



## How Well Can Land-Surface Models Represent the Diurnal Cycle of Turbulent Heat Fluxes?

MAIK RENNER,<sup>a,g</sup> AXEL KLEIDON,<sup>a</sup> MARTYN CLARK,<sup>b</sup> BART NIJSSEN,<sup>c</sup> MARVIN HEIDKAMP,<sup>d</sup> MARTIN BEST,<sup>e</sup> AND GAB ABRAMOWITZ<sup>f</sup>

<sup>a</sup> Max Planck Institute for Biogeochemistry, Jena, Germany

<sup>b</sup> Centre for Hydrology, University of Saskatchewan, Canmore, Alberta, Canada

<sup>c</sup> Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington

<sup>d</sup> Max Planck Institute for Meteorology, Hamburg, Germany

<sup>e</sup> Met Office, Exeter, United Kingdom

<sup>f</sup> ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, New South Wales, Australia

(Manuscript received 5 February 2020, in final form 8 September 2020)

**ABSTRACT:** The diurnal cycle of solar radiation represents the strongest energetic forcing and dominates the exchange of heat and mass of the land surface with the atmosphere. This diurnal heat redistribution represents a core of land–atmosphere coupling that should be accurately represented in land surface models (LSMs), which are critical parts of weather and climate models. We employ a diagnostic model evaluation approach using a signature-based metric that describes the diurnal variation of heat fluxes. The metric is obtained by decomposing the diurnal variation of surface heat fluxes into their direct response and the phase lag to incoming solar radiation. We employ the output of 13 different LSMs driven with meteorological forcing of 20 FLUXNET sites (PLUMBER dataset). All LSMs show a poor representation of the evaporative fraction and thus the diurnal magnitude of the sensible and latent heat flux under cloud-free conditions. In addition, we find that the diurnal phase of both heat fluxes is poorly represented. The best performing model only reproduces 33% of the evaluated evaporative conditions across the sites. The poor performance of the diurnal cycle of turbulent heat exchange appears to be linked to how models solve for the surface energy balance and redistribute heat into the subsurface. We conclude that a systematic evaluation of diurnal signatures is likely to help to improve the simulated diurnal cycle, better represent land–atmosphere interactions, and therefore improve simulations of the near-surface climate.

**KEYWORDS:** Atmosphere–land interaction; Heat budgets/fluxes; Land surface model; Model comparison; Model evaluation/performance; Diurnal effects

### 1. Introduction

#### a. Background and motivation

Land surface models simulate distinct diurnal cycles of turbulent heat fluxes, but they also show systematic deviations from observations, which were reported in early (Henderson-Sellers et al. 1995; Chen et al. 1997) and more recent model intercomparison studies (Holtzlag et al. 2013; Best et al. 2015). Best et al. (2015) used observational meteorological forcing to drive and evaluate state-of-the-art models at 20 different flux towers. A striking finding of Best et al. (2015) was that simple linear regression models with solar radiation as a predictor variable outcompeted all land surface models when evaluated

with standard statistical metrics. The simple linear response seen in observations can be regarded as a signature of complex land–atmosphere interactions, which are known to simplify the response of turbulent fluxes at certain scales (Jarvis and McNaughton 1986; De Bruin and Holtzlag 1982). Hence, these land surface models, which include many different process parameterizations, may not represent diurnal land–atmosphere interaction well. This is important because land–atmosphere feedbacks propagate to larger scales and may ultimately affect model sensitivity to global change (Miralles et al. 2019). Here, we employ a diagnostic model evaluation based on diurnal signatures of surface heat fluxes to quantify the performance of LSMs specifically for diurnal heat redistribution processes and point toward parameterizations that need to be improved.

#### b. Signatures of diurnal land–atmosphere exchange

The response of turbulent heat fluxes to solar radiation at the subdaily time scale represents a signature of heat redistribution

Denotes content that is immediately available upon publication as open access.

<sup>g</sup> Current affiliation: Hochwassermeldezentrale, Landesamt für Umwelt Brandenburg, Frankfurt, Germany.

Corresponding author: Maik Renner, maik.renner@lfu.brandenburg.de



This article is licensed under a [Creative Commons Attribution 4.0 license](http://creativecommons.org/licenses/by/4.0/) (<http://creativecommons.org/licenses/by/4.0/>).

DOI: 10.1175/JHM-D-20-0034.1

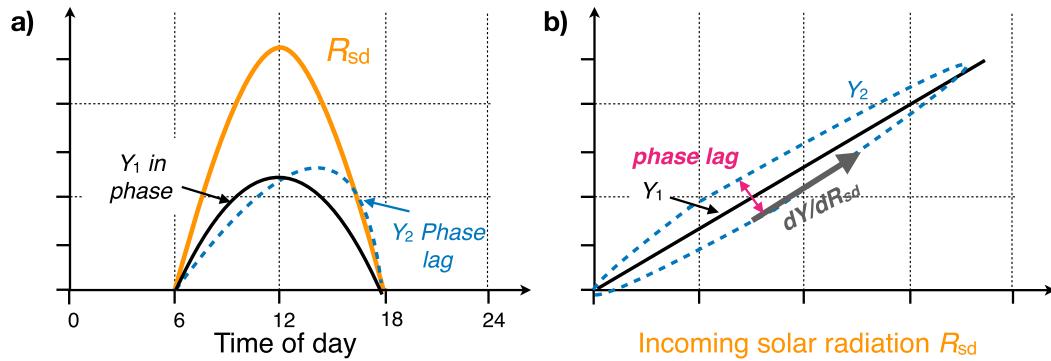


FIG. 1. Illustration of the signature-based metric of the phase lag. (a) A diurnal cycle of solar radiation  $R_{sd}$  and two variables;  $Y_1$  is in phase with  $R_{sd}$  while  $Y_2$  is lagging  $R_{sd}$ . (b) The variables plotted as a function of  $R_{sd}$ . The variable  $Y_1$  shows a linear relationship, but the lagged variable reveals a counterclockwise hysteresis loop. This variation can be decomposed into a direct response  $dY/dR_{sd}$  and a second-order response of  $dY/(\partial R_{sd}/\partial t)$  that can be expressed as a phase lag in time units.

processes and is at the core of local land–atmosphere interactions. The diurnal imbalance of solar heating of the surface is redistributed into the lower atmosphere and into the subsurface, which is well constrained by the energy balance of the whole system (Kleidon and Renner 2018). Heat redistribution into the lower atmosphere is achieved by turbulent heat transfer driving the atmospheric boundary layer development (Oke 1987), while subsurface heat redistribution includes the conduction and storage of heat in the soil and canopy. These processes are linked through the surface energy balance [in simplified form following Ohmura (2014)]

$$R_{sd} - R_{su} + R_{ld} - H - \lambda E - dU_s/dt = \epsilon_s \sigma T^4, \quad (1)$$

where  $R_{sd}$  is the incoming and  $R_{su}$  the reflected solar radiation at the surface,  $R_{ld}$  is downwelling longwave radiation, and  $H$  and  $\lambda E$  are the turbulent heat fluxes of sensible and latent heat, respectively. The term  $dU_s/dt$  summarizes heat storage fluxes at and into the subsurface such as the soil heat flux  $G$  or the change in biomass heat storage. The emission of longwave radiation from the surface  $\epsilon_s \sigma T^4$  on the right-hand side of Eq. (1) responds passively to the fluxes on the left-hand side. Here,  $T$  refers to the surface temperature,  $\sigma$  is the Stefan–Boltzmann constant, and  $\epsilon_s$  is the emissivity of the surface.

While downward solar radiation can be regarded as an independent driver (at subdaily time scale), all other terms are influenced by surface properties and the exchange with the atmosphere. Hence, the resulting diurnal course of turbulent heat fluxes is an outcome of the response to the distinct diurnal course of solar radiation, influenced by land–atmosphere interaction and biophysical properties of the surface. Although many studies focus on turbulent heat fluxes as key heat fluxes in terms of local land–atmosphere interaction (Best et al. 2015; Haughton et al. 2016; Zhou and Wang 2016; Dirmeyer et al. 2018), there has been relatively little focus on the diurnal signal itself. Exceptions are Wilson et al. (2003) and Nelson et al. (2018), who used the diurnal centroid as a measure of timing of turbulent heat fluxes, and Renner et al. (2019b), who analyzed the

diurnal magnitude and phase lag of turbulent fluxes to solar radiation in observations and evapotranspiration schemes.

The diurnal magnitude of turbulent heat fluxes is related to the partitioning into sensible and latent fluxes, which is controlled by many processes, especially the redistribution of water (Dirmeyer et al. 2018). The phase lag of the turbulent heat fluxes to solar radiation is influenced by diurnal heat storage changes in the land–atmosphere system (Renner et al. 2019b). Although the absorption of solar radiation at the surface comprises a dominant diurnal forcing, the land–atmosphere system appears to be well buffered (Taylor 2012) through the exchange of turbulent heat and radiation within the lower atmosphere and heat conduction with the subsurface (Kleidon and Renner 2018). Heat storage changes at some depth in the soil, within vegetation and the lower atmosphere are relatively slow and typically show large lags to solar radiation (Renner et al. 2019b). The heat fluxes that arise from the instantaneous surface heating by absorption of solar radiation redistribute heat into and interact with the larger heat reservoirs of the soil and the atmosphere. Land surface model parameterizations must resolve these interactions through the (i) surface energy balance solution (Viterbo and Beljaars 1995), and the subgrid-scale parameterizations of the (ii) surface or canopy conductance (Monteith 1965; Medlyn et al. 2011), (iii) the turbulent exchange in the surface layer (Monin and Obukhov 1954; Businger et al. 1971; Maronga and Reuder 2017), and (iv) the planetary boundary layer (Baklanov et al. 2011). These subgrid-scale parameterizations are critical for accurate simulation of the surface climate (Baklanov et al. 2011; Heidkamp et al. 2018), including heat and mass exchange with atmosphere. Therefore, an evaluation of the diurnal phase lag of simulated turbulent heat fluxes can identify important model deficiencies and point toward improvements.

### c. Research questions and design

Here we aim to assess the performance of LSMs at the subdaily time scale when driven with meteorological observations. In particular, we address the open question of why the complex land surface models can be outcompeted by simple

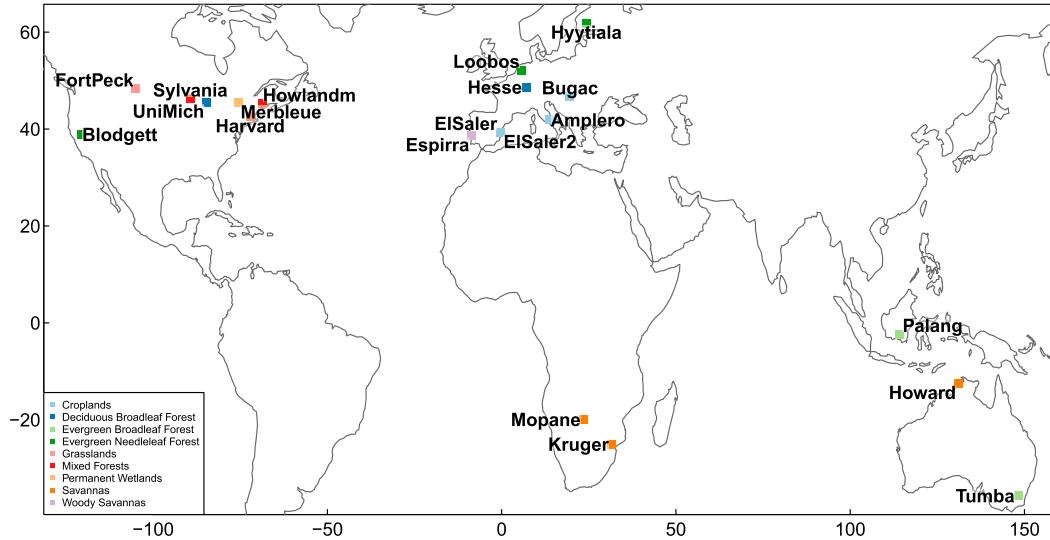


FIG. 2. Map of the FLUXNET sites used in this study. Further information of the sites can be found in Table 1.

linear regression models and point toward key parameterizations of heat redistribution processes that can be improved given the information on observed diurnal signatures.

To assess the performance of current state of the art LSMs, we use the signature based metric of the phase lag of heat fluxes to solar radiation following Renner et al. (2019b). This method has important advantages to standard model observation evaluations. First, as a signature approach it does not evaluate the variability over time, but the response to a common reference variable such as solar radiation. Second, the approach effectively decomposes diurnal variability into two metrics whose statistical significance can be tested and which can be directly interpreted. We will employ the PLUMBER dataset, which was established by the Land Surface Model Benchmarking Evaluation Project published by Best et al. (2015). The PLUMBER dataset includes 13 different land surface models at 20 different FLUXNET sites. This dataset thus represents a wide range of observations and includes key land surface models used in climate and weather models.

