Local scale evaluation of the simulated interactions between energy, water and vegetation in ISBA, ORCHIDEE and a diagnostic model.

Jan De Pue[1], José Miguel Barrios[1], Liyang Liu[2], Philippe Ciais[2], Alirio Arboleda[1], Rafiq Hamdi[1], Manuela Balzarolo[3], Fabienne Maignan[2], and Françoise Gellens-Meulenberghs[1]

[1]Department of Meteorological and Climatological Research, Royal Meteorological Institute, Belgium
[2]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France
[3]Department of Biology, University of Antwerp, Belgium

Correspondence: Jan De Pue (jan.depue@meteo.be)

Abstract. The processes involved in the exchange of water, energy and carbon in terrestrial ecosystems are strongly intertwined. To accurately represent the terrestrial biosphere in land surface models (LSM), the intrinsic coupling between these processes is required. Soil moisture and leaf area index (LAI) are two key variables at the nexus of water, energy and vegetation. Here, we evaluated two prognostic LSM (ISBA and ORCHIDEE) and a diagnostic model (based on the LSA SAF algorithms) in their ability to simulate the latent heat flux (LE) and gross primary production (GPP) coherently, and their interactions through LAI and soil moisture. The models were validated using in situ eddy covariance observations, soil moisture measurements and remote-sensing based LAI. It was found that the diagnostic model performed consistently well, regardless land cover, whereas important shortcomings of the prognostic models were revealed for in herbaceous/dry sites. Despite their different architecture and parametrization, ISBA and ORCHIDEE shared some key weaknesses. In both models, LE and GPP were found to be oversensitive to drought stress. Though the simulated soil water dynamics could be improved, this was not the main cause of errors in the surface fluxes. Instead, these errors were strongly correlated to errors in LAI. The simulated phenological cycle in ISBA and ORCHIDEE was delayed compared to observations, and failed to capture the observed seasonal variability. The feedback mechanism between GPP and LAI (i.e. the biomass allocation scheme) was identified as a key element to improve the intricate coupling between energy, water and vegetation in LSM.

1 Introduction

Terrestrial ecosystems modulate the surface fluxes of heat, water and carbon, and are thereby an essential driver of weather and climate (Pielke et al., 1998). They are a substantial dynamic component of the global carbon budget, with 15% of the global atmospheric CO_2 being yearly exchanged through the stomata of leaves and assimilated through photosynthesis (Ciais et al., 2013). Furthermore, the pivotal role of vegetation on the global climate is mediated by its impact on the hydrological cycle (Falkenmark et al., 2004). Despite its importance in the frame of the global changing climate, large uncertainties remain in our understanding of the coupling of the energy, water and carbon cycle in the terrestrial biosphere (Piao et al., 2013; Kauwe et al., 2017; Shukla et al., 2019).

Land surface models (LSM) are key tools to quantify these surface fluxes, and to better the representation of their interactions. They allow the coupled simulation of the fluxes of water, energy and carbon between the surface and the atmosphere, and are a crucial component of numerical weather models and earth system models. Over the past decades, they have evolved from their initial simple biophysical configuration to include more complex feedback mechanisms, such as soil moisture dynamics, dynamic vegetation, plant phenology etc. (Delire et al., 2020; Fisher and Koven, 2020).

The processes involved in the surface fluxes from the terrestrial biosphere, such as photosynthesis, transpiration, soil hydrology and leaf phenology, are deeply intertwined with each other. Soil moisture and leaf area index (LAI) are two key variables at the nexus between energy, water and vegetation processes.

Root zone soil moisture affects the leaf exchange of water and carbon by modulating the stomatal closure (Raschke, 1979). Although the physiological processes involved are well-described, there is a substantial disagreement in the stomatal behaviour across various models (Kauwe et al., 2017). An evaluation of the impact of soil moisture in the CMIP5 models (Taylor et al., 2012) indicated that the LSM were generally oversensitive to drought stress and wet events (Huang et al., 2016). Whereas several other studies have reported similar outcomes (Piao et al., 2013; Kolus et al., 2019), Rebel et al. (2012) found an underestimation of ORCHIDEE response to drought. Some of the key challenges lie in: the upscaling of leaf-level processes to canopy-scale and ecosystem-scale simulations (Kauwe et al., 2017), the broad range of processes contributing to evapotranspiration (ET), along with numerous feedback mechanisms (Bonan et al., 2014; Fisher and Koven, 2020), and the difficulty to simulate soil moisture dynamics and infiltration itself (Li et al., 2018; Vereecken et al., 2019). Furthermore, the validation of these simulations is hampered due to the scale mismatch between flux footprint and model grid and the challenge in accurately observing the partitioning of the surface fluxes (transpiration, soil evaporation, canopy intercept evaporation, etc.; Nelson et al., 2020).

Leaf area index is another key variable in terrestrial ecosystem models. It is used to represent the abundance of foliar vegetation and its canopy state. Many leaf-scale processes are scaled to canopy-scale surface fluxes, proportional to LAI. Over the past decades, simulations with prognostic LAI have become an established approach to account for interseasonal variability of the terrestrial vegetation in land surface models (Calvet et al., 1998; Dickinson et al., 1998; Krinner et al., 2005; Gibelin et al., 2006). The coupling of the carbon assimilation to a biomass allocation scheme allows to simulate the variable phenological cycle, and the vegetation response to atmospheric forcings. The degree of complexity of this scheme is very variable amongst models, and range from fairly simplistic (e.g. in ISBA (Le Moigne et al., 2018) or CHTESSEL (Boussetta et al., 2013)) to advanced, with dedicated phenology modules or non-structural carbohydrate dynamics (e.g. in ORCHIDEE (Krinner et al., 2005), CLM (Lawrence et al., 2019) or CLASS (Asaadi et al., 2018)). Previous studies have concluded that LSM are capable of representing the amplitude of the seasonal LAI cycle with reasonable accuracy (Gibelin et al., 2008), but substantial shortcomings are found in the timing of the phenological cycle and the interseasonal variability (Lafont et al., 2012). The disagreement amongst models (and observations) can be attributed to our limited knowledge on the drivers of budburst and senescence, biomass allocation, reserve dynamics and below-ground processes (Le Roux et al., 2001; Fatichi et al., 2019). As a consequence of the coupling of the vegetation dynamics with the water and carbon cycles, the uncertainty associated with the seasonal cycle of LAI propagates back to the surface fluxes.

The resulting feedbacks from the coupling are summarized in Fig. 1. Soil moisture and LAI are state variables which determine the exchange of heat, water and carbon. Through the feedback to soil moisture in prognostic models, uncertainties in the exchange of heat and water (e.g. sensible heat flux (H), latent heat flux (LE), or evapotranspiration) propagate to the carbon assimilation. Inversely, uncertainties in gross primary production (GPP) or the vegetation growth affect the heat and water fluxes over LAI. Finally, through phenology equations in some models (e.g. ORCHIDEE for grass), soil moisture can also affect LAI directly.

[Figure 1 about here.]

This study focuses on the representation of these interactions in two well-established prognostic LSM: ORCHIDEE (Krinner et al., 2005) and ISBA (Le Moigne et al., 2018) and one diagnostic model (Ghilain et al., 2011; Martínez et al., 2020). The evaluation of LSM is typically achieved by validating components of the LSM individually, as mediated by the ever-increasing availability of long-term in situ measurements of energy, water and carbon fluxes from eddy covariance (EC) tower networks (Balzarolo et al., 2014; Napoly et al., 2017; Dirmeyer et al., 2018). In situ observations of surface fluxes, meteorological conditions and soil moisture are an essential resource in the study of terrestrial ecosystems and the development of LSM. In combination with remote-sensing based observations of LAI, they provide key insights in the interactions between the surface fluxes and the biosphere.

Beyond the validation of the model outputs, the assessment of the model dynamics and internal interactions is needed to further advance LSM development. Approaches to tackle this include sensitivity analyses, anomaly analysis or isolation of extreme events (e.g. Alton, 2016; Huang et al., 2016). Additionally, the quality of the prognostic state variables can be assessed through a functional evaluation. Here, the diagnostic LSM is used as a vehicle to test the impact of the prognostic soil moisture and LAI on the surface fluxes. Diagnostic LSM are typically designed to estimate fluxes from observed state variables, such as remote-sensing based soil moisture and LAI. Replacing the observed states by the prognostic states allows to test their impact on the surface fluxes. To our knowledge, this is the first study to perform such a functional evaluation of LSM.

The objective of this paper is to evaluate the performance and internal dynamics of three LSM at local scale. Our focus is the relation between the surface fluxes (LE, GPP), and important state variables (soil moisture, LAI). This is done by 1) validation and intercomparison of the simulated surface fluxes and prognostic states in these models, 2) comparison of the model dynamics (phenology and flux partitioning) and 3) evaluation of the interactions with soil moisture and LAI. Given the degree of coupling in the current LSM, we try to disentangle the relation between key facets of the terrestrial vegetation in a holistic way.

2 Materials & Methods

2.1 Models

Three well-established models were used to simulate the intrinsically coupled fluxes of water, energy and carbon from terrestrial vegetation: a diagnostic model based on the LSA SAF algorithms (hereafter referred to as DiagMod), ISBA and ORCHIDEE. Each model has a different approach to represent plant phenology. Whereas ISBA has a fairly simple biomass allocation scheme to represent the phenological cycle, ORCHIDEE relies on dedicated phenology modules, and DiagMod is driven by remote-sensing based forcing variables, such as LAI.

Simulations were performed for a wide range of hydro-climatic biomes and plant functional types at local scale (i.e. a single grid point). The simulated fluxes were validated using eddy-covariance measurements, and the simulated phenology was compared to remote-sensed observations of LAI.

For adequate intercomparison, the models were configured to run with identical land cover and atmospheric forcing. The land cover at each site was derived from ECOCLIMAP 2 (Faroux et al., 2013) and corrected manually if this was not representative for the tower footprint area (based on ICOS and FLUXNET metadata, and satellite imagery). The sources of the forcing variables are listed in Table 1. ERA5 was used to replace tower variables with large gaps in the time series (e.g. relative humidity; Hersbach et al., 2020). It was verified that the impact of the use of ERA5 instead of local forcings was limited (not shown here). The forcing from ERA5 (hourly resolution) was linearly interpolated to match the 30 minute temporal resolution from the tower observations.

[Table 1 about here.]

An overview of some key plant physiology parameters and soil physical parameters is given in the supplement material, along with full option namelists of the ISBA and ORCHIDEE runs to allow reproducibility.

Most of the vegetation parameters in ISBA are derived from the TRY plant traits database (Kattge et al., 2011; Delire et al., 2020). Parameters in ORCHIDEE are regularly calibrated using various data types, including satellite observations and in situ observations of fluxes and atmospheric CO_2 concentration (e.g. Kuppel et al., 2012, 2014; MacBean et al., 2015; Peylin et al., 2016). Kuppel et al. (2014) used 78 FLUXNET sites to optimize parameters related to the NEE (net ecosystem exchange) and LE fluxes (see their Table S2). Hence, whereas the ORCHIDEE parameters were not optimized using the specific dataset of this study, a part of it may have been used formerly in this regard. Similarly, key parameters of the diagnostic model have been (indirectly) derived from a subset of the global eddy covariance network (Garbulsky et al., 2010; Martínez et al., 2020).

2.1.1 Diagnostic Model (DiagMod)

The diagnostic model used in this study is based on the algorithms applied in the LSA SAF products. The LSA SAF algorithm to simulate surface turbulent energy fluxes was developed in the framework of EUMETSAT deployment of 'Satellite Appli-

cations Facilities' (SAF; https://www.eumetsat.int/about-us/satellite-application-facilities-safs), and is used to generate LSA SAF ET, LE and H products operationally (i.e in near real time). It is a Soil Vegetation Atmosphere Transfer (SVAT) model, largely driven by remote-sensing based observations of downwelling long- and shortwave radiation, LAI and albedo. It relies on the Jarvis (1976) approach to calculate the stomatal response to environmental factors.

In operational modus, the observations of the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard the Meteosat Second Generation Satellite (MSG) are the primary source of the forcing variables. A more in-depth outline of the algorithm is given in Ghilain et al. (2011) and Ghilain et al. (2012). Consequently, it was designed to run at the resolution of MSG-observations, but its capabilities at sub-kilometer scale were recently demonstrated (Barrios et al., 2020). For this study, the model was configured to run at kilometre-scale (i.e. the local scale corresponding to the footprint of eddy covariance measurements), using LAI from the European Copernicus Global Land Service (CGLS) and soil moisture from ERA5.

More recently, a LSA SAF GPP product was developed, based on the Monteith light use efficiency (LUE) concept (Martínez et al., 2020). This product is calculated at the end of the LSA SAF pipeline, as it relies on several other LSA SAF products, such as ET, reference ET, LAI and FAPAR. The same formulation was adopted in the diagnostic model in our study, resulting in coherent surface fluxes.

Contrary to ISBA and ORCHIDEE, the calculations for LE and GPP in the diagnostic model do not share parameters like stomatal resistance. Instead, the GPP calculations are coupled to LE by using the actual evapotranspiration as an input variable.

2.1.2 ISBA

Within the Surfex (SURFace Externalisée) land surface model, ISBA (Interactions between Soil, Biosphere and Atmosphere) is the component dedicated to modeling the exchange of water, energy and carbon fluxes between the soil-vegetation-snow continuum and the atmosphere (Masson et al., 2013; Le Moigne et al., 2018). In this case, a configuration of ISBA with interactive carbon cycling is used, i.e. ISBA-CC (Gibelin et al., 2008; Delire et al., 2020). The fluxes of water and carbon from the vegetation are coupled through the stomatal resistance. This shared parameter is calculated through the A-gs surface scheme, and largely depends on soil moisture stress and air temperature (Calvet et al., 2004). The parametrization for this scheme is based on plant traits derived from the TRY-database (Kattge et al., 2011; Delire et al., 2020).

The assimilation of carbon results in the evolution of LAI through a biomass allocation scheme. The growth and senescence of leaves is purely photosynthesis-driven. The biomass reservoirs are coupled to a soil organic matter module to calculate the respiration terms.

The simulations with ISBA were performed on the Surfex v8.1 platform[1]. The soil profile was discretized in 14 layers (up to 12 m depth), using a diffusion scheme for soil heat and water transfer and an exponential decrease of hydraulic conductivity through the profile. The nitrogen dilution scheme (Calvet and Soussana, 2001) and canopy radiation transfer scheme (Carrer et al., 2013) were enabled. In the forest patches, the energy fluxes were calculated with the recently developed multi-energy balance scheme (MEB, Boone et al., 2017). Contrary to the standard soil-vegetation composite version of ISBA (which was

[1] https://www.umr-cnrm.fr/surfex/

used for the non-forest patches), MEB explicitly solves the transfer of mass and energy between the soil surface, the snowpack, the canopy and the atmosphere. At the time of this study, the combination of MEB and prognostic LAI modelling is still considered experimental (Le Moigne et al., 2018). A spin-up period of 3 years was sufficient to eliminate effects from the initial model state on the surface fluxes (respiration is not analysed in this study). ISBA was not coupled to a hydrological model (e.g. CTRIP Decharme et al., 2019). Consequently, there was no lateral ground water flow or a water table, only free drainage at the bottom of the soil profile.

2.1.3 ORCHIDEE

ORCHIDEE is the land surface model of the Institut Pierre Simon Laplace (IPSL) earth system model, and was initially described in Krinner et al. (2005). We used the version prepared for the 6th Coupled Model Inter-comparison Project (CMIP6) (Boucher et al., 2020; Cheruy et al., 2020).

The LAI is prognostic, and the phenology models used for the various plant functional types (PFT) are described in Botta et al. (2000) and MacBean et al. (2015). The canopy is discretized in layers of increasing thickness from the top to the bottom of the canopy. The incoming light is attenuated through the canopy following a Beer-Lambert extinction law. The photosynthesis is modelled at the leaf level following Farquhar et al. (1980) for C3 species and Collatz et al. (1992) for C4 species. The maximum carboxylation rate at $25°C$ is a PFT-dependent parameter. The maximum carboxylation rate varies with the temperature following Medlyn et al. (2002) and Kattge and Knorr (2007). A water stress function depending on soil moisture and root profile (Rosnay and Polcher, 1998) is applied to the maximum carboxylation rate, the stomatal and the mesophyll conductances. An analytical solution to the three equations linking CO_2 assimilation, stomatal conductance and CO_2 leaf intercellular concentration is computed following Yin and Struik (2009). The assimilation is then upscaled over the layers to calculate the GPP.

