The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally-gridded forcing data

Matthew F. McCabe[1]*, Ali Ershadi[1], Carlos Jimenez[2], Diego G. Miralles[3], Dominik Michel[4] and Eric F. Wood[5]

[1] {Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia}

[2] {Estellus, Paris, France}

[3] {Department of Earth Sciences, VU University Amsterdam, Amsterdam, The Netherlands}

[4] {Institute for Atmospheric and Climate Sciences, ETH Zurich, Zurich, Switzerland}

[5] {Department of Civil and Environmental Engineering, Princeton University, Princeton, United States of America}

[*] Correspondence to: matthew.mccabe@kaust.edu.sa

Abstract

Determining the spatial distribution and temporal development of evaporation at regional and global scales is required to improve our understanding of the coupled water and energy cycles and to better monitor any changes in observed trends and variability of linked hydrological processes. With recent international efforts guiding the development of long-term and globally distributed flux estimates, continued product assessments are required to inform upon the selection of suitable model structures and also to establish the appropriateness of these multi-model simulations for global application. In support of the objectives of the GEWEX LandFlux project, four commonly used evaporation models are evaluated against data from tower-based eddy-covariance observations, distributed across a range of biomes and climate zones. The selected schemes include the Surface Energy Balance System (SEBS) approach, the Priestley-Taylor Jet Propulsion Laboratory (PT-JPL) model, the Penman-Monteith based Mu model (PM-

Mu) and the Global Land Evaporation Amsterdam Model (GLEAM). Here we seek to examine the fidelity of global evaporation simulations by examining the multi-model response to varying sources of forcing data. To do this, we perform parallel and collocated model simulations using tower-based data together with a global-scale grid-based forcing product. Through quantifying the multi-model response to high-quality tower data, a better understanding of the subsequent model response to the coarse-scale globally gridded data that underlies the LandFlux product can be obtained, while also providing a relative evaluation and assessment of model performance.

Using surface flux observations from forty-five globally distributed eddy-covariance stations as independent metrics of performance, the tower-based analysis indicated that PT-JPL provided the highest overally statistical performance (*0.72; 61* $W.m^{-2}$*; 0.65*), followed closely by GLEAM (*0.68; 64* $W.m^{-2}$*; 0.62*), with values in parenthesis representing the R^2, *RMSD* and Nash-Sutcliffe Efficiency (*NSE*), respectively. PM-Mu (*0.51; 78* $W.m^{-2}$*; 0.45*) tended to underestimate fluxes, while SEBS (*0.72; 101* $W.m^{-2}$*; 0.24*) overestimated values relative to observations. A focused analysis across specific biome types and climate zones showed considerable variability in the performance of all models, with no single model consistently able to outperform any other. Results also indicated that the global gridded data tended to reduce the performance for all of the studied models when compared to the tower data, likely a response to scale mismatch and issues related to forcing quality. Rather than relying on any single model simulation, the spatial and temporal variability at both the tower- and grid-scale highlighted the potential benefits of developing an ensemble or blended evaporation product for global scale LandFlux applications. Challenges related to the robust assessment of the LandFlux product are also discussed.

1 Introduction

Characterizing the exchange of water between the land surface and the atmosphere is a topic of multi-disciplinary interest, as the processes that comprise this dynamic cycling of water determine the spatial and temporal variability of hydrological responses across local and global scales. In recent years, there has been significant progress in the development of regional and global datasets based largely on remote sensing retrievals. These data have provided a wealth of spatially and temporally varying information across a range of Earth system processes, including soil moisture (Liu et al., 2011a), vegetation change (Tucker et al., 2005; Liu et al., 2011b; Liu et al., 2013), groundwater (Famiglietti et al., 2011; Richey et al., 2015) and precipitation (Huffman et al., 1995; Nesbitt et al., 2004), enabling a capacity to enhance our understanding and description of regional- and global-scale water cycles and their spatial and temporal variability. While evaporation represents the key process returning the Earth's surface water to the overlying atmosphere and provides the linking mechanism between the water and energy cycles, it is only in relatively recent times that effort has been directed towards the development of global products (Mu et al., 2007; Fisher et al., 2008; Vinukollu et al., 2011a).

To address this observation limitation, a number of evaporation modelling approaches have been developed over the past few years to enable estimation at scales beyond the field, using satellite remote sensing (Sheffield et al., 2010; Miralles et al., 2011a) and other data sources (Douville et al., 2013). The models tend to differ in their level of empiricism and in the desired scale of application, with some exclusively developed for farm-scale operation and requiring local calibration (Bastiaanssen et al., 1998; Allen et al., 2007). Others have been developed for broader scale application and are built on physical relationships describing the water and energy transfer at the land surface (Norman et al., 1995; Su, 2002; Fisher et al., 2008; Miralles et al., 2011a). While traditional applications of evaporation estimates have been directed towards agricultural monitoring (Allen, 2000), catchment water budgets and basin-scale water management (Kustas et al., 1994; Granger, 2000), more recent applications of evaporation products have included detection and prediction of heatwaves (Hirschi et al., 2011; Miralles et al., 2014a), droughts (Mu et al., 2012; Otkin et al., 2014) and in resolving the likely contribution of human-induced climate change on such events (Greve et al., 2014).

Despite the importance of understanding the magnitude and spatial and temporal variability of evaporation, the availability of long-term products required to do this are rather limited. Characterizing the long-term trends and variability in independent observations of the Earth's coupled water and energy cycles is a key objective of the World Climate Research Programmes (WCRP) Global Energy and Water Cycle Exchanges (GEWEX) project. Towards this task, the

GEWEX Data and Assessments Panels (GDAP) LandFlux project has coordinated two interrelated research efforts that seek to: i) intercompare long-term gridded surface flux data sets and identify their skill and reliability (i.e. product-benchmarking), and ii) simulate and intercompare evaporation models to identify algorithms appropriate for developing a global flux product (i.e. model-benchmarking). In one of the first global-scale product assessments, Jiménez et al. (2011) examined twelve evaporation products obtained from satellite-based, reanalyses and off-line land surface model (LSM) simulations for a 3 year period (1993-1995), identifying large correlations between the products, similarity in their spatial distributions, as well as large absolute differences in the annual average evaporation. A complementary investigation of the inter-product differences was undertaken by Mueller et al. (2011), which included forty-one global evaporation data sets across a range of satellite-based simulations, LSMs, Global Circulation Models (GCMs), atmospheric reanalyses datasets, empirical up-scaling of eddy-covariance measurements, as well as atmospheric water budget data sets. In that study, Mueller et al. (2011) used seven years of monthly mean data for the period 1989-1995 and found strong similarity in the absolute magnitude and spatial distribution of evaporation amongst the products. More recently, Mueller et al. (2013) examined multi-annual trends and variations in evaporation products from a range of diagnostic data sets, LSMs and reanalysis products and showed consistency in inter-annual variations of evaporation products that corresponded well with previous investigations (Jung et al., 2010).

These benchmarking studies provided a thorough (and much needed) assessment of available global evaporation products and the varying approaches used to derive them. However, evaluation of the models for their predictive skill was challenging due to inconsistencies in the forcing data used to drive the models, as well as to the different parameterization schemes employed. That is, the analysis was performed on the published evaporation output, rather than re-running simulations from a common forcing dataset. In these benchmarking studies, the evaporation data sets were also aggregated to similar spatial and temporal resolutions for a common analysis period, to enable unbiased comparison. Uncertainties emerging from such aggregations can often reduce the confidence in any subsequent model performance ranking. One initial effort addressing this was the study of Vinukollu et al. (2011a), which used the

Surface Energy Balance System (SEBS) model (SEBS; Su, 2002), a two-source Penman-Monteith scheme by Mu et al. (2007) and a three-source model based on parameterizing the Priestley-Taylor model (PT-JPL) (Fisher et al., 2008) to estimate global evaporation for the period 2003-2004. The Vinukollu et al. (2011a) analysis revealed that the modelled instantaneous evaporation (coinciding with the time of satellite overpass) was in reasonable agreement with locally-observed evaporation at twelve eddy-covariance towers across the United States, with correlations ranging from 0.43 to 0.54. However, uncertainties resulting from scale mismatch between satellite data and the validation tower footprint reduced the confidence and skill ranking of the models. One of the unique aspects of the present study is that tower data are consistent across all model simulations: that is, tower-bias is minimized, by ensuring that all models are assessed against the same tower records. Further, even though sub-grid scale variability is not explored here (since none of the models explicitly account for this), the tower-to-grid scale analysis acts as a diagnostic of representativeness and point-to-pixel error.

Recently, Ershadi et al. (2014) examined a number of models including SEBS, PT-JPL, the Advection-Aridity model of Brutsaert and Stricker (1979) and a single-source Penman-Monteith (PM) model (Monteith, 1965), using a set of twenty flux towers distributed across a range of biome types and climate zones to force the models with tower-based data directly. Based on common forcing and considering overall results, the study found that PT-JPL was the best performing model, followed by SEBS, PM and Advection-Aridity. In a related contribution, Ershadi et al. (2015) provided a more focused analysis on the influence of model structure and resistance parameterization on single, two-layer and three-source Penman-Monteith models. The authors identified considerable variability in the performance of models due to their structure and parameterization choices. While establishing a baseline level of performance at the tower scale is important, understanding the impact of using the large-scale globally-gridded forcing that will ultimately drive the global products is key. Indeed, undertaking a parallel assessment between the tower and grid scales, while imposing consistency in the forcing data and sampling locations used, allows for a much greater understanding of model response than can be achieved through either assessment in isolation: an important extension upon recent tower-only analyses, such as Ershadi et al. (2014) and related contributions.

A parallel effort to the LandFlux project is the European Space Agency (ESA) funded WAter Cycle Multi-mission Observation Strategy for EvapoTranspiration (WACMOS-ET; see http://wacmoset.estellus.eu/). WACMOS-ET, which is focused on an analysis period covering 2005-2007, seeks to better understand the impacts of model structure on flux estimation, with an additional focus on developing a consistent forcing dataset using predominantly European Space Agency developed products. A key result from these early works and the preliminary outcomes from WACMOS-ET support the finding that no single model or parameterization consistently outperformed any other across different biomes. Further details on these complimentary efforts can be found in Michel et al. (2015) and Miralles et al. (2015).

The focus of the current investigation is to build upon these recent efforts as well as to complement ongoing WACMOS-ET investigations, by simulating state-of-the-art evaporation models using a joint assessment of tower-based meteorology and gridded data, and comparing results with available eddy-covariance flux observations. Understanding how application of gridded forcing data might influence the performance of the different models, relative to their performance when forced with (presumably) higher-quality tower data, is a motivating rationale for this work. Such evaluations are important as they offer insight into the sensitivity of the models to input data uncertainties, provide a relative assessment of model quality and also inform upon issues of spatial scale and footprint mismatch (McCabe and Wood, 2006). Establishing model suitability for large-scale operational application as part of the GEWEX Landflux project is a further motivating goal for this work. As such, a major objective is to evaluate the individual model responses across a large range of biomes and climate zones. The models selected for assessment include SEBS, PT-JPL, the Penman-Monteith based Mu model (PM-Mu) (Mu et al., 2011) as well as the Global Land Evaporation Amsterdam Methodology (GLEAM) (Miralles et al., 2011a). These models satisfy a number of criteria that were considered important for global model selection, including reliance on a minimum number of forcing variables, capacity to use remote sensing based observations, as well as previous application at either the regional or global scale.

2 Data and Methodology

2.1 Data

For this analysis, model simulations cover the period from 1997 to 2007 and are performed at a 3-hourly temporal resolution. To examine model response and inter-product variability, a parallel tower- and grid-based analysis was performed. Data for the tower-based analysis are derived from a set of forty-five eddy-covariance towers (see Table A1), while the gridded data are extracted from a compilation of available globally distributed satellite, meteorological and land surface characteristics products. Compared to the 0.5 degree and 3-hourly gridded data, the use of tower-based forcing is expected to minimize issues related to footprint uncertainties when evaluating simulations against the observed eddy-covariance based flux data. The primary purpose of the grid-based analysis is to better understand the effects of large-scale forcing data on the accuracy of global retrievals, relative to the tower-based evaluations.

2.1.1 Description of tower-based forcing data

Data for the tower-based analyses are derived from forty-five eddy-covariance towers selected from within the FLUXNET database (Baldocchi et al., 2001). *Table A1* lists the key attributes of the selected towers and Figure A1 describes the varying temporal lengths of the tower records used in this study. The requirement that towers only be used if they are able to provide the input data required by all models (see Table 1) was a strong limiting criterion that significantly reduced the number of available study sites. In particular, the availability of land surface temperature data, which is required for SEBS, drastically constrained the choice of towers. However, ensuring data consistency within the towers used for simulation and assessment was an important component of this work, as it removes the impact of tower bias in subsequent model assessment. Even with this reduced number, the selected towers represent a considerable spatial spread encompassing a variety of biome types and climate zones (see Figure 1).

In terms of forcing data requirements, tower-based variables that were used for model simulations include air temperature, relative humidity, wind speed, net radiation, ground heat flux and precipitation. A summary of the forcing data requirements for each model is provided in Table 1. Land surface emissivity, leaf area index and fractional vegetation cover were estimated from Normalized Difference Vegetation Index (NDVI) data obtained from the Global Inventory Monitoring and Modelling Study (GIMMS) dataset (Tucker et al., 2005), at 8 km spatial and bi-monthly temporal resolutions. Here, the emissivity was calculated using the approach of Sobrino et al. (2004), leaf area index was estimated following Fisher et al. (2008) and the fractional vegetation cover was estimated using the technique described in Jiménez-Muñoz et al. (2009). Land surface temperature was calculated using tower-observed longwave upward radiation and by inverting the Stefan-Boltzmann equation (Brutsaert, 2005). Atmospheric pressure data, which are absent from many towers, were calculated based on ground elevation of tower locations using an equation presented in Bos et al. (2008). Canopy height (h_c), which is needed for the SEBS model, was obtained from tower metadata and was assumed constant during the simulation period. Although h_c varies over many vegetation types, accounting for its within- and inter-annual variability is usually not possible, as observed data of h_c variations are rarely recorded. Tower data were aggregated (i.e. summed for precipitation and averaged for other input variables) from their native resolution of half-hourly or hourly to 3-hourly, to match the temporal resolution of the gridded data.

2.1.2 Description of grid-based forcing data (LandFlux Version 0 forcing dataset)

Grid-based data were developed by Princeton University for the LandFlux Version 0 (V-0) dataset. The variables in the V-0 include air temperature, land surface temperature, wind speed, atmospheric pressure, specific humidity, precipitation, net radiation, NDVI and leaf area index. Net radiation data derive from the GEWEX Surface Radiation Budget (SRB) Version-3 (Stackhouse et al., 2011), while land surface temperature is determined by employing a Bayesian post-processing procedure that merges High-Resolution Infrared Radiation Sounder (HIRS) retrievals with the land surface temperature data from the National Centers for Environment Prediction (NCEP) Climate Forecast System Reanalysis (CFSR) (Saha et al., 2010), as described in Coccia et al. (2015). Precipitation data are also from the NCEP CFSR product and have been bias-corrected to the Global Precipitation Climatology Project (GPCP) V2.2 dataset (Adler et al., 2003). Likewise, atmospheric pressure, specific humidity and wind speed data were extracted from the CFSR reanalysis data. For vegetation based parameters, NDVI data were prepared by aggregating 8-km resolution GIMMS NDVI data to 0.5° resolution, while leaf area index data were developed by Zhu et al. (2013) through fitting GIMMS NDVI data to the Moderate Resolution Imaging Spectroradiometer (MODIS) MOD15A2 NDVI product, using a neural network technique.

The majority of variables in the global LandFlux V-0 forcing dataset are at 0.5° spatial and 3-hourly temporal resolution. Exceptions include the net radiation (1° and 3-hourly), NDVI (0.5° and bi-monthly) and leaf area index (0.5° and monthly). For net radiation, the 1° data were linearly interpolated to a 0.5° resolution. The bi-monthly NDVI data were assumed constant for all 3-hourly time steps during each 15-day interval, while the leaf area index data were assumed constant during each month. The canopy height over shrubland and forest biomes was assumed fixed and was estimated using a static canopy height product developed by Simard et al. (2011). For grassland and cropland biomes, where the dynamics of canopy height can be considerable, canopy height was calculated using Equation 1, derived from Chen et al. (2012):

$$h_c = h_c^{min} + \frac{h_c^{max} - h_c^{min}}{NDVI_{max} - NDVI_{min}} \times (NDVI - NDVI_{min}) \quad (1)$$

where h_c^{min} and h_c^{max} are the minimum and maximum canopy height and were obtained from the static vegetation table of the North American Data Assimilation System (NLDAS) (available from http://ldas.gsfc.nasa.gov/nldas/web/web.veg.table.html). $NDVI_{min}$ and $NDVI_{max}$ are the minimum and maximum NDVI, respectively, and were calculated on a pixel-wise basis for each calendar year. The JPL static vegetation height was aggregated linearly from 1 km to 0.5°. Likewise, the NDVI derived canopy height was calculated at 8 km resolution and then aggregated to 0.5°. Similar to the tower-based data, the methodology of Jiménez-Muñoz et al. (2009) was used for the gridded forcing to estimate the fractional vegetation cover data from NDVI data. The ground heat flux at the grid-scale was calculated as a fraction of net radiation using fractional vegetation cover, following Su (2002).

