AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Biogeosciences

RESEARCH ARTICLE

10.1002/2016JG003591

Key Points:

- We present a ground heat flux model intercomparison using six models evaluated in situ across land cover and climate zones and globally
- For instantaneous applications, ground heat flux should be modeled differently for bare soil compared to moderate to high vegetation cover
- For daily applications a globally calibrated thermal diffusion model demonstrates the best ability to capture seasonal dynamics

Supporting Information:

- Supporting Information S1
- Table S1

Correspondence to:
A. J. Purdy,
ajpurdy@uci.edu

Citation:
Purdy, A. J., J. B. Fisher, M. L. Goulden, and J. S. Famiglietti (2016), Ground heat flux: an analytical review of 6 models evaluated at 88 sites and globally, *J. Geophys. Res. Biogeosci.*, *121*, doi:10.1002/2016JG003591.

Received 17 AUG 2016
Accepted 22 NOV 2016
Accepted article online 28 NOV 2016

Ground heat flux: an analytical review of 6 models evaluated at 88 sites and globally

A. J. Purdy[1], J. B. Fisher[2], M. L. Goulden[1], and J. S. Famiglietti[1,2]

[1]Department of Earth System Science, University of California, Irvine, California, USA, [2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA

Abstract Uncertainty in ground heat flux (G) means that evaluation of the other terms in the surface energy balance (e.g., latent and sensible heat fluxes (LE and H)) remains problematic. Algorithms that calculate LE and H require available energy, the difference between net radiation, R_{NET}, and G. There are a wide range of approaches to model G for large-scale applications, with a subsequent wide range of estimates and accuracies. We provide the largest review of these methods to date ($N = 6$), evaluating modeled G against measured G from 88 FLUXNET sites. The instantaneous midday variability in G is best captured by models forced with net radiation, while models forced by temperature show the least error at both instantaneous and daily time scales. We produce global decadal data sets of G to illustrate regional and seasonal sensitivities, as well as uncertainty. Global model mean midmorning instantaneous G is highest during September, October, and November at 63.42 ($\pm$16.84) Wm^{-2}, while over December, January, and February G is lowest at 53.86 ($\pm$18.09) Wm^{-2} but shows greater intermodel uncertainty. Results from this work have the potential to improve evapotranspiration estimates and guide appropriate G model selection and development for various land uses.

1. Introduction and Background

Ground heat flux (G) is an integral part of the surface energy budget ($R_{NET} - G = LE + H + V$). Net radiation, R_{NET}, and G are balanced by latent heat (LE), sensible heat (H), and chemical energy provided by metabolism or used by photosynthesis in plants (V), a negligible amount. Ground heat flux accounts for the energy gained or lost during belowground warming or cooling. Commonly used approaches to calculate LE and evapotranspiration (ET), such as the Penman-Monteith equation, the Priestley-Taylor equation, and the residual of the energy balance, not only need high-fidelity R_{NET} but also require G to calculate the available energy [*Penman*, 1948; *Monteith*, 1965; *Priestley and Taylor*, 1972]. The magnitude of G varies greatly across different landscapes. In wet areas with dense canopy G is small, while in arid regions with sparse canopy midday G reaches comparable amounts of energy to H and often larger amounts than ET. With G varying orders of magnitude across different landscapes and being an essential part of available energy to support global ET applications, the need for robust and accurate estimates of G is evident. However, many approaches to model G were formulated with limited spatiotemporal sampling and have since been broadly applied. A clear characterization of the discrepancies and potential sources of bias in current G models has the potential to improve estimates of available energy, increase the accuracy and consistency of ET estimates, and facilitate scrutiny of mechanistic model differences across ET algorithms, which is an actively ongoing focus of research [*Vinukollu et al.*, 2011; *Jiménez et al.*, 2011; *McCabe et al.*, 2013; *Mueller et al.*, 2013; *Ershadi et al.*, 2014; *Chen et al.*, 2014; *Michel et al.*, 2016; *Miralles et al.*, 2016].

Remote sensing algorithms developed to calculate daily ET at high spatial resolutions created a need for spatially explicit G estimates. Subsequently, many methods to quantify G using satellite data were developed. Initial approaches assumed G to be a constant fraction of R_{NET} or G to be negligible at daily or longer times [*Seguin and Itier*, 1983]. Later work derived linear and nonlinear empirical relationships between G/R_{NET} and vegetation indices [*Reginato et al.*, 1985; *Clothier et al.*, 1986; *Choudhury et al.*, 1987; *Kustas and Daughtry*, 1990; *Kustas et al.*, 1993] and G/R_{NET} and surface temperatures [*Jacobsen and Hansen*, 1999; *Mu et al.*, 2011]. Recently, G models have employed physically based analytical solutions to thermal diffusion equations [*Bennett et al.*, 2008; *Holmes et al.*, 2008]. The assumption that G is always negligible is not appropriate, especially at midmorning times near satellite overpasses or in areas with sparse vegetation cover (Figures 1 and 2) [*Daughtry et al.*, 1990]. Setting G as a constant fraction of R_{NET} discounts the impact of

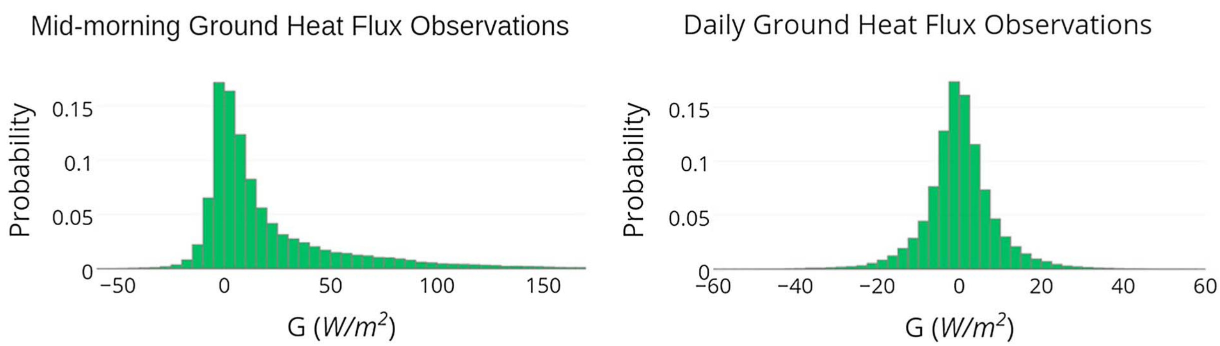

Figure 1. Probability density distributions of observed instantaneous midmorning and daily G from the 88 FLUXNET towers used in this analysis. (left) Probability density distribution of midmorning instantaneous G. Although G has the highest probability of being near 0, the long tail at the positive end of G indicates that G is an integral term in the instantaneous energy balance. (right) Probability density distribution of daily G. Daily G is more often than not 0. However, a normal distribution around G that spans from $-40\ \mathrm{W m^{-2}}$ to $40\ \mathrm{W m^{-2}}$ demonstrates the seasonality of G at daily time steps.

spatially varying soil properties, neglects the influence from vegetation insulation, and disregards conservation of energy, unless efforts are taken to equally weight periods when G is positive (daytime/summer) and negative (nighttime/winter). Variables that influence the magnitude of G/R_{NET} include soil properties, vegetation cover and height, and temperature fluctuations [*Santenello and Friedl*, 2002]. These different factors impart different magnitudes of influence at instantaneous or daily time scales. Consequently, models have been developed to quantify G instantaneously or aggregated across daily or longer times. The G model formulations and variable selection can be seen in Table 1, but models primarily use vegetation characteristics or temperature.

Vegetation cover density impacts G by attenuating incoming radiation and temperature fluctuations at the soil surface. As vegetation cover increases, the ratio of G/R_{NET} decreases. Despite the general agreement of this relationship, G/R_{NET} varies for distinct amount vegetation cover across models, especially for areas with sparse to no vegetation (Figure 2). Previous research reported G/R_{NET} to range from 0 at complete canopy cover to 0.39 for bare soil [*Mu et al.*, 2011]. Other studies indicated smaller ranges: *Kustas and Daughtry* [1990] found G/R_{NET} spanning from 0.15 at full canopy to 0.30 for bare soil, while *Reginato et al.* [1985] observed a much smaller range from 0.05 for full canopy and 0.1 for bare soil. The breadth of these ranges demonstrates how localized measurements for one particular environment do not translate toward broader applications and that factors other than vegetation cover regulate G/R_{NET}.

