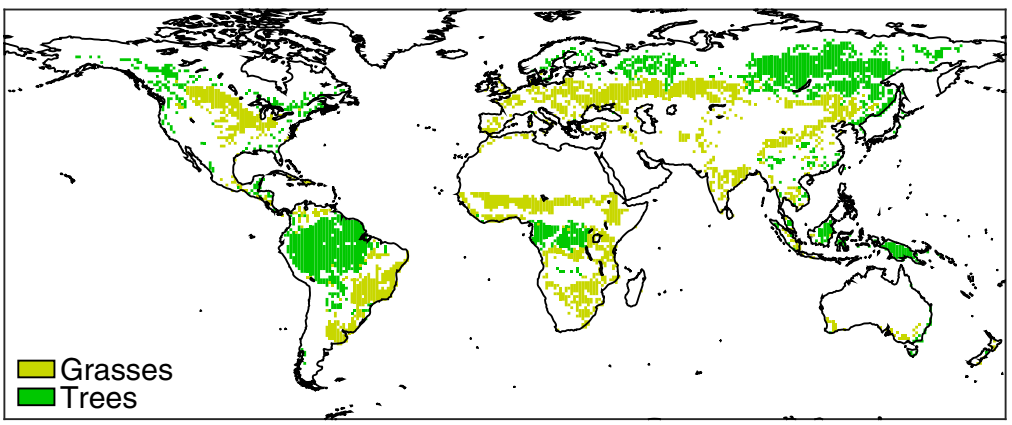

Figure 1. Spatial domains of grasses and trees. Selection of pixels used in this work, which have a cover fraction ≥60% for the broad PFT classes "grasses" and "trees" derived from the ESA-CCI land cover map and with <10% of irrigated area derived from the GMIA product.

Table 1
Observation-Based Products Used to Derive Process-Oriented Diagnostics

Variable	Product	Temporal resolution	Temporal coverage	Domain	Spatial resolution
Net radiation	CERES	3 hourly	2000–2015	Global	1°
Evapotranspiration	GLEAM	Daily	2000–2015	50°S–50°N	0.25°
Leaf area index	GIMMS3g	15 day	1982–2011	Global	1/12°
Precipitation	CRU-NCEP	Daily	1901–2014	Global	0.5°
Air temperature	CRU-NCEP	Daily	1901–2014	Global	0.5°
Plant functional type	ESA-CCI	Static	2008–2012	Global	300 m
Irrigated area fraction	GMIA	Static	2005	Global	5 min

affected by irrigation. This map was derived from statistical census data and provides the percentage of irrigated area per grid cell in 2005. Spatial and temporal characteristics of the observation-based products are listed in Table 1.

2.2. Land Surface Simulations

We assess the performance of five LSMs: JSBACH (Raddatz et al., 2007), JULES (Best et al., 2011; Clark et al., 2011), ORCHIDEE (Krinner et al., 2005), CLM version 4.5 (Lawrence et al., 2011), and LPJ-GUESS (Smith et al., 2014). LSMs are mechanistic models of the land surface and simulate the water, energy, and carbon exchanges between the atmosphere and the Earth surface, and are designed to be incorporated into Earth system models. Amongst the LSMs tested, only CLM and LPJ-GUESS incorporate carbon-nitrogen interactions and only LPJ-GUESS includes competition amongst PFTs.

We used a dedicated set of LSM simulations forced with historic climatic CRU-NCEP data and CO_2 concentrations consistent with the TRENDY S2 protocol (Sitch et al., 2015). A 500 years spin-up period was used to balance the biomass pools and LAI, then transient climate and CO_2 from 1901 onward. Simulations for the period 2000–2014 were used for the model performance evaluation (CO_2 concentrations raised from 369 to 399 ppm over the reference period). To better disentangle the dynamics associated to the land surface processes from the climate signal, and to isolate the effect of individual PFTs, the runs considered an idealized configuration with grid cells containing only fully homogeneous vegetation cover. These PFT-specific model runs were used to overcome the limitation of LSMs that typically do not compute the energy budget for each PFT in a grid cell and instead output the average surface fluxes of the mixture of coexisting PFTs. Similarly, LSMs do not generally represent differences in soil hydrology for each PFT, instead a single soil column is modeled in each grid cell and all PFTs interact with this hydrology (Schultz et al., 2016). On the contrary, in our model experiment, the different PFTs have their specific soil column and the energy fluxes are not affected by soil moisture feedbacks from other PFTs. Ultimately, modeling pure PFTs and then recomposing them according to the actual fractions of the vegetation distribution allows to better isolate the signal of the vegetation cover on the surface energy budget. Following this logic, we performed as many runs as model PFTs, each providing the dynamics—prognostically calculated at hourly time step—of the land surface energy fluxes (RN, LE, and H + G) and LAI (note LPJ-GUESS does not explicitly solve the land surface energy balance and only provides daily values for LE and LAI). Model output were provided in equally spaced 0.5° geographic grid (JSBACH, CLM, LPJ-GUESS), in 1° Gaussian grid (ORCHIDEE) or in 1.875° × 1.25° (longitude × latitude) grid (JULES). Single PFT model runs were then aggregated to the 7 PFT class maps according to ESA-CCI (6 PFT class maps for models with no shrubs class) (Table 2). These maps incorporate climatic information based on the Köppen-Geiger classification to distinguish tropical, temperate, arid, and boreal types (Kottek et al., 2006), as well as different photosynthetic pathways as derived from the global distribution of C3 and C4 vegetation (Knorr & Heimann, 2001). The implicit assumptions with these simulations are that (1) the energy and water fluxes of a given PFT in a pure cell are comparable with those of the same PFT in a mixed cell and (2) potential feedbacks due to the land-atmosphere interactions were excluded since all simulations were forced with observed climate. More details on the model setup protocol are described in Duveiller et al. (2018).

From this data set, we created a representation of the land surface fluxes and LAI in the last decade accounting for the current mixture of diverse PFTs on the Earth surface as derived from the ESA-CCI land cover map. To this aim, a weighted sum of the PFT-specific runs was computed, separately for each model

Journal of Advances in Modeling Earth Systems 10.1002/2018MS001284

Table 2
Look-Up Table Used to Map Model-Specific PFTs into the 7 PFT Class Maps Derived From ESA-CCI

ESA-CCI PFTs	JSBACH	JULES	CLM	ORCHIDEE	LPJ-GUESS[d]
Broadleaf evergreen trees	Tropical evergreen trees are bounded in areas where either deciduous or evergreen trees are tropical in the default PFT distribution. Extra-tropical evergreen trees populate the remaining vegetated areas.[a]	Broadleaf trees	Broadleaf evergreen trees	Tropical broadleaf evergreen trees are bounded in tropical areas. Temperate broadleaf evergreen trees populate the remaining vegetated areas.[a]	Run with grass, temperate evergreen broadleaf, tropical evergreen broadleaf and tropical evergreen shade-intolerant broadleaf competing against each other. All other PFTs removed.
Broadleaf deciduous trees	Tropical deciduous trees are bounded in areas where either deciduous or evergreen trees are tropical in the default PFT distribution. Extra-tropical deciduous trees populate the remaining vegetated areas.[a]	Broadleaf trees	Broadleaf deciduous trees	Tropical broadleaf deciduous trees are bounded in tropical areas. Temperate broadleaf summer trees are bounded in arid and temperate areas. Boreal broadleaf summer green trees populate the remaining vegetated areas.[a]	Run with grass, tropical raingreen broadleaf, temperate deciduous broadleaf and temperate deciduous shade-intolerant broadleaf competing against each other. All other PFTs removed.
Needleleaf evergreen/ deciduous trees	Coniferous deciduous trees where dominant according to the default PFT distribution. Coniferous evergreen trees populate the remaining vegetated areas.	Needleleaf trees	Needleleaf trees	Boreal needleleaf summer green trees where dominant according to the default PFT distribution. Boreal needleleaf evergreen trees populate the remaining boreal and polar areas. Temperate needleleaf evergreen trees populate the remaining tropical, temperate and arid areas.[a]	Run with grass, boreal evergreen needleleaf, boreal evergreen shade-intolerant needleleaf and boreal deciduous needleleaf competing against each other. All other PFTs removed.
Shrubs	Deciduous shrubs are bounded as in the default PFT distribution. Raingreen shrubs populate the remaining vegetated areas.	Shrubs	Shrubs	No class[c]	No class[c]
Natural grasses	C3 grass or C4 grass, based on the dominant photosynthetic pathways[b]	C3 grass or C4 grass, based on the dominant photosynthetic pathways[b]	C3 grass or C4 grass, based on the dominant photosynthetic pathways[b]	C3 grass or C4 grass, based on the dominant photosynthetic pathways[b]	Run with C3 grass and C4 grass competing against each other. All other PFTs removed.
Managed grasses	C3 crop or C4 crop, based on the dominant photosynthetic pathways[b]	C3 grass or C4 grass, based on the dominant photosynthetic pathways[b]	Crop	C3 crop or C4 crop, based on the dominant photosynthetic pathways[b]	Weighted average of single runs with crops maize, crops maize irrigated, crops summer wheat, crops summer wheat irrigated, crops winter wheat, crops winter wheat irrigated based on their respective cover fraction. All other PFTs removed.
Bare soil	Bare soil	Bare soil	Bare soil	Bare soil	Bare soil. All other PFTs removed.

[a]Climatic information is retrieved from the Köppen-Geiger classification (Kottek et al., 2006) to distinguish arid, tropical, temperate, and boreal climate zones. [b]Different photosynthetic pathways are derived from the global distribution of C3 and C4 vegetation (Knorr & Heimann, 2001). [c]ORCHIDEE and LPJ-GUESS refer to the six PFT class maps derived from ESA-CCI because the shrubs class is not represented in the models. [d]Since LPJ-GUESS works with various PFTs competing to simulate a given ecosystems, several PFTs have been simultaneously used for each vegetation type. Model-specific nomenclature of PFT is used.

and variable, and based on the cover fractions of each grid cell. The objective of this approach was to minimize the intermodel spread due to the model-specific PFT distributions and to the different strategies used to represent subgrid land heterogeneity across models. The additive composition of the PFT-specific simulations implicitly assumed independent dynamics of each PFT. Note that since LPJ-GUESS works with various PFTs competing to simulate a given ecosystems, several PFTs have been simultaneously used for each vegetation type. The integration of competition amongst PFTs makes LPJ-GUESS output different from those of the other models; however, the specific model LPJ-GUESS runs performed were designed to allow for intermodel comparability (Table 2). We believe that these differences do not affect our results and conclusions.

2.3. Evaluation Framework

The proposed evaluation framework focused on the covariation between LAI and the terms of the energy balance equation:

$$RN = LE + (H + G), \tag{1}$$

In order to assess the interplay between biophysical processes and changes in vegetation density, we analyzed the year-to-year variations in each variable of equation (1) with respect to year-to-year variations in LAI. Such approach based on the difference between two consecutive years (Δ operator) preserved the resulting features from possible long-term dependencies on covariates (e.g., the combined effect of rising temperatures and CO_2 concentrations on long-term LAI trends).

Relations between ΔZ (where Z is one of the terms in equation (1)) and ΔLAI were explored across climatological gradients of temperature and precipitation. For this purpose ΔZ, ΔLAI and the climatological temperature (or precipitation) were calculated for each pixel and for each year-to-year time step. The set of multiyear values derived from the global domain was used to characterize the bioclimatic spaces representing the vegetation-biophysics relations under average climate conditions. The bioclimatic spaces are built by binning ΔZ values as a function of ΔLAI and the climatological median of air temperature (T-space) or precipitation (P-space) over a predefined grid with cell size of 1/30 of the range of variability corresponding to bands of $\sim$0.07 m^2 m^{-2}, $\sim$1.2°C, and $\sim$130 mm, respectively. Only bins with at least five values were accounted for and the corresponding median was considered representative of the given bin. The analysis was performed for annual and monthly average values from both remote sensing and land surface simulations, after a preliminary resampling of the data to the common 1° spatial resolution and masking over the same spatial domain. For the monthly scale investigation, the Δ values are calculated as difference between the same months of two consecutive years. To quantify the uncertainty in the observed bioclimatic spaces of given vegetation type and land surface flux, the *t* test was applied to the ΔZ distribution of each bin to assess the statistical significance of the difference from a 0-mean distribution. Similarly, ΔZ values derived from remote sensing data and model outputs were analyzed for each bin of the bioclimatic spaces by the Welch's unequal variances *t* test to assess the statistical significance of the differences between the two samples. To facilitate the integration of data from areas with largely different climate and LAI, the interannual differences were expressed in relative terms as well, i.e., divided by the local multiyear median.

In order to assess possible vegetation-specific patterns, the bioclimatic spaces were separately examined for the two broad PFT classes of grasses and trees. These classes were selected based on the simplified PFT maps in which pixels of grasses or trees covered at least 60% of the grid cell area and comprised less than 10% of irrigated area (Figure 1). We masked areas substantially affected by irrigation to better disentangle the observed LAI-biophysics relation across the precipitation gradient. To assess the robustness of our results to different degrees of pixel homogeneity and irrigation, bioclimatic spaces were also derived using cells with cover fractions ranging from 50% to 90% and irrigation fractions from 10% to 50%; areas with cover fraction of ice and water larger than 5% were masked out in all experiments.