A single-layer energy balance is computed per grid cell. LE is the weighted average of the snow sublimation, the soil evaporation, the canopy transpiration and the evaporation of foliage water; all these terms were initially computed following Ducoudré et al. (1993). The soil is now discretized over 2 meters into 11 layers of increasing thickness, and the hydrology scheme follows Richard's equation (De Rosnay et al., 2002; d'Orgeval et al., 2008). There is free drainage at the bottom. The soil thermodynamics are described in Wang et al. (2016), and the snow scheme is detailed in Wang et al. (2013).

To initialize the simulations, a first spin-up phase was performed, where we cycled over the available FLUXNET years for at least 45 years. This enables to reach an equilibrium for the above-ground biomass and the water stocks and fluxes, as an initial state for the transient simulation.

2.2 Test sites

The performance of the models was evaluated at field-scale, using observations from flux towers. From the FLUXNET2015 dataset (Pastorello et al., 2020) and the ICOS '2018 drought initiative' dataset (Drought 2018 Team and ICOS Ecosystem Thematic Centre, 2019), sites were selected with adequate EC data quality (at least 1 year of carbon fluxes, dominated by

observations with quality flag 1 or better), homogeneous land cover (within 1 km radius from the tower, assessed via Google earth) and limited disturbance due to management. This resulted in the 56 sites listed in Table 2, and a total of 526 simulation years. 33 of these sites are dominated by forest land cover, whereas 18 are dominated by herbaceous vegetation and 5 are crop sites (the models are configured to run without management practices). The FLUXNET and ICOS data products had been pre-processed with the ONEFLUX processing pipeline (Pastorello et al., 2020). The test sites were classified per PFT (taken from the FLUXNET/ICOS IGBP metadata), dominant vegetation type (forest, herbaceous or crop) and hydro-climatic biome (HCB; Papagiannopoulou et al., 2018).

In addition to the classification based on land cover and meteorology, the sites were classified in 'aridity classes'. In the LSM, the rootzone soil moisture modulates the stomatal conductance when it drops below field capacity (ISBA) or below 80% of the difference between field capacity and wilting point (ORCHIDEE). As a proxy for aridity, the fraction of the simulation time that the simulated soil moisture in the topsoil (0-7 cm) drops below this threshold was used. It was found that this was significantly (Wilcoxon signed-rank test $p<0.05$) more frequent in ISBA (48% of the time, median value of all sites), compared to ORCHIDEE (26% of the time). Significant differences persisted deeper in the soil profile, until 70 cm depth. Using this metric, the sites were classified in four classes, equal in size, going from least (class 1) to most arid (class 4) (see Table 2). This classification was based on the ISBA simulations, but a similar classification was obtained with ORCHIDEE (despite the differences in absolute values). The vegetation at sites with aridity class 1 was mainly dominated by forest, whereas the aridity class 4 sites were mostly occupied by herbaceous vegetation. No evident relation with the hydro-climatic biomes was found.

[Table 2 about here.]

Not all sites are equipped with soil moisture sensors, nor is there a standardized setup or post-processing for soil moisture in the datasets used for this study. Consequently, the validation of the simulated soil moisture and the sensitivity analysis were only performed for the sites with sensors. Furthermore, some sites were equipped with multiple sensors in the soil profile. Here, only the median score of the sensors was used in the statistics (i.e. one score per site). For the validation, all sensors up to 2 m depth were used, whereas only the sensors up to 0.5 m depth (i.e. the shallow root zone) were used in the sensitivity analysis (though the impact on the results was minimal).

2.3 Validation

The simulated H and LE were validated with the observed daily mean fluxes from flux towers. The non-closure of the energy balance is a well-known issue in the eddy covariance observations (Cui and Chui, 2019). The turbulent fluxes in the FLUXNET and ICOS datasets were corrected for this, under the assumption that the measured Bowen ratio was correct (Pastorello et al., 2020). Due to missing observations of the ground heat flux, this correction was not possible for all sites. The validation of H and LE was only performed for the sites where all fluxes were available. The mean correction of LE of each site is listed in Table 2.

Similarly, the simulated GPP was validated with the FLUXNET/ICOS GPP data. The net ecosystem exchange (NEE) observed at the fluxtower was partitioned into its ecosystem respiration (RECO) and GPP components using the daytime fluxes and constant friction velocity (USTAR) threshold method (Pastorello et al., 2020). Only data with a quality flag indicating good quality (1) or better was used in this analysis. Though some authors have recommended to correct the carbon fluxes in a similar way as the turbulent fluxes, such a procedure was not included in the processing pipeline (Massman and Lee, 2002; Gao et al., 2019, see also section 4).

An important key to the feedback mechanism between the surface fluxes is the LAI. The simulated LAI from ISBA-CC and ORCHIDEE was validated using the remote-sensing based LAI from the European Copernicus Global Land Service [2]. The LAI data product used here is derived from SPOT-VGT and PROBA-V satellite data, it has a spatial resolution of 1 km and a temporal resolution of 10 days (Camacho et al., 2013). The sites were selected to be fairly homogeneous within the footprint area, and the observed LAI is assumed to be representative for the direct surroundings of the eddy covariance stations.

The simulated soil moisture profiles of ISBA and ORCHIDEE, and the ERA5 soil moisture (used in DiagMod), were validated where possible. To reduce biases caused by different soil physical properties of the soil profiles or differences in scale between models and observations, the observed and simulated volumetric soil moisture (θ) was converted to the effective saturation (S_e) as follows:

$$S_e = \frac{\theta - \theta_{min}}{\theta_{max} - \theta_{min}} \tag{1}$$

where θ_{min} and θ_{max} were assumed to be the 5th and 95th percentile of the observed soil moisture in a site for the observations, or the residual and saturated water content for the simulations.

For H, LE, GPP, LAI and S_e, the classical validation indices are calculated: mean error (ME), root mean square error (RMSE), Pearson correlation (r) and Nash-Sutcliffe model efficiency (NS). They were calculated as in Equations 2 - 5, in which y^* and y^o are the predicted and observed values, $\overline{y}$ the mean of y and n_o the number of observations:

$$\text{ME} = \frac{\sum^{n_o}(y^* - y^o)}{n_o} \tag{2}$$

$$\text{RMSE} = \sqrt{\frac{\sum^{n_o}(y^* - y^o)^2}{n_o}} \tag{3}$$

$$r = \frac{\sum^{n_o}(y^* - \overline{y^*})(y^o - \overline{y^o})}{\sqrt{\sum^{n_o}(y^* - \overline{y^*})^2 \sum^{n_o}(y^o - \overline{y^o})^2}} \tag{4}$$

$$\text{NS} = 1 - \frac{\sum(y^* - y^o)^2}{\sum(y^o - \overline{y^o})^2} \tag{5}$$

[2] http://land.copernicus.eu/global/

Taylor diagrams were constructed using the Pearson correlation (r) and standard deviation (σ) of the observed and simulated variables. The validation was performed using the daily totals/averages.

Furthermore, the same analysis was also performed on the anomalies to the mean annual cycles, to isolate the capability of the models to capture seasonal variability. The mean annual cycles were computed per site, across all its site-years. The validation indices of the seasonal anomalies have the subscript $_{\text{ANOM}}$, e.g. NS_{ANOM}.

Significant differences between the models were evaluated with the Wilcoxon signed-rank test (paired), and the significance of the PFT, HCB, aridity class and dominant land cover to classify the model performances was evaluated with the Kruskal-Wallis H-test. Differences between classes were tested with the Mann–Whitney U test (non-paired).

2.4 Model dynamics

2.4.1 Phenology

The capability of the models to reproduced the timing of the seasonal cycle of LE, GPP and LAI was evaluated. The detection of the start, maximum and end of the seasonal cycle (SOS, MOS and EOS) was achieved by applying a smoothing operation (20 day rolling mean), followed by a threshold procedure (Maleki et al., 2020). In this threshold procedure, the minima and maxima were used to delineate the growing and senescent phase of the season. MOS was defined as the date when the maximum of the season is reached, SOS and EOS were defined at the date where the growing or senescent phase crosses the threshold value T. T was calculated for each growing or senescent phase as $T = P_5 + 0.2(P_{95} - P_5)$, where P_5 and P_{95} are the 5th and 95th percentile.

2.4.2 Partitioning

To compare the model dynamics, the simulated LE flux partitioning, water balance and water use efficiency (WUE) were evaluated as well. Direct observations of the LE flux partitioning were not available, but it is possible to extract the Transpiration component from the total LE flux, using the underlying Water Use Efficiency (uWUE) method (Zhou et al., 2016; Nelson et al., 2020). From the GPP and transpiration (Tr), the WUE was derived:

$$\text{WUE} = \frac{\text{GPP}}{Tr} \tag{6}$$

2.5 Evaluation of prognostic LAI and soil moisture

2.5.1 Sensitivity and error correlation

To assess the sensitivity of the fluxes to the state variables (S_e and LAI), the slope of the seasonal anomalies of the fluxes against the anomalies of the state variables was determined. This analysis was performed for the observations and the simulations, and compared. Note that the linear slope was used here, though a linear response is not necessarily expected (e.g. the response to soil moisture anomalies depends on a wet/dry regime). The goal of this analysis was to investigate whether LSM

are capable of reproducing a similar relationship as found in the observations. Significant differences between the models were evaluated with the Wilcoxon signed-rank test.

To evaluate whether errors in the state variables result in errors in the surface fluxes (or vice versa), the Spearman rank correlation between both was calculated. Since Copernicus LAI was the reference LAI, this analysis was not possible for LAI in DiagMod.

2.5.2 Functional evaluation with DiagMod

The diagnostic model is a suitable vehicle to test the impact of the prognostic state variables from ISBA and ORCHIDEE on the surface fluxes. Given its architecture to easily ingest state variables, it can serve as an independent model platform to evaluate the quality of the soil moisture and LAI. DiagMod simulations were performed using soil moisture and/or LAI from ISBA and ORCHIDEE, and compared to simulations with soil moisture from ERA5 and CGLS LAI (resulting in 7 runs per site, see Table 4).

The fraction of absorbed photosynthetically active radiation (FAPAR) is an important variable in DiagMod to produce GPP, but it is no output of the prognostic models. In order to be consistent with the prognostic LAI, FAPAR was estimated using a simple Beer's law with a general purpose extinction coefficient value of 0.5 (Eq. 7; Monsi and Saeki, 2005).

$$\text{FAPAR} = 1 - \exp(-0.5\ \text{LAI}) \tag{7}$$

The soil moisture of the soil profiles in the prognostic models was integrated to match the 4 layers in DiagMod (0-7cm, 7-21cm, 21-72cm, 72-189cm). Furthermore, the soil moisture was rescaled using the wilting point and field capacity parameters of the models.

Prior to the evaluation of the prognostic state, the reproducibility of the prognostic models by the DiagMod was tested. The detailed results are shown in the supplement material. It was found that the surface fluxes produced by DiagMod, forced by the same atmospheric conditions, soil moisture and LAI, were more closely correlated to those from ISBA, compared to ORCHIDEE. Differences can be caused by different parametrization of the plant physiology, as well as the representation of processes (or lack thereof), such as rainfall interception, snow cover, canopy radiation transfer, etc.

3 Results

3.1 Validation

3.1.1 Surface fluxes: LE and GPP

The bias (ME) and accuracy (RMSE) of the simulated LE and GPP are shown in Fig. 2, together with Taylor diagrams of the simulated fluxes and their seasonal anomalies. It was evident that the inter-site variability of the model performance is much larger than the inter-model variability. In terms of bias and accuracy, the differences between the models were relatively limited. All models suffered a substantial underestimation of LE, whereas the overall bias in GPP was relatively small. Significant differences (Wilcoxon $p<0.05$) were found in the bias of GPP between DiagMod (overestimation) and ISBA (underestimation), and the simulated LE was significantly more accurate in ISBA, compared to ORCHIDEE.

Notably, no substantial bias was found in the simulated H of any model to compensate for the consistent bias in LE (results shown in supplementary material). In this study, the corrected fluxes from the FLUXNET/ICOS dataset were used as a reference. If the non-corrected fluxes were used instead, the bias in LE was reduced, but the simulated H was overestimated (not shown here). This points at the significant uncertainty associated with the observed fluxes from eddy-covariance measurements. The estimated observation uncertainty of the turbulent fluxes (associated with random measurement errors and energy balance correction) had the same order of magnitude as the model errors.

The Taylor diagrams in Fig. 2 show that the average variability of the simulated LE and GPP was in fair agreement with the observations. After removal of the mean seasonal cycle, the performance of the models decreased (r_{ANOM}, $\mathrm{NS}_{\mathrm{ANOM}}$), but the mean variability of the anomalies is reasonably accurate. In terms of r and r_{ANOM} of LE and GPP, ORCHIDEE was significantly outperformed by ISBA and DiagMod (Wilcoxon $p<0.05$). No significant differences were found between ISBA and DiagMod.

[Figure 2 about here.]

The impact of the land cover type of the test site on the model performance is illustrated in Fig. 3. Here, the test sites are classified by the dominant vegetation type. The NS and $\mathrm{NS}_{\mathrm{ANOM}}$ of the simulated LE was not significantly impacted in any of the models, whereas a significant influence (Kruskal $p<0.05$) was found on the quality of the simulated GPP in DiagMod and ORCHIDEE. The NS and $\mathrm{NS}_{\mathrm{ANOM}}$ of the simulated GPP in ORCHIDEE were significantly better (Mann-Whitney U $p<0.05$) for forest sites, compared to sites that were dominated by herbaceous vegetation. Inversely, the simulation of the seasonal GPP anomalies in DiagMod were significantly better in herbaceous test sites (Mann-Whitney U $p<0.05$). No significant impact was found in the ISBA simulations. Notably, the differences between the models were most pronounced in the herbaceous sites (see also Table 3). Yet, despite its poorer performance in the herbaceous sites, ORCHIDEE simulated GPP in forest sites most accurately, compared to the other models.

[Table 3 about here.]

[Figure 3 about here.]

Similar results were found with the other validation indices. A more detailed breakdown of the results per PFT and HCB is given in the supplement material. A significant impact of PFT and HCB on the NS of the simulated GPP (Kruskal p $<$0.05) was found in all models. This was contrasted by LE, where a significant impact of HCB (Kruskal p $<$0.05) was found only for ORCHIDEE.

3.1.2 State variables: Soil moisture and LAI

The validation results of S_e and LAI are shown in Fig. 4. The soil moisture from ERA5 (used in DiagMod) tended to be overestimated compared to in situ observations, whereas an overall negative bias was found in ISBA and ORCHIDEE. The simulated variability of soil moisture was too low in all models, in particular for ORCHIDEE. Notably, ERA5 outperformed ISBA (p>0.05) and ORCHIDEE (p<0.05) in terms of accuracy, despite their use of in situ meteorological forcings (e.g. precipitation). ORCHIDEE performed significantly worse than the other two models for all validation metrics (Wilcoxon $p < 0.05$). The highest correlation in the anomalies was simulated by ISBA.

Compared to the surface fluxes, the accuracy of the simulated soil moisture was substantially lower. The validation scores of S_e are given in Table 3, separated per dominant land cover type. In all models, the simulated S_e was significantly better for herbaceous sites, compared to forest sites. The herbaceous sites are generally found in a water-driven dryland climate, with strong precipitation-driven anomalies.

Similarly, the prognostic LAI was also of poorer quality than the simulated surface fluxes. ISBA had a significantly better ME and RMSE than ORCHIDEE, but both models overestimated LAI and strongly underestimated its variability. In particular, the variability of LAI in the evergreen needleleaf forests was strongly underestimated in both models, as well as the variability of LAI in evergreen broadleaf forests in ORCHIDEE. Furthermore, both models obtained only a poor correlation, and achieved a very poor correlation of the seasonal anomalies. In both models, the simulated LAI for forest sites was better than for the herbaceous sites, though not significantly (p$>$0.05). The simulated anomalies were modelled significantly better (p$<$0.05) in forest sites than herbaceous sites (Table 3).

[Figure 4 about here.]

