

Characterizing the diurnal patterns of errors in the prediction of evapotranspiration by  
several land-surface models: an NACP analysis

Ashley M. Matheny, Gil Bohrer, Paul C. Stoy, Ian Baker, Andy Black, Ankur R. Desai,  
Michael Dietze, Chris Gough, Valeriy Ivanov, Paul Jassal, Kim Novick, Karina Schäfer, Hans  
Verbeeck

Ashley M. Matheny and Gil Bohrer, Department of Civil, Environmental, and Geodetic  
Engineering, The Ohio State University, Columbus, Ohio, USA.

Paul C. Stoy, Department of Land Resources and Environmental Sciences, Montana State  
University, Bozeman, Montana, USA

Ian Baker, Atmospheric Science Department, Colorado State University, Fort Collins, Colorado,  
USA

Andy Black and Paul Jassal, Faculty of Land and Food Systems, University of British Columbia,  
Vancouver, British Columbia, Canada

Ankur R. Desai, Center for Climactic Research, University of Wisconsin-Madison, Madison,  
Wisconsin, USA

Michael Dietze, Department of Earth and Environment, Boston University, Boston,  
Massachusetts, USA

Chris Gough, Department of Biology, Virginia Commonwealth University, Richmond, Virginia,  
USA

Valeriy Ivanov, Department of Civil and Environmental Engineering, University of Michigan,  
Ann Arbor, Michigan, USA

Kim Novick, School of Public and Environmental Affairs, Indiana University, Bloomington,  
Indiana, USA

Karina Schäfer, Department of Biological Sciences, Rutgers University, Newark, New Jersey,  
USA

Hans Verbeeck, Department of Applied Ecology and Environmental Biology, Ghent University,  
Ghent, Belgium

Corresponding Author: A. M. Matheny, Department of Civil, Environmental, and Geodetic  
Engineering, The Ohio State University, Columbus, Ohio 43210, USA (matheny.44@osu.edu)

This article has been accepted for publication and undergone full peer review but has not  
been through the copyediting, typesetting, pagination and proofreading process which may  
lead to differences between this version and the Version of Record. Please cite this article  
as doi: 10.1002/2014JG002623

## Abstract

Land surface models use different formulations of stomatal conductance and plant hydraulics, and it is unclear which type of model best matches observed surface-atmosphere water flux. We use the NACP (North American Carbon Program) dataset of latent heat flux (LE) measurements from 25 sites and predictions from 9 models to evaluate models' ability to resolve sub-daily dynamics of transpiration. Despite overall good forecast at the seasonal scale, models have difficulty resolving the dynamics of intra-daily hysteresis. The majority of models tend to underestimate LE in the pre-noon hours and overestimate in the evening. We hypothesize that this is a result of un-resolved afternoon stomatal closure due to hydrodynamic stresses. Although no model or stomata parameterization was consistently best or worst in terms of ability to predict LE, errors in model-simulated LE were consistently largest and most variable when soil moisture was moderate and VPD was moderate to limiting. Nearly all models demonstrate a tendency to underestimate the degree of maximum hysteresis which, across all sites studied, is most pronounced during moisture limited conditions. These diurnal error patterns are consistent with models' diminished ability to accurately simulate the natural hysteresis of transpiration. We propose that lack of representation of plant hydrodynamics is, in part, responsible for these error patterns.

## Key Points

- Land surface models produce sub-daily patterns of latent heat flux error
- Error patterns are characterized by the stomatal conductance formulation used
- Current models lack a mechanism to simulate hysteretic transpiration

## Key Words

Ameriflux; Evapotranspiration; hysteresis; land-surface models; NACP; stomatal conductance

## 1. Introduction

Evapotranspiration (ET) accounts for 60% of precipitated water in some catchments and up to 95% in others, comprising the largest component of the terrestrial hydrologic cycle [Fisher *et al.*, 2005; Ford *et al.*, 2007; Jasechko *et al.*, 2013; Katul *et al.*, 2012]. Jasechko [2013] has recently shown that more than 80% of terrestrial ET must be partitioned to transpiration to maintain a mass balance between plant transpiration and CO<sub>2</sub> uptake.

Transpiration couples the carbon and water cycles and acts as a principal feedback between the land surface and atmosphere. Furthermore, stomatal conductance controls transpiration rates and acts as a principle control to the surface energy redistribution and, thus, couples transpiration with sensible heat and convection at the planetary surface layer. It is therefore crucial to understand its dynamics [Savabi and Stockle, 2001]. Concomitant to transpiration, terrestrial ecosystems take up an average of *ca.* 120 Pg of carbon annually, providing a critical sink for atmospheric CO<sub>2</sub> [Beer *et al.*, 2010; Pan *et al.*, 2011]. C uptake rates are also limited by stomatal conductance. With global and regional temperature increases and a potential rise of the variation in regional precipitation patterns, the ability to accurately model transpiration is essential to predicting land surface-climate change feedbacks in terms of heat, water, and carbon exchange [Allen *et al.*, 2010; Berry *et al.*, 2010; Choat *et al.*, 2012; Huntington *et al.*, 2009; Jasechko *et al.*, 2013; Niyogi *et al.*, 2009; Wu *et al.*, 2012].

Plants balance the uptake of CO<sub>2</sub> required for photosynthesis with the need to maintain sufficient moisture levels inside the leaf. Leaves close their stomata to avoid decreases in leaf water potential that could lead to desiccation or catastrophic cavitation within the xylem system [Brodribb and Holbrook, 2003; 2006; 2007; Brodribb and Cochard, 2008; Tyree and Zimmermann, 2002; Zimmerman, 1983]. Stomata operate dynamically to regulate water loss and C uptake, and this regulation – ultimately a function of controls over guard cell function – is related to a number of external environmental variables including

incoming solar radiation, ambient CO<sub>2</sub> concentration, temperature, humidity, and soil water availability, and internal physiological variables such as CO<sub>2</sub> concentration within the leaf, abscisic acid (ABA) concentration within the leaf, and leaf water potential [Ball, 1988; Sperry *et al.*, 2002; Tyree and Zimmermann, 2002]. Regulation of stomatal conductance due to leaf-level water stress is known to control transpiration when the soil is dry or when vapor pressure deficit (VPD) is large, but can also impact stomatal apertures under conditions of adequate soil moisture and lower evaporative demand if depletion of water storage in the leaves occurs at a faster rate than recharge through roots, stems, and branches [Brodribb and Holbrook, 2004; McCulloh *et al.*, 2012; Sperry *et al.*, 2002]. The imbalance, or lack thereof, between water demand in the leaf and water supply from the soil imposes hydrodynamic limitations on stomatal conductance [Bohrer *et al.*, 2005; Damour *et al.*, 2010; McCulloh and Sperry, 2005; Thomsen *et al.*, 2013; Tyree and Sperry, 1989; Zhang *et al.*, 2014].

Despite numerous advances in the understanding of water transport, current land-surface and ecosystem models do not adequately resolve observed hydrodynamic phenomena within the plants [Damour *et al.*, 2010; Egea *et al.*, 2011]. At present, the most widely used models for stomatal conductance ( $g_s$ ) are based on the net photosynthetic rate and environmental variables, most frequently PAR and relative humidity or VPD [e.g., Ball *et al.*, 1987; Collatz *et al.*, 1991; Leuning, 1995]. These models do not include a mechanistic link between leaf water potential and stomatal conductance, but instead force water supply limitations through empirical parameterizations of the response of  $g_s$  to soil moisture at the root zone, representing a ‘conceptual’ understanding of the soil control on plant water uptake [De Kauwe *et al.*, 2013]. This relationship is generally obtained through the use of a heuristic, Feddes-type or Jarvis-type parameterization [Feddes *et al.*, 1976; Jarvis, 1976], which describes threshold values of soil water potential, i.e., the onset of soil stress and the permanent wilting point, between which stomatal behavior leads to suppressed flux and gas

exchange. Within this range of values, stomatal conductance declines linearly or sigmoidally with declining water availability. Alternatively, hydraulically based models for stomatal conductance link stomatal function to leaf water potential [Bohrer *et al.*, 2005; Janott *et al.*, 2011; Tuzet *et al.*, 2003] but typically require extensive knowledge of plant hydraulic parameters such as a species' threshold for cavitation, hydraulic and tree architecture, tree water storage capacity, or the cumulative root-to-leaf resistance to water transport, all of which are difficult to measure, parameterize, and implement over across multiple species, canopy and crown structures, and plant functional types. Although these hydrodynamic models provide a more physically realistic description of the plant hydraulic system and thus can simulate the effects of diurnal water stress and drought limitations on transpiration, they have not been widely incorporated in land-surface models to date [Damour *et al.*, 2010; Egea *et al.*, 2011; Hickler *et al.*, 2006].

A hallmark feature of plant hydrodynamics is the daily hysteresis of transpiration (Figure 1). For the same VPD, plants transpire more during the morning, when internal storage levels are high, than during the afternoon when water storage in the stem is depleted [Bohrer *et al.*, 2005; Phillips *et al.*, 2003; Verbeeck *et al.*, 2007a; Zhang *et al.*, 2014]. We hypothesize that models will produce patterns of diurnal error that are representative of the missing mechanistic representation of the physiological response to plant hydrodynamics, namely a hysteresis as fluxes are underestimated during morning and overestimated in the afternoon. These patterns will be characteristic of the form and parameterization of the stomatal conductance model, the manner in which water stress limitations are enforced, and the ecosystem type, all of which are related to the features of plant hydrodynamics noted above. Errors in model predictions of stomatal conductance translate directly into errors in the prediction of transpiration and therefore evapotranspiration and latent heat flux (LE), the heat flux associated with the vaporization of water. Furthermore, when models are well

parameterized for the long-term overall transpiration (daily, weekly or monthly), we expect to find an underestimation of transpiration during the morning and early afternoon to compensate for the afternoon-early evening overestimation. It is likely therefore that errors in LE will additionally lead to errors in the sensible heat flux (H) on account of incorrect partitioning of the turbulent flux terms in the surface energy balance. Overestimations of afternoon evapotranspiration will cause afternoon temperature to appear lower, thus causing a “damping” effect on modeled H. Errors in resolving stomatal conductance will also influence a model’s ability to correctly predict photosynthesis and, therefore, may propagate into the model’s predictions of gross primary productivity (GPP) [Schaefer *et al.*, 2012] and net ecosystem exchange (NEE) [Dietze *et al.*, 2011; Schwalm *et al.*, 2010; Stoy *et al.*, 2013].

The specific objectives of our study are as follows: (1) characterize patterns of intra-daily errors in model-simulated LE, (2) if typical error patterns exist, determine whether these patterns are typical of site characteristics (biome, mean annual rainfall, soil texture, leaf area index) or model formulation (i.e. the parameterization of stomatal function and water stress), and (3) determine the extent to which errors in LE are propagated into errors in carbon flux. We test these predictions by comparing simulation results from 9 commonly used land-surface models with observations from 25 eddy covariance tower sites of different ecosystem types. The data was collected as part of the North American Carbon Program (NACP).