Model performance was assessed by comparing the bioclimatic spaces derived from simulations with those derived from remote sensing through a set of scoring metrics. The percent bias (PBIAS) was employed to quantify the average tendency of the simulated absolute values to be larger or smaller than observed absolute values (Gupta et al., 1999). The optimal value of PBIAS is 0, with low-magnitude values indicating accurate model simulation. Positive values (i.e., model outputs larger than observations) indicate overestimation bias, whereas negative values indicate model underestimation bias. The Root Mean Square Error (RMSE)

was utilized to measure the magnitude of the deviation between model and observations, while the Spearman coefficient (ρ) was used to assess the degree of spatial correlation between model output and observations. Scoring metrics were calculated by comparing pairs of modeled and observed ΔZ values derived from each bin of a given bioclimatic space. They integrate the model-data displacement over the whole climatological space and the full spectrum of LAI variability and allow identifying the agreement between observed and modeled sensitivities of net radiation, latent fluxes, and sensible and ground fluxes to variation in LAI across the different climates.

In order to retrieve spatially explicit information on the observed/modeled relation between energy fluxes and vegetation, temporal correlation maps between ΔZ and ΔLAI were computed in terms of Spearman rank, quantified for each pixel over a centered $3^\circ \times 3^\circ$ spatial window.

We point out that evaluating model performance with respect to functional relationships as derived from the bioclimatic spaces is particularly effective given the unknown local-scale uncertainties in data and the potential biases induced by the climate forcings in the modeled land surface responses (Luo et al., 2012; Randerson et al., 2009). This is particularly relevant in light of the differences in interannual variability observed across multiple LAI products (Jiang et al., 2017). In order to evaluate the functional relationships with respect to their single drivers, the afore-mentioned assessment was complemented with an evaluation on model capability to reproduce the year-to-year variability of single variables (ΔLAI_S, ΔRN_S, ΔLE_S, and $\Delta(H + G)_S$) against remote sensing estimates in terms of PBIAS, RMSE, and ρ. For this exercise grid cell values were used.

3. Results

3.1. Observation-Based and Model-Based Vegetation-Biophysics Interplay Across Climatological Gradients

We first focus on the observed bioclimatic spaces for trees and grasses (Figure 2). Then, we assess model performance in representing the interannual variability of the single variables (Figure 3) and the covariability of multiple variables as represented by the bioclimatic spaces (Figures 4–7) separately for each model and vegetation biome. In order to identify model strengths and deficiencies with respect to different climate background conditions, each modeled bioclimatic space is analyzed and compared with observations across a range of different temperature and precipitation fields. Model limitations and mechanisms potentially responsible of data-model discrepancies are diagnosed and synthesized in Table 3. The emphasis of the model-observation comparison is on the role of the background climate in mediating vegetation-biophysics relations. Temporal correlation maps (Figure 8) help to geolocalize the emerging patters from the bioclimatic spaces.

3.1.1. Remote Sensing

Tree biomes. In cold-temperate and boreal tree biomes ($T < 10°C$) and mostly under dry regime ($P < 1{,}000$ mm) observed interannual increases in LAI are associated with an increase in net radiation (Figures 2a and 2d, upper-left corner, p-value ≤ 0.05). The interplay likely results from the enhancement in absorbed radiation due to an increased masking of snow by vegetation and the consequent reduction of surface albedo. In warmer regions ($T > 20°C$), similar increases in LAI are moderately associated with positive variations in net radiation (p-value ≤ 0.05) and are partially reflected in the turbulent fluxes, prominently in latent heat (Figures 2b and 2c). Such positive relations between LAI and energy fluxes emerge distinctly from the temporal correlation maps over spatially consistent areas of boreal and tropical forests (Figures 8a–8c). In high and moderately high precipitation regimes ($P > 2{,}000$ mm), there is no clear relationship between LAI and the turbulent fluxes (Figures 2e and 2f), which suggests that LAI does not control the latent heat flux and the high availability of surface water reduces the impact of vegetation on turbulent fluxes. Overall, observed features of net radiation in temperate climates and turbulent fluxes over most of the observed climatological gradients should be critically evaluated in light of the considerable uncertainty of the emerging signal.

Grass biomes. For grasslands in warm climates ($T > 20°C$), the observed increase in net radiation in combination with rising LAI (Figure 2g, upper right corner) promotes latent heat fluxes (Figure 2h) thanks to the high sensitivity of canopy conductance at moderate and low precipitation levels ($P < 1{,}500$ mm, Figure 2k). As expected, sensible and ground heat fluxes show opposite patterns with respect to latent heat since they

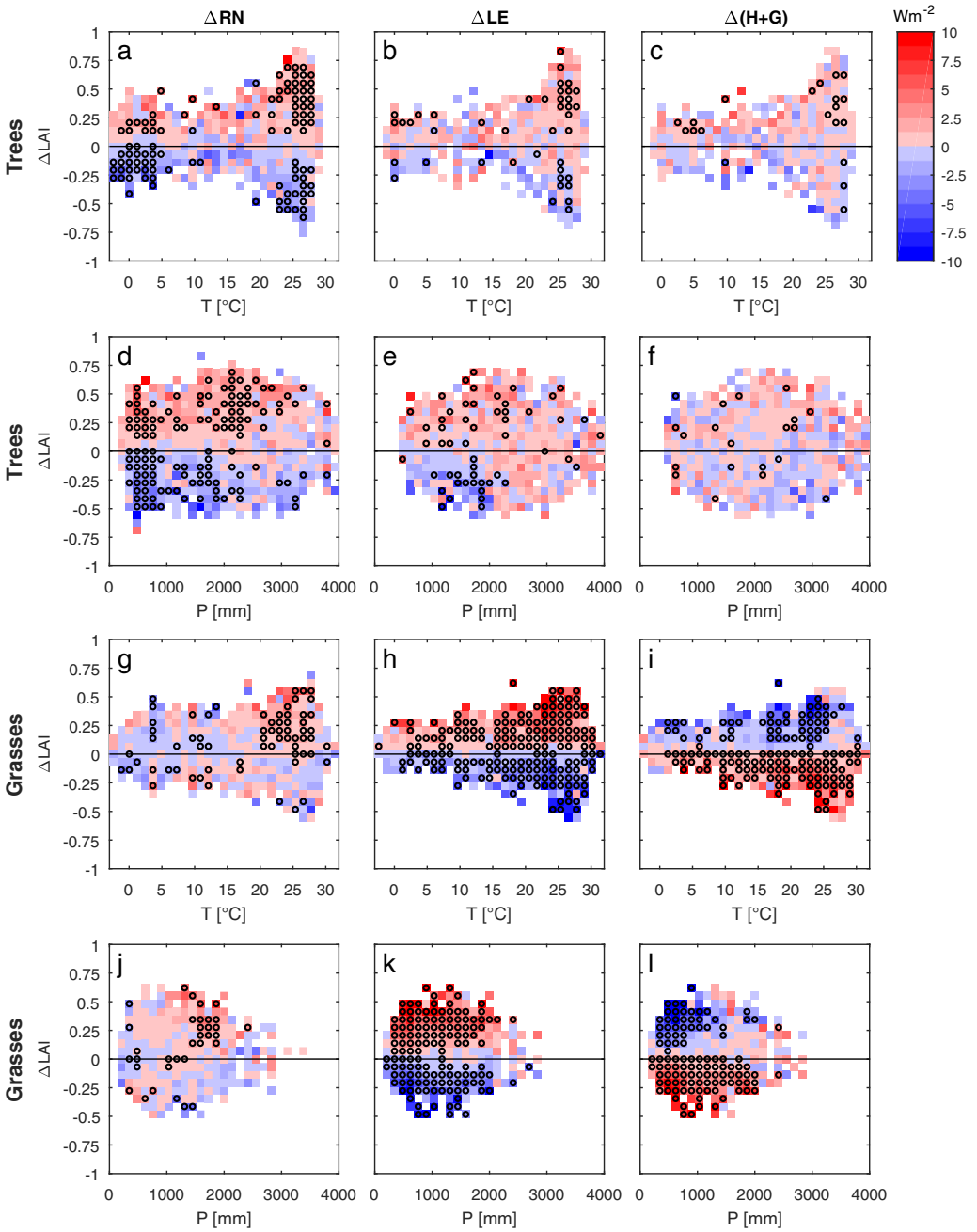

Figure 2. Observed interannual variations of LAI and the components of the surface energy balance across climatological gradients for trees and grasses. Remotely sensed-derived bioclimatic spaces for trees showing the interplay between interannual variations in (left column) annual average leaf area index (ΔLAI, *y* axis) and net radiation (ΔRN), (middle column) latent heat (ΔLE), and (right column) sensible and ground heat (Δ(H + G)) over the climatological median (*x* axis) of (a–c) air temperature gradient and (d–f) precipitation gradient. Figures 2g–2i and 2j–2l as 2a–2c and 3d–2f, respectively, but for grasses. Empty circles show bins where the *t* test indicates that the observed mean value is statistically different (p-value $\leq$0.05) from a 0-mean distribution.

represent alternative pathways for the release of energy from the land surface (Figures 2h, 2i and 2k and 2l). The interplay between turbulent fluxes and LAI emerges clearly with a more robust signal compared to trees (mostly p-value $\leq$0.05 over the explored climatological gradients). However, the vegetation-net radiation relation appears uncertain, particularly in cold and temperate regions (T $<$ 20°C), likely resulting from compensatory effects of radiative and nonradiative processes.

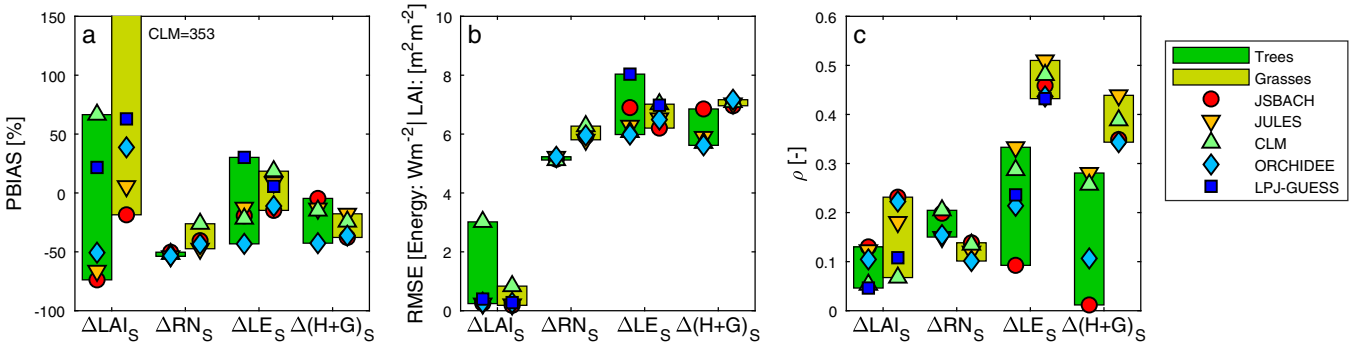

Figure 3. Overall model performance in reproducing single targeted variables. Model performance in simulating single variables, including leaf area index (ΔLAI_S), net radiation (ΔRN_S), latent heat (ΔLE_S), and sensible and ground heat ($\Delta(H+G)_S$). Performance is quantified in terms of (a) percent bias (PBIAS), (b) Root Mean Square Error (RMSE), and (c) Spearman rank (ρ). Model-specific performance is visualized with different symbols accordingly to the legend. Intermodel spreads on tree and grass coverages are shown in dark and light green, respectively. Note that the units of measurement of RMSE are different amongst leaf area index and energy fluxes and the PBIAS value of CLM on grasses is out of plot margins.

3.1.2. JSBACH

Tree biomes. The model tends to underestimate the year-to-year variability in LAI over tree biomes (PBIAS $\sim -74\%$, Figure 3a). The bias is reflected in all bioclimatic spaces prominently in boreal and temperate zones ($T < 20°C$, Figures 4a–4c to compare with Figures 2a–2c) under dry and wet regimes ($P < 1{,}000$ mm and $P > 3{,}000$ mm, Figures 5a–5c to compare with Figures 2d–2f). This likely originates from the model constraints based on threshold mechanisms: JSBACH has a fixed, PFT-specific, parameter for the maximum LAI that vegetation can reach. This parameter is quickly reached during the seasonal cycle, strongly influencing the annual mean value, such that the interannual variability in LAI is dampened. In addition, the parameterization of phenology may further influence the correct representation of the annual LAI average and then its year-to-year variations. According to Dalmonech et al. (2014), JSBACH LAI should be in best agreement with observations in the midlatitude to high latitude, since there phenology is mostly driven by the seasonal course of temperature. However, our analysis shows poor model performance in capturing ΔLAI for these regions. This could indicate limitations of the phenology model to properly determine when the vegetation shifts between growth, vegetative, and dormancy phase should occur.