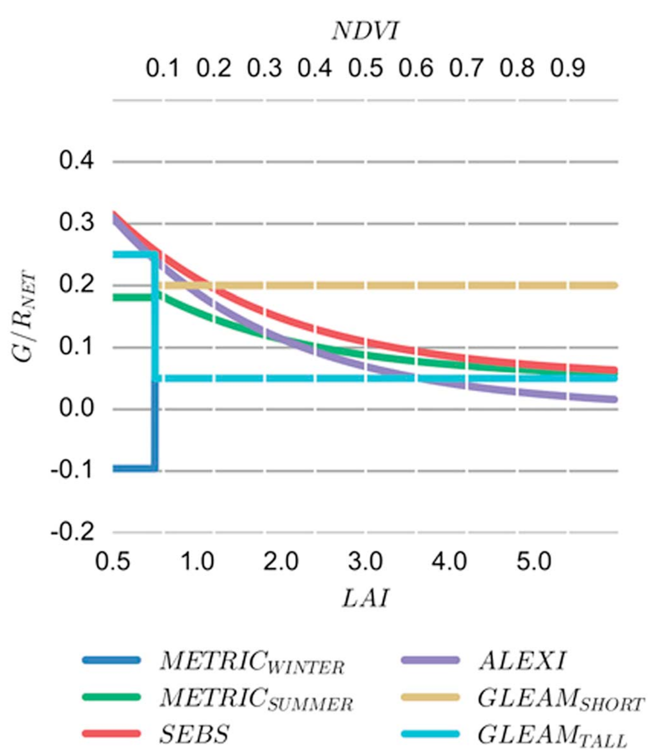

Figure 2. Vegetation influence on instantaneous modeled G/R_{NET} ratios. The separation between these models is greatest at low vegetation cover.

Soil properties and land surface temperature are also known to impact G. *Idso et al.* [1975] measured G/R_{NET} to be 0.5 for dry bare soil and 0.3 for the same bare soil when saturated. Land surface temperature (LST), also known as the radiometric skin temperature, has been shown to correlate to G.

Table 1. Table of Widely Used *G* Modules in Current *LE* Algorithms Including the Equations to Calculate Both Instantaneous *G* and Daily *G*[a]

Model	Equation	Source
	Developed for instantaneous applications	
METRIC	$G/R_{NET} = 0.05 + 0.18e^{-0.521 \times LAI}$ LAI > 0.5 $G/R_{NET} = 1.8 \times (T_s - 273.15)/R_{NET} + 0.084$ LAI < 0.5	*Allen et al.* [2007]
SEBS	$G/R_{NET} = \Gamma_C + (1 - f_c) \times (\Gamma_S - \Gamma_C)$ $\Gamma_C = 0.315$ fraction of $G : R_{NET}$ for full canopy cover $\Gamma_S = 0.05$ fraction of $G : R_{NET}$ for bare soil	*Su* [2002]
ALEXI	${}^{G}/_{R_{NET}} = 0.31 \times (1 - f_c)$	*Anderson et al.* [2007]
	Developed for daily applications	
GLEAM	$G/R_{NET} = 0.05$ tall canopy $G/R_{NET} = 0.20$ short canopy $G/R_{NET} = 0.25$ bare soil	*Miralles et al.* [2011]
MOD16	for $T_{an} > -8^oC$ and $T_{an} < 25^oC$ and $T_{dif} > 5^oC$: $G_{s(day/night)} = 4.73 \times (T_{day/night} - 273.15) - 20.8$ for $T_{dif} < 5^oC \mid T_{an} < -8^oC \mid T_{an} > 25^oC$: $G_{s(day/night)} = 0$ Then G is capped at $G_s[G_s > LE + H] = 0.39 \times (LE + H)$	*Mu et al.* [2011]
Thermal diffusion	$G(t) = \frac{I}{\sqrt{\pi}} \cdot \frac{dT(0,s)}{\sqrt{t-s}}$ $I = \sqrt{\rho c k}$ I = thermal inertia, c = specific heat T = skin temperature, ρ = soil bulk density k = thermal conductivity, s = integration variable	*Bennett et al.* [2008]

[a]f_c is the fractional cover, LAI is the leaf area index, T_s is the surface temperature, T_{an} is the mean annual temperature, T_{dif} is the daytime and nighttime temperature difference, and $T_{day/night}$ is the day or night surface air temperature, respectively.

Empirical relationships and thermal diffusion solutions have been successful at modeling *G* over aggregated daily time steps [*Bennett et al.*, 2008; *Holmes et al.*, 2008]. A linear empirical relationship developed for the Arctic tundra between simultaneously measured surface temperature and *G* exhibited high correlation and low error and has since been adopted for global applications [*Jacobsen and Hansen*, 1999; *Mu et al.*, 2011]. However, a linear relationship between skin temperature and *G* may not capture other factors that impact G, such as attenuation due to vegetation cover, the thermal conductivity of soil, or the temperature gradient of thawing tundra. Thermal diffusion solutions use the integrated time difference between land surface temperatures to calculate *G*. This approach mimics a physically accurate method to transfer heat from the atmosphere into the Earth surface while relying on knowledge of soil surface thermal inertia estimates. *Bennett et al.* [2008] bypass the need for spatially explicit global soil properties by parameterizing a constant thermal inertia for each location globally. To date, this method has only been applied globally at relatively coarse scales (1o–3o) using reanalysis data sets [*Bennett et al.*, 2008; *Vinukollu et al.*, 2011].

Many previous studies demonstrate success at tuning a *G* model for specific a location, but many of these *G* models and their optimized parameter sets have not been tested across a robust observation data set with a variety of land covers and various climates. With a push for global high-resolution spatiotemporal ET data, some ET models and their respective *G* representations that were constructed to function over specific land uses have since been applied to continental applications without scrutiny [*Allen et al.*, 2015]. The limited studies which have investigated differences in *G* models have only focused on irrigated agricultural land uses [*Cammelleri et al.*, 2009; *Irmak et al.*, 2011]. These studies found overall poor performance compared to mean in situ *G* observations and concluded that local calibration is necessary for successful model application. Additionally, the differences between methods to quantify *G* at both instantaneous and daily resolutions need to be better understood to aide appropriate *G* model selection in global ET algorithms.

Globally distributed observations of G at FLUXNET eddy covariance towers and global satellite observations of vegetation and LST facilitate the direct comparison of numerous G models across a robust global observation data set to address the limitations of previous work. Determining the best method to quantify G will lead to a high-fidelity G data set to apply to global ET algorithms and reduce the energy budget closure uncertainty at towers that have poor or missing G measurements. We compare several currently used methodologies ($N = 6$) to answer three main questions with this study: (1) What is the best G model structure for both instantaneous and daily ET algorithms? (2) What mechanisms govern G across instantaneous and daily time scales? (3) What is the impact of G uncertainty on ET globally?

2. Methods

Global energy flux and meteorological observations from the FLUXNET eddy covariance site network provide a robust data set to assess current remote sensing G models. In this section, we introduce the FLUXNET synthesized data set, describe the satellite vegetation and temperature data, the radiation data used to perform this analysis, and detail the statistical metrics used to evaluate model performance.

2.1. Data Sets

2.1.1. FLUXNET La Thuile Data Set and Validation Sites

The FLUXNET eddy covariance tower network provides a decadal set of carbon, water, and energy cycle observations across a numerous biomes and climates [*Baldocchi et al.*, 2001]. The La Thuile data set is a subset of this network providing harmonious quality control treatment and gap filling to limit potential biases arising from data-processing techniques. Data are available from the FLUXNET database (http:///www.fluxdata.org). Despite being the best available collection of globally distributed observations, many locations lack a full year of observations, experience instrument quality degradation, and locate ground heat flux plates and soil thermocouples to calculate storage at different depths (2–15 cm) to measure G. We subset and filtered the La Thuile data set for sites with data that met our requirements for remote sensing G model evaluation. Selected towers for this study contain at least 90% high-quality G and R_{NET} observations for 330 days for a given year based on the La Thuile table of core variables present for each year (http://www.fluxdata.org). Only original observed data or high-quality, gap-filled data for both G and R_{NET} are used in this analysis. Overall, we used measurements from 88 towers across 11 climates and 10 biomes to evaluate modeled G (Table S1 and Figure S1 in the supporting information). All tower data used in this analysis were open access.

The inherent uncertainty associated with small-scale variability of G due to soil moisture, soil conductivity, vegetation cover, sensor placement, and sensor accuracy contributes to the limited performance against coarser resolution remote sensing footprints in more heterogeneous landscapes. The large sample size ($N = 88$) mitigates potential bias from tower representativeness or sensor placement that may exist with a smaller sample size. Previous energy balance closure assessments have pointed out that limited sampling of G may contribute up to 15% of the closure uncertainty [*Twine et al.*, 2000]. Additionally, the variability for G measurements is highest in the early morning and midday, the time when many ET algorithms require high-fidelity energy balance flux observations [*Kustas et al.*, 2000]. To reduce bias from gridded forcing data and remote sensing observations to in situ tower comparison we use in situ observations of R_{NET} for forcing data for models that calculate G as a fraction of R_{NET}. Models are compared against instantaneous midmorning (9:30–10:30) G and daily G tower observations. The tower data are only used to assess models forced by high-resolution remote sensing data.

2.1.2. Moderate Resolution Infrared Spectroradiometer Data

The Moderate Resolution Infrared Spectroradiometer (MODIS) provided continuous high-resolution global coverage of vegetation phenology and land surface temperature. These observations span from 2000 to present at resolutions of 250 m–5600 m. Observations at 250 m and 1 km were used to evaluate G models against FLUXNET point observations of G and R_{NET}. We utilized the Oak Ridge National Laboratory MODIS land product subset tool and apply quality control filters to extract good to excellent quality MODIS normalized difference vegetation index (NDVI) and land surface temperature data to evaluate each remote sensing model (http://daac.ornl.gov/MODIS/).

