Version of Record: https://www.sciencedirect.com/science/article/pii/S0168192318301102
Manuscript_3d80d5127989b064be6770609fad8572

Comprehensive evaluation of empirical algorithms for estimating land surface evapotranspiration

Corinne Carter
Shunlin Liang*

Department of Geography, University of Maryland, College Park, MD 20742, USA

*Corresponding author at: Department of Geography, University of Maryland, College Park, College Park, MD 20742.
Email address: sliang@umd.edu

ABSTRACT

Many empirical algorithms for obtaining evapotranspiration (ET) from vegetation indices (VIs) have been developed, but there has been little work comparing these algorithms to each other or deriving coefficients for them using large data sets for training and validation. Twelve different vegetation index-based regression algorithms for retrieval of ET on a daily basis are reviewed and evaluated here. New coefficients have been derived for four of these algorithms using data from 181 Ameriflux and Fluxnet2015 sites and 1km MODIS subsets centered at each site location. Algorithm validation with previously published and new coefficients was

© 2018 published by Elsevier. This manuscript is made available under the Elsevier user license
https://www.elsevier.com/open-access/userlicense/1.0/

performed using one year of data from each Ameriflux and Fluxnet2015 site. There was a wide range of performance of these algorithms, with the median R^2 by site in the 0.6 to 0.7 range, median root mean square error (RMSE) about 25 W/m^2 and median bias within 10 W/m^2. When algorithm coefficients were re-derived, the RMSE and bias of the worst-performing algorithms were largely reduced, but R^2 was little changed. Agricultural and wetland sites had a low bias across most of the algorithms, and wetland sites had a higher RMSE. When several of the algorithms were re-tuned to obtain coefficients specific to each surface type, the biases of the agricultural and wetland sites were reduced to those more typical of other site types, and RMSE for agricultural and wetland sites was also reduced. The effects of linear interpolation of VIs to obtain daily LE and interpolation over periods of rapid VI change at agricultural sites were examined. No significant algorithm performance degradation was found in either case. It is recommended to use more detailed algorithms when possible, with inclusion of net radiation as a parameter along with VI at a minimum.

Research highlights:

- 12 regression algorithms tested with MODIS VI, MODIS albedo, and Fluxnet tower data
- Median statistics of best algorithms: R^2 0.6 to 0.7, RMSE ~ 25 W/m^2
- Re-deriving coefficients reduced RMSE and bias for some algorithms
- Re-deriving coefficients improved performance for wetland and agricultural sites

Keywords: Evapotranspiration; remote sensing; vegetation index; regression algorithms; MODIS; Fluxnet

1. Introduction

1.1. Background and motivation

Increasing demands are being made on water resources globally, and this trend is expected to continue due to anticipated changes in global climate and hydrology (Field et al. 2014). Evapotranspiration (ET) is a major component of the global water cycle and its measurement is also used in water resources, agricultural, and ecosystem health monitoring. Determination of ET on global and regional scales is crucial to understanding trends in the global hydrological cycle (Zeng et al. 2012; Jiménez et al. 2011; Jung et al. 2010; Wang et al. 2010b) and regional impacts of global hydrological change (e.g. Du et al. 2017; Spinoni et al. 2017; Garner et al. 2017; Haileslassie et al. 2009).

A broad review of LE measurement methods has been performed by Wang and Dickinson (2012). Two frequently used methods can provide ET on scales of tens of meters. Weighing lysimeters provide the most direct measurement of ET, and are used to calibrate ET found through other methods (Liu et al. 2017; Hirschi et al. 2017). The frequently-used method for obtaining LE presented in the Food and Agricultural Organization of the United Nations (FAO) Irrigation and Drainage Paper 56 (R. G. Allen 1998) (FAO56) depends only on meteorological observations and crop coefficients estimated based on surface conditions. The FAO56 method has the advantage of not depending on any instruments besides those used to collect standard weather observations. The lysimeter and FAO56 methods are most useful for estimating ET over scales where meteorological and land cover conditions are relatively uniform, such as that of an individual agricultural field.

ET measurements from eddy correlation flux towers such as the Fluxnet network (Baldocchi et al. 2001) typically have footprints on the order of hundreds of meters. This spatial scale is convenient for many purposes, including validation of ET obtained through remote sensing. There is an issue with energy balance closure (Foken 2008) for flux tower measurements, which is usually resolved by assuming conservation of energy at the surface and a consistent Bowen ratio between measured and actual sensible and latent heat fluxes. With this correction, flux tower measurements are estimated to be accurate within 20% or less (Perez-Priego et al. 2017; Hirschi et al. 2017; Wang and Dickinson 2012). However, they are limited in their applicability due to their relatively small scale and restricted areal coverage, as well as by the significant overrepresentation of northern hemisphere midlatitude sites. In addition, there are many sites with temporal records of a few years or less, and where there is no ongoing data collection. As a result, there is a great deal of interest in remote sensing of ET at larger spatial scales and in more remote areas.

There are many remote sensing methods for retrieving ET available (Zhang et al. 2016; Wang and Dickinson 2012; Kalma, McVicar, and McCabe 2008) The methods available require various combinations of visible and infrared band data or their derived products such as albedo, land surface temperature, or vegetation index. They also differ in the degree to which the land surface energy and moisture transport processes are modeled explicitly, and with which formulations. Some models, such as SEBAL and its descendants (Bastiaanssen et al. 1998), are based on finding the latent heat transfer rate from the surface (LE = λET, with ET of 1 mm/ day = LE of 26.3 W/m^2) residual of the surface energy balance

$$LE = R_n - H - G \tag{1}$$

where R_n is the net radiation at the surface, H is the sensible heat transfer rate, and G the rate of change in ground heat storage. These models consider the entire soil and canopy surface in bulk (one source models) or treat the soil and canopy separately (two source models). Energy balance residual models rely on thermal band observations as indicators of surface temperature. The two source time integrated model TSTIM, later renamed ALEXI (Anderson et al. 2007; Anderson 1997), relies on multiple daily surface temperature measurements, as a smaller range of surface temperature is indicative of greater moisture availability.

The Penman-Monteith formulation of turbulent heat transfer (Monteith 1965) is used as a basis for other methods of retrieving LE from remote sensing, such as that of Mu et al. (2011), now used to generate the global MOD16 product from MODIS data. The earlier Penman (1948) formulation was used as a basis for the model developed by Wang et al. (2010a). Another turbulent flux parameterization, the Priestley-Taylor formula (Priestley and Taylor 1972) has been used in combination with net radiation and vegetation indices (Yao et al. 2015, 2013; Fisher et al. 2008) to obtain ET. In the case of the Yao et al. (2015, 2013) and Wang et al.

(2010a) studies, the turbulent flux transfer parameterizations were used as a basis for formulas to which empirical regression coefficients were fitted.

There are also many simpler regression formulas that have been developed for estimation of ET. It has been found (Jiménez et al. 2011) that empirical regression formulas can produce ET values that are comparable in accuracy to more complex models, without as much computational demand or requirements for specific expertise. Many of these regression formulas are based on vegetation indices (VI), as reviewed by Glenn et al. (2010). The most frequently used vegetation indices in ET algorithms are the normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI). These ratios between near infrared, red, and blue band reflectances (ρ_{NIR}, ρ_{red}, and ρ_{blue} respectively) are as follows:

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} \tag{2}$$

$$EVI = G_{EVI} \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + C_1 \cdot \rho_{red} + C_2 \cdot \rho_{blue} + L} \tag{3}$$

The standard EVI product calculated from MODIS data has the constants G_{EVI}, C_1, C_2, and L set to values of 1.0, 6.0, 7.5, and 2.5 respectively.

Vegetation indices have several advantages for use in evapotranspiration algorithms. They are available from multiple instruments and at resolutions down to tens of meters. They have a high degree of consistency between instruments (Brown et al. 2006; Steven et al. 2003) Vegetation indices typically change on time scales of weeks to months, so interpolation can be used between observations separated by multiple days with some confidence. Algorithms that include a dependence on surface temperature are likely to be more responsive on shorter time scales, but the faster rate of change of surface temperature makes interpolation between observations more problematic. Overall, vegetation index-based methods have the advantages of simplicity, utility under a wide range of conditions, and resilience in the presence of data gaps.

Little work has been done evaluating these vegetation index-based algorithms under different conditions or comparing them to each other or to LE values derived through other methods. The goal of this paper is to provide a comprehensive evaluation of a range of VI- based evapotranspiration algorithms, identifying their strengths and weaknesses relative to each other.

1.2. Description of VI-based algorithms to be evaluated

A number of authors have proposed formulas for LE based on vegetation indices, ranging from highly simplified, depending only on the VI value with no additional data, to more complex formulas requiring ancillary data such as net radiation, surface and atmospheric temperatures, and other meteorological variables. All formulas to be evaluated in this paper are summarized in Table 1.

Table 1: Vegetation index based algorithms reviewed and compared, with full algorithm names and short names used to identify the algorithms in the figures. Key to variables: NDVI- Normalized difference vegetation index, EVI- Enhanced vegetation index, R_n- Net radiation at surface, G- Ground heat storage, T_{a_avg} – Daily average atmospheric temperature, T_{a_max}- Daily maximum atmospheric temperature, T_{a_dTr}- Daily atmospheric temperature range, T_{s_avg}- Daily average surface temperature, T_{s_max}- Daily maximum surface temperature, T_{s_dTr}- Daily surface temperature range, LE_0- Potential evapotranspiration, R_s- Incoming solar radiation at surface, RH- relative humidity, e_s- Saturation water vapor pressure, ws- Wind speed, VPD- vapor pressure deficit.