Relevant vegetation-biophysics interactions, such as the increase (decrease) in net radiation observed in combination with a positive (negative) LAI change in cold regions appear poorly represented in the model (compare upper-left/bottom-left corner in Figures 2a and 4a). This could be connected to the limited model capacity to simulate LAI variations and the consequent interplay with surface energy fluxes (albedo effect). Overall, modeled dynamics in net radiation are moderately correlated to observed patterns ($\rho \geq 0.32$, Figures 4a and 5a and spatial map in Figure 8d), albeit the corresponding energy partitioning into turbulent fluxes shows some discrepancies across the climate gradients. Results show a negative spatial correlation between the latent heat and sensible heat variability (Figures 4b, 4c and 5b, 5c), consistently with the alternative pathways for the release of energy from the land surface observed on grass (Figures 2h, 2i and 2k, 2l) but not on tree biomes (Figures 2b, 2c and 2e, 2f). The dominance of the transpiration-driven mechanism over the observed albedo-driven mechanism might suggest that JSBACH has the incorrect balance of these two processes (compare with observations in Figures 2a–2f), but it is difficult to provide a conclusive interpretation due to the lack of LAI variability in the model.

Estimates of interplay between turbulent fluxes and LAI appear twofold larger in modeled bioclimatic spaces compared to observations (PBIAS $\geq 140\%$ and RMSE ~ 7 W m^{-2}), particularly in warm and moderately wet climates ($T > 20°C$ and $1{,}000$ mm $< P < 2{,}000$ mm) (Figures 4b, 4c and 5b, 5c). Changes in sensible and ground heat fluxes may be of opposite sign compared to observations in a considerable number of different bioclimatic conditions (Figures 4c and 5c). These reverse relations emerge over time as well prominently in tree biomes of East Asia, Eastern North America, and part of the Amazonian forest (Figure 8f to compare with Figure 8c). Overall the model shows higher spatial correlations at bioclimatic level particularly for turbulent fluxes (Figures 4b, 4c and 5b, 5c), than at single-variable level (Figure 3c), likely because of compensatory effects.

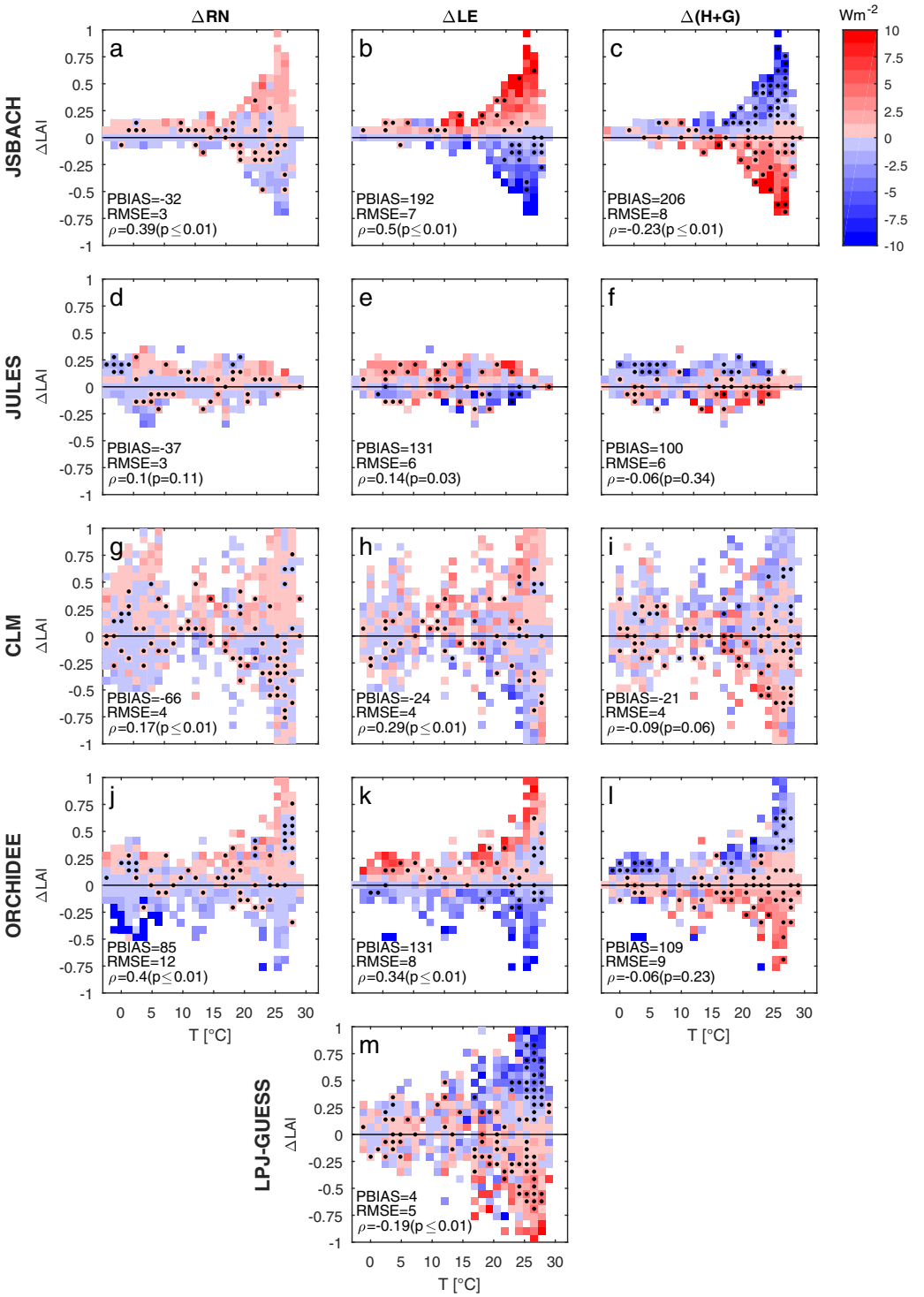

Figure 4. Simulated interannual variations of LAI and the components of the surface energy balance across the air temperature gradient for trees. Simulated bioclimatic spaces for trees showing the interplay between interannual variations in (left column) annual average leaf area index (ΔLAI, *y* axis) and net radiation (ΔRN), (middle column) latent heat (ΔLE), and (right column) sensible and ground heat (Δ(H + G)) over the climatological median of air temperature gradient (*x* axis). Results obtained from (a–c) JSBACH, (d–f) JULES, (g–i) CLM, (j–l) ORCHIDEE, and (m) LPJ-GUESS. Bins with statistically significant discordant sign (p-value ≤0.05) between observed and modeled variations in surface flux are labeled by black dots. Scoring metrics calculated over the whole bioclimatic space are reported in each plot.

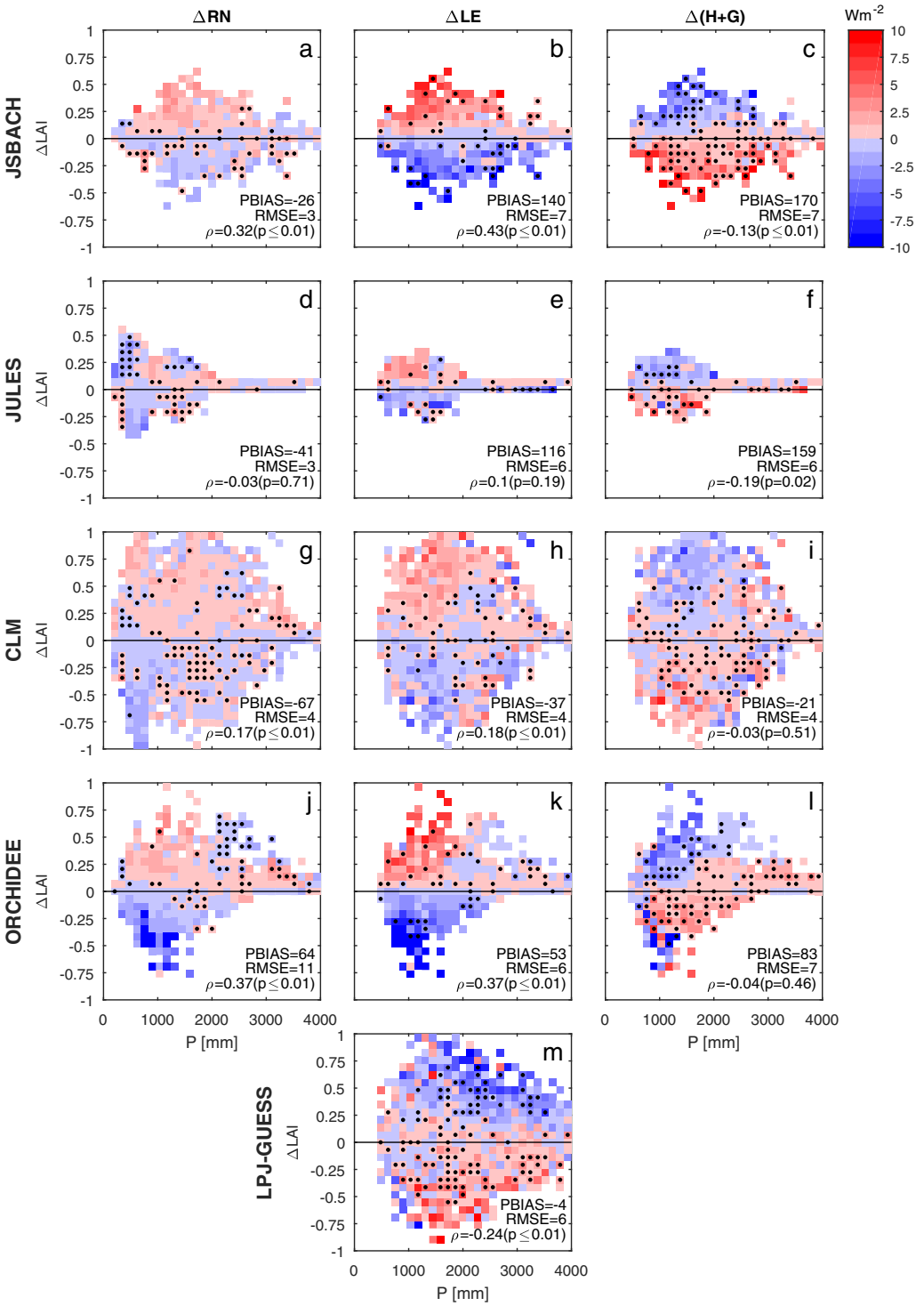

Figure 5. Simulated interannual variations of LAI and the components of the surface energy balance across the precipitation gradient for trees. Same as Figure 4, but bioclimatic spaces explored over the climatological median of precipitation gradient (*x* axis).

Grass biomes. Grasses growing season in JSBACH depends on temperature but also explicitly on soil moisture, and occurs whenever sufficient soil moisture is available and air temperature closed to the ground is larger than a critical air temperature. This approach seems to better represent processes in nature than tree phenology (summer green and evergreen) since modeled year-to-year variability of LAI in grasses appears

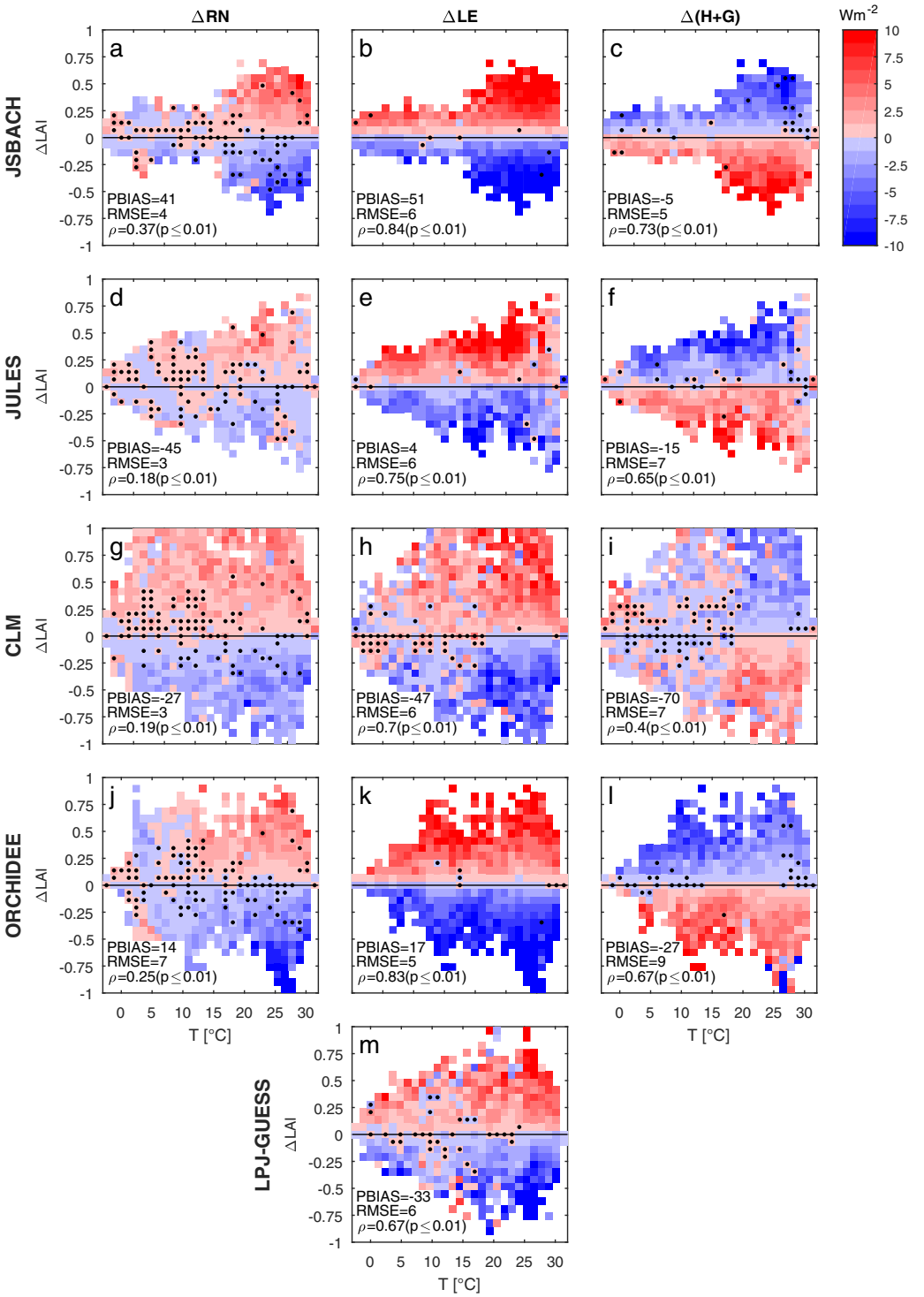

Figure 6. Simulated interannual variations of LAI and the components of the surface energy balance across the air temperature gradient for grasses. Simulated bioclimatic spaces for grasses showing the interplay between interannual variations in (left column) annual average leaf area index (ΔLAI, *y* axis) and net radiation (ΔRN), (middle column) latent heat (ΔLE), and (right column) sensible and ground heat (Δ(H + G)) over the climatological median of air temperature gradient (*x* axis). Results obtained from (a–c) JSBACH, (d–f) JULES, (g–i) CLM, (j–l) ORCHIDEE, and (m) LPJ-GUESS. Bins with statistically significant discordant sign (p-value ≤0.05) between observed and modeled variations in surface flux are labeled by black dots. Scoring metrics calculated over the whole bioclimatic space are reported in each plot.