2.1.3 Model specific forcing data and data sources

In addition to the data described above and shown in Table 1, both GLEAM and SEBS have some model specific forcing data requirements. For SEBS, information on land surface temperature, wind-speed and canopy height are required. At the tower-scale, these data are provided by available meteorological forcing or meta-data descriptions in the case of canopy height. At the grid-scale they are provided by a combination of the LandFlux V-O dataset and an adapted JPL static vegetation height, as described in Section 2.1.2. GLEAM based simulations require information on soil properties, vegetation optical depth (VOD), satellite soil moisture, snow water equivalent, lightning frequency and vegetation cover fraction. Soil properties data for GLEAM include field capacity, critical soil moisture and wilting point soil moisture thresholds. Data for these were obtained from the Global Gridded Surfaces of Selected Soil Characteristics dataset of the International Geosphere-Biosphere Programmes Data and Information System (IGBP-DIS), available from Oak Ridge National Laboratory Distributed Active Archive Center (http://www.daac.ornl.gov). Soil properties data were used in their native 5 arc-minute resolution for tower-based analysis, but were aggregated to 0.5° for grid-based assessment. Vegetation optical depth data was from Liu et al. (2011b) using a merged product from multiple microwave based satellite data. The 0.25° spatial and daily temporal resolutions VOD data were gap-filled as described by Miralles et al. (2011a). Soil moisture data assimilated in GLEAM comes from the CCI-WACMOS dataset (Liu et al., 2012) produced from both active and passive satellite microwave data at 0.25° and daily resolution. Snow water equivalent data are from the GlobSnow product version 1.0 (Luojus et al., 2010); as GlobSnow covers the northern hemisphere only, Global Monthly Snow Water Equivalent Climatology data from the National Snow and Ice Data Center (NSIDC) (Armstrong et al., 2005) are used for the BW-Ma1 tower (see Table A1) located in the southern hemisphere. Both GlobSnow data and the NSIDC product are at approximately 0.25° spatial and daily temporal resolutions. Lightning frequency data is based on the Combined Global Lightning Flash Rate Density monthly climatology at 0.5° (Mach et al., 2007) and it is used to calculate a climatology of rainfall rates (Miralles et al., 2010). Finally, vegetation cover fractions are derived from the MODIS MOD44B product (Hansen et al., 2005). The MODIS continuous cover factions describe every pixel as a combination of its fractions of water, tall canopy, short vegetation and bare soil. The temporal average of fractions is used here for the MODIS period, providing only a static cover fraction for the GLEAM simulations. The MOD44B product is available at 250 m and 0.25° resolution. For tower-based analysis, cover fractions are at 250 m resolution, but for grid-based analysis the 0.25° MOD44B product was aggregated to 0.5°.

Table 1 summarizes the different sources and spatio-temporal scales of the data that were used for both the tower- and grid-based flux simulations. As noted earlier, the temporal analysis encompasses the period 1997-2007, although as defined in Figure A1, the individual tower records do not necessarily provide uninterrupted observations during this time range.

2.1.4 Definition of selected biome type and climate zones

The specific biomes examined in this work include wetland (WET), grassland (GRA), cropland (CRO), shrubland (SHR), evergreen needleleaf forest (ENF), evergreen broadleaf forest (EBF) and deciduous broadleaf forest (DBF). Biome type was specified in Fluxnet metadata records for each of the individual tower sites and follows the International Geosphere-Biosphere Programme (IGBP) classification. For simplicity, the shrubland biome is comprised of closed shrubland, woody savannah and mixed forest biomes. The number of towers for each biome type varies, with fourteen for evergreen needleleaf forest, ten for grassland, seven for cropland, seven for deciduous broadleaf forest, four for shrubland, two for wetland and only one for evergreen broadleaf forest (see Table A1). The climate zones include boreal (BOR), sub-tropical (subTRO), temperate (TEMP), temperate-continental (TempCONT) and dry (DRY) for arid and semi arid regions. These zones were prescribed from the tower specific metadata, which were in turn derived from Rubel and Kottek (2010), based on a Köppen-Geiger climate classification. As with biome type, the towers are not evenly distributed across climate zones, with fifteen for temperate, eleven for sub-tropical, eight for temperate-continental, five for boreal and six for dry regions (see Table A1).

2.2 LandFlux Model Descriptions

Following are brief descriptions of the models employed in this analysis. For a more comprehensive explanation of the implementation of these different schemes, the reader is referred to the principal model references as well as the recent contributions of Ershadi et al. (2014) and Ershadi et al. (2015).

2.2.1 SEBS

SEBS is a widely employed process-based model used in the estimation of evaporation. The model uses a variety of land surface and atmospheric variables and parameters for simulating the transfer of heat and water vapor from the land surface to the atmosphere. To do so, the model first estimates the representative roughness of the land surface and then uses roughness parameters, temperature gradient and wind speed data to estimate sensible heat flux via a set of flux-gradient equations describing the transfer of heat from the land surface to the atmosphere. Depending on the atmospheric boundary layer height, the model uses either the Monin-Obukhov Similarity Theory or the Bulk Atmospheric Similarity Theory equations (Brutsaert, 2005). The model estimates the sensible heat flux of hypothetically wet and dry conditions and uses these extreme-cases to calculate the evaporative fraction. Evaporation is then calculated as a fraction of the available energy. The model requires accurate values of net radiation, land surface temperature, air temperature, humidity, wind speed and vegetation phenology to calculate surface fluxes. SEBS relaxes the need for parameterization of the surface resistance, but is sensitive to aerodynamic resistance parameterization (Ershadi et al., 2013). Further details on SEBS and its model formulation can be found in Su (2002).

2.2.2 PT-JPL

The PT-JPL model of evaporation uses a minimum of meteorological and remote sensing data and has been employed in a number of studies to estimate regional and global scales flux response (Fisher et al., 2008; Sahoo et al., 2011; Vinukollu et al., 2011b; Vinukollu et al., 2011a; Badgley et al., 2015). A key characteristic of the model is the use of bio-physiological properties of the land surface to reduce Priestley-Taylor potential evaporation to actual values. The PT-JPL is a three source model in which the total evaporation is partitioned into soil evaporation (λE_s), canopy transpiration (λE_t), and wet canopy evaporation (λE_i), i.e. $\lambda E = \lambda E_s + \lambda E_t + \lambda E_i$. The model first partitions the total net radiation to soil and vegetation components and calculates potential evaporation for soil, for canopy and for the wet canopy. The model then determines a set of constraint multipliers to represent the impacts of green canopy fraction, relative wetness of the canopy, air temperature, plant water stress and soil water stress on the evaporative process. The model uses the constraint multipliers to reduce the potential evaporation to actual values for each component of the system. PT-JPL does not calibrate or tune parameter values and does not use wind speed data or parameterizations of the aerodynamic and surface resistances. However, the model does require accurate estimates of optimum temperature (T_{opt}) (Potter et al., 1993) for canopy transpiration. The optimum temperature is the air temperature at the time of peak canopy activity, when the highest values of absorbed photosynthetically active radiation and minimum values of vapour pressure deficit occur. Further details of the PT-JPL model can be found in Fisher et al. (2008).

2.2.3 PM-Mu

The PM-Mu was expanded from a two-source Penman-Monteith implementation (Mu et al., 2007) to a three-source version (Mu et al., 2011), which forms the basis behind the near real-time estimation of global evaporation in the MOD16 product (Mu et al., 2013) (n.b. the PM-Mu nomenclature used herein reflects an identical description used in Michel et al. (2015) and Miralles et al. (2015), where it is referred to as PM-MOD). Evaporation in the PM-Mu model is the sum of soil evaporation, canopy transpiration and evaporation of the intercepted water in the canopy, i.e. ($\lambda E = \lambda E_s + \lambda E_t + \lambda E_i$). Estimation of evaporation for interception and transpiration components is based on the Penman-Monteith equation (Monteith, 1965). Actual soil evaporation is calculated using potential soil evaporation and a soil moisture constraint function from the Fisher et al. (2008) ET model. This function is based on the complementary hypothesis (Bouchet, 1963), which defines land-atmosphere interactions from air vapour pressure deficit and relative humidity. Evaporation components are weighted based on the fractional vegetation cover, relative surface wetness and available energy. Parameterization of aerodynamic and surface resistances for each source is based on extending biome specific conductance parameters from the stomata to the canopy scale, using vegetation phenology and meteorological data. In contrast to the majority of Penman-Monteith type of models, the PM-Mu does not require wind speed and soil moisture data for parameterization of resistances. However, global application of the model requires consideration of the fact that resistance parameters were calibrated against data from a set of eddy-covariance towers. One consideration that may influence model simulations is that this parameterization approach was developed at the daily-scale. However, both the present and also a recent related study (Miralles et al. 2015) suggest no obvious impact for sub-daily application. Further details on PM-Mu can be found in Mu et al. (2011) and Mu et al. (2013).

2.2.4 GLEAM

GLEAM (Miralles et al., 2011a) has been used not only in estimating global evaporation (Miralles et al., 2011b) but also in detection and evaluation of heatwaves (Miralles et al., 2014a), climate variability (Miralles et al., 2014b) and land-atmospheric feedbacks (Guillod et al., 2015). Designed as a satellite data based model, GLEAM first estimates interception loss using the analytical method of Gash (1979) and then applies the Priestley-Taylor equation to calculate potential evaporation for soil and vegetation. Like PT-JPL, the model constrains the potential evaporation values to actual values by applying a stress factor, although GLEAM is based on different assumptions and encompasses both moisture availability in a multi-layered soil system and vegetation water content inferred from vegetation optical depth data (Liu et al., 2011b). In contrast to SEBS, PT-JPL and PM-Mu, the GLEAM model is equipped with routines to quantify sublimation of snow-covered regions, to estimate open-water evaporation and to assimilate remote sensing soil moisture data. Routine application of GLEAM is usually performed in time-series mode, in which the model tracks the changes of soil moisture state across time steps. Here, to allow application of the model at the tower-scale, gaps in the tower data were filled by establishing correlation between the variables in tower- and grid-based data. Simulated evaporation values were filtered from the analysis for these gap-filled periods. Further details on GLEAM can be found in Miralles et al. (2011a;b).

2.3 Model Simulation and Analysis

The four selected models were forced with both tower- and grid-based data. The results were then filtered for daytime-only periods, defined as when the shortwave downward radiation exceeds 20 $W.m^{-2}$, to avoid issues associated with negative net radiation and night-time condensation. The data were also filtered for rain events, for negative sensible and latent heat flux observations, for low quality or gap-filled tower records, for frozen land surfaces and for times in which air temperature was less than or equal to 0 °C. The performance of the models was evaluated for individual towers, for the collection of data from all towers, for towers classified across biome types and for towers classified across climate zones.

To evaluate the skill of the models, we used traditional scatterplots and common statistical metrics including the coefficient of determination (R^2), slope (*m*) and y-intercept (*b*) of the linear regression, the root-mean-square difference (*RMSD*), relative error [*RE* = *RMSD*/mean(λE_{obs})] and the Nash-Sutcliffe Efficiency (*NSE*) (Nash and Sutcliffe, 1970). In developing these performance metrics, simulated evaporation was compared with tower-observed evaporation (λE_{obs}) that were corrected for non-closure using the energy residual technique, as described in Ershadi et al. (2014). Scatterplots of matching percentiles (referred to hereafter as percentile plots) of observed evaporation versus simulated values from the 1st to 99th percentile increment were also used (Section 3.1). The 25th percentile (Q_{25}), median (Q_{50}) and 75th percentile (Q_{75}) were used for further model assessment. To establish the response of the models to water availability at individual tower sites, we calculated an aridity index as $AI = P/E_p$, with P the annual precipitation ($mm.yr^{-1}$) and E_p the annual potential evaporation ($mm.yr^{-1}$), calculated using a Priestley-Taylor equation and assuming an alpha-coefficient of 1.26. LandFlux V-0 data (Section 2.1.2) at 3-hourly resolution were used to calculate aridity index values and an average value was calculated to represent the state of water availability at specific tower locations.

3 Results

3.1 Relative performance of the models when using tower-based and gridded data

Figure 2 and Figure 3 show scatterplots, percentile plots and relevant statistical metrics of the modelled evaporation for all of the available 3-hourly data records from across the forty-five towers (representing 115,148 records in total). For the tower-based analysis (see Figure 2), PT-JPL presents the best overall performance with lower model spread and an *RMSD* = 61 $W.m^{-2}$, *RE* = 0.41, R^2 = 0.71 and an *NSE* = 0.65. The model slightly underestimates evaporation, with a slope of linear regression equal to 0.91 and with the majority of the percentile plot (up to Q_{75}) located just under the 1:1 line. When considering results across all towers, GLEAM presents comparable statistical performance to PT-JPL, with an *RMSD* = 64 $W.m^{-2}$, *RE* = 0.43 and an *NSE* = 0.62. GLEAM tends to slightly underestimate evaporation, with the slope of linear regression equal to 0.84 and with the percentile plot being located under the 1:1 line. SEBS generally overestimates evaporation and has the lowest overall performance, with an *RMSD* = 101 $W.m^{-2}$, *RE* = 0.68 and *NSE* = 0.24, even though it has one of the highest R^2 values at 0.72. For PM-Mu, the model tends to underestimates evaporation, resulting in an *RMSD* = 78 $W.m^{-2}$, *RE* = 0.52 and an *NSE* = 0.45. Overall, the PT-JPL and GLEAM seem to present as more robust candidate models for estimation of evaporation, at least in terms of their statistical response at the tower scale. All models show a large spread around the fitted linear regression line. While the summary statistics are useful metrics of performance, the inter-tower variability of the models is an important element of this work and will be discussed further in the following sections.

The effect of using globally-gridded forcing data on the evaporation models is presented in Figure 3. Apart from providing a direct evaluation on the accuracy of the global LandFlux product, assessing flux response to a change in forcing aids in diagnosing the model sensitivity to data uncertainties (which are inherent in any data product). Likewise, an indirect assessment of the issue of footprint mismatch between the gridded data (0.5°) and the eddy-covariance tower (hundreds of meters) can also be inferred. Figure 3 clearly shows that use of the grid-based data reduces the performance of all models relative to the tower-based runs, with all statistics degrading with a change in forcing resolution. SEBS displayed the largest sensitivity to forcing data, with a 0.4 decrease in *NSE* and a 28 W.m^{-2} increase in *RMSD*. The sensitivity of PT-JPL and GLEAM to the use of gridded data was lower, with both showing an approximately 0.3 decrease in *NSE* and around 22 W.m^{-2} increase in *RMSD* when assessing the grid-based analysis. Overall, PM-Mu shows the lowest sensitivity to forcing, with a 0.26 decrease in *NSE* and 18 W.m^{-2} increase in RMSD, albeit presenting the lowest correlation and slope of linear regression for all model responses.

Overall, these results confirm that all models display a relatively high sensitivity to changes in the type and quality of input forcing data. While gridded forcing data are expected to have a mismatch with the tower-based forcing due to their larger pixel (and footprint) sizes, this spatial mismatch will impact all of the applied models, albeit to a lesser or greater extent, depending on forcing data requirements. While spatial scale no doubt plays a major role in decreasing model efficiencies at grid-scales, a key reason for the differences in tower- versus grid-based results relates to internal inconsistencies within the gridded forcing data. For instance, SEBS is known to be particularly sensitive to the temperature gradient between the land surface and the atmosphere (van der Kwast et al., 2009; Ershadi et al., 2013). While the temperature gradient at the tower scale is more reliable due to application of the tower-based sensors for air temperature and land surface temperature, obtaining such consistency is harder when different sources of forcing data are employed (see Section 2.1). Not surprisingly, results also indicate that those models that use fewer inputs show lower sensitivity to changes in the forcing. As such, any inconsistency between the tower and gridded data is likely to have less influence on the PT-JPL, GLEAM and PM-Mu models than it will on SEBS, which in addition to vegetation height, requires both land surface temperature and wind speed data: two variables with considerable spatial variability. Disentangling the varying influence of model structural and forcing data uncertainty requires focused attention and is examined further in the Discussion section.

The large spread of data in the scatterplots indicates that there is considerable variability in the performance of the models at individual towers, irrespective of whether tower or gridded data are used. Of course, it may also be indicative of systematic biases in the in-situ data, which vary from one tower to another and subsequently impact on model spread: however, this is nontrivial to determine. To investigate the nature of this variability, we extend the analysis by developing time series of R^2, *RE* and *NSE* at 3-hourly resolution for individual tower locations, as shown in Figure 4. To examine performance as a function of hydrological condition, the towers are arranged by degree of increasing aridity, as determined by calculation of an aridity index (see Section 2.3), with left-to-right representing the transition from wet-to-dry and describing an aridity index varying between approximately 2 and 0.

From Figure 4 it can be observed that there is a general downward trend in both R^2 and NSE as aridity increases, with a slight upward trend reflected in RE. In terms of R^2, most of the models (except for PM-Mu) show some consistency in performance until an aridity index of around 0.7, wherein models start to diverge. Similar agreement is seen in the relative error plot, although the outlier here is SEBS, which shows variable performance unrelated to aridity changes. Examining the Nash-Sutcliffe efficiency allows for a clearer evaluation of model response to be obtained. For this metric, PT-JPL and GLEAM display relatively good correspondence for most of the towers, but start to diverge more regularly for aridity indices below 0.8. Overall, PT-JPL presents a marginally better response than GLEAM, with higher values of *NSE* and R^2 and lowest values of *RE* produced across the majority of towers. Similar results are expressed in Figure A2, which presents the same tower based inter-comparison as in Figure 4, but for the grid-scale model simulations.

From Figure 2 it was observed that SEBS presented the lowest values of *NSE* and highest values of *RE*, while PM-Mu had the lowest values of R^2. Highlighting the importance of examining a range of statistical metrics, the R^2 values for SEBS are actually comparable to those of PT-JPL and GLEAM, or even higher for a majority of towers that have an aridity index less than 0.7. Inspection of individual tower-based scatterplots for each of the models (not shown) illustrated that while the SEBS evaporation has a strong linear relationship with observed values for a majority of towers, the linear regression line exhibits a large slope, indicating an overestimation in SEBS predictions. Those towers that exhibit drops in *NSE* (and rise in *RE*) for the SEBS model (e.g. DE-Tha, NL-Loo, US-Wrc, FR-Pue; see Table A1) are located mainly in shrubland and forest biomes, suggesting a dependency of SEBS model performance that is tied to land surface vegetation characteristics. Although statistical variations are evident in all models, the greater response variability in SEBS is likely due to problems in simulating heat transfer within the roughness sub-layer (RSL), which often forms over tall and heterogonous land surfaces (Harman, 2012). We explore the issue of skill dependency of certain models to biome type and climate zone in Sections 3.2 and 3.3.

As noted, Figure 4 shows a general decrease in the predictive skill in all models where towers have an aridity index less than 0.7, but particularly so for PM-Mu and SEBS. These reductions may in part be due to data uncertainties in tower observations that originate from the advection of dry air into the tower footprint, or to a reduced capacity of the models to reproduce the evaporative response when evaporation represents a small fraction of the total available energy. Two towers at which all models display poor performance are IT-Noe and IL-Yat (see Figure 1). It seems likely that IT-Noe is influenced by strong advection of moist air from the Mediterranean Sea, while IL-Yat is influenced by advection of hot and dry air from surrounding desert regions. None of the models in this study are able to specifically account for advection and are thus prone to misrepresenting the observed evaporative response.