In the following we will describe the metric and the sampling design. The results will first focus on the observed variations and controls of the phase lags and then report on the LSM performance. In the discussion we focus on the suitability of the phase lags as a signature and then diagnose which of the LSM components may cause the observed discrepancies.

## 2. Methods

### a. Diurnal signatures

The diurnal course of solar radiation is a dominant forcing of the land–atmosphere exchange fluxes and has already been identified as the dominant forcing of turbulent heat fluxes over land (Kleidon and Renner 2013; Renner et al. 2016; Kleidon and Renner 2018). This motivated the development of two related metrics to evaluate the diurnal variation of surface energy balance components and near-surface atmospheric variables in reference to the variation in solar

radiation (Renner et al. 2019b). This method is briefly explained next.

Timing differences between variables that exhibit large diurnal magnitudes can be difficult to discern in time series plots. However, when plotted against each other systematic deviations can be seen directly. This is illustrated for solar radiation  $R_{sd}$  and two variables  $Y_1$  and  $Y_2$  in Fig. 1 as a time series plot in Fig. 1a and hysteresis loop plot in Fig. 1b. The example shows a variable  $Y_1$  that has a different magnitude than  $R_{sd}$  but is perfectly in phase with  $R_{sd}$ . This results in a linear relationship as illustrated in Fig. 1b. In contrast, variable  $Y_2$  has a time lag with respect to  $R_{sd}$ . This time lag results in a counterclockwise hysteresis loop when plotted against  $R_{sd}$ , see Fig. 1b. Such a time lag is difficult to identify from a time series plot.

The diurnal variation of variable  $Y(t)$  can be decomposed into a direct response to solar radiation and a response to its time derivative  $\partial R_{sd}(t)/\partial t$ :

$$Y(t) = \beta_1 R_{sd}(t) + \beta_2 \frac{\partial}{\partial t} R_{sd}(t) + \beta_0 + \epsilon(t). \quad (2)$$

This form of regression is known as the Bernardi–Camuffo formula (Camuffo and Bernardi 1982). The regression coefficient  $\beta_1$  corresponds to the direct response of variable  $Y$  to  $R_{sd}$ . The second coefficient  $\beta_2$  reflects the width of the hysteresis loop,  $\beta_0$  is the intercept, and  $\epsilon(t)$  represents the residuals. The regression has the benefit that the significance of the coefficients can be statistically determined and tested. Assuming that both variables,  $Y$  and  $R_{sd}$ , can be represented by harmonic functions, one can link the two coefficients  $\beta_1$  and  $\beta_2$  to obtain a measure of the phase lag in time units (Renner et al. 2019b):

$$\phi = \tan^{-1} \left( \frac{\beta_2 \omega}{\beta_1} \right), \quad (3)$$

with the angular frequency  $\omega = (2\pi)/n_d$ , where  $n_d$  is the number of time steps per day. Equation (3) eliminates the  $n_d$  units of the two

TABLE 1. FLUXNET sites used in this study. Column names are the site codes used throughout the manuscript, the respective FLUXNET ID, duration of data used in this study, longitude and latitude, elevation above sea level, the Köppen climate classification, and vegetation cover following IGBP. Annual average air temperature and precipitation are also reported.

Site code	Code	Duration	Lon (°)	Lat (°)	Elev (m)	Köppen	Vegetation	$T_a$ (°C)	$P$ (mm yr <sup>-1</sup> )
Amplero	IT-Amp	2006–06	13.61	41.90	991	Cfa	Croplands	9.6	742
Blodgett	U.S.-Blo	2000–06	-120.63	38.90	1280	Csb	Open needleleaf forest	11.2	1383
Bugac	HU-Bug	2003–04	19.60	46.69	106	Cfb	Croplands	9.8	549
EISaler	ES-ES1	1999–2006	-0.32	39.35	1	Csa	Permanent wetlands	17.4	560
EISaler2	ES-ES2	2005–06	-0.32	39.28	7	Csa	Croplands	17.3	586
Espirra	PT-Esp	2002–05	-8.60	38.64	90	Csa	Woody savannas	16.0	660
FortPeck	U.S.-FPe	2000–06	-105.10	48.31	638	Bsk	Grasslands	5.7	392
Harvard	U.S.-Ha1	1994–2001	-72.17	42.54	353	Dfb	Deciduous broadleaf forest	8.3	1080
Hesse	FR-Hes	2001–06	7.07	48.67	293	Cfb	Deciduous broadleaf forest	10.2	940
Howard	AU-How	2003–05	131.15	-12.50	41	Aw	Savannas	26.4	1896
Howlandm	U.S.-Ho1	1996–2004	-68.74	45.20	72	Dfb	Deciduous broadleaf forest	6.6	817
Hyytiala	FI-Hyy	2001–04	24.30	61.85	185	Dfc	Closed needleleaf forest	4.1	459
Kruger	ZA-Kru	2002–03	31.50	-25.02	357	BSh	Savannas	22.0	271
Loobos	NL-Loo	1997–2006	5.74	52.17	25	Cfb	Closed needleleaf forest	10.2	943
Merbleue	CA-Mer	1999–2005	-75.52	45.41	65	Dfb	Permanent wetlands	5.9	844
Mopane	BW-Ma1	1999–2001	23.56	-19.92	950	BSh	Savannas	22.4	288
Palang	ID-Pag	2002–03	114.04	-2.35	52	Af	Evergreen broadleaf forest	26.6	2078
Sylvania	U.S.-Syv	2002–05	-89.35	46.24	539	Dfb	Deciduous broadleaf forest	4.3	491
Tumba	AU-Tum	2002–05	148.15	-35.66	932	Cfb	Evergreen broadleaf forest	9.1	1285
UniMich	U.S.-UMB	1999–2000	-84.71	45.56	228	Dfb	Deciduous broadleaf forest	7.5	618

variables  $Y$  and  $R_{sd}$  and allows easy transformation into time units. This way one can compare the phase lag of different variables such as temperatures or heat fluxes (Renner et al. 2019b).

### b. Data stratification

Conditional sampling of the data was applied to isolate diurnal heat exchange processes and to meaningfully compare the phase lag of models and observations. First, data were filtered for cloud-free days to avoid the influence of clouds on the diurnal cycle of turbulent fluxes. Second, the daily evaporative fraction was used to select days of similar evaporative conditions to avoid influences of biases from modeling of water redistribution processes.

Cloud-free days were determined by days that have more than 80% of the daily sum of surface clear-sky solar radiation  $R_{sd,cs}$ . Half-hourly values of  $R_{sd,cs}$  were determined using the method by Renner et al. (2019a). This method only requires measurements of observed total incoming shortwave radiation  $R_{sd}$  and shortwave incoming shortwave radiation at the top of atmosphere  $R_{s,in}$  as input. The latter can be calculated with information of location, date and time. The method estimates the fractional transmission of solar radiation under cloud-free conditions from monthly subsets of half-hourly data by quantile regression of observed  $R_{sd}$  against the theoretical maximum that is set by  $R_{s,in}$ . Using a quantile of 80% for the regression was found to yield the best agreement to the standard method of Long and Ackerman (2000), which requires additional data of diffuse radiation at a temporal resolution of minutes. Since such data are not commonly available, even at FLUXNET sites, we used the quantile regression approach of Renner et al. (2019a).

Renner et al. (2019b) found that evaporative conditions can alter the phase lag of turbulent heat fluxes at a grassland site. Therefore, we avoid a comparison of the phase lag on a day-by-day basis. Instead we compare the distribution of daily phase

lag estimates for a given site and a given range of evaporative fraction (EF). Daily EF was estimated by linear regression of half-hourly values of  $\lambda E(t)$  and  $H(t)$  for each day:

$$\lambda E(t) = \text{EF}[H(t) + \lambda E(t)] + \beta_0 + \epsilon(t). \quad (4)$$

The slope of this regression represents the daily evaporative fraction. This method is more robust than a simple ratio of daily average values of turbulent heat fluxes [ $\text{EF} = \overline{\lambda E}/(\overline{H} + \overline{\lambda E})$ ] because it can handle a few outliers and its significance can be tested.

We then use the daily EF estimates and stratify the data into 10% bins of EF for each site. Since models and observations differ in the estimated EF (see section 4) there are different sample sizes between models and observations. We choose a minimum of five samples for further statistical analysis of the phase lag per EF bin and site. The agreement of the models was assessed with a test for the center of the distributions. Since the models should be able to reproduce the phase lag of both the sensible and the latent heat fluxes, we use the two-sample Hotelling  $T^2$  test (Hotelling 1931). This is a two-dimensional Student's  $t$  test, which evaluates differences in the center of two linked distributions. We use the implementation “rrcov::T2.test” in the R programming language (R Core Team 2015) within the package “rrcov” (Todorov and Filzmoser 2009). We choose a minimum requirement of five samples, and we then count the frequency of nonrejections at a conservative significance level  $\alpha = 0.01$  per site and EF bin to assess model agreement.

### 3. Data

The analysis is based on the PLUMBER dataset published by Best et al. (2015). It contains site observations from the

TABLE 2. Overview of the land surface models and reference models of the PLUMBER dataset, taken from Best et al. (2015) and Haughton et al. (2016). The column “surface layer” indicates if a skin layer approach or the top soil layer is used to solve for the surface energy balance.

Model ID	Institution, model name, and details	Surface layer	Version	Reference
CABLE	CSIRO, Community Atmosphere Biosphere Land Exchange	Top soil	2	Wang et al. (2011)
CABLE_SLI	CABLE Soil–Litter–Iso	Top soil	2.0_SLI 1.0	Haverd and Cuntz (2010)
CHTESSEL	Tiled ECMWF Scheme for Surface Exchanges over Land uncoupled version	Skin layer	2	Balsamo et al. (2009) and Boussetta et al. (2013)
COLA-SSiB	Center for Ocean–Land–Atmosphere Studies Simplified SiB	Skin layer	1.0	Guo and Dirmeyer (2013)
ISBA-3L	CNRM, Interactions between Soil, Biosphere, and Atmosphere–Surface Externalisée	Top soil	3L-7.3	Masson et al. (2013)
ISBA_dif	14-layer soil moisture–temperature model (Richard’s equation, heat diffusion, root profile)	Top soil	dif-7.3	Decharme et al. (2011) and Masson et al. (2013)
JULES	Met Office, Joint U.K. Land Environment Simulator	Top soil	3.1	Best et al. (2011)
JULES_altP	Alternative (sub) surface parameterizations	Top soil	3.1_alp	Best et al. (2011)
Mosaic	NASA, used in NLDAS-2	Top soil	1	Koster and Suarez (1992, 1994)
Noah2.7.1	NCAR, Community Noah land surface model	Skin layer	2.7.1	Ek et al. (2003)
Noah3.2	Updated schemes	Skin layer	3.2	<a href="http://www.ral.ucar.edu/">www.ral.ucar.edu/</a>
Noah3.3	Same as Noah3.2 but time varying roughness	Skin layer	3.3	<a href="http://www.ral.ucar.edu/">www.ral.ucar.edu/</a>
ORCHIDEE	IPSL, Organizing Carbon and Hydrology in Dynamic Ecosystems	Top soil	Trunk rev. 1401	Krinner et al. (2005)
Penman–Monteith	FAO grass reference, no water stress	—		Best et al. (2015)
Manabe bucket	Simple bucket (Manabe 1969) no surface resistance	—		Best et al. (2015)
1lin	Linear regression using solar radiation ( $R_{sd}$ )	—		Best et al. (2015)
2lin	Multiple linear regression using $R_{sd}$ , $T_{air}$	—		Best et al. (2015)
3km27	Nonlinear regression using $R_{sd}$ , $T_{air}$ , $r_H$	—		Best et al. (2015)
Observation	FLUXNET La-Thuille free fair-use subset			<a href="https://fluxnet.fluxdata.org/data/la-thuille-dataset/">https://fluxnet.fluxdata.org/data/la-thuille-dataset/</a>

FLUXNET La Thuile dataset (<https://fluxnet.fluxdata.org/data/la-thuille-dataset/>) and LSM simulations using the meteorological forcing of these sites. A brief overview is presented below, but the reader is referred to the PLUMBER publications for further detail (Best et al. 2015; Haughton et al. 2016).

### a. Observations

PLUMBER contains 20 sites distributed across continents with different climate and land cover conditions, see the map in Fig. 2 and details in Table 1. For the analysis of the observations we removed data that did not pass quality control provided by the PLUMBER dataset (see Best et al. 2015). Further data screening for the present analysis revealed unrealistic values for  $R_{sd}$  in some years and at some sites. In particular, daily maximum  $R_{sd}$  showed drifts at sites Amplero, UniMich, Mopane, and Kruger, and these years were excluded from further analysis. Single outliers at some sites were set to missing.