## **2. Methods**

### **2.1 Data sources and preparation**

This study uses a large-scale inter-comparison between eddy-covariance observations of surface H, LE, and CO<sub>2</sub> fluxes from multiple sites, and simulation results from an ensemble of land-surface models, to characterize the patterns of intra-daily error in simulated LE. The North American Carbon Program (NACP) (<http://www.nacarbon.org/nacp/>) site-level interim synthesis project began in 2008 and was developed with the purpose of

comparing eddy-covariance measurements and model outputs across different biomes [Schwalm *et al.*, 2010]. One of the purposes of the NACP synthesis was to compare the differences between modeled and observed values for as many site-model pairs as possible as a way of assessing the confidence of observed and modeled data, and of identifying ways to improve the confidence of both simulated and empirical measurements [Dietze *et al.*, 2011; Richardson *et al.*, 2012; Schaefer *et al.*, 2012; Stoy *et al.*, 2013]. In this study, we used a subset of sites and models that are included in the NACP interim site synthesis dataset. Models were run in accordance with the NACP protocol ([http://nacp.ornl.gov/docs/Site\\_Synthesis\\_Protocol\\_v7.pdf](http://nacp.ornl.gov/docs/Site_Synthesis_Protocol_v7.pdf)). We selected all sites for which flux data were available from at least 10 models, and all models that provided simulation data on a sub-daily basis for a minimum of 20 sites that encompass a range of North American ecosystem types in boreal to subtropical climate zones. Table 1 and 2 list the sites ( $n = 25$ ) and models ( $n = 9$ ) included in our analysis, though not all models were run for every site. Overall, we analyze 209 model-site pairs. It is important to acknowledge that under the NACP, modelers were given the choice of which sites to run or not run.

We used half-hourly or hourly observations, depending on the averaging time available from the site, of LE, H, and NEE for all the sites to compare with the models' simulation results for the same sites. These data are based on contributions by site PIs from the Ameriflux and Fluxnet Canada networks. As this paper focuses on the effects of hydrodynamic regulation of stomatal conductance, we used only data from the growing season, when transpiration was expected to be an important component of the site water balance and dominate the total evapotranspiration flux. We did not attempt to partition LE into transpiration and evaporation components [Stoy *et al.*, 2006]. While models or additional measurements may be able to provide the separate estimates for these two components of the total LE, the flux sites reported only direct measurements of total LE. In an effort to avoid

additional error due to partitioning uncertainty, we discuss only errors regarding model predictions of total evapotranspiration, in terms of LE, during a continuous period defined by the peak of the growing season when we expect transpiration from vegetation to be the dominant term of evapotranspiration. The length of the growing season was determined using the carbon-flux phenology approach [Garrity *et al.*, 2012] from gap-filled observational NEE provided in the NACP dataset. We further focused on the peak growing seasons by using observations during periods when carbon uptake flux was at least 50% of the peak one month moving average flux. It is worth noting that during the peak of the growing season, leaf area index is high at most of these sites (i.e. > 3.0), such that the contribution of soil evaporation to total ET should be minimal [Oishi *et al.*, 2008].

All flux observations in the dataset were filtered to remove periods of low turbulent mixing, during which eddy-covariance measurements are inaccurate. We conducted that filtering using site-specific friction velocity ( $u_*$ ) thresholds calculated following Reichstein [2005]. Only actual observations (not gap filled) were used, and analysis was only conducted where observations were available. Carbon fluxes were provided as the net ecosystem exchange of CO<sub>2</sub> (NEE), and its components, gross primary productivity (GPP) and ecosystem respiration. GPP was modeled from observations of NEE and daytime respiration estimated the relationship between nighttime NEE and microclimate variables by the site PIs contributing data to the NACP analysis.

Soil moisture and VPD were obtained from <http://public.ornl.gov/ameriflux/> for US sites and at <http://fluxnet-canada.ccrp.ec.gc.ca/> for the Canadian sites. In order to compare the amount of soil water available for plant uptake among sites, volumetric soil water content and site-specific soil properties were used to calculate soil water matric potential,  $\Psi_s$  following Rawls [1982]. Soil characteristics provided by the Ameriflux and Fluxnet Canada websites were used to approximate soil texture used in the determination of soil water content



at the permanent wilting point and field capacity following *Campbell and Norman* [1998].

Soil conditions were then divided on the basis of  $\Psi_s$  into three categorical levels, representing times when the soil was near the field capacity, near the permanent wilting point, and in the region of intermediate moisture conditions. VPD was divided to three categorical levels in a similar manner based on each site's seasonal mean daily VPD. Days for which the daily mean VPD fell within the highest 33% of the seasonal range of VPD for that site were considered 'high VPD' days; similarly, in 'low VPD' days the mean daily VPD fall within the bottom 33% and 'intermediate VPD' refers to the middle 33%.

## 2.2 Analysis of model errors

Model outputs of LE were compared to observed data for all sites at a half hourly resolution. For the few sites where observations were provided in hourly time steps (US-Ha1, US-MMS, US-Ne1, US-Ne2, US-Ne3, US-PFa, US-UBS), observations were linearly interpolated to half-hourly. The daily cycle was interpreted on the basis of local time in all sites, and UTC timestamps were converted to local time where needed. Error was calculated as the difference between modeled and observed values for each half hour. Errors in H, GPP, and NEE were also calculated and analyzed for correlations with LE error.

To determine the diurnal pattern of the errors, we calculated the mean error for each model-site pair for every half hour of the day using only peak growing-season data averaged over all days and years of observation at each site and normalized by the mean observed LE over all years for each site. We repeated the same analysis conditioned on the three levels (wet, intermediate, dry) of soil water matric potential or VPD. Grand average errors in LE for each site-model pair were calculated by averaging all normalized half-hourly errors for the growing seasons in all years and all soil water matric potential/VPD conditions. The average daily error per model was calculated as the mean across all sites of the mean daily error, and the average daily error per site was calculated as the mean error across models. Sites were

grouped based on ecosystem type and climate. Models were grouped based on the type of stomatal conductance parameterization scheme employed, the type of water stress function used, and to what terms in the stomata formulation water stress limitations were applied.

Models that enforced water stress limitations in the calculation for  $g_s$  were considered to be ‘direct’, models that enforce limitations through the carboxylation capacity were considered to be ‘indirect’ (Table 2).

### 2.3 Analysis of hysteresis of the diurnal water flux cycle

Transpiration exhibits a distinct hysteresis in the diurnal pattern that can be seen by plotting LE as a function of VPD during the course of each individual day [Chen *et al.*, 2011; O’Grady *et al.*, 2008]. The plot traces a clockwise loop characteristic of the system’s hysteresis (Figure 1a) [Verbeeck *et al.*, 2007a; Zhang *et al.*, 2014]. In this work, the ‘degree of hysteresis’ of LE was defined as the integrated area encompassed by the loop. A (0,0) point, representing minimal nighttime transpiration, was included as a boundary condition to numerically close the loop and allow calculation of the loops’ integrated interior area [Novick *et al.*, 2009]. Observed VPD and modeled LE were used to calculate “model simulated degree of hysteresis” shown as an example for the ED2 model at Metolius Intermediate field site, US-Me2, in Figure 1b. This metric of hysteresis was calculated for each day in each model or site data and then averaged across days, models, and sites. Degree of hysteresis was also averaged for each of the three soil water potential categories and VPD categories using the same categorical breakdown as discussed above based on the daily mean soil water matric potential and daily mean VPD.

(Figure 1 here)

(Tables 1 and 2 here)

### 3. Results and Discussion

#### 3.1 Errors in simulated LE at the long (seasonal) time-scale

Models were evaluated by their ability to simulate LE across multiple sites representing a variety of ecosystems. The mean daily growing-season errors across all sites for each model and across all models for each site are shown in Figure 2. No stomatal conductance formulation type was found to provide consistently the best or worst simulated LE across all sites or ecosystem types, (Figure 2b). However, the model that used *Leuning* [1995] formulation consistently produced an underestimation of LE, though this underestimation was not extreme in its magnitude relative to mean errors of other parameterization approaches and may be due to factors other than the model for stomatal conductance. Overall, all models produced a rather good estimation of LE, with mean half-hourly errors of less than 25% of the mean flux. The error of most models at most sites was less than 10% (Figure 2), similar to the range of error suggested for LE observations in flux sites [*Richardson et al.*, 2006]. The ecosystem type did not affect overall errors (Figure 2a). No significant correlations were found between mean growing season LE errors for each site and mean annual rainfall, elevation, mean annual temperature, ecosystem type, or LAI (data not shown). Errors in model-simulated LE were larger and more variable when daily average soil water content was moderate to low (Figure 3), and when daily average VPD was moderate to high (VPD data not shown). This suggests that models have difficulties simulating the dynamics that cause stomata to close during moderate soil-water content, and are probably not driven by soil water availability limitation of transpiration. Models performed best at extreme conditions, for example, when conditions were very wet, or when conditions are very dry and the stomata exhibit high resistance to gas/vapor flows. Larger errors in LE, in absolute terms, occurred during the peak hours of photosynthesis and transpiration (9am to 3pm, consistent with *Richardson et al.* [2006], i.e. when LE was largest).

(Figure 2 here)

(Figure 3 here)

### 3.2 Relationships between LE and carbon flux errors

Overall, errors in LE were found to have a positive, linear correlation with errors in GPP; however, the amount of variation explained by this correlation was low (Figure 4).

Although errors in modeled carbon flux may stem from a variety of sources, inclusive of errors in the observations of the forcing and the carbon fluxes [Richardson *et al.*, 2006] and model assumptions and parameterization, some of the error is related to the inaccuracy in the prediction of stomatal conductance, which directly affects both simulated LE and photosynthesis and thus creates a relationship between errors in C and errors latent heat fluxes. Errors in LE are not decoupled from errors in other model-simulated C flux variables.

For example, Dietze [2011] showed that errors in NEE relate to phenology and vegetative processes at multiple time scales. However, here we focused on the peak growing season to avoid the potential effect of errors in the modeled phenology.

(Figure 4 here)

Errors in LE and GPP can also be indicative of errors in the way radiation affects stomatal conductance, and carbon flux [Foken, 2008; Huntzinger *et al.*, 2012; Richardson *et al.*, 2012; Schaefer *et al.*, 2012]. For example, Foken [2008] found that most land-surface observational experiments show consistently lower combined LE and H than the available energy (net radiation + soil heat flux), which will result in lower simulated LE and GPP than observed.

### 3.3 Relationships between LE errors and the surface energy budget

Apparent errors in LE can be the result of errors in measurement of LE and not necessarily be indicative of a non-physical result by the model. Both observation errors in fluxes and errors in energy closure in models will lead to lack of energy closure, i.e., the sum

of the fluxes will be smaller (or larger) than the net radiation. Lack of energy closure in eddy covariance observation, particularly at the sub-daily time scale, may be a result of changes of the within canopy heat and humidity storage, unaccounted for ground heat flux and lateral advective fluxes. A simple assumption driven by conservation of energy, which is at the basis of any land-surface model, is that the net radiation equals the sum of sensible, latent, and soil heat fluxes. We can make the simplifying assumption that during the growing season when leaves are fully expanded, the soil heat flux is low relative to the contribution of latent and sensible heat fluxes, and does not vary greatly at a particular site (again, relative to latent and sensible heat fluxes). We can therefore approximate the energy deficit,  $E_x$ , as

$$E_x = R_d - LE - H \quad (1)$$

where  $R_d$  is the net radiation at the land surface.