3.2 Performance of the models across biomes

The variability in model performance across the tower sites observed in Figure 4 and Figure A2, indicates that a biome-specific assessment could be useful to determine whether the performance of the models is also correlated to the underlying land cover, in addition to any aridity influence. Figure 5 presents the R^2, *RE* and *NSE* for each of the models for the seven different biome classes. The analysis was conducted using the higher quality tower-based simulations for all available 3-hourly data. One immediate highlight from Figure 5 is the relatively poor performance of all models over shrubland sites, where low values of *NSE* (i.e. *NSE* $\leq$ 0.05) and reduced R^2 can be observed. Ershadi et al. (2014) observed a similarly poor response over shrublands in a separate tower-based analysis that employed some of the same models examined here. They attributed the result to difficulties in the parameterization of the models over such landscapes due to the strong heterogeneities present in these environments, as well as inherent water limitations. For instance, the capacity of the GIMMS NDVI data with 8 km spatial resolution is clearly insufficient in effectively parameterizing the roughness for SEBS, resistances for PM-Mu and constraint functions for the PT-JPL.

Excluding shrublands from the analysis, the PT-JPL is one of the best performing models across the remaining biomes, having the highest values of *NSE* and R^2 and lowest relative errors. Consistency in the performance of PT-JPL across biome types has been reported in earlier studies (Vinukollu et al., 2011a; Ershadi et al., 2014) and was variously ascribed to the formulation of its constraint functions (see Section 2.2.2) and the minimal forcing data requirements, which reduce its sensitivity to uncertainties in input data. GLEAM closely follows PT-JPL for evergreen needleleaf forest and grassland biomes, but shows marginally lower *NSE* values for other biomes. Figure 5 also indicates that while SEBS has relatively high values of R^2 over the majority of biome types, it fails to provide sufficient predictive skill for the estimation of evaporation over shrublands and forest biomes. These biome types are characterized by tall and heterogeneous canopies, within which the roughness sub-layer forms. The reduced capacity of the SEBS flux gradient functions in simulating heat transfer within the roughness sub-layer has been highlighted previously (Weligepolage et al., 2012; Ershadi et al., 2014). Although performing poorly in shrubland and forest biomes, the SEBS model exhibits a comparatively good performance across wetlands, grasslands and croplands, where shorter canopies dominate. PM-Mu presents the lowest values of R^2 across all biomes, although the model presents reasonable *NSE* values over cropland (0.64) and broadleaf forest (>0.54) biomes. Improved performance of the PM-Mu model over croplands has been observed in a recent study (Ershadi et al., 2015), but the key reasons for low R^2 values of the model across other biomes is not immediately apparent and requires further investigation.

Percentile plots of the 3-hourly tower-based results were used to identify whether a model under- or over-estimates evaporation across its distribution function. From Figure 6 it can be seen that SEBS clearly overestimates while PM-Mu underestimates evaporation across all biome types, reflecting those results presented in Figure 2. The percentile plots for SEBS are close to the 1:1 line for grassland and cropland biomes that have short canopy height, confirming the observations made for Figure 4 and Figure 5. PT-JPL shows good model reproduction of observed values over grassland and deciduous broadleaf forest biomes, with the percentile plots close to the 1:1 line. However, the model slightly underestimated evaporation for croplands and overestimated evaporation for wetlands, with the tails (percentiles greater than Q_{75}) reflecting greater divergence than the bulk of the distribution. The rate of overestimation was higher for evergreen needleleaf forest, evergreen broadleaf forest and for shrubland biomes. Figure 6 also shows that GLEAM presents strong performance over grasslands, croplands and evergreen needleleaf forest sites, underestimated evaporation across deciduous broadleaf forest sites and tended to overestimate evaporation across the remaining biomes (wetlands, shrublands and evergreen broadleaf forests).

Overall, all models show a tendency towards reduced performance when applied over forest biomes, but improved performance over shorter canopies. These results may be reflecting the fundamental physical basis behind approaches such as the base Penman-Monteith (Penman, 1948), Priestley-Taylor (Priestley and Taylor, 1972) and Monin-Obukhov flux gradient functions, which were developed for such surface types (Brutsaert, 1982), highlighting the challenges inherent in global scale application of such models, especially over diverse land cover types.

To further evaluate the influence of biome type on evaporation estimation and to discriminate the role of individual forcing variables in impacting model efficiencies, the *NSE* and R^2 values between tower- and grid-based data were calculated for the flux response, as well as for key forcing variables such as net radiation, land surface temperature, air temperature, wind speed, specific humidity, fractional vegetation cover and leaf area index. As can be seen in Figure 7, agreement between tower-based and grid-based net radiation data is relatively high across all biomes, but especially so over forest biomes ($NSE \geq 0.67$). Grid-based wind speed data have the most variable agreement with tower data, with R^2 and *NSE* values generally lower than other selected variables across all of the examined biomes. Air temperature shows good agreement, with both high *NSE* values ($NSE \geq 0.7$) and high R^2 values ($R^2 \geq 0.84$). Specific humidity data are also well reproduced ($NSE \geq 0.72$), as is land surface temperature with an $NSE \geq 0.80$ for all biomes. In sharing a common GIMMS-NDVI based derivation, the agreement for fractional vegetation cover and leaf area index data is reasonable over the majority of biomes, except over evergreen broadleaf forest, where both the R^2 and *NSE* are low.

The lower panel of Figure 7 show R^2 and NSE values for both the tower- and grid-based simulations against eddy-covariance observations for each of the models, discriminated by biome type. As can be seen, the performance of all models is reduced across all biomes when grid-based forcing data is used, a result reflected in all cases by relatively lower *NSE* and R^2 values. PM-Mu had the smallest and SEBS had the largest decrease in performance over a majority of the biomes, in accordance with the findings of Section 3.1. PT-JPL and PM-Mu had a relatively constant decrease in *NSE* and R^2 for the grid-based simulations. Decreased modelling performance was also maintained for GLEAM, except over the single evergreen broadleaf forest tower, where a more significant departure (relative to the other biome types), was observed. SEBS showed a much larger variability in performance reduction, with smaller variations due to forcing over forest biomes and larger reductions over biomes with shorter canopies. The significant decrease in *NSE* for SEBS over grassland, cropland and to some extent the wetland biome, cannot be immediately associated with *NSE* or R^2 changes in any of the forcing variables. It is interesting that the agreement over grassland and cropland biomes between tower- and grid-based variables is amongst the highest (especially for wind speed, fractional vegetation cover and for leaf area index data), yet the subsequent model performance is among the worst. The use of global statistics to evaluate model response makes discriminating the cause of this variability difficult. It is possible that the statistics are biased low due to the influence of one or a few individual towers, by errors in the forcing fields driving model parameterizations (i.e. vegetation height) or in response to model sensitivities to particular forcing variables. Either way, these results highlight the difficulties in diagnosing the cause of performance response and related sensitivity to forcing data variables in complex process-based models, which often display a high degree of interactions between the variables. Indeed, diagnosing the forcing variables responsible for reducing the efficiency of particular models is not feasible with a simple correlation analysis of the input data fields, but requires a separate and focused sensitivity analysis.

3.3 Performance of the Models over Climate Zones

Similar to the biome-wise analyses, an evaluation of the models was conducted across a number of distinct climate zones, with R^2, *RE* and *NSE* values for tower-based 3-hourly evaporation estimations shown in Figure 8. Yet again, the results highlight the importance of considering a range of evaluation metrics, as the models display some variability relative to the statistical measure being employed. Overall, both PT-JPL and GLEAM maintain a consistently good performance over the majority of climate zones, with PT-JPL expressing a slightly improved response over all zones except temperate, where GLEAM shows an improved simulation. In terms of R^2, PM-Mu presents the lowest values overall, while SEBS exhibits high values over the majority of climate zones, similar to the biome based analysis. However, SEBS generally fails to reproduce the observed evaporation response, with high *RE* and low *NSE*. All models have their best performance over the temperate-continental climate zone, with high *NSE* and R^2 and low *RE*, which was followed closely by the temperate climate zone. The lowest overall performance for all models corresponded to the dry climate zone, again reflecting the aridity based results in Figure 4. As discussed in Section 3.1, data uncertainties due to the role of advection in dry regions and difficulties in the accurate estimation when confronted with low evaporative fractions are likely reasons behind such performance reductions in dry regions.

Figure 9 displays the corresponding percentile plots of model performance over the five different climate zones. As can be seen, PT-JPL and GLEAM provide generally good performance over all climate zones, although GLEAM slightly underestimates evaporation for temperate-continental and boreal climate zones. SEBS overestimates relative to tower-based evaporation across all biomes, while PM-Mu generally underestimates, except over temperate and temperate-continental climate zones, for which the percentile plot of PM-Mu are relatively close to the 1:1 line.

Similar to Figure 7, Figure 10 outlines the model response differentiated for the different climate zones when using grid-based forcing data. As can be seen from the lower panel, the simulation performance is reduced across all climate zones, relative to the tower data. In particular, SEBS is significantly impacted across the majority of climate zones, with both a reduction in *NSE* and R^2, except over boreal forests. One possible reason for this smaller variation over boreal forests could be due to lower surface-to-air temperature gradients over forests, which contributes to smaller sensible heat fluxes and consequently larger evaporative fraction values (in contrast to model performance over dry climates, where the temperature gradient is large). Nevertheless, the relationship between uncertainty in individual variables and the reduction of modelling performances is not able to be determined here. Further analysis examining the sensitivity of individual models to their forcing is required.

4 Discussion

Understanding the role of model forcing in influencing simulation results, as well as examining the impacts of biome type and climate zone on flux response, are important elements in the development of robust globally-distributed evaporation products. The focus of this study was on evaluating a set of process-based models, to support the development of globally distributed and long term observations of surface fluxes as part of the GEWEX LandFlux project. Overall, the PT-JPL and GLEAM models provided the most consistent performance, while PM-Mu tended to underestimate and SEBS overestimate evaporation relative to the forty-five eddy-covariance tower observations examined here. However, while statistical analysis allows a pseudo-ranking of model performance, more detailed evaluation across towers, and biome and climate types highlighted the considerable within-model variability in performance. Results also demonstrated that changing the scale of input forcing data from tower- to grid-based reduced the quality of model estimates in all cases, but especially for SEBS, where a sensitivity to surface-air temperature gradients plays a strong role. In the following, we examine these results and interpret any implications for large-scale global applications.

With its relatively simple modelling structure, PT-JPL performed consistently well relative to the other models that have more complex structures and parameterization configurations. One possible reason for this response may relate to the constraint functions of PT-JPL serving a wide range of hydro-meteorological conditions, encompassing energy-limited (e.g. boreal climate) to water-limited (e.g. dry climate) conditions. The good performance of PT-JPL was also observed in a recent multi-model evaluation study, with a summary of the strengths and limitations of the model presented in Ershadi et al. (2014). GLEAM also performed well, both at the tower and at the grid-scale (see Figure 4 and Figure A2). Previous studies have shown that the model is sensitive to the accuracy of precipitation data (Miralles et al., 2011b), as this determines the partitioning of intercepted evaporation in the model and the root-zone soil moisture. Unfortunately, testing for such sensitivities was not possible here, as both tower- and grid-based records were filtered for rainfall events in post-processing steps, in response to the limitation of eddy-covariance observations during such events.

In terms of the *NSE*, R^2 and *RE*, PM-Mu followed PT-JPL and GLEAM, with the model tending to underestimate evaporation when applied to most of the tower- and grid-based records. While reasons for this underestimation are not immediately clear, a recent study examining the structure and parameterization of Penman-Monteith type models (Ershadi et al., 2015) showed that the PM-Mu, which has a three-source structure, underperformed relative to a single-source (Monteith, 1965) and a two-layer approach (Shuttleworth and Wallace, 1985) across all studied biome types except croplands. An interesting aspect of Ershadi et al. (2015) was that application of the canopy transpiration resistance scheme of the PM-Mu in those simpler models improved their prediction skills. As such, the reduced performance of the PM-Mu predictions might relate to underlying structural and parameterization issues in the model. As the operational model behind the generation of the current MOD16 global evaporation product (Mu et al., 2013), further studies to diagnose the cause of these responses are required.

Regarding assessment against the tower-based eddy-covariance observations, SEBS performed relatively poorly in most statistical metrics when compared to the other models, as it overestimated evaporation across a majority of studied biomes and climate zones, except over grasslands and cropland sites with short canopies (e.g. less than 3 m). Interestingly, even though generally over-predicting results, it had one of the highest R^2 values, indicating good correlation with the eddy-covariance observations. Findings from Ershadi et al. (2014) confirm the good performance of the model over short canopies and its lack of performance over shrublands and forests. In terms of performance against underlying biome type, it was observed that any performance reduction was observed mainly across shrublands and forest biomes, where the roughness sub-layer forms above the canopy (Harman, 2012). Importantly, the flux-gradient functions of the SEBS model are not parameterized to effectively simulate the heat transfer process in the roughness sub-layer, and hence the model fails to perform well (Weligepolage et al., 2012). The reliance of SEBS on an accurate representation of the surface-air temperature gradient also limits the effectiveness of the model for global application, demanding improvements in characterizing the spatial and temporal representativeness of such variables.

It is apparent from Sections 3.2 and 3.3 that the application of gridded data for modelling evaporation inevitably reduces the predictive performance of all models, regardless of their complexity in the evaporation process or their economy in forcing data requirements. In fact, the footprint mismatch between the tower- and grid-based simulations is likely to increase uncertainties in the forcing data and cause discrepancies between the simulated and tower-based evaporation values. Importantly, comparing the models for their relative performance (see Figure 7 and Figure 10) reveals that the performance decrease for grid-based analysis was not equal amongst all of the models. For instance, SEBS was observed to be more sensitive to the use of gridded forcing data, most likely as a result of inconsistencies in temperature gradient fields: an aspect that has been noted previously (van der Kwast et al., 2009; Ershadi et al., 2013). Although input uncertainty also impacts the performance of PT-JPL, PM-Mu and GLEAM, the *NSE* and R^2 of gridded simulations for those models are closer to their tower-based counterparts. Apart from indicating a robust model structure, the reduced impact seen in these schemes may also be a consequence of avoiding the use of forcing data such as land surface temperature and wind speed data, which are known to be uncertain at both the grid and tower-scale. Regardless of the culprit behind the observed performance discrepancy between tower and grid-based simulations, it is clear that some models are better suited to global application than others – at least given the quality of currently available global forcing datasets.

Importantly, the results presented in Sections 3.2 and 3.3 showed that evaluating tower or grid-based statistical responses alone is not enough to identify those forcing variables most impacting model performance. Diagnosing forcing sensitivity is not trivial given non-linearities in the models and the high level of interaction within model variables and parameters. Indeed, caution is warranted in any approaches seeking to evaluate evaporation models using gridded data in isolation, as this is likely to yield unreliable performance metrics of the models. It is important to perform a parallel tower-based data assessment to increase confidence in any single models performance (Su et al., 2005) in any evaluation approach, particularly those occurring at global scales.

Although the largest possible set of eddy-covariance towers and a common set of forcing data was used to evaluate the different model simulations, there are still inevitable limitations in the evaluations. Identifying such limitations is important not only for the current evaluations, but also in guiding future contributions. One such example relates to the period of tower data used for evaluation in this study (see Figure A1), as the data record length varies amongst the towers and the data are not uniformly distributed across seasons. Moreover, the towers are not evenly distributed across the studied biomes and climate zones (see Figure 1, Table A1), with only one tower covering the entire evergreen broadleaf forest biome and two towers covering the wetland biome. Further, no towers were available for use in arctic and tropical climate zones. Although the tropical climate zone, especially Amazonian forests, is accounted as a critical component in studies of the global water and energy cycles (Chahine, 1992; Wohl et al., 2012), relatively few towers in this zone provide land surface temperature and longwave upward radiation data needed for the SEBS model. An additional limitation is the coarse (8 km) spatial resolution of the GIMMS NDVI data used in the models for the tower-based analysis, as this resolution certainly does not correspond with the footprint of eddy-covariance sensors at any of the towers. Developments towards improving the availability and access to long-term high-resolution Landsat images (e.g. via Google Earth Engine; https://earthengine.google.org) might be one way to improve model forcing and evaluation exercises, especially with the development of high-resolution vegetation products (Houborg et al. 2016).

While the accuracy of individual variables in the LandFlux dataset were enhanced by bias correction against independent data sources (see Section 2.1), diagnosing the internal consistency of the data fields (McCabe et al., 2008), especially for air temperature, land surface temperature, wind speed and humidity, is a concept that has not received much attention to date and demands more considered investigations and analysis. Internal consistency is an extremely challenging objective, but is critically important for flux estimation, where so many different forcing data are required. Essentially it demands that all required model data are derived from a common set of forcing variables, rather than by the standard approach of compilation based on availability and accessibility. The most illustrative example would be in the development of radiation data, derived here from NASA-GEWEX SRB sources (Stackhouse et al., 2011). Calculation of radiation components requires air temperature, surface temperature, land surface and vegetation features, as well as numerous other elements. However, these underlying variables are rarely if ever retained to provide a consistent overall forcing data set (i.e. the meteorological variables used in producing the SRB data are not subsequently used to drive the models). Interdependencies in forcing affect many variables in the estimation of evaporation, yet products are not developed with this simple consistency principle in mind. Apart from introducing further biases and uncertainties into model simulations, until such consistency is attained, discriminating between the impact of forcing versus the model sensitivity to that forcing will remain extremely challenging.

From one perspective, the performance of the evaporation models examined here seems relatively poor, even when they are forced with high-quality tower-based data. PT-JPL, which was identified as one of the most consistent and best performing models, still presented a relative error of 41%, with errors for GLEAM, PM-Mu and SEBS of 43%, 52% and 72%, respectively. However, it is important to recognise that tower-based evaluation represents one of the strictest measure of model performance and comes with its own caveats. One question that remains unanswered is whether it is even appropriate to expect models run with large-scale gridded forcing to replicate the small-scale response observed by eddy-covariance towers. The alternative perspective, given inherent uncertainties in forcing, observations and specification of model parameters, is that these results are encouraging. Broader scale metrics such as hydrological consistency (McCabe et al., 2008), catchment based assessments or water budget closure approaches would provide a better guide (Sheffield et al., 2009) and indeed, such evaluations will need to be performed. These questions highlight the difficulties in not just producing global estimates, but perhaps more importantly, in evaluating their quality.