2.1.2.1. MODIS Vegetation Data

The MOD13Q1 16-daily 250 m normalized difference vegetation index (NDVI) data set was sampled at each of the FLUXNET. Linear interpolation from 16-daily to daily NDVI was used for daily analysis. This interpolation method is commonly used to fill missing data gaps in current *LE* algorithms [*Ershadi et al.*, 2014]. For global spatial comparisons, the 16-day MOD13C1 0.05o NDVI data set was applied. Fractional cover is calculated assuming a linear relationship with NDVI. This is based on the fraction of photosynthetic active radiation intercepted by total vegetation cover [*Fisher et al.*, 2008]. LAI is calculated from fractional cover as LAI $= \frac{-\ln(1-f_c)}{0.5}$ [*Ross*, 1976; *Fisher et al.*, 2008].

2.1.2.2. MODIS Land Surface Temperature

Land surface temperature (LST) at 1 km from MOD11A1 was sampled at FLUXNET site locations daily, while daily MOD11C1 0.05° LST was used in spatial comparisons. MODIS quality control flags were used to filter each data set to avoid cloud contamination. The LST data encompass both the soil skin temperature and the canopy skin temperature for partially vegetated areas.

2.1.3. Reanalysis Data

The National Center for Environmental Prediction (NCEP) provides global reanalysis data sets including short-wave radiation, long-wave radiation, skin temperature, and ground heat flux. The reanalysis data set uses data assimilation to combine observations and model simulations. We used daily (24 hourly) radiation, skin temperature, and ground heat flux at 2.5° by 2.5° gridded data to complete this analysis. Skin temperature was used to calibrate the thermal inertia parameters by land classification for the heat diffusion approach [*Bennett et al.*, 2008].

2.2. Models

Models from six widely used ET algorithms that also calculate *G* are compared. Three of these models were developed with the intent to model instantaneous *G*, while two other models were developed for use at daily time steps, and one model was developed to model *G* separately for day and night. The *G* models either apply vegetation properties to reduce R_{NET} to *G* or use skin temperature to model *G* (Table 1).

For instantaneous applications, we compare three models that use different vegetation properties (LAI and fractional cover) and R_{NET} to calculate *G*. The Mapping EvapoTRanspiration using Inverse Calibration (METRIC) algorithm relies on measures of LAI to partition controls of *G*, where for sparse cover a linear relationship of LST normalized by R_{NET} estimates *G*. At moderate and high vegetation cover, the fraction of G/R_{NET} decreases exponentially with increasing LAI. The Surface Energy Balance System (SEBS) model uses fractional cover to determine the portion of R_{NET} that contributes to *G* [*Monteith*, 1973; *Kustas and Daughtry*, 1990; *Su*, 2002]. Similar to SEBS, the Atmospheric Land EXchange Inverse (ALEXI) model assumes *G* to be a constant fraction of the R_{NET} that reaches the soil surface [*Anderson et al.*, 2007]. All the instantaneous approaches (METRIC, SEBS, and ALEXI) incorporate vegetation phenology through calculating and removing radiation intercepted by the canopy, after which a fraction of the energy which reaches the soil determines *G*.

Three distinct theoretical approaches are used to compare modeled *G* at daily resolution. The Global Land-surface Evaporation: the Amsterdam Methodology (GLEAM) quantifies the daily *G* from set fractions of R_{NET} based on canopy height and canopy cover. Tall canopies reduce the magnitude of *G* more than short canopies [*Kustas and Daughtry*, 1990; *Miralles et al.*, 2011]. The MOD16 ET algorithm [*Mu et al.*, 2011] models *G* at both daytime and night using a linear relationship with surface temperature. Additionally, this method includes temperature constraints to set *G* equal to 0 for extremely hot climates, extremely cold climates, and in areas with small diurnal temperature changes. Furthermore, a maximum fraction of G/R_{NET} is set for the incoming radiation that reaches the soil surface. Daily *G* for MOD16 is computed from the average of daytime and night values. Lastly, a thermal diffusion (T-DIFF) approach is applied to quantify *G* using the amount of heat that is transferred from the atmosphere to the soil. This approach requires parameterization of soil properties that represent the soil thermal inertia [*Bennett et al.*, 2008]. We force the T-DIFF model with nighttime LST from MODIS. Model equations and variables are described in detail in Table 1.

We evaluate all models against each other and use the original model parameterization while changing the temporal resolution of the forcing data to compare instantaneous and daily *G* separately. At midmorning instantaneous times we evaluate five models (ALEXI, METRIC, SEBS, MOD16, and GLEAM), while at daily time steps we compare all six models. The T-DIFF model structure and forcing data requirement prevent modeling *G* at the instantaneous time steps.

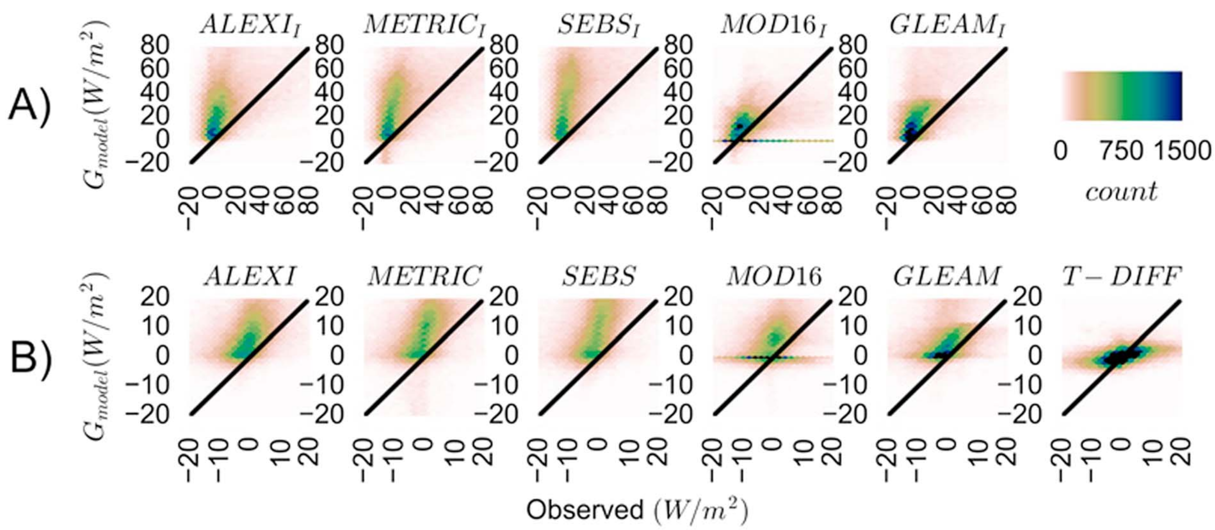

Figure 3. Top 2 rows with grey text (subscript I), frequency scatterplots of modeled and observed midday instantaneous *G*. The instantaneous models all overestimate instantaneous *G* as seen by deviation above 1:1 line. Bottom 2 rows with black text, frequency scatterplots of daily *G*. Scatterplots are truncated from 80 Wm^{-2} to -20 Wm^{-2} for instantaneous comparison and 20 Wm^{-2} to -20 Wm^{-2} for daily comparison to maximize *G* observations.

Many of the *G* models evaluated here were originally calibrated for specific land uses; therefore, model performance should vary across these different plant functional types (PFTs). For example, the METRIC and ALEXI *G* models were developed for use over cropland cover and grasslands, the *G* model in MOD16 was originally developed for Arctic tundra, and the GLEAM and T-DIFF *G* models were developed for global applications. Because sampling across PFTs for FLUXNET towers is limited with respect to high-quality G/R_{NET} observations, we evaluate model performance across the four most sampled PFTs: grassland (GRA; $N=25$), cropland (CRO; $N=15$), evergreen forest (ENF; $N=23$), and deciduous forest (DBF; $N=13$). As in the above global comparison, models are evaluated at both the instantaneously and daily temporal resolutions.

2.3. Statistical Evaluation

The above models and data are used to evaluate modeled *G*. Statistical metrics, including the mean bias (BIAS), root-mean-square error (RMSE), and the Kendall's tau (KT) coefficient, are used to objectively rank instantaneous and daily *G* models against in situ observations. Model performance and skill are evaluated using in situ observations from FLUXNET. Measurement errors may degrade model comparison to in situ observations, but these errors do not impact the relative ranking of model performance because all models are subjected to error equally. Intermodel uncertainty is quantified from the standard deviation of modeled *G* normalized by R_{NET}. This allows for global seasonal model assessments to identify where high model disagreement exists. The spatial comparison calculates the difference in *G* models normalized by R_{NET} to highlight the times and regions with the largest model disagreement. These steps will help determine the optimal *G* model to capture instantaneous midmorning *G* and daily G, the mechanisms that control *G* at these different time scales, and the potential impact of modeled *G* uncertainty on ET algorithms.