3.2 Model dynamics

3.2.1 Phenology

The timing of the start, maximum and end of the seasonal cycle was validated for LE, GPP and LAI. Fig. 5 shows the boxplots of the mean errors in all sites. In all models, the bias and accuracy of the seasonality of LE and GPP was comparable, whereas the leaf phenology (i.e. LAI) was poorer. The simulated phenology of LAI was delayed substantially, in particular in ISBA.

This bias was most pronounced by the MOS, and to a lesser extent in EOS.

ISBA performed significantly worse than ORCHIDEE (Wilcoxon $p < 0.05$) for ME of MOS GPP and MOS LAI, and RMSE of MOS LAI. The prognostic LAI in both models tended to peak towards the end of the growing season, whereas the maximum LAI was reached in the beginning of the season according to the observations. This is illustrated in Fig. 6, where the mean annual LE, GPP and LAI cycles of ENF and DBF sites are shown. The delayed GPP phenology in ISBA is a feedback effect of the delayed prognostic LAI. However, the effect is dampened, since GPP is largely driven by atmospheric forcings as well.

In forest sites, EOS of LAI tended to be simulated with the highest accuracy. The phenology of herbaceous sites had a higher variability (median standard deviation of EOS LAI in forest sites was 7.7 days, compared to 20.6 days in herbaceous sites), which turned out to be challenging to capture for ISBA and ORCHIDEE. An example is shown for the Savanna sites in Fig. 7. DiagMod relied on the remote-sensing based LAI and was significantly more accurate than the prognostic models to capture EOS of GPP (Wilcoxon p $<$0.05).

As the models were configured to run without dedicated management practices for the crop sites, EOS was estimated too late due to the harvest practice (Fig. 7). Even in DiagMod, EOS of GPP was delayed.

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

3.2.2 Water balance, WUE & LE partitioning

The water balance partitioning in ISBA and ORCHIDEE is shown in Fig. 8. In both models, the evapotranspiration fraction across PFT was similar, but substantial differences were found in the drainage and runoff in both models. Whereas nearly no water was lost through runoff in the ISBA simulations, a substantial amount of runoff was simulated with ORCHIDEE. On the other hand, the drainage in ISBA was consistently larger than in ORCHIDEE. DiagMod does not compute a water balance, so could not be included in this comparison.

[Figure 8 about here.]

Both models agreed that the largest fraction of LE is through transpiration of the vegetation (Fig 9). Aside from a few exceptions, T/ET in ORCHIDEE was larger than in ISBA. The median T/ET in ISBA (0.53) is lower than in ORCHIDEE (0.68), and is closer to the values derived from the tower observations with the uWUE method (0.54). However, measurements by Lian et al. (2018) indicate that this is an underestimation, and suggest 0.62$\pm$0.06 as a global estimate.

[Figure 9 about here.]

When the observed average water use efficiency is plotted versus the average LE flux, a pattern emerges in which the sites are grouped per PFT (Fig. 10). A similar pattern was found in the ISBA simulations, but not in the ORCHIDEE simulations. The range in WUE across the test sites was much smaller in ORCHIDEE than in the observations.

[Figure 10 about here.]

The difference in water use efficiency can be attributed to differences in the modelled plant physiology, or the amount of drought stress experienced by the vegetation. As mentioned above, the rootzone soil moisture dropped significantly more frequently below field capacity in ISBA, compared to ORCHIDEE.

3.3 Evaluation of prognostic LAI and soil moisture

3.3.1 Sensitivity and error correlation

The sensitivity of the surface fluxes to soil moisture and LAI was quantified with a simple linear regression between their anomalies. The slope of these regressions indicates the strength of the response to the state variables.

It was found that the sensitivity of the fluxes to the soil moisture was strongly dependent on the land cover type, both in the observations and in the models (Fig. 11). A stronger response was found in the herbaceous sites, compared to the forest sites. ISBA and ORCHIDEE have a too high sensitivity to soil moisture, whereas the response in the diagnostic model was closer to that in the observations. In Fig. 12, the same data is plotted, but classified per aridity class. This illustrates the oversensitivity of ISBA and ORCHIDEE to drought stress. Despite their differences in implementation and parametrisation, a striking similarity in their sensitivity to drought was found, both for LE and GPP. The observations did not show an increase in sensitivity of GPP to S_e in dryer sites.

In the forest sites, the response of GPP to S_e anomalies is counter-intuitively negative. This might indicate that soil moisture anomalies in forest sites were more dominated by wet anomalies, associated with rainfall events. These events coincide with a reduction in solar radiation, hence resulting in a negative GPP response. In herbaceous sites soils were generally drier, so the positive impact of the reduced drought stress after the rainfall event was more dominant, resulting in a positive response. This behaviour was mimicked well in the models.

The sensitivity of LE and GPP to LAI was generally higher in the herbaceous sites. Here, the models tended to underestimate the sensitivity to LAI. In the forest sites, the sensitivity was lower according to the observations. The modelled sensitivity of LE to LAI was reasonably accurate, whereas the sensitivity of GPP to LAI was too strong.

[Figure 11 about here.]

[Figure 12 about here.]

To evaluate the impact of the quality of S_e and LAI on the simulated surface fluxes, the Spearman correlation of the errors in the state variables and the fluxes was calculated (Fig. 13). It was found in both ISBA and ORCHIDEE that LAI had a stronger error correlation to LE and GPP, compared to S_e. Grouped per dominant land cover type (Fig. 14), both models agree that the error correlation between LAI and GPP was higher in the herbaceous sites, compared to the forest sites. Notably, this was not the case for LAI-LE.

Furthermore, the errors in LE were strongest correlated to those in S_e for all models. The highest error correlation was found in DiagMod, where this was most pronounced for the herbaceous sites. In these sites, the S_e-GPP error correlation was also the strongest for DiagMod, whereas no strong S_e-GPP error correlation was found in the other models.

[Figure 13 about here.]

[Figure 14 about here.]

3.3.2 Functional evaluation with DiagMod

The simulated LE and GPP from the DiagMod runs with soil moisture and/or LAI from the prognostic models was validated with tower observations. The resulting NS is shown in Table 4 and Fig. 15. Similar tendencies were found in RMSE, Pearson r, and validation of the seasonal anomalies (not shown here). The DiagMod run with CGLS LAI and ERA5 soil moisture serves as a reference to evaluate the prognostic state variables.

Soil moisture had a stronger impact on LE, compared to LAI. A significant (Wilcoxon $p < 0.05$) reduction in NS was found when using soil moisture from ORCHIDEE. This effect is most pronounced in the herbaceous (more water-limited) sites. This is in contrast with the runs using soil moisture from ISBA, which seemed to improve the simulated LE in herbaceous sites (though not significantly). Similar, but smaller effects were found in the forest (less water-limited) sites. On the other hand, the opposite was found for the crop sites (n=4) where simulations with soil moisture from ISBA reduced the NS of the simulated LE significantly.

Despite strong differences in LAI, no significant impact was found on LE (with the exception of crop sites with ISBA LAI). A stronger sensitivity to LAI was found in the DiagMod simulations of GPP. A significant reduction of NS was found in all DiagMod runs, but most explicitly in the runs using the prognostic LAI. As in the simulations of LE, this was most pronounced for the herbaceous sites. The use of LAI from ISBA and ORCHIDEE strongly degraded the simulated GPP in these sites, whereas it was unaffected by injecting the prognostic soil moisture.

Overall these results are in line with the error correlations in Fig. 13. The higher error correlation of LAI to LE and GPP compared to the error correlations of soil moisture was confirmed. Additionally, the stronger impact of prognostic LAI on errors in GPP, and of soil moisture on LE was found in both analyses.

[Table 4 about here.]

[Figure 15 about here.]

4 Discussion

4.1 Model performance

The validation metrics of the three models were generally in agreement with previously performed local scale evaluations. Similar simulations with the diagnostic model were done in the validations reports of both the LSA SAF evapotranspiration and surface fluxes products (Ghilain et al., 2018) on one hand and the LSA SAF GPP product (Martínez et al., 2020) on the other hand. The accuracy and Pearson correlation obtained here were better than the ones previously reported. This can be attributed to the use of local forcings in this study, which are not used in the LSA SAF products. The weaker performance of the algorithm for the sensible heat flux was also identified by Ghilain et al. (2018).

The GPP product is a recent addition to the ensemble of LSA SAF MSG products. It was demonstrated to outperform similar products which also rely on the Montheith light-use efficiency method (Martínez et al., 2020). Here, it was found to perform consistently well for forest and herbaceous sites, and achieve a comparable model performance as ISBA.

In previous intercomparison studies at local scale (Balzarolo et al., 2014) or global scale (Friedlingstein et al., 2021), GPP was simulated more accurately with ORCHIDEE than with ISBA, but this was not confirmed here. Since these studies, substantial improvements have been made to ISBA: introduction of the MEB scheme, parametrization update, diffuse multilayer soil scheme, etc. (Boone et al., 2017; Delire et al., 2020). The introduction of the MEB scheme for forests on the energy fluxes was evaluated in-depth by Napoly et al. (2017) at local scale (though prognostic LAI was not included in that study). Substantial improvements to G and H were reported, thanks to the addition of an insulating litter layer. The introduction of the MEB scheme improved the mechanistic representation of the canopy, and issues due to a shared roughness length of the vegetation and bare soil in the composite scheme were circumvented. Our findings agree with that outcome, but the bias we found for LE, is not in agreement with previous findings.

4.1.1 Observation uncertainties

With the emergence of freely available data from eddy covariance networks, the use of local datasets is an increasingly standardized approach to evaluate the performance of land surface models (Balzarolo et al., 2014; Napoly et al., 2017; Williams et al., 2020; Chen et al., 2018; Joetzjer et al., 2015). However, the eddy covariance observations notoriously suffer from substantial biases and non-closure of the energy balance (Foken, 2008; Mauder et al., 2020). The non-closure of the energy balance is attributed to 1) large advective fluxes caused by surface heterogeneities, 2) systematic measurement errors due to mismatch in observation footprint or inadequate sample rate, 3) thermal processes, such as heat storage or vegetation metabolism (Mauder et al., 2020; Chu et al., 2021; Liu et al., 2021). The test sites in this study were selected to have a relatively homogeneous landcover. Regardless, the resulting uncertainty in the observations was of the same order of magnitude as the model errors. The turbulent fluxes are typically underestimated, as is the GPP (Massman and Lee, 2002; Gao et al., 2019). Note that GPP is not corrected for this possible bias in the ONEFLUX processing pipeline (Pastorello et al., 2020). Furthermore, some studies have indicated that the eddy covariance observations are closer to lysimeter data if the energy balance is closed by correcting

LE only (Wohlfahrt et al., 2010). Considering this, the negative bias of the simulated LE (and GPP) in this study could be even underestimated. Conversely, others suggest that most or all of the deficit might be related to H (Ingwersen et al., 2011), or found a good match with independent reference data without LE correction (Graf et al., 2014). Validation results of the turbulent fluxes without energy balance closure correction are given in the supplement material.

4.1.2 Forest vs Herbaceous

Generally, the differences between the accuracy of the simulated surface fluxes was most distinct in the sites dominated by herbaceous vegetation (excluding crop sites). These sites have the most pronounced inter-annual variability, and seasonal anomalies are strongly driven by precipitation events (Weber et al., 2009). This can be largely attributed to their natural occurrence in dryer climates and shallower root system, compared to forests. The seasonal cycle of LAI in the herbaceous sites, and its variability, was simulated poorly with the prognostic models. The error correlation analysis indicated that these errors were strongly related to errors in the surface fluxes.

In the crop sites, management practices were missing in the prognostic models. In the mean annual cycle of LAI (Fig. 7), it is evident that no harvest occurs. Despite this, the simulations of LE were not significantly less accurate compared to other land cover types. After harvest, LE consists largely out of bare soil evaporation. Though vegetation was still present in the models, the bulk LE was still reasonably accurate. More evident degradation of the results was found in GPP after harvest, which was overestimated. Even in the diagnostic model, where management practices were incorporated implicitly in the forcing variables, GPP was overestimated. Notably, despite the missing management practices in the prognostic models, the quality of the simulated LE and GPP (and their anomalies) was not significantly different from that in natural herbaceous sites.

Still, the diagnostic model performed consistently well for all types of land cover, contrary to the prognostic models. Only the seasonal variability of GPP in forest sites was simulated less accurately than with the prognostic models. Whereas the remote-sensing based observations adequately captured this variability for the herbaceous and crop sites, they seemed to fall short for the forest sites.

4.2 Interactions

LAI and soil moisture are two key variables in the interaction between water, energy and vegetation. Though our understanding of the involved processes at leaf-level scale is advanced, it remains challenging to scale these relations to the canopy level. This was illustrated by erroneous sensitivity of the models to LAI and soil moisture. As in previous studies, the sensitivity of LE and GPP to soil moisture was generally overestimated (Piao et al., 2013; Huang et al., 2016) in ISBA and ORCHIDEE, whereas the diagnostic model represented the observed sensitivity relatively well.

The interplay between LE and LAI was analysed in detail by Forzieri et al. (2018, 2020). The estimated global sensitivity of LE to LAI ($3.66 \pm 0.45\ \mathrm{W\,m^{-2}}$ / $\mathrm{m^2\,m^{-2}}$, according to Forzieri et al., 2020) is lower than the one reported here, but the applied methodology was not the same. Contrary to our study, anomalies due to climatic drivers (i.e. precipitation, temperature etc.)

were factored out, resulting in a different response. The oversensitivity of LE to LAI in ORCHIDEE was also not confirmed in our study. Still, in accordance to these studies, a stronger response between LE and LAI was found for herbaceous/soil moisture supply-driven sites, compared to forest/demand-driven sites.

Despite the differences in their architecture and parametrization, ISBA and ORCHIDEE demonstrated similar behaviour in the interaction between water, energy and vegetation. Comparable sensitivities and error correlations were found in both models, indicating that they share common weaknesses in their implementation.

4.2.1 LAI

The errors in the surface fluxes were strongly correlated to errors in LAI for both prognostic models, even though their sensitivity to LAI reflects the observed sensitivity reasonably well (compared to the sensitivity to soil moisture). This seemed to indicate that the source of the errors in the fluxes lies in the feedback mechanism between GPP and LAI (i.e. biomass allocation and phenology), rather than in the forward link between GPP and LAI (i.e. photosynthesis and leaf to canopy upscaling).

The prognostic simulation of LAI in ISBA was introduced by Gibelin et al. (2006) and uses a fairly simple scheme. The latest update was the revision of plant trait parameters according to the TRY database (Kattge et al., 2011; Delire et al., 2020). It has frequently been reported that the seasonal cycle of the simulated LAI in ISBA is delayed by a month or more (Lafont et al., 2012; Gibelin et al., 2008; Joetzjer et al., 2015). Delire et al. (2020) attributes this to the leaf longevity parameter, and (Szczypta et al., 2014) mentions the vegetation undergrowth dynamics as a possible cause for the mismatch between remote-sensing based LAI and the prognostic LAI in LSM. However, the issue seems to be related to the architecture of the biomass allocation scheme as well. The assimilated carbon is attributed to the leaf biomass pool first, from where it trickles down to the other pools. No carbon reserve dynamics are implemented. The consequence is that the simulated LAI in ISBA starts slow during spring, as GPP is underestimated due to a low LAI. It continues to build up LAI until late in the second half of the season, when photosynthetic conditions become suboptimal, and leaf senescence is triggered. In contrast, the observed seasonal LAI cycles reach a maximum in the first half of the growing season.

The functional evaluation with the diagnostic model demonstrated that a fairly simple model is capable of simulating the surface fluxes accurately, given accurate observations of LAI. The prognostic LAI generally degraded the results, compared to simulations with remotely sensed LAI. Data assimilation experiments have demonstrated the potential of remotely sensed LAI to improve the surface fluxes (Albergel et al., 2017). Improvements to prognostic LAI schemes are required to increase the skill of the LSM to simulate surface fluxes.