The observed variability of modelling performance across the studied biomes and climate zones implies that caution is required in advocating any single model for large-scale or global application. These results are consistent with previous findings undertaken across a smaller number of towers and biome and climate types, that any one modelling approach is incapable of accurately reflecting the range of flux responses occurring across diverse landscapes (Ershadi et al., 2014; Ershadi et al., 2015). One possible solution to address this inherent model limitation is to assemble a mosaicked product based on the predictive skill of the model(s) over particular biomes or climate zones. Another approach might be to develop an ensemble product using a suitable multi-model blending technique, such as a Bayesian Model Averaging approach (Hoeting et al., 1999; Yao et al., 2014). Either way, it is clear that further multi-model assessments are required for progressing global scale flux characterisation and to ensure a robust and representative product is developed.

5 Conclusions

It is something of a contradiction that the global-scale estimation of surface fluxes is both straightforward and extremely challenging at the same time. It is more straightforward than ever due to the availability of needed forcing data from various sources, such as numerical weather prediction or other operational products, as well as the increased development of global satellite based datasets. However, the comparative ease with which products can be developed belies the difficulties in actually developing robust and coherent simulations. Uncertainties in the use of internally inconsistent forcing data, the influence of untested model parameterizations over different land surface and climate types, violation of model assumptions in their graduation from the local scale to global scale and the perennial question on how to best evaluate model output all seek to confound global flux efforts.

The evaluation of four process-based evaporation models as part of the GEWEX LandFlux project undertaken here over a range of biome types and climate zones, highlighted the variable performance and verified the sentiment that no single model is able to consistently outperform any other. While individual model results at the tower scale allowed for a relative performance ranking, the overall model errors when considered globally were high. Of those models assessed here and being considered as potential candidates for a GEWEX LandFlux product, PT-JPL and GLEAM represent the most likely schemes for providing consistent simulation response over a range of biome and climate types. In a challenge for the development of more accurate global flux products, application of gridded data reduces the performance of all models, even if the overall performance ranking does not change between simulation runs. Such a response has obvious implications when model simulations at the continental and global scales are increasingly required in many applications and where not only the forcing data have large uncertainties, but also the underlying assumptions of the models themselves are likely to be questioned. Further investigations on the reasons for such variable performance and ways to offset the inherent uncertainties in global forcing are required. Additional research is also needed to improve the structure and parameterization of some of these candidate models, to understand model sensitivities to forcing (by conducting a thorough sensitivity analysis) and to develop and implement an appropriate ensemble modelling and merging technique that takes advantage of individual model performance over defined regions. Further detailed comparisons against estimates from more complex modelling systems, such as reanalysis and numerical weather prediction models, are needed to provide greater context and additional benchmarking metrics to guide future investigations.

Appendix A: Description of Tower Locations

Table A1: Selected eddy-covariance and their attributes. Further details and information on individual tower sites can be found via the Fluxnet data portal (http://fluxnet.fluxdata.org/)

Site-ID	Country	Lat.	Lon.	Ground Elev. (masl)	Tower height (m)	IGBP	Climate Class	Climate Zone	Reference
BW-Ma1	Botswana	-19.9	23.6	947	12.6	WSA	BSh	Dry	(Veenendaal et al., 2004)
CA-Ca1	Canada	49.9	-125.3	324	43	ENF	Cfb	Temperate	(Humphreys et al., 2006)
CA-Mer	Canada	45.4	-75.5	68	3	WET	Dfb	Temperate-Continental	(Kross et al., 2013)
CA-Oas	Canada	53.6	-106.2	594	39	DBF	Dfc	Boreal	(Fu et al., 2014)
CA-Obs	Canada	54.0	-105.1	593	25	ENF	Dfc	Boreal	(Fu et al., 2014)
CA-Ojp	Canada	53.9	-104.7	517	28	ENF	Dfc	Boreal	(Hilton et al., 2014)
CA-Qfo	Canada	49.7	-74.3	389	25	ENF	Dfc	Boreal	(Flanagan et al., 2012)
CN-Do2	China	31.6	121.9	4	5	WET	Cfa	Sub-Tropical	(Yan et al., 2008)
DE-Geb	Germany	51.1	10.9	159	6	CRO	Cfb	Temperate	(Smith et al., 2010)
DE-Hai	Germany	51.1	10.5	458	43.5	DBF	Cfb	Temperate	(Rebmann et al., 2005)
DE-Kli	Germany	50.9	13.5	480	3.5	CRO	Cfb	Temperate	(Smith et al., 2010)
DE-Meh	Germany	51.3	10.7	289	3	GRA	Cfb	Temperate	(Don et al., 2009)
DE-Tha	Germany	51.0	13.6	387	42	ENF	Cfb	Temperate	(Delpierre et al., 2009)
DE-Wet	Germany	50.5	11.5	789	27	ENF	Cfb	Temperate	(Richardson et al., 2010)
FR-LBr	France	44.7	-0.8	71	41	ENF	Cfb	Temperate	(Göckede et al., 2008)
FR-Lam	France	43.5	1.2	182	3.65	CRO	Cfb	Temperate	(Merlin et al., 2011)
FR-Pue	France	43.7	3.6	271	13	EBF	Csa	Sub-Tropical	(Soudani et al., 2014)
IL-Yat	Israel	31.3	35.1	654	18	ENF	BSh	Dry	(Sprintsin et al., 2011)
IT-BCi	Italy	40.5	15.0	9	2	CRO	Csa	Sub-Tropical	(Reichstein et al., 2003)
IT-Col	Italy	41.8	13.6	1534	25	DBF	Cfa	Sub-Tropical	(Chiti et al., 2010)
IT-Lav	Italy	46.0	11.3	1367	33	ENF	Cfb	Temperate	(Stoy et al., 2013)
IT-MBo	Italy	46.0	11.0	1563	2.5	GRA	Cfb	Temperate	(Gamon et al., 2010)
IT-Noe	Italy	40.6	8.2	27	3.6	CSH	Csa	Sub-Tropical	(Carvalhais et al., 2010)
IT-Ro1	Italy	42.4	11.9	174	20	DBF	Csa	Sub-Tropical	(Chiti et al., 2010)
JP-Tom	Japan	42.7	141.5	133	42	MF	Dfb	Temperate-Continental	(Saigusa et al., 2010)
NL-Ca1	Netherlands	52.0	4.9	-1	5	GRA	Cfb	Temperate	(Gioli et al., 2004)
NL-Loo	Netherlands	52.2	5.7	34	27	ENF	Cfb	Temperate	(Sulkava et al., 2011)
PT-Mi2	Portugal	38.5	-8.0	191	2.5	GRA	Csa	Sub-Tropical	(Gilmanov et al., 2007)
RU-Fyo	Russia	56.5	32.9	274	29	ENF	Dfb	Temperate-Continental	(Smith et al., 2010)
SE-Nor	Sweden	60.1	17.5	35	103	ENF	Dfb	Temperate-Continental	(Zierl et al., 2007)

Site-ID	Country	Lat.	Lon.	Ground Elev. (masl)	Tower height (m)	IGBP	Climate Class	Climate Zone	Reference
US-ARM	USA	36.6	-97.5	318	60	CRO	Cfa	Sub-Tropical	(Lokupitiya et al., 2009)
US-Aud	USA	31.6	-110.5	1474	4	GRA	BSk	Dry	(Horn and Schulz, 2011)
US-Bkg	USA	44.3	-96.8	496	4	GRA	Dfa	Temperate-Continental	(Hollinger et al., 2010)
US-Bo1	USA	40.0	-88.3	218	10	CRO	Dfa	Temperate-Continental	(Hollinger et al., 2010)
US-Bo2	USA	40.0	-88.3	220	10	CRO	Dfa	Temperate-Continental	(Hollinger et al., 2010)
US-CaV	USA	39.1	-79.4	993	4	GRA	Cfb	Temperate	(Hollinger et al., 2010)
US-FPe	USA	48.3	-105.1	632	3.5	GRA	BSk	Dry	(Horn and Schulz, 2011)
US-Goo	USA	34.3	-89.9	94	4	GRA	Cfa	Sub-Tropical	(Hollinger et al., 2010)
US-MMS	USA	39.3	-86.4	290	48	DBF	Cfa	Sub-Tropical	(Dragoni et al., 2011)
US-MOz	USA	38.7	-92.2	238	30	DBF	Cfa	Sub-Tropical	(Hollinger et al., 2010)
US-NR1	USA	40.0	-105.5	3053	26	ENF	Dfc	Boreal	(Hilton et al., 2014)
US-SRM	USA	31.8	-110.9	1120	6.4	WSA	BSk	Dry	(Cavanaugh et al., 2011)
US-WCr	USA	45.8	-90.1	524	30	DBF	Dfb	Temperate-Continental	(Curtis et al., 2002)
US-Wkg	USA	31.7	-109.9	1522	6.4	GRA	BSk	Dry	(Scott, 2010)
US-Wrc	USA	45.8	-122.0	391	85	ENF	Csb	Temperate	(Wharton et al., 2009)

Figure A1: Temporal duration of the eddy-covariance based flux and tower meteorological observations for each of the 45 sites used in this study

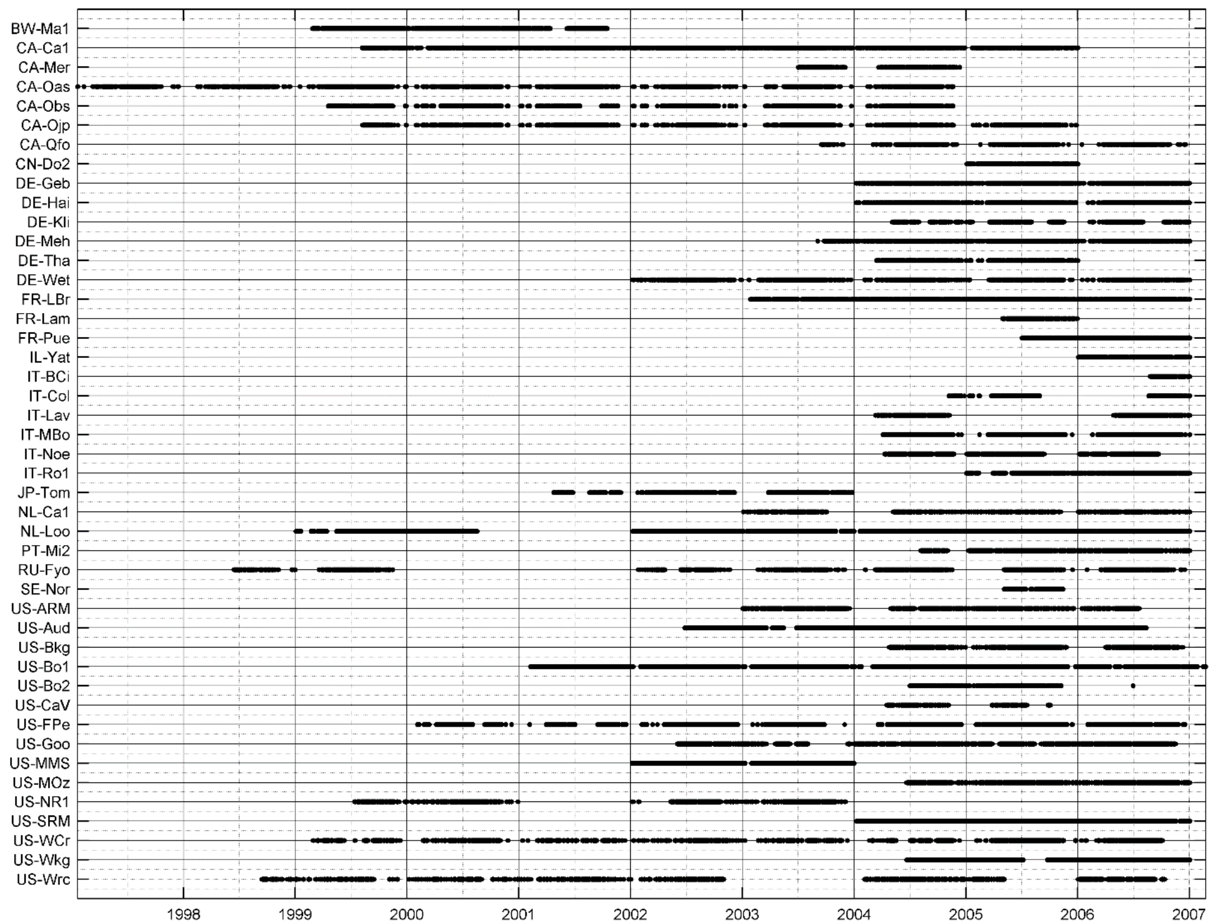
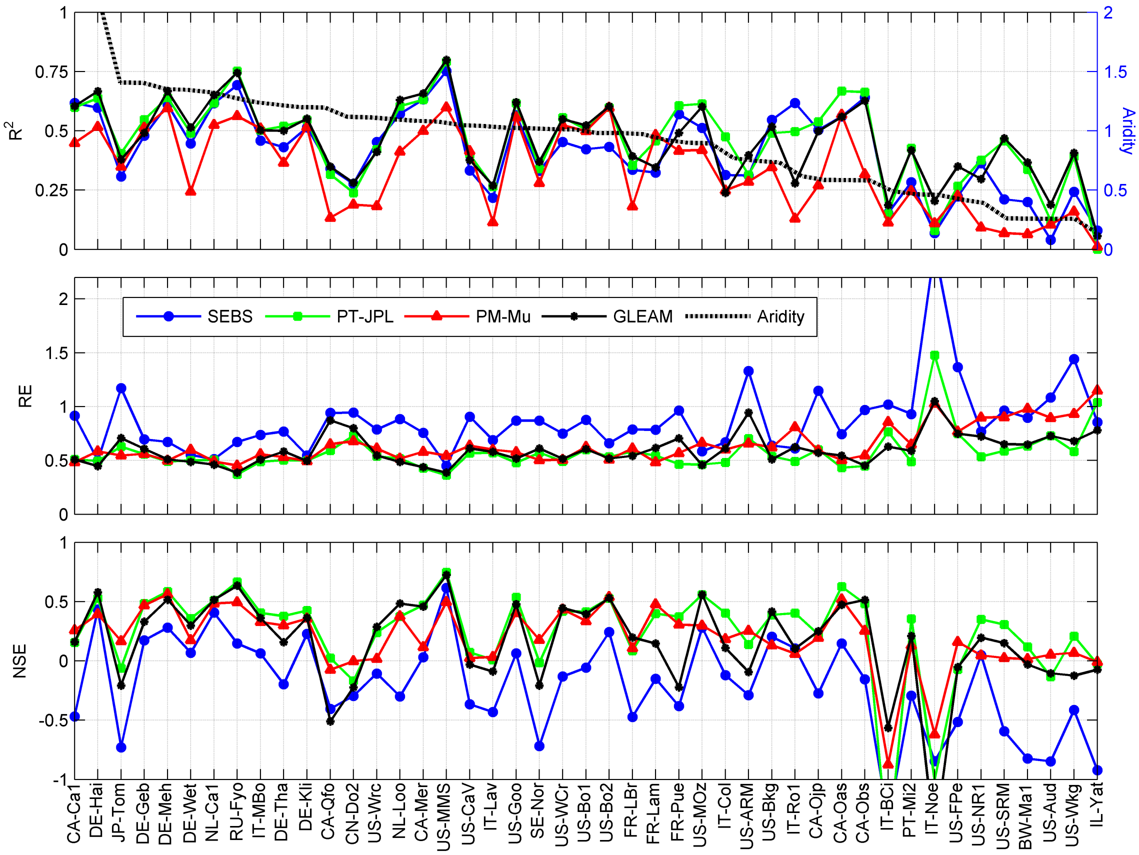

BW-Ma1
CA-Ca1
CA-Mer
CA-Oas
CA-Obs
CA-Ojp
CA-Qfo
CN-Do2
DE-Geb
DE-Hai
DE-Kli
DE-Meh
DE-Tha
DE-Wet
FR-LBr
FR-Lam
FR-Pue
IL-Yat
IT-BCi
IT-Col
IT-Lav
IT-MBo
IT-Noe
IT-Ro1
JP-Tom
NL-Ca1
NL-Loo
PT-Mi2
RU-Fyo
SE-Nor
US-ARM
US-Aud
US-Bkg
US-Bo1
US-Bo2
US-CaV
US-FPe
US-Goo
US-MMS
US-MOz
US-NR1
US-SRM
US-WCr
US-Wkg
US-Wrc
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007

Figure A2: Comparison of the performance skill of the models in reproducing evaporation for the grid-based analyses. R^2 is the coefficient of determination, RE is relative error (lower is better) and NSE is the Nash-Sutcliffe Efficiency (higher is better). Towers are arranged from left to right based on an aridity index (secondary y-axis).


R²
Aridity
RE
NSE
SEBS
PT-JPL
PM-Mu
GLEAM
Aridity
CA-Ca1 DE-Hai JP-Tom DE-Geb DE-Meh DE-Wet NL-Ca1 RU-Fyo IT-MBo DE-Tha DE-Kli CA-Qfo CN-Do2 US-Wrc NL-Loo CA-Mer US-MMS US-CaV IT-Lav US-Goo SE-Nor US-WCr US-Bo1 US-Bo2 FR-LBr FR-Lam FR-Pue US-MOz IT-Col US-ARM US-Bkg IT-Ro1 CA-Ojp CA-Oas CA-Obs IT-BCi PT-Mi2 IT-Noe US-FPe US-NR1 US-SRM BW-Ma1 US-Aud US-Wkg IL-Yat

Code Availability

The PM-Mu, SEBS and PT-JPL models were coded in MATLAB as part of the GEWEX LandFlux and WACMOS-ET projects, in discussion with (but independent of) the principal model authors, as referenced in the relevant publications. The GLEAM model was developed in MATLAB by Diego Miralles and Brecht Martens. All model code can be made available upon an emailed request to hydrology@kaust.edu.sa, including a brief description of the intended purpose and application.

Data Availability

Evaporation model output presented here for both the gridded and tower based analyses can be provided upon an emailed request to hydrology@kaust.edu.sa. The request should include a brief description of the intended purpose and application of the model data. Further details can be found at http://hydrology.kaust.edu.sa/landflux

Acknowledgements

Research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST). D.G.M. acknowledges the financial support from The Netherlands Organization for Scientific Research through grant 863.14.004. We appreciate the support of the ESA funded WACMOS-ET project for both fruitful scientific discussions and guidance in ensuring complementarity of these joint efforts. We thank the FLUXNET site investigators for allowing the use of their meteorological data. This work used eddy-covariance data acquired by the FLUXNET community and in particular by the AmeriFlux program (U.S. Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program: DE-FG02-04ER63917 and DE-FG02-04ER63911), AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCan), GreenGrass, KoFlux, LBA, NECC, TCOS-Siberia, USCCC. We acknowledge the financial support to the eddy-covariance data harmonization provided by CarboEuropeIP, FAO-GTOS-TCO, iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia, Université Laval and Environment Canada and US Department of Energy and the database development and technical support from Berkeley Water Centre, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California - Berkeley, University of Virginia.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present), Journal of Hydrometeorology, 4, 1147-1167, 2003.