We raise two alternative hypotheses: (1) errors in LE are the results of errors in the partitioning of the surface energy budget into LE and H, mostly due to an error in the value of stomatal conductance which leads to over or under estimation of LE and thus a compensation with an under or over estimation of H driven by the assumption of conservation of energy in the model; (2) errors in LE are a result of observation or model error in energy closure and thus an overall over/under estimation of the total surface fluxes will be distributed to errors in both LE and H. In the case of hypothesis (1) we predict that errors in H and errors in LE will be negatively correlated, while in the case of (2) the LE and H errors should be positively correlated. Additionally, in the case of (1) errors in LE or H should not be correlated with the observed  $E_x$ , while in the case of (2) both H, and LE errors (the sum of which defines the error in modeled  $E_x$ ) should be positively correlated with the observed  $E_x$ .

We analyzed the relationship between model errors in LE, model errors in H, observed  $E_x$ , and modeled  $E_x$  during the daytime hours of the growing season (between 9am – 6pm and when  $R_d > 20 \text{ W/m}^2$ ) and when turbulence mixing is high (above the  $u^*$

threshold). In these times, changes of heat or humidity storage in the canopy should be minimal and these time periods are when transpiration is the major component of LE (Table 3). Errors in LE and errors in H were very weakly correlated (all  $R^2 < 0.05$ ) with an inconsistent trend, positive in roughly half of the models and negative in the others (Table 3). Errors in LE and errors in H were positively correlated with errors in observed  $E_x$  ( $R^2 < 0.14$  and  $0.14$ , respectively). Modeled  $E_x$  was also positively correlated with observed  $E_x$  ( $R^2 < 0.33$ ) indicating that models perform less well when the site observations include a large energy closure error. The significance, strength and sign of these correlations show that it is impossible to reject either hypothesis and that models' errors are likely a combination of errors in LE that are largely related to models' or observations ability to close the energy balance; and that errors in LE and H are driven by compensation to conserve energy in several of the models we studied. Therefore, we can assume that, at least to some degree, causes of model errors in LE are independent of the energy budget and related to the calculation of transpiration and stomatal conductance.

(Table 3 here)

### 3.4 Diurnal patterns of LE error and hysteresis in the LE-VPD curve

Although the majority of site-model pairs tend to produce underestimations of LE, intra-daily patterns of errors in LE show both under and overestimations, frequently in a compensatory manner. These diurnal patterns of LE errors, for the most part, tend to be grouped well on the basis of the formulation of stomatal conductance in the different models (Figure 5). ED2, which uses the *Leuning* [1995] formulation, consistently underestimated LE, with most of the underestimation during the early part of the day. As the only model in its class, it is not clear what components of its formulation drive this type of error pattern compared to patterns in other models. CN-CLASS, ORCHIDEE, and Teco, which use the *Ball-Woodrow-Berry* [1987] scheme, tend to underestimate transpiration early in the day, and

overestimate it during the afternoon and early evening. Similar patterns are shown by Ecosys, which uses the *Grant-Black* [2006] scheme and exhibits the smallest overall bias. This error pattern corresponds with the pattern of error we predict to be caused due to the lack of plant hydrodynamics – overestimation of the flux late in the day may be driven by the lack of hydrodynamic restriction of stomatal conductance and therefore photosynthesis due to reduced storage levels and low leaf water potential. Parametrization using the long-term (daily, weekly, monthly) totals of the water flux adjusts the entire curve downwards (assumably through changes to parameters controlling light-use efficiency and reduction of maximal photosynthesis and stomatal conductance) and leads to underestimation of LE in the earlier parts of the day. Models using *Collatz-Ball* [1991] scheme show the inverted patterns and tend to overestimate LE early in the day, and then underestimate afternoon LE (Figure 5) which could be caused by an over-sensitivity to VPD.

(Figure 5 here)

At the site level, the majority of the largest diurnal errors in half-hourly LE occurred when the daily average soil conditions were in the range of intermediately wet and when VPD was high (data not shown). The majority of these large errors were underestimations of flux during the hours of 10am to 3pm. Sites did not tend to demonstrate similar patterns of error throughout the day when grouped by ecosystem type (Figure 6). Regardless of ecosystem type, the majority of sites showed an underestimation of LE during the morning and afternoon followed by an overestimation during the evening. However, some sites (e.g., CA-Let, CA-Gro, US-PFa, US-IB2, US-NR1, US-Dk3, US-Me2, US-UMB) show an underestimation during the entire day with slight to no overestimation in the evening and nighttime. Other sites show very little error or an over estimation during the early part of the day (e.g., CA-Oas, US-MMS, US-Ton, US-WCr). Crop sites show the largest intra-daily errors, indicating that models have the most difficulty simulating transpiration for these

ecosystems.

(Figure 6 here)

The underestimation of LE in the first half of the day and overestimation in the second, together with the observation most models approximate LE very well at the integrated daily time step, demonstrates that model parameterizations offset opposing errors.

The largest errors among the four models with the least bias (Ecosys, CN-CLASS, SibCASA, ORCHIDEE, Figure 2b) occur when soil is wet and intermediately wet (Figure 3a). The pattern of error indicates that they tend to underestimate evapotranspiration moderately throughout the entire day and more so during the hours of peak transpiration when the soil is wet and intermediately wet. These models tend to overestimate LE when the soil is dry. Conversely, the relatively largest errors for the models with the highest bias (SiB3, CanIBIS, SSiB2, Teco, ED2, Figure 2b) tend to occur when the soil is dry or intermediately wet (Figure 3b). They show a similar tendency to underestimate LE during peak transpiration but overestimate during the mornings and evenings for all soil conditions. Though nighttime observations of fluxes have larger errors than during the day, due to weak wind and turbulent mixing and thus larger fraction of filtered observations [Novick *et al.*, 2009], the fact that models consistently overestimate nighttime evaporation, rather than fluctuate with larger errors, suggests that at least in part, this nighttime overestimation is driven by parameterization of the models for the total water budget that compensates for underestimation of daytime transpiration. This compensation, particularly during early morning and late evening hours, when transpiration is low, allows models to be reasonably accurate at the overall daily time scale, while having consistent bias patterns of the diurnal dynamics.

The differences between the simulated and observed diurnal dynamics of LE are apparent when analyzed within the framework of the daily hysteresis of LE as a function of



VPD. We observed positive hysteresis in all sites in most days (Figure 7). Reversible hysteresis of this type (i.e., the next day can repeat a similar cycle) is usually the result of a numerical reduction to a single dimension of a multi-dimensional process. There is a debate as to the cause of this hysteresis, whether it is the result of the lag between the daily pattern of VPD and the level of photosynthetic active radiation (PAR), or due to stomatal conductance inherent dependence on the plant hydrodynamics through hydration status at the leaf and branch level [Tuzet *et al.*, 2003; Verbeeck *et al.*, 2007a; Verbeeck *et al.*, 2007b; Zhang *et al.*, 2014]. In the first case, LE is a repeatable function of both VPD and PAR and the same level of LE is expected when a specific combination of LE and PAR would occur. Hysteresis in the LE curve will be a result of plotting LE as a single axis function of VPD while ignoring the PAR axis. The daily peak of PAR is earlier than the daily peak of VPD and thus higher transpiration levels occur before the daily max of VPD than later in the day. In the case of plant hydrodynamics, LE would be a function of VPD and water storage levels in the plant, as expressed by leaf water potential. Hysteresis may be driven by midday stomata closure [Brodribb and Holbrook, 2004; Sperry *et al.*, 2002; Tyree and Sperry, 1989; Tyree and Zimmermann, 2002]. Water transport from the root is, in most cases, slower than the maximal rate of transpirational water loss from the leaf and therefore, when storage is depleted, maximal transpiration rates cannot be sustained. Throughout the day, capacity becomes depleted due to ongoing transpiration. As this hydraulic process occurs stomata respond to reduced leaf water potential by reducing their aperture. This pattern of stomatal response leads to a decline in the afternoon transpiration relative to the pre-noon rates and appears as hysteresis in the LE-VPD curve (Figure1). In the physical world, LE is a function of VPD, leaf water potential and PAR, along with other variables (temperature, plant physiology, phenology, etc.), and therefore it is plausible that the apparent hysteresis is a result of a combination of drivers.

Following our assumptions about the potential sources for the hysteresis, if the source of hysteresis is mostly due to the lag between PAR and VPD, models should be able to resolve the hysteresis well because the stomata parameterizations primarily describe the PAR and VPD response of stomata. In fact, models that overemphasize the role of VPD or PAR in modulating stomatal conductance will predict larger hysteresis than observed. However, the component of the hysteresis that is driven by the lag between VPD-driven stomatal conductance response and plant water storage level will not be resolved by models and thus, will lead to lower simulated hysteresis than observed.

Overall, most models tend to under-predict the degree of hysteresis in all sites (Table 4). Underestimation of hysteresis was relatively greater for sites where the observed hysteresis was either very low or very high (Figure 7). Although, for our study, we do not have data available by which to directly calculate plant water status, we use soil water potential as a measure of how much water is available in the soil for root uptake. Similarly to the findings of *O'Grady* [2008] and *Chen* [2011], we find that the largest values for observed hysteresis occur when daily mean VPD is moderate to large while the smallest occur when VPD is lowest (Table 4). We also find that as observed hysteresis becomes large, model performance declines relative to periods characterized by moderate hysteresis (Figure 7). We suggest that as hysteresis becomes large, it is more dependent on plant-hydrodynamic factors than abiotic factors and that this “tailing-off” of models’ ability to accurately simulate these effects is also related to the lack of a mechanistic manner with which to simulate changing plant water status and the related stomatal closure.

Evergreen needleleaf forests tended to have the lowest observed degree of hysteresis, while crop sites tended to have the highest (Figure 7). The effects of different components of the conductive system, such as the area of conductive xylem, the xylem maximal conductance and capacitance, the shape of the cavitation curve, the sensitivity of stomata to

leaf water potential, as well as the response to foliar abscisic acid all combine to describe the hydrodynamic component of stomatal conductance. In this study, we do not have the observations needed to determine which of these are specifically responsible for the differences between site types. We also cannot fully determine to what degree the errors in the hysteresis are driven by errors in the temporal hydrodynamics or errors in the modelled lag between the timing of max VPD and timing of max transpiration (as suggested by *Zhang et al* [2014]). Nonetheless, at a very simplistic level, our observations are consistent with the fact that tall conifer forests (large, low wood density) are expected to have the largest amount of storage per transpiring leaf area and thus encounter less and later daily hydrodynamic restriction of stomatal conductance and therefore lower hysteresis, while crops have the lowest storage per leaf area. These results are further supported by a recent study from *McAdam and Brodribb* [2014] that showed no hysteresis in the coniferous species *Metasequoia glyptostroboides* due to lack of ABA response under evaporative stress.