In that context, processes from ORCHIDEE and other LSM could be adopted to improve the fairly simple biomass allocations scheme in ISBA. The importance of non-structural carbohydrates to capture the leaf phenology in LSM is well-known, though rarely implemented (Asaadi et al., 2018). Fatichi et al. (2019) indicates that a full-grown canopy of a deciduous broadleaf forest contains approximately 30% of the total yearly assimilated carbon, yet it is grown in 1 month (1/5 - 1/6 of the growing season). This rough simplification illustrates that reserve dynamics are essential to simulate the seasonal cycle of the vegetation accurately. Such dynamics are implemented in ORCHIDEE: once certain phenological criteria are fulfilled, the carbon in

a reserve pool is allocated to leaf biomass to kickstart the phenological cycle. Still, despite the dedicated phenology modules, non-structural carbohydrates reserve dynamics, and a more advanced leaf demography, simulating LAI remained challenging in ORCHIDEE. The timing of the phenological cycle was more accurate in ORCHIDEE, though the accuracy of the simulated LAI was significantly poorer than ISBA. This was the case in particular for herbaceous vegetation. This tendency towards delayed phenology (and in particular a delayed leaf senescence) is found in most earth system models in CMIP5 and CMIP6 (Park and Jeong, 2021; Song et al., 2021).

The discrepancy in complexity between the modelling of photosynthesis and that of the biomass allocation has been highlighted by several authors (Fatichi et al., 2016; Friend et al., 2019), though the main challenge lies in the parametrization of those processes. The allocation of carbon in terrestrial vegetation is an important knowledge gap, hindering the advancement of earth system models.

Finally, there are several important differences between the remote-sensing based vegetation and the idealized vegetation in the models which need to be recognized when comparing both. Firstly, the role of the understory has a well-known impact on the remote-sensing based LAI (Camacho et al., 2013), whereas the LSM do not consider the separate evolution of an understory. This can result in substantial differences in the seasonal cycle of LAI. This was illustrated by the differences in the simulations and observations of the LAI cycle at ENF sites. Continuous in situ LAI observations with hemispherical photography in ENF sites are rare, but Rautiainen et al. (2012) reported that the effective canopy LAI (including non-green foliage) in FI-Hyy (boreal ENF site) remained constant from June till mid-September. This is in agreement with the flat LAI cycle for ENF in ORCHIDEE, but is in contrast with the remote-sensing based LAI and the prognostic LAI in ISBA. In an empirical model based on in situ observations for the FR-LBr site, LAI demonstrated a seasonal cycle. The understory was responsible for most of the seasonal variation, and 30% of the LAI was attributed to the understory during the summer (Rivalland, 2003). The seasonal cycle in the remote-sensing based LAI seems exaggerated (ranging between 1 $\mathrm{m^2\,m^{-2}}$ in winter and 4 $\mathrm{m^2\,m^{-2}}$ in summer). However, considering the understory and seasonal variation in needleleaf greenness (Seyednasrollah et al., 2021), assuming a flat LAI does not seem accurate nether, in the context of simulating GPP.

This brings up a second issue: the remote-sensing based LAI is the 'green' LAI, i.e. photosynthetically active leaves (Camacho et al., 2013). Whereas LAI in LSM is a key variable which wears many hats. A single LAI variable is used to represent the role of leaves in several processes (photosynthesis, interception, canopy radiative transfer, surface roughness, etc.), in which the greenness of the canopy is not always important. These discrepancies contribute to the mismatch between LAI in the observations and the models. Addressing them might further advance the representation of vegetation in LSM.

4.2.2 Soil moisture

A significant difference between ISBA and ORCHIDEE is found in the simulated soil moisture dynamics, the water partitioning and the water use efficiency. The simulated WUE in ISBA was in fair agreement with what is deduced from the eddy covariance observations. In contrast, the WUE in ORCHIDEE had a much narrower range. The comparison of the LE partitioning shows also that a larger fraction of the water was transpired in ORCHIDEE, compared to ISBA. The differences in WUE and flux partitioning could be attributed to differences in the simulated plant physiology, or to the quality of the simu-

lated soil water content. The variability of the simulated water content in ORCHIDEE was strongly underestimated, and the vegetation experienced significantly less drought stress in ORCHIDEE. It is likely that this translated to a low variability in WUE as well. Furthermore, a substantial part of the precipitation was lost as surface runoff, compared to ISBA. Though we did not have validation data to evaluate the water partitioning, it seems that the simulations of ORCHIDEE could be improved significantly by addressing the soil moisture dynamics. The superior simulation of soil moisture in ISBA contributes to the good performance in simulating the surface fluxes, in particular for sites with herbaceous vegetation and water-driven climate. The functional evaluation demonstrated that the prognostic soil moisture from ISBA even resulted in an improvement of the simulated LE for these sites, compared to simulations with ERA5 soil moisture.

Overall, the accurate simulation of soil moisture and water infiltration is a challenge (perhaps one of the main challenges) in land surface models (Vereecken et al., 2019). The poor quality of the simulated soil moisture compared to in situ observations is also evident in this study, despite the use of the multi-layer diffuse water transport scheme. The soil physical parameters are determined using a global pedotransfer function (PTF), using only texture as input. New, advanced PTF with global coverage have emerged in recent years, using not only texture, but also climatology and land use as predictors (Gupta et al., 2021). As soil moisture is at the basis of many processes in LSM, incorporating these PTF seems the logical new step forward in LSM (Fatichi et al., 2020).

The local scale simulations in this study were not coupled to a hydrological model, thus ground water dynamics were lacking. Though only a limited effect of capillary rise was found in studies with a coupled groundwater hydrology, the impact can be non-negligible for forest ecosystems with a deep root system (Decharme et al., 2019; MacBean et al., 2020). The further development of ground water dynamics in LSM is indispensable for the accurate coupling of energy, water and carbon in forest vegetation and its response to severe drought events.

Several efforts have already explored the potential of improving soil moisture dynamics in LSM. Substantial improvements to soil moisture have indeed been obtained by calibrating the pedotransfer functions, or soil physical parameters. Yet, the impact thereof on the surface fluxes has been found to be relatively limited (Pinnington et al., 2021), or in some cases even negative (Raoult et al., 2021). Though many parameters in ISBA and ORCHIDEE are derived from databases (Delire et al., 2020), the LSM have been calibrated to produce accurate surface fluxes using (amongst others) eddy covariance observations. The limited accuracy of the soil moisture dynamics might have been overcompensated in the resulting parametrization (Raoult et al., 2021). The oversensitivity to drought stress in ISBA and ORCHIDEE is possibly an illustration of this. Improvements to the intricate network of gears under the hood of LSM are a delicate matter. Addressing the soil moisture dynamics should go hand in hand with corrections to the oversensitivity to drought stress.

5 Conclusions

Three land surface models were compared at local scale, using identical meteorological forcing and prescribed land cover. The goal was to evaluate their skill to simulate surface fluxes (LE and GPP), as well as their simulated interaction between water, energy and vegetation. It was found that the diagnostic model (based on LSA SAF algorithms) performed consistently

well for all land covers. The prognostic models (ISBA and ORCHIDEE) performed similarly well for the forest sites, but the simulations for herbaceous sites revealed some important shortcomings. The sensitivity analysis demonstrated that both models overestimate the sensitivity to drought stress, which was occurring most frequently in herbaceous sites. On the other hand, the error analysis showed that errors in the prognostic LAI (and not soil moisture) were the dominant source of errors for LE and GPP in ISBA and ORCHIDEE. This was underlined by the functional evaluation with the diagnostic model. Given the acceptable sensitivity to LAI, the source of these errors is likely found in the feedback mechanism between GPP and LAI. Compared to observations, the simulated phenological cycle in both models was delayed and failed to capture the observed seasonal variability. Processes describing carbon reserve dynamics during spring and leaf senescence were found to be falling short or missing. Improvements in the leaf phenology and biomass allocation scheme are required to improve the simulated surface fluxes.

The analysis here demonstrated key strengths and weaknesses of each LSM. Most notably, we showed that ISBA and ORCHIDEE shared key deficiencies concerning the coupling of the water, energy and vegetation, despite their differences in architecture and parametrization. Improving the feedback between GPP and LAI, the soil moisture dynamics, and the oversensitivity to drought might advance the performance of these LSM significantly.

Code and data availability. The scripts and datasets used in this study are freely available upon request to the authors.

Author contributions. JDP: Conceptualization, Investigation, Analysis, Writing - original draft preparation; MB and LL: Investigation, Analysis, Writing - review & editing;PC, AA, RH: Writing - review & editing; MB, FM, FGM: Supervision, Project administration, Writing - review & editing

Competing interests. The authors declare to have no competing interests

Acknowledgements. The research presented in this paper is funded by BELSPO (Belgian Science Policy Office) in the frame of the STEREO III programme – project ECOPROPHET (SR/00/334), and co-funded by EUMETSAT (LSA SAF program for CDOP-3) and the Belspo/ESA Prodex Program (PEA 4000110695). This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The FLUXNET eddy covariance data processing and harmonization was carried out by the ICOS Ecosystem Thematic Center, AmeriFlux Management Project and Fluxdata project of FLUXNET, with the support of CDIAC, and the OzFlux, ChinaFlux and AsiaFlux offices. The assistance of the researchers who developed and maintain

LSA-SAF, Surfex and ORCHIDEE was much appreciated. This work stands on the shoulders of the many who offer free, open source data, knowledge and tools. We thank Sci-hub for making scientific knowledge available to everyone.

References

Albergel, C., Munier, S., Leroux, D. J., Dewaele, H., Fairbairn, D., Barbu, A. L., Gelati, E., Dorigo, W., Faroux, S., Meurey, C., et al.: Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8. 0: LDAS-Monde assessment over the Euro-Mediterranean area, Geoscientific Model Development, 10, 3889–3912, 2017.

Alton, P. B.: The sensitivity of models of gross primary productivity to meteorological and leaf area forcing: A comparison between a Penman–Monteith ecophysiological approach and the MODIS Light-Use Efficiency algorithm, Agricultural and forest meteorology, 218, 11–24, 2016.

Asaadi, A., Arora, V. K., Melton, J. R., and Bartlett, P.: An improved parameterization of leaf area index (LAI) seasonality in the Canadian Land Surface Scheme (CLASS) and Canadian Terrestrial Ecosystem Model (CTEM) modelling framework, Biogeosciences, 15, 6885–6907, 2018.

Balzarolo, M., Boussetta, S., Balsamo, G., Beljaars, A., Maignan, F., Calvet, J.-C., Lafont, S., Barbu, A., Poulter, B., Chevallier, F., et al.: Evaluating the potential of large-scale simulations to predict carbon fluxes of terrestrial ecosystems over a European Eddy Covariance network, Biogeosciences, 11, 2661–2678, 2014.

Barrios, J. M., Arboleda, A., De Pue, J., Chormanski, J., and Gellens-Meulenberghs, F.: Continuous Daily Evapotranspiration with Optical Spaceborne Observations at Sub-Kilometre Spatial Resolution, Remote Sensing, 12, 2218, 2020.

Bonan, G., Williams, M., Fisher, R., and Oleson, K.: Modeling stomatal conductance in the earth system: linking leaf water-use efficiency and water transport along the soil–plant–atmosphere continuum, Geoscientific Model Development, 7, 2193–2222, 2014.

Boone, A., Samuelsson, P., Gollvik, S., Napoly, A., Jarlan, L., Brun, E., and Decharme, B.: The interactions between soil–biosphere–atmosphere land surface model with a multi-energy balance (ISBA-MEB) option in SURFEXv8–Part 1: Model description, Geoscientific Model Development, 10, 843–872, 2017.

Botta, A., Viovy, N., Ciais, P., Friedlingstein, P., and Monfray, P.: A global prognostic scheme of leaf onset using satellite data, Global Change Biology, 6, 709–725, 2000.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., et al.: Presentation and evaluation of the IPSL-CM6A-LR climate model, Journal of Advances in Modeling Earth Systems, 12, e2019MS002 010, 2020.

Boussetta, S., Balsamo, G., Beljaars, A., Panareda, A.-A., Calvet, J.-C., Jacobs, C., van den Hurk, B., Viterbo, P., Lafont, S., Dutra, E., et al.: Natural land carbon dioxide exchanges in the ECMWF integrated forecasting system: Implementation and offline validation, Journal of Geophysical Research: Atmospheres, 118, 5923–5946, 2013.

Calvet, J.-C. and Soussana, J.-F.: Modelling CO2-enrichment effects using an interactive vegetation SVAT scheme, Agricultural and forest meteorology, 108, 129–152, 2001.

Calvet, J.-C., Noilhan, J., Roujean, J.-L., Bessemoulin, P., Cabelguenne, M., Olioso, A., and Wigneron, J.-P.: An interactive vegetation SVAT model tested against data from six contrasting sites, Agricultural and Forest Meteorology, 92, 73–95, 1998.

Calvet, J.-C., Rivalland, V., Picon-Cochard, C., and Guehl, J.-M.: Modelling forest transpiration and CO2 fluxes—Response to soil moisture stress, Agricultural and forest meteorology, 124, 143–156, 2004.

Camacho, F., Cernicharo, J., Lacaze, R., Baret, F., and Weiss, M.: GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference products, Remote Sensing of Environment, 137, 310–329, 2013.

Carrer, D., Roujean, J.-L., Lafont, S., Calvet, J.-C., Boone, A., Decharme, B., Delire, C., and Gastellu-Etchegorry, J.-P.: A canopy radiative transfer scheme with explicit FAPAR for the interactive vegetation model ISBA-A-gs: Impact on carbon fluxes, Journal of Geophysical Research: Biogeosciences, 118, 888–903, 2013.

Carsel, R. F. and Parrish, R. S.: Developing joint probability distributions of soil water retention characteristics, Water resources research, 24, 755–769, 1988.

Chen, L., Dirmeyer, P. A., Guo, Z., and Schultz, N. M.: Pairing FLUXNET sites to validate model representations of land-use/land-cover change, Hydrology and Earth System Sciences, 22, 111–125, 2018.

Cheruy, F., Ducharne, A., Hourdin, F., Musat, I., Vignon, É., Gastineau, G., Bastrikov, V., Vuichard, N., Diallo, B., Dufresne, J.-L., et al.: Improved near-surface continental climate in IPSL-CM6A-LR by combined evolutions of atmospheric and land surface physics, Journal of Advances in Modeling Earth Systems, 12, e2019MS002 005, 2020.

Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., Torn, M. S., Metzger, S., Kumar, J., Arain, M. A., et al.: Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites, Agricultural and Forest Meteorology, 301, 108 350, 2021.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., et al.: Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, K., Tignor, M., Allen, SK, Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, PM, Eds, 2013.

Clapp, R. B. and Hornberger, G. M.: Empirical equations for some soil hydraulic properties, Water resources research, 14, 601–604, 1978.

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

Cui, W. and Chui, T. F. M.: Temporal and spatial variations of energy balance closure across FLUXNET research sites, Agricultural and Forest Meteorology, 271, 12–21, 2019.

De Rosnay, P., Polcher, J. d., Bruen, M., and Laval, K.: Impact of a physically based soil water flow and soil-plant interaction representation for modeling large-scale land surface processes, Journal of Geophysical Research: Atmospheres, 107, ACL–3, 2002.

Decharme, B., Delire, C., Minvielle, M., Colin, J., Vergnes, J.-P., Alias, A., Saint-Martin, D., Séférian, R., Sénési, S., and Voldoire, A.: Recent changes in the ISBA-CTRIP land surface system for use in the CNRM-CM6 climate model and in global off-line hydrological applications, Journal of Advances in Modeling Earth Systems, 11, 1207–1252, 2019.

Delire, C., Séférian, R., Decharme, B., Alkama, R., Calvet, J.-C., Carrer, D., Gibelin, A.-L., Joetzjer, E., Morel, X., Rocher, M., et al.: The global land carbon cycle simulated with ISBA-CTRIP: improvements over the last decade, Journal of Advances in Modeling Earth Systems, 12, e2019MS001 886, 2020.

Dickinson, R. E., Shaikh, M., Bryant, R., and Graumlich, L.: Interactive canopies for a climate model, Journal of Climate, 11, 2823–2836, 1998.

Dirmeyer, P. A., Chen, L., Wu, J., Shin, C.-S., Huang, B., Cash, B. A., Bosilovich, M. G., Mahanama, S., Koster, R. D., Santanello, J. A., et al.: Verification of land–atmosphere coupling in forecast models, reanalyses, and land surface models using flux site observations, Journal of hydrometeorology, 19, 375–392, 2018.

d'Orgeval, T., Polcher, J., and Rosnay, P. d.: Sensitivity of the West African hydrological cycle in ORCHIDEE to infiltration processes, Hydrology and Earth System Sciences, 12, 1387–1401, 2008.