Allen, R. G.: Using the FAO-56 dual crop coefficient method over an irrigated region as part of an evapotranspiration intercomparison study, Journal of Hydrology, 229, 27-41, 2000.

Allen, R. G., Tasumi, M., and Trezza, R.: Satellite-Based Energy Balance for Mapping Evapotranspiration with Internalized Calibration (METRIC)-Model, Journal of Irrigation and Drainage Engineering, 133, 380-394, 2007.

Armstrong, R. L., Brodzik, M. J., Knowles, K., and Savoie, M.: Global monthly EASE-Grid snow water equivalent climatology, Boulder, CO: National Snow and Ice Data Center, Digital media, 2005. 2005.

Badgley, G., Fisher, J. B., Jiménez, C., Tu, K. P., and Vinukollu, R.: On uncertainty in global terrestrial evapotranspiration estimates from choice of input forcing datasets, Journal of Hydrometeorology, doi: 10.1175/JHM-D-14-0040.1, 2015. 2015.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem–Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities, Bulletin of the American Meteorological Society, 82, 2415-2434, 2001.

Bastiaanssen, W. G. M., Menenti, M., Feddes, R. A., and Holtslag, A. A. M.: A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation, Journal of Hydrology, 212-213, 198-212, 1998.

Bos, M. G., Kselik, R. A. L., Allen, R. G., and Molden, D. J.: Water Requirements for Irrigation and the Environment, Springer, Dordrecht, 2008.

Bouchet, R. J.: Evapotranspiration réelle et potentielle, signification climatique. General Assembly Berkeley, International Association for Hydrological Sciences. Gentbrugge, Belgium. **62:** 134-142, 1963.

Brutsaert, W.: Evaporation Into the Atmosphere : theory, history, and applications, Reidel Publishing, Dordrecht etc., 1982.

Brutsaert, W.: Hydrology : An Introduction, Cambridge University Press, Cambridge, 2005.

Brutsaert, W. and Stricker, H.: An advection-aridity approach to estimate actual regional evapotranspiration, Water Resour. Res., 15, 443-450, 1979.

Carvalhais, N., Reichstein, M., Collatz, G. J., Mahecha, M. D., Migliavacca, M., Neigh, C. S. R., Tomelleri, E., Benali, A. A., Papale, D., and Seixas, J.: Deciphering the components of regional net ecosystem fluxes following a bottom-up approach for the Iberian Peninsula, Biogeosciences, 7, 3707-3729, 2010.

Cavanaugh, M. L., Kurc, S. A., and Scott, R. L.: Evapotranspiration partitioning in semiarid shrubland ecosystems: a two ‐ site evaluation of soil moisture control on transpiration, Ecohydrology, 4, 671-681, 2011.

Chahine, M. T.: The hydrological cycle and its influence on climate, Nature, 359, 373-380, 1992.

Chen, X., Su, Z., Ma, Y., Yang, K., Wen, J., and Zhang, Y.: An Improvement of Roughness Height Parameterization of the Surface Energy Balance System (SEBS) over the Tibetan Plateau, Journal of Applied Meteorology and Climatology, 52, 607-622, 2012.

Chiti, T., Papale, D., Smith, P., Dalmonech, D., Matteucci, G., Yeluripati, J., Rodeghiero, M., and Valentini, R.: Predicting changes in soil organic carbon in mediterranean and alpine forests during the Kyoto Protocol commitment periods using the CENTURY model, Soil use and management, 26, 475-484, 2010.

Coccia, G., Siemann, A., Pan, M., and Wood, E. F.: Creating consistent datasets by combining remotely-sensed data and land surface model estimates through Bayesian uncertainty post-processing: the case of Land Surface Temperature from HIRS, Remote Sensing of Environment, 170, 290-305, doi:10.1016/j.rse.2015.09.010, 2015.

Curtis, P. S., Hanson, P. J., Bolstad, P., Barford, C., Randolph, J. C., Schmid, H. P., and Wilson, K. B.: Biometric and eddy-covariance based estimates of annual carbon storage in five eastern North American deciduous forests, Agricultural and Forest Meteorology, 113, 3-19, 2002.

Delpierre, N., Soudani, K., Francois, C., Köstner, B., Pontailler, J. Y., Nikinmaa, E., Misson, L., Aubinet, M., Bernhofer, C., and Granier, A.: Exceptional carbon uptake in European forests during the warm spring of 2007: a data–model analysis, Global Change Biology, 15, 1455-1474, 2009.

Don, A., Rebmann, C., Kolle, O., Scherer ‐ Lorenzen, M., and Schulze, E. D.: Impact of afforestation ‐ associated management changes on the carbon balance of grassland, Global Change Biology, 15, 1990-2002, 2009.

Douville, H., Ribes, A., Decharme, B., Alkama, R., and Sheffield, J.: Anthropogenic influence on multidecadal changes in reconstructed global evapotranspiration, Nature Clim. Change, 3, 59-62, 2013.

Dragoni, D., Schmid, H. P., Wayson, C. A., Potter, H., Grimmond, C. S. B., and Randolph, J. C.: Evidence of increased net ecosystem productivity associated with a longer vegetated season in a deciduous forest in south ‐ central Indiana, USA, Global Change Biology, 17, 886-897, 2011.

Ershadi, A., McCabe, M. F., Evans, J. P., Chaney, N. W., and Wood, E. F.: Multi-site evaluation of terrestrial evaporation models using FLUXNET data, Agricultural and Forest Meteorology, 187, 46-61, 2014.

Ershadi, A., McCabe, M. F., Evans, J. P., Mariethoz, G., and Kavetski, D.: A Bayesian analysis of sensible heat flux estimation: Quantifying uncertainty in meteorological forcing to improve model prediction, Water Resources Research, 49, 2343-2358, 2013.

Ershadi, A., McCabe, M. F., Evans, J. P., and Wood, E. F.: Impact of model structure and parameterization on Penman–Monteith type evaporation models, Journal of Hydrology, 525, 521-535, 2015.

Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K. J., Syed, T. H., Swenson, S. C., de Linage, C. R., and Rodell, M.: Satellites measure recent rates of groundwater depletion in California's Central Valley, Geophysical Research Letters, 38(3), 10.1029/2010GL046442, 2011.

Fisher, J. B., Tu, K. P., and Baldocchi, D. D.: Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites, Remote Sensing of Environment, 112, 901-919, 2008.

Flanagan, L. B., Cai, T., Black, T. A., Barr, A. G., McCaughey, J. H., and Margolis, H. A.: Measuring and modeling ecosystem photosynthesis and the carbon isotope composition of ecosystem-respired CO_2 in three boreal coniferous forests, Agricultural and Forest Meteorology, 153, 165-176, 2012.

Fu, D., Chen, B., Zhang, H., Wang, J., Black, T. A., Amiro, B. D., Bohrer, G., Bolstad, P., Coulter, R., and Rahman, A. F.: Estimating landscape net ecosystem exchange at high spatial–temporal resolution based on Landsat data, an improved upscaling model framework, and eddy covariance flux measurements, Remote Sensing of Environment, 141, 90-104, 2014.

Gamon, J. A., Coburn, C., Flanagan, L. B., Huemmrich, K. F., Kiddle, C., Sanchez-Azofeifa, G. A., Thayer, D. R., Vescovo, L., Gianelle, D., and Sims, D. A.: SpecNet revisited: bridging flux and remote sensing communities, Can. J. Remote Sens., 36, S376-S390, 2010.

Gash, J. H.: An analytical model of rainfall interception by forests quarterly, Journal of Royal Meteorological So 鄄 ciety, 105, 43-45, 1979.

Gilmanov, T., Soussana, J., Aires, L., Allard, V., Ammann, C., Balzarolo, M., Barcza, Z., Bernhofer, C., Campbell, C., Cernusca, A., Cescatti, A., Clifton-Brown, J., Dirks, B., Dore, S., Eugster, W., Fuhrer, J., Gimeno, C., Gruenwald, T., Haszpra, L., Hensen, A., Ibrom, A., Jacobs, A., Jones, M., Lanigan, G., Laurila, T., Lohila, A., Manca, G., Marcolla, B., Nagy, Z., Pilegaard, K., Pinter, K., Pio, C., Raschi, A., Rogiers, N., Sanz, M., Stefani, P., Sutton, M., Tuba, Z., Valentini, R., Williams, M., and Wohlfahrt, G.: Partitioning European grassland net ecosystem CO2 exchange into gross primary productivity and ecosystem respiration using light response function analysis, Agriculture, Ecosystems and Environment, 121, 93 - 120, 2007.

Gioli, B., Miglietta, F., De Martino, B., Hutjes, R. W. A., Dolman, H. A. J., Lindroth, A., Schumacher, M., Sanz, M. J., Manca, G., and Peressotti, A.: Comparison between tower and aircraft-based eddy covariance fluxes in five European regions, Agricultural and Forest Meteorology, 127, 1-16, 2004.

Göckede, M., Foken, T., Aubinet, M., Aurela, M., Banza, J., Bernhofer, C., Bonnefond, J.-M., Brunet, Y., Carrara, A., and Clement, R.: Quality control of CarboEurope flux data–Part 1: Coupling footprint analyses with flux data quality assessment to evaluate sites in forest ecosystems, Biogeosciences, 5, 433-450, 2008.

Granger, R. J.: Satellite-derived estimates of evapotranspiration in the Gediz basin, Journal of Hydrology, 229, 70-76, 2000.

Greve, P., Orlowsky, B., Mueller, B., Sheffield, J., Reichstein, M., and Seneviratne, S. I.: Global assessment of trends in wetting and drying over land, Nature geoscience, 7, 716-721, 2014.

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

Hansen, M. C., Townshend, J. R. G., DeFries, R. S., and Carroll, M.: Estimation of tree cover using MODIS data at global, continental and regional/local scales, Int. J. Remote Sens., 26, 4359-4380, 2005.

Harman, I.: The Role of Roughness Sublayer Dynamics Within Surface Exchange Schemes, Boundary-Layer Meteorology, 142, 1-20, 2012.

Hilton, T. W., Davis, K. J., and Keller, K.: Evaluating terrestrial CO2 flux diagnoses and uncertainties from a simple land surface model and its residuals, Biogeosciences, 11, 217-235, 2014.

Hirschi, M., Seneviratne, S. I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., Formayer, H., Orlowsky, B., and Stepanek, P.: Observational evidence for soil-moisture impact on hot extremes in southeastern Europe, Nature Geoscience, 4, 17-21, 2011.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T.: Bayesian Model Averaging: A Tutorial, Statistical Science, 14, 382-401, 1999.

Hollinger, D. Y., Ollinger, S. V., Richardson, A. D., Meyers, T. P., Dail, D. B., Martin, M. E., Scott, N. A., Arkebauer, T. J., Baldocchi, D. D., and Clark, K. L.: Albedo estimates for land surface models and support for a new paradigm based on foliage nitrogen concentration, Global Change Biology, 16, 696-710, 2010.

Horn, J. E. and Schulz, K.: Identification of a general light use efficiency model for gross primary production, Biogeosciences, 8, 999 - 1021, 2011.

Houborg, R., McCabe, M. F., and Gao, F.: A Spatio-Temporal Enhancement Method for medium resolution LAI (STEM-LAI), International Journal of Applied Earth Observation and Geoinformation, 47, 15-29, 2016.

Huffman, G. J., Adler, R. F., Rudolph, B., Schneider, U., and Keehn, P.: Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information, J. Climate, 8, 1284-1295, 1995.

Humphreys, E. R., Black, T. A., Morgenstern, K., Cai, T., Drewitt, G. B., Nesic, Z., and Trofymow, J. A.: Carbon dioxide fluxes in coastal Douglas-fir stands at different stages of development after clearcut harvesting, Agricultural and Forest Meteorology, 140, 6-22, 2006.

Jiménez, C., Prigent, C., Mueller, B., Seneviratne, S. I., McCabe, M. F., Wood, E. F., Rossow, W. B., Balsamo, G., Betts, A. K., Dirmeyer, P. A., Fisher, J. B., Jung, M., Kanamitsu, M., Reichle, R. H., Reichstein, M., Rodell, M., Sheffield, J., Tu, K., and Wang, K.: Global intercomparison of 12 land surface heat flux estimates, J. Geophys. Res., 116, D02102, 2011.

Jiménez-Muñoz, J., Sobrino, J., Plaza, A., Guanter, L., Moreno, J., and Martinez, P.: Comparison Between Fractional Vegetation Cover Retrievals from Vegetation Indices and Spectral Mixture Analysis: Case Study of PROBA/CHRIS Data Over an Agricultural Area, Sensors, 9, 768-793, 2009.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., and de Jeu, R.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951-954, 2010.

Kross, A., Seaquist, J. W., Roulet, N. T., Fernandes, R., and Sonnentag, O.: Estimating carbon dioxide exchange rates at contrasting northern peatlands using MODIS satellite data, Remote Sensing of Environment, 137, 234-243, 2013.

Kustas, W. P., Perry, E. M., Doraiswamy, P. C., and Moran, M. S.: Using satellite remote sensing to extrapolate evapotranspiration in time and space over a semiarid rangeland, Remote Sens. Environ., 49, 275-286, 1994.

Liu, Y. Y., de Jeu, R. A. M., McCabe, M. F., Evans, J. P., and van Dijk, A. I. J. M.: Global long-term passive microwave satellite-based retrievals of vegetation optical depth, Geophys. Res. Lett., 38, L18402, 2011a.

Liu, Y. Y., Dorigo, W. A., Parinussa, R. M., De Jeu, R. A. M., Wagner, W., McCabe, M. F., Evans, J. P., and Van Dijk, A. I. J. M.: Trend-preserving blending of passive and active microwave soil moisture retrievals, Remote Sensing of Environment, 123, 280-297, 2012.

Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., M. Van Dijk, A. I. J., McCabe, M. F., and Evans, J. P.: Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, Hydrol. Earth Syst. Sci., 15, 425-436, 2011b.

Liu, Y. Y., van Dijk, A. I. J. M., McCabe, M. F., Evans, J. P., and de Jeu, R. A. M.: Global vegetation biomass change (1988-2008) and attribution to environmental and human drivers, Global Ecology and Biogeography, 22, 692-705, 2013.

Lokupitiya, E., Denning, S., Paustian, K., Baker, I., Schaefer, K., VERMA, S., MEYERS, T., Bernacchi, C. J., SUYKER, A., and Fischer, M.: Incorporation of crop phenology in Simple Biosphere Model (SiBcrop) to improve land-atmosphere carbon exchanges from croplands, Biogeosciences, 6, 1103 - 1103, 2009.

Luojus, K., Pulliainen, J., Takala, M., Lemmetyinen, J., Derksen, C., and Wang, L.: Snow water equivalent (SWE) product guide, Global Snow Monitoring for Climate Research, 1, 2010.

Mach, D. M., Christian, H. J., Blakeslee, R. J., Boccippio, D. J., Goodman, S. J., and Boeck, W. L.: Performance assessment of the optical transient detector and lightning imaging sensor, Journal of Geophysical Research: Atmospheres (1984–2012), 112, 2007.

McCabe, M. F. and Wood, E. F.: Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors, Remote Sensing of Environment, 105, 271-285, 2006.

McCabe, M. F., Wood, E. F., Wójcik, R., Pan, M., Sheffield, J., Gao, H., and Su, H.: Hydrological consistency using multi-sensor remote sensing data for water and energy cycle studies, Remote Sensing of Environment, 112, 430-444, 2008.

Merlin, O., Al Bitar, A., Rivalland, V., Béziat, P., Ceschia, E., and Dedieu, G.: An analytical model of evaporation efficiency for unsaturated soil surfaces with an arbitrary thickness, Journal of Applied Meteorology and Climatology, 50, 457-471, 2011.

Michel, D., Jiménez C., Miralles D. G., Jung M., Hirschi M., Ershadi A., Martens B., McCabe M. F., Fisher J. B., Mu Q., Seneviratne S. I., Wood E. F. and Fernández-Prieto D.: The WACMOS-ET project – Part 1: Tower-scale evaluation of four remote sensing-based evapotranspiration algorithms, Hydrol. Earth Syst. Sci. Discuss., 12(10): 10739-10787, 2015

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H., and Dolman, A. J.: Magnitude and variability of land evaporation and its components at the global scale, Hydrol. Earth Syst. Sci., 15, 967-981, 2011a.

Miralles, D. G., Gash, J. H., Holmes, T. R. H., de Jeu, R. A. M., and Dolman, A.: Global canopy interception from satellite observations, Journal of Geophysical Research, 115, D16122, 2010.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15, 453-469, 2011b.

Miralles, D. G., Teuling, A. J., van Heerwaarden, C. C., and de Arellano, J. V.-G.: Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation, Nature Geoscience, 7, 345-349, 2014a.

Miralles, D. G., van den Berg, M. J., Gash, J. H., Parinussa, R. M., de Jeu, R. A. M., Beck, H. E., Holmes, T. R. H., Jiménez, C., Verhoest, N. E. C., and Dorigo, W. A.: El Niño–La Niña cycle and recent trends in continental evaporation, Nature Climate Change, 4, 122-126, 2014b.

Miralles, D. G., Jiménez C., Jung M., Michel D., Ershadi A., McCabe M. F., Hirschi M., Martens B., Dolman A. J., Fisher J. B., Mu Q., Seneviratne S. I., Wood E. F. and Fernaìndez-Prieto D.: The WACMOS-ET project – Part 2: Evaluation of global terrestrial evaporation data sets, Hydrol. Earth Syst. Sci. Discuss., 12(10): 10651-10700, 2015

Monteith, J. L.: Evaporation and environment, Symp. Soc. Exp. Biol., 19, 205-234, 1965.

Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sensing of Environment, 111, 519-536, 2007.

Mu, Q., Zhao, M., Kimball, J. S., McDowell, N. G., and Running, S. W.: A Remotely Sensed Global Terrestrial Drought Severity Index, Bulletin of the American Meteorological Society, 94, 83-98, 2012.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sensing of Environment, 115, 1781-1800, 2011.