3. Results

3.1. Model Evaluation Against In Situ Observations

3.1.1. Midmorning Instantaneous Model Evaluation

The *G* models in currently applied ET algorithms exhibit a wide range in performance across the FLUXNET sites. Site-wide analysis reveals that the ALEXI, METRIC, SEBS, GLEAM, and MOD16 models more often than not overestimate instantaneous *G* observations with the slope between modeled *G* compared measured *G* greater than 1.0 coinciding with a positive BIAS (Figure 3). Models' individual performances vary across all sites with average RMSE ranging from the least error from MOD16 (RMSE = 26.93 Wm^{-2}) to the highest error from SEBS (RMSE = 42.08 Wm^{-2}; Table 2). The site-wide average absolute BIAS ranges from a low of 14.96 Wm^{-2} from MOD16 to a maximum BIAS of 31.59 Wm^{-2} from SEBS. The average model KT spans from

Table 2. Instantaneous Model Performance Across All FLUXNET Sites[a]

Model	RMSE	BIAS	KT	Slope	Int	R^2
ALEXI	35.20	23.54	0.36	0.37	22.27	0.12
GLEAM	35.14	24.86	0.45	0.48	15.50	0.21
METRIC	32.40	21.31	0.43	0.40	22.49	0.11
MOD16	26.93	14.85	0.41	0.27	9.62	0.10
SEBS	42.08	31.59	0.40	0.42	26.12	0.10

[a]RMSE is the root-mean-square error, BIAS is the mean absolute difference between model mean and observed mean, and KT is the mean Kendall's tau statistic across all sites. Slope and intercept (Int) are the coefficients of the model *G* versus observed *G*. R^2 is the coefficient of determination for the linear relationship described by Slope and Int.

the worst KT at 0.36 from ALEXI to the highest KT at 0.45 from GLEAM. Of all the *G* models analyzed at the towers used in this study, MOD16 exhibits the strongest performance with the lowest overall error (RMSE = 26.93 $Wm^{-2}$and BIAS = 14.85 Wm^{-2}), while maintaining similar ability to other models at capturing *G* variability (KT = 0.41; Table 2).

3.1.2. Daily Model Evaluation

Of the six models, the thermal diffusion model (T-DIFF) best fits the observations with both the lowest average error given the towers and conditions for this analysis (RMSE = 7.34 Wm^{-2}; BIAS = 1.45 Wm^{-2}) and the second highest explanation of variance (KT = 0.38; Table 3). The T-DIFF model underestimates the magnitude of daily *G* with a negative bias and a slope between modeled *G* and measured *G* less than 1, while the other models all overestimate *G* with slopes greater than 1 and a positive BIAS (Figure 3). The GLEAM approach only explains slightly more variance (KT = 0.40) than T-DIFF, but GLEAM has average errors across all sites that are twice as large (RMSE = 14.00 Wm^{-2}; BIAS = 10.21 Wm^{-2}) as T-DIFF. Like GLEAM, MOD16 has twice as much error compared to the T-DIFF model and explains the least *G* variability (KT = 0.27). The other models originally suited to quantify instantaneous *G*, ALEXI, GLEAM, METRIC, and SEBS all explain a similar amount of variance to T-DIFF but exhibit larger errors (Table 3). The RMSEs of each model are at least 1.75 times greater than T-DIFF, and the absolute BIASs are at least 5 times greater than T-DIFF.

3.1.3. Model Performance by Land Use

We evaluate *G* models across the four most sampled land covers, grassland, cropland, evergreen forest, and deciduous forest. Four FLUXNET sites were selected to provide an example of the wide range in modeled *G* across a year for these distinct land covers (Figure 4). For the instantaneous model statistics averaged over grasslands, METRIC results in the lowest errors (RMSE = 32.20 Wm^{-2}; BIAS = 15.75 Wm^{-2}), and GLEAM explains slightly more variance (KT = 0.51) than other models. For cropland cover, METRIC again has the lowest errors (RMSE = 30.03 Wm^{-2}; BIAS = 10.82 Wm^{-2}), while METRIC and GLEAM share the highest explanation of variance (KT = 0.52). For evergreen needleleaf forests MOD16 results in the lowest error (RMSE = 15.90 Wm^{-2}; BIAS = 8.10 Wm^{-2}) and is again followed by GLEAM (RMSE 16.13 Wm^{-2}; BIAS 10.49 Wm^{-2}). METRIC explains the most variance (KT = 0.45) for evergreen needleleaf forests. In deciduous broadleaf forests, MOD16 has the lowest error (RMSE = 14.75 Wm^{-2}; BIAS = 8.79 Wm^{-2}) followed by GLEAM (RMSE = 15.41 Wm^{-2}; BIAS = 11.30 Wm^{-2}), while GLEAM explains the most variance (KT = 0.36). Figure 5 shows the range in model performance for each of these statistics across all sites for each PFT. The MOD16 and GLEAM models exhibit more consistent performance over deciduous broadleaf forest and evergreen needleleaf forest with tighter error statistic box plots compared to other PFTs. Explanation of variance is generally higher for grassland and cropland cover compared to forests. Model errors, specifically, MOD16

Table 3. Daily Model Performance Across All FLUXNET Sites[a]

Model	RMSE	BIAS	KT	Slope	Int	R^2
ALEXI	12.63	9.02	0.36	0.32	8.86	0.12
GLEAM	14.00	10.21	0.40	0.23	5.02	0.08
METRIC	12.91	8.07	0.37	0.39	7.07	0.12
MOD16	10.75	6.39	0.26	0.22	4.78	0.06
SEBS	14.98	11.51	0.37	0.07	9.90	0.01
T-DIFF	7.34	1.45	0.38	0.16	0.61	0.17

[a]Statistics are the same as indicated in Table 2.

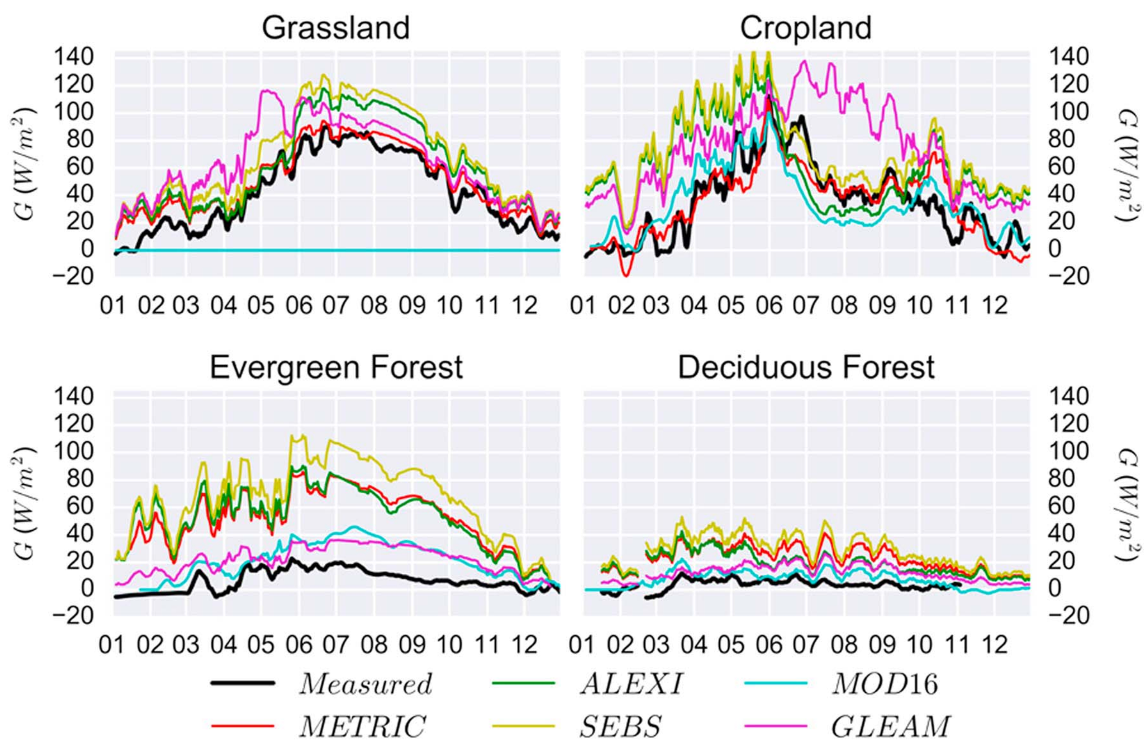

Figure 4. Instantaneous model performance compared across four different land covers at individual representative sites ((top left) grassland U.S.-Var, (top right) cropland U.S.-Ne1, (bottom left) evergreen forest U.S.-Blo, and (bottom right) deciduous forest UK-Ham). Data are plotted for one calendar year; the numbers along the bottom axis indicate each month in the year. The METRIC model most closely matches measured *G* for grassland and cropland cover, while GLEAM and MOD16 more closely match measured *G* over forested land cover. Model spread is high for each site demonstrating need for appropriate model choice for certain PFTs and potential for model improvement through global calibration.

