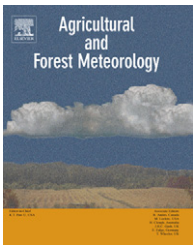
AGRICULTURAL AND FOREST METEOROLOGY 148 (2008) 268–287

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/agrformet




# Sensitivity, uncertainty and time dependence of parameters in a complex land surface model

*L. Prihodko* [a,*], *A.S. Denning* [b], *N.P. Hanan* [a], *I. Baker* [b], *K. Davis* [c]

[a] *Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, Colorado, 80523-1499, USA*
[b] *Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado, 80523-1371, USA*
[c] *Department of Meteorology, Pennsylvania State University, University Park, Pennsylvania, 16802-5013, USA*

ARTICLE INFO

*Keywords:*
Parameter sensitivity
Uncertainty
Data assimilation
Monte Carlo
GLUE
SiB
WLEF

ABSTRACT

This paper explores the use of Monte Carlo carbon cycle data assimilation within the generalized likelihood uncertainty estimation (GLUE) framework to evaluate the sensitivities of a well-known complex land surface model (SiB v2.5) to its parameterization and the predictive uncertainty of simulated fluxes on a monthly basis, and for an entire year, at the WLEF tall-tower site in Park Falls, Wisconsin. An analysis is described wherein randomly generated parameter sets were ranked based on their capacity to simulate fluxes of latent (LE) and sensible heat (H) and the net ecosystem exchange of carbon (NEE) for each month of the year 1997 and for the entire year. Two criteria were used to evaluate the success of the simulations; the first evaluated the ability of SiB2.5 to simulate LE and H, the second included NEE as an additional constraint. The best-performing parameter sets for each criterion were used to assess model sensitivity to parameters, to calculate uncertainty bounds for predicted LE, H and NEE and to assess the information content of eddy covariance data on the analyzed time scales. Patterns in the sensitivity of the model to its parameterization and the uncertainty of the predictions were related to the physiological and phenological characteristics of the ecosystem, model structure and the relationship between deterministic models and comparatively stochastic measurements. The results show that model sensitivity varies through time for a larger set of parameters than those typically considered time varying in LSMs, thus optimization of model parameters on tower flux data should allow for variability at sub-annual time scales in order to capture the most information and best simulate fluxes. Further, constraining predictions annually versus monthly showed that some quantities (e.g. night-time NEE) were on average better constrained annually, whereas other quantities that show more variability with vegetation phenology and structure (e.g. daytime NEE and LE) were better constrained monthly. The addition of the net ecosystem exchange of carbon in the data assimilation scheme improved model results by (a) constraining model parameterization (optimal parameter values), particularly during times of the year when the land surface was rapidly changing (spring and fall), and increasing the number of influential parameters, and (b) decreasing the uncertainty in NEE simulations (but not appreciably reducing uncertainty in LE and H simulations). We also found that there was an irreducible level of mismatch between the simulated and observed fluxes that could not be overcome through optimization due to variability in the observations and/or structural problems with the model. The uncertainty estimates can be used to characterize uncertainty in the simulations at multiple time scales.



* *Corresponding author*. Tel.: +970-491-1964; fax: +970-491-1965.
E-mail address: lara@nrel.colostate.edu (L. Prihodko).
0168-1923/$ – see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.agrformet.2007.08.006

## 1. Introduction

Simulation models can never fully represent every component and process in a complex ecosystem, it would be impractical and unproductive to do so. Thus, approximations and compromise solutions are always necessary at some level. Consequently, a principal characteristic of simulation models is that they utilize parameters. Parameters in models are either measurable terms that are used to define the system and determine its behavior or are terms that represent important processes that for some reason (practical or theoretical) cannot be described in detail. Like models, parameters themselves can be imperfect representations of reality. If a parameter value is poorly defined, if it cannot be accurately measured, or if it is not sufficiently representative, it will introduce uncertainty into modeled results depending on the sensitivity of the model to that parameter. Further, complex models with many parameters are often over-parameterized. That is, they contain parameters that contribute little to the simulation but have a cost associated with them in terms of computation time or definition. A consequence of over-parameterization in complex models is a tendency towards equifinality, where compensation between parameters can result in equally good simulations across a wide range of parameter values (Beven and Freer, 2001; Fedra et al., 1981; Franks, 1998; Franks et al., 1997; Spear et al., 1994). Equifinality leads to increased uncertainty in predictions as a result of uncertainty in parameterization (Schulz and Beven, 2003).

Biophysical land surface models (LSMs), which simulate fluxes of energy and mass between the atmosphere and the land surface, are used to study land surface dynamics, to generate surface conditions in atmospheric models and to elucidate the processes contributing to land surface fluxes. Typically these models are quite complex and can easily have more than 40 parameters, putting them at risk of over-parameterization and equifinality. At present, biophysical LSMs coupled to regional and global atmospheric models are playing a large role in advancing our understanding of carbon cycle dynamics and associated changes in weather and climate (Bounoua et al., 2002; Collatz et al., 2000; Cox et al., 2000; Denning et al., 1995; Pielke et al., 1998). Because of their importance in determining potential trajectories for global and regional climate change, it is important that we understand the strengths and weaknesses of these models, how they could be improved and what level of uncertainty is associated with their predictions.