Algorithm	**Short name**	**Reference**	**Required input data**
Yebra direct (ET)	YET	Yebra et al. (2013)	NDVI or EVI
Yebra evaporative fraction (EF)	YEF	Yebra et al. (2013)	NDVI or EVI, R_n, G
Helman exponential	HEx	Helman et al. (2015)	NDVI or EVI
Helman scaled	HSc	Helman et al. (2015)	EVI, T_{s_avg}
Wang 2007	W07	Wang et al. (2007)	NDVI or EVI, R_n, one of T_{a_avg}, T_{a_max}, T_{s_avg}, or T_{s_max}
Wang/ Liang	WL	Wang and Liang (2008)	NDVI or EVI, R_n, T_{s_dTr}, one of T_{a_avg}, T_{a_max},

			T_{s_avg}, or T_{s_max}
Choudhury/ FAO56	Ch	Choudhury et al. (1994) Allen et al. (1998)	EVI, LE_0
Kamble/ FAO56	Kmb	Kamble et al (2013) Allen et al. (1998)	NDVI, LE_0
Wang 2010	W10	Wang et al. (2010a)	NDVI or EVI, R_s, RH, e_s, ws, T_{a_avg}
Yao 2011	Y11	Yao et al. (2011)	NDVI, R_n, T_{a_avg}, T_{a_dTr}
Yao 2013	Y13	Yao et al. (2013)	NDVI, R_n, G, T_{a_avg}, T_{a_dTr} or T_{s_dTr},
Yao 2015	Y15	Yao et al. (2015)	NDVI, R_n, G, T_{a_avg}, RH, VPD

A total of 12 algorithms, based on 11 separate publications, are reviewed and evaluated in this paper. For each algorithm, Table 1 gives a short name, the source publication(s), and required input data. Some of the publications listed also include other algorithms that depend on remote sensing parameters other than NDVI or EVI, but only the VI-based algorithms are included here.

Two of the algorithms, Yebra ET (Yebra et al. 2013) and Helman exponential (Helman et al. 2015), depend on the vegetation index alone. These algorithms were trained using 16 Fluxnet sites each. The Yebra algorithm sites were distributed over six different land cover types with forest and cropland sites most common, while the Helman algorithms were developed specifically for Mediterranean ecosystems with cropland and grassland sites most represented.

The Yebra ET formula

$$LE_{YET} = a + b * VI \tag{4}$$

is a linear function of a vegetation index VI (NDVI or EVI), while the Helman exponential formula

$$LE_{HEx} = a * \exp(b * VI) \tag{5}$$

is an exponential function of either NDVI or EVI. For each of these algorithms, regression coefficients were found for NDVI and EVI separately.

The Yebra EF formula (Yebra et al. 2013) treats the evaporative fraction

$$EF = LE/(R_n - G) \tag{6}$$

as a linear function of NDVI or EVI, resulting in

$$LE_{YEF} = (R_n - G)(a + b * VI) \tag{7}$$

The Helman scaled algorithm (Helman et al. 2015), trained with the same data set as the Helman exponential algorithm, depends on a EVI and daily mean surface temperature T_{s_avg}, scaled according to:

$$EVI_{scl} = EVI - b \tag{8}$$

$$LST_{scl} = c - (d * T_{s\_avg})$$

$$if\ (LST/e) < LST_{scl}, \qquad LST_{scl} = LST/e$$

then obtaining LE as the product of these scaled parameters:

$$LE_{Hsc} = a * EVI_{scl} * LST_{scl} \tag{9}$$

Wang et al. (2007) and Wang and Liang (2008) have published two empirical algorithms:

$$LE_{W07} = R_n * (a_1 + a_2 * VI + a_3 * T) \tag{10}$$

and

$$LE_{WL} = R_n * (a_1 + a_2 * VI + a_3 * T + a_4 T_{s\_dTr}) \tag{11}$$

respectively. Eight sets of coefficients were derived for each of these formulas, for each possible combination of MODIS NDVI or EVI, and average or maximum daily surface temperature (T_{s_avg}, T_{s_max}), or average or maximum atmospheric temperature (T_{a_avg}, T_{a_max}). The Wang and Liang (2008) formula also includes daily surface temperature range (T_{s_dTr}) as a proxy for moisture availability. These formulas are based on the maximum correlations between LE and other variables measured at eight Bowen ratio tower sites in the US Southern Great Plains, and, in the case of Wang and Liang (2008), four additional eddy correlation tower sites also in the US. In both studies, the strongest correlation was with net radiation, with VI and temperature variables following.

Two of the published formulas parameterize evapotranspiration as a function of the potential evapotranspiration ET_0, or the equivalent latent heat transfer LE_0, defined as the ET that would occur from a standardized, well-watered ground cover given a set of atmospheric conditions. LE_0 is often derived from the standard surface conditions and the Penman-Monteith formula for LE (Monteith 1965):

$$LE = \frac{\Delta(R_n - G) + \rho_a c_p \frac{VPD}{r_a}}{\Delta + \gamma\left(1 + \frac{r_s}{r_a}\right)} \tag{12}$$

where Δ is the derivative of saturation vapor pressure with temperature, ρ_a is the density of air, c_p the specific heat of air at constant pressure, VPD the vapor pressure deficit ($e_s - e_a$, where e_s is the saturation vapor pressure and e_a is actual vapor pressure), γ the psychrometric constant, and r_s and r_a are bulk aerodynamic resistance factors characterizing surface and atmospheric conditions respectively. A frequently-used formula for estimation of ET_0 is given in FAO56 (Allen et al. 1998) After conversion to units of LE, the FAO56 formula becomes

$$LE_0 = 26.3 * \left[\frac{0.408\Delta(R_n - G) + \gamma\left(\frac{900}{T+273}\right) ws * VPD}{\Delta + \gamma(1 + 0.34ws)} \right] \tag{13}$$

where ws represents wind speed.

Choudhury et al. (1994) combined observations of agricultural fields in an arid climate with surface and radiative transfer modeling to obtain a transpiration coefficient as a function of vegetation index. Glenn et al. (2010) proposed neglecting the bare soil evaporation in this formula, resulting in a formula for LE in terms of LE_0:

$$LE_{Ch} = LE_0 \left(1.0 - \frac{EVI_{max} - EVI}{EVI_{max} - EVI_{min}}\right) \tag{14}$$

Choudhury et al. (1994) suggested using EVI_{max} = 0.95 and EVI_{min}=0.05.

Kamble et al. (2013) suggested a linear function of NDVI for obtaining LE based on LE_0, and derived coefficients based on agricultural sites in the US Great Plains:

$$LE_{Kmb} = LE_0(a * NDVI - b) \tag{15}$$

Wang et al. (2010a) developed their formula based on the approach of Penman (1948), estimating LE as consisting of two components, one controlled by available energy and another by atmospheric resistance. They developed the regression formula

$$LE_E = \frac{\Delta}{\Delta+\gamma} R_s[a_1 + a_2 VI + RHD(a_3 + a_4 VI)] \quad (16)$$

$$LE_A = \frac{\gamma}{\Delta+\gamma} ws * VPD[a_5 + RHD(a_6 + a_7 VI)]$$

$$LE_{W10} = a_8(LE_E + LE_A) + a_9(LE_E + LE_A)^2$$

with an energy control component LE_E dependent on incoming shortwave flux R_s and an atmospheric transmission control component LE_A. RHD represents the relative humidity deficit (as a function of relative humidity RH in percent: (100 – RH) / 100). This regression formula was trained using 64 eddy correlation and Bowen ratio ground stations, with the goal of obtaining globally-applicable coefficients. Unlike many of the other formulas, which contain an R_n or R_n – G term as a measure of available energy at the surface, the Wang formula uses the incoming solar radiation at the surface R_s. R_s may be measured directly, or estimated based on R_n, albedo, temperature, and relative humidity through the formula given in Wang and Liang (2009).

The three Yao et al. formulas considered here (2015, 2013, 2011), like the Wang et al (2010) model, are regressions based on pre-existing physical LE models. The Yao 2011 formula, developed for drought monitoring from a two-source LE model and data from 22 flux tower sites and global radiation and NDVI products, takes the form

$$LE_{Y11} = R_n^2(a_1 NDVI - a_2) + R_n\left(a_3 + a_4 T_{a\_avg} + \frac{a_5}{T_{a\_dTr}}\right) + R_n NDVI\left(a_6 + a_7 T_a + \frac{a_8}{T_{a\_dTr}}\right) \quad (17)$$

where T_{a_dTr} is the daily range of near-surface atmospheric temperature.

The Yao 2013 and Yao 2015 formulas are both based on the Priestley-Taylor (Priestley and Taylor 1972) parameterization, where r_s and r_a are combined into an empirically determined coefficient α with a value of 1.26 representing a well-covered and watered surface and a function f(e) ranging from 0 to 1 representing constraints on LE:

$$LE = \alpha\left(\frac{\Delta}{\Delta+\gamma}\right) f(e) * (R_n - G) \quad (18)$$

The Yao 2013 formula represents each of four separate components of LE through individual Priestley-Taylor parameterizations. These are a canopy transpiration component LE_c, a soil evaporation component LE_s, and components for evaporation from wet canopy and soil surfaces, LE_{ic} and LE_{ws}:

$$LE_{Y13} = LE_c + LE_s + LE_{ic} + LE_{ws} \quad (19)$$

$$LE_c = \alpha\left(\frac{\Delta}{\Delta+\gamma}\right)(1 - f_{wet}) f_v f_T R_{nc}$$

$$LE_s = \alpha\left(\frac{\Delta}{\Delta+\gamma}\right)(1 - f_{wet}) f_{sm} (R_{ns} - G)$$

$$LE_{ic} = \alpha\left(\frac{\Delta}{\Delta+\gamma}\right) f_{wet} R_{nc}$$

$$LE_{ws} = \alpha\left(\frac{\Delta}{\Delta+\gamma}\right)f_{wet}(R_{ns}-G)$$

The parameters f_{sm} and f_T represent soil moisture and temperature constraints respectively, f_v is fractional vegetation cover, f_{wet} is relative surface wetness, R_{nc} is net radiation to the vegetation canopy, and R_{ns} is net radiation to the soil. These variables are in turn parameterized in terms of vegetation index, daily average temperature, and daily temperature range. Separate sets of coefficients were derived using atmospheric and surface daily temperature ranges.