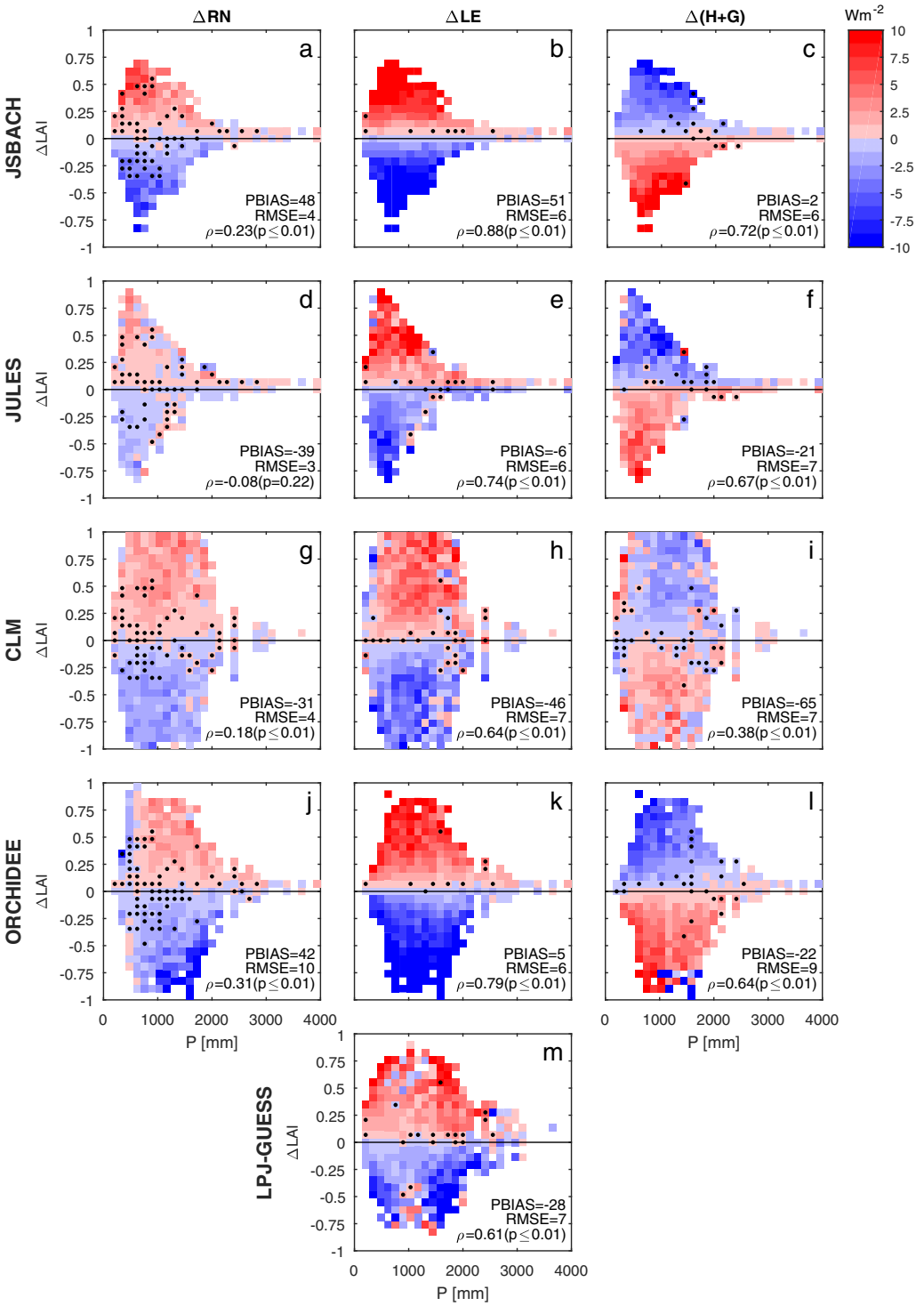

Figure 7. Simulated interannual variations of LAI and the components of the surface energy balance across the precipitation gradient for grasses. Same as Figure 6, but bioclimatic spaces explored over the climatological median of precipitation gradient (*x* axis).

largely consistent with observations (Figure 3). This is well reflected by the range of LAI changes sampled in the bioclimatic spaces except in wet regions where JSBACH tend to underestimate ΔLAI (Figures 6a–6c). In particular, the simulated bioclimatic spaces of turbulent fluxes show high level of agreement (ρ up to 0.84 and RMSE $\sim$ 5 W m^{-2}). In contrast, the interplay between net radiation and LAI changes largely shows opposite pattern compared to observations (Figures 6a and 7a).

Table 3
Model Limitations and Potential Mechanisms Responsible of the Emerging Biases Separately Reported for Each Land Surface Models and Vegetation Biome

		Model limitation	Potential causes
JSBACH	Trees	Underestimation of the ΔLAI particularly in boreal and temperate climates ($T < 20°C$) under dry ($P < 1{,}000$ mm) and wet ($P > 3{,}000$ mm) regimes.	Saturation of LAI dynamics due to PFT-specific maximum LAI threshold. Limitations of the phenology model to properly determine onset and end of the growing season.
		Interplay between LAI and net radiation poorly represented in boreal and temperate climates ($T < 10°C$).	Likely linked to the underestimation of ΔLAI (see previous model limitation).
		Dominance of the transpiration-driven mechanism over the observed albedo-driven mechanisms (observational signal mostly not significant, p-value >0.05).	Incorrect balance of these two processes.
		Sensitivity of turbulent fluxes to LAI appears twofold larger in modeled bioclimatic spaces compared to observations, particularly in warm and moderately wet climates ($T > 20°C$ and 1,000 mm $< P <$ 2,000 mm).	Unclear
		Changes in sensible and ground heat fluxes may be of opposite sign compared to observations (observational signal mostly not significant, p-value >0.05).	Unclear
	Grasses	Sensitivity of net radiation to LAI changes shows opposite pattern compared to observations over large part of the explored climatological gradients (observational signal mostly not significant, p-value >0.05).	Unclear
		Underestimation of ΔLAI in wet regions	Unclear
JULES	Trees	Underestimation of ΔLAI across all the explored climate gradients except very dry regimes ($P < 800$ mm).	Seasonal maximum LAI constrained to the changes in total vegetation carbon. Limitations of the phenology model to properly determine onset and end of the growing season.
		Interplay between LAI and net radiation poorly represented in boreal and temperate climates ($T < 20°C$).	Likely linked to the underestimation of ΔLAI (see previous model limitation).
		Modeled dynamics in energy fluxes are weakly correlated to observed patterns, with possible reverse link (observational signal mostly not significant, p-value >0.05).	Unclear
		Dominance of the transpiration-driven mechanism over the observed albedo-driven mechanism (observational signal mostly not significant, p-value >0.05).	Incorrect balance of these two processes.
	Grasses	Sensitivity of net radiation to LAI changes shows opposite pattern compared to observations over large part of the explored climatological gradients (observational signal mostly not significant, p-value >0.05).	Possible errors in simulated snow cover in cold regions.
		Underestimation of ΔLAI in wet regions.	Model underestimates, or does not include, a process which limits the productivity and LAI of grasses.
CLM	Trees	Overestimation of ΔLAI particularly in cold ($T < 10°C$) and warm ($T > 20°C$) regions under moderately wet regimes ($P < 2{,}000$ mm).	Parameterization of the dynamic carbon allocation scheme leading to an increased LAI sensitivity to interannual changes in carbon fluxes.
		Interplay between LAI and net radiation poorly represented in boreal and temperate climates ($T < 20°C$).	Misrepresentation of the masking effect of snow by vegetation due to partial modeling of understory vegetation in low-density boreal forests.
		Underestimation of the sensitivity of biophysical processes to LAI changes.	Likely linked to the overestimation of ΔLAI (see previous model limitation).
		Modeled dynamics in net radiation are weakly correlated to observed patterns, with reverse link in high-latitude areas.	Unclear
		Changes in sensible and ground heat fluxes may be of opposite sign compared to observations (observational signal mostly not significant, $p > 0.05$).	Unclear
	Grasses	Overestimation of ΔLAI	Parameterization of the dynamic carbon allocation scheme leading to an increased LAI sensitivity to interannual changes in carbon fluxes. Limitations of the phenology model to properly determine onset and end of the growing season.
		Sensitivity of net radiation to LAI changes shows opposite pattern compared to observations in cold and temperate climates ($T < 20°C$) under moderately wet conditions ($P < 2{,}000$ mm) (observational signal mostly not significant, $p > 0.05$).	Unclear

Table 3. (continued)

		Model limitation	Potential causes
ORCHIDEE	Trees	Underestimation of ΔLAI in wet regimes (P > 2,000 mm)	Saturation of LAI dynamics due to PFT-specific maximum LAI threshold.
		Interplay between LAI and net radiation poorly represented in boreal and temperate climates (T < 20°C).	Misrepresentation of the masking effect of snow by vegetation due to partial modeling of understory vegetation in low-density boreal forests.
		Dominance of the transpiration-driven mechanism over the observed albedo-driven mechanism (observational signal mostly not significant, p > 0.05).	Incorrect balance of these two processes.
		Sensitivity of latent heat fluxes to LAI appears twofold larger in modeled bioclimatic spaces compared to observations, particularly in cold (T < 10°C) and warm (T > 20°C) moderately wet (P < 2,000 mm) climates	Too strong control of water stress on surface biophysics in dry areas.
		Changes in sensible and ground heat fluxes may be of opposite sign compared to observations (observational signal mostly not significant, p > 0.05).	Unclear
	Grasses	Sensitivity of net radiation to LAI changes shows opposite pattern compared to observations over large part of the explored climatological gradients (observational signal mostly not significant, p > 0.05).	Unclear
LPJ-GUESS	Trees	Overestimation of ΔLAI in warm (T > 20°C) moderately wet regimes (P < 2,000 mm).	Unclear
		Changes in latent heat are of opposite sign compared to observations in a considerable number of different bioclimatic conditions, particularly in tropical forests (observational signal mostly not significant, p > 0.05).	The annual carbon allocation scheme introduces an 1 year lag in the response of the plant leaf mass to the carbon uptake.
	Grasses	Slight tendency to overestimate year-to-year variability of LAI	Unclear

3.1.3. JULES

Tree biomes. JULES tends to underestimate substantially the year-to-year variability in LAI (PBIAS $\sim -66\%$, Figure 3a). The bias emerges across all the explored climate gradients but in very dry regimes (P < 800 mm) (Figures 4d–4f and 5d–5f to compare with Figures 2a–2c and 2d–2f, respectively). There is little year-to-year variability in LAI because JULES does not contain a mechanism to change a tree's seasonal maximum LAI on this timescale. Seasonal maximum LAI is determined by an allometric relation to total vegetation carbon (Clark et al., 2011); trees grow slowly and have large stores of vegetation carbon, so the year-to-year percentage change in tree vegetation carbon is very small. The parameterization of phenology provides a mechanism by which LAI can deviate from its seasonal maximum; phenology is represented by a binary leaf-on/leaf-off response to a PFT-dependent cold temperature threshold and so variations in the timing of leaf-on and leaf-off could cause year-to-year variation in annual mean LAI. The small variations in LAI in JULES mean that other processes are likely to control the year-to-year variations in surface energy fluxes.

The strong underestimation of interannual variations in LAI is likely the main responsible of the poor representation of the covariability between increase (decrease) in net radiation and positive (negative) LAI change (compare upper-left/bottom-left corner in Figures 4d and 2a). Modeled dynamics in energy fluxes are weakly correlated to observed patterns (Figures 4d–4f and 5d–5f) and in some regions may even show opposite relations over time (e.g., net radiation in boreal trees in Russia and Canada, Figure 8g to compare with Figure 8a). Large errors exist in the relationships between turbulent fluxes and LAI changes (PBIAS $\geq 100\%$ and RMSE $= 6$ W m^{-2}, Figures 4e, 4f and 5e, 5f); however, the model shows reasonable good spatial correlation values at single-variable level in simulating latent heat, and to a less extent of sensible and ground heat (Figure 3c). JULES may be getting the right answer for the wrong reason; flux variability is well captured, but the mechanisms driving the variability are not correct.