The data have been filtered for a set of criteria. First, we only sampled cloud-free days that range between 93 days per year in Loobos and 246 days per year in Blodgett. Since the number of cloud-free days varies across sites this leads to a slightly imbalanced dataset. Second, we restricted the analysis to days with high potential solar radiation of more than  $300 \text{ W m}^{-2}$ . Further days were excluded that have fewer than 30 observations per day, show nonsignificant slopes in determining EF [Eq. (4)] or nonsignificant slopes of sensible and latent heat fluxes to  $R_{sd}$  [Eq. (2)]. We also exclude days where the Bernardi–Camuffo regression explains less than 50% of the diurnal variation in  $H$  and  $\lambda E$ . These filter criteria help to exclude atypical conditions. After applying the filter criteria, the dataset contained between 83 days at Amplero (with one year of data) and 973 days at Blodgett (spanning 7 years of data), see Table 3.

The energy balance requires that the observed heat fluxes should be balanced by the net radiation budget. However, most eddy

TABLE 3. Statistics of observed data for each site, with the number of years, the number of cloud-free days per year, and the number of days after filtering was applied. Also reported are the 10th, 50th, and 90th percentiles of the daily EF estimates. The median of the daily energy balance closure from a linear regression of the total turbulent heat fluxes against available energy is reported in column EB closure. The phase lags  $\phi$  are reported in minutes relative to incoming solar radiation. We report the median for the phase lag of net radiation and the energy balance closure gap  $Q_{\text{Gap}}$  and the 10th, 50th, and 90th percentiles for the latent and the sensible heat fluxes.

Site code	$n_{\text{yr}}$	$n_{\text{yr,cf}}$	$n_{\text{filter}}$	EF <sub>q10</sub>	EF <sub>q50</sub>	EF <sub>q90</sub>	EB closure	$\phi_{R_n}$	$\phi_{\lambda E,q10}$	$\phi_{\lambda E,q50}$	$\phi_{\lambda E,q90}$	$\phi_{H,q10}$	$\phi_{H,q50}$	$\phi_{H,q90}$	$\phi_{Q_{\text{Gap}},q50}$
Amplero	1	174	83	0.52	0.57	0.69	0.73	4.2	-1.1	6.6	26.5	-37.5	-11.9	0.1	12.6
Blodgett	7	246	973	0.32	0.45	0.61	0.78	0.0	-7.3	7.9	23.2	9.9	30.2	42.2	-97.5
Bugac	2	155	143	0.18	0.44	0.65	0.77	-3.5	-22.0	1.0	16.1	-16.9	-0.7	11.0	-52.1
EISaler	8	219	784	0.18	0.26	0.37	0.93	-1.5	-28.1	1.1	27.8	-5.3	9.9	27.9	-28.9
EISaler2	2	230	227	0.27	0.70	0.82	0.55	-2.2	-8.4	14.7	50.0	-40.3	10.9	41.3	-15.7
Espirra	4	209	420	0.15	0.37	0.50	—	-4.6	-39.0	18.6	59.0	-6.2	41.8	99.6	—
FortPeck	7	179	455	0.10	0.31	0.62	0.82	-8.3	-12.7	10.3	29.9	-17.5	-0.4	9.5	-134.0
Harvard	8	146	471	0.13	0.42	0.65	—	3.6	-14.8	11.1	28.9	-31.1	-8.1	9.1	—
Hesse	6	121	348	0.14	0.47	0.65	0.61	-4.0	1.7	22.0	50.5	-35.7	-5.0	15.9	-26.1
Howard	3	201	393	0.32	0.55	0.75	0.87	-1.8	-11.6	11.8	32.3	-15.3	5.2	29.1	-31.8
Howlandm	9	140	543	0.15	0.38	0.49	0.90	-1.6	0.0	18.4	51.0	-8.3	1.7	16.8	-29.9
Hyytiala	4	101	140	0.23	0.39	0.53	—	-1.5	3.2	16.0	30.4	-4.3	7.0	28.0	—
Kruger	2	178	70	0.09	0.23	0.45	—	-4.3	-20.0	4.3	21.0	0.1	8.1	24.6	-67.9
Loobos	10	93	355	0.26	0.36	0.47	0.76	-1.0	-12.8	7.8	23.5	-4.1	6.1	17.4	-29.0
Merbleue	7	133	381	0.32	0.54	0.66	0.75	-3.1	8.9	24.8	51.8	-20.8	2.0	16.9	-52.3
Mopane	3	217	345	0.05	0.16	0.51	0.71	-4.3	-40.0	-2.4	17.2	-2.0	6.8	18.5	-35.7
Palang	2	168	288	0.64	0.71	0.76	—	29.2	0.1	9.7	21.8	-28.2	-14.0	-0.2	—
Sylvania	4	130	190	0.14	0.48	0.60	0.62	7.5	1.7	21.2	48.2	-27.6	-5.7	10.4	22.2
Tumba	4	187	402	0.29	0.42	0.52	0.81	33.3	-14.0	4.9	25.3	-5.6	7.1	20.5	101.6
UniMich	2	150	170	0.11	0.48	0.67	—	1.4	-8.4	11.4	34.4	-9.8	3.5	19.9	—
Average	5	169	359	0.23	0.43	0.60	0.80	1.9	-11.2	11.1	33.4	-15.3	4.7	22.9	-31.0

covariance measurement sites show a systematic underestimation of the turbulent heat fluxes (Foken 2008; Stoy et al. 2013) and an energy balance correction is often applied to the observed turbulent heat fluxes. The energy balance correction is strongly debated since (i) the magnitude of heat storage changes in the canopy and soil is difficult to assess (Moderow et al. 2009; Eshonkulov et al. 2019) and (ii) it is unclear how the gap should be partitioned between the sensible and the latent heat flux (e.g., Roo et al. 2018). For a broad assessment of the closure gap we used the Bowen ratio correction (Twine et al. 2000), where the energy balance closure gap  $Q_{\text{gap}} = R_n - (H + \lambda E + dU_s/dt)$  is partitioned to sensible and latent heat fluxes by the Bowen ratio  $B_r$ . We used the daily EF estimates [Eq. (4)] to correct the latent and sensible heat fluxes.

However, note that the energy balance closure gap may itself show a distinct phase shift to solar radiation. This could easily be caused by errors in estimated heat storage. In this case the Bowen ratio correction will add this phase shift to the corrected turbulent heat fluxes. Therefore, the main part of the analysis of this paper focuses on the uncorrected values, which is also true for the previous publications using the PLUMBER dataset (Best et al. 2015; Haughton et al. 2016; Nearing et al. 2018).

### b. Land surface models

The PLUMBER dataset contains output from eight different core models with multiple model versions for some such that 13 different LSMs are finally compared (see Table 2). All models were run with meteorological forcing data from each site including  $R_{\text{sd}}$ , incoming longwave radiation  $R_{\text{ld}}$ , near surface air temperature  $T_a$ , wind speed  $u$ , relative humidity  $r_H$ , and precipitation  $P$ , all at half-hourly resolution (Best et al.

2015; Nearing et al. 2018). The vegetation type, vegetation height and reference height of each site were used. Otherwise, default parameter values were used with soil parameters from internal routines. The models differ substantially in their structure and have thus different simulated output variables. Sensible  $H$  and latent heat  $\lambda E$  fluxes were reported for all models, which is why we focus on these surface heat fluxes. For further interpretation in terms of surface energy balance, we also included net radiation balance  $R_n$  and the soil heat flux of these models. In

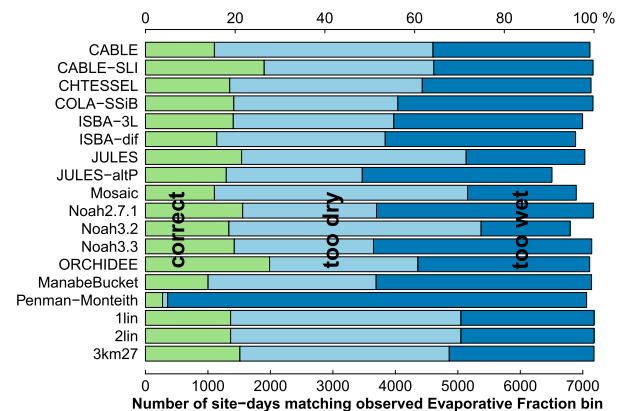


FIG. 3. Bar plot of the number of days matching the observed evaporative fraction within the filtered dataset. Green bars show the number of days with correct evaporative fraction, light blue refers to too dry (low EF), and dark blue too wet model simulation. A bin width of 0.1 was used to evaluate if a model is correct, too low, or too high.

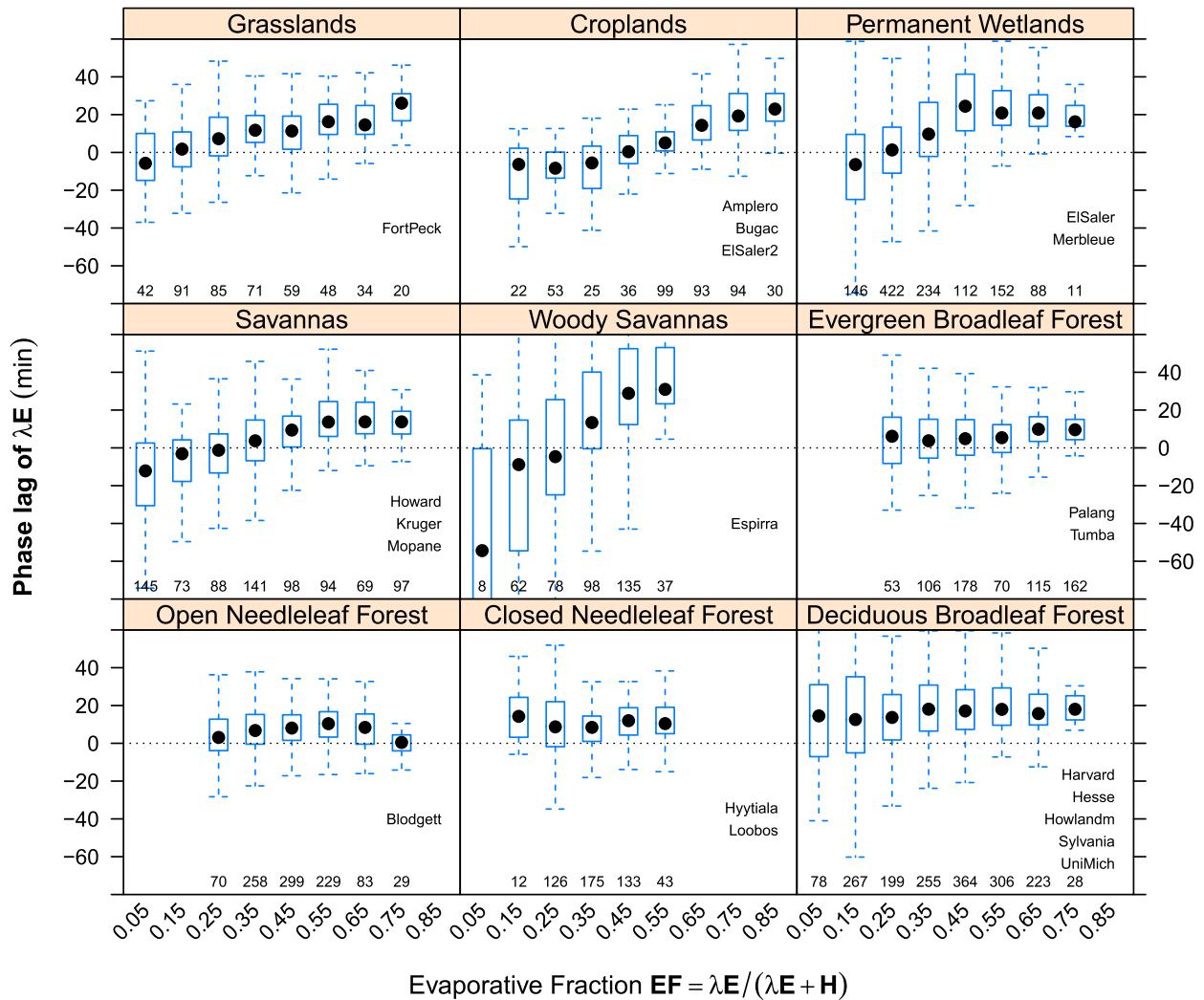


FIG. 4. Boxplot of observed daily phase lags of the latent heat flux  $\lambda E$  for different EF bins and aggregated by land-cover classes. Small numbers at the bottom of each panel show the number of samples per bin.

addition to the LSMs, Best et al. (2015) included benchmark approaches consisting of the physical benchmarks, the Manabe bucket model and the Penman–Monteith potential evapotranspiration (no water limitation), as well as three empirical benchmarks that include a simple linear regression to  $R_{sd}$ , a multiple-linear regression with  $R_{sd}$  and  $T_a$  (2lin), and a cluster-based regression using relative humidity as a third variable (3km27). These regressions were established out-of-sample and not for each site specifically.