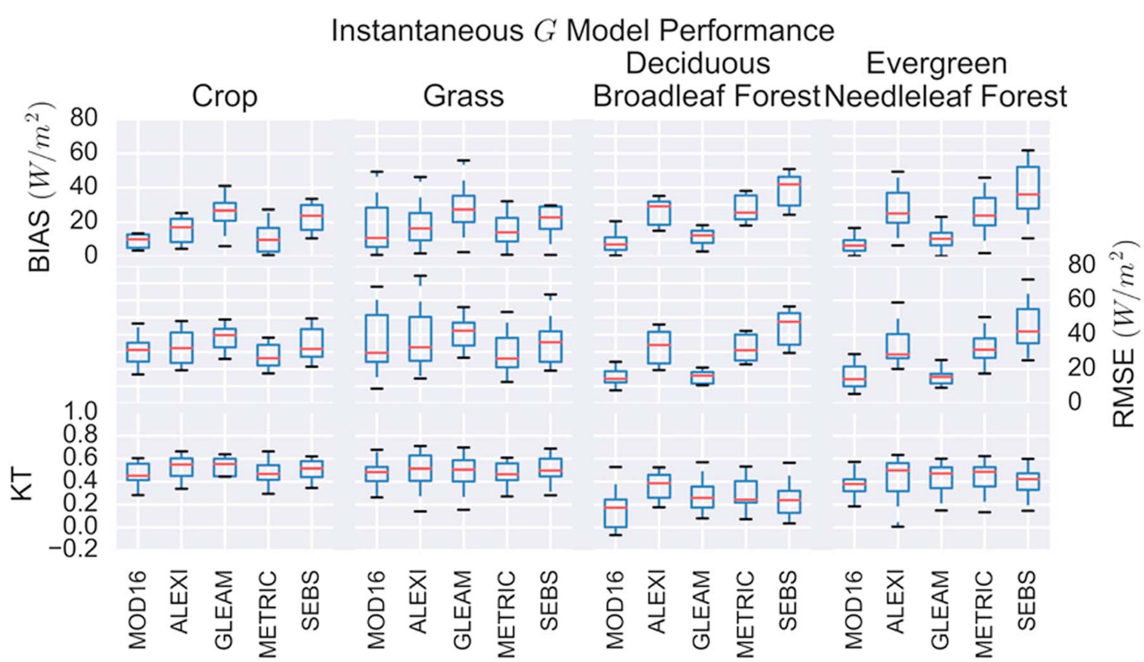

Figure 5. Instantaneous model performance across four most sampled PFTs in the observation data set. The METRIC *G* model is strongest for cropland and grassland cover. The MOD16 and GLEAM models show the strongest performance for both deciduous broadleaf forest and evergreen needleleaf forest.

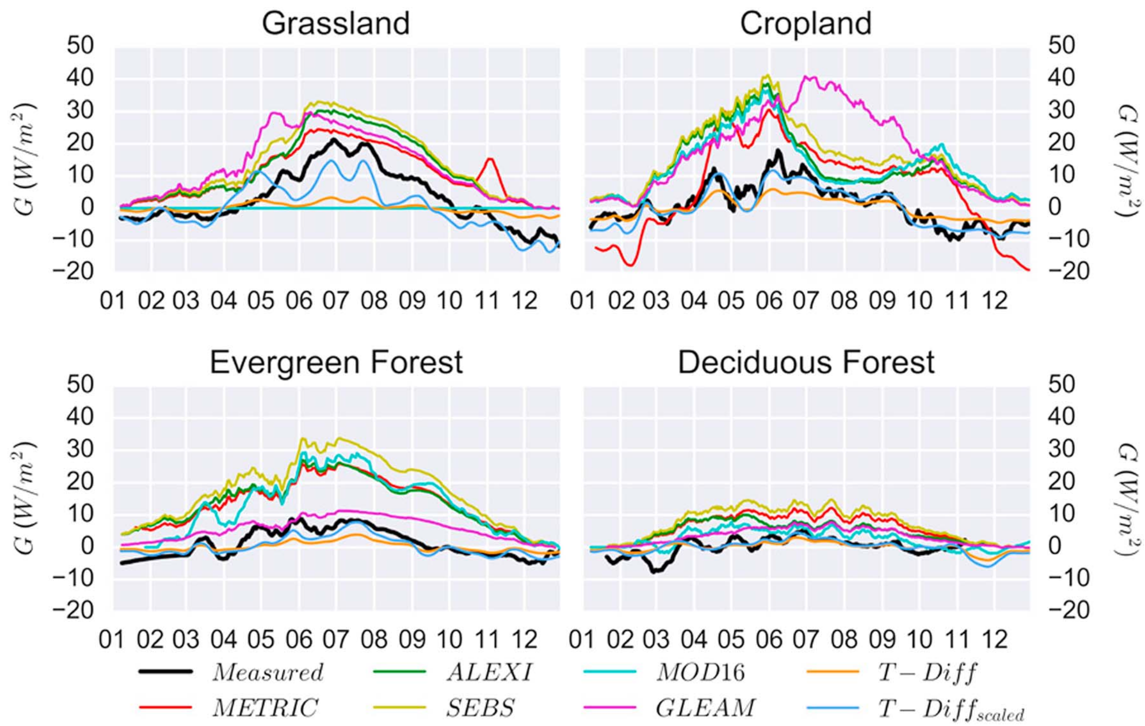

Figure 6. Daily model performance compared across four different land covers at same representative sites as seen in Figure 4. The T-DIFF model most closely models G throughout the year across all sites when the thermal inertia parameter is scaled.

and ALEXI BIAS and RMSE have the widest ranges over grassland, while ALEXI and SEBS BIAS and RMSE have the widest range for evergreen needleleaf forests.

The daily G models show more varied performance among the different PFTs (Figures 6 and 7). The T-DIFF model consistently results in the lowest RMSE and BIAS and exhibits comparable explanation of variance

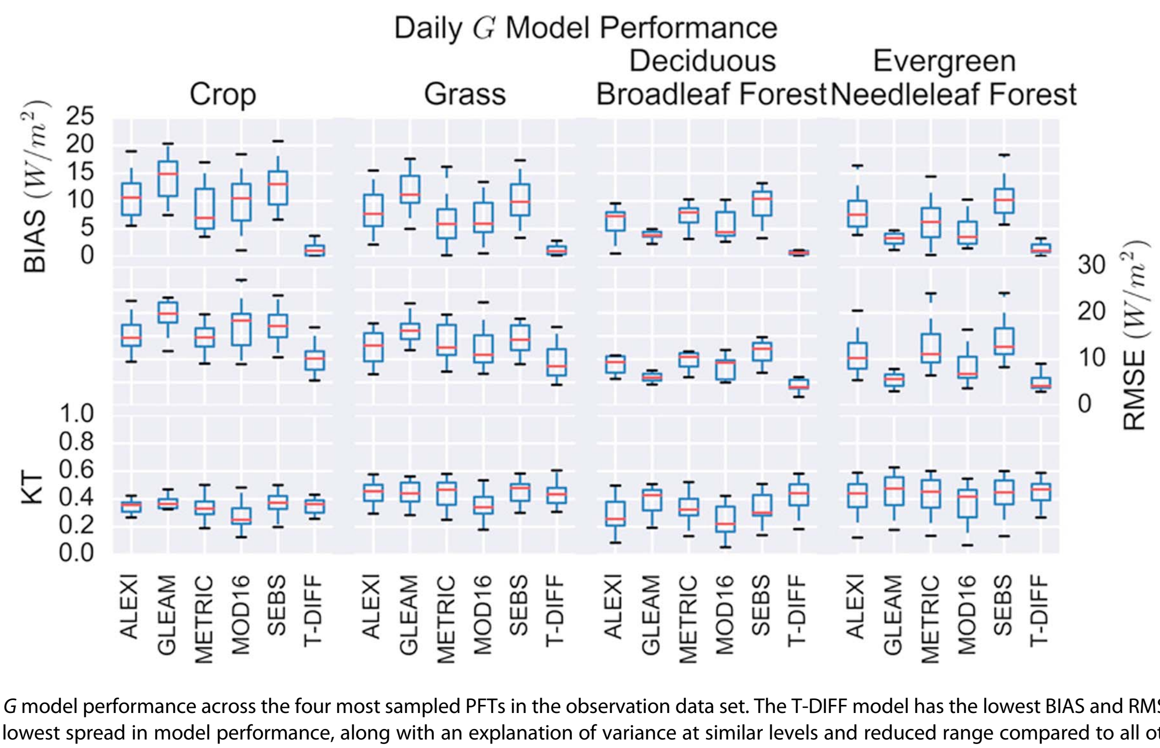

Figure 7. Daily G model performance across the four most sampled PFTs in the observation data set. The T-DIFF model has the lowest BIAS and RMSE across all sites along with the lowest spread in model performance, along with an explanation of variance at similar levels and reduced range compared to all other models.