(Figure 7 here)

(Table 4 here)

Models using the *Collatz-Ball* [1991] formulation tended to produce larger-than-observed hysteresis, while all other model types consistently underestimate hysteresis. None of the models are driven by a mechanistically-physical representation of the plant hydrodynamics and therefore, all are expected to produce at least some degree of underestimation of the hysteresis. We propose that the Collatz-Ball-based models overemphasize the effects of VPD and, as we predicted above, generate an overall overestimation the hysteresis. This is further supported by the fact that all and only these models produce the specific pattern of overestimation of transpiration in the early part of the day when light is the major driver of stomatal conductance (Figure 5c). In contrast, all other models over-estimate stomatal conductance in the later part of the day, as we hypothesize is

the characteristic pattern of lack of modeled plant hydrodynamics (Figure 5a, b, and d). It is noteworthy that Ecosys, the only model employing the *Grant-Black* [2006] formulation, shows the least amount of overall bias (Figure 2b) in the calculation of LE and has a relatively subtle pattern of diurnal error (Figure 5d). The Grant-Black formulation is distinguished in two ways: (1) it provides no direct link between VPD and  $g_s$  and (2) it is the only formulation to directly employ a diagnostic variable representing the plant water potential that controls the stomata response to water limitations, rather than to use a Jarvis-like or Feddes-like water stress factor, which relates stomatal conductance directly to the soil. By subtly altering the formulation of stomatal conductance to include a slightly more physical approach to link conductance to plant water status, Grant-Black may have made an increase in model accuracy. However, with only one model in our study employing this stomatal formulation, it remains unclear if this is the only factor driving its accuracy.

#### **4. Conclusions**

Land-surface models are essential tools for predicting how rising atmospheric CO<sub>2</sub> concentrations and climate change will affect our environment and the ways in which we interpret it. While we found that, overall, models resolve the long-term water budget relatively well, errors in simulated LE were found to be correlated with errors in gross primary productivity and net ecosystem exchange. We suggest that model performance in terms of surface heat fluxes, as well as carbon fluxes could be improved by improving the simulation of the latent heat flux, and particularly stomatal conductance. Our results indicate that intra-daily patterns of errors in LE as simulated by current land-surface models are likely driven to some degree by models lack of ability to account for the daily hysteretic response of transpiration to VPD. This deficiency along with the patterns of the errors in modeled LE is indicative in most models of missing hydrodynamic responses at the leaf level, which limit accurate predictions of stomatal conductance and therefore transpiration. We propose that the

poor representation in present land-surface models of plant hydrodynamic fluctuations stems from the lack of a mechanistic tool with which to simulate water flux from the soil through the vegetative system. In order to improve model robustness, accuracy in the simulation of the sub-daily time scale must be achieved.

Overall, at daily, monthly and seasonal time scales, the models were accurate with model error rate close to the observation uncertainty, which defines the limit of achievable accuracy. Nonetheless, the diurnal pattern of errors at the hourly time scale may accumulate to affect particular large-scale dynamics in weather models. At short time scales, these errors in the modeled LE will lead to errors in model prediction of temperature which will be reflected in simulated H and in the response of photosynthesis to temperature and thermal injury [Dietze, 2014; Sage and Kubien, 2007]. We hypothesize that in areas with strong diurnal patterns of wind, such as places with lake or sea breeze, or mountains with thermally driven slope flows, trading off morning-time transpiration for afternoon transpiration will lead to different advection paths for humidity which will consequently change the long-term mean humidity downwind of the transpiring vegetation. Cloud formation is another process with a strong diurnal cycle where differences in the intra-daily dynamics of transpiration may affect moisture availability and thermal uplift strength at key times for specific cloud processes [de Arellano *et al.*, 2014; LeMone and Pennell, 1976]. Similarly, the diurnal dynamics of stomatal conductance affect processes other than evaporation and photosynthesis. For example emissions rates of some volatile organic compounds (VOC) from certain species of plants (e.g. *Fagus sylvatica*, *Quercus robur*, *Betula pendula*, *Eucalyptus camaldulensis*) are dependent on stomatal conductance [Folkers *et al.*, 2008; Hüve *et al.*, 2007; Niinemets and Reichstein, 2003; Winters *et al.*, 2009]. The large-scale consequences of these VOC emissions are determined by the reactions and interactions of these VOC with other chemicals in the atmosphere (e.g., ozone), light availability and

temperature, all of which have strong diurnal dynamics. Therefore, changing the diurnal temporal dynamics of stomatal conductance and VOC emissions may result in non-linear interactions and drive a large-scale difference in the resulting atmospheric chemistry. We propose that an approach for mechanistic representation of plant water potential and its effects on stomatal conductance at a plot scale is needed to increase accuracy at a sub-daily time scale in addition to longer periods of drought, which are predicted to become longer and more frequent across a number of areas with global climate change.

### **Acknowledgements**

We acknowledge the contribution of site PIs in the Ameriflux and FluxCanada networks that contributed site data to this analysis and model PIs that contributed simulation results. This research was supported by the U.S. Department of Energy (DOE) awards #DE-SC0006708 and DE-SC0007041 and the National Science Foundation grants #DEB-0911461 including an REU supplement award for AM, and additional support by the National Oceanic and Atmospheric Administration award #NA11OAR4310190. Data analysis was conducted in part at the Ohio Supercomputer Center, as part of computation allocation grant #PAS0409-4. PCS acknowledges support from the US NSF EF #1241881.

## References

- Allen, C. D., et al. (2010), A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests, *For. Ecol. Manage.*, 259(4), 660-684.
- Arain, M. A., F. M. Yuan, and T. A. Black (2006), Soil-plant nitrogen cycling modulated carbon exchanges in a western temperate conifer forest in Canada, *Agric. For. Meteorol.*, 140(1-4), 171-192.
- Baker, I. T., L. Prihodko, A. S. Denning, M. Goulden, S. Miller, and H. R. da Rocha (2008), Seasonal drought stress in the Amazon: reconciling models and observations, *J. Geophys. Res.*, 113.
- Ball, J. T. (1988), An analysis of stomatal conductance, 145 pp, Stanford University.
- Ball, J. T., I. E. Woodrow, and J. A. Berry (1987), A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions, *Progress in Photosynthesis Research*, 4, 221-224.
- Barr, A. G., T. A. Black, E. H. Hogg, N. Kljun, K. Morgenstern, and Z. Nesic (2004), Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production, *Agric. For. Meteorol.*, 126(3-4), 237-255.
- Beer, C., et al. (2010), Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate, *Science*, 329(5993), 834-838.
- Bergeron, O., H. A. Margolis, T. A. Black, C. Coursolle, A. L. Dunn, A. G. Barr, and S. C. Wofsy (2007), Comparison of carbon dioxide fluxes over three boreal black spruce forests in Canada, *Glob. Change Biol.*, 13(1), 89-107.
- Berry, J. A., D. J. Beerling, and P. J. Franks (2010), Stomata: key players in the earth system, past and present, *Curr. Opin. Plant Biol.*, 13(3), 232-239.
- Bohrer, G., H. Mourad, T. A. Laursen, D. Drewry, R. Avissar, D. Poggi, R. Oren, and G. G. Katul (2005), Finite element tree crown hydrodynamics model (FETCH) using porous media flow within branching elements: A new representation of tree hydrodynamics, *Water Resour. Res.*, 41(11).
- Bradford, J. B., R. A. Birdsey, L. A. Joyce, and M. G. Ryan (2008), Tree age, disturbance history, and carbon stocks and fluxes in subalpine Rocky Mountain forests, *Glob. Change Biol.*, 14(12), 2882-2897.
- Brodribb, T. J., and N. M. Holbrook (2003), Stomatal closure during leaf dehydration, correlation with other leaf physiological traits, *Plant Physiol.*, 132(4), 2166-2173.
- Brodribb, T. J., and N. M. Holbrook (2004), Stomatal protection against hydraulic failure: a comparison of coexisting ferns and angiosperms, *New Phytol.*, 162(3), 663-670.
- Brodribb, T. J., and N. M. Holbrook (2006), Declining hydraulic efficiency as transpiring leaves desiccate: two types of response, *Plant Cell Environ.*, 29(12), 2205-2215.
- Brodribb, T. J., and N. M. Holbrook (2007), Forced depression of leaf hydraulic conductance in situ: effects on the leaf gas exchange of forest trees, *Funct. Ecol.*, 21(4), 705-712.
- Brodribb, T. J., and H. Cochard (2008), Hydraulic failure defines the recovery and point of death in water-stressed conifers, *Plant Physiol.*, 149(1), 575-584.
- Campbell, G. S., and J. M. Norman (1998), *An introduction to environmental biophysics*, Second edition, xxi+286p pp., Springer-Verlag, Heidelberger Platz 3, D-1000 Berlin, Germany; Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, New York 10010, USA.
- Chen, L. X., Z. Q. Zhang, Z. D. Li, J. W. Tang, P. Caldwell, and W. J. Zhang (2011), Biophysical control of whole tree transpiration under an urban environment in Northern China, *J. Hydrol.*, 402(3-4), 388-400.
- Choat, B., et al. (2012), Global convergence in the vulnerability of forests to drought, *Nature*, 491(7426), 752-756.



- Collatz, G. J., J. T. Ball, C. Grivet, and J. A. Berry (1991), Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminal boundary layer, *Agric. For. Meteorol.*, 54, 107-136.
- Cook, B. D., et al. (2004), Carbon exchange and venting anomalies in an upland deciduous forest in northern Wisconsin, USA, *Agric. For. Meteorol.*, 126(3-4), 271-295.
- Damour, G., T. Simonneau, H. Cochard, and L. Urban (2010), An overview of models of stomatal conductance at the leaf level, *Plant Cell Environ.*
- Davis, K. J., P. S. Bakwin, C. X. Yi, B. W. Berger, C. L. Zhao, R. M. Teclaw, and J. G. Isebrands (2003), The annual cycles of CO<sub>2</sub> and H<sub>2</sub>O exchange over a northern mixed forest as observed from a very tall tower, *Glob. Change Biol.*, 9(9), 1278-1293.
- de Arellano, J. V. G., H. G. Ouwensloot, D. Baldocchi, and C. M. J. Jacobs (2014), Shallow cumulus rooted in photosynthesis, *Geophys. Res. Lett.*, 41(5), 1796-1802.
- De Kauwe, M. G., et al. (2013), Forest water use and water use efficiency at elevated CO<sub>2</sub>: a model-data intercomparison at two contrasting temperate forest FACE sites, *Glob. Change Biol.*, 19(6), 1759-1779.
- Desai, A. R., P. V. Bolstad, B. D. Cook, K. J. Davis, and E. V. Carey (2005), Comparing net ecosystem exchange of carbon dioxide between an old-growth and mature forest in the upper Midwest, USA, *Agric. For. Meteorol.*, 128(1-2), 33-55.
- Dietze, M. C. (2014), Gaps in knowledge and data driving uncertainty in models of photosynthesis, *Photosynth. Res.*, 119(1-2), 3-14.
- Dietze, M. C., et al. (2011), Characterizing the performance of ecosystem models across time scales: A spectral analysis of the North American Carbon Program site-level synthesis, *J. Geophys. Res.*, 116(G4).
- Egea, G., A. Verhoef, and P. L. Vidale (2011), Towards an improved and more flexible representation of water stress in coupled photosynthesis-stomatal conductance models, *Agric. For. Meteorol.*, 151(10), 1370-1384.
- Feddes, R. A., P. Kowalik, K. Kolinska-Malinka, and H. Zarandy (1976), Simulation of field water uptake by plants using a soil water dependent root extraction function, *J. Hydrol.*, 31, 13-26.
- Fischer, M. L., D. P. Billesbach, J. A. Berry, W. J. Riley, and M. S. Torn (2007), Spatiotemporal variations in growing season exchanges of CO<sub>2</sub>, H<sub>2</sub>O, and sensible heat in agricultural fields of the Southern Great Plains, *Earth Interact.*, 11.
- Fisher, J. B., T. A. DeBiase, Y. Qi, M. Xu, and A. H. Goldstein (2005), Evapotranspiration models compared on a Sierra Nevada forest ecosystem, *Environ. Modell. Softw.*, 20(6), 783-796.
- Flanagan, L. B., L. A. Wever, and P. J. Carlson (2002), Seasonal and interannual variation in carbon dioxide exchange and carbon balance in a northern temperate grassland, *Glob. Change Biol.*, 8(7), 599-615.
- Foken, T. (2008), The energy balance closure problem: an overview, *Ecol. Appl.*, 18(6), 1351-1367.
- Folkers, A., K. Huve, C. Ammann, T. Dindorf, J. Kesselmeier, E. Kleist, U. Kuhn, R. Uerlings, and J. Wildt (2008), Methanol emissions from deciduous tree species: dependence on temperature and light intensity, *Plant biology (Stuttgart, Germany)*, 10(1), 65-75.
- Ford, C. R., R. M. Hubbard, B. D. Kloeppel, and J. M. Vose (2007), A comparison of sap flux-based evapotranspiration estimates with catchment-scale water balance, *Agric. For. Meteorol.*, 145(3-4), 176-185.
- Garrity, S. R., K. Meyer, K. D. Maurer, B. Hardiman, and G. Bohrer (2012), Estimating plot-level tree structure in a deciduous forest by combining allometric equations, spatial wavelet analysis and airborne LiDAR, *Remote Sens. Lett.*, 3(5), 443-451.