#### 4. Results

##### a. Evaporative conditions

The PLUMBER dataset provides a wide range of evaporative conditions, expressed by the daily EF [Eq. (4)]. The dataset contains rather dry sites like Mopane and Kruger with a median EF  $\leq 0.23$ , and rather wet sites like Palang and ElSaler with median EF = 0.7. The wetter sites show less variation in EF on the order of 20%, while drier sites show variations on the order of 50%, see Table 3.

The fractional energy balance closure of the sites was obtained by linear regression of the sum of the sensible and latent heat fluxes against the available energy for each day. Available energy was derived from the difference of net radiation and soil heat flux data, when available. The median energy balance closure is on average 80%. However, it ranges between 55% at the site ElSaler2 (irrigated cropland) and 93% at the neighboring site ElSaler, which is classified as a wetland, see Table 3.

Next, we evaluate the LSM ability to predict the evaporative fraction for a given day, cut into 10% bins. We find very low agreement of about 20% for the LSMs, see Fig. 3. While the LSMs show seasonal variation in EF, the empirical benchmarks show rather constant EF across season and sites, since these have been obtained from an out-of-sample fit using the remaining 19 sites. The Penman–Monteith physical benchmark, which predicts potential evapotranspiration, clearly overestimates EF agreeing in only about 5% of all days in the filtered dataset.

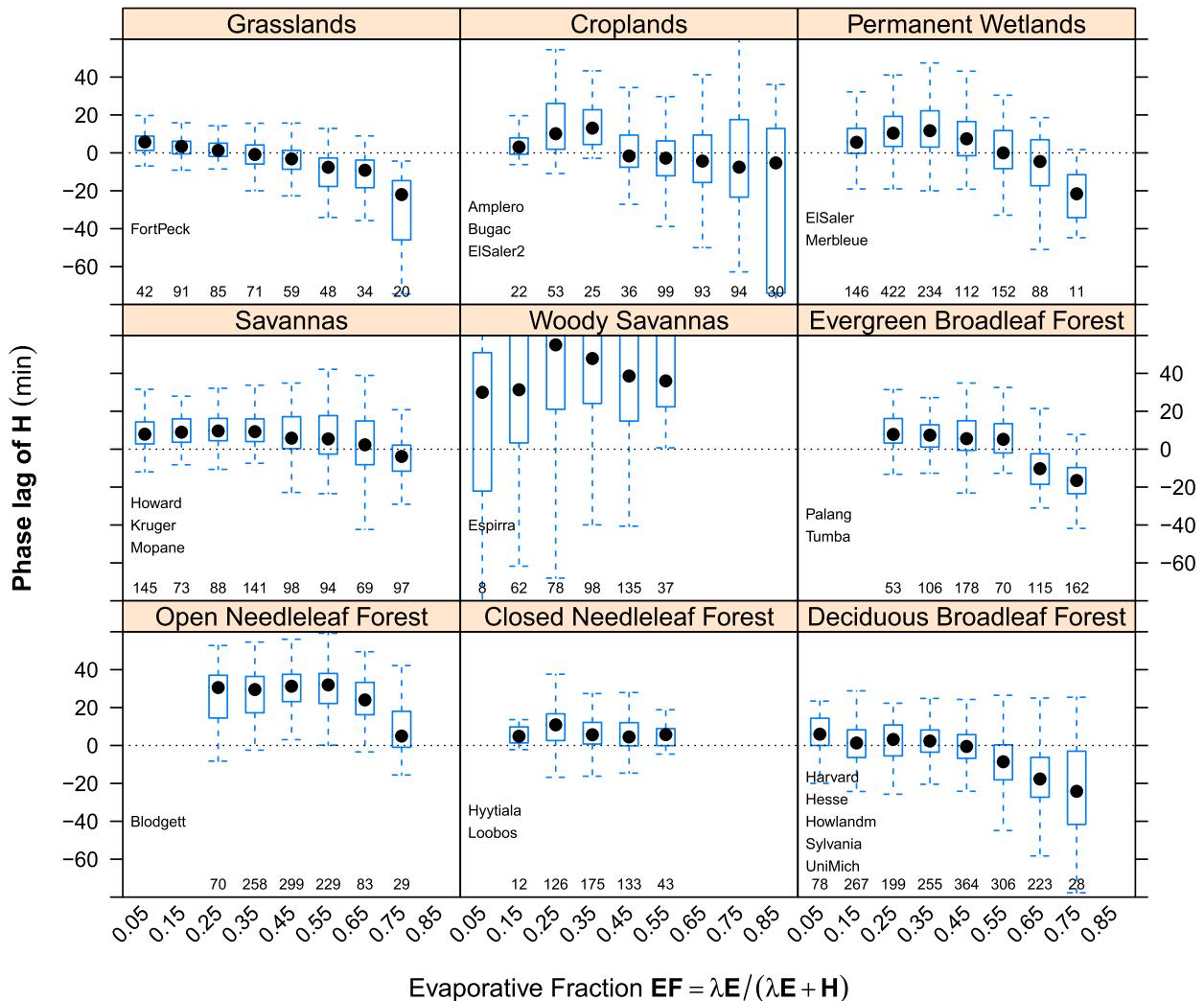


FIG. 5. Boxplot of daily phase lags of sensible heat fluxes  $H$  for different EF and aggregated by IGBP land-use classes.

The poor agreement of simulated EF is directly related to the response of the sensible and latent heat fluxes to solar radiation, since the sum of the turbulent heat fluxes is strongly and almost linearly related to solar radiation (Kleidon and Renner 2018; Conte et al. 2019). Here, daily EF explains 79% of the slope of the latent heat flux to solar radiation in the observations. The response is strongly linked to the diurnal magnitude of the turbulent heat fluxes. The model error in EF and the error in the diurnal magnitudes of the latent heat flux (derived from the diurnal range) is strongly linearly related with an explained variance of 69% across all models. Therefore, the poor performance in EF is in line with the poor simulation of the diurnal magnitude of the sensible as well as the latent heat fluxes by the LSMs and the benchmarks.

#### b. Phase lag of observations

Phase lags of the sensible and latent heat fluxes ( $\phi_H$  and  $\phi_{\lambda E}$ , respectively) were computed for each day and are evaluated

for the filtered days, see Table 3. The Bernardi–Camuffo regression is able to explain the dominant variation of the diurnal cycle of the latent heat flux (average across all sites  $R^2 = 0.91$ , lowest median  $R^2 = 0.75$  at Kruger) and the sensible heat flux (average across all sites  $R^2 = 0.95$ , lowest median  $R^2 = 0.91$  at Howard and Hyttiala). This emphasizes the strong control of incoming solar radiation on turbulent heat fluxes and the statistical robustness of our signature approach. On average, the phase lag of  $H$  is about +5 min, while  $\lambda E$  has a lag of +11 min relative to  $R_{sd}$ . There is considerable variation of site median values and within each site. Generally,  $\phi_H$  and  $\phi_{\lambda E}$  show different patterns whereby  $\phi_{\lambda E}$  is larger than  $\phi_H$  for 14 out of 20 sites. Systematically larger phase lags of the sensible heat flux are found at Blodgett, Espirra, El-saler, Kruger, Tumba, and Mopane, all being rather dry ecosystems.

To investigate the effects of water availability on the diurnal turbulent heat exchange we separated the days by bins of 0.1 of the daily evaporative fraction. We also grouped the sites by the

land-cover characteristics following the IGBP standards, see Fig. 4 for  $\phi_{\lambda E}$  and Fig. 5 for  $\phi_H$ . We find that  $\phi_{\lambda E}$  shows a significant effect of wetness at open vegetation sites (grasslands, cropland, wetlands, and savannas; Fig. 4). Under dry conditions a smaller phase lag (around 0) exists, which increases with EF to about 20 min. For the site Espirra, which is classified as woody savanna, this effect of water availability is further enhanced. Interestingly, forest ecosystems do not show such a reduction of the phase lag  $\phi_{\lambda E}$  when dry (see Fig. 4) and remain rather stable with positive phase lags. The pattern for the phase lag of the sensible heat flux appears to be rather different, see Fig. 5. Under dry conditions (EF < 0.5) we find phase lags around 0 for most sites, except at Espirra and Blodgett, which show rather large positive phase lags at around 30 min. Under wetter conditions with EF > 0.6,  $\phi_H$  becomes negative at the grassland, cropland, wetland and broadleaf forest sites. At needleleaf forest sites  $\phi_H$  remains relatively stable across different EF.

As a reference we also computed the phase lag of net radiation  $\phi_{R_n}$  (Table 3). Net radiation comprises incoming and outgoing longwave as well as shortwave radiation and balances the turbulent heat fluxes and heat storage fluxes such as the ground heat flux. Usually, incoming shortwave radiation dominates the diurnal variation of  $R_n$ , since longwave radiation has much smaller diurnal amplitude but typically with positive phase lags with respect to shortwave radiation (Renner et al. 2019b). Therefore, we expect  $\phi_{R_n} \approx 0$ . It appears that this is violated at sites Palang and Tumba, which show a lag of 30 min, most likely a timing offset in the radiation data. The timing of the turbulent heat fluxes, however, appears to be correct.

c. Ranking model performance with phase lags

The Bernardi–Camuffo regression also explains a large part of the diurnal cycle of heat fluxes simulated by the LSMs. To quantify the performance of the models to reproduce the observed phase lags of the turbulent fluxes, we evaluated the distribution of phase lags of both the sensible and the latent heat flux for a given site and bin of EF. Model agreement was assessed with the two-sample Hotelling’s  $T^2$  test at a significance level  $\alpha = 0.01$ . Figure 6 shows the frequency of agreement for each model aggregated across all sites and bins of EF. The LSMs are ordered by agreement, with the highest performance of the Mosaic model with 33% agreement. Models with more than 15% agreement are the versions of the JULES models, CABLE and CHTESSEL. The empirical benchmark with three input variables (3km27), including humidity, shows relatively good performance, while the other two empirical benchmarks show no agreement since these have no lag by definition (1lin) or very small lags and hence a very small variation. A range of LSMs show rather poor agreement and fall between that of the physical model benchmarks, namely, the Penman–Monteith and Manabe bucket models. This includes all the Noah3.2 LSM, ORCHIDEE, and COLA-SSiB models as well as the two versions of the ISBA LSM.

The evaluation of the models was done with the observed turbulent heat fluxes without the energy balance correction. Applying a correction relies on an accurate estimation of the

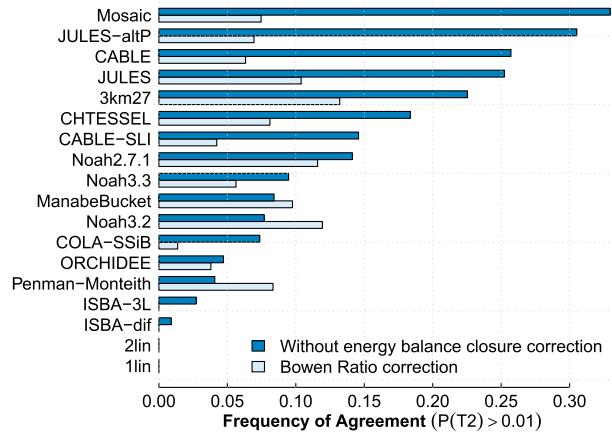


FIG. 6. Ranking of model performance for phase lags of sensible and latent heat by frequency of agreement across all sites and evaporative conditions. The agreement is tested by a Hotelling’s  $T^2$  test based on the distribution of phase lags of both  $\phi_H$  and  $\phi_{\lambda E}$  for a given site and 10% bin of EF. Dark blue bars show the statistics without correction for energy balance in observations and light blue shows with Bowen ratio correction.

available energy and heat storage fluxes. As a preliminary assessment of whether energy balance played a role in our results, we reevaluated the models with the Bowen ratio correction of the turbulent heat fluxes, using site net radiation and soil heat flux estimates. Other heat storage terms could not be included, since the observational dataset does not contain sufficient data. It appears that the closure gap itself showed large phase lags at many sites (see Table 3). Such a phase lag in the closure gap changes the diurnal course of the corrected fluxes. In some cases this resulted in significantly different phase lags of both fluxes. However, we find that for most models the agreement with corrected observations is much lower (see Fig. 6). Hence we will focus on the uncorrected fluxes in the following.

The ability to reproduce phase lags at the site level is rather mixed. We ordered the performance per model and site in a model performance matrix, shown in Fig. 7. Some sites are easier to predict like Amplero and Kruger, while only the Mosaic model achieves some agreement at the sites Tumba and Blodgett.