The Yao (2015) formula, which is similar in its basis to that of Fisher et al. (2008), is also based on the Priestley-Taylor equation, in this case with constraints on all sources of LE combined into one formulation. It was also developed for global applications, and the coefficients were trained with data from 240 Fluxnet sites.

$$LE_{Y15} = \phi\frac{\Delta}{\Delta+\gamma}(R_n-G)\left[a_1+a_2T_{a\_avg}+a_3\left(\frac{RH}{100}\right)^{VPD}+VPD(a_4NDVI-a_5)\right] \tag{20}$$

In summary, a range of formulas for obtaining LE from VI exist with different theoretical bases, degrees of complexity, and other input variables required. Some have forms that have a physical basis, but all ultimately depend on empirical regression for training of coefficients. In most cases they were trained with a limited number of ground sites, so it is desirable to test whether improvements can be made to their performance by using a larger training data set.

2. Data

2.1. Ground-based

A total of 184 flux tower sites were used, 119 from the Ameriflux network (http://ameriflux.lbl.gov) and 65 from the Fluxnet2015 data set (http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/). All available sites with at least 3 continuous years of data were included. Most of the Ameriflux sites were within the United States, with good representation of the latitude range and land cover types of the continental US and Alaska. Eleven of the Ameriflux sites are Canadian, one Mexican, and one Brazilian. The Fluxnet2015 sites are mostly in Europe with some in Asia and Africa, cover a wide range of surface types and climates, but have the northern midlatitude bias typical of flux tower records. A total of 1166 site-years of data from 181 sites was used. The global distribution of these sites is shown in Figure 1. The IGBP surface types represented in the combined Ameriflux and Fluxnet2015 data, the categories used for further analysis here, and the number of sites and total site-years in each category are listed in Table 2.

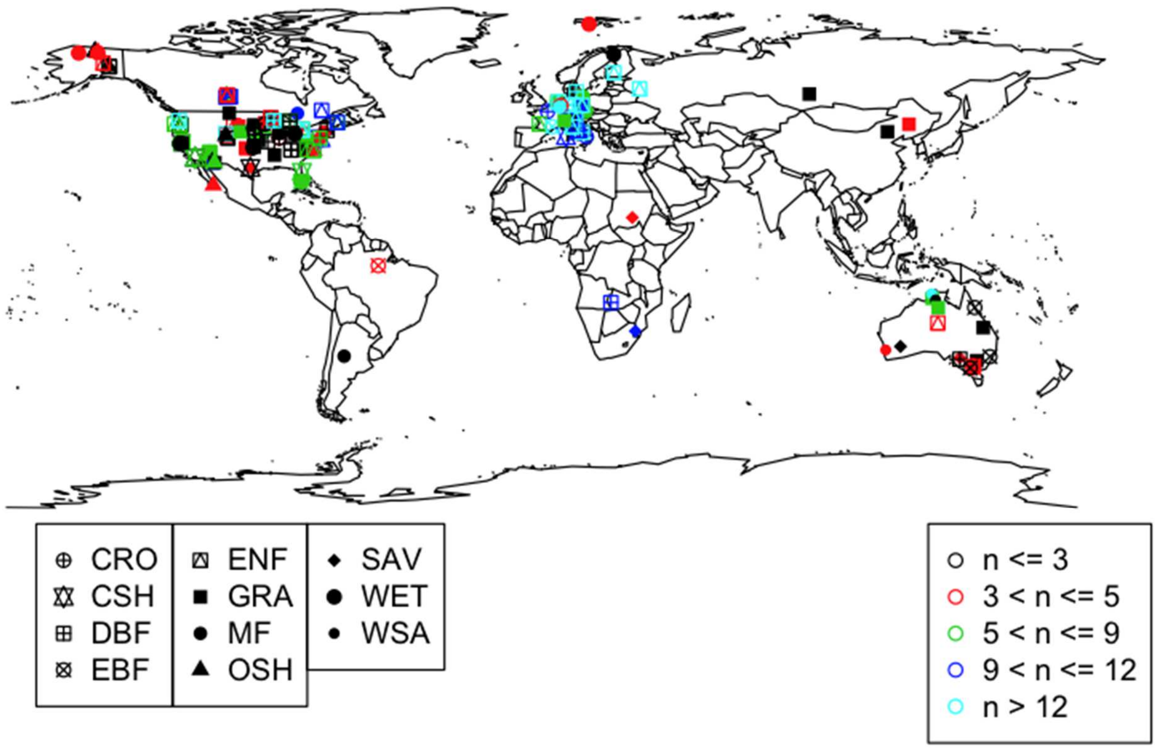


Figure 1: Global distribution of flux tower sites used in this study. Colors of points indicate number of years of data used from each site. Shapes of points indicate IGBP ecosystem type: CRO- crop, CSH- closed shrubland, DBF- deciduous broadleaf forest, EBF- evergreen broadleaf forest, ENF- evergreen needleleaf forest, GRA- grassland, MF- mixed forest, OSH- open shrubland, SAV- savannah, WET- wetland, WSA- woody savannah

NOTE: Figure 1 should be in color electronically but not in print.

Table 2: Land cover type categories used for algorithm evaluation, with IGBP classes included, number of sites available, and total site-years of data used for each.

Category	Included IGBP classes	Number of sites	Total site-years
Agricultural	CRO	23	115
Grassland	GRA	35	181
Deciduous	DBF, DNF, MF	29	228
Evergreen	EBF, ENF	50	392
Savannah	SAV, WSA	13	80
Shrub	CSH, OSH	18	76
Wetland	WET	13	94

The flux tower observations were preprocessed to obtain daily values of LE and all parameters required by the algorithms except for vegetation indices and albedo. For those days with at least 40 of 48 half hourly observations available for all variables, daily mean values of all required meteorological and energy balance variables were calculated. No modeled or gap-filled data were used, so days with insufficient flux tower data are not represented in our analysis. For atmospheric and surface temperatures, daily maximum and minimum values were also found and daily temperature ranges calculated.

2.2. Remote sensing

MODIS Terra NDVI and EVI products (MOD13Q1, Didan 2015) and Terra/ Aqua combined albedo (MCD43A, Schaaf and Wang 2015) time series were obtained for each site, for the same time period as the available flux tower data where it overlaps with the MODIS record. Subsets of each product were obtained from the Oak Ridge National Laboratory DAAC (https://daac.ornl.gov/MODIS/modis.shtml) Standard QC screening was applied. A 1km subset size was used, and all pixels that passed QC screening were included in calculations of mean NDVI, EVI, and albedo. (Preliminary testing with 0 km (same pixel), 1 km, and 3 km subset sizes indicated very little difference in LE algorithm results. Restricting included pixels to those with the same surface type as the central pixel also had a negligible effect.) Under ideal conditions VI is available every 16 days and albedo every 8 days, but longer data gaps exist in some locations due to insufficient high-quality pixels. VI and albedo were both linearly interpolated to generate daily time series.

3. Methods

Each model was first used to calculate LE (LE_{mod}) for each day where sufficient flux tower data was available at every site with the original published coefficients then compared against the ground observation LE (LE_{obs}). The coefficients for each algorithm were then re-derived using Levenberg-Marquardt fitting initialized with the published coefficient values. For purposes of algorithm evaluation, the last year of each site time series was reserved for testing and coefficients were trained with the remaining data. The algorithm evaluation results shown below all use this division of training and test data. In addition, a set of coefficients for each algorithm was derived using all available data, with results shown in Table 3. The coefficients for each algorithm from its original publication are given in Table S1 in the Supplementary Material.

Table 3: Re-derived coefficients for each algorithm using all available data from all sites. For the Yao (2013) and Yao (2015) algorithms, a set of coefficients was derived using a variable value of the Priestley-Taylor coefficient α and a constant α of 1.26.

Algorithm	Short name	Version	Re-derived coefficients
Yebra ET	YET	NDVI	a = -0.4589, b = 81.7987
		EVI	a = -1.2841, b = 149.9876
Yebra EF	YEF	NDVI	a = 0.02867, b = 0.6131
		EVI	a = 0.04879, b = 1.0316
Helman exponential	HEx	NDVI	a = 13.3611, b = 2.0344
		EVI	a = 17.0592, b = 2.8873
Helman scaled	HSc		a = -1518.3715, b = 0.001387, c = 33.6520, d = -1.1212, e = -4807.2619
Wang 2007	W07	EVI, T_{a_avg}	a_1 = -0.04417, a_2 = 0.9481, a_3 = 0.006516
		EVI, T_{a_max}	a_1 = -0.06821, a_2 = 0.9715, a_3 = 0.005585
		EVI, T_{s_avg}	a_1 = -0.02849, a_2 = 1.0189, a_3 = 0.004237
		EVI, T_{s_max}	a_1 = 0.0004923, a_2 = 1.0416, a_3 = 0.001707
		NDVI, T_{a_avg}	a_1 = -0.09575, a_2 = 0.5815, a_3 = 0.007896
		NDVI, T_{a_max}	a_1 = -0.1300, a_2 = 0.5995, a_3 = 0.006939
		NDVI, T_{s_avg}	a_1 = -0.09734, a_2 = 0.6438, a_3 = 0.005862
		NDVI, T_{s_max}	a_1 = -0.05442, a_2 = 0.6493, a_3 = 0.002534
Wang/Liang	WL	EVI, T_{a_avg}	a_1 = 0.07223, a_2 = 0.6681, a_3 = 0.009505 a_4 = -0.009441
		EVI, T_{a_max}	a_1 = 0.03066, a_2 = 0.6862, a_3 = 0.008800, a_4 = -0.009861
		EVI, T_{s_avg}	a_1 = 0.08232, a_2 = 0.7360, a_3 = 0.008243,

			a_4 = -0.01089
		EVI, T_{s_max}	a_1 = 0.08224, a_2 = 0.7293, a_3 = 0.008534, a_4 = -0.01610
		NDVI, T_{a_avg}	a_1 = 0.05191, a_2 = 0.3879, a_3 = 0.01077, a_4 = -0.01048
		NDVI, T_{a_max}	a_1 = 0.0005417, a_2 = 0.4030, a_3 = 0.0101, a_4 = -0.01097
		NDVI, T_{s_avg}	a_1 = 0.04231, a_2 = 0.4534, a_3 = 0.009886, a_4 = -0.01223
		NDVI, T_{s_max}	a_1 = 0.04353, a_2 = 0.4484, a_3 = 0.01015, a_4 = -0.01837
Choudhury/ FAO56	Ch		EVI_{min} = 0.02355, EVI_{max} = 0.6117
Kamble/ FAO56	Kmb		a = 1.0452, b = -0.08478
Wang 2010	W10	NDVI	a_1 = -0.1387, a_2 = 1.9938, a_3 = 0.1542, a_4 = -2.1872, a_5 = 54.5977, a_6 = -79.8249, a_7 = 67.8465, a_8 = 0.6891, a_9 = -0.001150
		EVI	a_1 = -0.06988, a_2 = 3.1684, a_3 = 0.05535, a_4 = -3.2777, a_5 = 60.6141, a_6 = -99.1790, a_7 =