- Gough, C. M., B. S. Hardiman, L. E. Nave, G. Bohrer, K. D. Maurer, C. S. Vogel, K. J. Nadelhoffer, and P. S. Curtis (2013), Sustained carbon uptake and storage following moderate disturbance in a Great Lakes forest, *Ecol. Appl.*, 23(5), 1202-1215.
- Grant, R. F., T. A. Black, D. Gaumont-Guay, N. Kljun, A. G. Barr, K. Morgenstern, and Z. Nesic (2006), Net ecosystem productivity of boreal aspen forests under drought and climate change: Mathematical modelling with Ecosys, *Agric. For. Meteorol.*, 140(1-4), 152-170.
- Griffis, T. J., T. A. Black, K. Morgenstern, A. G. Barr, Z. Nesic, G. B. Drewitt, D. Gaumont-Guay, and J. H. McCaughey (2003), Ecophysiological controls on the carbon balances of three southern boreal forests, *Agric. For. Meteorol.*, 117(1-2), 53-71.
- Gu, L. H., T. Meyers, S. G. Pallardy, P. J. Hanson, B. Yang, M. Heuer, K. P. Hosman, J. S. Riggs, D. Sluss, and S. D. Wullschlegel (2006), Direct and indirect effects of atmospheric conditions and soil moisture on surface energy partitioning revealed by a prolonged drought at a temperate forest site, *J. Geophys. Res.*, 111(D16).
- Hickler, T., I. C. Prentice, B. Smith, M. T. Sykes, and S. Zaehle (2006), Implementing plant hydraulic architecture within the LPJ Dynamic Global Vegetation Model, *Global Ecology and Biogeography*, 15(6), 567-577.
- Huntington, T. G., A. D. Richardson, K. J. McGuire, and K. Hayhoe (2009), Climate and hydrological changes in the northeastern United States: recent trends and implications for forested and aquatic ecosystems, *Canadian Journal of Forest Research*, 39(2), 199-212.
- Huntzinger, D. N., et al. (2012), North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison, *Ecol. Model.*, 232, 144-157.
- Hüve, K., M. M. Christ, E. Kleist, R. Uerlings, Ü. Niinemets, A. Walter, and J. Wildt (2007), Simultaneous growth and emission measurements demonstrate an interactive control of methanol release by leaf expansion and stomata, *J. Exp. Bot.*, 58(7), 1783-1793.
- Janott, M., S. Gayler, A. Gessler, M. Javaux, C. Klier, and E. Priesack (2011), A one-dimensional model of water flow in soil-plant systems based on plant architecture, *Plant Soil*, 341(1-2), 233-256.
- Jarvis, P. G. (1976), The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field *Philos. Trans. R. Soc. B-Biol. Sci.*, 273, 593-610.
- Jasechko, S., Z. D. Sharp, J. J. Gibson, S. J. Birks, Y. Yi, and P. J. Fawcett (2013), Terrestrial water fluxes dominated by transpiration, *Nature*, 496(7445), 347-350.
- Katul, G. G., R. Oren, S. Manzoni, C. Higgins, and M. B. Parlange (2012), Evapotranspiration: a process driving mass transport and energy exchange in the soil-plant-atmosphere-climate system, *Rev. Geophys.*, 50(3).
- Kucharik, C. J., J. A. Foley, C. Delire, V. A. Fisher, M. T. Coe, J. D. Lenters, C. Young-Molling, N. Ramankutty, J. M. Norman, and S. T. Gower (2000), Testing the performance of a dynamic global ecosystem model: water balance, carbon balance, and vegetation structure, *Glob. Biogeochem. Cycle*, 14(3), 795-825.
- LeMone, M. A., and W. T. Pennell (1976), The relationship of trade wind cumulus distribution to subcloud layer fluxes and structure, *Mon. Weather Rev.*, 104(5), 524-539.
- Leuning, R. (1995), A critical appraisal of a combined stomatal-photosynthesis model for C3 plants, *Plant Cell Environ.*, 18, 339-355.
- Ma, S. Y., D. D. Baldocchi, L. K. Xu, and T. Hehn (2007), Inter-annual variability in carbon dioxide exchange of an oak/grass savanna and open grassland in California, *Agric. For. Meteorol.*, 147(3-4), 157-171.
- McAdam, S. A. M., and T. J. Brodribb (2014), Separating active and passive influences on stomatal control of transpiration, *Plant Physiol.*, 164(4), 1578-1586.
- McCaughey, J. H., M. R. Pejam, M. A. Arain, and D. A. Cameron (2006), Carbon dioxide

- and energy fluxes from a boreal mixedwood forest ecosystem in Ontario, Canada, *Agric. For. Meteorol.*, 140(1-4), 79-96.
- McCulloh, K. A., and J. S. Sperry (2005), Patterns in hydraulic architecture and their implications for transport efficiency, *Tree Physiol.*, 25, 257-267.
- McCulloh, K. A., D. M. Johnson, F. C. Meinzer, S. L. Voelker, B. Lachenbruch, and J.-C. Domec (2012), Hydraulic architecture of two species differing in wood density: opposing strategies in co-occurring tropical pioneer trees, *Plant Cell Environ.*, 35(1), 116-125.
- Medvigy, D., S. C. Wofsy, J. W. Munger, D. Y. Hollinger, and P. R. Moorcroft (2009), Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2, *J. Geophys. Res.*, 114.
- Nave, L. E., et al. (2011), Disturbance and the resilience of coupled carbon and nitrogen cycling in a north temperate forest, *J. Geophys. Res.*, 116(G4).
- Niinemets, U., and M. Reichstein (2003), Controls on the emission of plant volatiles through stomata: Differential sensitivity of emission rates to stomatal closure explained, *J. Geophys. Res.-Atmos.*, 108(D7), 17.
- Niyogi, D., K. Alapaty, S. Raman, and F. Chen (2009), Development and evaluation of a coupled photosynthesis-based gas exchange evapotranspiration model (GEM) for mesoscale weather forecasting applications, *J. Appl. Meteorol. Climatol.*, 48(2), 349-368.
- Novick, K. A., R. Oren, P. C. Stoy, M. B. S. Siqueira, and G. G. Katul (2009), Nocturnal evapotranspiration in eddy-covariance records from three co-located ecosystems in the Southeastern US: Implications for annual fluxes, *Agric. For. Meteorol.*, 149(9), 1491-1504.
- O'Grady, A. P., D. Worledge, and M. Battaglia (2008), Constraints on transpiration of *Eucalyptus globulus* in southern Tasmania, Australia, *Agric. For. Meteorol.*, 148(3), 453-465.
- Oishi, A. C., R. Oren, and P. C. Stoy (2008), Estimating components of forest evapotranspiration: A footprint approach for scaling sap flux measurements, *Agric. For. Meteorol.*, 148(11), 1719-1732.
- Pan, Y., et al. (2011), A large and persistent carbon sink in the world's forests, *Science*, 333(6045), 988-993.
- Phillips, N. G., M. G. Ryan, B. J. Bond, N. G. McDowell, T. M. Hinckley, and J. Cermak (2003), Reliance on stored water increases with tree size in three species in the Pacific Northwest, *Tree Physiol.*, 23(4), 237-245.
- Post, W. M., R. C. Izaurralde, J. D. Jastrow, B. A. McCarl, J. E. Amonette, V. L. Bailey, P. M. Jardine, T. O. West, and J. Z. Zhou (2004), Enhancement of carbon sequestration in US soils, *Bioscience*, 54(10), 895-908.
- Rawls, W. J., D. L. Brakensiek, and K. E. Saxton (1982), Estimation of soil water properties, *Trans. ASAE*, 25(5), 1316-1328.
- Reichstein, M., et al. (2005), On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Glob. Change Biol.*, 11(9), 1424-1439.
- Richardson, A. D., D. Y. Hollinger, D. B. Dail, J. T. Lee, J. W. Munger, and J. O'Keefe (2009), Influence of spring phenology on seasonal and annual carbon balance in two contrasting New England forests, *Tree Physiol.*, 29(3), 321-331.
- Richardson, A. D., et al. (2006), A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes, *Agric. For. Meteorol.*, 136(1-2), 1-18.
- Richardson, A. D., et al. (2012), Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis, *Glob. Change Biol.*, 18(2), 566-584.
- Sage, R. F., and D. S. Kubien (2007), The temperature response of C-3 and C-4