Next, we evaluated whether the models tend to over or underestimate the observed phase lags. Therefore, we assessed the mean difference in phase lags for each site and evaporative conditions. The boxplot in Fig. 8 shows that all LSMs tend to overestimate  $\phi_{\lambda E}$ . The median bias ranges between +10 and +30 min for the poorer models. The three empirical benchmarks underestimate both  $\phi_{\lambda E}$  and  $\phi_H$ , because solar radiation is used as the key predictor variable and there is no energy balance constraint. The best models (Mosaic, CABLE, and the two JULES versions), show a large variation, but no overall bias in  $\phi_H$ . The models with a medium rank in the  $T^2$  statistics (CHTESSEL, CABLE-SLI, the Noah LSMs) show systematic underestimation of  $\phi_H$ . COLA-SSiB overestimates both  $\phi_{\lambda E}$  and  $\phi_H$ . The LSMs with poorest ranks show systematically large  $\phi_{\lambda E}$  on the order of

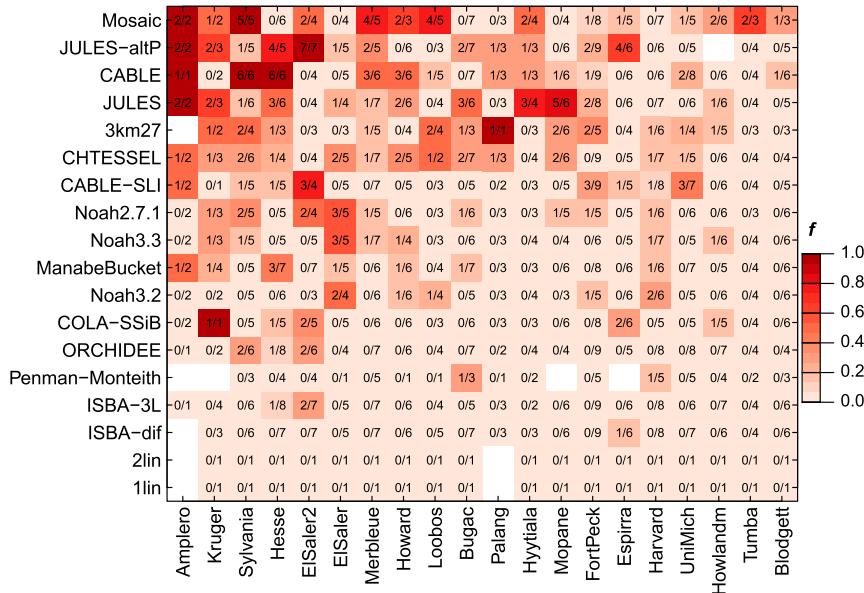


FIG. 7. Model performance matrix of phase lags of sensible and latent heat per model and site. Models are ordered by their overall performance across sites. Sites are ordered by their relative agreement from high at the left to low at the right side. The color represents the relative agreement between nonrejected Hotelling's  $T_2$  tests and the number of tests applied per site. The total number of tests differs between models and sites. This is because the EF of a model may have no agreement with the EF in observations, e.g., Penman-Monteith at the dry site Kruger. These missing conditions are shown with white color.

30 min, whereby the ISBA models also show a systematic overestimation of the sensible heat flux.

We also evaluated the phase lag of the soil heat flux, which shows large discrepancies with the observed values (see Fig. 8). All models show a negative phase lag, i.e., higher values during the morning than in the afternoon. The differences are on the order of 90 min and can be as large as 3 h. A bias in the phase lag of the modeled soil heat flux will also influence the turbulent heat fluxes in a LSM because of the surface energy balance constraint. The effect on the turbulent fluxes, however, is only large when the magnitude of the soil heat flux is also large. Therefore, we assessed the diurnal magnitude of the soil heat fluxes, through the daily range between the minimum and maximum soil heat flux. The results are shown in aggregated form for short vegetation, savanna, and forest in Fig. 9. This shows that the observed soil heat fluxes are much smaller than the modeled fluxes at forest and short vegetation sites. For savannas there is no clear systematic bias across all the models. Hence, the modeled soil heat fluxes have rather large magnitudes and a different sign for the phase lags, which reveals a potential misrepresentation of heat storage changes.

#### d. Diurnal composites

To illustrate the performance of models at the diurnal cycle we classified the sites into short vegetation (three sites) and forest vegetation (four sites). Then hourly composites of the data were computed using the median for a given bin of EF (bin width 0.1). In Fig. 10 we show examples of these composites for two poor performing models and two relatively well performing models in short vegetation and forests. The figure

shows the latent, sensible, and soil heat fluxes against the observed solar radiation. Under dry conditions at sites with short vegetation, the observations show a rather linear response of the turbulent heat fluxes without a distinct hysteresis (Figs. 10a,c). Only the soil heat flux shows an anticlockwise hysteresis that is smaller in the morning than in the afternoon, which corresponds to a positive phase lag.

Figure 10a shows the corresponding results for CABLE, which reveals a much too large anticlockwise hysteresis of the sensible heat flux while the almost linear response of the latent heat flux corresponds well to the observations. The large hysteresis of the sensible heat flux is linked to clockwise hysteresis of the soil heat flux, which is higher in the morning than in the afternoon. The discrepancies of the phase lags of  $H$  are linked to the high diurnal magnitude of the soil heat flux of CABLE (Fig. 9) and its negative phase lag (Fig. 8). Results for the JULES model are shown in Fig. 10c for short vegetation under dry conditions. The JULES model exhibits a very small hysteresis of the heat fluxes and corresponds much better to observations. Only the phase lag of the latent heat flux is slightly overestimated.

As a further example, we chose forests sites under ample moisture conditions (Figs. 10b,d). Observations show rather large hysteresis loops for the turbulent heat fluxes, where the latent heat flux shows a pronounced anticlockwise hysteresis, which is compensated by a smaller clockwise hysteresis of the sensible heat flux. Observations of the soil heat flux in forests show very small values both in magnitude and in terms of phase lag. However, there is no assessment of the heat storage terms

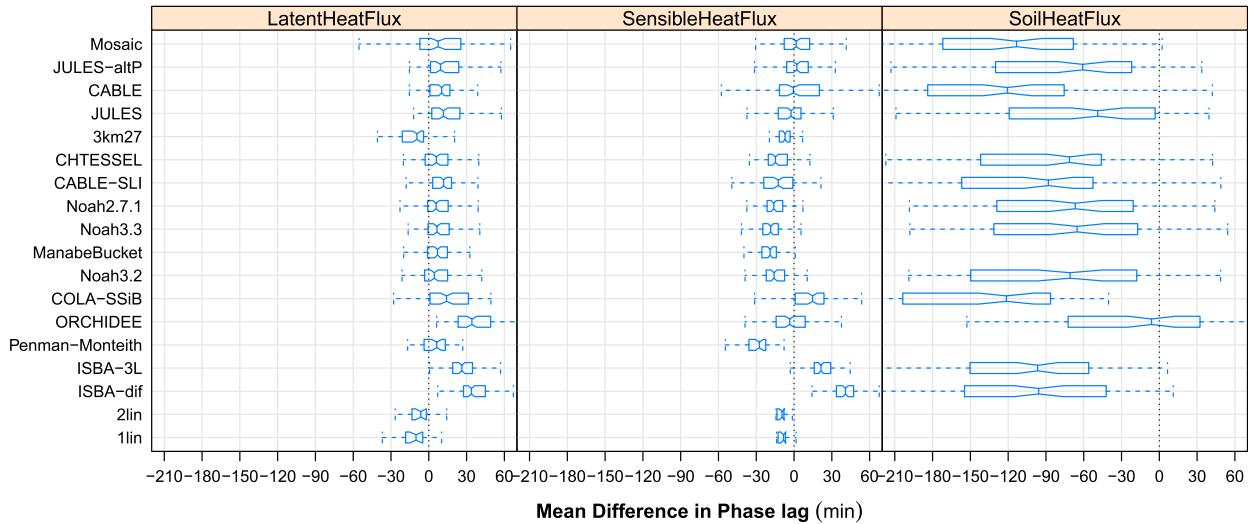


FIG. 8. Average difference of the simulated minus observed phase lags for the latent, sensible, and soil heat fluxes. Data are based on the average per EF bin and site. Note that not all sites have soil heat flux observations.

in the biomass, which can be quite large (Meier et al. 2019). Figure 10b shows the ISBA-dif model that exhibits a large clockwise hysteresis and magnitude of the soil heat flux, a common issue of this model (Figs. 8 and 9). For these reasons the ISBA-dif model produces very large hysteresis loops for the turbulent heat fluxes. Their direction, however, corresponds well with the observations. Figure 10d shows the ORCHIDEE LSM, which better reproduces the diurnal patterns. This is related to a smaller hysteresis loop of the soil heat flux, see also Fig. 8. ORCHIDEE, however, consistently overestimates the hysteresis of the latent heat flux, which results in poor agreement when evaluated at the site level, see Fig. 7.

### 5. Discussion

We employed the approach of *diagnostic signatures* (Gupta et al. 2008) to evaluate complex models by a typical behavior that can be quantified both from observations and models. As diagnostic signature we employed the metric of a phase lag of heat fluxes to solar radiation at the subdaily time scale. In the following we discuss the typical signatures obtained from observations, and discuss the LSM performance to diagnose which process representations need to be improved.

#### a. Phase lag in observations

We focused on the diurnal variation of the turbulent heat fluxes that were obtained from eddy covariance observations from the FLUXNET La-Thuille synthesis dataset. The diurnal variation of turbulent heat fluxes can be rather well captured by its direct response to solar radiation and its time derivative, explaining more than 90% of subdaily variation under the cloud-free, summertime conditions evaluated here. This emphasizes the strong control of solar radiation for land-atmosphere exchange but also highlights the presence of heat storage effects that are captured by the phase lags.

Both can be well identified on a daily basis. Our results further show significant temporal variation in the phase lags as well as differences between sites. Still we find convergence of phase lag behavior within broad vegetation categories, consisting of forests, savannas and short vegetation. Short vegetated sites show a significant influence of water limitation on the diurnal course of the turbulent fluxes. Under ample moisture supply and high EF there is a positive phase lag in the latent heat flux and a tendency for negative phase lags of the sensible heat flux. Under drier conditions the positive phase lag of the latent heat flux is reduced toward zero. This behavior is consistent with a recent analysis of turbulent heat fluxes at a grassland site by Renner et al. (2019b), who showed a reduction of phase lags during a dry-down.

The savanna type ecosystems show an enhanced response of  $\phi_{\lambda E}$  to EF with negative phase lags under very dry conditions ( $EF < 0.2$ ). This behavior is in agreement with Nelson et al. (2018), who used the method of diurnal centroids (Wilson et al. 2003) on the FLUXNET2015 dataset. The forest ecosystems evaluated here, however, show no response of  $\phi_{\lambda E}$  to EF. Needleleaf forests also show no response of  $\phi_H$  to EF, while broadleaf forests show a decline toward negative  $\phi_H$  under high EF.

The positive  $\phi_{\lambda E}$  can be regarded as a response to the diurnal cycle of the vapor pressure deficit (VPD) of the air as a driver of evapotranspiration. The diurnal cycle of VPD is strongly related to air temperature and both variables show large positive phase lags with respect to solar radiation. This is due to the heat storage changes in the lower, well mixed atmosphere. Hence we argue that VPD, which is highest during early afternoon, enhances  $\lambda E$  leading to a positive  $\phi_{\lambda E}$  (Renner et al. 2019b). This explanation is in line with the phase lag observations in forests and under wet conditions in short vegetation and savannas. The significant reduction of  $\phi_{\lambda E}$  in short vegetation and savannas toward water limited conditions can be related to the enhanced surface resistance to evapotranspiration leading to a decoupling of the surface from the