			194.5842, a_8 = 0.6498, a_9 = -0.0009489
Yao 2011	Y11		a_1 = -0.0009580, a_2 = -0.0004328, a_3 = 0.03625, a_4 = -0.003210, a_5 = 2.0066, a_6 = 0.5167, a_7 = 0.02503, a_8 = -2.7852
Yao 2013	Y13	T_{s_dTr}	α = 0.7888, $NDVI_{max}$ = 0.7052, $NDVI_{min}$ = -0.08551, T_{opt} = 32.8330, dTr_{max} = 30.9849
		T_{a_dTr}	α = 0.9987, $NDVI_{max}$ = 0.9198, $NDVI_{min}$ = -0.3712, T_{opt} = 25.5854, dTr_{max} = 22.9378
		T_{s_dTr}, α constant	$NDVI_{max}$ = 0.6486, $NDVI_{min}$ = -0.2723, T_{opt} = 141.0440, dTr_{max} = 10.9068
		T_{a_dTr}, α constant	$NDVI_{max}$ = 1.1234, $NDVI_{min}$ = -0.4696, T_{opt} = 25.7667, dTr_{max} = 15.7136
Yao 2015	Y15		α = 1.6445, a_1 = -0.002953, a_2 = 0.007440, a_3 = 0.4299, a_4 = 0.05653, a_5 = 0.01933

		α constant	a_1 = -0.003854, a_2 = 0.009711, a_3 = 0.5611, a_4 = 0.07379, a_5 = 0.02523

For each site and algorithm, RMSE, R^2, and bias were calculated based on LE_{mod} and LE_{obs}, where n is the number of days with valid data available:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(LE_{mod\_i} - LE_{obs\_i}\right)^2}{n}} \tag{21}$$

$$Bias = \frac{\sum_{i=1}^{n}(LE_{mod\_i} - LE_{obs\_i})}{n} \tag{22}$$

$$R^2 = \left\{\frac{n\ \sum_{i=1}^{n}\left(LE_{mod\_i}LE_{obs\_i}\right) - \sum_{i=1}^{n} LE_{mod\_i} \sum_{i=1}^{n} LE_{obs\_i}}{\left[n\sum_{i=1}^{n} LE_{mod\_i}^2 - \left(\sum_{i=1}^{n} LE_{mod\_i}\right)^2\right]\left[n\sum_{i=1}^{n} LE_{obs\_i}^2 - \left(\sum_{i=1}^{n} LE_{obs\_i}\right)^2\right]}\right\}^2 \tag{23}$$

These results were then used to generate boxplots by algorithm. Boxplots were generated using all available sites, separately for the initial published and re-derived coefficients.

Similar statistical comparisons between algorithms were also conducted for the individual surface types specified in Table 2. Based on the results from the analyses with all surface types, four relatively well-performing algorithms with different theoretical bases (Yebra EF, Wang and Liang, Wang 2010, and Yao 2013) were selected for this evaluation. Coefficients were re-derived for each surface type using only data from sites with that type, again reserving the last year of each site for testing. Boxplots similar to those for all types were generated with the surface type specific coefficients and compared to results from the coefficients previously derived from all available sites in order to evaluate whether use of data from only the same surface type improved algorithm performance.

Two additional tests were made of algorithm performance. In order to test whether linear interpolation was artificially improving algorithm statistics by introducing large numbers of non-independent data points, a subset of sites was selected and only data from the vegetation index composite dates were considered. Statistics from only the composite dates were compared to results including all days with sufficient flux tower data for each algorithm. An analysis was also performed for agricultural sites to assess whether interpolation over periods with sudden changes in vegetation index introduces error. To test for this effect, algorithm performance for agricultural sites was evaluated with dates with steep vegetation index slope (> 0.015/ day in NDVI or > 0.01/ day in EVI) excluded, then compared to agricultural site performance without this exclusion.

4. Results analysis

4.1. Global comparison of algorithms and coefficient tuning

Boxplots of RMSE, R^2, and bias by site for all surface types and for the original and re-derived coefficients are shown in Figures 2, 3, and 4. The algorithms are arranged left to right roughly in order of increasing complexity and number of input variables required. Figure 2a shows that the Yebra ET and Helman scaled algorithms have the highest median RMSEs. It is notable that these algorithms are the only ones that do not have any dependence on R_n. The best performing algorithms have median RMSEs that cluster around 25-30 W/m^2 with the original coefficients.

Figure 2b shows the RMSE for all sites with the re-derived coefficients. All algorithms except Yao 2011 had similar or improved RMSE performance, with the best-performing models again having median RMSE in the 25-30 W/m^2 range. The most significant changes were for the Yebra and Helman algorithms, which have the simplest form and fewest required inputs. Most of the other algorithms had little change in median RMSE values, but RMSE tended to decrease for those algorithms that had higher RMSE using the original coefficients.

There are a significant number of outlier sites in the RMSE (Figure 2) and bias (Figure 4) results. Further investigation showed that different sites were outliers for different algorithms with the original coefficients (Figure 2a, 4a), with no systematic patterns apparent. With the re-derived coefficients (Figure 2b, 4b), six sites were responsible for most of the outliers. These sites either had 1 km subset areas that were unrepresentative of the area immediately surrounding the flux tower or were wetland sites. Wetland sites have greater bias and RMSE than other sites, as shown in Figure 5. The difference in performance between wetland sites and others is discussed in greater detail below.

2a) 2b)

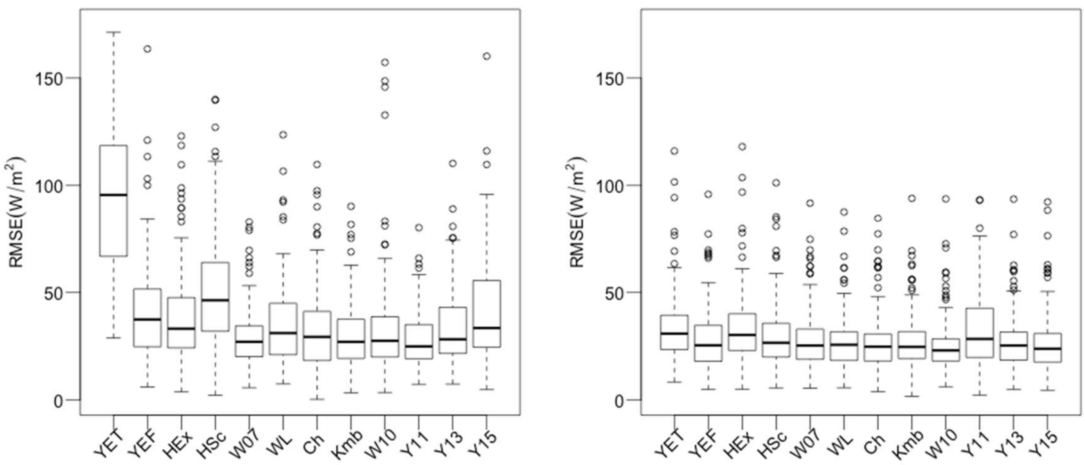


Figure 2: RMSE for each algorithm by site for all cover types. 2a) Using original published coefficients. 2b) Using re-derived coefficients. Key to algorithms: YET - Yebra ET, YEF - Yebra EF, HEx - Helman exponential, HSc - Helman scaled, W07 - Wang 2007, WL - Wang and Liang, Ch - Choudhury/ FAO56, Kmb - Kamble/ FAO56, W10 - Wang 2010, Y11 - Yao 2011, Y13 - Yao 2013, Y15 - Yao 2015

R^2 values for each site and algorithm are shown in Figure 3, with results for the original coefficients shown in Figure 3a and for the re-derived coefficients in Figure 3b. The median R^2 values for the best performing algorithms are between 0.6 and 0.7, with others, usually the simpler algorithms, having significantly lower values. Unlike the results for RMSE, re-fitting the coefficients did not have a strong impact on median R^2 or its distribution for any of the algorithms.

3a) 3b)

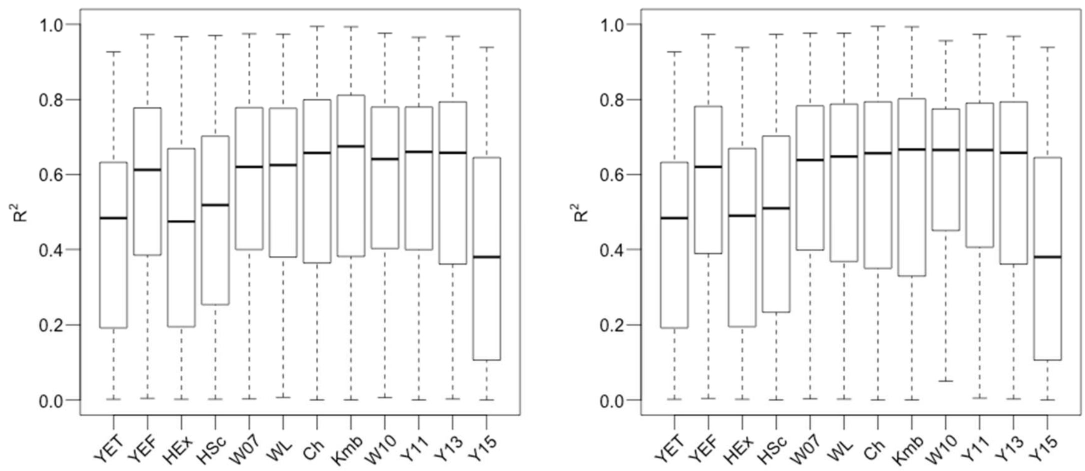


Figure 3: R^2 values by site for each algorithm with the original (3a) and re-derived (3b) coefficients. Algorithm legend on horizontal axis is the same as for Figure 2.