Drought 2018 Team and ICOS Ecosystem Thematic Centre: Drought-2018 ecosystem eddy covariance flux product in FLUXNET-Archive format - release 2019-1, https://doi.org/10.18160/PZDK-EF78, 2019.

Ducoudré, N. I., Laval, K., and Perrier, A.: SECHIBA, a new set of parameterizations of the hydrologic exchanges at the land-atmosphere interface within the LMD atmospheric general circulation model, Journal of Climate, pp. 248–273, 1993.

Falkenmark, M., Rockstrom, J., and Rockström, J.: Balancing water for humans and nature: the new approach in ecohydrology, Earthscan, 2004.

Faroux, S., Kaptué Tchuenté, A., Roujean, J.-L., Masson, V., Martin, E., and Moigne, P. L.: ECOCLIMAP-II/Europe: A twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models, Geoscientific Model Development, 6, 563–582, 2013.

Farquhar, G. D., von Caemmerer, S. v., and Berry, J. A.: A biochemical model of photosynthetic CO 2 assimilation in leaves of C 3 species, Planta, 149, 78–90, 1980.

Fatichi, S., Pappas, C., and Ivanov, V. Y.: Modeling plant–water interactions: an ecohydrological overview from the cell to the global scale, Wiley Interdisciplinary Reviews: Water, 3, 327–368, 2016.

Fatichi, S., Pappas, C., Zscheischler, J., and Leuzinger, S.: Modelling carbon sources and sinks in terrestrial vegetation, New Phytologist, 221, 652–668, 2019.

Fatichi, S., Or, D., Walko, R., Vereecken, H., Young, M. H., Ghezzehei, T. A., Hengl, T., Kollet, S., Agam, N., and Avissar, R.: Soil structure is an important omission in Earth System Models, Nature communications, 11, 1–11, 2020.

Fisher, R. A. and Koven, C. D.: Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems, Journal of Advances in Modeling Earth Systems, 12, e2018MS001 453, 2020.

Foken, T.: The energy balance closure problem: an overview, Ecological Applications, 18, 1351–1367, 2008.

Forzieri, G., Duveiller, G., Georgievski, G., Li, W., Robertson, E., Kautz, M., Lawrence, P., Garcia San Martin, L., Anthoni, P., Ciais, P., et al.: Evaluating the interplay between biophysical processes and leaf area changes in Land Surface Models, Journal of advances in modeling earth systems, 10, 1102–1126, 2018.

Forzieri, G., Miralles, D. G., Ciais, P., Alkama, R., Ryu, Y., Duveiller, G., Zhang, K., Robertson, E., Kautz, M., Martens, B., et al.: Increased control of vegetation on global terrestrial energy fluxes, Nature Climate Change, 10, 356–362, 2020.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C., Hauck, J., Le Quéré, C., Peters, G. P., Peters, W., Pongratz, J., et al.: Global Carbon Budget 2021, Earth System Science Data Discussions, pp. 1–191, 2021.

Friend, A. D., Eckes-Shephard, A. H., Fonti, P., Rademacher, T. T., Rathgeber, C. B., Richardson, A. D., and Turton, R. H.: On the need to consider wood formation processes in global vegetation models and a suggested approach, Annals of Forest Science, 76, 1–13, 2019.

Gao, Z., Liu, H., Missik, J. E., Yao, J., Huang, M., Chen, X., Arntzen, E., and Mcfarland, D. P.: Mechanistic links between underestimated CO2 fluxes and non-closure of the surface energy balance in a semi-arid sagebrush ecosystem, Environmental Research Letters, 14, 044 016, 2019.

Garbulsky, M. F., Peñuelas, J., Papale, D., Ardö, J., Goulden, M. L., Kiely, G., Richardson, A. D., Rotenberg, E., Veenendaal, E. M., and Filella, I.: Patterns and controls of the variability of radiation use efficiency and primary productivity across terrestrial ecosystems, Global Ecology and Biogeography, 19, 253–267, 2010.

Ghilain, N., Arboleda, A., and Gellens-Meulenberghs, F.: Evapotranspiration modelling at large scale using near-real time MSG SEVIRI derived data, Hydrology and Earth System Sciences, 15, 771–786, 2011.

Ghilain, N., Arboleda, A., Sepulcre-Cantò, G., Batelaan, O., Ardö, J., and Gellens-Meulenberghs, F.: Improving evapotranspiration in a land surface model using biophysical variables derived from MSG/SEVIRI satellite, Hydrology and Earth System Sciences, 16, 2567–2583, 2012.

Ghilain, N., Arboleda, A., and Meulenberghs, F.: Validation Report Evapotranspiration and Surface Fluxes, PRODUCTS: LSA-311 (MET V2), LSA-312 (DMET V2) LSA-304 (MH), LSA-305 (MLE), Tech. Rep. SAF/LAND/RMI/VR/1.1, EUMETSAT LSA SAF, 2018.

Gibelin, A.-L., Calvet, J.-C., Roujean, J.-L., Jarlan, L., and Los, S. O.: Ability of the land surface model ISBA-A-gs to simulate leaf area index at the global scale: Comparison with satellites products, Journal of Geophysical Research: Atmospheres, 111, 2006.

Gibelin, A.-L., Calvet, J.-C., and Viovy, N.: Modelling energy and CO2 fluxes with an interactive vegetation land surface model-Evaluation at high and middle latitudes, agricultural and forest meteorology, 148, 1611–1628, 2008.

Goudriaan, J., Van Laar, H., Van Keulen, H., and Louwerse, W.: Photosynthesis, CO 2 and plant production, in: Wheat growth and modelling, pp. 107–122, Springer, 1985.

Graf, A., Bogena, H. R., Drüe, C., Hardelauf, H., Pütz, T., Heinemann, G., and Vereecken, H.: Spatiotemporal relations between water budget components and soil water content in a forested tributary catchment, Water resources research, 50, 4837–4857, 2014.

Gupta, S., Lehmann, P., Bonetti, S., Papritz, A., and Or, D.: Global Prediction of Soil Saturated Hydraulic Conductivity Using Random Forest in a Covariate-Based GeoTransfer Function (CoGTF) Framework, Journal of Advances in Modeling Earth Systems, 13, e2020MS002 242, 2021.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, 2020.

Huang, Y., Gerber, S., Huang, T., and Lichstein, J. W.: Evaluating the drought response of CMIP5 models using global gross primary productivity, leaf area, precipitation, and soil moisture data, Global Biogeochemical Cycles, 30, 1827–1846, 2016.

Ingwersen, J., Steffens, K., Högy, P., Warrach-Sagi, K., Zhunusbayeva, D., Poltoradnev, M., Gäbler, R., Wizemann, H.-D., Fangmeier, A., Wulfmeyer, V., et al.: Comparison of Noah simulations with eddy covariance and soil water measurements at a winter wheat stand, Agricultural and Forest Meteorology, 151, 345–355, 2011.

Jacobs, C., Van den Hurk, B., and De Bruin, H.: Stomatal behaviour and photosynthetic rate of unstressed grapevines in semi-arid conditions, Agricultural and Forest Meteorology, 80, 111–134, 1996.

Jarvis, P.: The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field, Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 273, 593–610, 1976.

Joetzjer, E., Delire, C., Douville, H., Ciais, P., Decharme, B., Carrer, D., Verbeeck, H., Weirdt, M. D., and Bonal, D.: Improving the ISBA CC land surface model simulation of water and carbon fluxes and stocks over the Amazon forest, Geoscientific Model Development, 8, 1709–1727, 2015.

Kattge, J. and Knorr, W.: Temperature acclimation in a biochemical model of photosynthesis: a reanalysis of data from 36 species, Plant, cell & environment, 30, 1176–1190, 2007.

Kattge, J., Diaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P. B., Wright, I. J., et al.: TRY–a global database of plant traits, Global change biology, 17, 2905–2935, 2011.

Kauwe, M. G. D., Medlyn, B. E., Knauer, J., and Williams, C. A.: Ideas and perspectives: how coupled is the vegetation to the boundary layer?, Biogeosciences, 14, 4435–4453, 2017.

Kolus, H. R., Huntzinger, D. N., Schwalm, C. R., Fisher, J. B., McKay, N., Fang, Y., Michalak, A. M., Schaefer, K., Wei, Y., Poulter, B., et al.: Land carbon models underestimate the severity and duration of drought's impact on plant productivity, Scientific reports, 9, 1–10, 2019.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochemical Cycles, 19, 2005.

Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., and Richardson, A.: Constraining a global ecosystem model with multi-site eddy-covariance data, Biogeosciences, 9, 3757–3776, 2012.

Kuppel, S., Peylin, P., Maignan, F., Chevallier, F., Kiely, G., Montagnani, L., and Cescatti, A.: Model–data fusion across ecosystems: from multisite optimizations to global simulations, Geoscientific Model Development, 7, 2581–2597, 2014.

Lafont, S., Zhao, Y., Calvet, J.-C., Peylin, P., Ciais, P., Maignan, F., and Weiss, M.: Modelling LAI, surface water and carbon fluxes at high-resolution over France: comparison of ISBA-A-gs and ORCHIDEE, Biogeosciences, 9, 439–456, 2012.

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., et al.: The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty, Journal of Advances in Modeling Earth Systems, 11, 4245–4287, 2019.

Le Moigne, P., Boone, A., Calvet, J., Decharme, B., Faroux, S., Gibelin, A., Lebeaupin, C., Mahfouf, J., Martin, E., Masson, V., et al.: SURFEX v8.1 scientific documentation, Note de centre (CNRM/GMME), Météo-France, Toulouse, France, 2018.

Le Roux, X., Lacointe, A., Escobar-Gutiérrez, A., and Le Dizès, S.: Carbon-based models of individual tree growth: a critical appraisal, Annals of Forest Science, 58, 469–506, 2001.

Li, L., Wang, Y., Arora, V. K., Eamus, D., Shi, H., Li, J., Cheng, L., Cleverly, J., Hajima, T., Ji, D., et al.: Evaluating global land surface models in CMIP5: Analysis of ecosystem water-and light-use efficiencies and rainfall partitioning, Journal of Climate, 31, 2995–3008, 2018.

Lian, X., Piao, S., Huntingford, C., Li, Y., Zeng, Z., Wang, X., Ciais, P., McVicar, T. R., Peng, S., Ottlé, C., et al.: Partitioning global land evapotranspiration using CMIP5 models constrained by observations, Nature Climate Change, 8, 640–646, 2018.

Liu, H., Gao, Z., and Katul, G. G.: Non-Closure of Surface Energy Balance Linked to Asymmetric Turbulent Transport of Scalars by Large Eddies, Journal of Geophysical Research: Atmospheres, p. e2020JD034474, 2021.

MacBean, N., Maignan, F., Peylin, P., Bacour, C., Bréon, F.-M., and Ciais, P.: Using satellite data to improve the leaf phenology of a global terrestrial biosphere model, Biogeosciences, 12, 7185–7208, 2015.

MacBean, N., Scott, R. L., Biederman, J. A., Ottlé, C., Vuichard, N., Ducharne, A., Kolb, T., Dore, S., Litvak, M., and Moore, D. J.: Testing water fluxes and storage from two hydrology configurations within the ORCHIDEE land surface model across US semi-arid sites, Hydrology and Earth System Sciences, 24, 5203–5230, 2020.

Maleki, M., Arriga, N., Barrios, J. M., Wieneke, S., Liu, Q., Peñuelas, J., Janssens, I. A., and Balzarolo, M.: Estimation of Gross Primary Productivity (GPP) Phenology of a Short-Rotation Plantation Using Remotely Sensed Indices Derived from Sentinel-2 Images, Remote Sensing, 12, 2104, 2020.

Martínez, B., Gilabert, M., Sánchez-Ruiz, S., Campos-Taberner, M., García-Haro, F., Brümmer, C., Carrara, A., Feig, G., Grünwald, T., Mammarella, I., et al.: Evaluation of the LSA-SAF gross primary production product derived from SEVIRI/MSG data (MGPP), ISPRS Journal of Photogrammetry and Remote Sensing, 159, 220–236, 2020.

Massman, W. and Lee, X.: Eddy covariance flux corrections and uncertainties in long-term studies of carbon and energy exchanges, Agricultural and Forest Meteorology, 113, 121–144, 2002.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The

SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of Earth surface variables and fluxes, Geoscientific Model Development, 6, 929–960, 2013.

Mauder, M., Foken, T., and Cuxart, J.: Surface-energy-balance closure over land: a review, Boundary-Layer Meteorology, 177, 395–426, 2020.

Medlyn, B., Dreyer, E., Ellsworth, D., Forstreuter, M., Harley, P., Kirschbaum, M., Le Roux, X., Montpied, P., Strassemeyer, J., Walcroft, A., et al.: Temperature response of parameters of a biochemically based model of photosynthesis. II. A review of experimental data, Plant, Cell & Environment, 25, 1167–1179, 2002.

Monsi, M. and Saeki, T.: On the factor light in plant communities and its importance for matter production, Annals of botany, 95, 549, 2005.

Monteith, J.: Solar radiation and productivity in tropical ecosystems, Journal of applied ecology, 9, 747–766, 1972.

Nachtergaele, F., van Velthuizen, H., Verelst, L., Batjes, N., Dijkshoorn, K., van Engelen, V., Fischer, G., Jones, A., and Montanarela, L.: The harmonized world soil database, in: Proceedings of the 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia, 1-6 August 2010, pp. 34–37, 2010.

Napoly, A., Boone, A., Samuelsson, P., Gollvik, S., Martin, E., Seferian, R., Carrer, D., Decharme, B., and Jarlan, L.: The interactions between soil–biosphere–atmosphere (ISBA) land surface model multi-energy balance (MEB) option in SURFEXv8–Part 2: Introduction of a litter formulation and model evaluation for local-scale forest sites, Geoscientific Model Development, 10, 1621–1644, 2017.

Nelson, J. A., Pérez-Priego, O., Zhou, S., Poyatos, R., Zhang, Y., Blanken, P. D., Gimeno, T. E., Wohlfahrt, G., Desai, A. R., Gioli, B., et al.: Ecosystem transpiration and evaporation: Insights from three water flux partitioning methods across FLUXNET sites, Global change biology, 26, 6916–6930, 2020.

Papagiannopoulou, C., Gonzalez Miralles, D., Demuzere, M., Verhoest, N., and Waegeman, W.: Global hydro-climatic biomes identified via multitask learning, Geoscientific Model Development, 11, 4139–4153, 2018.

Park, H. and Jeong, S.: Leaf area index in Earth system models: how the key variable of vegetation seasonality works in climate projections, Environmental Research Letters, 16, 034 027, 2021.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., et al.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Scientific data, 7, 1–27, 2020.

Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P., Kuppel, S., Koffi, E., Kane, A., Maignan, F., Chevallier, F., et al.: A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle, Geoscientific Model Development, 9, 3321–3346, 2016.

Piao, S., Sitch, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., Ahlström, A., Anav, A., Canadell, J. G., Cong, N., et al.: Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO2 trends, Global change biology, 19, 2117–2132, 2013.

Pielke, R. A., Avissar, R., Raupach, M., Dolman, A. J., Zeng, X., and Denning, A. S.: Interactions between the atmosphere and terrestrial ecosystems: influence on weather and climate, Global change biology, 4, 461–475, 1998.

Pinnington, E., Amezcua, J., Cooper, E., Dadson, S., Ellis, R., Peng, J., Robinson, E., Morrison, R., Osborne, S., and Quaife, T.: Improving soil moisture prediction of a high-resolution land surface model by parameterising pedotransfer functions through assimilation of SMAP satellite data, Hydrology and Earth System Sciences, 25, 1617–1641, 2021.

Raoult, N., Ottlé, C., Peylin, P., Bastrikov, V., and Maugis, P.: Evaluating and Optimizing Surface Soil Moisture Drydowns in the ORCHIDEE Land Surface Model at In Situ Locations, Journal of Hydrometeorology, 22, 1025–1043, 2021.

Raschke, K.: Movements of stomata., Encyclopedia of Plant Physiology., 7, 383–441, 1979.

Rautiainen, M., Heiskanen, J., and Korhonen, L.: Seasonal changes in canopy leaf area index and MODIS vegetation products for a boreal forest site in central Finland, Boreal Environment Research, 17, 72–84, 2012.