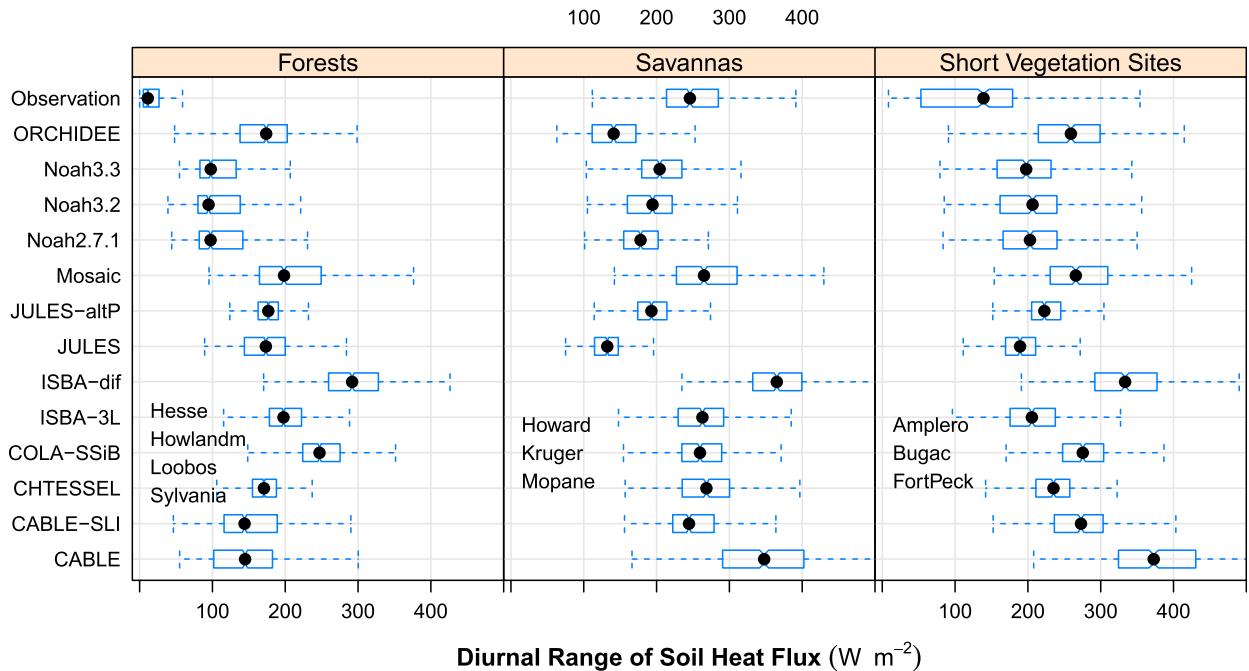


FIG. 9. Distribution of observed and modeled diurnal range of the soil heat fluxes of the filtered data, aggregated by sites belonging to broad land-cover groups.

atmosphere (McNaughton and Jarvis 1983). As a result, surface skin temperature increases more strongly than air temperature (Mildrexler et al. 2011; Panwar et al. 2019), which also enhances subsurface heat storage changes. This decoupling is typical for sites with short vegetation or an open canopy where the ground is not sufficiently shaded from solar radiation. Negative phase lags of  $\lambda E$  observed under very dry conditions could also be related to water saving strategies of plants that control gas exchange by stomata and close under very high VPD conditions leading to a morning shift of  $\lambda E$  (Nelson et al. 2018).

The pattern of the phase lag of the sensible heat flux is strongly coupled to the phase lag of the latent heat flux. For most sites there is only a small phase lag of the sensible heat flux and this explains the good performance of the simple linear regression benchmark (1lin) used by Best et al. (2015). The shift to negative phase lags under wet conditions can be explained by a compensation due to the positive phase lags of the latent heat flux to keep the surface energy balance. A deviation of this explanation is found for Blodgett and Espirra that show relatively large positive phase lags of the sensible heat flux throughout. Both sites have a relatively open canopy (Blodgett was reforested at that time, Misson et al. 2005) and we speculate that these patterns are induced by relatively large diurnal soil heat fluxes.

#### b. Energy balance closure

One of the key assumptions of this work is that the turbulence measurements correctly capture the diurnal cycle of the sensible and latent heat fluxes. The sites studied in this analysis achieve an energy balance closure between 55% and 93% of available energy ( $R_n - dU_s/dt$ ), assuming that reported soil heat fluxes capture heat storage components. The estimated

energy balance closure gap revealed significant phase shifts for many sites (Table 3), with the implication that the phase lag of the corrected fluxes also changed. Comparison with the modeled phase lags revealed an even poorer comparison than without correction (Fig. 6). We believe that the unaccounted heat storage in the subsurface and canopy are a crucial limitation of the energy balance correction. These heat storage components, such as in the upper soil layer, can have significant diurnal magnitudes and phase shifts. Therefore, the diurnal cycle and especially the phase lag of the corrected heat fluxes may not be reliable. On the other hand, we have no reason to believe that there are systematic measurement errors with a time lag in eddy covariance measurements, since these are based on the instantaneous turbulent fluctuations of wind speed, temperature and moisture.

Given the lack of further data to calculate these heat storage components, we cannot further assess this issue within the scope of this paper. However, we strongly recommend a detailed measurement of subsurface and canopy heat storage components, which are important for the diurnal heat redistribution (Moderow et al. 2009; Gao et al. 2017; Heidkamp et al. 2018; Meier et al. 2019; Eshonkulov et al. 2019). Furthermore, alternative measurement technologies should be used in parallel to assess the validity of the diurnal cycle of eddy covariance turbulence measurements.

#### c. Diurnal cycle of simulated heat fluxes

All LSMs and benchmarks have been driven with meteorological forcing data specific to each site. Prescribing near surface meteorological states such as air temperature and humidity strongly constrains the model response and thus focusses on the partitioning of the surface energy balance terms.

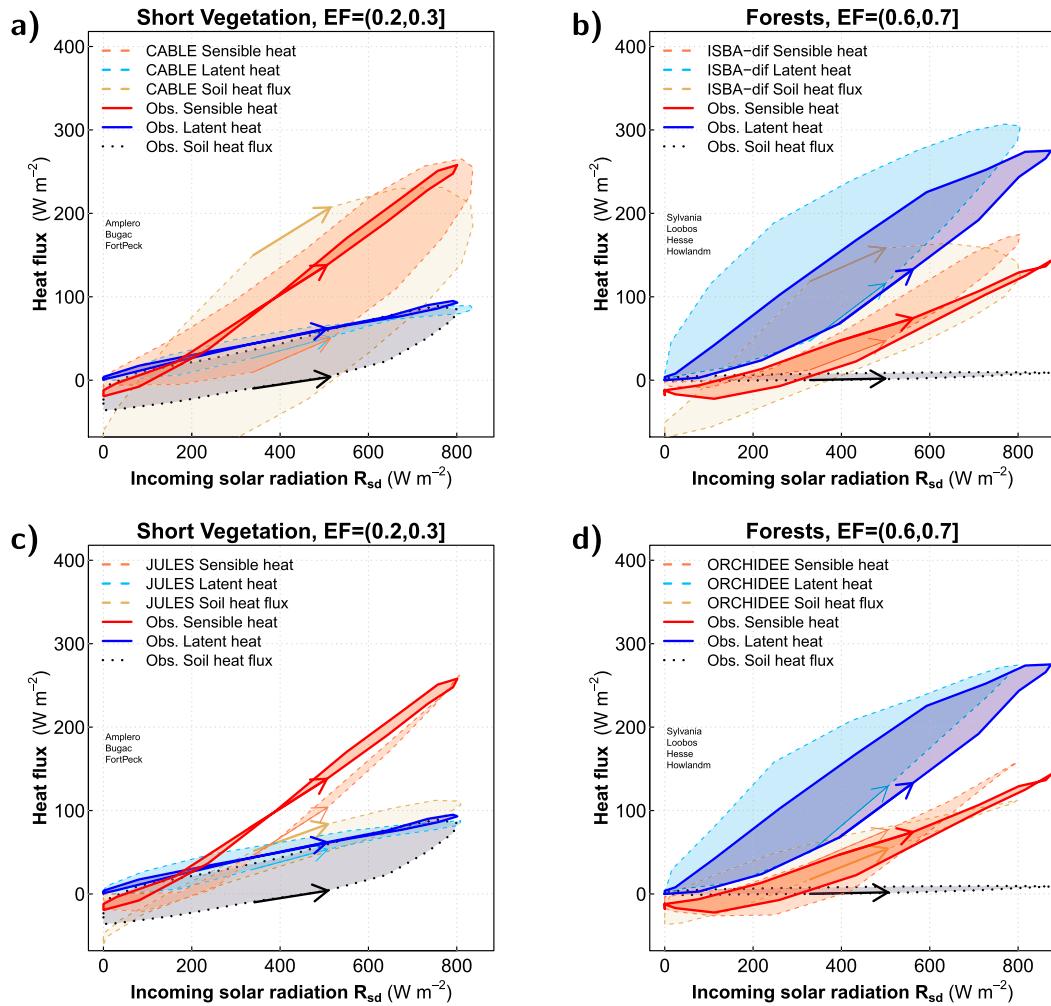


FIG. 10. Model to observation comparison based on diurnal composites of median diurnal cycles. Composites are derived for a specific bin of EF aggregated over a set of sites (a),(c) with short vegetation under dry conditions and (b),(d) for forested sites under moist conditions. Panels (a) and (b) show typical LSMs, which strongly overestimate the phase lag of turbulent heat fluxes. Panels (c) and (d) show better performing models. Arrows show the rising flux during morning hours.

All LSMs and benchmarks show large discrepancies in the simulated magnitude of the diurnal cycle of the sensible and latent heat fluxes. This is apparent from the poor simulation of evaporative fraction (Fig. 3), which is strongly influenced by water redistribution processes and the large memory of soil moisture. The poor performance of the PLUMBER models in terms of EF was noted by Ukkola et al. (2016) and is also a key issue in coupled model simulations and reanalysis (Dirmeyer et al. 2018). Therefore, we focused on the phase lag of turbulent heat fluxes, which is influenced by the LSM parameterizations that control heat redistribution.

Common to all models is a tendency to overestimate the phase lag of  $\phi_{\lambda E}$ , while there is no common bias for the sensible heat flux, see Fig. 8. When we evaluated both phase lags, the model performance was rather low (Figs. 6 and 7). Hence, although the models evaluated in PLUMBER show

distinct diurnal cycles, these are shifted in time and thus do not capture heat redistribution processes adequately.

Although there is incomplete observational data on the soil heat storage for the sites in the PLUMBER dataset, we find very large discrepancies for diurnal magnitude and the phase lag of the soil heat flux. In particular, the models show a negative phase lag of the soil heat flux, with higher values during the morning than in the afternoon. This does not correspond with the observations, which show positive phase lags. Consequently, the phase lag in the soil heat flux is compensated by the turbulent heat fluxes since net radiation does not show large phase lags, see Table 3. In addition, the models overestimate the diurnal magnitude of the soil heat flux compared to the observations. A higher magnitude of the soil heat flux then enhances this effect on the turbulent fluxes. This is readily seen, for example, for the CABLE model in Fig. 10a.

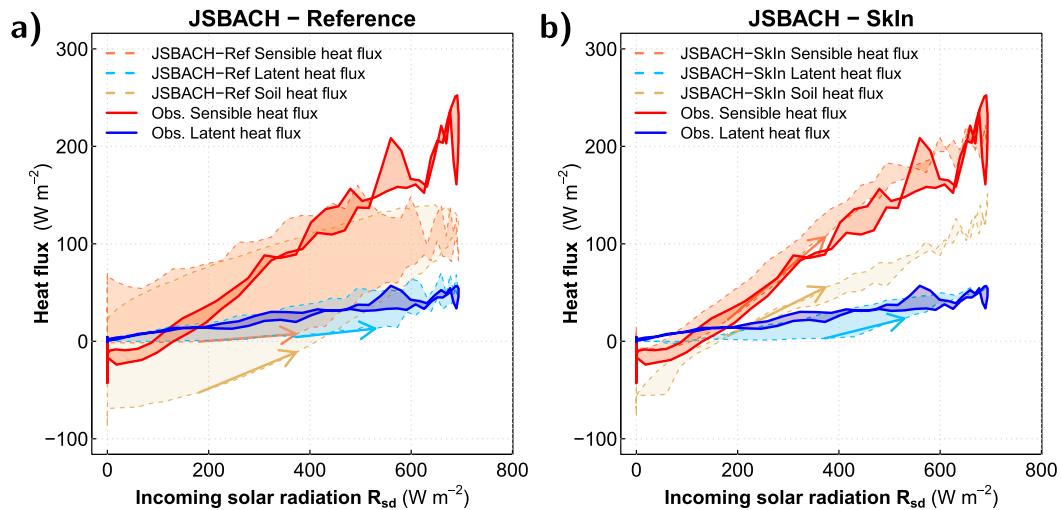


FIG. 11. Simulations and observations of turbulent heat fluxes during DICE. (a) Simulations with the reference JSBACH land surface model using the upper soil layer to solve for the surface energy balance. (b) The new version of the JSBACH model with a Skin layer. Data are obtained from Heidkamp et al. (2018) and DICE (<http://appconv.metoffice.com/dice/dice.html>).

#### d. Model diagnostics with phase lags

In general the diurnal phase lag signature will be sensitive to the following parameterizations of LSMs: (i) the method to solve the surface energy balance, (ii) the dynamic parameterization of surface/canopy conductance, and (iii) the parameterization of surface layer exchange of heat, moisture and momentum. Since the soil heat flux seems to be the main cause for the observed biases we believe that the way in which the surface energy balance is solved is of greatest importance to improve the modeled diurnal cycles.

##### 1) SURFACE ENERGY BALANCE SOLUTION

The surface energy balance is the central component of any LSM. It links all different parameterizations particularly the transfer of heat into the subsurface and into the atmosphere. This partitioning is critical for the phase lags as it relates the strong diurnal variation of solar radiation to storage changes below and above the surface.