Bias values for all sites and algorithms are shown in Figure 4, with results for the original coefficients in Figure 4a and for the re-derived coefficients in Figure 4b. The patterns here are similar to those seen for RMSE, with the simpler algorithms, especially Yebra ET, usually having the greatest absolute values of median bias with the original coefficients. Figure 4b shows that re-fitting the coefficients reduced the absolute value of median bias for many of the algorithms and reduced the range of bias values in many cases as well.

4a) 4b)

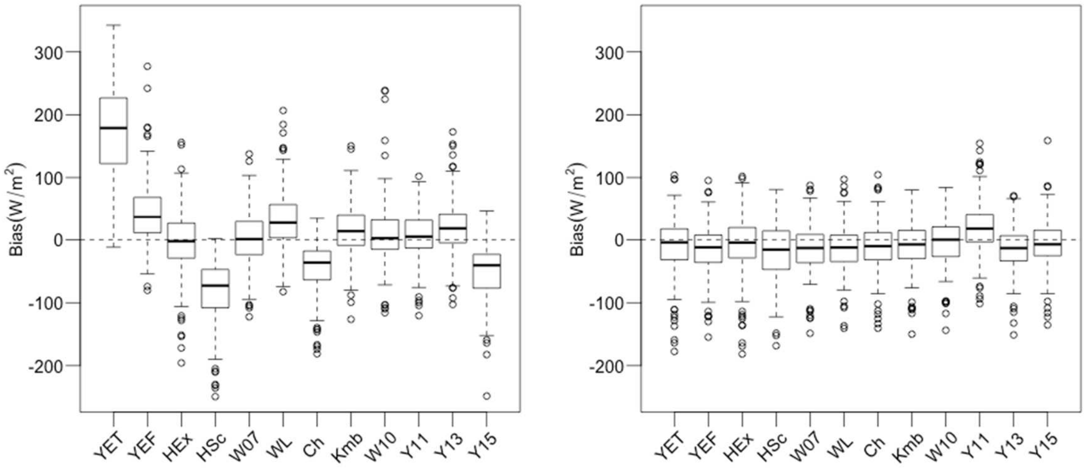


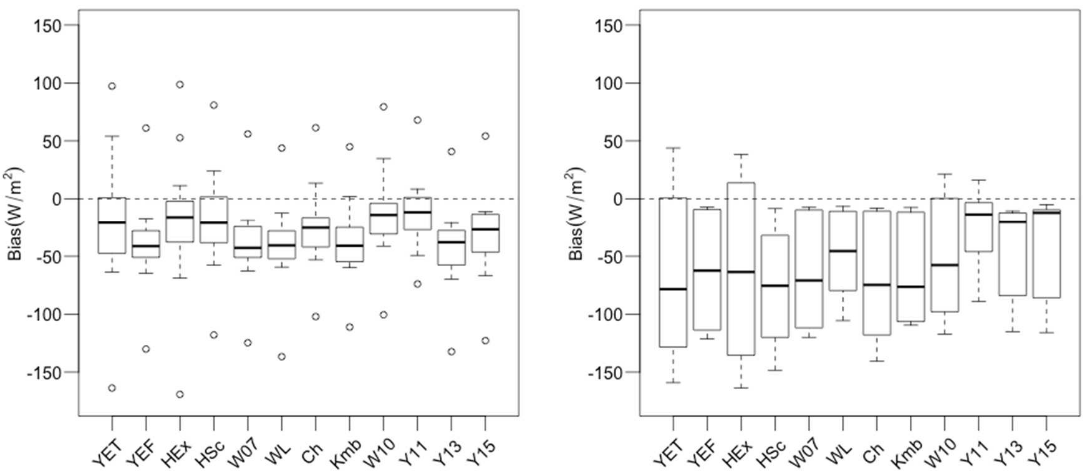
Figure 4: Bias by site for all algorithms and land cover types. Results for original coefficients are shown in Figure 4a, and for re-derived coefficients in Figure 4b. Algorithm legend on horizontal axis is the same as for Figure 2.

4.2. *Evaluation of algorithms by land cover typ*e

In general, there was little difference in the patterns of RMSE, R^2, and bias performance when the re-derived coefficients were used between surface types considered individually and what was shown in the previous section for all sites together. Exceptions to this overall pattern include higher R^2 values for agricultural, deciduous, evergreen, and grassland sites than for all sites considered together, and lower R^2 values for savannah, shrub, and wetland sites. There are also differences in bias and RMSE for agricultural and wetland sites.

Bias differences for agricultural and wetland sites, and RMSE differences for wetland sites, are shown below in Figure 5. Wetland sites (Figure 5a), and to a lesser degree agricultural sites (Figure 5b), showed a consistent low bias across algorithms, with typical bias values of around -25 W/m^2 for agricultural sites and -50 W/m^2 for wetland sites. The Yao 2011, Yao 2013, and Yao 2015 algorithms had a less pronounced bias than the others for wetland sites, but not for agricultural sites. In addition, RMSE for wetland sites was significantly higher than was typical for other surface types, with values of around 40 W/m^2 or more not being unusual (Figure 5c). The Yao algorithms had lower median RMSE, but RMSE was still relatively high for the sites where it was greatest.

5a) 5b)



5c)

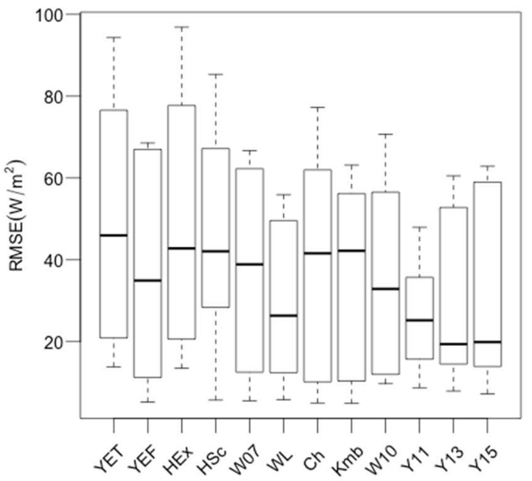


Figure 5: Bias and RMSE by site for those surface types where performance differed significantly from all sites with globally-derived coefficients. Figure 5a: Bias for agricultural sites. Figure 5b: Bias for wetland sites. Figure 5c: RMSE for wetland sites. Algorithm legend on horizontal axis is the same as for Figure 2.

4.3. *Re-training of coefficients by surface type*

For the four algorithms tested (Yebra EF, Wang and Liang, Wang et al. 2010, and Yao et al. 2013), training with data from sites from only one surface type did not result in much change from globally-trained coefficients for most surface types in most cases. (See Figures S1- S3 in the Supplementary Material). The most pronounced exceptions occurred for bias and RMSE for agricultural and wetland sites, paralleling the results when comparing those surface types to the global results as described above. There were also modest improvements in RMSE for deciduous, grassland, and savannah sites (Figures S1-b, S1-d, and S1-e), some modest increase in R^2 for savannah and decrease in R^2 for deciduous sites (Figures S2-e and S2-b) and modest reductions in absolute bias values for deciduous, grassland, and shrub sites (Figures S3-b, S3-d, and S3-f). For evergreen sites, bias values became somewhat more negative (Figure S3-c). In all other cases, there was little change to the statistics, or performance improved for some algorithms and was reduced for others.

The results of surface type specific training for agricultural and wetland sites are shown in Figures 6- 9. Figures 6 and 7 show a decrease in RMSE for agricultural sites and a reduction in the maximum RMSE by site for wetland sites, Figure 8 shows a decrease in bias for agricultural sites, and Figure 9 shows a decrease in bias for wetland sites.

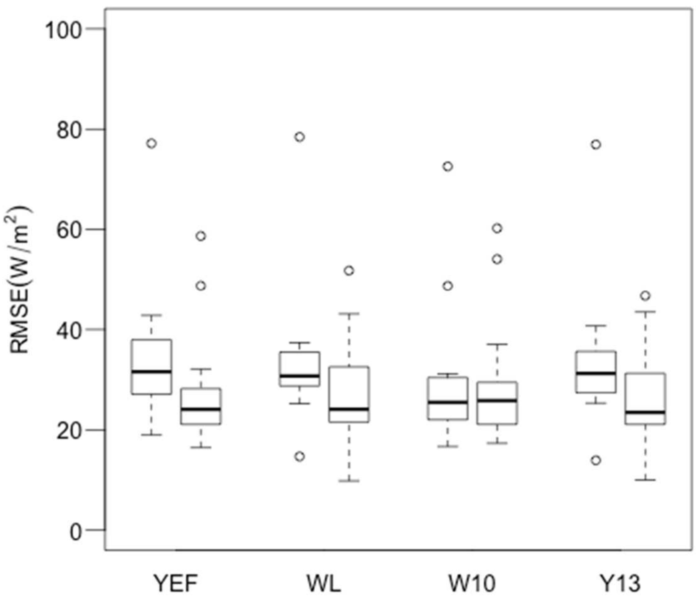
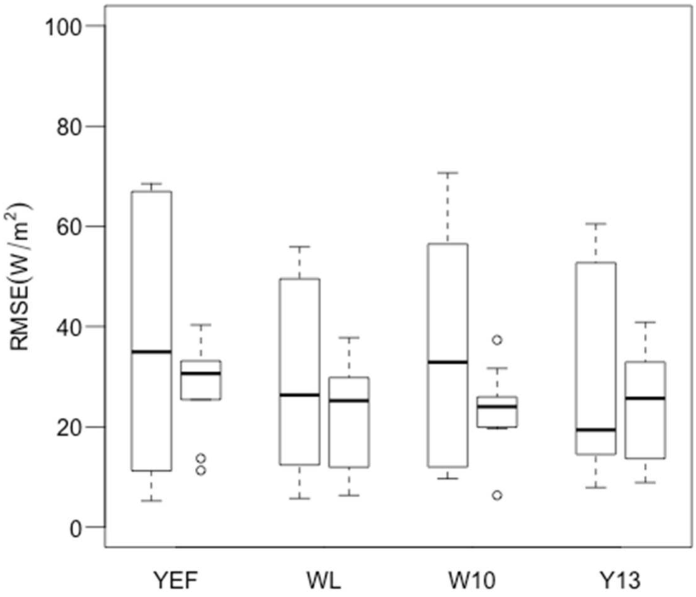


Figure 6: RMSE for agricultural sites for Yebra EF (YEF), Wang and Liang (WL), Wang et al. 2010 (W10) and Yao et al. 2013 (Yao13) algorithms. For each algorithm, left box is for training with data from all sites, and right box is for training with agricultural sites only.