Rebel, K., De Jeu, R., Ciais, P., Viovy, N., Piao, S., Kiely, G., and Dolman, A.: A global analysis of soil moisture derived from satellite observations and a land surface model, Hydrology and Earth System Sciences, 16, 833–847, 2012.

Reynolds, C., Jackson, T., and Rawls, W.: Estimating soil water-holding capacities by linking the Food and Agriculture Organization soil map of the world with global pedon databases and continuous pedotransfer functions, Water Resources Research, 36, 3653–3662, 2000.

Rivalland, V.: Amelioration et validation du modele de fonctionnement de la végétation ISBA-A-gs: stress hydrique et flux de CO2, Ph.D. thesis, Université Paul Sabatier-Toulouse III, 2003.

Rosnay, P. d. and Polcher, J.: Modelling root water uptake in a complex land surface scheme coupled to a GCM, Hydrology and Earth System Sciences, 2, 239–255, 1998.

Seyednasrollah, B., Bowling, D. R., Cheng, R., Logan, B. A., Magney, T. S., Frankenberg, C., Yang, J. C., Young, A. M., Hufkens, K., Arain, M. A., et al.: Seasonal variation in the canopy color of temperate evergreen conifer forests, New Phytologist, 229, 2586–2600, 2021.

Shukla, P., Skea, J., Calvo Buendia, E., Masson-Delmotte, V., Pörtner, H., Roberts, D., Zhai, P., Slade, R., Connors, S., Van Diemen, R., Ferrat, M., Haughey, E., Luz, S., Neogi, S., Pathak, M., Petzold, J., Portugal Pereira, J., Vyas, P., Huntley, E., Kissick, K., Belkacemi, M., and Malley, J.: IPCC, 2019: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems, 2019.

Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., Doney, S. C., Graven, H., Heinze, C., Huntingford, C., et al.: Recent trends and drivers of regional sources and sinks of carbon dioxide, Biogeosciences, 12, 653–679, 2015.

Song, X., Wang, D.-Y., Li, F., and Zeng, X.-D.: Evaluating the performance of CMIP6 Earth system models in simulating global vegetation structure and distribution, Advances in Climate Change Research, 12, 584–595, 2021.

Szczypta, C., Calvet, J.-C., Maignan, F., Dorigo, W., Baret, F., and Ciais, P.: Suitability of modelled and remotely sensed essential climate variables for monitoring Euro-Mediterranean droughts, Geoscientific Model Development, 7, 931–946, 2014.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American meteorological Society, 93, 485–498, 2012.

Vereecken, H., Weihermüller, L., Assouline, S., Šimůnek, J., Verhoef, A., Herbst, M., Archer, N., Mohanty, B., Montzka, C., Vanderborght, J., et al.: Infiltration from the pedon to global grid scales: An overview and outlook for land surface modeling, Vadose Zone Journal, 18, 1–53, 2019.

Wang, F., Cheruy, F., and Dufresne, J.-L.: The improvement of soil thermodynamics and its effects on land surface meteorology in the IPSL climate model, Geoscientific Model Development, 9, 363–381, 2016.

Wang, T., Ottle, C., Boone, A., Ciais, P., Brun, E., Morin, S., Krinner, G., Piao, S., and Peng, S.: Evaluation of an improved intermediate complexity snow scheme in the ORCHIDEE land surface model, Journal of Geophysical Research: Atmospheres, 118, 6064–6079, 2013.

Weber, U., Jung, M., Reichstein, M., Beer, C., Braakhekke, M., Lehsten, V., Ghent, D., Kaduk, J., Viovy, N., Ciais, P., et al.: The interannual variability of Africa's ecosystem productivity: a multi-model analysis, Biogeosciences, 6, 285–295, 2009.

Williams, I. N., Lee, J. M., Tadić, J., Zhang, Y., and Chu, H.: Modeling Spatial Heterogeneity in Surface Turbulent Heat Flux in the US Southern Great Plains, Journal of Geophysical Research: Atmospheres, 125, e2019JD032 255, 2020.

Wohlfahrt, G., Irschick, C., Thalinger, B., Hörtnagl, L., Obojes, N., and Hammerle, A.: Insights from independent evapotranspiration estimates for closing the energy balance: a grassland case study, Vadose Zone Journal, 9, 1025–1033, 2010.

Wösten, J., Lilly, A., Nemes, A., and Le Bas, C.: Development and use of a database of hydraulic properties of European soils, Geoderma, 90, 169–185, 1999.

Yin, X. and Struik, P.: C3 and C4 photosynthesis models: An overview from the perspective of crop modelling, NJAS-Wageningen Journal of Life Sciences, 57, 27–38, 2009.

Zhou, S., Yu, B., Zhang, Y., Huang, Y., and Wang, G.: Partitioning evapotranspiration based on the concept of underlying water use efficiency, Water Resources Research, 52, 1160–1175, 2016.

List of Figures

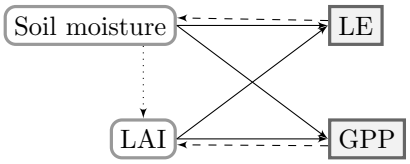

Figure 1. First order relations (plain lines) and feedbacks (dashed lines) of the state variables and surface fluxes in prognostic LSM. The feedback mechanisms are not present in diagnostic models, and the soil moisture-LAI relation (dotted line) occurs only in prognostic LSM with dedicated phenology schemes.

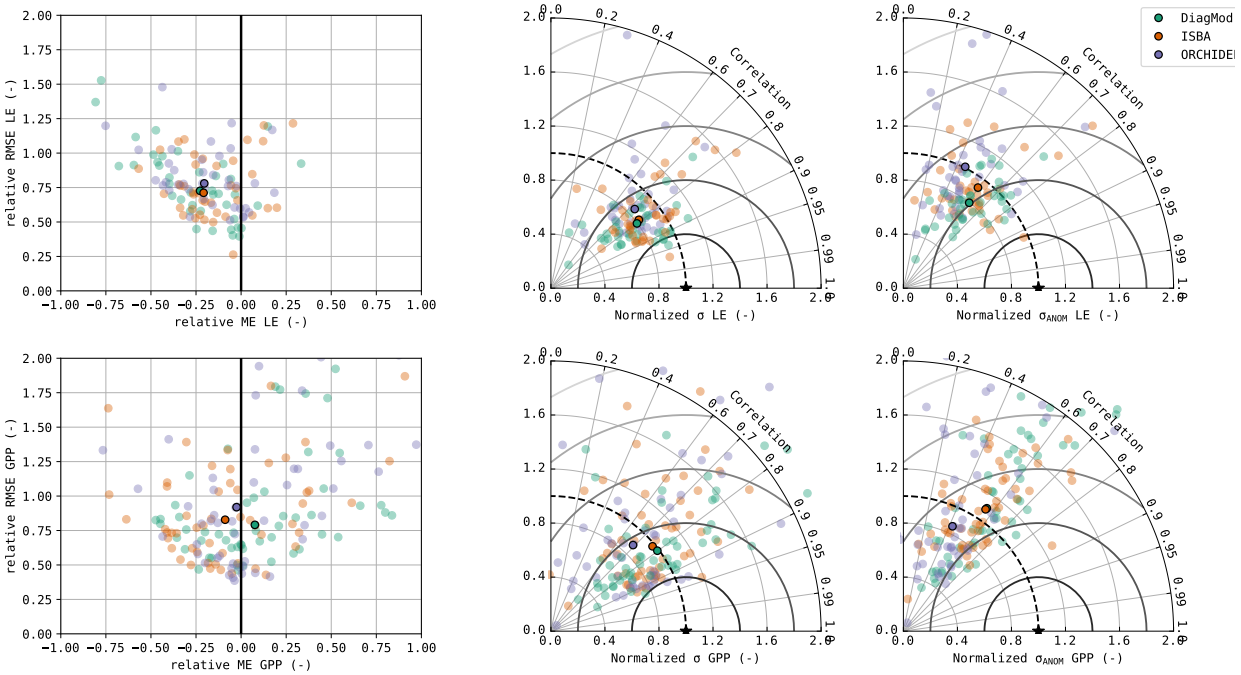

Figure 2. Accuracy plot (left), Taylor diagram (middle) and Taylor diagram of the seasonal anomalies (right) of the simulated daily mean LE (top) and GPP (bottom). The median performance is shown with the opaque markers.

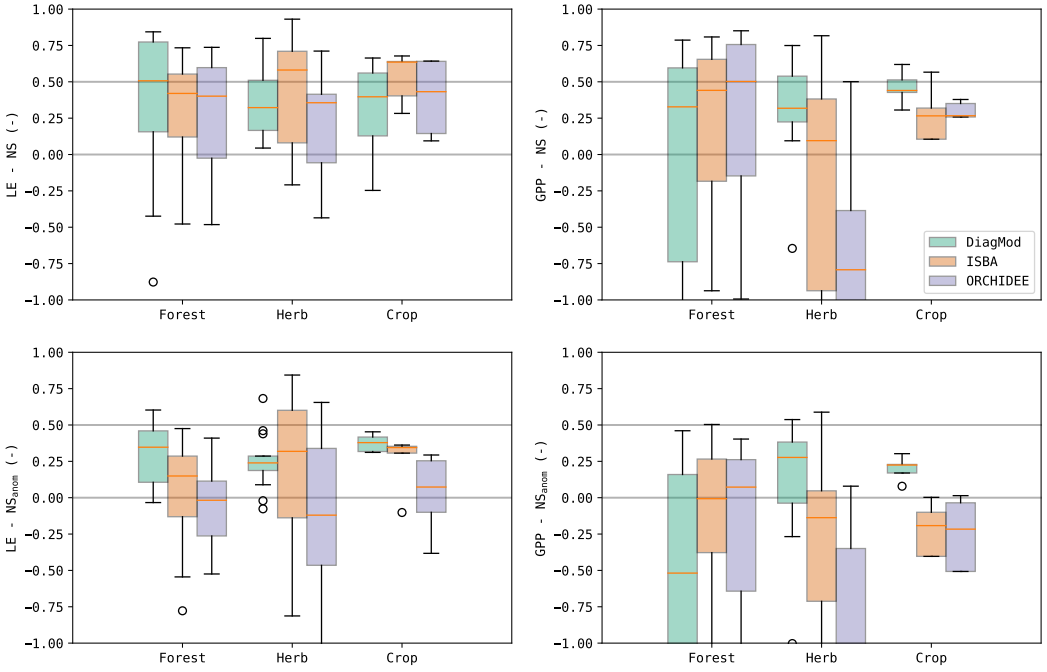

Figure 3. NS and NS_{ANOM} of the simulated daily LE and GPP, grouped per land cover type

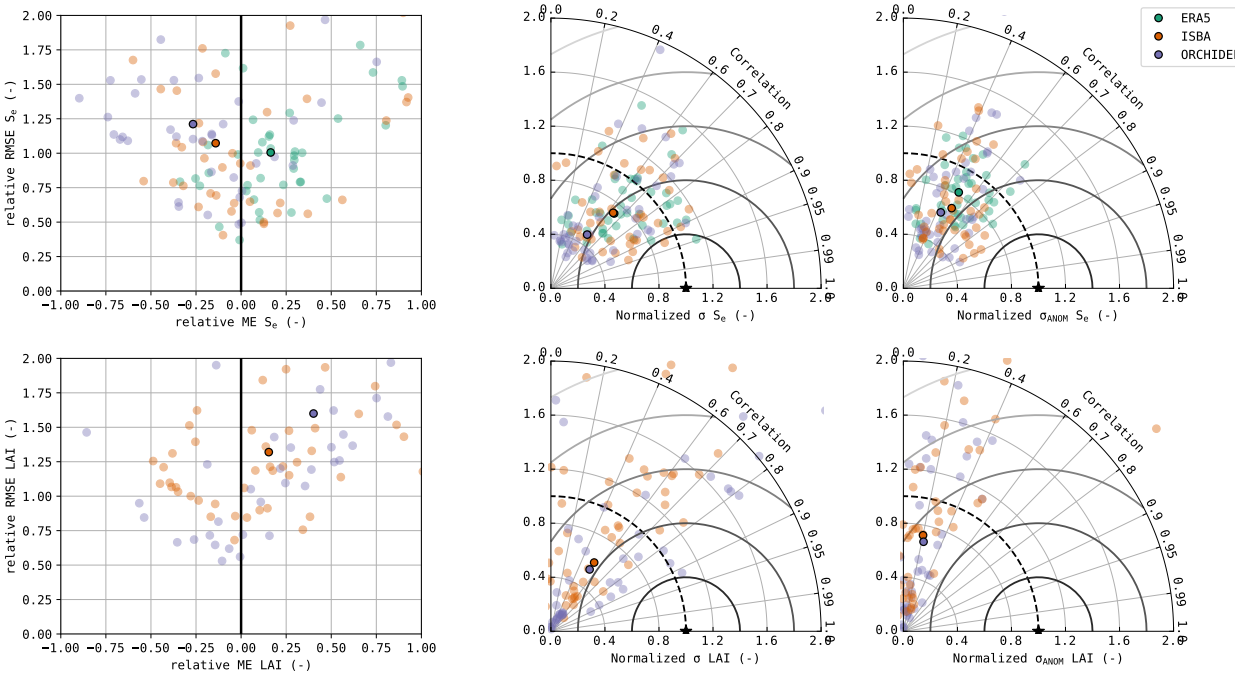

Figure 4. Accuracy plot (left), Taylor diagram (middle) and Taylor diagram of the seasonal anomalies (right) of S_e (top) and LAI (bottom). The median performance is shown with the opaque markers.

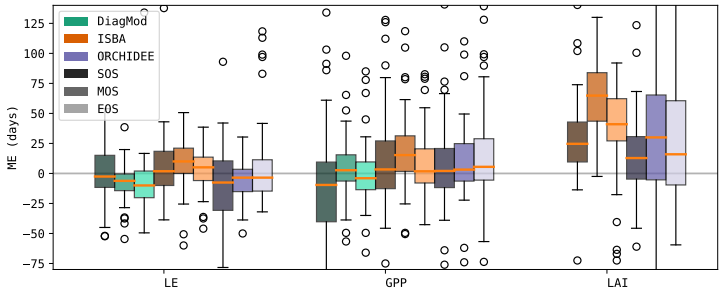

Figure 5. Mean error in the timing of the simulated seasonal cycle (Start, Max and End of season) for LE, GPP and LAI

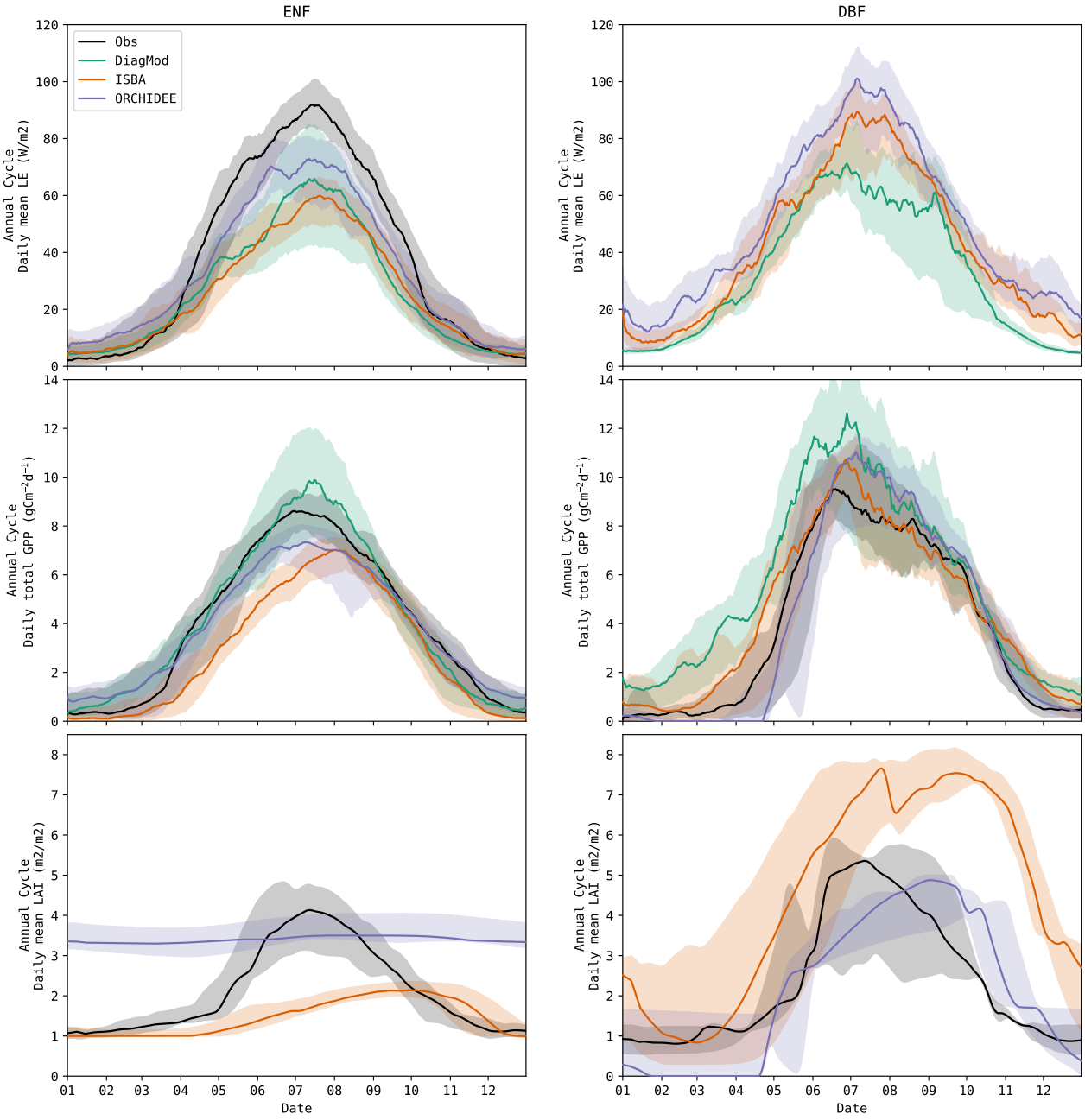

Figure 6. Mean annual cycle for LE, GPP and LAI in all evergreen needleleaf forest (left) and deciduous broadleaf forest (right) sites, observed and simulated. Note: corrected LE observations were missing in all DBF sites.