Mu, Q., Zhao, M., and Running, S. W.: MODIS Global Terrestrial Evapotranspiration (ET) Product (NASA MOD16A2/A3), Algorithm Theoretical Basis Document, Collection, 5, 2013.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., Miralles, D., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K. C., Wood, E. F., Zhang, Y., and Seneviratne, S. I.: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-dataset synthesis, Hydrol. Earth Syst. Sci. Discuss., 10, 769-805, 2013.

Mueller, B., Seneviratne, S. I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J. B., Guo, Z., Jung, M., Maignan, F., McCabe, M. F., Reichle, R., Reichstein, M., Rodell, M., Sheffield, J., Teuling, A. J., Wang, K., Wood, E. F., and Zhang, Y.: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, Geophysical Research Letters, 38, 2011.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models: Part I - a discussion of principles, Journal of Hydrology, 10, 282-290, 1970.

Nesbitt, S. W., Zipser, E. J., and Kummerow, C. D.: An examination of version-5 rainfall estimates from the TRMM Microwave Imager, precipitation radar, and rain gauges on global, regional, and storm scales, J. Appl. Meteorol., 43, 1016-1036, 2004.

Norman, J. M., Kustas, W. P., and Humes, K. S.: Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature, Agricultural and Forest Meteorology, 77, 263-293, 1995.

Otkin, J. A., Anderson, M. C., Hain, C., and Svoboda, M.: Examining the Relationship between Drought Development and Rapid Changes in the Evaporative Stress Index, Journal of Hydrometeorology, 15, 938-956, 2014.

Penman, H. L.: Natural Evaporation from Open Water, Bare Soil and Grass, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 193, 120-145, 1948.

Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., and Klooster, S. A.: Terrestrial ecosystem production: a process model based on global satellite and surface data, Global Biogeochemical Cycles, 7, 811-841, 1993.

Priestley , C. H. B. and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, Mon. Weather Rev., 100, 81-92, 1972.

Rebmann, C., Göckede, M., Foken, T., Aubinet, M., Aurela, M., Berbigier, P., Bernhofer, C., Buchmann, N., Carrara, A., and Cescatti, A.: Quality analysis applied on eddy covariance measurements at complex forest sites using footprint modelling, Theor Appl Climatol, 80, 121-141, 2005.

Reichstein, M., Rey, A., Freibauer, A., Tenhunen, J., Valentini, R., Banza, J., Casals, P., Cheng, Y., Grünzweig, J. M., and Irvine, J.: Modeling temporal and large‐scale spatial variability of soil respiration from soil water availability, temperature and vegetation productivity indices, Global biogeochemical cycles, 17, 2003.

Richardson, A. D., Black, T. A., Ciais, P., Delbart, N., Friedl, M. A., Gobron, N., Hollinger, D. Y., Kutsch, W. L., Longdoz, B., and Luyssaert, S.: Influence of spring and autumn phenological transitions on forest ecosystem productivity, Philosophical Transactions of the Royal Society B: Biological Sciences, 365, 3227-3246, 2010.

Richey, A. S., Thomas, B. F., Lo, M.-H., Reager, J. T., Famiglietti, J. S., Voss, K., Swenson, S., and Rodell, M.: Quantifying renewable groundwater stress with GRACE, Water Resources Research, doi: 10.1002/2015WR017349, 2015. n/a-n/a, 2015.

Rubel, F. and Kottek M.: Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification, Meteorologische Zeitschrift, 19(2): 135-141, 2010.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., and Behringer, D.: The NCEP climate forecast system reanalysis, Bulletin of the American Meteorological Society, 91, 1015-1057, 2010.

Sahoo, A. K., Pan, M., Troy, T. J., Vinukollu, R. K., Sheffield, J., and Wood, E. F.: Reconciling the global terrestrial water budget using satellite remote sensing, Remote Sensing of Environment, 115, 1850-1865, 2011.

Saigusa, N., Ichii, K., Murakami, H., Hirata, R., Asanuma, J., Den, H., Han, S. J., Ide, R., Li, S. G., and Ohta, T.: Impact of meteorological anomalies in the 2003 summer on Gross Primary Productivity in East Asia, Biogeosciences, 7, 641-655, 2010.

Scott, R. L.: Using watershed water balance to evaluate the accuracy of eddy covariance evaporation measurements for three semiarid ecosystems, Agricultural and Forest Meteorology, 150, 219-225, 2010.

Sheffield, J., Ferguson, C. R., Troy, T. J., Wood, E. F., and McCabe, M. F.: Closing the terrestrial water budget from satellite remote sensing, Geophysical Research Letters, 36, n/a-n/a, 2009.

Sheffield, J., Wood, E. F., and Munoz-Arriola, F.: Long-term regional estimates of evapotranspiration for Mexico based on downscaled ISCCP data, Journal of Hydrometeorology, 11, 253-275, 2010.

Shuttleworth, W. J. and Wallace, J. S.: Evaporation from sparse crops-an energy combination theory, Q. J. R. Meteorol. Soc., 111, 839-855, 1985.

Simard, M., Pinto, N., Fisher, J. B., and Baccini, A.: Mapping forest canopy height globally with spaceborne lidar, Journal of Geophysical Research: Biogeosciences, 116, G04021, 2011.

Smith, P., Lanigan, G., Kutsch, W. L., Buchmann, N., Eugster, W., Aubinet, M., Ceschia, E., Béziat, P., Yeluripati, J. B., and Osborne, B.: Measurements necessary for assessing the net ecosystem carbon budget of croplands, Agriculture, ecosystems & environment, 139, 302-315, 2010.

Sobrino, J. A., Jiménez-Muñoz, J. C., and Paolini, L.: Land surface temperature retrieval from LANDSAT TM 5, Remote Sensing of Environment, 90, 434-440, 2004.

Soudani, K., Hmimina, G., Dufrêne, E., Berveiller, D., Delpierre, N., Ourcival, J.-M., Rambal, S., and Joffre, R.: Relationships between photochemical reflectance index and light-use efficiency in deciduous and evergreen broadleaf forests, Remote Sensing of Environment, 144, 73-84, 2014.

Sprintsin, M., Cohen, S., Maseyk, K., Rotenberg, E., Grünzweig, J., Karnieli, A., Berliner, P., and Yakir, D.: Long term and seasonal courses of leaf area index in a semi-arid forest plantation, Agricultural and Forest Meteorology, 151, 565-574, 2011.

Stackhouse, P. W., Gupta, S. K., Cox, S. J., Zhang, T., Mikovitz, J. C., and Hinkelman, L. M.: The NASA/GEWEX surface radiation budget release 3.0: 24.5-year dataset, GEWEX News, 21, 10-12, 2011.

Stoy, P. C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., Arain, M. A., Arneth, A., Aurela, M., and Bernhofer, C.: A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity, Agricultural and forest meteorology, 171, 137-152, 2013.

Su, H., McCabe, M. F., Wood, E. F., Su, Z., and Prueger, J. H.: Modeling evapotranspiration during SMACEX: Comparing two approaches for local- and regional-scale prediction, Journal of Hydrometeorology, 6, 910-922, 2005.

Su, Z.: The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, Hydrol. Earth Syst. Sci., 6, 85-100, 2002.

Sulkava, M., Luyssaert, S., Zhehle, S., and Papale, D.: Assessing and improving the representativeness of monitoring networks: The European flux tower network example, Journal of Geophysical Research, 116, 2011.

Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., Vermote, E. F., and El Saleous, N.: An extended AVHRR 8 ‐ km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data, Int. J. Remote Sens., 26, 4485-4498, 2005.

van der Kwast, J., Timmermans, W., Gieske, A., Su, Z., Olioso, A., Jia, L., Elbers, J., Karssenberg, D., and de Jong, S.: Evaluation of the Surface Energy Balance System (SEBS) applied to ASTER imagery with flux-measurements at the SPARC 2004 site (Barrax, Spain), Hydrol. Earth Syst. Sci., 13, 1337-1347, 2009.

Veenendaal, M., Kolle, O., and Lloyd, J.: Seasonal variation in energy fluxes and carbon dioxide exchange for a broad leaved semi-arid savanna (Mopane woodland) in Southern Africa, Global Change Biology, 10, 318 - 328, 2004.

Vinukollu, R. K., Sheffield, J., Wood, E. F., Bosilovich, M. G., and Mocko, D.: Multimodel Analysis of Energy and Water Fluxes: Intercomparisons between Operational Analyses, a Land Surface Model, and Remote Sensing, Journal of Hydrometeorology, 13, 3-26, 2011a.

Vinukollu, R. K., Wood, E. F., Ferguson, C. R., and Fisher, J. B.: Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches, Remote Sensing of Environment, 115, 801-823, 2011b.

Weligepolage, K., Gieske, A. S. M., van der Tol, C., Timmermans, J., and Su, Z.: Effect of sub-layer corrections on the roughness parameterization of a Douglas fir forest, Agricultural and Forest Meteorology, 162–163, 115-126, 2012.

Wharton, S., Schroeder, M., Paw U, K. T., Falk, M., and Bible, K.: Turbulence considerations for comparing ecosystem exchange over old-growth and clear-cut stands for limited fetch and complex canopy flow conditions, Agricultural and Forest Meteorology, 149, 1477-1490, 2009.

Wohl, E., Barros, A., Brunsell, N., Chappell, N. A., Coe, M., Giambelluca, T., Goldsmith, S., Harmon, R., Hendrickx, J. M. H., Juvik, J., McDonnell, J., and Ogden, F.: The hydrology of the humid tropics, Nature Clim. Change, 2, 655-662, 2012.

Yan, Y., Zhao, B., Chen, J., Guo, H., Gu, Y., Wu, Q., and Li, B.: Closing the carbon budget of estuarine wetlands with tower ‐ based measurements and MODIS time series, Global Change Biology, 14, 1690-1702, 2008.

Yao, Y., Liang, S., Li, X., Hong, Y., Fisher, J. B., Zhang, N., Chen, J., Cheng, J., Zhao, S., and Zhang, X.: Bayesian multimodel estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations, Journal of Geophysical Research: Atmospheres, 119, 4521-4545, 2014.

Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., Samanta, A., Piao, S., Nemani, R. R., and Myneni, R. B.: Global data sets of vegetation leaf area index (LAI) 3g and Fraction of Photosynthetically Active Radiation (FPAR) 3g derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the period 1981 to 2011, Remote Sensing, 5, 927-948, 2013.

Zierl, B., Bugmann, H., and Tague, C. L.: Water and carbon fluxes of European ecosystems: An evaluation of the ecohydrological model RHESSys, Hydrological processes, 21, 3328-3339, 2007.

Table 1: Summary of data sources for tower-based and grid-based analysis and their spatial and temporal resolutions.

Variable	Tower-based	Grid-based	Model
Air temperature	Tower data aggregated to 3-hourly	LandFlux data at 0.5° and 3-hourly	All models
Humidity	Tower-based relative humidity converted to specific humidity and aggregated to 3-hourly	Specific humidity from LandFlux data at 0.5° and 3-hourly	All except GLEAM
Pressure	Calculated as a function of ground elevation	LandFlux data at 0.5° and 3-hourly	All models
Net radiation	Tower data aggregated to 3-hourly	LandFlux data from SRB v3 at 1° and 3-hourly	All models
Ground heat flux	Tower data aggregated to 3-hourly	Calculated from net radiation and fractional vegetation cover data, 0.5° and 3-hourly	All models
Land surface temperature	Calculated from tower-based longwave upward radiation and aggregated to 3-hourly	LandFlux data at 0.5° and 3-hourly	SEBS only
Wind speed	Tower data aggregated to 3-hourly	LandFlux data at 0.5° and 3-hourly	SEBS only
Canopy height	Tower meta data	JPL product and Equation 1	SEBS only
NDVI	GIMMS NDVI at 8km and bi-monthly	GIMMS NDVI at 0.5° and bi-monthly	All except GLEAM
Leaf area index	Calculated from NDVI	LandFlux data at 0.5° and monthly	SEBS and PM-Mu
Fractional vegetation cover	Not used as ground heat flux is available.	Calculated from NDVI	All except GLEAM
Precipitation	Tower data aggregated to 3-hourly	LandFlux data at 0.5° and 3-hourly	GLEAM only
Soil properties	IGBP-DIS at 5 arc-minutes	IGBP-DIS data aggregated to 0.5°	GLEAM only
Soil moisture	CCI-WACMOS data at 0.25° and daily	Same as tower-based	GLEAM only
Soil depth	GlobSnow (daily and 25 km)	Same as tower-based	GLEAM only
Vegetation optical depth	From Liu et al. (2011b) at 0.25° and daily	Same as tower-based	GLEAM only
Snow water equivalent	GlobSnow and NSIDC at 0.25° and daily	Same as tower-based	GLEAM only
Lightning frequency	Monthly climatology at 0.5°	Same as tower-based	GLEAM only
Cover fractions	MOD44B data at 250 m	MOD44B data at 0.5°	GLEAM only

Figure 1: Location of the selected towers and their distributions for various biomes

Figure 2: Scatterplots of observed versus simulated latent heat flux for tower-based data. Colors show the frequency of values from high (red) to low (yellow). The thick black line represents the linear regression, while the thin line is the 1:1 line. The series of small circles show the percentile increments of data from the 1st to 99th, with large circles denoting the 25th, 50th and 75th percentiles. The statistics shown on each figure provide coefficient of determination (R^2), slope (m), y-intercept (b), number of data records (n), the root-mean-squared difference (RMSD), relative error (RE) and the Nash-Sutcliffe Efficiency (NSE).

Figure 3: Scatterplots of observed versus simulated evaporation for grid-based data. Colors show the frequency of values from high (red) to low (yellow). The thick black line is the linear regression and the thin line is the 1:1 line. The series of small circles show the percentile increments of data from the 1st to 99th, with large circles denoting the 25th, 50th and 75th percentiles. The statistics shown on the graphs are coefficient of determination (R^2), slope (m), y-intercept (b), number of data records (n), the root-mean-squared difference (RMSD), relative error (RE) and the Nash-Sutcliffe Efficiency (NSE).

Figure 4: Comparison of the performance skill of the models in reproducing evaporation for the tower-based analyses. R^2 is the coefficient of determination, RE is relative error (lower is better) and NSE is the Nash-Sutcliffe Efficiency (higher is better). Towers are arranged from left to right based on an aridity index (secondary y-axis).

Figure 5: Coefficient of determination (R^2), relative error (RE) and Nash-Sutcliffe Efficiency (NSE) for models across different biome types. Each point represents the collection of all available 3-hourly records of towers located within the selected biome, with the number of towers shown on the secondary y-axis of the R^2 plot in red. NSE for the shrubland response of SEBS is printed.

Figure 6: Percentile plots of observed (x-axis) versus estimated latent heat flux (y-axis) at 3-hourly resolution for the tower-based analysis across the seven studied biomes. Percentiles encompass the 1st to 99th range in 1 percent increments, with Q_{25}, Q_{50} and Q_{75} denoted by large coloured circles.

Figure 7: The upper panel presents Nash-Sutcliffe Efficiency (NSE; x-axis) and R^2 (color tone) between tower- and grid-based values for net radiation, land surface temperature, air temperature, wind speed, specific humidity, fractional vegetation cover and leaf area index, across the seven studied biome types. The lower panel presents the NSE (x-axis) and R^2 of model simulated evaporation against closure-corrected observed values. The number of towers for each biome type used in the analysis are shown in red font on the secondary (right) axis in each of the plots. Statistics for those results beyond the range of the x-axis are printed separately on the plot.

Figure 8: Coefficient of determination (R^2), relative error (RE) and Nash-Sutcliffe Efficiency (NSE) for model simulated results across the five different climate zones (y-axis). The zones are represented by dryland (DRY), temperate continental (TempCONT), temperate (TEMP), sub-tropical (subTRO) and boreal (BOR). Each point represents the collection of all towers located within the selected climate zone, with the number of towers shown on the secondary y-axis of the R^2 panel in red.

Figure 9: Percentile plots of observed (x-axis) versus estimated latent heat flux (y-axis) at 3-hourly resolution for tower-based analysis and across the different climate zones. Percentiles encompass the 1st to 99th range in 1 percent increments. Q_{25}, Q_{50} and Q_{75} are denoted by large circles.

Figure 10: The upper panel shows Nash-Sutcliffe Efficiency (NSE; x-axis) and R^2 (color tone) between tower-based and grid-based values for net radiation, land surface temperature, air temperature, wind speed, specific humidity, fractional vegetation cover and leaf area index across the five different climate zones. The lower panel shows NSE (x-axis) and R^2 of model simulated evaporation against closure-corrected observed values across climate zones. The number of towers for each biome are shown in red font on the secondary (right) axis of the plots. Statistics for the grid-based SEBS result over dry climate zone are printed.