There appears to be a negative spatial correlation between the latent heat and sensible heat variability (Figures 4e, 4f and 5e, 5f), suggesting that increases in latent heat reduce surface temperature and thus reduce the sensible heat flux, as observed for grasslands (Figures 2h, 2i and 2k, 2l) but not evident on tree biomes (Figures 2b, 2c and 2e, 2f). The dominance of the transpiration-driven mechanism over the observed albedo-driven mechanism might suggest that JULES has the incorrect balance of these two processes (compare with observations in Figures 2a–2f), but it is hard to be sure, due to the lack of LAI variability in the model.

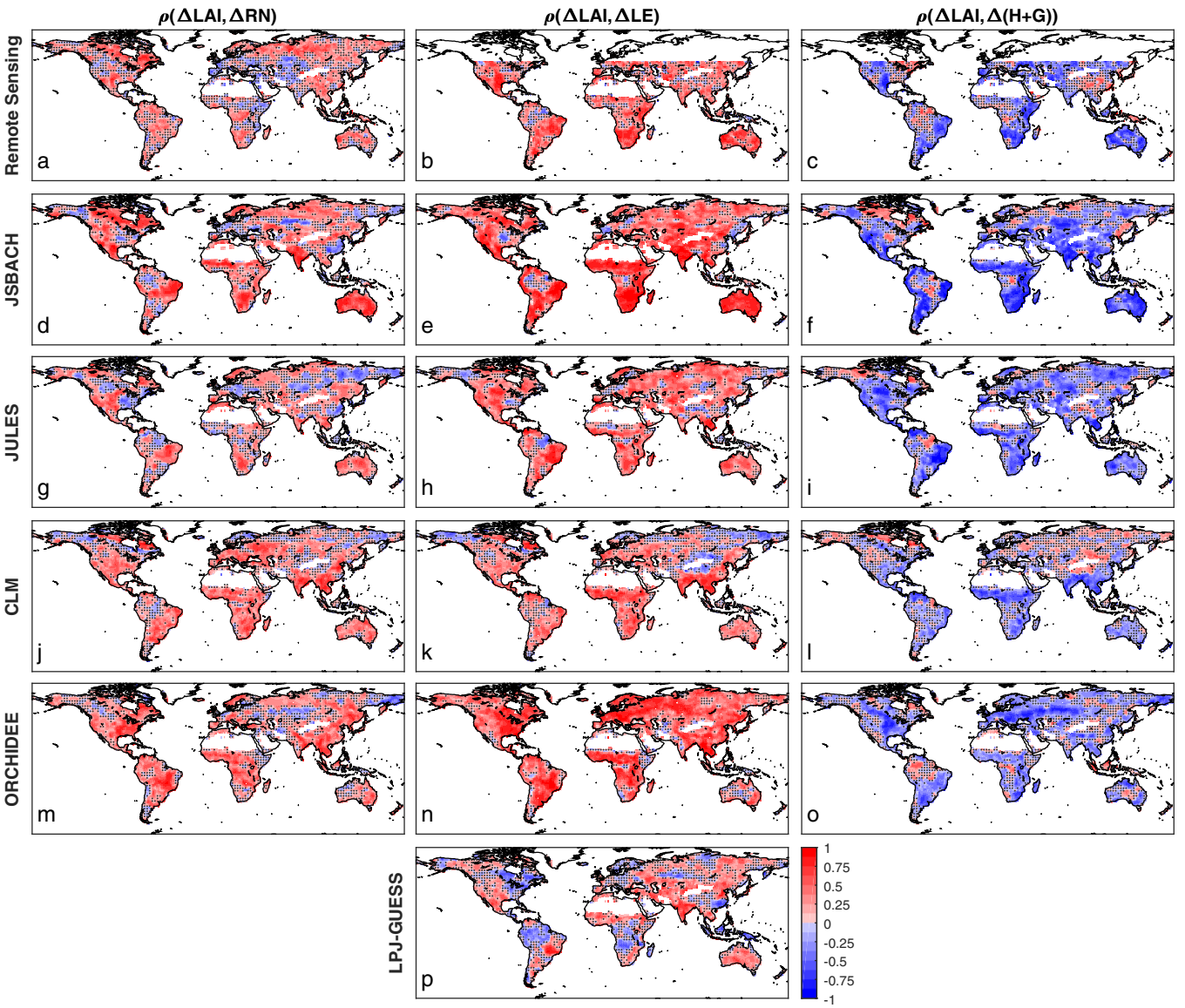

Figure 8. Maps of temporal correlation between interannual variations of LAI and the components of the surface energy balance. Temporal correlation maps between interannual variations in (left column) annual average leaf area index (ΔLAI) and net radiation (ΔRN), (middle column) latent heat (ΔLE), and (right column) sensible and ground heat (Δ(H + G)) quantified from remote sensing data and model simulations. Grey dots indicate correlations that are not statistically significant (p-value >0.05).

Grass biomes. Modeled year-to-year variability of LAI in grasses appears largely consistent with observations (Figure 3) as also reflected across the explored climatological gradients (Figures 6d–6f and 7d–7f). Grass PFTs grow rapidly and have low biomass, allowing LAI to vary by much more than in tree PFTs. Interplay between turbulent fluxes and LAI appears well represented and shows high level of agreement (ρ up to 0.75 and RMSE $\sim$ 6 W m^{-2}). In contrast, the interplay between net radiation and LAI changes shows opposite pattern compared to observations under a number of different bioclimatic conditions (Figures 6d and 7d). JULES tends to absorb more radiation when LAI is higher, while there is no clear signal in the observations. Over cold regions, the differences may be due to errors in simulated snow cover which could dominate year-to-year variability in net radiation. JULES does a good job at capturing the relationships between LAI and surface fluxes, but it underestimates LAI variability in wet regions, suggesting that the model underestimates, or does not include, a process which limits the productivity and LAI of grasses in humid areas.

3.1.4. CLM

Tree biomes. CLM substantially overestimates the year-to-year variability in LAI (PBIAS $\sim$ 66%, RMSE = 3 m^2 m^{-2}, Figures 3a and 3b) particularly in cold (T < 10°C) and warm (T > 20°C) regions and under moderately wet regimes (P < 2,000 mm) (Figures 4g–4i and 5g–5i to compare with Figures 2a–2c and 2d–2f, respectively). This can potentially originate from the dynamic carbon allocation scheme, and its parameterized limitations for some plant functional types, like deciduous PFTs (Montané et al., 2017). The nitrogen limitation constraints in the calculation of productivity play an important role in the sensitivity to climate changes in boreal regions. The tight link with the water cycle via stomata conductance could lead to smaller changes in LAI in relatively moist regions while having a larger effect in dryer regions by reducing water stress (Lee et al., 2013). CLM predicts LAI using an increasing profile with canopy depth, with no LAI threshold, which could lead to unrealistically high LAI values for certain forest PFTs (Ghimire et al., 2016). We argue that such effects can potentially cause an increased LAI sensitivity to interannual changes in carbon fluxes, ultimately resulting in an overestimation of the year-to-year variations in LAI.

The positive spatial correlations between interannual variations in LAI and net radiation observed in cold climates appears poorly represented in the model (compare upper-left/bottom-left corner in Figures 4g and 2a). Reverse patterns compared to observations are prominently evident in high-latitude boreal areas and in large part of Europe (Figure 8j). This may result from a misrepresentation of the masking effect of snow by vegetation (and thus of surface albedo) that ultimately affects the amount of absorbed radiation, and could be linked to the fact that the model does not represent understory vegetation in low-density boreal forests. Opposite relations emerge between interannual LAI and turbulent fluxes as well compared to observations (Figures 4h, 4i and 5h, 5i). Substantial divergences emerge in sensible and ground heat fluxes, particularly over Eastern and interior Asia and Eastern North America (Figure 8l to compare with Figure 8a). In contrast, reasonable good model performance is found in reproducing single energy fluxes (Figure 3). This, in combination with an overestimation of ΔLAI, leads to underestimate the interplay between biophysical processes and LAI changes (PBIAS < −20 and RMSE = 4 W m^{-2}) and to poorly reflect the covariability of vegetation and energy fluxes as suggested by low spatial correlation values (Figures 4g–4i and 5g–5i).

Grass biomes. CLM tends to substantially overestimate LAI interannual variations over grasses (Figures 6g–6i and 7g–7i to compare with Figures 2g–2i and 2j–2l) as furthermore confirmed by the very high positive PBIAS found for the single ΔLAI variable (353%, Figure 3a). Such overestimation appears consistent with what found for tree biomes; even though it is mainly driven by the same causes, it could be amplified by other factors. Onset and offset of phenological cycle of grass biomes are controlled in CLM by soil water availability, soil temperature, and day length. The parameterization adopted can potentially lead to several growing periods in 1 year or multiyear growing season that might, in some locations, not correspond with observations. This may ultimately affects the model capacity to properly reproduce observed interannual variability of LAI. The general patterns of the modeled bioclimatic spaces appear largely consistent with observations (Figures 6g–6i and 7g–7i to compare with Figures 2g–2i and 2j–2l). In particular, sensitivity of turbulent fluxes to LAI shows high level of agreement (ρ up to 0.7) with a slight tendency to underestimate fluctuations (PBIAS < 0, RMSE $\sim$ 6 W m^{-2}) under large variations of LAI. Modeled net radiation show opposite direction of change compared to observations in cold and temperate climates (T < 20°C) under moderately wet conditions (P < 2,000 mm). Such reverse relations emerge between interannual changes in LAI and turbulent fluxes as well, although limitedly to low variations in LAI.

3.1.5. ORCHIDEE

Tree biomes. ORCHIDEE captures the year-to-year variability of LAI reasonably well compared to observations (Figure 3a). This emerges across all the climatological gradients explored except in wet regimes (P > 2,000 mm) where a considerable underestimation occurs (Figures 5j–5l to compare with Figures 2d–2f), mainly associated to an asymmetric distribution of ΔLAI with prevalent positive values. In ORCHIDEE, there is a maximum LAI threshold for each PFT. After reaching the maximum LAI, there is no biomass allocated to leaf and then LAI stop increasing. Such threshold mechanisms are likely responsible of the inconsistencies observed in tropical regions.

Modeled dynamics in net radiation are moderately correlated to observed patterns ($\rho \geq 0.37$ in Figures 4j and 5j and spatial patterns in Figure 8m to compare with Figure 8a), albeit considerable divergences emerge in terms of magnitude of error (RMSE $\sim$ 11 W m^{-2}). The interplay between changes in LAI and net radiation appears poorly represented; particularly in cold regions (compare upper-left/bottom-left corner in

Figures 4j and 2a). This pattern could originates from a misrepresentation of the masking effect of snow by vegetation (and thus of surface albedo) that ultimately affects the amount of absorbed radiation, and could be linked to the fact that the model does not represent understory vegetation in low-density boreal forests. The corresponding energy partitioning into turbulent fluxes shows some discrepancies across the climate gradients. A strong negative spatial correlation between the latent heat and sensible heat variability emerges clearly from model results (Figures 4k, 4l and 5k, 5l) that does not find correspondence with observations (Figures 2b, 2c and 2e, 2f). As already mentioned for certain LSMs, there appears a clear dominance of the transpiration-driven mechanism over the observed albedo-driven mechanism (compare with observations in Figures 2a–2f). This might suggest that ORCHIDEE has the incorrect balance of these two processes.

Estimates of sensitivity of latent heat fluxes to LAI appear twofold larger in modeled bioclimatic spaces compared to observations (PBIAS up to 131% and RMSE up to 8 W m^{-2}), particularly in cold ($T < 10°C$) and warm ($T > 20°C$) moderately wet ($P < 2{,}000$ mm) climates (Figures 4k and 5k). This may reflect that the water stress plays a too strong control so that LE is more tightly negatively coupled to LAI than observed in dry areas. Furthermore, changes in sensible and ground heat fluxes may be of opposite sign compared to observations in a considerable number of different bioclimatic conditions (Figures 4l and 5l) prominently in Eastern Asia and Eastern North America (Figure 8o to compare with Figure 8c).

Grass biomes. Modeled year-to-year variability of LAI in grasses appears largely consistent with observations with a slight tendency to overestimate observed patterns (Figures 3a, 6j–6l, and 7j–7l). In particular, the simulated bioclimatic spaces of turbulent fluxes show high level of agreement (ρ up to 0.83 in Figure 6k). Sensitivity of net radiation to LAI changes shows opposite pattern compared to observations over large part of the explored climatological gradients (Figures 6j and 7j).

3.1.6. LPJ-GUESS

Tree biomes. LPJ-GUESS captures well the year-to-year variability of LAI (PBIAS = 22%, Figure 3a). This emerges across all the climatological gradients explored except in warm ($T > 20°C$) moderately wet regimes ($P < 2{,}000$ mm), where a marked overestimation occurs (Figures 4m and 5m to compare with Figures 2b and 2e, respectively). Changes in latent heat are of opposite sign compared to observations and the other tested LSMs in a considerable number of different bioclimatic conditions (Figures 4m and 5m). This is also reflected in negative spatial correlation values at bioclimatic space level ($\rho \sim -0.2$). The negative relation between ΔLE and ΔLAI is evident in tropical forests, as emerging by the spatially consistent patterns of temporal correlation (Figure 8p). The cause for the discrepancies to the observed interannual changes potentially originates from the annual carbon allocation scheme for trees, which introduces a 1 year lag in the response of the plant leaf mass to the carbon uptake.