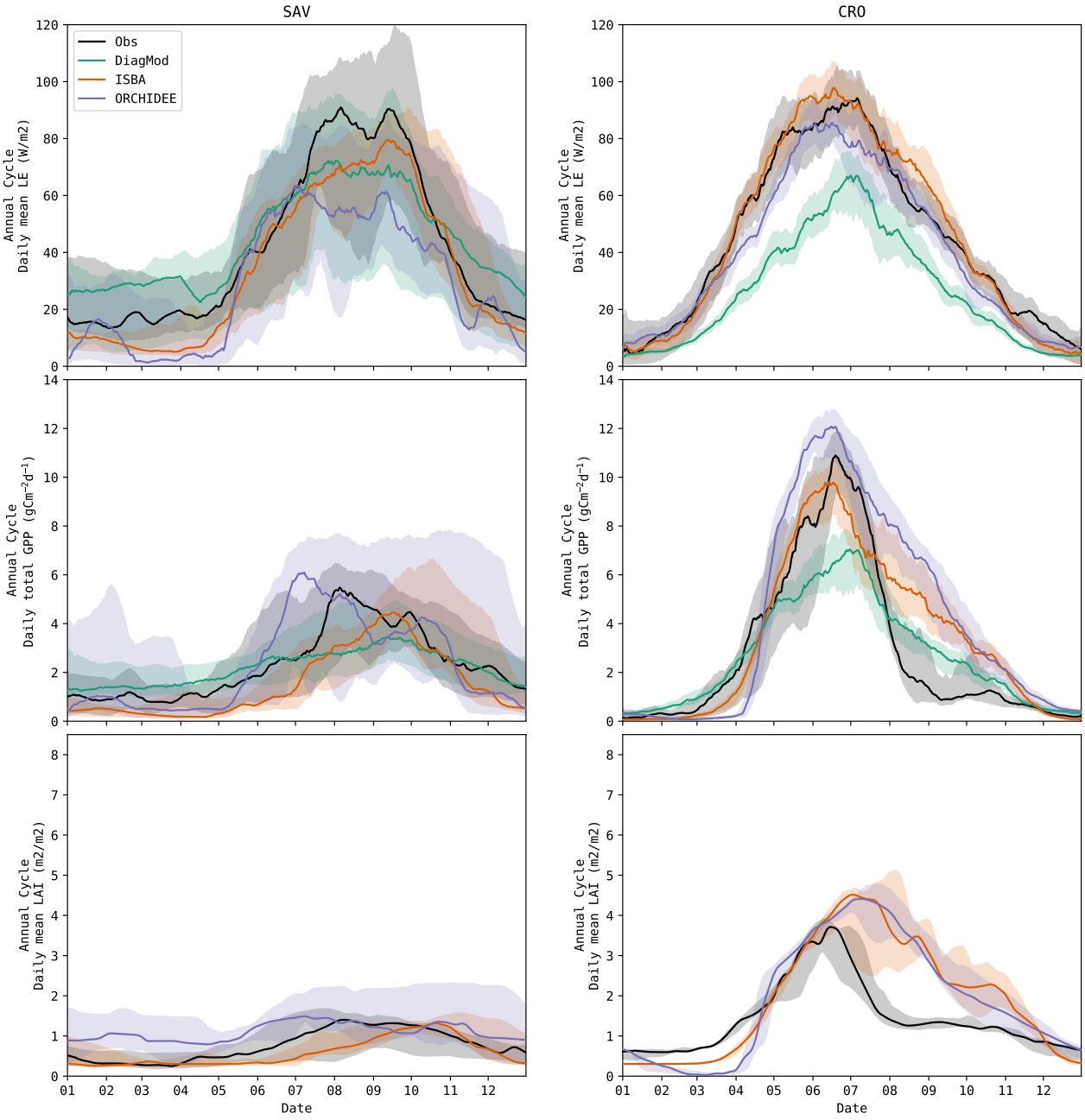

Figure 7. Mean annual cycle for LE, GPP and LAI in all savanna (left) and crop (right) sites, observed and simulated

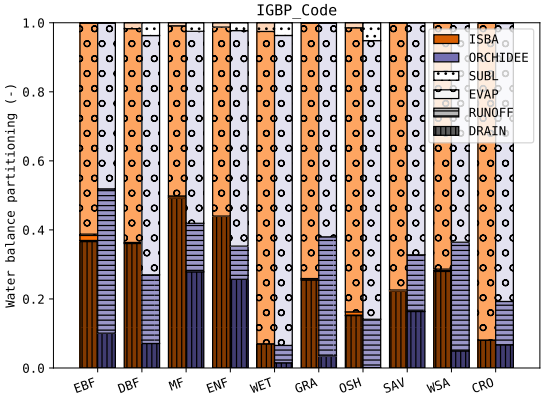

Figure 8. Average water balance partitioning (deep drainage, runoff, evapotranspiration and sublimination) per PFT class in ISBA and ORCHIDEE

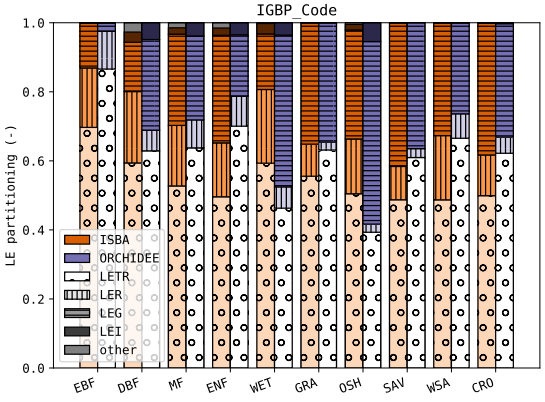

Figure 9. Average LE partitioning per PFT class in ISBA and ORCHIDEE. LETR: Transpiration, LER: intercept evaporation, LEG: soil evaporation, LEI: ice/snow evaporation and other, including evaporation from flooded surfaces.

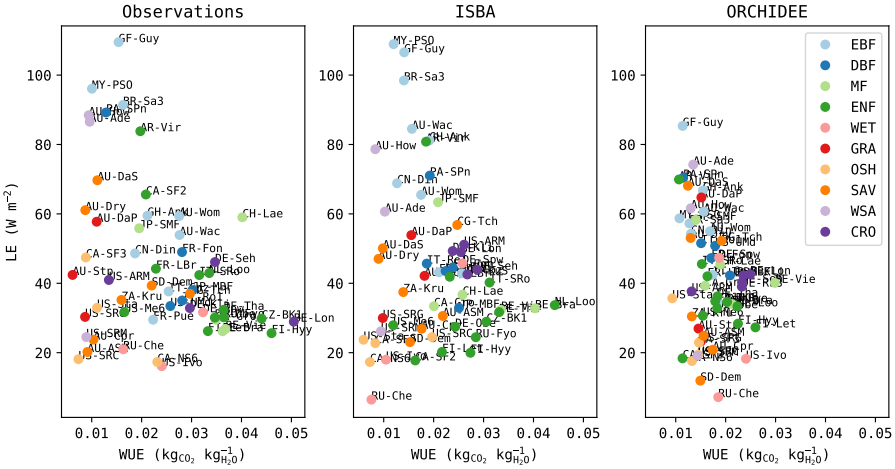

Figure 10. Median water use efficiency and LE in observations and simulations. Sites classified per PFT.

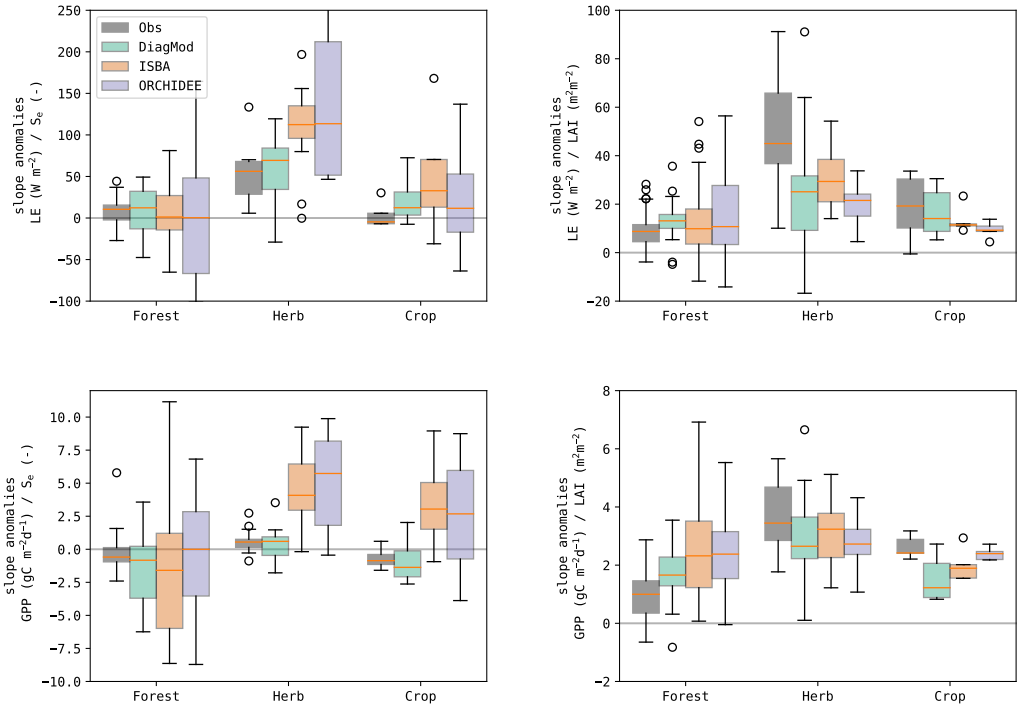

Figure 11. Boxplots of the slope of the linear regression between the anomalies in the state variables (S_e and LAI) and the fluxes (LE and GPP) in the test sites, grouped per dominant land cover

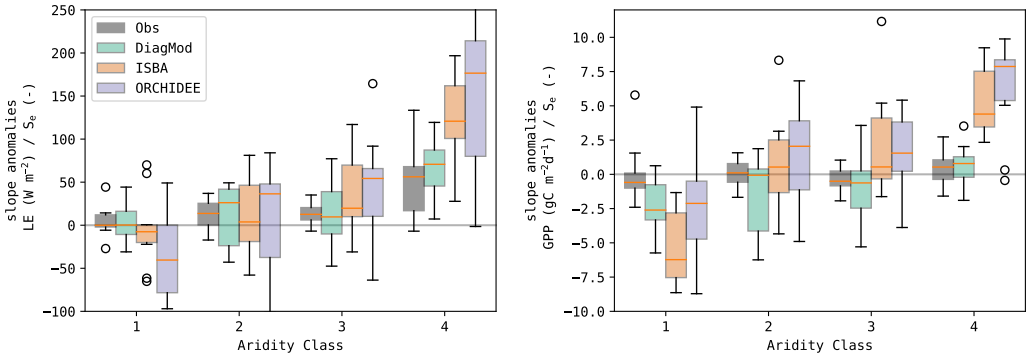

Figure 12. Boxplots of the slope of the linear regression between the anomalies in the state variables (S_e and LAI) and the fluxes (LE and GPP) in the test sites, grouped per aridity class (1: least - 4: most frequent drought stress)

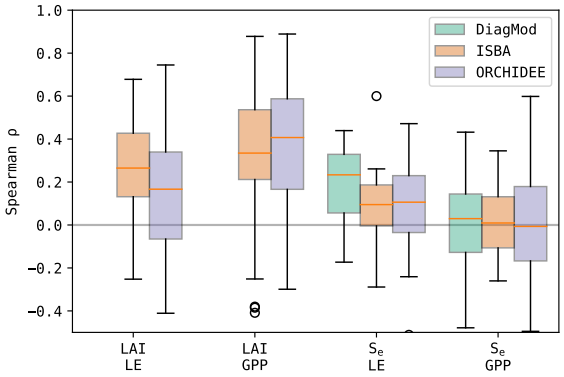

Figure 13. Boxplots of the Spearman correlation between the errors in the state variables (S_e and LAI) and the fluxes (LE and GPP) in all test sites

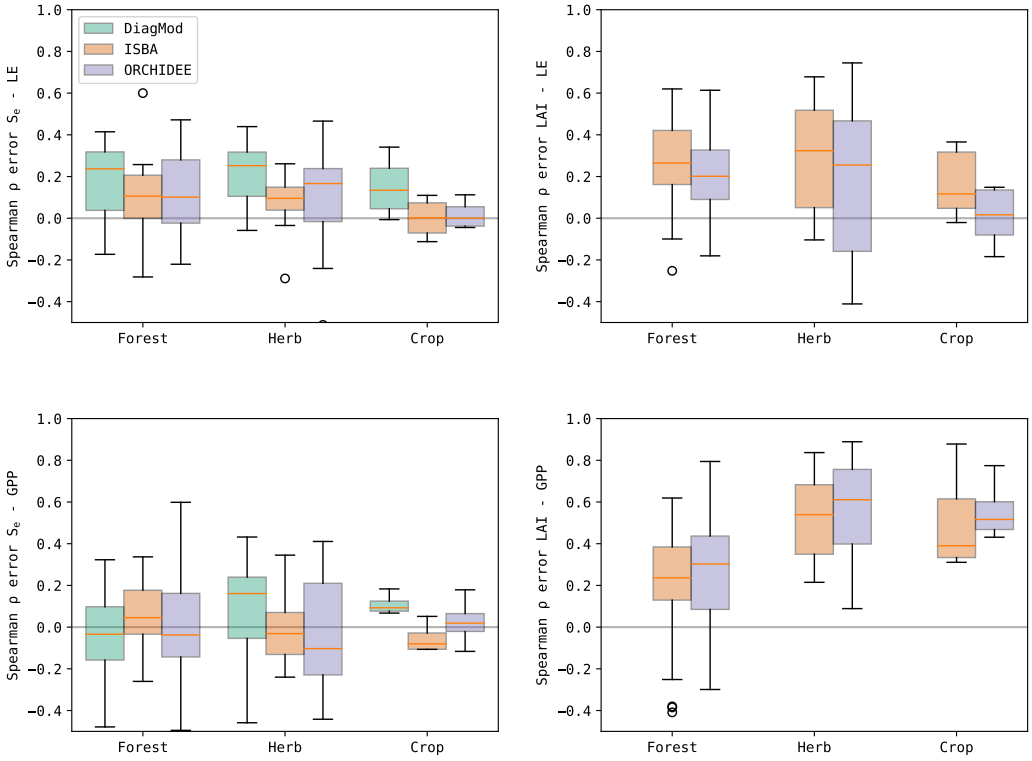

Figure 14. Boxplots of the Spearman correlation between the errors in the state variables (S_e and LAI) and the fluxes (LE and GPP) in the test sites, grouped per dominant land cover