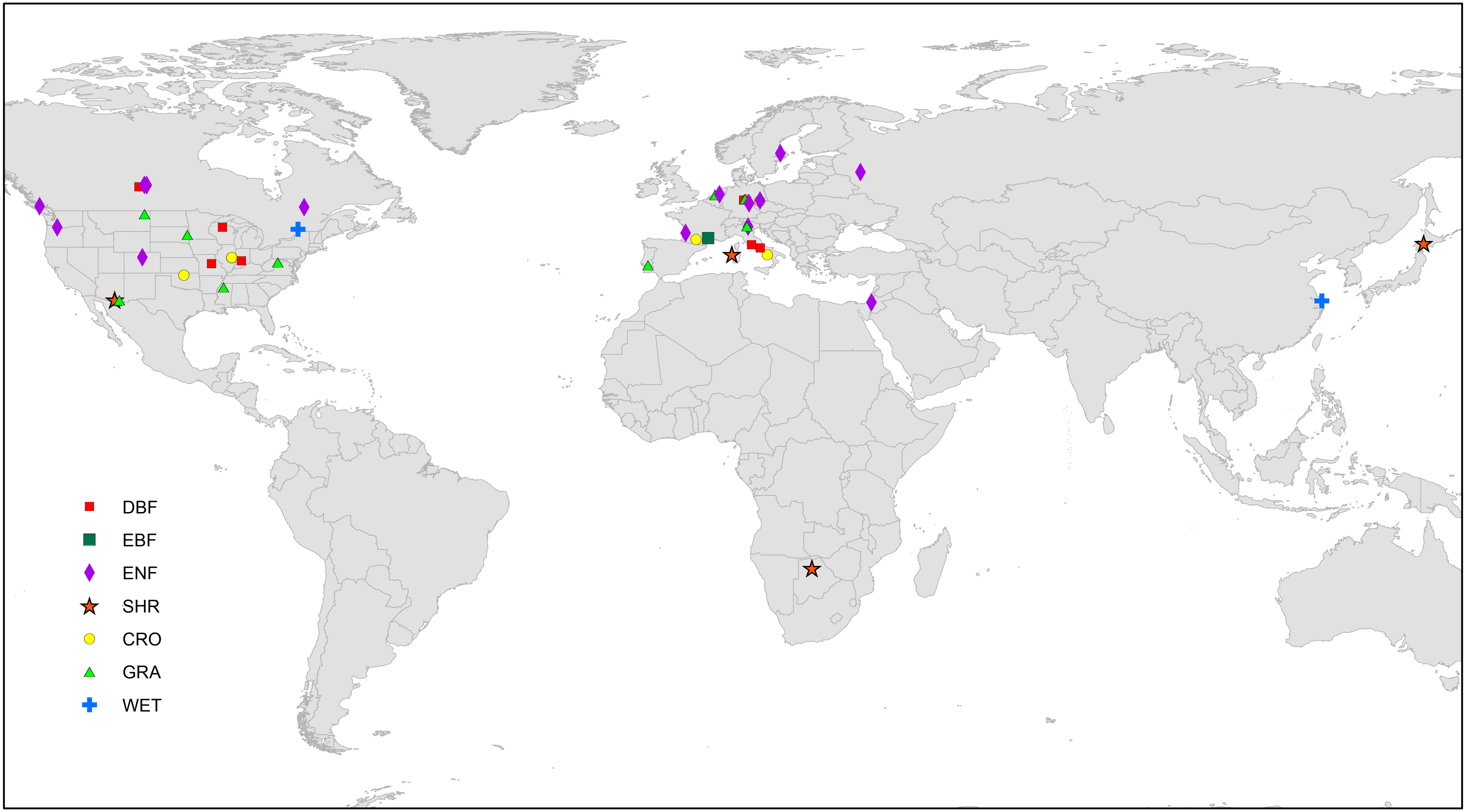


Figure 1

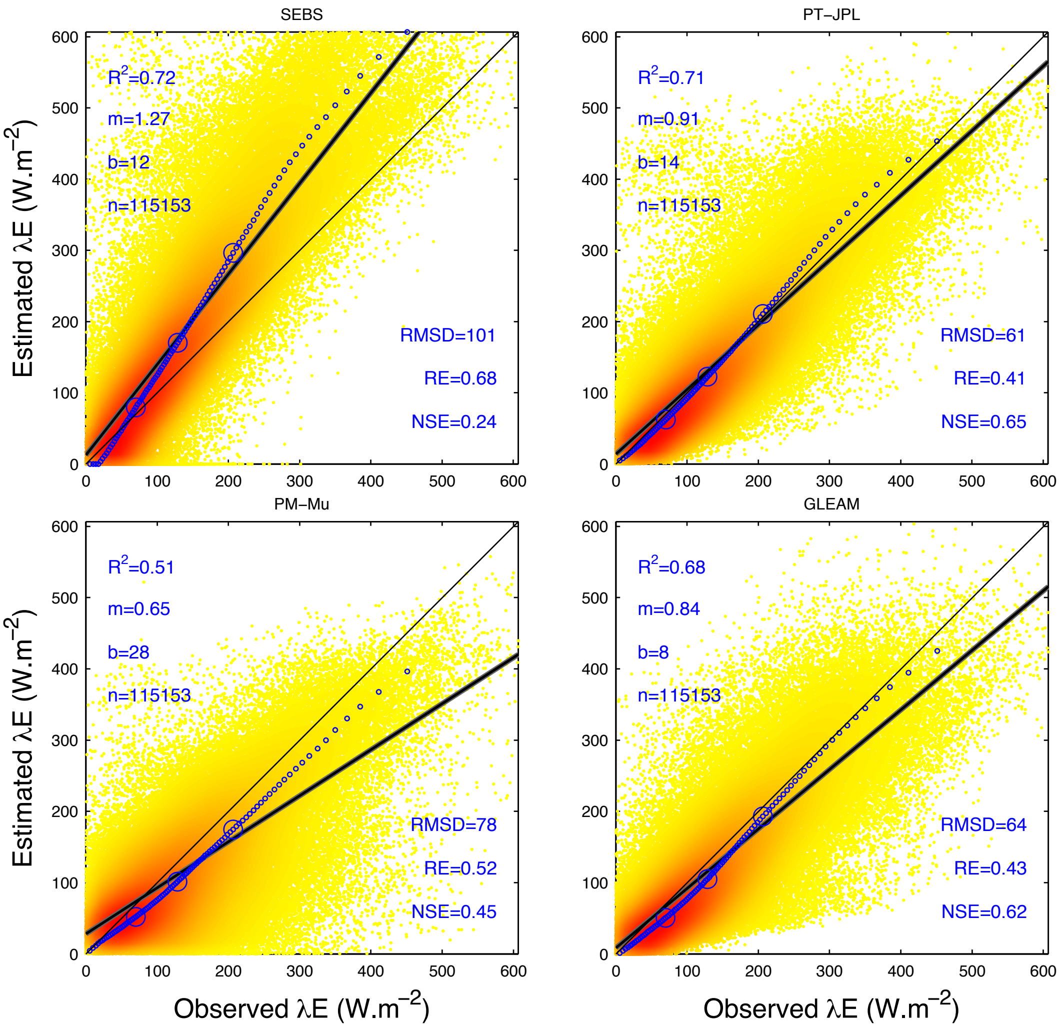
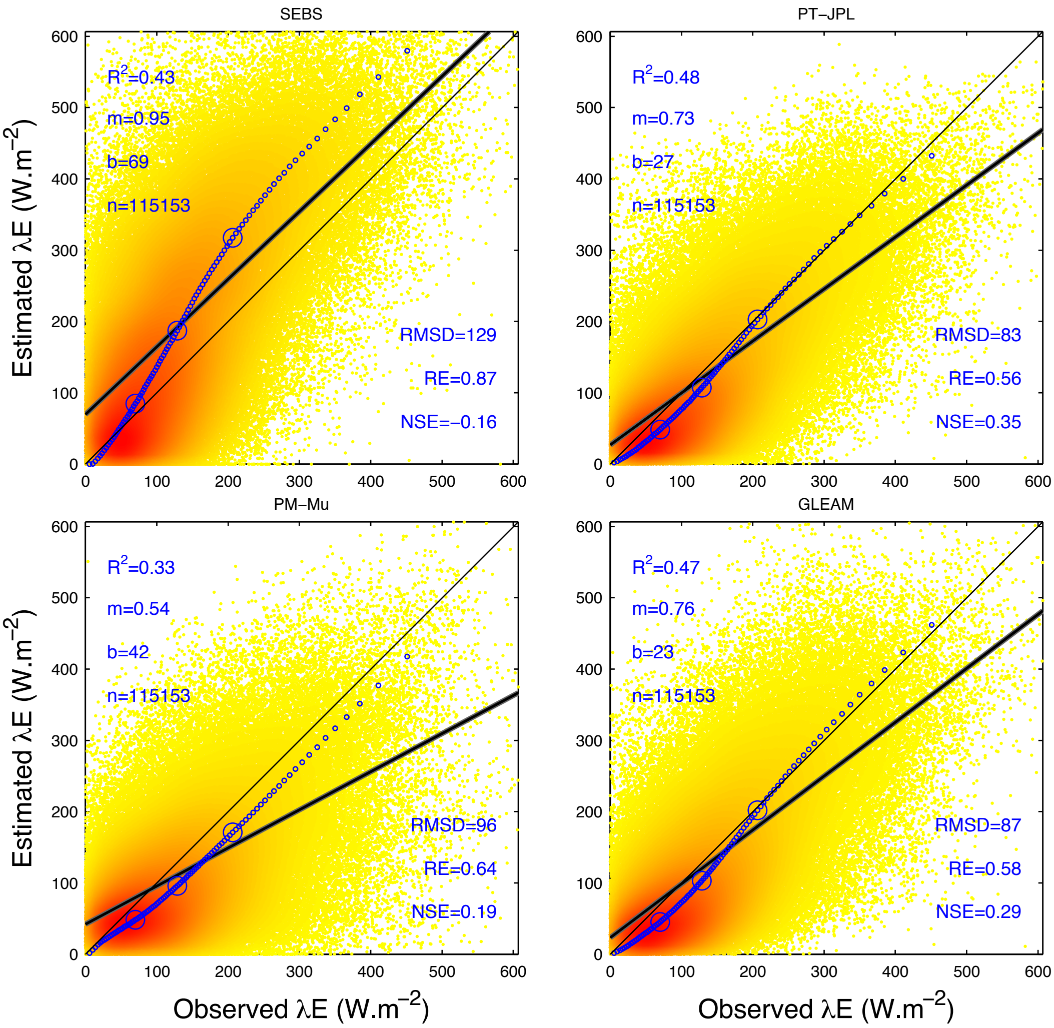
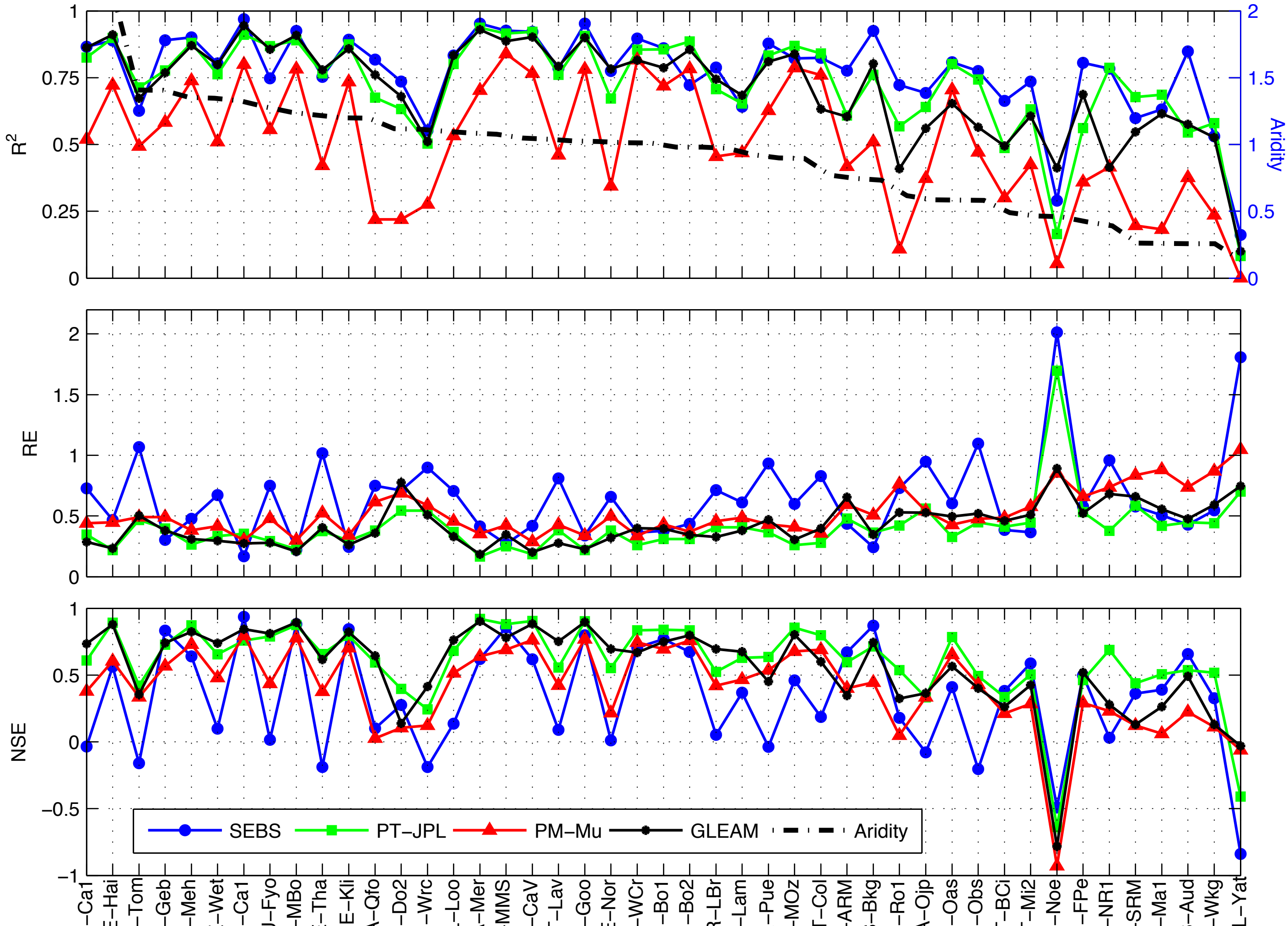

SEBS
R²=0.72
m=1.27
b=12
n=115153
RMSD=101
RE=0.68
NSE=0.24
PT−JPL
R²=0.71
m=0.91
b=14
n=115153
RMSD=61
RE=0.41
NSE=0.65
PM−Mu
R²=0.51
m=0.65
b=28
n=115153
RMSD=78
RE=0.52
NSE=0.45
GLEAM
R²=0.68
m=0.84
b=8
n=115153
RMSD=64
RE=0.43
NSE=0.62
Estimated λE (W.m⁻²)
Observed λE (W.m⁻²)

Figure 2



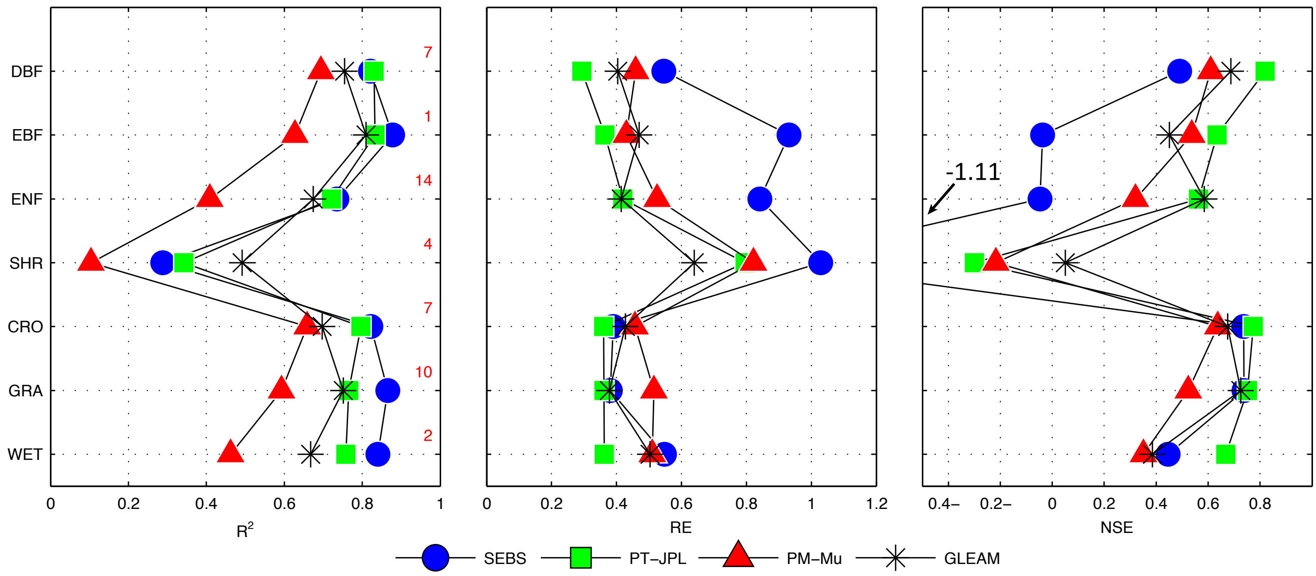
SEBS
PT−JPL
PM−Mu
GLEAM
Estimated λE (W.m−2)
Observed λE (W.m−2)
R2=0.43
m=0.95
b=69
n=115153
RMSD=129
RE=0.87
NSE=−0.16
R2=0.48
m=0.73
b=27
n=115153
RMSD=83
RE=0.56
NSE=0.35
R2=0.33
m=0.54
b=42
n=115153
RMSD=96
RE=0.64
NSE=0.19
R2=0.47
m=0.76
b=23
n=115153
RMSD=87
RE=0.58
NSE=0.29