Grass biomes. Modeled year-to-year variability of LAI in grasses appears largely consistent with observations with a slight tendency to overestimate (Figures 3a, 6m, and 7m). Simulated bioclimatic spaces of latent fluxes show a high level of agreement with observations (PBIAS ~ -30, RMSE ~ 6 W m^{-2} and ρ up to 0.67).

3.2. Intermodel Comparison

Single LSMs show similar performance across temperature and precipitation gradients (T-space versus P-space), whereas a larger variability emerges across different scoring metrics, energy budget terms, and vegetation covers (Figure 9). Sensitivity of turbulent fluxes simulated above woody vegetation show larger bias compared to that one of net radiation while over grasses they present comparable patterns (Figures 9a and 9b). Deviations in interannual variability of turbulent fluxes are up to twofold bigger than those on net radiation (RMSE ~ 7 W m^{-2} in ΔLE and $\Delta(H + G)$ versus ~ 3.5 W m^{-2} in ΔRN, Figures 9c and 9d), except for ORCHIDEE that shows the contrary. Models mostly provide similar deviations across different vegetation covers; however, when the year-to-year differences in the latent fluxes are normalized by the annual average fluxes, grasses may show higher relative error compared to trees, given the larger absolute values of the latter vegetation type (supporting information Figures S2c and S2d). Simulated bioclimatic spaces of grasses appear well reproduced and strongly correlated with observations, particularly for turbulent fluxes ($\rho \sim 0.7$, Figures 9e and 9f). A lower agreement emerges for trees that may even present negative spatial correlations for certain models/energy terms (e.g., notably for sensible and ground heat fluxes), thus suggesting a reversed LAI control on simulated biophysical processes compared to observations (discussed in the previous sections).

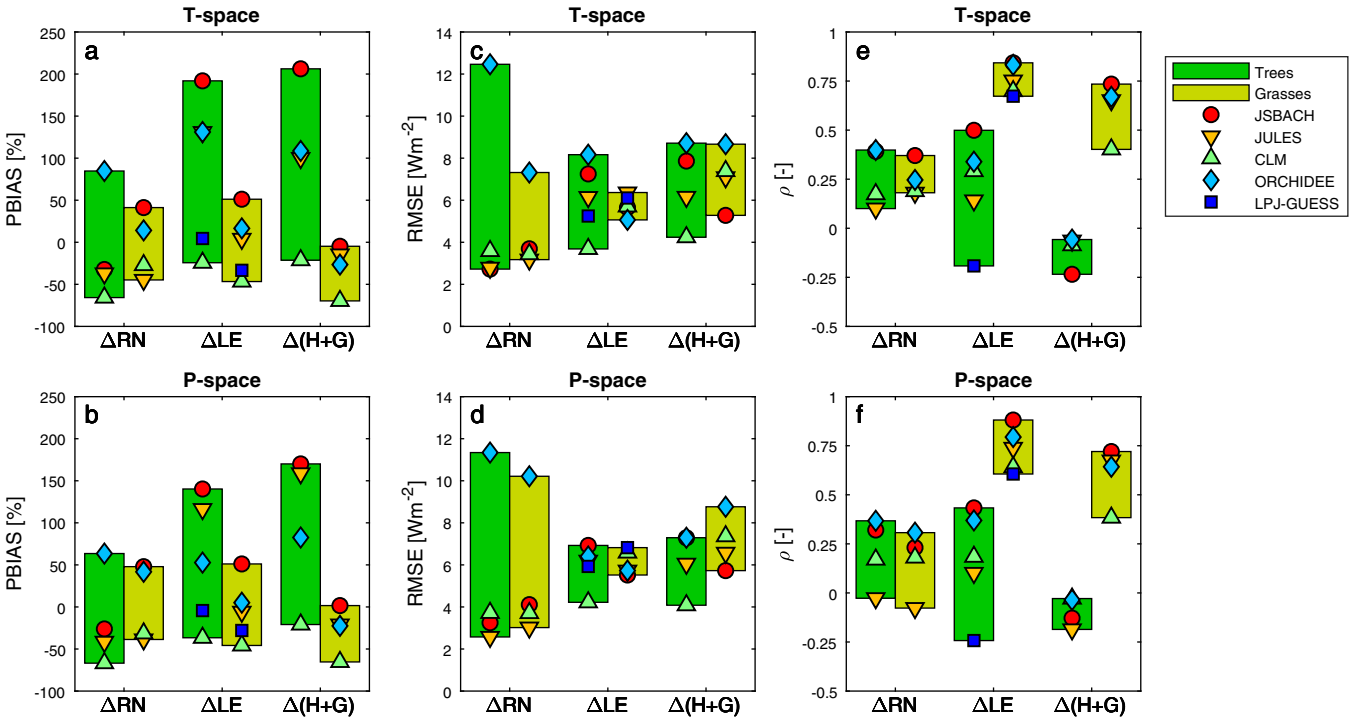

Figure 9. Overall model performance in reproducing the covariability of interannual variations of LAI and the components of the surface energy balance at the annual aggregation scale. Model performance in simulating the interplay between interannual variations in annual average leaf area index and net radiation (ΔRN), latent heat (ΔLE) and sensible and ground heat (Δ(H + G)) over the climatological medians of temperature (T-space, first row) and precipitation (P-space, second row) gradients. Performance is quantified in terms of (a and b) percent bias (PBIAS), (c and d) Root Mean Square Error (RMSE), and (e and f) Spearman rank (ρ). Model-specific performance is visualized with different symbols accordingly to the legend. Intermodel spreads on tree and grass coverages are shown in dark and light green, respectively.

There is no overall best performing LSM resulting from the intermodel comparison, but individual models may outperform the others in specific metrics, surface energy fluxes, vegetation covers, and climate drivers. For instance, JSBACH and ORCHIDEE show generally higher capacity to reproduce the spatial patterns of the observed bioclimatic spaces (high spatial correlation values); JULES and CLM provide lower deviations in absolute terms (low RMSE values); and LPJ-GUESS generates near-zero PBIAS for latent heat flux. In contrast, ORCHIDEE shows lower performance in terms of RMSE; JSBACH and CLM have generally higher systematic biases (overestimating and underestimating observations, respectively); JULES and LPJ-GUESS are poorly correlated with observations compared to the other LSMs.

Comparison of observed-modeled bioclimatic spaces at seasonal scale reveals some common features across models (Figure 10). LSMs tend to overestimate more prominently vegetation-biophysics relations during autumn-winter and produce more pronounced deviations of fluxes in spring-summer, particularly in net radiation and latent fluxes over grasses (Figures 10a–10l). On the one hand, this apparent contradiction can be attributed to higher cloud cover during cold seasons that limits the acquisition capacity of surface properties from satellite sensors and potentially leads to an underestimation of the observational-based surface heat fluxes (observational bias). On the other hand, a more prominent model amplification of variability in fluxes during warm seasons can contribute to the emerging fluctuations in the scoring metrics (model bias). In addition, to the afore-mentioned reasons, the information conveyed by the chosen metrics could also partially explain the apparent contradiction between PBIAS and RMSE. LAI and energy fluxes are generally lower in autumn-winter so that small deviations likely produce large PBIAS. In contrast, LAI and energy fluxes in spring-summer are higher thus potentially leading to small percentage errors still resulting in large RSME (metric bias). No seasonal patterns emerge in terms of spatial correlations amongst observations and models (Figures 10m–10r).

3.3. Sensitivity Analysis on Vegetation Cover and Irrigation Fraction

In order to evaluate the robustness of our results with respect to the masking procedure of the vegetation signal, we performed a sensitivity analysis of the annual-scale scoring metrics to variations in the threshold

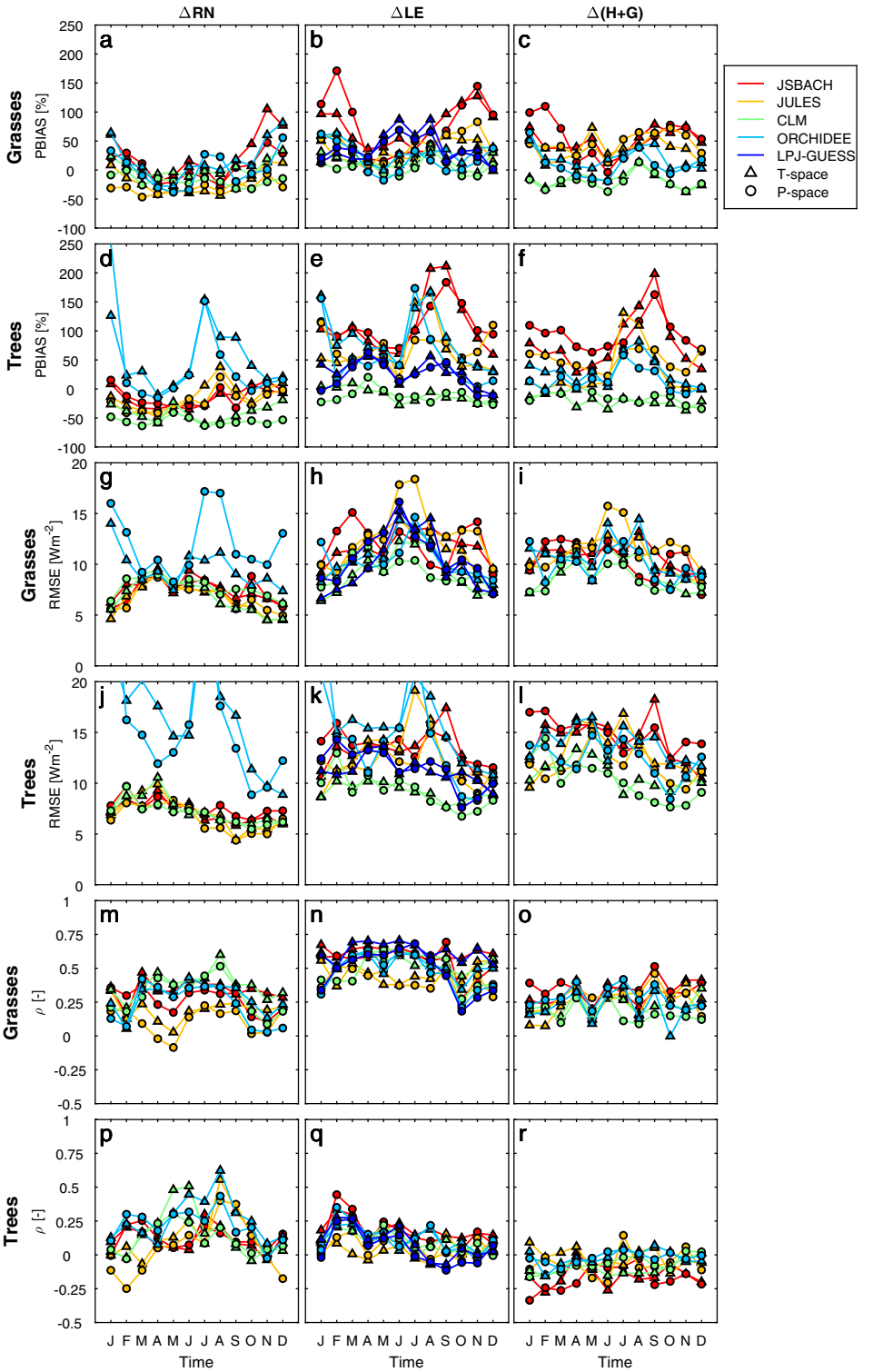

Figure 10. Overall model performance in reproducing the covariability of interannual variations of LAI and the components of the surface energy balance at the monthly aggregation scale. Seasonal patterns of overall model performance in reproducing the interplay between interannual variations in (left column) monthly average leaf area index and net radiation (ΔRN), (middle column) latent heat (ΔLE), and (right column) sensible and ground heat (Δ(H + G)) quantified in terms of (a–f) percent bias (PBIAS), (g–l) Root Mean Square Error (RMSE), and (m–r) Spearman rank (ρ), separately for grasses and trees. Model-specific performance is visualized with different color lines accordingly to the legend, estimates over climatological medians of temperature and precipitation gradients are visualized by triangles and circles, respectively.

of cover fraction used to select grass and tree-covered grid cells. Results mostly reveal low variations in PBIAS and RMSE for increasing levels of within-pixel vegetation cover homogeneity (supporting information Figures S3a–S3l). ORCHIDEE represents an exception showing a considerable reduction in the modeling errors for a threshold of trees cover fraction $\geq$70% to allocate a forest in a grid cell. Spatial correlation values tend to decrease with the assumed increasing threshold levels of cover fraction, more evident for trees (supporting information Figures S3m–S3r), possibly as effect of the strong reduction in the sample size (supporting information Figure S3s). Overall, the selected 60% threshold appears a good compromise to isolate the vegetation type-specific signal and preserve samples of suitable size for robust statistics.