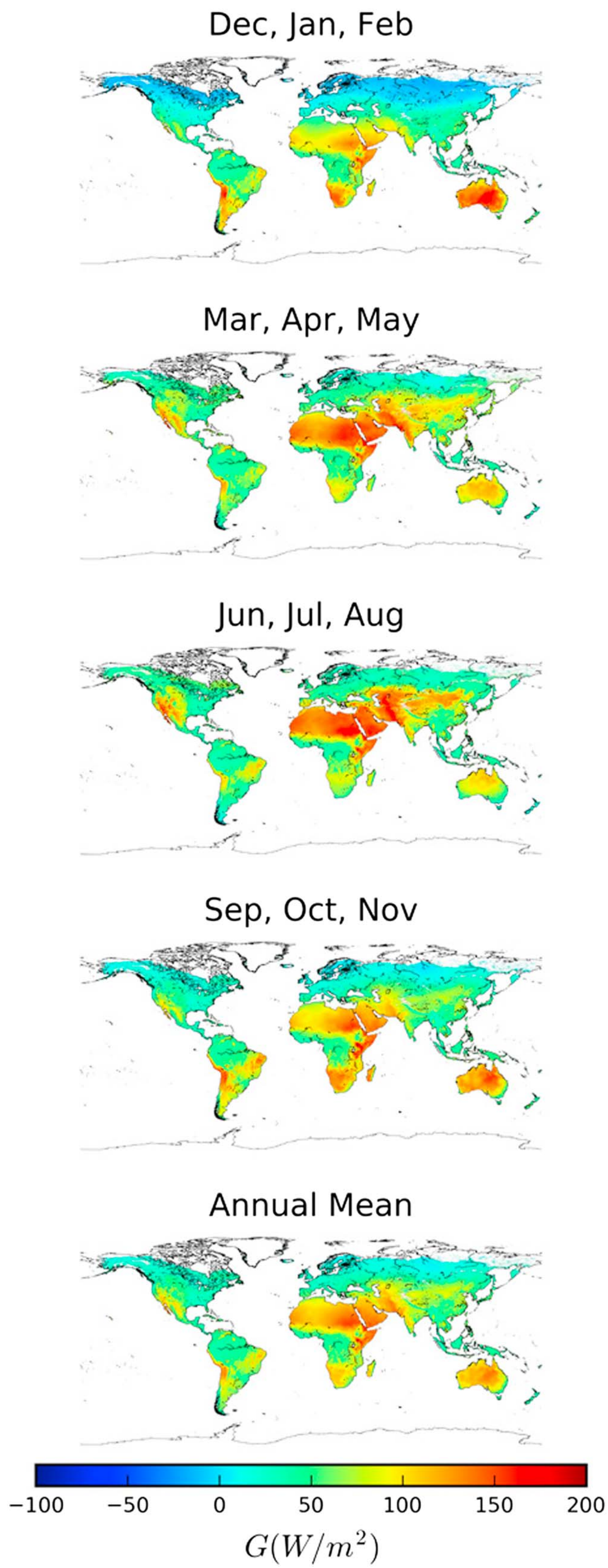

Figure 8. Seasonal and annual multimodel mean instantaneous *G* averaged from 2001 to 2006.

to GLEAM, MOD16, ALEXI, METRIC, and SEBS for the four land covers (Figure 7). The T-DIFF model has more varied performance for cropland and grassland compared to other land uses due to underestimation of thermal inertia (Figures 6 and 7). We also compare a scaled version of the T-DIFF model at each of the four locations to show that the parameterized coarse resolution thermal inertia inhibits the T-DIFF model's ability to capture high and low *G* values. The scaled version demonstrates the strength of the model's structure by improvement in model performance with local calibration of thermal inertia.

The GLEAM and SEBS models show similar explanation of variance with the highest average Kendall's tau at three of the four PFTs (GRA, CRO, and ENF; Figure 7). Despite high explanation of variance across these particular PFTs, the GLEAM and SEBS *G* models have errors more than twice as large as the T-DIFF over GRA and CRO due to overestimation of daily *G* (Figures 6 and 7). For GLEAM, model error improves in deciduous forest and evergreen forest (Figures 6 and 7). The MOD16 model shows the widest range in model error across both grassland and cropland covers, with the highest error for one site out of all the models, most likely from setting $G = 0$ at a location where this is not appropriate as seen in the annual plot of grassland in Figure 6. For evergreen forest and deciduous forest, the MOD16 model shows reduced error and interquartile range in errors but exhibits a wide range in explanation of variance. The differences in model formulation yield a wide range of results for *G* estimation across these different biomes and climates.

3.2. Spatial and Seasonal Model Intercomparison

We model *G* from 2001 to 2006 at 5 km globally using MODIS NDVI, LST, and NCEP net radiation. The 2001–2006 model average midmorning *G* is not negligible in all areas globally (Figure 8). Areas with dense vegetation such as

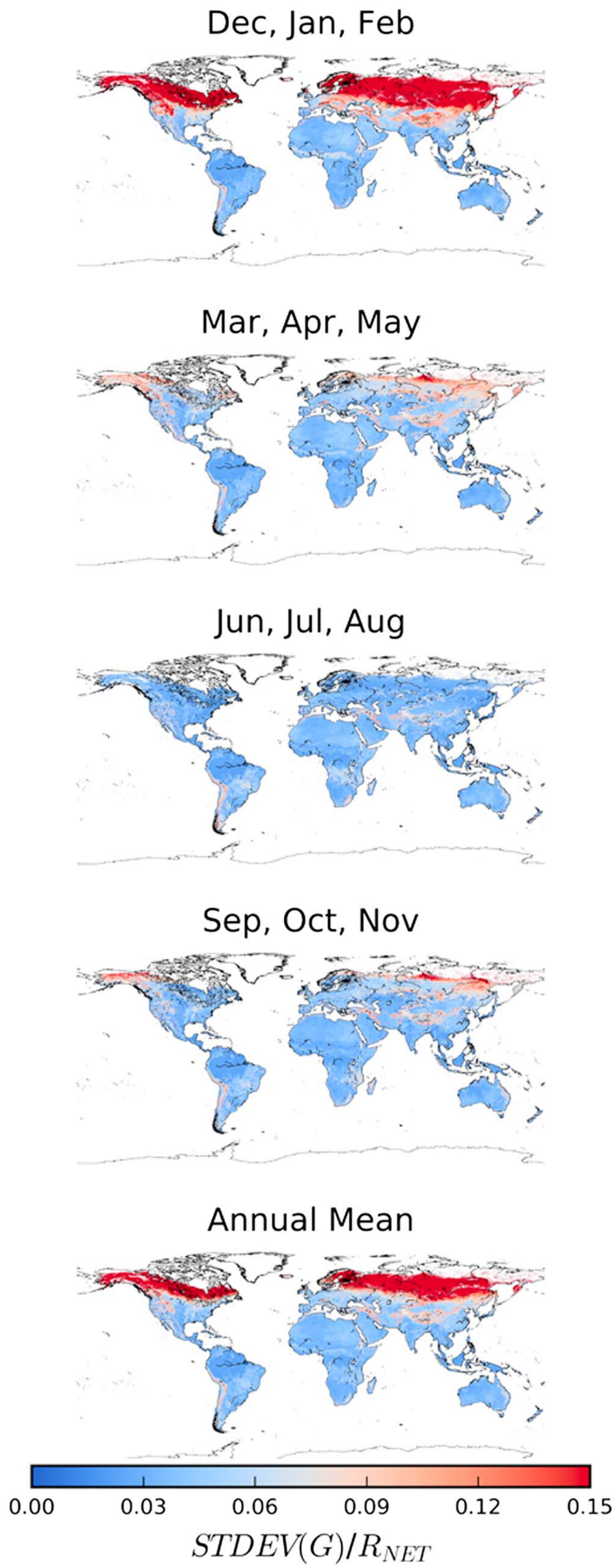

Figure 9. Average seasonal *G* model uncertainty from 2001 to 2006. The red areas indicate the greater model disagreement.

the Amazon and boreal forested regions exhibit low *G*, but in areas with little vegetation *G* is greater than 150 Wm^{-2}. Global *G* is lowest during the boreal winter and is at maximum during the boreal fall. High latitudes during winter months have the lowest *G*. Less vegetated regions, such as the southwestern United States, the Saharan desert, the Arabian Peninsula, central Australia, and southern Africa, show the highest average modeled *G* during warmer months. Modeled *G* differs more at low vegetation cover compared to high vegetation cover (Figure 2). A spatiotemporal comparison highlights these seasonally driven model differences globally. We evaluate the *G* model uncertainty from 2001 to 2006 by normalizing the multimodel standard deviation of *G* by the mean R_{NET} for each season (Figure 9). During December, January, and February the largest regions of model disagreement are boreal Canada, Siberia, the southwest United States, and high-mountain Asia. Over the boreal spring and summer (March, April, and May and June, July, and August) models generally agree globally. For the months of September, October, and November, similar to the winter months high-latitude areas in the northern hemisphere experience more disagreement. The regions of disagreement are predominantly areas where bare soil, dormant vegetation, low radiation, and low temperature drive model divergence. The model formulations (Table 1) which quantify G/R_{NET} from empirical relationships to vegetation cover or temperature disagree most under periods of low vegetation cover and low R_{NET} (Figure 2). For areas with peak seasonal greenness the models converge to estimate similar magnitudes of G/R_{NET}.