The key state variable is surface (skin) temperature, which is often part of the other parameterizations such as for soil heat transfer but also for canopy and surface layer exchange. There are two different approaches to solve this nonlinear problem. One way is to use the temperature of the upper soil layer. This layer has a large heat capacity and dampens the temporal variation of the surface temperature depending on the depth of the layer. The second approach is to use a skin layer that has no heat capacity but requires iteration to get a stable solution (Viterbo and Beljaars 1995; Best et al. 2005; Heidkamp et al. 2018).

##### 2) TOP SOIL VERSUS SKIN LAYER APPROACH

Next we illustrate the impact of surface energy balance schemes on the phase lags of the heat fluxes using output of two

different versions of the JSBACH model, which is the LSM scheme of the ECHAM climate model. Heidkamp et al. (2018) used the JSBACH model in an offline mode and simulated the diurnal cycle with the reference version of the JSBACH model (i.e., with upper soil layer to solve the surface energy balance) and the new version with a skin layer (JSBACH-Skin). For illustration Heidkamp et al. (2018) used observational data from the Diurnal Land/Atmosphere Coupling Experiment (DICE, <http://appconv.metoffice.com/dice/dice.html>) project. This is based on the well-known CASES99 campaign at the Southern Great Plains sites in the United States over crop land. These simulations are ideal to illustrate the impact of these schemes on the phase lags of the heat fluxes. We obtained the simulation output data from Heidkamp et al. (2018) and compared the hysteresis patterns. Figure 11a shows the hysteresis on a single day for the observations and the reference model. The observations reveal the linear response of the turbulent heat fluxes under dry conditions ( $EF = 0.17$ ) as seen in the PLUMBER dataset (Fig. 10a). The JSBACH-Reference model, however, shows a large phase lag in the simulated sensible heat flux as well as in the soil heat flux. In contrast, the new model version with a skin layer does not show such large hysteresis of any fluxes and corresponds much better to observations, see Fig. 11b. This highlights the impact of how the surface energy balance is solved on the phase lag of heat fluxes.

Although Heidkamp et al. (2018) show that the skin layer approach is superior for the JSBACH model, we do not find that PLUMBER models that use a skin layer approach (CHTESSEL, COLA, Noah LSMs) are generally superior to models that use the top soil layer (see Table 2). Figure 10 highlights that the JULES model performs well on average even though it employs a top soil layer approach. However, JULES uses a soil heat transfer scheme in which the vertical layers are optimized to account for the diurnal to annual

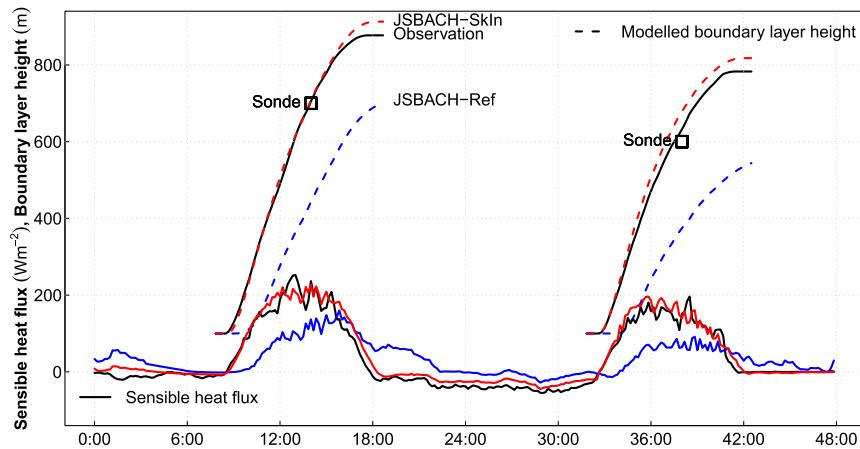


FIG. 12. Sensible heat fluxes (continuous lines) and derived planetary boundary layer height (dashed) for two days of the DICE experiment. Sensible heat flux data are obtained from Heidkamp et al. (2018), DICE (<http://appconv.metoffice.com/dice/dice.html>), and radio-sounding data of PBL height is taken from Steeneveld et al. (2008).

frequencies of the atmospheric forcings (Best et al. 2005, 2011). The optimized vertical discretization scheme includes four layers with thinner layers on top and is based on solutions of the soil heat diffusion equation (Best et al. 2005).

### 3) POTENTIAL BIAS ON PLANETARY BOUNDARY LAYER DYNAMICS

To illustrate how biases in the diurnal cycle of turbulent heat fluxes propagate when coupled to the atmosphere we used the two cases of Heidkamp et al. (2018) and looked at what these imply for the diurnal development of the convective planetary boundary layer. Using a simple bulk mixed layer scheme we simulated the growth of the boundary layer given the sensible heat flux as an input. Example results are shown for the two days of the DICE dataset, for which the height of the boundary layer was estimated by radio soundings (Steenefeld et al. 2008). Figure 12 shows the sensible heat fluxes whereby the JSBACH-Reference model has a damped diurnal cycle and lags the simulated growth of the PBL compared to JSBACH-Skin. In contrast the simulated PBL growth for the observed and the simulated sensible heat flux from JSBACH-Skin agree well with the observed height.

This case study highlights the importance of how and where the diurnal forcing of solar radiation is buffered in the land-atmosphere system. Misrepresentation of heat storage changes, such as in the subsurface, directly affect the exchange of heat and water with the atmosphere. Systematic offsets cannot only influence the diurnal cycle of heat fluxes, but also affect temperature and will propagate into the lower atmosphere in coupled models (Heidkamp et al. 2018). Offsets in boundary layer height affect the mixing of heat, momentum and moisture and potentially the timing of the development of boundary layer clouds (de Arellano et al. 2014). While modeling these interactive processes is a challenge, the strong local coupling of the land-atmosphere system also leads to simple emergent relationships such as the almost linear response of the sensible heat flux to solar radiation,

which can serve as a valuable signature to evaluate land-atmosphere feedbacks represented in complex models.

## 6. Conclusions

We performed a diagnostic signature based analysis of state-of-the-art land surface models. We focused on the simulated diurnal cycle of the turbulent heat fluxes and used the phase lag to solar radiation as a metric to quantify the model's ability to resolve a signature that is characteristic of the diurnal cycle. The evaluation is based on a set of 20 FLUXNET sites with diverse climatic and land surface conditions. As known from previous assessments the models show a poor ability to resolve the partitioning of sensible and latent heat fluxes, which directly affects the diurnal magnitude of these fluxes. This issue is related to difficulties in simulating water redistribution processes. Hence we focus on the phase lag of the heat fluxes, which is influenced by diurnal heat redistribution processes and found to be influenced by vegetation type and water limitation. We find that all models tend to overestimate the phase lag of the latent heat flux, combined with varying biases in the sensible heat flux. These biases appear to be systematic and thus provide a further reason for the poor performance of the PLUMBER LSMs found in earlier studies (Best et al. 2015; Haughton et al. 2016; Nearing et al. 2018).

Modeled soil heat fluxes have much larger diurnal magnitudes than the available observations and their phase lags to solar radiation are strongly biased. These larger diurnal fluxes and large phase lags propagate to the phase lag of turbulent heat fluxes. We argue that this is related to issues in jointly solving the surface energy balance and soil heat transfer.

Our findings have important implications: 1) the models are not able to accurately capture diurnal heat exchange processes that are at the core of land-atmosphere interactions. The biases will propagate into the lower atmosphere when coupled and to longer time scales when aggregated. 2) The phase lags of the turbulent heat fluxes appear to be quite sensitive to biases

induced by the surface energy balance scheme including subsurface heat transfer. 3) After resolving these issues, we expect that the diurnal signature of the phase lag can help to identify and calibrate important subgrid-scale parameterizations of the canopy and surface layer, enabling the use of the large set of available data at the diurnal time scale to better constrain complex model parameterizations.

*Acknowledgments.* We are very grateful that the PIs of the respective FLUXNET sites shared their data openly. This work benefited from a workshop on “Improving the underpinnings of hydrological models” held at the University of Sopron, Hungary in April 2018. MR and AK were funded by the Max Planck Society.

*Data availability statement.* All data used were previously published as part of the respective manuscripts cited in the text. Code to perform the diurnal analysis can be obtained through Renner (2019b) or <https://github.com/laubblatt/phaselag>. Code to identify cloud-free days (Renner 2019a) is accessible at <https://github.com/laubblatt/clearskyQuantileRegression>.

#### REFERENCES

- Baklanov, A. A., and Coauthors, 2011: The nature, theory, and modeling of atmospheric planetary boundary layers. *Bull. Amer. Meteor. Soc.*, **92**, 123–128, <https://doi.org/10.1175/2010AMS2797.1>.
- Balsamo, G., A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts, 2009: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *J. Hydrometeorol.*, **10**, 623–643, <https://doi.org/10.1175/2008JHM1068.1>.
- Best, M. J., P. M. Cox, and D. Warrilow, 2005: Determining the optimal soil temperature scheme for atmospheric modelling applications. *Bound.-Layer Meteorol.*, **114**, 111–142, <https://doi.org/10.1007/s10546-004-5075-3>.
- , and Coauthors, 2011: The Joint UK Land Environment Simulator (JULES), model description - Part 1: Energy and water fluxes. *Geosci. Model Dev.*, **4**, 677–699, <https://doi.org/10.5194/gmd-4-677-2011>.
- , and Coauthors, 2015: The plumbing of land surface models: Benchmarking model performance. *J. Hydrometeorol.*, **16**, 1425–1442, <https://doi.org/10.1175/JHM-D-14-0158.1>.
- Boussetta, S., and Coauthors, 2013: Natural land carbon dioxide exchanges in the ECMWF integrated forecasting system: Implementation and offline validation. *J. Geophys. Res. Atmos.*, **118**, 5923–5946, <https://doi.org/10.1002/jgrd.50488>.
- Businger, J. A., J. C. Wyngaard, Y. Izumi, and E. F. Bradley, 1971: Flux-profile relationships in the atmospheric surface layer. *J. Atmos. Sci.*, **28**, 181–189, [https://doi.org/10.1175/1520-0469\(1971\)028<0181:FPRITA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<0181:FPRITA>2.0.CO;2).
- Camuffo, D., and A. Bernardi, 1982: An observational study of heat fluxes and their relationships with net radiation. *Bound.-Layer Meteorol.*, **23**, 359–368, <https://doi.org/10.1007/BF00121121>.
- Chen, T. H., and Coauthors, 1997: Cabauw experimental results from the project for intercomparison of land-surface parameterization schemes. *J. Climate*, **10**, 1194–1215, [https://doi.org/10.1175/1520-0442\(1997\)010<1194:CERFTP>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<1194:CERFTP>2.0.CO;2).
- Conte, L., M. Renner, P. Brando, C. O. Santos, D. Silvério, O. Kolle, S. E. Trumbore, and A. Kleidon, 2019: Effects of tropical deforestation on surface energy balance partitioning in Southeastern Amazonia estimated from maximum convective power. *Geophys. Res. Lett.*, **46**, 4396–4403, <https://doi.org/10.1029/2018GL081625>.
- de Arellano, J. V.-G., H. G. Ouwensloot, D. Baldocchi, and C. M. J. Jacobs, 2014: Shallow cumulus rooted in photosynthesis. *Geophys. Res. Lett.*, **41**, 1796–1802, <https://doi.org/10.1002/2014GL059279>.
- De Bruin, H. R., and A. M. Holtstlag, 1982: A simple parameterization of the surface fluxes of sensible and latent heat during daytime compared with the penman-Monteith concept. *J. Appl. Meteorol.*, **21**, 1610–1621, [https://doi.org/10.1175/1520-0450\(1982\)021<1610:ASPOTS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1982)021<1610:ASPOTS>2.0.CO;2).
- Decharme, B., A. Boone, C. Delire, and J. Noilhan, 2011: Local evaluation of the Interaction between Soil Biosphere Atmosphere soil multilayer diffusion scheme using four pedotransfer functions. *J. Geophys. Res.*, **116**, D20126, <https://doi.org/10.1029/2011JD016002>.
- Dirmeyer, P. A., and Coauthors, 2018: Verification of land-atmosphere coupling in forecast models, reanalyses, and land surface models using flux site observations. *J. Hydrometeorol.*, **19**, 375–392, <https://doi.org/10.1175/JHM-D-17-0152.1>.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **10**, 8851, <https://doi.org/10.1029/2002JD003296>.
- Eshonkulov, R., A. Poyda, J. Ingwersen, A. Pulatov, and T. Streck, 2019: Improving the energy balance closure over a winter wheat field by accounting for minor storage terms. *Agric. For. Meteorol.*, **264**, 283–296, <https://doi.org/10.1016/j.agrformet.2018.10.012>.
- Foken, T., 2008: The energy balance closure problem: An overview. *Ecol. Appl.*, **18**, 1351–1367, <https://doi.org/10.1890/06-0922.1>.
- Gao, Z., and Coauthors, 2017: A novel approach to evaluate soil heat flux calculation: An analytical review of nine methods. *J. Geophys. Res. Atmos.*, **122**, 6934–6949, <https://doi.org/10.1002/2017JD027160>.
- Guo, Z., and P. A. Dirmeyer, 2013: Interannual variability of land-atmosphere coupling strength. *J. Hydrometeorol.*, **14**, 1636–1646, <https://doi.org/10.1175/JHM-D-12-0171.1>.
- Gupta, H. V., T. Wagener, and Y. Liu, 2008: Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrol. Processes*, **22**, 3802–3813, <https://doi.org/10.1002/hyp.6989>.
- Haughton, N., and Coauthors, 2016: The plumbing of land surface models: Is poor performance a result of methodology or data quality? *J. Hydrometeorol.*, **17**, 1705–1723, <https://doi.org/10.1175/JHM-D-15-0171.1>.
- Haverd, V., and M. Cuntz, 2010: Soil-Litter-Iso: A one-dimensional model for coupled transport of heat, water and stable isotopes in soil with a litter layer and root extraction. *J. Hydrol.*, **388**, 438–455, <https://doi.org/10.1016/j.jhydrol.2010.05.029>.
- Heidkamp, M., A. Chlond, and F. Ament, 2018: Closing the energy balance using a canopy heat capacity and storage concept – A physically based approach for the land component JSBACHv3.11. *Geosci. Model Dev.*, **11**, 3465–3479, <https://doi.org/10.5194/gmd-11-3465-2018>.
- Henderson-Sellers, A., A. J. Pitman, P. K. Love, P. Irannejad, and T. H. Chen, 1995: The Project for Intercomparison of Land