A similar exercise has been developed to assess the sensitivity of results to the threshold used to exclude irrigated areas. Results show negligible variations of results in all scoring metrics, bioclimatic spaces, and vegetation biomes, to varying percentages of irrigation fraction, thus corroborating the robustness of our findings with respect to the 10% of irrigated area selected as threshold in this study (supporting information Figure S4).

4. Discussion and Conclusions

The evaluation framework proposed in this study has the objective to assess the capabilities of an ensemble of LSMs to reproduce the observed interplay between variation in vegetation density, expressed as LAI, and surface energy fluxes. In the following paragraphs, we present a synthesis on the possible strategies for model improvements suggested by this analysis. The strategies may be useful for other LSMs which share common features and parameterizations with JSBACH, JULES, CLM, ORCHIDEE, or LPJ-GUESS.

The analysis helped to (a) determining model-specific deficiencies, (b) diagnosing systematic errors common across models, (c) identifying vegetation-biophysics processes correctly represented across models, and (d) discriminating between models.

a. Model-specific mechanisms listed in Table 3 help to identify potential critical areas for model development and suggest directions for further investigation to improve our understanding and modeling capabilities. Deviations of different magnitude of the sensitivity of biophysical processes mediated by vegetation are found across the different land surface schemes. This leads to overestimation, underestimation, or even a change in sign of the energy fluxes attributed to changes in LAI compared to observations. Such incorrect representations of the biophysical processes affect model simulations of current and future response to climate and environmental change. Identifying rigorously the missing components or a relationship between components or the flawed representation of one or more components in LSMs is extremely complex due to the coevolution of multiple interacting processes and the diverse ways how they are implemented in each land surface scheme. The potential causes listed in Table 3 are those considered as the most plausible based on our knowledge; however, we cannot exclude that additional model biases/artifacts/processes may affect the performance in reproducing the sensitivity of biophysics to LAI changes. Furthermore, we recognize that for certain emerging model-data inconsistencies specific causes have not been clearly identified, thus additional inspections are needed to explain the potential sources of model limitations.
b. Certain systematic errors have been found across the ensemble of LSMs. A marked misrepresentation of the interplay between the year-to-year variability in LAI and net radiation in cold and temperate climates emerges for tree coverage in 4 out of 5 models (JSBACH, JULES, CLM, and ORCHIDEE) (Figures 4 and 5). This likely reflects an imprecise modeling of LAI-related effects on surface albedo and ultimately impacts the amount of absorbed radiation. Furthermore, a dominance of the transpiration-driven mechanism over the observed albedo-driven mechanism is found in 3 out of 5 models (JSBACH, JULES, and ORCHIDEE) possibly linked to an incorrect balance of these two processes. Sensitivity of net radiation to LAI changes appears poorly represented in grass coverage as well in cold and temperate climates, whose effects result on the energy partitioning (Figures 6 and 7). Relations between interannual variations in sensible heat fluxes and LAI are of opposite sign than observed for trees across the whole set of LSMs and may lead to large biases in biophysical processes. However, a conclusive statement of the possible responsible mechanisms cannot be made.

   The emerging systematic biases across models shed light on the current partial understanding of some key vegetation-mediated biophysical processes and how their (flawed) conceptualization is likely to be

shared across models (Haughton et al., 2016). In this respect, previous data-model comparison studies suggests that the misrepresentation of turbulent fluxes in LSMs is due to a limited use of the information content in the atmospheric forcing data (Best et al., 2015). Based on our results, we argue that the representation of the vegetation control on the surface energy balance may represent a further important limitation in LSMs hampering their capability to mimic energy exchanges between the Earth surface and the atmosphere.

Possible common additional causes of discrepancies between observations and simulations could lie with the model capacity to reproduce the heterogeneity of biophysical processes, for instance due to orographic effects or variability in species composition, which our evaluation system does not allow to properly quantify. Noteworthy, the observation-based diagnostics have shown a certain degree of uncertainty in cold-temperate climates especially on sensible and ground heat fluxes for trees, and net radiation for grasses, as evident from the number of bioclimatic conditions where the observational signal is not significant ($p > 0.05$, Figures 2c, 2f and 2g, 2j). Furthermore, the variance in any time series of observed data is necessarily affected by noise in space and time due to instrumental accuracy (Vinnikov et al., 1996). As a consequence, random errors in the observations are likely leading to an underestimation of the actual relationships between LAI and the components of the energy balance (Figure 2). Such additional source of observational uncertainty could partially explain some emerging data-model discrepancies, in particular with regards to the overestimation of the modeled LAI control on turbulent fluxes compared to observations (Figures 4–7 versus 2). However, in this study the magnitude of random errors in satellite retrievals has been minimized by the averaging of the observed data across large spatial domains. Additional techniques to account for the instrumental noise effects could be further integrated to improve the assessment of model performance (Dirmeyer et al., 2016). The synergic use of multiple alternative and independent global scale observation-based products and detailed in situ measurements (Kumar et al., 2012) could surely enhance this assessment. Deviations between models and observations may also originate from uncertainty in the climate forcing (CRU-NCEP) used for LSMs, in particular with respect to radiation, whose values are retrieved from cloudiness visual observations at sparse CRU stations.

c. We found a reasonably good representation of the LAI effect on turbulent fluxes on grasses over most of the climatological gradients and on net radiation in warm climates (Figures 6 and 7). In general higher model performance emerge under large year-to-year variations in LAI both at annual and seasonal scale (e.g., arid regions, beginning of the growing season, Figure 10), thus emphasizing the key role of LAI in mediating biophysical processes in LSMs. These proven strengths increase the confidence on how certain processes are represented in LSMs.
d. An overall best or worst performing model does not emerge from the intermodel comparison (e.g., Figures 9 and 10). The analysis suggests what models work better/worse with respect to certain climate conditions and land surface energy fluxes. We recognize that the employed scoring metrics may only partially characterize the model performance (Bennett et al., 2013). However, we argue that the selected metrics help to characterize the typology of model error, key information to develop suitable strategies for model improvements. Furthermore, the analysis of the vegetation-biophysics relationships through bioclimatic spaces offers a detailed inspection of the model capability to reproduce surface energy fluxes under gradients in vegetation cover and different climate backgrounds. The range in model performance gives a clear indication that improvements are expected to be achievable for those metrics and bioclimatic conditions where another model already performs better. We point out that differences in spatial resolution amongst LSMs might have influenced the results of the model intercomparison. For instance, due the lower resolution JULES necessarily exhibits fewer spatial degrees of freedom compared to the other models, up to a level that may have ultimately affected the spatial correlation tests against observations. However, in our analysis scoring metrics are evaluated over the climatological space and the full spectrum of LAI variability, whose spatial resolution is coarser than the model grids and therefore have likely minimized the potential artifact induced the heterogeneous model resolution. For this reason, we believe that our results and conclusions have not been substantially affected by the differences in spatial resolution between model runs.

Previous studies on data-model comparison of energy balance terms have been mostly devoted to assessing model capability to simulate radiative and nonradiative fluxes, not explicitly accounting for

the interactions with vegetation changes (e.g., Abramowitz, 2012; Abramowitz et al., 2008; Best et al., 2015; Kelley et al., 2013). The few studies that have investigated model performance in representing the vegetation effects have focused on evaluating land surface phenology and trends in greenness (e.g., Murray-Tortarolo et al., 2013). Improved parameterization and phenology schemes have been developed to optimize the agreement between model simulations and observation-derived estimates of seasonal vegetation dynamics (Forkel et al., 2014; MacBean et al., 2015). We argue that such approaches can only partially characterize the model capability to represent the covariability of multiple LAI-related biophysical processes. By focusing on the relationships between concurrent year-to-year variations in LAI and surface energy fluxes, and thus filtering out possible long-term dependencies on covariates, this study provides a novel and complementary assessment of model performances. Our findings help to identify knowledge gaps and improve model representation of the sensitivity of biophysical processes to changes in leaf area density. This is of crucial importance due to the observed widespread greening of Earth occurred in the last decades (Zhu et al., 2016) and its feedbacks into the climate system (Forzieri et al., 2017; Zeng et al., 2017). In particular, comparing models and observations over a wide range of climate and vegetation conditions, as analyzed here, allowed capturing the nonlinearity of system responses that may emerge more frequently in future climate scenarios (Luo et al., 2011; Pitman et al., 2011). Areas with low variation in LAI and high sensitivity of energy terms could particularly benefit from such model developments. An enhanced sensitivity of modeled biophysical processes to changes in leaf area based on the reported diagnostics could significantly help to improve their representation of land-atmosphere interactions. Climate projections could also benefit from an increased confidence in our capacity to model the energy balance response to interannual variations in LAI, not least with respect to simulating the greening of the planet (Mahowald et al., 2016) and forest expansion (Davies-Barnard et al., 2015; Liddicoat et al., 2013), both of which are expected to contribute significantly to future carbon sinks.

Acknowledgments

The authors declare no competing financial interests. All information about the data used can be found in sections 2.1 and 2.2. G.F. and A.C. conceived and designed the study, G.D. developed the protocol for LSM runs, G.G. provided JSBACH simulations, W.L. provided ORCHIDEE simulations, E.R. provided JULES simulations, M.K. provided LPJ-GUESS simulations, P.L. provided CLM simulations, G.F. developed the evaluation framework and analyzed the data, and G.F. interpreted the results and wrote the manuscript with contributions from all coauthors. The study was funded by the FP7 LUC4C project (grant 603542). P. C. acknowledges support the European Research Council Synergy grant ERC-2013-SyG-610028 IMBALANCE-P. J.P. acknowledges support by the German Research Foundation's Emmy Noether Program.

References

Abramowitz, G. (2005). Towards a benchmark for land surface models. *Geophysical Research Letters*, *32*, L22702. https://doi.org/10.1029/2005GL024419

Abramowitz, G. (2012). Towards a public, standardized, diagnostic benchmarking system for land surface models. *Geoscientific Model Development*, *5*(3), 819–827. https://doi.org/10.5194/gmd-5-819-2012

Abramowitz, G., Leuning, R., Clark, M., & Pitman, A. (2008). Evaluating the performance of land surface models. *Journal of Climate*, *21*(21), 5468–5481. https://doi.org/10.1175/2008JCLI2378.1

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., et al. (2015). Spatiotemporal patterns of terrestrial gross primary production: A review. *Reviews of Geophysics*, *53*, 785–818. https://doi.org/10.1002/2015RG000483

Arora, V. K., & Montenegro, A. (2011). Small temperature benefits provided by realistic afforestation efforts. *Nature Geoscience*, *4*(8), 514–518. https://doi.org/10.1038/ngeo1182

Arora, V. K., Scinocca, J. F., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., et al. (2011). Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophysical Research Letters*, *38*, L05805. https://doi.org/10.1029/2010GL046270

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., et al. (2001). FLUXNET: A new tool to study the temporal and spatial variability of ecosystem—Scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, *82*(11), 2415–2434. https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., et al. (2013). Characterising performance of environmental models. *Environmental Modelling & Software*, *40*, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., et al. (2015). The plumbing of land surface models: Benchmarking model performance. *Journal of Hydrometeorology*, *16*(3), 1425–1442. https://doi.org/10.1175/JHM-D-14-0158.1

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., et al. (2011). The Joint UK Land Environment Simulator (JULES), model description—Part 1: Energy and water fluxes. *Geoscientific Model Development*, *4*(3), 677–699. https://doi.org/10.5194/gmd-4-677-2011

Blyth, E., Clark, D. B., Ellis, R., Huntingford, C., Los, S., Pryor, M., et al. (2011). A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. *Geoscientific Model Development*, *4*(2), 255–269. https://doi.org/10.5194/gmd-4-255-2011

Blyth, E., Gash, J., Lloyd, A., Pryor, M., Weedon, G. P., & Shuttleworth, J. (2009). Evaluating the JULES land surface model energy fluxes using FLUXNET data. *Journal of Hydrometeorology*, *11*(2), 509–519. https://doi.org/10.1175/2009JHM1183.1

Bonan, G. B. (2008). *Ecological climatology: Concepts and applications* (2nd ed.). Cambridge, NY: Cambridge University Press.

Bright, R. M., Zhao, K., Jackson, R. B., & Cherubini, F. (2015). Quantifying surface albedo and other direct biogeophysical climate forcings of forestry activities. *Global Change Biology*, *21*(9), 3246–3266. https://doi.org/10.1111/gcb.12951

Chen, T. H., Henderson-Sellers, A., Milly, P. C. D., Pitman, A. J., Beljaars, A. C. M., Polcher, J., et al. (1997). Cabauw experimental results from the project for intercomparison of land-surface parameterization schemes. *Journal of Climate*, *10*(6), 1194–1215. https://doi.org/10.1175/1520-0442(1997)010<1194:CERFTP>2.0.CO;2

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., et al. (2011). The Joint UK Land Environment Simulator (JULES), model description—Part 2: Carbon fluxes and vegetation dynamics. *Geoscientific Model Development*, *4*(3), 701–722. https://doi.org/10.5194/gmd-4-701-2011

Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., et al. (2003). The common land model. *Bulletin of the American Meteorological Society*, *84*(8), 1013–1023. https://doi.org/10.1175/BAMS-84-8-1013

Dalmonech, D., Zaehle, S., Schürmann, G. J., Brovkin, V., Reick, C., & Schnur, R. (2014). 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*(1), 272–291. https://doi.org/10.1175/JCLI-D-13-00593.1

Davies-Barnard, T., Valdes, P. J., Singarayer, J. S., Wiltshire, A. J., & Jones, C. D. (2015). Quantifying the relative importance of land cover change from climate and land use in the representative concentration pathways. *Global Biogeochemical Cycles*, *29*(6), 842–853. https://doi.org/10.1002/2014GB004949

de Noblet-Ducoudré, N., Boisier, J.-P., Pitman, A., Bonan, G. B., Brovkin, V., Cruz, F., et al. (2012). Determining robust impacts of land-use-induced land cover changes on surface climate over North America and Eurasia: Results from the first set of LUCID experiments. *Journal of Climate*, *25*(9), 3261–3281. https://doi.org/10.1175/JCLI-D-11-00338.1

Dingman, S. L. (2015). *Physical hydrology* (3rd ed.). Long Grove, IL: Waveland Press.