554 

f

555 Figure 7: RMSE for wetland sites. Algorithm labels on X axis are the same as for Figure 6. For
556 each algorithm, left box is for training with data from all sites, and right box is for training with
557 wetland sites only.

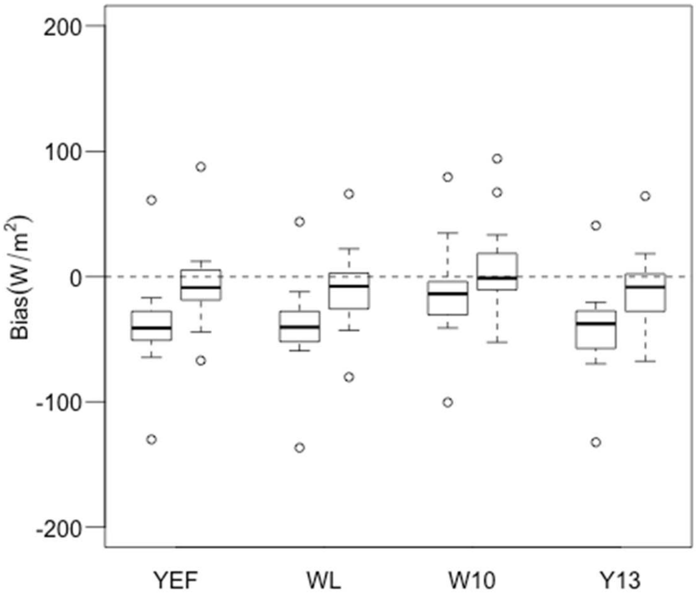


Figure 8: Bias for agricultural sites. Algorithm labels on X axis are the same as for Figure 6. For each algorithm, left box is for training with data from all sites, and right box is for training with agricultural sites only.

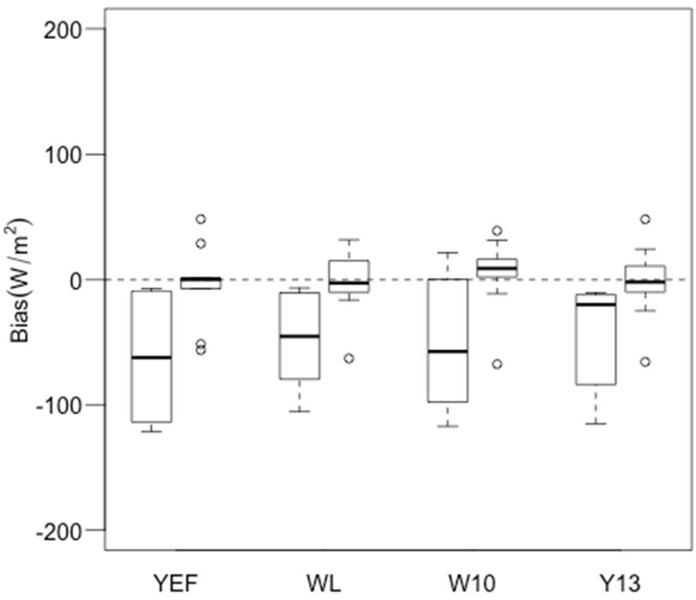


Figure 9: Bias for wetland sites. Algorithm labels on X axis are the same as for Figure 6. For each algorithm, left box is for training with data from all sites, and right box is for training with agricultural sites only.

4.4. Test of effect of linear interpolation of vegetation indices

The possibility that the statistical results of this analysis are being affected by the large number of non-independent data points introduced by linear interpolation of vegetation indices was tested. This was done using seven stations that each had a long data record, in order to obtain a significant number (659) station-days where that were both a composite date and had sufficiently complete Fluxnet records. These stations, listed in Table 4, also represent seven different land cover types. The analysis was conducted for seven of the best-performing algorithms.

Table 4: Stations used for comparison of results from all dates to day of composite only.

Station	Site ID	IGBP class
Audubon Ranch	US-Aud	Grassland (GRA)
Blodgett Forest	US-Blo	Evergreen needleleaf forest (ENF)
Lost Creek	US-Los	Wetland (WET)
Rosemount G21 conventional corn/ soy	US-Ro1	Cropland (CRO)
Santa Rita mesquite	US-SRM	Woody savannah (WSA)
Soroe	DK-Sor	Deciduous broadleaf forest (DBF)
Walnut Gulch Lucky Hills Shrub	US-Whs	Open shrub (OSH)

The results of this analysis are shown in Table 5. It was found that R^2 was higher and RMSE lower when only the composite days were used. The bias was a few W/m^2 more negative in most cases. These results could be because accuracy was lost through interpolation, or because composites were taken on clear weather days and the algorithms performed better under those conditions. It appears not to be the case that the interpolation artificially improved the apparent performance of the algorithms.

Table 5: Results of comparison between all dates and day of composite only.

Algorithm	RMSE all days	RMSE composite days	Bias all days	Bias composite days	R^2 all days	R^2 composite

	(W/m²)	(W/m²)	(W/m²)	(W/m²)		days
Yebra EF (YEF)	32.628	28.871	-5.555	-8.038	0.474	0.619
Choudhury (Ch)	38.761	37.958	-17.583	-21.059	0.473	0.559
Wang 2010 (W10)	31.243	27.028	-5.581	-7.665	0.523	0.673
Wang and Liang (WL)	33.279	29.033	-6.809	-8.952	0.454	0.618
Yao 2011 (Y11)	33.850	29.000	6.432	4.137	0.432	0.586
Yao 2013 (Y13)	32.213	28.789	-6.776	-9.805	0.502	0.656
Yao 2015 (Y15)	31.830	26.886	-2.258	-2.914	0.489	0.657

4.5. Test of effect of rapid VI changes at agricultural sites

At agricultural sites, there are periods where vegetation indices change rapidly, notably at harvest but also during greenup at the beginning of the growing season. The possibility that the vegetation index interpolation might not be as accurate at those times and degrade algorithm performance as a result was examined. The significance of this effect was tested using the 23 agricultural sites and seven algorithms. The median site RMSE, bias, and R^2 were found excluding those times where absolute value of the slope of NDVI > 0.015/ day, or of EVI > 0.01/ day, and compared against the results when all days were included. The results of this analysis are shown in Table 6. The performance of the algorithms was not much different between the cases, or slightly worse when the steep VI slope periods were excluded. It does not appear that periods with steep VI slope are introducing additional error to the results for agricultural sites.

Table 6: Median site statistics of 23 agricultural sites, comparing results with and without exclusion of steep slope in vegetation indices.

Algorithm	RMSE all days (W/m^2)	RMSE VI slope exclusion (W/m^2)	Bias all days (W/m^2)	Bias VI slope exclusion (W/m^2)	R^2 all days	R^2 VI slope exclusion
Yebra EF (YEF)	28.892	29.699	-38.533	-39.340	0.685	0.682
Choudhury (Ch)	36.017	36.651	-51.922	-54.332	0.622	0.616
Wang 2010 (W10)	23.459	24.557	-7.470	-9.063	0.645	0.647
Wang and Liang (WL)	30.560	31.386	-36.285	-37.540	0.694	0.692
Yao 2011 (Y11)	24.746	25.386	-22.921	-23.666	0.666	0.676
Yao 2013 (Y13)	29.944	31.098	-34.811	-35.823	0.664	0.664
Yao 2015 (Y15)	24.056	24.125	-25.712	-26.290	0.688	0.688

5. Discussion

There has been a significant amount of effort devoted to measurement of evapotranspiration at regional to global scales, due to the parameter's importance for a wide range of applications. At these scales, remote sensing is required for at least some of the input data. A large number of remote sensing methods to obtain LE have been developed, and the empirical methods evaluated here are just a subset of those available. There has been a significant amount of work evaluating different LE data sets at global (Jiménez et al. 2011; Mueller et al. 2011), and regional scales (e.g. Mao and Wang 2017; Chen et al. 2014) The focus of these studies has usually been on comparing different “families” of data sets (models vs. reanalyses vs. different remote sensing techniques), but less work has been done comparing results within each “family”. The work done here was performed to fill in this gap for the “family” of regression-based models.

We found that most of the regression methods yielded useful estimates of LE with errors of similar magnitude to those from other methods. This is consistent with the results provided by the original developers of these algorithms (references given in Table 1) as well as with the intercomparison studies cited above and the evaluation of VI-based LE retrieval methods by Glenn et al. (2010). Aside from the effect of inclusion of net radiation as an input parameter, the differences in performance were relatively modest, consistent with Mueller et al. (2011), where the two regression-based models included in the comparison had similar results.

The finding that, while increasing the number of input variables included improved the results, the specific formulation of the regression formula did not, was somewhat surprising. However, this is consistent with the fact that a broad range of different LE algorithms with different theoretical bases are all able to work with some skill, with no particular formulation coming out ahead consistently. The finding that net radiation is the most significant forcing variable is consistent with Badgley et al. (2015), who found that changing the source of net radiation data used by a Priestley-Taylor model resulted in a greater change to its results than changing the source of meteorological or vegetation index data. In addition, the finding of the high significance of the net radiation variable is also consistent with Wang et al. (2007), who found a greater correlation of flux tower LE measurements to net radiation than to temperatures or vegetation indices.

The effect of land surface type on the performance of a range of empirical algorithms has not been examined in detail before this study. We found that there was some variation in performance, which is not unexpected, since different land cover types have different degrees of annual variation in vegetation index, and probably different relationships between VI and LE.