Figure 3



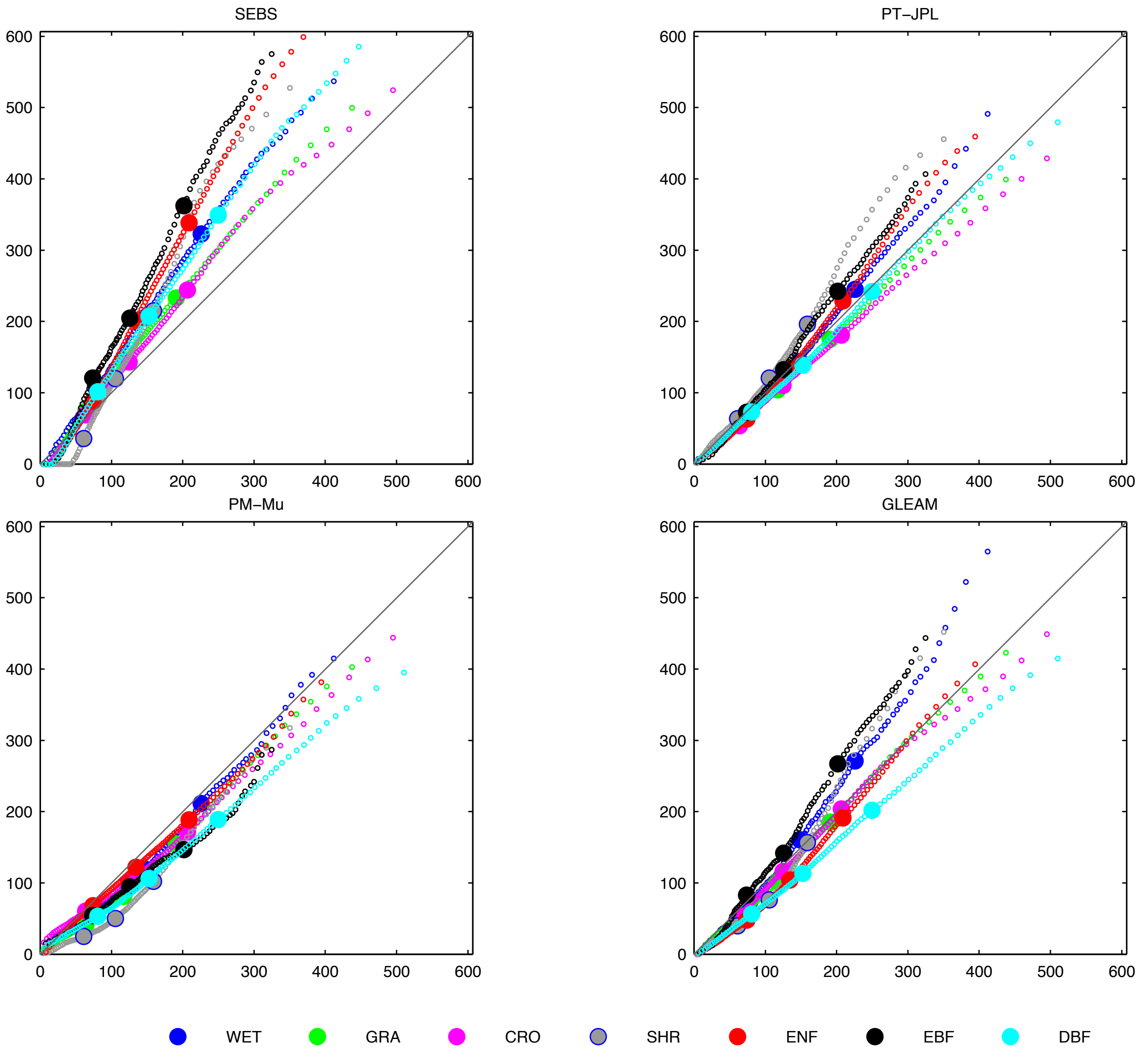
R^2
Aridity
RE
NSE
SEBS
PT−JPL
PM−Mu
GLEAM
Aridity
CA−Ca1
DE−Hai
JP−Tom
DE−Geb
DE−Meh
DE−Wet
NL−Ca1
RU−Fyo
IT−MBo
DE−Tha
DE−Kli
CA−Qfo
CN−Do2
US−Wrc
NL−Loo
CA−Mer
US−MMS
US−CaV
IT−Lav
US−Goo
SE−Nor
US−WCr
US−Bo1
US−Bo2
FR−LBr
FR−Lam
FR−Pue
US−MOz
IT−Col
US−ARM
US−Bkg
IT−Ro1
CA−Ojp
CA−Oas
CA−Obs
IT−BCi
PT−Mi2
IT−Noe
US−FPe
US−NR1
US−SRM
BW−Ma1
US−Aud
US−Wkg
IL−Yat

Figure 4



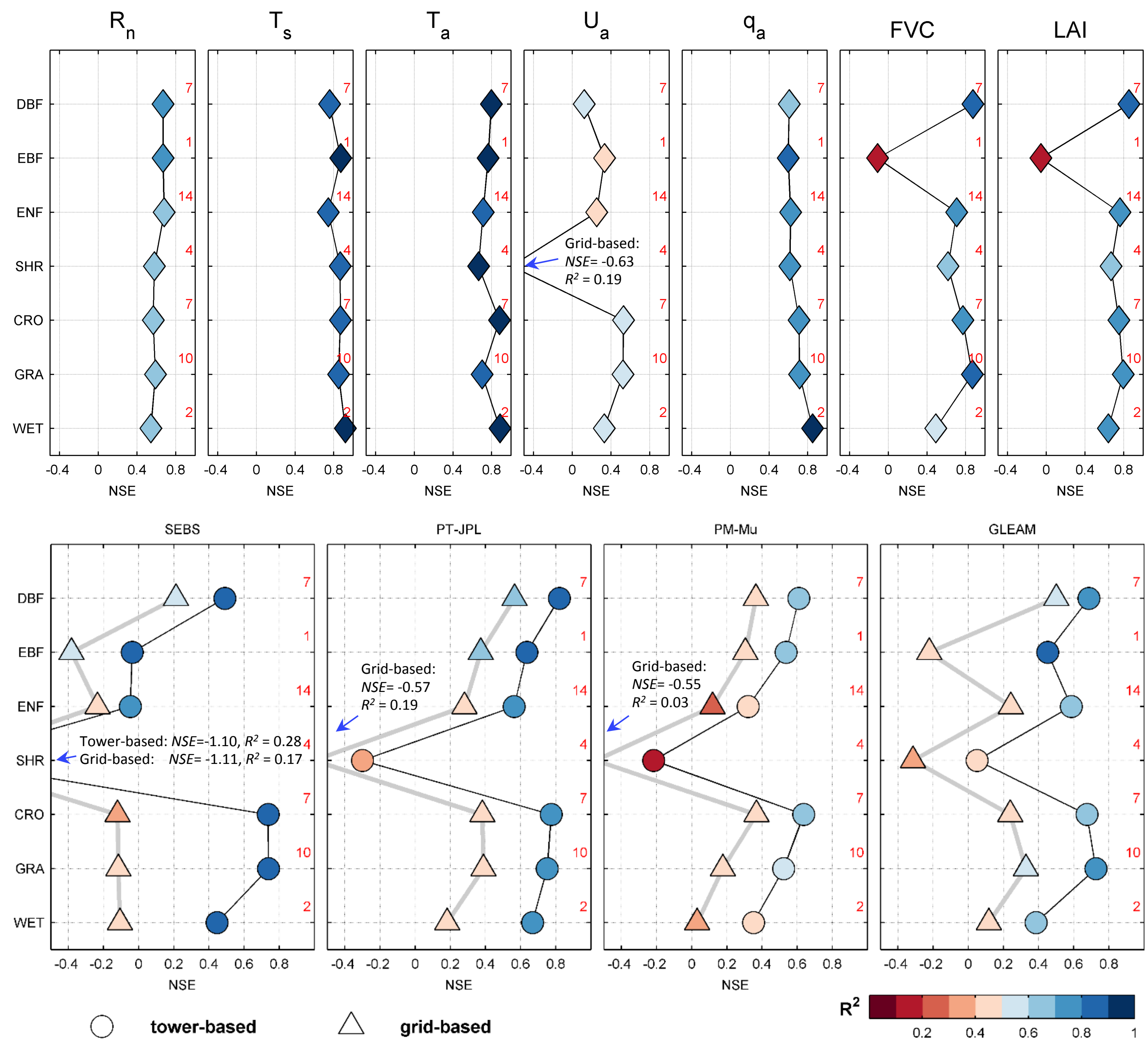
DBF
EBF
ENF
SHR
CRO
GRA
WET
7
1
14
4
7
10
2
0
0.2
0.4
0.6
0.8
1
R2
0
0.2
0.4
0.6
0.8
1
1.2
RE
-1.11
0.4−
0.2−
0
0.2
0.4
0.6
0.8
NSE
SEBS
PT−JPL
PM−Mu
GLEAM

Figure 5



SEBS
PT−JPL
PM−Mu
GLEAM
0
100
200
300
400
500
600
WET
GRA
CRO
SHR
ENF
EBF
DBF

Figure 6



Rn
Ts
Ta
Ua
qa
FVC
LAI
DBF
EBF
ENF
SHR
CRO
GRA
WET
NSE
Grid-based:
NSE= -0.63
R² = 0.19
SEBS
PT-JPL
PM-Mu
GLEAM
Tower-based: NSE=-1.10, R² = 0.28
Grid-based: NSE= -1.11, R² = 0.17
Grid-based:
NSE= -0.57
R² = 0.19
Grid-based:
NSE= -0.55
R² = 0.03
tower-based
grid-based
R²

Figure 7

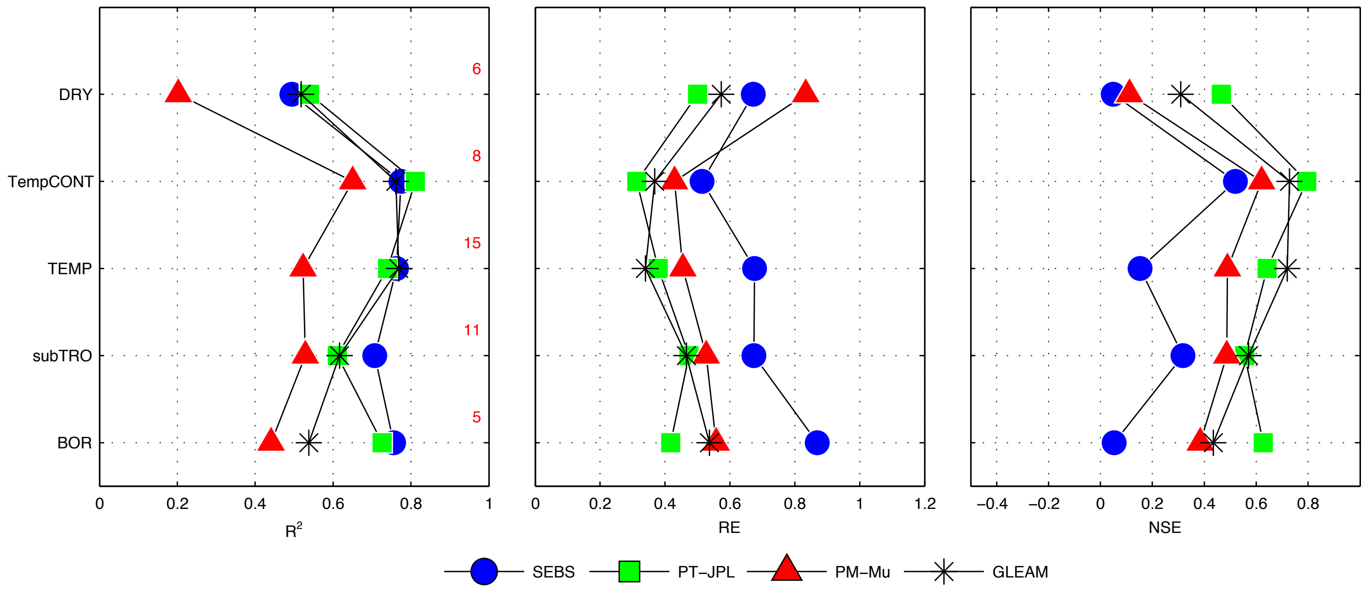
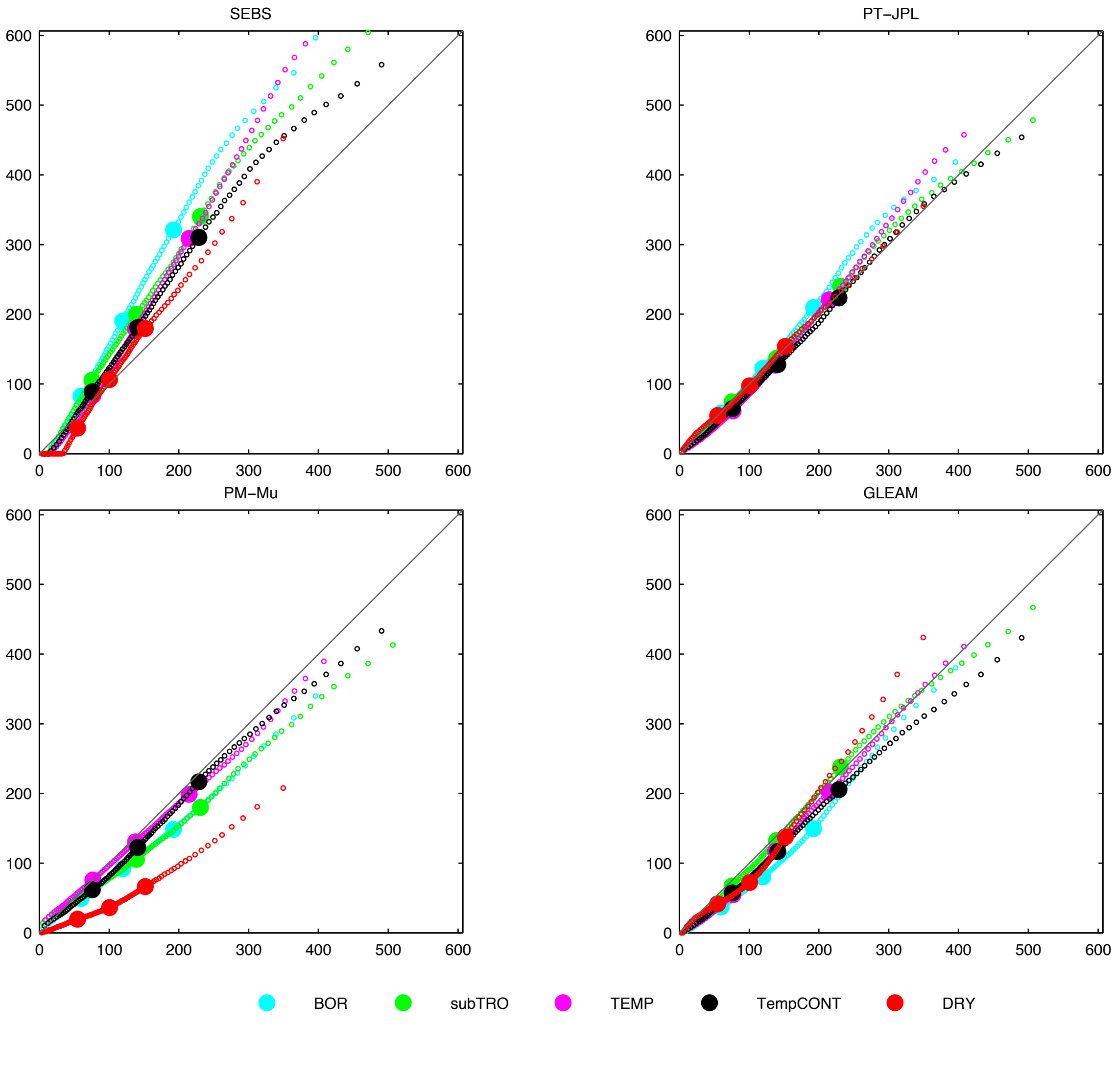
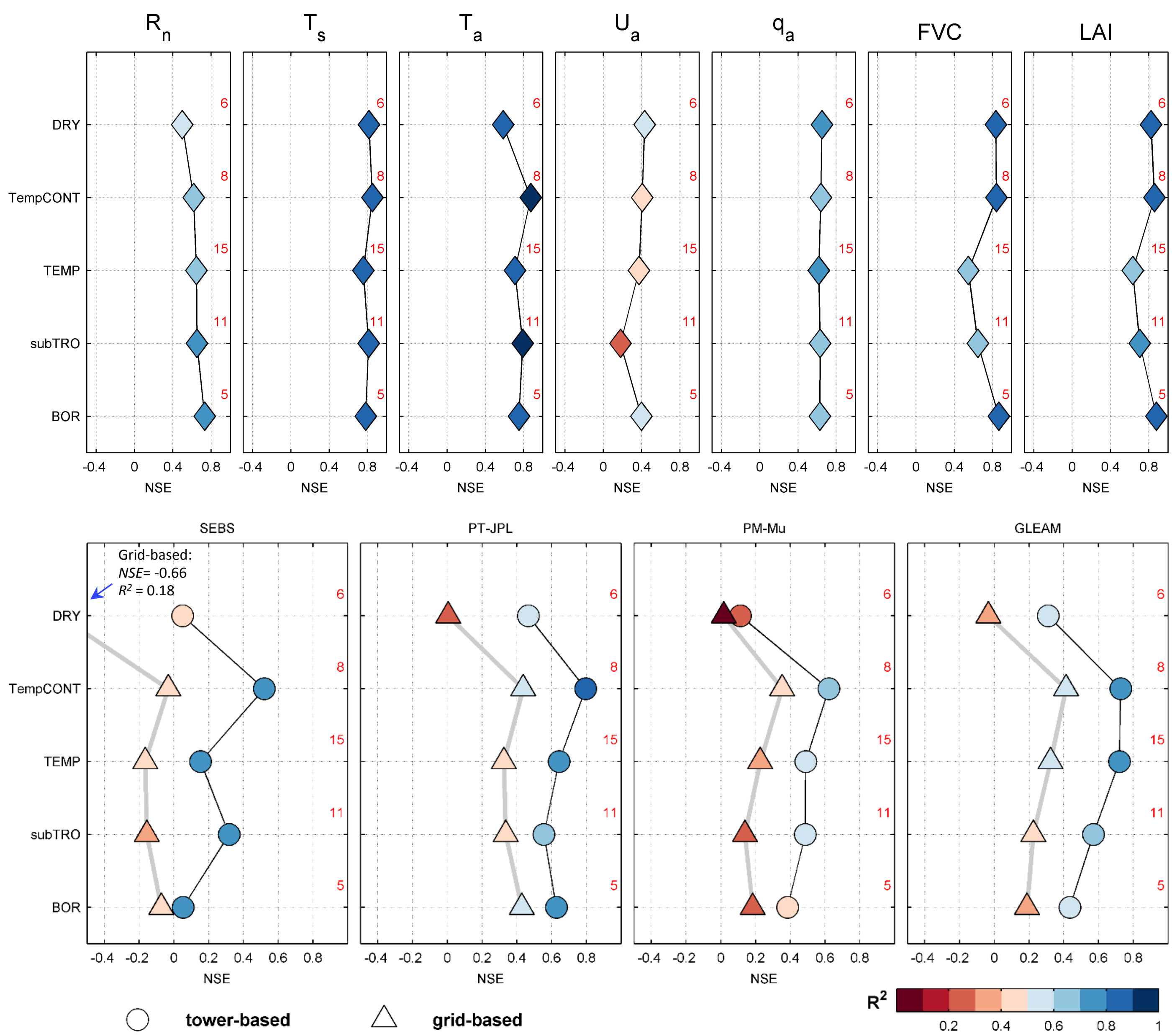

DRY
TempCONT
TEMP
subTRO
BOR
6
8
15
11
5
0
0.2
0.4
0.6
0.8
1
R^2
0
0.2
0.4
0.6
0.8
1
1.2
RE
−0.4
−0.2
0
0.2
0.4
0.6
0.8
NSE
SEBS
PT−JPL
PM−Mu
GLEAM

Figure 8



SEBS
PT–JPL
PM–Mu
GLEAM
0
100
200
300
400
500
600
BOR
subTRO
TEMP
TempCONT
DRY

Figure 9



Rn
Ts
Ta
Ua
qa
FVC
LAI
DRY
TempCONT
TEMP
subTRO
BOR
NSE
SEBS
PT-JPL
PM-Mu
GLEAM
Grid-based:
NSE= -0.66
R² = 0.18
tower-based
grid-based
R²

Figure 10