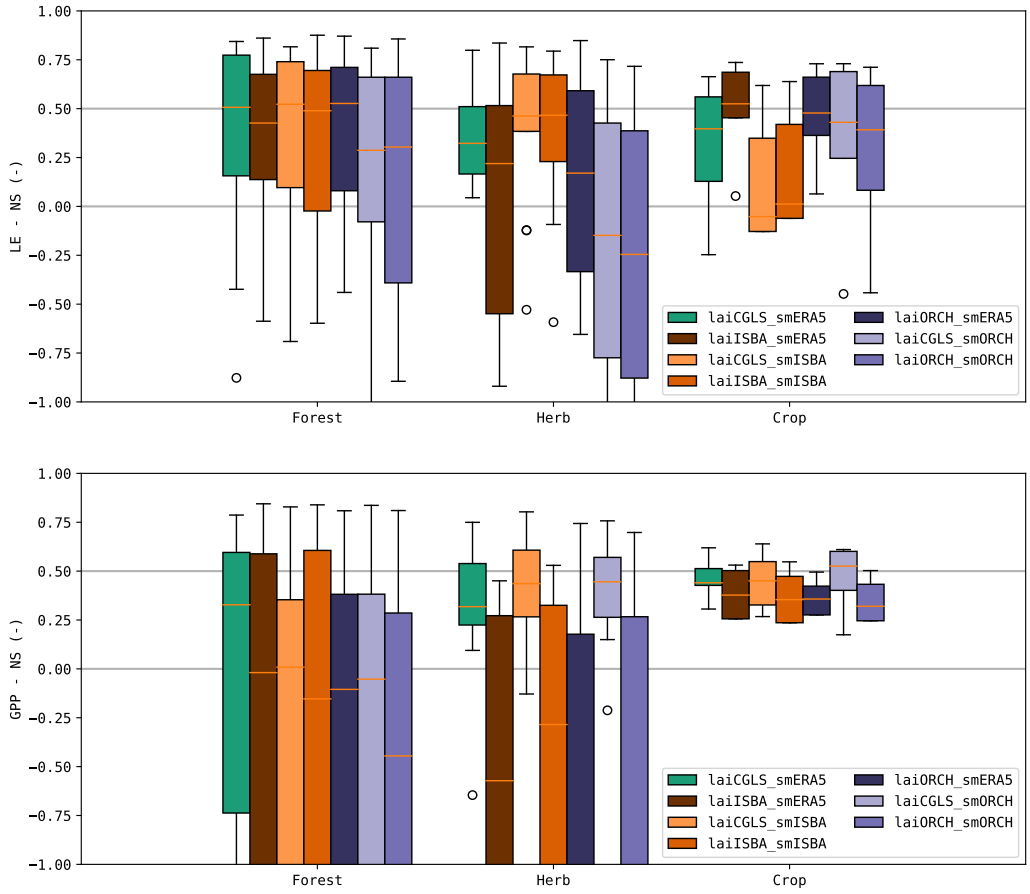

Figure 15. Nash-Sutcliffe model efficiency index of the DiagMod runs for the functional evaluation.

List of Tables

Forcing	**DiagMod**	**ISBA**	**ORCHIDEE**
Air Temperature	ERA5	ERA5	ERA5
Air Humidity	ERA5	ERA5	ERA5
Wind	Tower	ERA5	ERA5
Wind Direction	-	ERA5	ERA5
Atmospheric pressure	ERA5	ERA5	ERA5
Precipitation Rain	-	Tower	Tower
Precipitation Snow	-	Tower	Tower
Short Wave radiation	Tower	Tower	Tower
Long Wave radiation	Tower	Tower	Tower
CO_2 concentration	-	TRENDY	TRENDY
Soil Moisture	ERA5	prognostic	prognostic
LAI	CGLS	prognostic	prognostic
FAPAR	CGLS	-	-
Photosynthesis model	Monteith (1972)	Goudriaan et al. (1985); Jacobs et al. (1996)	Farquhar et al. (1980); Collatz et al. (1992)
Phenology	-	Photosynthesis-driven	Dedicated modules
Soil Layers	4	14	12
Soil Type	ECMWF	HWSD	FAO/USDA
Pedotransfer function	Wösten et al. (1999)	Clapp and Hornberger (1978)	Carsel and Parrish (1988)
Waterlimiting threshold	θ(h=-100cm)	θ(h=-100cm)	0.8 * θ(h=-330cm)

Table 1. Source of forcing variables. Tower: Fluxtower observations from FLUXNET2015 dataset (Pastorello et al., 2020) and the ICOS '2018 drought initiative' dataset (Drought 2018 Team and ICOS Ecosystem Thematic Centre, 2019), TRENDY (Sitch et al., 2015, https://sites.exeter.ac.uk/trendy), CGLS: Copernicus Global Land Service (Camacho et al., 2013), ECMWF: soil texture used in the ECMWF Integrated Forecast System (https://apps.ecmwf.int/codes/grib/param-db?id=43), HWSD: harmonized world soil database (Nachtergaele et al., 2010), FAO/USDA: USDA texture map based on FAO digital Soil Map of the World (Reynolds et al., 2000). θ(h=-100cm) refers to the water content at matric head=-100cm, derived from the water retention curve.

Code	Name	Country	Database	Start	End	PFT	HCB	Köppen	LE corr	Aridity
AR-Vir	Virasoro	Argentina	FLUXNET2015	2009	2013	ENF	Trans_W	Cfa	-	2
AU-ASM	Alice Springs	Australia	FLUXNET2015	2009	2013	SAV	SubTr_W	BSh	0.09	4
AU-Ade	Adelaide River	Australia	FLUXNET2015	2006	2009	WSA	Trans_W	As	0.29	3
AU-Cpr	Calperum	Australia	FLUXNET2015	2009	2014	SAV	Trans_W	BSk	0.02	4
AU-DaP	Daly River Savanna	Australia	FLUXNET2015	2006	2013	GRA	Trans_E	As	0.22	3
AU-DaS	Daly River Cleared	Australia	FLUXNET2015	2007	2014	SAV	Trans_E	Aw	0.03	3
AU-Dry	Dry River	Australia	FLUXNET2015	2007	2014	SAV	Trans_E	As	0.27	4
AU-How	Howard Springs	Australia	FLUXNET2015	2000	2014	WSA	Trans_E	As	0.19	3
AU-Stp	Sturt Plains	Australia	FLUXNET2015	2007	2014	GRA	Trans_E	As	0.08	4
AU-Wac	Wallaby Creek	Australia	FLUXNET2015	2004	2008	EBF	Trans_E	Cfb	0.12	1
AU-Wom	Wombat	Australia	FLUXNET2015	2009	2012	EBF	Trans_W	Cfb	0.27	2
BE-Bra	Brasschaat	Belgium	ICOS Drought	1995	2018	MF	MidL_T	Cfb	0.17	1
BE-Lon	Lonzee	Belgium	ICOS Drought	2003	2018	CRO	MidL_T	Cfb	0.48	4
BE-Vie	Vielsalm	Belgium	ICOS Drought	1995	2018	MF	MidL_T	Cfb	-0.03	1
BR-Sa3	Santarem	Brazil	FLUXNET2015	2000	2005	EBF	Tropic	Aw	0.17	2
CA-Gro	Ontario	Canada	FLUXNET2015	2003	2015	MF	Bor_T	Dfb	0.42	1
CA-NS6	UCI-1989 burn site	Canada	FLUXNET2015	2001	2006	OSH	Bor_T	BSk	-	3
CA-SF2	Saskatchewan	Canada	FLUXNET2015	2001	2006	ENF	Bor_T	Dwc	0.41	3
CA-SF3	Saskatchewan	Canada	FLUXNET2015	2001	2007	OSH	Bor_T	Dwc	0.35	2
CG-Tch	Tchizalamou	Congo	FLUXNET2015	2005	2009	SAV	Tropic	As	-	2
CH-Lae	Laegeren	Switzerland	ICOS Drought	2003	2018	MF	MidL_T	Dfb	-	2
CN-Din	Dinghushan	China	FLUXNET2015	2002	2005	EBF	SubTr_E	Cwa	-	3
CZ-BK1	Bily Kriz forest	Czech Rep.	ICOS Drought	2003	2018	ENF	MidL_T	Dfb	-	1
DE-Kli	Klingenberg	Germany	ICOS Drought	2003	2018	CRO	MidL_T	Dfb	0.46	3
DE-Obe	Oberbrenburg	Germany	ICOS Drought	2007	2018	ENF	MidL_T	Dfb	0.21	1
DE-RuS	Selhausen Juelich	Germany	ICOS Drought	2010	2018	CRO	MidL_T	Cfb	0.47	4
DE-Seh	Selhausen	Germany	FLUXNET2015	2006	2010	CRO	MidL_T	Cfb	0.14	4
DE-Spw	Spreewald	Germany	FLUXNET2015	2009	2014	WET	MidL_T	Cfb	-	3
DE-Tha	Tharandt	Germany	ICOS Drought	1995	2018	ENF	MidL_T	Dfb	0.26	1
FI-Hyy	Hyytiala	Finland	ICOS Drought	1995	2018	ENF	Bor_WT	Dfb	0.03	1
FI-Let	Lettosuo	Finland	ICOS Drought	2008	2018	ENF	Bor_WT	Dfb	-0.26	1
FR-Fon	Fontainebleau	France	FLUXNET2015	2004	2014	DBF	MidL_T	Cfb	-	3
FR-LBr	Le Bray	France	FLUXNET2015	1995	2008	ENF	Trans_E	Cfb	0.21	2
FR-Pue	Puechabon	France	FLUXNET2015	1999	2014	EBF	Trans_E	Csa	0.42	2
GF-Guy	Guyaflux	Fr. Guiana	FLUXNET2015	2004	2015	EBF	Tropic	As	-	2
GH-Ank	Ankasa	Ghana	FLUXNET2015	1989	1989	EBF	Tropic	Aw	0.56	1
IT-Cpz	Castelporziano	Italy	FLUXNET2015	1996	2009	EBF	Trans_E	Csa	0.07	2
IT-Ro1	Roccarespampani	Italy	FLUXNET2015	1999	2008	DBF	Trans_E	Csa	-	3
IT-SRo	San Rossore	Italy	FLUXNET2015	1998	2012	ENF	Trans_E	Csa	0.37	2
JP-MBF	Moshiri	Japan	FLUXNET2015	2002	2005	DBF	Bor_T	Dfb	-	1
JP-SMF	Seto	Japan	FLUXNET2015	2001	2006	MF	Trans_E	Cfa	-	1
MY-PSO	Pasoh	Malaysia	FLUXNET2015	2002	2009	EBF	Tropic	Af	-0.01	1
NL-Loo	Loobos	Netherlands	ICOS Drought	1995	2018	ENF	MidL_T	Cfb	0.05	1
PA-SPn	Sardinilla	Panama	FLUXNET2015	2007	2010	DBF	Trans_E	Aw	-	2
RU-Che	Cherski	Russia	FLUXNET2015	2001	2005	WET	Bor_E	Dwc	-	4
RU-Fyo	Fyodorovskoye	Russia	ICOS Drought	1997	2018	ENF	Bor_WT	Dfb	-0.18	3
SD-Dem	Demokeya	Sudan	FLUXNET2015	2004	2009	SAV	SubTr_W	Aw	0.62	4
US-ARM	Lamont	United States	FLUXNET2015	2003	2013	CRO	MidL_W	Cfa	0.19	4
US-Ivo	Ivotuk	United States	FLUXNET2015	2004	2008	WET	Bor_E	Dwc	-0.15	2
US-Me6	Metolius	United States	FLUXNET2015	2010	2015	ENF	Trans_E	Dsb	0.46	3
US-SRC	Santa Rita Creosote	United States	FLUXNET2015	2008	2015	OSH	Trans_E	BSh	0.65	4
US-SRG	Santa Rita Grassland	United States	FLUXNET2015	2008	2015	GRA	Trans_E	BSh	0.30	4
US-SRM	Santa Rita Mesquite	United States	FLUXNET2015	2004	2015	WSA	Trans_E	BSh	0.26	4
US-Sta	Saratoga	United States	FLUXNET2015	2005	2010	OSH	MidL_W	Dfb	-	2
US-UMd	UMBS Disturbance	United States	FLUXNET2015	2007	2015	DBF	Bor_T	Dfb	-	3
ZA-Kru	Skukuza	South Africa	FLUXNET2015	1999	2013	SAV	Trans_W	Csa	0.21	4

Table 2. Selection of 56 FLUXNET/ICOS sites used in this study. Classification by PFT, HCB (Boreal/Mid-Latitude/Transitional/Subtropical/Tropical+Energy/Water/Temperature-driven) and Köppen. LE corr: relative change of the mean LE flux after correction for energy balance closure (no value: correction not available). Aridity: aridity class, derived from ISBA simulations

		DiagMod				ISBA				ORCHIDEE			
			Forest	Herb	Crop		Forest	Herb	Crop		Forest	Herb	Crop
NS	LE	**0.47**	0.51	0.32	0.40	**0.49**	0.42	0.58	0.64	**0.39**	0.40	0.36	0.43
	GPP	**0.37**	0.33	0.32	0.44	**0.31**	0.44	0.10	0.27	**0.15**	0.46	-0.79	0.27
	S_e	**-0.01**	-0.11	0.37	-0.70	**-0.09**	-0.74	0.52	-3.07	**-0.37**	-0.71	0.14	-2.89
	LAI					**-0.74**	-0.47	-1.05	-0.77	**-1.56**	-1.12	-3.91	-0.83
NS_{ANOM}	LE	**0.30**	0.35	0.24	0.38	**0.21**	0.15	0.32	0.34	**-0.03**	-0.02	-0.12	0.07
	GPP	**0.04**	-0.52	0.28	0.22	**-0.07**	-0.01	-0.14	-0.19	**-0.19**	0.11	-1.83	-0.22
	S_e	**0.09**	0.01	0.36	0.03	**0.14**	0.06	0.47	-0.14	**0.06**	0.02	0.43	-0.12
	LAI					**-0.54**	-0.32	-1.34	-3.74	**-0.50**	-0.03	-4.06	-2.68

Table 3. Nash-Sutcliffe model efficiency coefficient of LE, GPP, S_e and LAI, and their seasonal anomalies. Median scores given for all sites, and grouped per dominant land cover type. The scores for the DiagMod S_e are computed using the ERA5 S_e.

			NS – LE				NS – GPP			
	LAI	SM	**All**	Forest	Herb	Crop	**All**	Forest	Herb	Crop
DiagMod	CGLS	ERA5	**0.47**	0.51	0.32	0.40	**0.37**	0.33	0.32	0.44
laiISBA_smERA5	ISBA	ERA5	**0.42**	0.43	0.22	0.52*	**0.11***	-0.02*	-0.57*	0.38
laiCGLS_smISBA	CGLS	ISBA	**0.45**	0.52	0.46	-0.05*	**0.27***	0.01*	0.44	0.45
laiISBA_smISBA	ISBA	ISBA	**0.45**	0.49	0.47	0.01*	**-0.08***	-0.15*	-0.28*	0.35
laiORCH_smERA5	ORCHIDEE	ERA5	**0.42**	0.53	0.17	0.48	**-0.24***	-0.10*	-1.92*	0.36
laiCGLS_smORCH	CGLS	ORCHIDEE	**0.24***	0.29*	-0.15*	0.43	**0.29***	-0.05*	0.45	0.53
laiORCH_smORCH	ORCHIDEE	ORCHIDEE	**0.27***	0.30*	-0.25*	0.39	**-0.50***	-0.45*	-1.28*	0.32

Table 4. Median Nash-Sutcliffe model efficiency index of the DiagMod runs (functional evaluation of the prognostic LAI and soil moisture). Results presented for all sites, and classified per dominant land cover. Significant differences (Wilcoxon $p < 0.05$) with reference DiagMod runs are marked.