A probable reason for the low bias in wetland sites is that evaporation from the surface makes a more significant contribution to LE than for other site types, while vegetation indices are more of an indicator of transpiration. Multiple studies (S. T. Allen et al. 2017; Runkle et al. 2014; Malone et al. 2014) have shown that H is a much smaller component of the surface energy budget than LE for wetland sites, and at least one study (Beigt et al. 2008) indicates that sensible heating can make a positive contribution to available energy at a wetland site. High values of LE relative to H are also seen in the wetland flux tower energy balance measurements used in this study. In addition, S. T. Allen et al. (2017) have shown that release of stored energy from the surface can contribute to available energy in the autumn season for a wetland. These sources of energy are available for evaporation but not transpiration. Along with higher surface moisture availability, these effects can result in high evaporative fraction and high rates of evaporation relative to transpiration from wetlands. Vegetation indices are not a good indicator of surface evaporation, as in the limiting case of open water where VIs are very low but surface evaporation is high.

There are other variables, such as precipitation and soil moisture, that are strongly related to LE but not incorporated into any of the regression formulas reviewed. It should be possible to include precipitation and soil moisture from surface or microwave measurements, but it would be important to consider scaling effects when using these data. Surface precipitation and soil moisture measurements are in effect point measurements, limiting the possibilities for upscaling. On the other hand, while the footprint of microwave observations is typically greater than the resolution of vegetation indices. For example, the resolution of the microwave-based Global Precipitation Measurement (GPM) is about 5 km. Global microwave soil moisture observations are currently available at scales of around 25 km, although there are ongoing efforts to downscale remote sensing soil moisture data sets, as reviewed by Peng et al. (2017). If precipitation is used as an input variable, a lag effect must be considered as the moisture made available in a precipitation event may remain available for several days. By contrast, soil moisture is a more immediate measure of water availability and a lag effect would not be expected.

Overall, the performance of the VI algorithms is consistent with what has been seen in previous work with those algorithms and with other methods for obtaining ET from remote sensing.

Where possible, it is preferable to use algorithms with more input data parameters and a more realistic basis to their parameterization, although the specifics of the underlying basis appear to matter little. Simpler algorithms can perform almost as well as more complex ones, but it is more important that they be tuned with appropriate training data. At a minimum, inclusion of R_n as a parameter along with VI is recommended wherever possible.

6. Conclusions

In this study, we have confirmed that many simple regression methods can work to obtain LE on daily time scales with error levels comparable to those from more complex methods. We have noted certain patterns in the performance of these algorithms. Increasing the number of variables included in regression formulas tends to improve performance, although the specific form of model used is not as significant. Those algorithms in which net radiation was one of the input variables produced much less error than those that did not, as demonstrated by the difference between the Yebra (2013) ET (YET) algorithm, and Yebra (2013) EF (YEF) algorithms, which are very similar to each other except that YEF has net radiation as an input while YET does not. (Figures 2, 3, 4). Tuning of the regression coefficients to the global data set improved performance in most cases, which is also demonstrated in Figures 2-4. This improvement was most significant for those models with fewer input variables. For wetland and agricultural surface types, tuning with data specific to that surface type produced improved results (Figures 6-8), but this was not the case for other surface types.

There are multiple opportunities for adaptation and improvement of the methods evaluated here. All of the input variables to the regression formulas are potentially available through remote sensing (Liang 2007, Liang et al. 2012) or reanalyses, so there is the potential for removing all dependence on ground-based observations. In addition, additional variables such as soil moisture and precipitation that are not included in the set of empirical algorithms evaluated here could be included in similar algorithms in the future if issues with spatial resolution can be addressed.

Acknowledgements: This work is partially funded by NASA and NOAA grants. Funding for AmeriFlux data resources was provided by the U.S. Department of Energy's Office of Science. Fluxnet 2015 data were provided by the European Fluxes Database and the AmeriFlux Management Project. MODIS subset data were provided by the Oak Ridge National Laboratory Distributed Active Archvie Center (ORNL DAAC).

Conflicts of interest: None.

References

Allen, R. G., and Food and Agriculture Organization of the United Nations, eds. 1998. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*. FAO Irrigation and Drainage Paper 56. Rome: Food and Agriculture Organization of the United Nations.

Allen, Scott T., Michele L. Reba, Brandon L. Edwards, and Richard F. Keim. 2017. "Evaporation and the Subcanopy Energy Environment in a Flooded Forest." *Hydrological Processes* 31 (16): 2860–71. https://doi.org/10.1002/hyp.11227.

Anderson, M. 1997. "A Two-Source Time-Integrated Model for Estimating Surface Fluxes Using Thermal Infrared Remote Sensing." *Remote Sensing of Environment* 60 (2): 195–216. https://doi.org/10.1016/S0034-4257(96)00215-5.

Anderson, Martha C., John M. Norman, John R. Mecikalski, Jason A. Otkin, and William P. Kustas. 2007. "A Climatological Study of Evapotranspiration and Moisture Stress across the Continental United States Based on Thermal Remote Sensing: 1. Model Formulation." *Journal of Geophysical Research* 112 (D10).

https://doi.org/10.1029/2006JD007506.

Badgley, Grayson, Joshua B. Fisher, Carlos Jiménez, Kevin P. Tu, and Raghuveer Vinukollu. 2015. "On Uncertainty in Global Terrestrial Evapotranspiration Estimates from Choice of Input Forcing Datasets*." *Journal of Hydrometeorology* 16 (4): 1449–55. https://doi.org/10.1175/JHM-D-14-0040.1.

Baldocchi, Dennis, Eva Falge, Lianhong Gu, Richard Olson, David Hollinger, Steve Running, Peter Anthoni, et al. 2001. "FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem?Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities." *Bulletin of the American Meteorological Society* 82 (11): 2415–34. https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2.

Bastiaanssen, W.G.M., H. Pelgrum, J. Wang, Y. Ma, J.F. Moreno, G.J. Roerink, and T. van der Wal. 1998. "A Remote Sensing Surface Energy Balance Algorithm for Land (SEBAL)." *Journal of Hydrology* 212–213 (December): 213–29. https://doi.org/10.1016/S0022-1694(98)00254-6.

Beigt, D., M. C. Piccolo, and G.M.E. Perillo. 2008. "Surface Heat Budget of an Estuarine Tidal Flat (Bah´ıa Blanca Estuary, Argentina)." *Cienc. Mar.* 34 (1): 1–15.

Brown, M.E., J.E. Pinzon, K. Didan, J.T. Morisette, and C.J. Tucker. 2006. "Evaluation of the Consistency of Long-Term NDVI Time Series Derived from AVHRR,SPOT-Vegetation, SeaWiFS, MODIS, and Landsat ETM+ Sensors." *IEEE Transactions on Geoscience and Remote Sensing* 44 (7): 1787–93. https://doi.org/10.1109/TGRS.2005.860205.

Chen, Yang, Jiangzhou Xia, Shunlin Liang, Jinming Feng, Joshua B. Fisher, Xin Li, Xianglan Li, et al. 2014. "Comparison of Satellite-Based Evapotranspiration Models over Terrestrial Ecosystems in China." *Remote Sensing of Environment* 140 (January): 279–93. https://doi.org/10.1016/j.rse.2013.08.045.

Choudhury, Bj, Nu Ahmed, Sb Idso, Rj Reginato, and Cst Daughtry. 1994. "Relations Between Evaporation Coefficients and Vegetation Indexes Studied by Model Simulations." *Remote Sensing of Environment* 50 (1): 1–17. https://doi.org/10.1016/0034-4257(94)90090-6.

Didan. K. 2015. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD13Q1.006

Du, Yiheng, Ronny Berndtsson, Dong An, Linus Zhang, Zhenchun Hao, and Feifei Yuan. 2017. "Hydrologic Response of Climate Change in the Source Region of the Yangtze River, Based on Water Balance Analysis." *Water* 9 (2): 115. https://doi.org/10.3390/w9020115.

Field, Christopher B., Vicente R. Barros, David Jon Dokken, Katharine J. Mach, and Michael D. Mastrandrea, eds. 2014. *Climate Change 2014 Impacts, Adaptation, and Vulnerability: Working Group II Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107415379.

Fisher, Joshua B., Kevin P. Tu, and Dennis 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 Sensing of Environment* 112 (3): 901–19. https://doi.org/10.1016/j.rse.2007.06.025.

Foken, Thomas. 2008. "THE ENERGY BALANCE CLOSURE PROBLEM: AN OVERVIEW." *Ecological Applications* 18 (6): 1351–67. https://doi.org/10.1890/06-0922.1.

Garner, Grace, David M Hannah, and Glenn Watts. 2017. "Climate Change and Water in the