- Surface Parameterization Schemes (PILPS): Phases 2 and 3. *Bull. Amer. Meteor. Soc.*, **76**, 489–504, [https://doi.org/10.1175/1520-0477\(1995\)076<0489:TPFIOL>2.0.CO;2](https://doi.org/10.1175/1520-0477(1995)076<0489:TPFIOL>2.0.CO;2).
- Holtlag, A. A. M., and Coauthors, 2013: Stable atmospheric boundary layers and diurnal cycles: Challenges for weather and climate models. *Bull. Amer. Meteor. Soc.*, **94**, 1691–1706, <https://doi.org/10.1175/BAMS-D-11-00187.1>.
- Hotelling, H., 1931: The generalization of Student's ratio. *Ann. Math. Stat.*, **2**, 360–378, <https://doi.org/10.1214/aoms/1177732979>.
- Jarvis, P. G., and K. G. McNaughton, 1986: Stomatal control of transpiration: Scaling up from leaf to region. *Adv. Ecol. Res.*, **15**, 1–49, [https://doi.org/10.1016/S0065-2504\(08\)60119-1](https://doi.org/10.1016/S0065-2504(08)60119-1).
- Kleidon, A., and M. Renner, 2013: Thermodynamic limits of hydrologic cycling within the earth system: Concepts, estimates and implications. *Hydrol. Earth Syst. Sci.*, **17**, 2873–2892, <https://doi.org/10.5194/hess-17-2873-2013>.
- , and —, 2018: Diurnal land surface energy balance partitioning estimated from the thermodynamic limit of a cold heat engine. *Earth Syst. Dyn.*, **9**, 1127–1140, <https://doi.org/10.5194/esd-9-1127-2018>.
- Koster, R. D., and M. J. Suarez, 1992: Modeling the land surface boundary in climate models as a composite of independent vegetation stands. *J. Geophys. Res.*, **97**, 2697–2715, <https://doi.org/10.1029/91JD01696>.
- , and —, 1994: The components of a 'SVAT' scheme and their effects on a GCM's hydrological cycle. *Adv. Water Resour.*, **17**, 61–78, [https://doi.org/10.1016/0309-1708\(94\)90024-8](https://doi.org/10.1016/0309-1708(94)90024-8).
- Krinner, G., and Coauthors, 2005: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochem. Cycles*, **19**, GB1015, <https://doi.org/10.1029/2003GB002199>.
- Long, C. N., and T. P. Ackerman, 2000: Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. *J. Geophys. Res.*, **105**, 15 609–15 626, <https://doi.org/10.1029/2000JD900077>.
- Manabe, S., 1969: Climate and the ocean circulation: I. The atmospheric circulation and the hydrology of the Earth's surface. *Mon. Wea. Rev.*, **97**, 739–774, [https://doi.org/10.1175/1520-0493\(1969\)097<0739:CATOC>2.3.CO;2](https://doi.org/10.1175/1520-0493(1969)097<0739:CATOC>2.3.CO;2).
- Maronga, B., and J. Reuder, 2017: On the formulation and universality of Monin–Obukhov similarity functions for mean gradients and standard deviations in the unstable surface layer: Results from surface-layer-resolving large-eddy simulations. *J. Atmos. Sci.*, **74**, 989–1010, <https://doi.org/10.1175/JAS-D-16-0186.1>.
- Masson, V., and Coauthors, 2013: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geosci. Model Dev.*, **6**, 929–960, <https://doi.org/10.5194/gmd-6-929-2013>.
- McNaughton, K., and P. Jarvis, 1983: Predicting effects of vegetation changes on transpiration and evaporation. Additional Woody Crop Plants, T. T. Kozlowski, Ed., Vol. **7**, *Water Deficits and Plant Growth*, Academic Press, 1–47.
- Medlyn, B. E., and Coauthors, 2011: Reconciling the optimal and empirical approaches to modelling stomatal conductance. *Global Change Biol.*, **17**, 2134–2144, <https://doi.org/10.1111/j.1365-2486.2010.02375.x>.
- Meier, R., E. L. Davin, S. C. Swenson, D. M. Lawrence, and J. Schwaab, 2019: Biomass heat storage dampens diurnal temperature variations in forests. *Environ. Res. Lett.*, **14**, 084026, <https://doi.org/10.1088/1748-9326/ab2b4e>.
- Mildrexler, D. J., M. Zhao, and S. W. Running, 2011: A global comparison between station air temperatures and MODIS land surface temperatures reveals the cooling role of forests. *J. Geophys. Res.*, **116**, G03025, <https://doi.org/10.1029/2010JG001486>.
- Miralles, D. G., P. Gentile, S. I. Seneviratne, and A. J. Teuling, 2019: Land–atmospheric feedbacks during droughts and heatwaves: State of the science and current challenges. *Ann. N. Y. Acad. Sci.*, **1436**, 19–35, <https://doi.org/10.1111/nyas.13912>.
- Misson, L., J. Tang, M. Xu, M. McKay, and A. Goldstein, 2005: Influences of recovery from clear-cut, climate variability, and thinning on the carbon balance of a young ponderosa pine plantation. *Agric. For. Meteorol.*, **130**, 207–222, <https://doi.org/10.1016/j.agrformet.2005.04.001>.
- Moderow, U., and Coauthors, 2009: Available energy and energy balance closure at four coniferous forest sites across Europe. *Theor. Appl. Climatol.*, **98**, 397–412, <https://doi.org/10.1007/s00704-009-0175-0>.
- Monin, A. S., and A. M. Obukhov, 1954: Basic laws of turbulent mixing in the atmosphere near the ground. *Tr. Akad. Nauk SSSR Geofiz. Inst.*, **24**, 163–187.
- Monteith, J. L., 1965: Evaporation and environment. *Symp. Soc. Exp. Biol.*, **19**, 205–234.
- Nearing, G. S., B. L. Ruddell, M. P. Clark, B. Nijssen, and C. Peters-Lidard, 2018: Benchmarking and process diagnostics of land models. *J. Hydrometeorol.*, **19**, 1835–1852, <https://doi.org/10.1175/JHM-D-17-0209.1>.
- Nelson, J. A., N. Carvalhais, M. Migliavacca, M. Reichstein, and M. Jung, 2018: Water-stress-induced breakdown of carbon–water relations: indicators from diurnal FLUXNET patterns. *Biogeosciences*, **15**, 2433–2447, <https://doi.org/10.5194/bg-15-2433-2018>.
- Ohmura, A., 2014: The development and the present status of energy balance climatology. *J. Meteor. Soc. Japan*, **92**, 245–285, <https://doi.org/10.2151/jmsj.2014-401>.
- Oke, T., 1987: *Boundary Layer Climates*. Routledge, 435 pp.
- Panwar, A., A. Kleidon, and M. Renner, 2019: Do surface and air temperatures contain similar imprints of evaporative conditions? *Geophys. Res. Lett.*, **46**, 3802–3809, <https://doi.org/10.1029/2019GL082248>.
- R Core Team, 2015: R: A language and environment for statistical computing. R Foundation for Statistical Computing, <https://www.R-project.org/>.
- Renner, M., 2019a: laubblatt/cleaRskyQuantileRegression: First public release of R package cleaRskyQuantileRegression, <https://doi.org/10.5281/zenodo.3380327>.
- , 2019b: Phaselag: R package to compute the phase lag between two variables. <https://doi.org/10.5281/zenodo.2540534>.
- , S. K. Hassler, T. Blume, M. Weiler, A. Hildebrandt, M. Guderle, S. J. Schymanski, and A. Kleidon, 2016: Dominant controls of transpiration along a hillslope transect inferred from ecohydrological measurements and thermodynamic limits. *Hydrol. Earth Syst. Sci.*, **20**, 2063–2083, <https://doi.org/10.5194/hess-20-2063-2016>.
- , M. Wild, M. Schwarz, and A. Kleidon, 2019a: Estimating shortwave clear-sky fluxes from hourly global radiation records by quantile regression. *Earth Space Sci.*, **6**, 1532–1546, <https://doi.org/10.1029/2019EA000686>.
- , and Coauthors, 2019b: Using phase lags to evaluate model biases in simulating the diurnal cycle of evapotranspiration: A case study in Luxembourg. *Hydrol. Earth Syst. Sci.*, **23**, 515–535, <https://doi.org/10.5194/hess-23-515-2019>.
- Roo, F. D., S. Zhang, S. Huq, and M. Mauder, 2018: A semi-empirical model of the energy balance closure in the surface layer. *PLOS ONE*, **13**, e0209022, <https://doi.org/10.1371/journal.pone.0209022>.
- Steenveld, G. J., T. Mauritsen, E. I. F. de Bruijn, J. Vilà-Guerau de Arellano, G. Svensson, and A. A. M. Holtlag, 2008: Evaluation

- of limited-Area models for the representation of the diurnal cycle and contrasting nights in CASES-99. *J. Appl. Meteor. Climatol.*, **47**, 869–887, <https://doi.org/10.1175/2007JAMC1702.1>.
- Stoy, P. C., and Coauthors, 2013: A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity. *Agric. For. Meteorol.*, **171–172**, 137–152, <https://doi.org/10.1016/j.agrformet.2012.11.004>.
- Taylor, P. C., 2012: Tropical outgoing longwave radiation and longwave cloud forcing diurnal cycles from CERES. *J. Atmos. Sci.*, **69**, 3652–3669, <https://doi.org/10.1175/JAS-D-12-088.1>.
- Todorov, V., and P. Filzmoser, 2009: An object-oriented framework for robust Multivariate analysis. *J. Stat. Software*, **32**, 1–47, <https://doi.org/10.18637/JSS.V032.I03>.
- Twine, T. E., and Coauthors, 2000: Correcting eddy-covariance flux underestimates over a grassland. *Agric. For. Meteorol.*, **103**, 279–300, [https://doi.org/10.1016/S0168-1923\(00\)00123-4](https://doi.org/10.1016/S0168-1923(00)00123-4).
- Ukkola, A. M., M. G. D. Kauwe, A. J. Pitman, M. J. Best, G. Abramowitz, V. Haverd, M. Decker, and N. Houghton, 2016: Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts. *Environ. Res. Lett.*, **11**, 104012, <https://doi.org/10.1088/1748-9326/11/10/104012>.
- Viterbo, P., and A. C. M. Beljaars, 1995: An improved land surface parameterization scheme in the ECMWF model and its validation. *J. Climate*, **8**, 2716–2748, [https://doi.org/10.1175/1520-0442\(1995\)008<2716:AILS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<2716:AILS>2.0.CO;2).
- Wang, Y. P., E. Kowalczyk, R. Leuning, G. Abramowitz, M. R. Raupach, B. Pak, E. Gorsel, and A. Luhar, 2011: Diagnosing errors in a land surface model (CABLE) in the time and frequency domains. *J. Geophys. Res. Biogeosci.*, **116**, G01034, <https://doi.org/10.1029/2010JG001385>.
- Wilson, K. B., and Coauthors, 2003: Diurnal centroid of ecosystem energy and carbon fluxes at FLUXNET sites. *J. Geophys. Res.*, **108**, 4664, <https://doi.org/10.1029/2001JD001349>.
- Zhou, C., and K. Wang, 2016: Evaluation of surface fluxes in ERA-Interim using flux tower data. *J. Climate*, **29**, 1573–1582, <https://doi.org/10.1175/JCLI-D-15-0523.1>.