- photosynthesis, *Plant Cell Environ.*, 30(9), 1086-1106.
- Savabi, M. R., and C. O. Stockle (2001), Modeling the possible impact of increased CO<sub>2</sub> and temperature on soil water balance, crop yield and soil erosion, *Environ. Modell. Softw.*, 16(7), 631-640.
- Schaefer, K., G. J. Collatz, P. Tans, A. S. Denning, I. Baker, J. Berry, L. Prihodko, N. Suits, and A. Philpott (2008), Combined Simple Biosphere/Carnegie-Ames-Stanford Approach terrestrial carbon cycle model, *J. Geophys. Res.*, 113(G3).
- Schaefer, K., et al. (2012), A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis, *J. Geophys. Res.*, 117(G3).
- Schmid, H. P., C. S. B. Grimmond, F. Cropley, B. Offerle, and H. B. Su (2000), Measurements of CO<sub>2</sub> and energy fluxes over a mixed hardwood forest in the mid-western United States, *Agric. For. Meteorol.*, 103(4), 357-374.
- Schwalm, C. R., T. A. Black, K. Morgenstern, and E. R. Humphreys (2007), A method for deriving net primary productivity and component respiratory fluxes from tower-based eddy covariance data: a case study using a 17-year data record from a Douglas-fir chronosequence, *Glob. Change Biol.*, 13(2), 370-385.
- Schwalm, C. R., et al. (2010), A model-data intercomparison of CO<sub>2</sub> exchange across North America: Results from the North American Carbon Program site synthesis, *J. Geophys. Res.*, 115.
- Sellers, P. J., M. D. Heiser, F. G. Hall, S. J. Goetz, D. E. Strebel, S. B. Verma, R. L. Desjardins, P. M. Schuepp, and J. I. MacPherson (1995), Effects of spatial variability in topography, vegetation cover and soil moisture on area-averaged surface fluxes: a case study using the FIFE 1989 data, *J. Geophys. Res.*, 100(D12), 25607-25629.
- Siqueira, M. B., G. G. Katul, D. A. Sampson, P. C. Stoy, J. Y. Juang, H. R. McCarthy, and R. Oren (2006), Multiscale model intercomparisons of CO<sub>2</sub> and H<sub>2</sub>O exchange rates in a maturing southeastern US pine forest, *Glob. Change Biol.*, 12(7), 1189-1207.
- Sperry, J. S., U. G. Hacke, R. Oren, and J. P. Comstock (2002), Water deficits and hydraulic limits to leaf water supply, *Plant Cell Environ.*, 25(2), 251-263.
- Stoy, P. C., G. G. Katul, M. B. S. Siqueira, J.-Y. Juang, K. A. Novick, H. R. McCarthy, A. C. Oishi, J. M. Uebelherr, H.-S. Kim, and R. Oren (2006), Separating the effects of climate and vegetation on evapotranspiration along a successional chronosequence in the southeastern US, *Glob. Change Biol.*, 12(11), 2115-2135.
- Stoy, P. C., et al. (2013), Evaluating the agreement between measurements and models of net ecosystem exchange at different times and time scales using wavelet coherence: an example using data from the North American Carbon Program Site-Level Interim Synthesis, *Biogeosciences Discuss.*, 10(2), 3039-3077.
- Thomas, C. K., B. E. Law, J. Irvine, J. G. Martin, J. C. Pettijohn, and K. J. Davis (2009), Seasonal hydrology explains interannual and seasonal variation in carbon and water exchange in a semiarid mature ponderosa pine forest in central Oregon, *J. Geophys. Res.*, 114.
- Thomsen, J., G. Bohrer, A. M. Matheny, V. Y. Ivanov, L. He, H. Renninger, and K. Schäfer (2013), Contrasting hydraulic strategies during dry soil conditions in *Quercus rubra* and *Acer rubrum* in a sandy site in Michigan, *Forests*, 4(4), 1106-1120.
- Tuzet, A., A. Perrier, and R. Leuning (2003), A coupled model of stomatal conductance, photosynthesis and transpiration, *Plant Cell Environ.*, 26, 1097-1116.
- Tyree, M. T., and J. S. Sperry (1989), Vulnerability of xylem to cavitation and embolism, *Annu. Rev. Plant Phys. Mol. Bio.*, 40, 19-36.
- Tyree, M. T., and M. H. Zimmermann (2002), Xylem structure and the ascent of sap, in *Xylem structure and the ascent of sap*, edited, p. i, Springer-Verlag New York Inc., 175 Fifth Avenue, New York, NY, 10010-7858, USA; Springer-Verlag GmbH & Co. KG,

- Heidelberger Platz 3, D-14197, Berlin, Germany.
- Urbanski, S., C. Barford, S. Wofsy, C. Kucharik, E. Pyle, J. Budney, K. McKain, D. Fitzjarrald, M. Czikowsky, and J. W. Munger (2007), Factors controlling CO<sub>2</sub> exchange on timescales from hourly to decadal at Harvard Forest, *J. Geophys. Res.*, *112*(G2).
- Verbeeck, H., P. Peylin, C. Bacour, D. Bonal, K. Steppe, and P. Ciais (2011), Seasonal patterns of CO<sub>2</sub> fluxes in Amazon forests: fusion of eddy covariance data and the ORCHIDEE model, *J. Geophys. Res.*, *116*.
- Verbeeck, H., K. Steppe, N. Nadezhdina, M. O. De Beeck, G. Deckmyn, L. Meiresonne, R. Lemeur, J. Cermak, R. Ceulemans, and I. A. Janssens (2007a), Model analysis of the effects of atmospheric drivers on storage water use in Scots pine, *Biogeosciences*, *4*(4), 657-671.
- Verbeeck, H., K. Steppe, N. Nadezhdina, M. Op de Beeck, G. Deckmyn, L. Meiresonne, R. Lemeur, J. Cermak, R. Ceulemans, and I. A. Janssens (2007b), Stored water use and transpiration in Scots pine: a modeling analysis with ANAFORE, *Tree Physiol.*, *27*(12), 1671-1685.
- Verma, S. B., et al. (2005), Annual carbon dioxide exchange in irrigated and rainfed maize-based agroecosystems, *Agric. For. Meteorol.*, *131*(1-2), 77-96.
- Weng, E. S., and Y. Q. Luo (2008), Soil hydrological properties regulate grassland ecosystem responses to multifactor global change: a modeling analysis, *J. Geophys. Res.*, *113*(G3).
- Winters, A. J., M. A. Adams, T. M. Bleby, H. Rennenberg, D. Steigner, R. Steinbrecher, and J. Kreuzwieser (2009), Emissions of isoprene, monoterpene and short-chained carbonyl compounds from *Eucalyptus* spp. in southern Australia, *Atmospheric Environment*, *43*(19), 3035-3043.
- Wu, Y. P., S. G. Liu, and O. I. Abdul-Aziz (2012), Hydrological effects of the increased CO<sub>2</sub> and climate change in the Upper Mississippi River Basin using a modified SWAT, *Clim. Change*, *110*(3-4), 977-1003.
- Zhang, Q., S. Manzoni, G. G. Katul, A. Porporato, and D. Yang (2014), The hysteretic evapotranspiration—vapor pressure deficit relation, *Journal of Geophysical Research: Biogeosciences*, 2013JG002484.
- Zimmerman, M. H. (1983), *Xylem structure and the ascent of sap*, 125 pp., Springer-Verlag, Berlin.

**Table 1: Site Summary -site details (Latitude, Longitude, Elevation, Soil Class, and IGBP for USA sites from <http://public.ornl.gov/ameriflux/> and for Canada sites from <http://fluxnet-canada.ccrp.ec.gc.ca/> Mean annual rainfall as calculated by Schwalm et al. [2010].**

Site	Name	Lat.	Long.	Elev. (m a.s.l.)	Ann. precip (mm)	Years	Soil class	IGBP	Reference
CA-Ca1	Campbell River	49.87	-125.33	300	1256	1998-2006	Sandy loam	ENF	<i>Schwalm</i> [2007]
CA-Gro	Groundhog River	48.22	-82.16	200	427	2004-2006	Sandy loam	MF	<i>McCaughey</i> [2006]
CA-Let	Lethbridge	49.71	-112.94	960	335	1997-2007	Clay loam	GRA	<i>Flanagan</i> [2002]
CA-Oas	SSA Old Aspen	53.63	-106.20	530	460	1997-2006	Loam	DBF	<i>Barr</i> [2004]
CA-Obs	SSA Old Black Spruce	53.99	-105.12	629	470	2000-2006	Sandy loam	ENF	<i>Griffis</i> [2003]
CA-Ojp	SSA Old Jack Pine	53.92	-104.69	579	461	2000-2006	Sand	ENF	<i>Griffis</i> [2003]
CA-Qfo	Quebec Mature Forest	49.69	-74.34	382	819	2004-2006	Sandy loam	ENF	<i>Bergeron</i> [2007]
US-ARM	Lamot S. Great Plains	36.61	-97.49	310	629	2000-2007	Silty clay loam	CRO	<i>Fischer</i> [2007]
US-Dk3	Duke Forest	35.98	-79.09	163	1126	1998-2005	Sandy loam	ENF	<i>Siqueira</i> [2006]
US-Ha1	Harvard Forest	42.54	-72.17	303	1122	1991-2006	Sandy loam	DBF	<i>Urbanski</i> [2007]
US-Ho1	Howland Forest	45.20	-68.74	60	818	1996-2004	Sandy loam	ENF	<i>Richardson</i> [2009]
US-IB2	Batavia Prairie	41.84	-88.24	227	818	2004-2007	Silty clay loam	GRA	<i>Post</i> [2004]
US-Me2	Metolius Intermediate	44.45	-121.56	1253	434	2002-2007	Sand	ENF	<i>Thomas</i> [2009]



**Table 1: cont.**

Site	Name	Lat.	Long.	Elev. (m a.s.l.)	Annual precip. (mm)	Years	Soil class	IGBP	Reference
US-MMS	Morgan Monroe State Forest	39.32	-86.41	275	1109	1999-2006	Clay loam	DBF	<i>Schmid</i> [2000]
US-MOz	Missouri Ozark	38.74	-92.20	219	730	2004-2007	Silt loam	DBF	<i>Gu</i> [2006]
US-Ne1	Mead Irrigated Maize	41.47	-96.48	361	832	2001-2006	Silty clay loam	CRO	<i>Verma</i> [2005]
US-Ne2	Mead Irrigated Rotation	41.16	-96.47	361	823	2001-2006	Silty clay loam	CRO	<i>Verma</i> [2005]
US-Ne3	Mead Rainfed Rotation	41.18	-96.44	361	627	2001-2006	Silty clay loam	CRO	<i>Verma</i> [2005]
US-NR1	Niwot Ridge Forest	40.03	-105.55	3050	663	1998-2007	Clay	ENF	<i>Bradford</i> [2008]
US-PFa	Park Falls/WLEF	45.95	-90.27	485	736	1995-2005	Sandy loam	MF	<i>Davis</i> [2003]
US-Syv	Sylvania Wilderness	46.24	-89.35	540	700	2001-2006	Sandy loam	MF	<i>Desai</i> [2005]
US-Ton	Tonzi Ranch	38.43	-120.97	177	549	2001-2007	Silt loam	WSA	<i>Ma</i> [2007]
US-UMB	U of M Biological Station	45.56	-84.71	234	629	1999-2006	Loamy sand	DBF	<i>Gough</i> [2013] <i>Nave</i> [2011]
US-Var	Vaira Ranch	38.41	-120.95	129	563	2000-2007	Silt loam	GRA	<i>Ma</i> [2007]
US-Wcr	Willow Creek	45.81	-90.08	520	712	1998-2006	Sandy loam	DBF	<i>Cook</i> [2004]

**Table 2: Summary of models – stomatal conductance formulation, water stress function and moisture limitation enforcement, temporal resolution, phenology scheme, and GPP methodology.**

Model	Stomata parameterization	Water stress function	Temporal resolution	Canopy phenology	GPP	Reference
Can IBIS	<i>Collatz</i> [1991]	Jarvis-type, direct and indirect	Half-hourly	Prognostic	Enzyme Kinetic	<i>Kucharik</i> [2000]
CN-CLASS	<i>Ball</i> [1987]	Feddes-type, direct	Half-hourly	Prognostic	Enzyme Kinetic	<i>Arain</i> [2006]
Ecosys	<i>Grant</i> [2006]	Turgor potential, direct	Hourly	Prognostic	Enzyme Kinetic	<i>Grant</i> [2006]
ED2	<i>Leuning</i> [1995]	Bucket model, direct	Half-hourly	Prognostic	Enzyme Kinetic	<i>Medvigy</i> [2009]
ORCHIDEE	<i>Ball</i> [1987]	Jarvis-type, indirect	Half-hourly	Prognostic	Enzyme Kinetic	<i>Verbeeck</i> [2011]
SiB3	<i>Collatz</i> [1991]	Feddes-type, indirect	Half-hourly	Prescribed	Enzyme Kinetic	<i>Baker</i> [2008]
SibCASA	<i>Collatz</i> [1991]	Jarvis-type, indirect	Half-hourly	Prescribed	Enzyme Kinetic	<i>Schaefer</i> [2008]
SSiB2	<i>Collatz</i> [1991]	Jarvis-type, indirect	Half-hourly	Prescribed	Stomatal Conductance	<i>Sellers</i> [1995]
Teco	<i>Ball</i> [1987]	Feddes-type, direct	Hourly	Prognostic	Stomatal Conductance	<i>Weng</i> [2008]