UK: Recent Scientific Evidence for Past and Future Change." *Progress in Physical Geography* 41 (2): 154–70. https://doi.org/10.1177/0309133316679082.
Glenn, Edward P., Pamela L. Nagler, and Alfredo R. Huete. 2010. "Vegetation Index Methods for Estimating Evapotranspiration by Remote Sensing." *Surveys in Geophysics* 31 (6): 531–55. https://doi.org/10.1007/s10712-010-9102-2.
Haileslassie, Amare, Don Peden, Solomon Gebreselassie, Tilahun Amede, and Katrien Descheemaeker. 2009. "Livestock Water Productivity in Mixed Crop-Livestock Farming Systems of the Blue Nile Basin: Assessing Variability and Prospects for Improvement." *Agricultural Systems* 102 (1–3): 33–40. https://doi.org/10.1016/j.agsy.2009.06.006.
Helman, D., A. Givati, and I. M. Lensky. 2015. "Annual Evapotranspiration Retrieved from Satellite Vegetation Indices for the Eastern Mediterranean at 250 M Spatial Resolution." *Atmospheric Chemistry and Physics* 15 (21): 12567–79. https://doi.org/10.5194/acp-15-12567-2015.
Hirschi, Martin, Dominik Michel, Irene Lehner, and Sonia I. Seneviratne. 2017. "A Site-Level Comparison of Lysimeter and Eddy Covariance Flux Measurements of Evapotranspiration." *Hydrology and Earth System Sciences* 21 (3): 1809–25. https://doi.org/10.5194/hess-21-1809-2017.
Jiménez, C., C. Prigent, B. Mueller, S. I. Seneviratne, M. F. McCabe, E. F. Wood, W. B. Rossow, et al. 2011. "Global Intercomparison of 12 Land Surface Heat Flux Estimates." *Journal of Geophysical Research* 116 (D2). https://doi.org/10.1029/2010JD014545.
Jung, Martin, Markus Reichstein, Philippe Ciais, Sonia I. Seneviratne, Justin Sheffield, Michael L. Goulden, Gordon Bonan, et al. 2010. "Recent Decline in the Global Land Evapotranspiration Trend due to Limited Moisture Supply." *Nature* 467 (7318): 951–54. https://doi.org/10.1038/nature09396.
Kalma, Jetse D., Tim R. McVicar, and Matthew F. McCabe. 2008. "Estimating Land Surface Evaporation: A Review of Methods Using Remotely Sensed Surface Temperature Data." *Surveys in Geophysics* 29 (4–5): 421–69. https://doi.org/10.1007/s10712-008-9037-z.
Kamble, Baburao, Ayse Irmak, and Kenneth Hubbard. 2013. "Estimating Crop Coefficients Using Remote Sensing-Based Vegetation Index." *Remote Sensing* 5 (4): 1588–1602. https://doi.org/10.3390/rs5041588.
Liang, S. (2007). Recent developments in estimating land surface biogeophysical variables from optical remote sensing. *Progress in Physical Geography, 31*, 501-516
Liang, S., Li, X., & Wang, J. (Eds.) (2012). *Advanced remote sensing: Terrestrial information extraction and applications*: Academic Press
Liu, Xiaoying, Chunying Xu, Xiuli Zhong, Yuzhong Li, Xiaohuan Yuan, and Jingfeng Cao. 2017. "Comparison of 16 Models for Reference Crop Evapotranspiration against Weighing Lysimeter Measurement." *Agricultural Water Management* 184 (April): 145–55. https://doi.org/10.1016/j.agwat.2017.01.017.
Malone, Sparkle L., Christina L. Staudhammer, Henry W. Loescher, Paulo Olivas, Steven F. Oberbauer, Michael G. Ryan, Jessica Schedlbauer, and Gregory Starr. 2014. "Seasonal Patterns in Energy Partitioning of Two Freshwater Marsh Ecosystems in the Florida Everglades: ENERGY DYNAMICS IN EVERGLADES ECOSYSTEMS." *Journal of Geophysical Research: Biogeosciences* 119 (8): 1487–1505. https://doi.org/10.1002/2014JG002700.
Mao, Yuna, and Kaicun Wang. 2017. "Comparison of Evapotranspiration Estimates Based on the Surface Water Balance, Modified Penman-Monteith Model, and Reanalysis Data Sets

for Continental China: Terrestrial Evapotranspiration in China." *Journal of Geophysical Research: Atmospheres*. https://doi.org/10.1002/2016JD026065.

Monteith, J. 1965. "Evaporation and Environment." *Symposia of the Society for Experimental Biology* 19: 205–34.

Mu, Qiaozhen, Maosheng Zhao, and Steven W. Running. 2011. "Improvements to a MODIS Global Terrestrial Evapotranspiration Algorithm." *Remote Sensing of Environment* 115 (8): 1781–1800. https://doi.org/10.1016/j.rse.2011.02.019.

Mueller, B., S. I. Seneviratne, C. Jimenez, T. Corti, M. Hirschi, G. Balsamo, P. Ciais, et al. 2011. "Evaluation of Global Observations-Based Evapotranspiration Datasets and IPCC AR4 Simulations: GLOBAL LAND EVAPOTRANSPIRATION DATASETS." *Geophysical Research Letters* 38 (6): n/a-n/a. https://doi.org/10.1029/2010GL046230.

Peng, Jian, Alexander Loew, Olivier Merlin, and Niko E. C. Verhoest. 2017. "A Review of Spatial Downscaling of Satellite Remotely Sensed Soil Moisture: Downscale Satellite-Based Soil Moisture." *Reviews of Geophysics* 55 (2): 341–66. https://doi.org/10.1002/2016RG000543.

Penman, H. L. 1948. "Natural Evporation from Open Water, Bare Soil and Grass." *The Royal Society* 193 (1032): 120–45.

Perez-Priego, Oscar, Tarek S. El-Madany, Mirco Migliavacca, Andrew S. Kowalski, Martin Jung, Arnaud Carrara, Olaf Kolle, et al. 2017. "Evaluation of Eddy Covariance Latent Heat Fluxes with Independent Lysimeter and Sapflow Estimates in a Mediterranean Savannah Ecosystem." *Agricultural and Forest Meteorology* 236 (April): 87–99. https://doi.org/10.1016/j.agrformet.2017.01.009.

Priestley, C., and R. Taylor. 1972. "On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters." *Monthly Weather Review* 100: 81–92.

Runkle, B.R.K., C. Wille, M. Gazovic, M. Wilmking, and L. Kutzbach. 2014. "The Surface Energy Balance and Its Drivers in a Boreal Peatland Fen of Northwestern Russia." *Journal of Hydrology* 511: 359–73. https://doi.org/https://doi.org/10.1016/j.jhydrol.2014.01.056.

Schaaf, C., and Z. Wang. 2015. MCD43A1 MODIS/Terra+Aqua BRDF/Albedo Model Parameters Daily L3 Global - 500m V006. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MCD43A1.006

Spinoni, Jonathan, Gustavo Naumann, and J?rgen V. Vogt. 2017. "Pan-European Seasonal Trends and Recent Changes of Drought Frequency and Severity." *Global and Planetary Change* 148 (January): 113–30. https://doi.org/10.1016/j.gloplacha.2016.11.013.

Steven, Michael D, Timothy J Malthus, Fr?d?ric Baret, Hui Xu, and Mark J Chopping. 2003. "Intercalibration of Vegetation Indices from Different Sensor Systems." *Remote Sensing of Environment* 88 (4): 412–22. https://doi.org/10.1016/j.rse.2003.08.010.

Wang, Kaicun, and Robert E. Dickinson. 2012. "A Review of Global Terrestrial Evapotranspiration: Observation, Modeling, Climatology, and Climatic Variability: GLOBAL TERRESTRIAL EVAPOTRANSPIRATION." *Reviews of Geophysics* 50 (2). https://doi.org/10.1029/2011RG000373.

Wang, Kaicun, Robert E. Dickinson, Martin Wild, and Shunlin Liang. 2010a. "Evidence for Decadal Variation in Global Terrestrial Evapotranspiration between 1982 and 2002: 1. Model Development." *Journal of Geophysical Research* 115 (D20). https://doi.org/10.1029/2009JD013671.

———. 2010b. "Evidence for Decadal Variation in Global Terrestrial Evapotranspiration between 1982 and 2002: 2. Results.” *Journal of Geophysical Research* 115 (D20). https://doi.org/10.1029/2010JD013847.

Wang, Kaicun, and Shunlin Liang. 2008. “An Improved Method for Estimating Global Evapotranspiration Based on Satellite Determination of Surface Net Radiation, Vegetation Index, Temperature, and Soil Moisture.” *Journal of Hydrometeorology* 9 (4): 712–27. https://doi.org/10.1175/2007JHM911.1.

———. 2009. “Estimation of Daytime Net Radiation from Shortwave Radiation Measurements and Meteorological Observations.” *Journal of Applied Meteorology and Climatology* 48 (3): 634–43. https://doi.org/10.1175/2008JAMC1959.1.

Wang, Kaicun, Pucai Wang, Zhanqing Li, M. Cribb, and Michael Sparrow. 2007. “A Simple Method to Estimate Actual Evapotranspiration from a Combination of Net Radiation, Vegetation Index, and Temperature.” *Journal of Geophysical Research* 112 (D15). https://doi.org/10.1029/2006JD008351.

Yao, Yunjun, Shunlin Liang, Jie Cheng, Shaomin Liu, Joshua B. Fisher, Xudong Zhang, Kun Jia, et al. 2013. “MODIS-Driven Estimation of Terrestrial Latent Heat Flux in China Based on a Modified Priestley-Taylor Algorithm.” *Agricultural and Forest Meteorology* 171–172 (April): 187–202. https://doi.org/10.1016/j.agrformet.2012.11.016.

Yao, Yunjun, Shunlin Liang, Xianglan Li, Jiquan Chen, Kaicun Wang, Kun Jia, Jie Cheng, et al. 2015. “A Satellite-Based Hybrid Algorithm to Determine the Priestley–Taylor Parameter for Global Terrestrial Latent Heat Flux Estimation across Multiple Biomes.” *Remote Sensing of Environment* 165 (August): 216–33. https://doi.org/10.1016/j.rse.2015.05.013.

Yao, Yunjun, Shunlin Liang, Qiming Qin, Kaicun Wang, and Shaohua Zhao. 2011. “Monitoring Global Land Surface Drought Based on a Hybrid Evapotranspiration Model.” *International Journal of Applied Earth Observation and Geoinformation* 13 (3): 447–57. https://doi.org/10.1016/j.jag.2010.09.009.

Yebra, Marta, Albert Van Dijk, Ray Leuning, Alfredo Huete, and Juan Pablo Guerschman. 2013. “Evaluation of Optical Remote Sensing to Estimate Actual Evapotranspiration and Canopy Conductance.” *Remote Sensing of Environment* 129 (February): 250–61. https://doi.org/10.1016/j.rse.2012.11.004.

Zeng, Zhenzhong, Shilong Piao, Xin Lin, Guodong Yin, Shushi Peng, Philippe Ciais, and Ranga B Myneni. 2012. “Global Evapotranspiration over the Past Three Decades: Estimation Based on the Water Balance Equation Combined with Empirical Models.” *Environmental Research Letters* 7 (1): 14026. https://doi.org/10.1088/1748-9326/7/1/014026.

Zhang, Ke, John S. Kimball, and Steven W. Running. 2016. “A Review of Remote Sensing Based Actual Evapotranspiration Estimation: A Review of Remote Sensing Evapotranspiration.” *Wiley Interdisciplinary Reviews: Water* 3 (6): 834–53. https://doi.org/10.1002/wat2.1168.