4. Discussion

4.1. Strengths and Weaknesses of Models for Instantaneous and Daily *G* Calculations

Despite potential scaling issues in relating remote sensing footprints to in situ data, we find that most models provide reasonable estimates of G across a variety of land uses and climates. For instantaneous estimation of G, we identify the linear relationship to LST used in MOD16 to provide the lowest error across all the sites (minimum average BIAS and RMSE). However, the current model formulation results in unrealistic G for regions with very hot ($T_{an} > 25°C$) or cold ($T_{an} < -5°C$) temperatures or low diurnal temperature swings ($T_{dif} < 5°C$) and can result in misclassification for some locations (Figure 6). This approach assumes that vegetation cover does not impact G for these regions, which is contradictory to much preliminary work developing G models [*Choudhury et al.*, 1987; *Kustas and Daughtry*, 1990; *Kustas et al.*, 1993]. Therefore, the METRIC formulation presents an attractive alternative where G is modeled using vegetation cover for areas of moderate to high vegetation (LAI > 0.5), a common characteristic of models that capture more day-to-day instantaneous G variance, while still using a linear relationship to temperature for barren and sparsely covered areas (LAI < 0.5). The METRIC G model exhibits lower errors over grasslands and croplands compared to other land covers, which was the environment under which the original empirical relationship was calibrated [*Allen et al.*, 2007]. Initial exploration into recalibration for the different PFTs indicated parameters for this formulation convergence for tall (mixed forest, DBF, evergreen broadleaf forest, and ENF) and short canopy (GRA and CRO) covers, respectively. This suggests that canopy structure not just leaf cover density influences G at these time scales [*Reginato et al.*, 1985; *Clothier et al.*, 1986; *Miralles et al.*, 2011]. For midmorning instantaneous G calculations, the METRIC formulation shows the most potential from parameter recalibration for a more robust representation by incorporating canopy height.

For G needed at daily or longer time scales the T-DIFF model bests all other models. The T-DIFF model exhibits low errors while explaining a similar level of variance to models forced by R_{NET}. We find the optimal thermal inertia parameters to minimize error are greater than reported values for daily time steps (Figure 6) [*Bennett et al.*, 2008]. Despite the difference, the integrated time difference in nighttime LST is the best method to model G at these time scales. Further work on parameterization of the soil thermal properties and the incorporation of changing properties such as surface soil moisture creates the opportunity to enhance model performance globally [*Idso et al.*, 1975; *Santenello and Friedl*, 2002].

4.2. Mechanisms That Control *G* Across Instantaneous and Aggregated Time Scales

The instantaneous and daily G models each share a common set of variables: R_{NET}, vegetation properties (NDVI, LAI, and f_c), and LST (Table 1). These variables are used to model the environmental processes that control G. Vegetation impacts the magnitude of G in multiple ways. Dense vegetation reduces G through shading the ground from incoming radiation and by buffering temperature gradients in areas with high rates of ET. We analyze the results from the instantaneous and daily model evaluations to determine if these processes are appropriately represented.

For instantaneous models the highest explanation of day-to-day variability is achieved by GLEAM, a model forced with R_{NET} (Table 2). The models forced by only LST miss higher-frequency variability in G that is captured by models forced with R_{NET}; this is in contrast to previously published correlation coefficients for a model forced only by LST data ($R^2 = 0.90$) [*Jacobsen and Hansen*, 1999]. Models that use vegetation properties to scale R_{NET} to G (ALEXI and SEBS) result in more error compared to models that use both vegetation properties and LST to estimate G (METRIC and MOD16). Comparing the explanation of variance between the GLEAM scalar approach and the models that only use vegetation phenology (ALEXI and SEBS) would imply that vegetation changes might not even play a major role in the calculation of instantaneous G (Table 2 and Figure 5). This suggests that the phenological changes in vegetation are less indicative of seasonal changes in G compared to seasonal fluctuations in R_{NET} and LST or that LST implicitly incorporates phenological changes from the impact of vegetation on LST. This finding is contrary to early work built on the foundation that G/R_{NET} is proportional to vegetation density [*Choudhury et al.*, 1987; *Kustas et al.*, 1993]. However, for specific locations, such as the cropland cover where large changes of vegetation occur with cultivation practices, this is not the case (Figure 4). Here the GLEAM model deviates from the observation and the other G models between May and June due to a change in vegetative cover unaccounted for by annually invariant canopy height forcing data sets (Figure S11). Not every model applies vegetation properties to scale R_{NET}; instead, models rely on LST data to reflect changes in canopy conditions appropriate for modeling G.

For daily or longer times, temperature is the most important variable for accurately modeling G. As evidenced above the T-DIFF model outperforms all other models, suggesting that the integrated time differential in nighttime LST is a sound method to model G at daily time steps. Additionally, the T-DIFF structure implicitly accounts for energy storage and conserves energy by summing both increases and decreases in the time derivative of LST to calculate G. For these time scales, models that reduce R_{NET} to G using vegetation attributes (NDVI, LAI, or canopy height) have errors close to twice as large as T-DIFF and do not explain considerably more variability in daily G (Table 3 and Figure 7). The poor performance by these models might be traced to model formulations that neglected energy storage. Energy storage is important for aggregated times but not necessary for snapshots of energy partitioning. At daily or longer frequencies temperature fluctuations explain just as much variability in G as R_{NET} does. The GLEAM G model, which directly scales R_{NET} to G, only explains slightly more daily G variability than T-DIFF (Table 3). The T-DIFF model error suggests that other factors, such as soil properties, may help explain the magnitude of G, but parameterizing these uncertainties using PFT specific calibrations is an appropriate way to handle the uncertainty surrounding spatially explicit specific soil properties. However, since a relatively low amount of daily G variability is explained by this model ($KT = 0.36$), the assumption that the thermal inertia calibrated for this model evaluation remains constant across a year warrants further exploration. Future work to calibrate soil thermal inertia with soil moisture might enhance model performance.

We find that G models forced by R_{NET} better portray diel variability while models forced by temperature changes better capture seasonal variation. Since R_{NET} and temperature are largely independent, we hypothesize that a model incorporating both variables will best represent the environmental processes that drive G, such as absorption of incoming R_{NET} and energy gain or loss through the temperature gradient. Additionally, we posit that models that incorporate physical principles such as conservation of energy will outperform models that do not. Taking these environmental processes into account might offer a path toward G model reformulation and improvement.

4.3. Impact of *G* Uncertainty on Global ET

The G model disagreement is highest for areas and seasons when the magnitude of global ET is lowest. We quantify G uncertainty as a proportion of R_{NET}. ET is directly impacted by errors in available energy ($R_{NET} - G$ for evaporation or R_{NET} for transpiration) in both the Penman-Monteith and Priestley-Taylor equations. Depending on the radiation partitioning method for evaporation and transpiration for each algorithm the magnitude of the impact will differ across models. The spread in model disagreement is greatest in seasons with low R_{NET} and cold weather and in areas with low vegetation (Figure 2). Snow cover negatively impacts remote sensing observations of vegetation and should decrease G due to lower net radiation as a result of a high albedo and smaller thermal inertia of snow relative to soil [*Bennett et al.*, 2008]. The spatiotemporal occurrences of minimum ET occur across the seasons and areas where model differences in G/R_{NET} are largest. Therefore, the timing of the differences in G/R_{NET} would have minimal impact on annual magnitude of ET models, especially because of smaller differences in available energy during seasons of peak ET (i.e., June, July, and August) [*Vinukollu et al.*, 2011; *Miralles et al.*, 2016]. However, during periods of low energy the importance of G is amplified since it is a larger component of available energy ($R_{NET} - G$) and subsequently the impact on ET. The occurrences of the largest modeled G disparities highlight seasons and areas where mechanistic differences in modeled ET and H would be most difficult to disentangle. Despite potential limitation to the impact on ET estimates globally, this analysis reveals regions ripe for G model refinement and demonstrates the importance to choose and calibrate the correct G model for regional studies, especially those focusing on the arctic regions (December-January-February and September-October-November; Figure 9). Moreover, the limited performance by each instantaneous G model compared to in situ observations reveals a great opportunity to reduce uncertainty in ET model performance from uncertain G by recalibrating models to broader more robust data sets.

5. Conclusions

This study evaluates a suite of different G remote sensing methods in current ET models. Six different G models from highly regarded ET algorithms were compared. Previous studies have not evaluated G models against a global set of in situ observations or investigated seasonal model divergence globally. We use in situ G observations from the FLUXNET La Thuile synthesis data set in combination with MODIS vegetation and LST data to evaluate model performance across 88 sites globally. Global G flux estimates produced using MODIS NDVI and LST data along with NCEP R_{NET} were evaluated globally from 2001 to 2006. We identify which G

models perform best globally, which models perform best across different plant functional types, and the areas with the largest seasonal disagreement among *G* models.

The results from this study uncover potential for improvement in energy balance closure at the site level through leveraging remote sensing data, provide guidance on how best to model *G* at different time scales for different land covers, and identify regional and seasonal model biases with implications for global ET and *H* estimates. Poor model performance at some locations (KT < 0) reveals potential issues of comparing models to in situ *G* observations. These potential issues include the scaling of situ observations to remote sensing footprints or the observation technique used to measure *G*, the depth of heat flux plate, and how soil heat storage above the sensor is modeled using soil thermocouples. For regional studies, utilizing finer spatial resolution remote sensing products would help minimize scaling issues in heterogeneous areas such as croplands and harvested forests. Much work is still needed to enhance understanding on how limited point measurements impact scaling this flux to much larger areas. Additional high-fidelity observations at high latitudes, where model disagreement is greatest (Figure 9), would provide opportunity to scrutinize and improve *G* models under low-energy conditions.