Dirmeyer, P. A., Wu, J., Norton, H. E., Dorigo, W. A., Quiring, S. M., Ford, T. W., et al. (2016). Confronting weather and climate models with observational data from soil moisture networks over the United States. *Journal of Hydrometeorology*, *17*(4), 1049–1067. https://doi.org/10.1175/JHM-D-15-0196.1

Duveiller, G., Forzieri, G., Robertson, E., Li, W., Georgievski, G., Lawrence, P., et al. (2018). Biophysics and vegetation cover change: A process-based evaluation framework for confronting land surface models with satellite observations. *Earth System Science Data Discussions*, 1–24. https://doi.org/10.5194/essd-2018-24

Forkel, M., Carvalhais, N., Schaphoff, S., Bloh, W. V., Migliavacca, M., Thurner, M., et al. (2014). Identifying environmental controls on vegetation greenness phenology through model-data integration. *Biogeosciences*, *11*(23), 7025–7050. https://doi.org/10.5194/bg-11-7025-2014

Forzieri, G., Alkama, R., Miralles, D. G., & Cescatti, A. (2017). Satellites reveal contrasting responses of regional climate to the widespread greening of Earth. *Science*, *356*(6343), 1180–1184. https://doi.org/10.1126/science.aal1727

Ghimire, B., Riley, W. J., Koven, C. D., Mu, M., & Randerson, J. T. (2016). Representing leaf and root physiological traits in CLM improves global carbon and nitrogen cycling predictions. *Journal of Advances in Modeling Earth Systems*, *8*, 598–613. https://doi.org/10.1002/2015MS000538

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, *4*(2), 135–143. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135)

Haughton, N., Abramowitz, G., Pitman, A. J., Or, D., Best, M. J., Johnson, H. R., et al. (2016). The plumbing of land surface models: Is poor performance a result of methodology or data quality? *Journal of Hydrometeorology*, *17*(6), 1705–1723. https://doi.org/10.1175/JHM-D-15-0171.1

IPCC (2013). *Climate change 2013 the physical science basis*. Cambridge, UK: Cambridge University Press.

Jiang, C., Ryu, Y., Fang, H., Myneni, R., Claverie, M., & Zhu, Z. (2017). Inconsistencies of interannual variability and trends in long-term satellite leaf area index products. *Global Change Biology*, *23*(10), 4133–4146. https://doi.org/10.1111/gcb.13787

Jiménez, C., Prigent, C., Mueller, B., Seneviratne, S. I., McCabe, M. F., Wood, E. F., et al. (2011). Global intercomparison of 12 land surface heat flux estimates. *Journal of Geophysical Research*, *116*, D02102. https://doi.org/10.1029/2010JD014545

Jung, M., Reichstein, M., & Bondeau, A. (2009). Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, *6*(10), 2001–2013. https://doi.org/10.5194/bg-6-2001-2009

Kelley, D. I., Prentice, I. C., Harrison, S. P., Wang, H., Simard, M., Fisher, J. B., et al. (2013). A comprehensive benchmarking system for evaluating global vegetation models. *Biogeosciences*, *10*(5), 3313–3340. https://doi.org/10.5194/bg-10-3313-2013

Knorr, W., & Heimann, M. (2001). Uncertainties in global terrestrial biosphere modeling: 1. A comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme. *Global Biogeochemical Cycles*, *15*(1), 207–225. https://doi.org/10.1029/1998GB001059

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, *15*(3), 259–263. https://doi.org/10.1127/0941-2948/2006/0130

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., et al. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, *19*, GB1015. https://doi.org/10.1029/2003GB002199

Kumar, S. V., Peters-Lidard, C. D., Santanello, J., Harrison, K., Liu, Y., & Shaw, M. (2012). Land surface Verification Toolkit (LVT)—A generalized framework for land surface model evaluation. *Geoscientific Model Development*, *5*(3), 869–886. https://doi.org/10.5194/gmd-5-869-2012

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., et al. (2011). Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems*, *3*, M03001. https://doi.org/10.1029/2011MS00045

Lee, E., Felzer, B. S., & Kothavala, Z. (2013). Effects of nitrogen limitation on hydrological processes in CLM4-CN. *Journal of Advances in Modeling Earth Systems*, *5*, 741–754. https://doi.org/10.1002/jame.20046

Liddicoat, S., Jones, C., & Robertson, E. (2013). CO_2 emissions determined by HadGEM2-ES to be compatible with the representative concentration pathway scenarios and their extensions. *Journal of Climate*, *26*(13), 4381–4397. https://doi.org/10.1175/JCLI-D-12-00569.1

Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDeau, S., et al. (2011). Ecological forecasting and data assimilation in a data-rich era. *Ecological Applications*, *21*(5), 1429–1442. https://doi.org/10.1890/09-1275.1

Luo, Y. Q., Randerson, J. T., Abramowitz, G., Bacour, C., Blyth, E., Carvalhais, N., et al. (2012). A framework for benchmarking land models. *Biogeosciences*, *9*(10), 3857–3874. https://doi.org/10.5194/bg-9-3857-2012

MacBean, N., Maignan, F., Peylin, P., Bacour, C., Bréon, F.-M., & Ciais, P. (2015). Using satellite data to improve the leaf phenology of a global terrestrial biosphere model. *Biogeosciences*, *12*(23), 7185–7208. https://doi.org/10.5194/bg-12-7185-2015

Mahowald, N., Lo, F., Zheng, Y., Harrison, L., Funk, C., Lombardozzi, D., et al. (2016). Projections of leaf area index in earth system models. *Earth System Dynamics*, *7*(1), 211–229. https://doi.org/10.5194/esd-7-211-2016

Maignan, F., Bréon, F.-M., Chevallier, F., Viovy, N., Ciais, P., Garrec, C., et al. (2011). Evaluation of a Global Vegetation Model using time series of satellite vegetation indices. *Geoscientific Model Development*, *4*(4), 1103–1114. https://doi.org/10.5194/gmd-4-1103-2011

Marcolla, B., Rödenbeck, C., & Cescatti, A. (2017). Patterns and controls of inter-annual variability in the terrestrial carbon budget. *Biogeosciences*, *14*(16), 3815–3829. https://doi.org/10.5194/bg-14-3815-2017

Martens, B., Miralles, D., Lievens, H., Fernández-Prieto, D., & Verhoest, N. E. C. (2016). Improving terrestrial evaporation estimates over continental Australia through assimilation of SMOS soil moisture. *International Journal of Applied Earth Observation and Geoinformation*, *48*, 146–162. https://doi.org/10.1016/j.jag.2015.09.012

McPherson, R. A. (2007). A review of vegetation-atmosphere interactions and their influences on mesoscale phenomena. *Progress in Physical Geography*, *31*(3), 261–285. https://doi.org/10.1177/0309133307079055

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., & Dolman, A. J. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, *15*(2), 453–469.

Montané, F., Fox, A. M., Arellano, A. F., MacBean, N., Alexander, M. R., Dye, A., et al. (2017). Evaluating the effect of alternative carbon allocation schemes in a land surface model (CLM4.5) on carbon fluxes, pools, and turnover in temperate forests. *Geoscientific Model Development*, *10*(9), 3499–3517. https://doi.org/10.5194/gmd-10-3499-2017

Mueller, B., Seneviratne, S. I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., et al. (2011). Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophysical Research Letters*, *38*, L06402. https://doi.org/10.1029/2010GL046230

Murray-Tortarolo, G., Anav, A., Friedlingstein, P., Sitch, S., Piao, S., Zhu, Z., et al. (2013). Evaluation of land surface models in reproducing satellite-derived LAI over the high-latitude Northern Hemisphere. Part I: Uncoupled DGVMs. *Remote Sensing*, *5*(10), 4819–4838. https://doi.org/10.3390/rs5104819

Piao, S., Sitch, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., et al. (2013). Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO_2 trends. *Global Change Biology*, *19*(7), 2117–2132. https://doi.org/10.1111/gcb.12187

Pitman, A. J., Avila, F. B., Abramowitz, G., Wang, Y. P., Phipps, S. J., & de Noblet-Ducoudré, N. (2011). Importance of background climate in determining impact of land-cover change on regional climate. *Nature Climate Change*, *1*(9), 472–475. https://doi.org/10.1038/nclimate1294

Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., et al. (2015). Plant functional type classification for earth system models: Results from the European Space Agency's Land Cover Climate Change Initiative. *Geoscientific Model Development*, *8*(7), 2315–2328. https://doi.org/10.5194/gmd-8-2315-2015

Quéré, C. L., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., et al. (2018). Global carbon budget 2017. *Earth System Science Data*, *10*(1), 405–448. https://doi.org/10.5194/essd-10-405-2018

Raddatz, T. J., Reick, C. H., Knorr, W., Kattge, J., Roeckner, E., Schnur, R., et al. (2007). Will the tropical land biosphere dominate the climate–carbon cycle feedback during the twenty-first century? *Climate Dynamics*, *29*(6), 565–574. https://doi.org/10.1007/s00382-007-0247-8

Randerson, J. T., Hoffman, F. M., Thornton, P. E., Mahowald, N. M., Lindsay, K., Lee, Y.-H., et al. (2009). Systematic assessment of terrestrial biogeochemistry in coupled climate–carbon models. *Global Change Biology*, *15*(10), 2462–2484. https://doi.org/10.1111/j.1365-2486.2009.01912.x

Schimel, D., Pavlick, R., Fisher, J. B., Asner, G. P., Saatchi, S., Townsend, P., et al. (2015). Observing terrestrial ecosystems and the carbon cycle from space. *Global Change Biology*, *21*(5), 1762–1776. https://doi.org/10.1111/gcb.12822

Schultz, N. M., Lee, X., Lawrence, P. J., Lawrence, D. M., & Zhao, L. (2016). Assessing the use of subgrid land model output to study impacts of land cover change. *Journal of Geophysical Research: Atmospheres*, *121*, 6133–6147. https://doi.org/10.1002/2016JD025094

Sellers, P. J., Hall, F. G., Kelly, R. D., Black, A., Baldocchi, D., Berry, J., et al. (1997). BOREAS in 1997: Experiment overview, scientific results, and future directions. *Journal of Geophysical Research*, *102*(D24), 28731–28769. https://doi.org/10.1029/97JD03300

Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., et al. (2015). Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences*, *12*(3), 653–679. https://doi.org/10.5194/bg-12-653-2015

Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., et al. (2014). Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. *Biogeosciences*, *11*(7), 2027–2054. https://doi.org/10.5194/bg-11-2027-2014

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., et al. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, *13*(14), 4291–4313. https://doi.org/10.5194/bg-13-4291-2016

Vinnikov, K. Y., Robock, A., Speranskaya, N. A., & Schlosser, C. A. (1996). Scales of temporal and spatial variability of midlatitude soil moisture. *Journal of Geophysical Research*, *101*(D3), 7163–7174. https://doi.org/10.1029/95JD02753

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B. III, Smith, G. L., & Copper, J. E. (1996). Clouds and the Earth's Radiant Energy System (CERES): An Earth observing system experiment. *Bulletin of the American Meteorological Society*, *77*, 853–868.

Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., et al. (2009). Improving land surface models with FLUXNET data. *Biogeosciences*, *6*(7), 1341–1359. https://doi.org/10.5194/bg-6-1341-2009

Winckler, J., Reick, C. H., & Pongratz, J. (2016). Robust identification of local biogeophysical effects of land-cover change in a global climate model. *Journal of Climate*, *30*(3), 1159–1176. https://doi.org/10.1175/JCLI-D-16-0067.1

Zeng, Z., Piao, S., Li, L. Z. X., Zhou, L., Ciais, P., Wang, T., et al. (2017). Climate mitigation from vegetation biophysical feedbacks during the past three decades. *Nature Climate Change*, *7*(6), 432–436. https://doi.org/10.1038/nclimate3299

Zhang, Y., Peña-Arancibia, J. L., McVicar, T. R., Chiew, F. H. S., Vaze, J., Liu, C., et al. (2016). Multi-decadal trends in global terrestrial evapotranspiration and its components. *Scientific Reports*, *6*, 19124. https://doi.org/10.1038/srep19124

Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., et al. (2013). Global data sets of Vegetation Leaf Area Index (LAI)3g and Fraction of Photosynthetically Active Radiation (FPAR)3g derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the period 1981 to 2011. *Remote Sensing*, *5*(2), 927–948. https://doi.org/10.3390/rs5020927

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., et al. (2016). Greening of the Earth and its drivers. *Nature Climate Change*, *6*(8), 791–795. https://doi.org/10.1038/nclimate3004