**Table 3: Correlation statistics between the modeled and observed energy budget deficit ( $E_x$ ) and errors in LE and H for each model. ED2 and Teco did not provide model outputs of H. All  $P < 0.0001$**

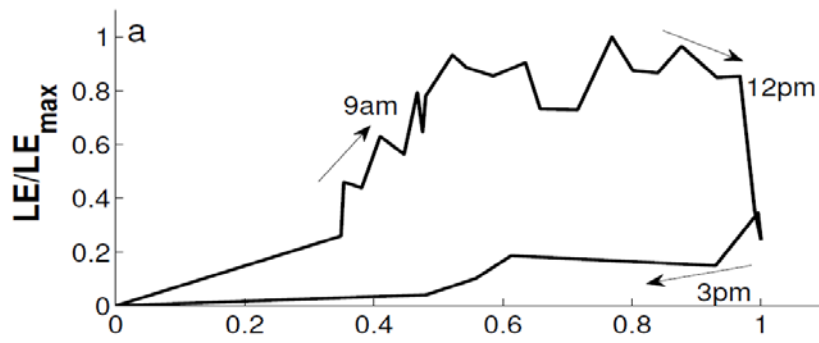
(y vs. x)	Modeled $E_x$ vs. observed $E_x$		LE error vs. observed $E_x$		H error vs. observed $E_x$		LE error vs. H error	
Model	$R^2$	slope	$R^2$	slope	$R^2$	slope	$R^2$	slope
1. CanIBIS	0.15	0.59	0.01	0.09	0.12	0.32	0.00	-0.06
2. CN-CLASS	0.28	0.73	0.07	0.19	0.01	0.09	0.01	-0.07
3. Ecosys	0.24	0.86	0.04	0.15	0.00	-0.01	0.04	0.14
4. ED2	-	-	0.04	0.35	-	-	-	-
5. ORCHIDEE	0.22	0.64	0.02	0.13	0.07	0.24	0.00	-0.04
6. SiB3	0.10	0.38	0.14	0.28	0.13	0.35	0.03	0.13
7. SibCASA	0.14	0.42	0.12	0.26	0.14	0.32	0.02	0.11
8. SSiB2	0.33	0.64	0.12	0.24	0.02	0.13	0.05	-0.16
9. Teco	-	-	0.00	-0.11	-	-	-	-



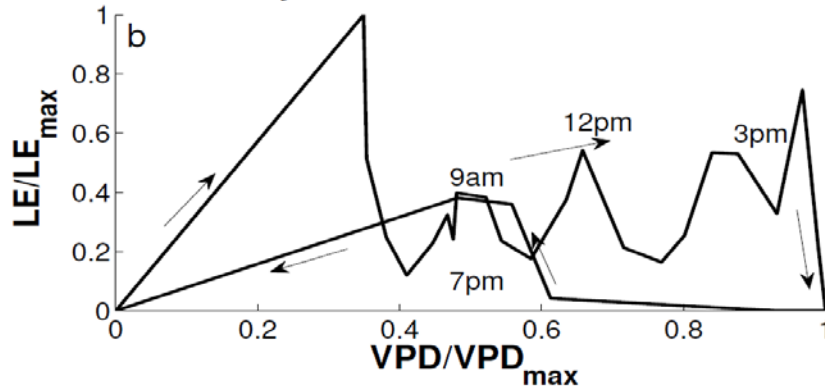
**Table 4: Degree of Hysteresis errors across multiple forcing conditions, averaged over all sites studied. Errors are calculated as modeled - observed.**

	Average observed degree of hysteresis	Error: Ball [1987]	Error: Collatz-Ball [1991]	Error: Leuning [1995]	Error: Grant-Black [2006]	Grand averaged error
All	0.217	-0.183	-0.009	-0.186	-0.106	-0.107
Wet soil	0.214	-0.176	0.002	-0.168	-0.104	-0.097
Low VPD	0.193	-0.202	-0.033	-0.207	-0.131	-0.129
Med soil	0.215	-0.188	-0.026	-0.181	-0.108	-0.116
Med VPD	0.251	-0.157	0.023	-0.155	-0.063	-0.076
Dry soil	0.247	-0.175	0.005	-0.178	-0.071	-0.095
High VPD	0.247	-0.156	0.036	-0.183	-0.030	-0.073

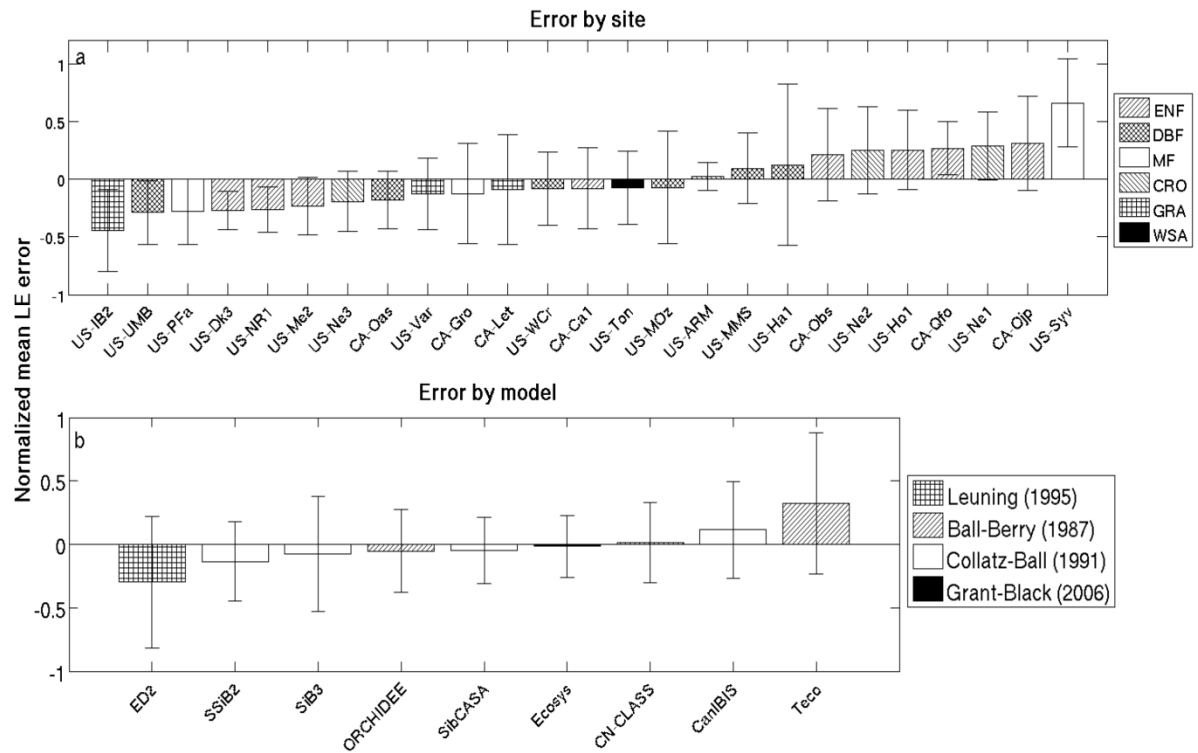
# Observed hysteresis at US-Me2, on DOY 171, 2001



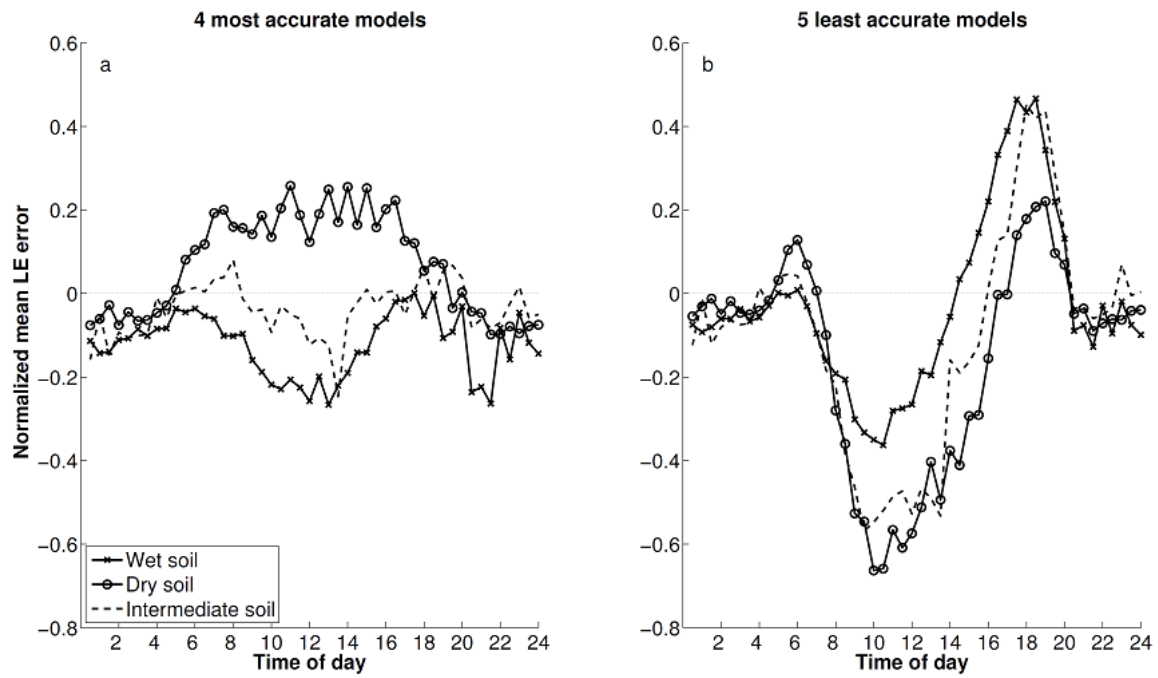
# ED2 simulated hysteresis for US-Me2, on DOY 171, 2001