The *G* models that only rely on vegetation may introduce seasonal biases in areas with large phenology swings, and areas with small vegetation changes may mask a seasonal cycle directly observed by LST. The *G* models that only use LST may overestimate *G* in areas with tall canopies and/or dense vegetation cover. Future work should incorporate physical principals and the environmental processes that control *G* and optimize the best structural formulations to many land uses and canopy heights. Preliminary results for these efforts suggest that calibration by PFT delivers optimal performance for both instantaneous and daily estimates. Despite often being the smallest term in the surface energy budget, this study reveals that *G* is not negligible and warrants appropriate representation in ET algorithms.

Acknowledgments
This work used eddy covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux (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, OzFlux, TCOS-Siberia, and 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, Environment Canada, and U.S. Department of Energy and the database development and technical support from Berkeley Water Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California—Berkeley, and the University of Virginia. Research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Support was provided by NASA's Science Utilization of the Soil Moisture Active-Passive Mission (SUSMAP) and NASA's Earth and Space Science Fellowship (NESSF). Code used for this analysis will be available upon requests to corresponding author. All data used in this study are open access and available through FLUXNET or NASA. The MOD13C1 and MOD11C1 were retrieved from the online Data Pool, courtesy of the NASA EOSDIS Land Processes Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota.

References

Allen, R. G., M. Tasumi, and R. Trezza (2007), Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)—Model, *J. Irrig. Drain. Eng.*, *133*(4), 380–394.

Allen, R. G., et al. (2015), EEFlux: A Landsat-based evapotranspiration mapping tool on the Google Earth Engine. 2015 ASABE/IA Irrigation Symposium, Emerging Technologies for Sustainable Irrigation – A tribute to the career of Terry Howell, Sr. Conference Proceedings.

Anderson, M. C., J. M. Norman, J. R. Mecikalski, J. P. Otkin, and W. P. Kustas (2007), A climatological study of evapotranspiration and moisture stress across the continental U.S. based on the thermal remote sensing: I. Model formulation, *J. Geophys. Res.*, *112*, D10117, doi:10.1029/2006JD007506.

Baldocchi, D., et al. (2001), FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, *Bull. Am. Meteorol. Soc.*, *82*, 2415–2434, doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2.

Bennett, W. B., J. Wang, and R. L. Bras (2008), Estimation of global ground heat flux, *J. Hydrometeorol.*, *9*, 744–759.

Cammelleri, C., G. la Loggia, A. Loggia, and A. Maltese (2009), Critical analysis of empirical ground heat flux equations on a cereal field using micrometeorological data, *Proc. SPIE*, *7472*, 747225, 1–12.

Chen, Y., et al. (2014), Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China, *Remote Sens. Environ.*, *140*, 279–293.

Choudhury, B. J., S. B. Idso, and R. J. Reginato (1987), Analysis of an empirical model for soil heat flux under a growing wheat crop for estimating evapotranspiration by an infrared-temperature based energy balance equation, *Agric. For. Meteorol.*, *39*(4), 283–297.

Clothier, B. E., K. L. Clawson, P. J. Pinter, M. S. Moran, R. J. Reginato, and R. D. Jackson (1986), Estimation from soil heat flux from net radiation during the growth of alfalfa, *Agric. For. Meteorol.*, *37*, 319–329.

Daughtry, C. S. T., W. P. Kustas, M. S. Moran, P. J. Pinter, R. D. Jackson, P. W. Brown, W. D. Nichols, and L. W. Gay (1990), Spectral estimates of net radiation and soil heat flux, *Remote Sens. Environ.*, *32*, 111–124.

Ershadi, A., M. F. McCabe, J. P. Evans, N. W. Chaney, and E. F. Wood (2014), Multi-site evaluation of terrestrial evapotranspiration models using FLUXNET data, *Agric. For. Meteorol.*, *187*(15), 46–61.

Fisher, J. B., K. P. Tu, and D. Baldocchi (2008), Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites, *Remote Sens. Environ.*, *112*(3), 901–919.

Holmes, T. R. H., M. Owe, R. A. M. De Jeu, and H. Kooi (2008), Estimating the soil temperature profile from a single depth observation: A simple empirical heatflow solution, *Water Resour. Res.*, *44*, W02412, doi:10.1029/2007WR005994.

Idso, S. B., T. J. Schmugge, R. D. Jackson, and R. J. Reginato (1975), The utility of surface temperature measurements for the remote sensing of surface soil water status, *J. Geophys. Res.*, *80*, 3044–3049, doi:10.1029/JC080i021p03044.

Irmak, A., R. K. Singh, E. A. Walter-Shea, S. Verma, and A. E. Suyker (2011), Comparison and analysis of empirical equations for soil heat flux for different cropping systems and irrigation methods. Papers in Natural Resources Paper 334. [Available at http://digitalcommons.unl.edu/natrespapers/334.]

Jacobsen, A., and B. U. Hansen (1999), Estimation of the soil heat flux/net radiation ratio based on spectral vegetation indexes at high latitude Arctic areas, *Int. J. Remote Sens.*, *20*(2), 445–461, doi:10.1080/014311699213532.

Jiménez, C., et al. (2011), Global intercomparison of 12 land surface heat flux estimates, *J. Geophys. Res.*, *116*, D02102, doi:10.1029/2010JD014545.

Kustas, W. P., and C. S. T. Daughtry (1990), Estimation of the soil heat flux/net radiation ratio from spectral data, *Agric. For. Meteorol.*, *49*, 205–223.

Kustas, W. P., C. S. T. Daughtry, and P. J. Van Oevelen (1993), Analytical treatment of the relationships between soil heat flux/net radiation ratio and vegetation indices, *Remote Sens. Environ.*, *46*, 319–330.

Kustas, W. P., J. H. Prueger, J. L. Hatfield, K. Ramalingam, and L. E. Hi (2000), Variability of soil heat flux from a mesquite dune site, *Agric. For. Meteorol.*, *103*, 249–264.

McCabe, M., et al. (2013), in *Global-Scale Estimation of Land Surface Heat Fluxes from Space: Product Assessment and Inter-Comparison, Remote Sensing of Energy Fluxes and Soil Moisture Content*, edited by G. P. Petropoulos, pp. 538, CRC Press, Taylor & Francis Group.

Michel, D., et al. (2016), The WACMOS-ET project—Part 1: Tower-scale evaluation of four remote sensing-based evapotranspiration algorithms, *Hydrol. Earth Syst. Sci.*, *20*, 803–822.

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

Miralles, D. G., et al. (2016), The WACMOS-ET project—Part 2: Evaluation of global terrestrial evaporation data sets, *Hydrol. Earth Syst. Sci.*, *20*, 823–842.

Monteith, J. L. (1965), Evaporation and the environment, *Symp. Soc. Exp. Biol.*, *19*, 205–234.

Monteith, J. L. (1973), *Principles of Environmental Physics*, Edward Arnold Press, London.

Mu, Q., M. Zhao, and S. W. Running (2011), Improvements to a MODIS global terrestrial evapotranspiration algorithm, *Remote Sens. Environ.*, *115*, 1781–1800.

Mueller, B., et al. (2013), Benchmark products for land evapotranspiration: LandFlux-EVAL multi-dataset synthesis, *Hydrol. Earth Syst. Sci.*, *17*, 3707–3720.

Penman, H. L. (1948), Natural evaporation from open water, bare soil and grass, *Proc. R. Soc. London Ser. A*, *193*, 120–146.

Priestley, C. H., and R. J. Taylor (1972), Assessment of surface heat flux and evaporation using large-scale parameters, *Mon. Weather Rev.*, *100*, 81–92.

Reginato, R. J., R. D. Jackson, and P. J. Pinter (1985), Evapotranspiration calculated from remote sensing multispectral and ground station meteorological data, *Remote Sens. Environ.*, *18*, 75–89.

Ross, J. (1976), Radiative transfer in plant communities, in *Vegetation and the Atmosphere*, edited by J. L. Monteith, pp. 13–56, Academic Press, London.

Santenello, J. A., and M. A. Friedl (2002), Diurnal variation in soil heat flux and net radiation, *J. Appl. Meteorol.*, *42*, 851–862.

Seguin, B., and B. Itier (1983), Using midday surface temperature to estimate daily evaporation from satellite thermal IR data, *Int. J. Remote Sens.*, *4*, 371–384.

Su, Z. (2002), The Surface Energy Balance System (SEBS) for turbulent heat fluxes, *Hydrol. Earth Syst. Sci. Discuss.*, *6*(1), 85–100.

Twine, T. E., W. P. Kustas, J. M. Norman, D. R. Cook, P. R. Houser, T. P. Meyers, J. H. Prueger, P. J. Starks, and M. L. Wesely (2000), Correcting eddy-covariance flux underestimates over a grassland, *Agric. For. Meteorol.*, *103*(3), 279–300.

Vinukollu, R. K., R. Meynadier, J. Sheffield, and E. F. Wood (2011), Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches, *Remote Sens. Environ.*, *115*(3), 801–823.