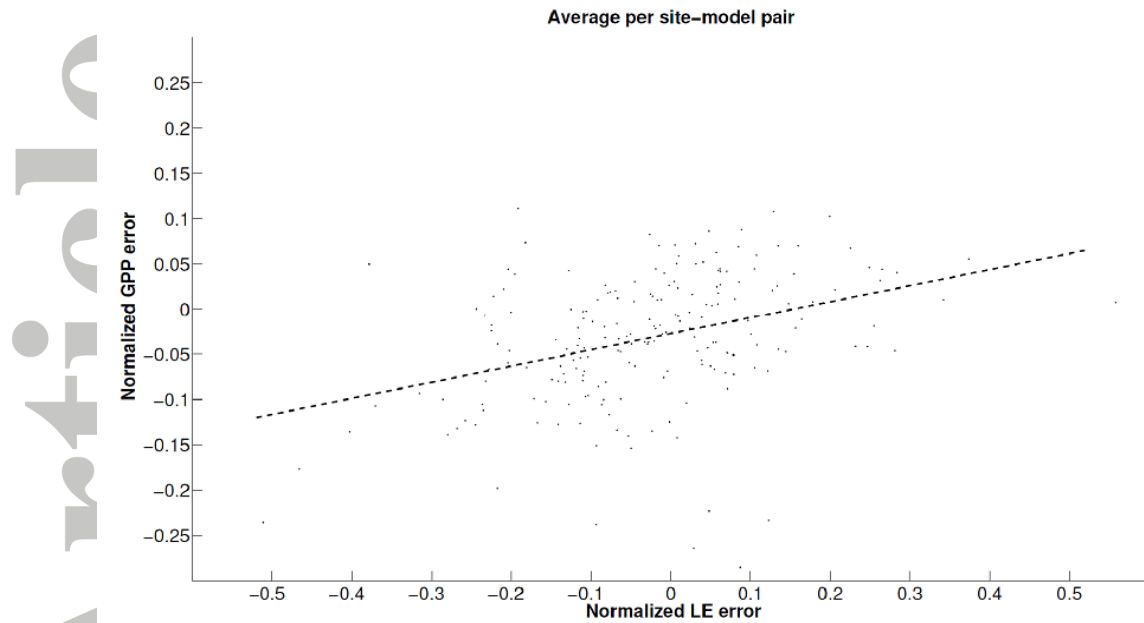
**Figure 1: Examples of observed hysteresis (a) at US-ME2 on DOY 171 of 2001 and model ‘simulated’ hysteresis (b) from ED2 for the same day. Arrows and times indicate the direction of the hysteric cycle.  $LE_{max}$  and  $VPD_{max}$  are the respective maxima for the day shown.**



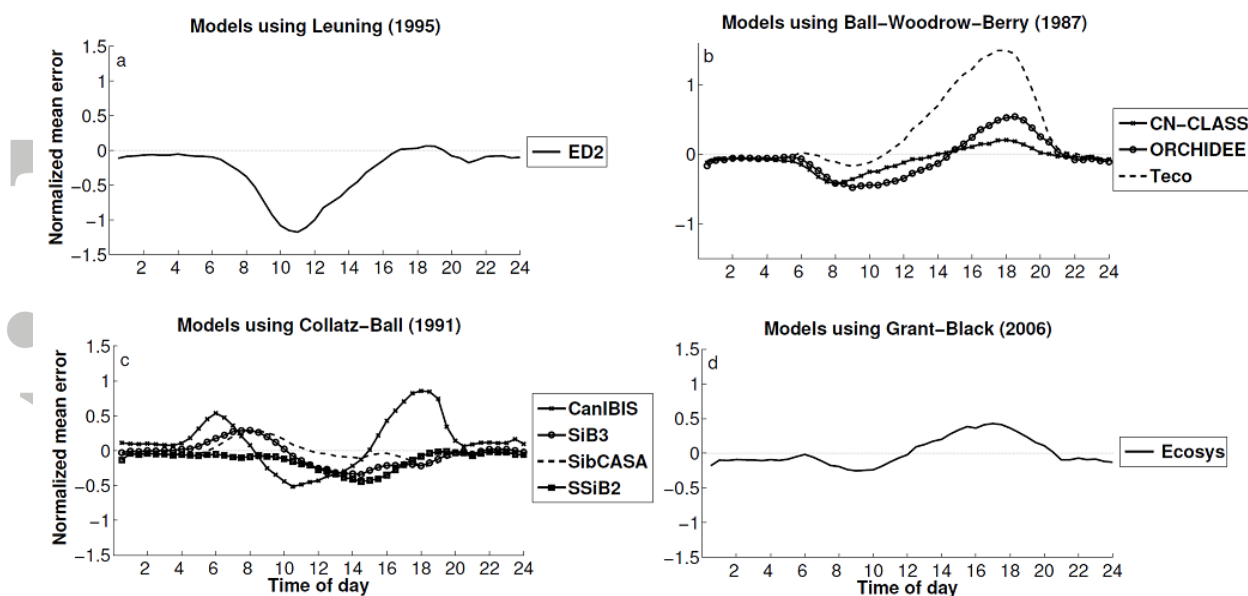
**Figure 2: Average normalized growing season LE error across all models: (a) for each site studied; and (b) across all sites for each model studied. (a) Patterns represent each site's IGBP classification: ENF – Evergreen Needleleaf Forest, DBF – Deciduous Broadleaf Forest, MF – Mixed Forest, CRO – Crop, GRA – Grassland, WSA – Woody Savanna. (b) Patterns represent the stomatal conductance parameterization employed in each model. Error bars represent the standard deviation for each site among models (a) and for each model among sites (b).**



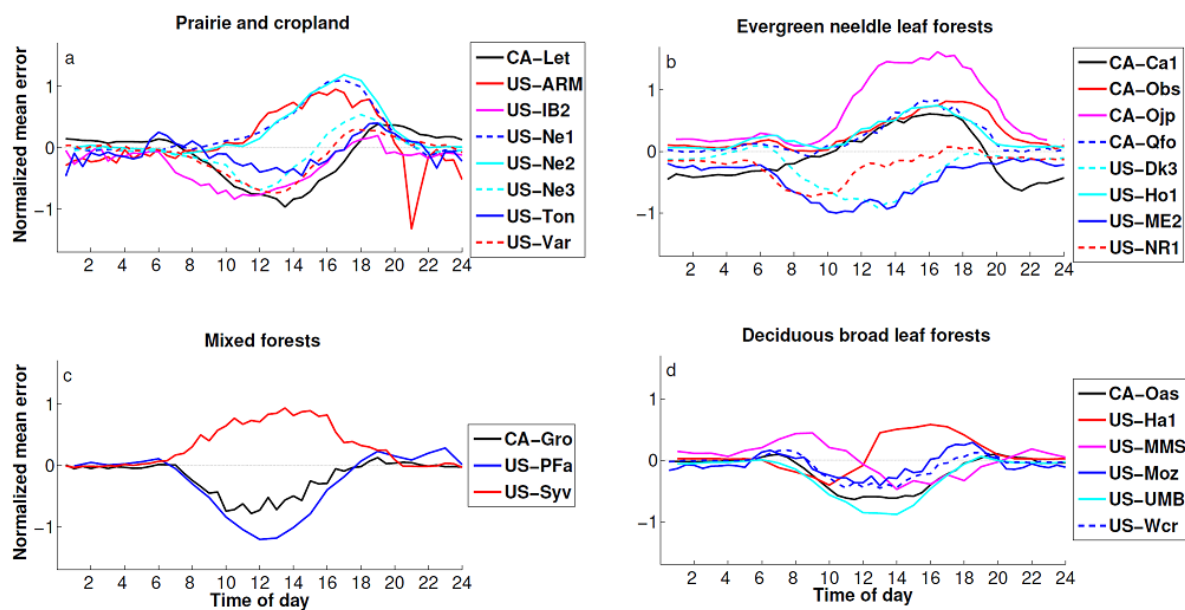
**Figure 3: Diurnal model error patterns for the 4 models with the smallest bias (a) and the 5 with the largest bias (b). Error is broken down based on soil wetness conditions: soil wetness is near the field capacity ('Wet Soil'), soil moisture is near the permanent wilting point ('Dry Soil'), soil conditions are in the intermediate ranges of conditions ('Intermediate Soil').**



**Figure 4: Average error in GPP vs. average error in LE for each site-model pair. Flux errors were normalized by the respective maximum error for each site-model pair.  $R^2 = 0.14$ , slope = 0.18,  $P < 0.0001$ .**



**Figure 5: Diurnal model error patterns for specific stomatal conductance schemes. Leuning (1995) (a), Ball-Woodrow-Berry (1987) (b), Collatz-Ball (1991) (c), and Grant-Black (2006) (d).**



**Figure 6: Diurnal average model error patterns for site type. Agricultural, prairie, and grasslands (a), evergreen needleleaf forests (b), mixed forests (c), and deciduous broadleaf forests (d). Due to the similarity of their error patterns, agricultural (CRO), grassland (GRA), and savanna (WSA) sites were grouped together.**

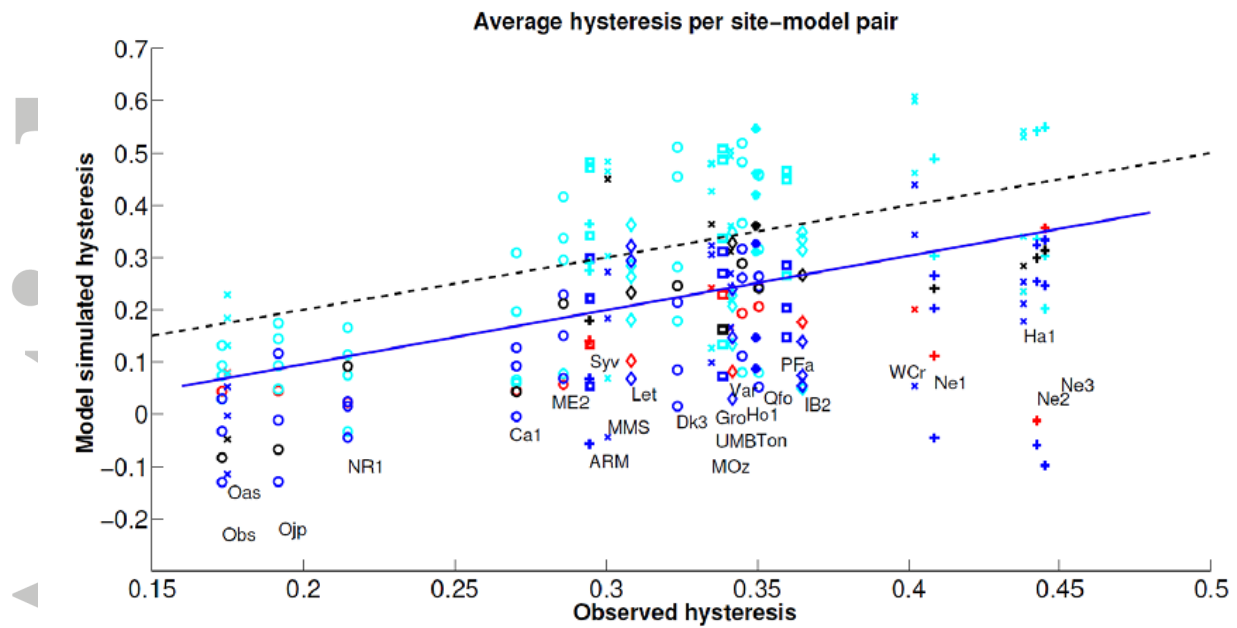


Figure 7: Model simulated and observed hysteresis for each site-model pair with respect to IGBP site classification (ENF (o), DBF (x), MF (square), CRO (+), GRA (diamond), WSA (\*)) and models' stomatal conductance parameterization (Ball (1987) – blue, Collatz-Ball (1991) – cyan, Leuning (1995) red, Grant-Black (2006) black. The black dashed line is the 1:1 line and the blue line is a linear regression,  $R^2 = 0.23$ ,  $P < 0.00001$ , slope = 1.039.