A relatively new development, and promising approach to this problem, has been termed carbon cycle data assimilation (CCDA) (Braswell et al., 2005; Kaminski et al., 2002, 2003; Knorr and Kattge, 2005; Rayner et al., 2005; Wang et al., 2001), where atmospheric $CO_2$ concentration data, $CO_2$ flux measurements and other observational data are used to optimize a land surface model. The primary goals of CCDA have been to improve the parameterization and predictive power of land surface process models, to quantify the uncertainty in the surface fluxes and models, and to increase the amount and kinds of data that can be used to constrain atmospheric inversions (Kaminski et al., 2002). Initial work with a relatively simple land surface model (two parameters to optimize) has shown that both $CO_2$ concentration data and $CO_2$ flux measurements can provide powerful constraints on model parameters and the resulting predictions of fluxes (Kaminski et al., 2002). Both non-linear inversion and Monte Carlo methods of CCDA (Braswell et al., 2005; Kaminski et al., 2002; Knorr and Kattge, 2005; Mo and Beven, 2004; Rayner et al., 2005; Schulz et al., 2001; Wang et al., 2001) show more promise than standard methods of sensitivity and uncertainty analysis such as varying one parameter at a time (El Maayar et al., 2002; Potter et al., 2001) or factorial methods (Henderson-Sellers, 1993) because these methods can better account for non-linearity in the model and parameter interactions, while also providing more robust estimates of uncertainty in model predictions.

This paper explores the use of Monte Carlo CCDA within the generalized likelihood uncertainty estimation (GLUE; Beven and Binley, 1992) framework to evaluate the sensitivity and uncertainty of a well-known land surface model, the Simple Biosphere Model (SiB2.5; Baker et al., 2003; Sellers et al., 1996b) that has been variously used to simulate water, heat and carbon fluxes for single ecosystems as well as run coupled to regional and global circulation models (Baker et al., 2003; Colello et al., 1998; Denning et al., 1996a,b, 2003; Nicholls et al., 2004; Randall et al., 1996). SiB2.5 is a complex, highly non-linear model with a large number of parameters and interactions between them are expected to be important. Operational values for these parameters are often poorly known and so a wide range of possible (but realistic) values for them should be tested. The strengths and flexibility of Monte Carlo style data assimilation are well suited to the problem of analyzing the sensitivity of SiB2.5 to its parameterization because these methodologies allow for the assessment of non-linear interactions between parameters (Bastidas et al., 1999; Beven and Freer, 2001; Fedra et al., 1981; Franks, 1998; Franks et al., 1997; Hornberger and Spear, 1981; O'Neill et al., 1982; Spear and Hornberger, 1980). Further, even though Monte Carlo methodologies are computationally expensive, they are relatively simple to set up and run in a distributed computing system.

We chose the GLUE methodology over others because rather than model optimization per se, we wanted to explore the model error space to determine which parameters are most influential and therefore should be most carefully evaluated for effective simulations. We tried to cast a wider net than Markov Chain Monte Carlo (Braswell et al., 2005; Knorr and Kattge, 2005) or non-linear inversion methods (Wang et al., 2001), which tend to focus on optimization, by looking at 10s of thousands of possible parameter sets and exploring the impact of these on simulated fluxes. Further, the GLUE methodology is well suited to cases where there are many parameters, little a priori knowledge of parameter distributions and a risk of equifinality. Similar to other methods, the robustness of sensitivities and uncertainties are dependant on how the problem is set up, but because these choices are clearly defined the results can be interpreted accordingly. The GLUE methodology is well established in hydrological model applications (Beven and Binley, 1992; Beven and Freer, 2001; Brazier et al., 2000, 2001; Franks et al., 1998; Freer and Beven, 1996; Romanowicz et al., 1994) and is also well suited to land surface-atmosphere models, since it

implicitly handles non-linear interactions. The number of studies using this framework for land surface-atmosphere modeling is steadily increasing (Franks and Beven, 1999; Franks et al., 1999, 1997; Mo and Beven, 2004; Schulz and Beven, 2003; Schulz et al., 2001).

In this paper we evaluate model sensitivities and uncertainties associated with parameters evaluated using remotely sensed data, literature values and measurements at monthly and annual time scales for the WLEF tall-tower site in Park Falls, Wisconsin. In particular, the objectives of this study are to identify parameters that, if more accurately known, would improve the results of the model; to evaluate the appropriateness of conditioning model parameters on different time scales and to assess whether the assimilation of NEE measurements helps to constrain model parameterization and uncertainty and thus improve the simulation of energy and water fluxes as well.

Analyzing the sensitivity and uncertainty of model simulations to parameterization can lead to insights about model structure and system behavior that would not ordinarily be evident simply from simulation results (Franks et al., 1997; O'Neill et al., 1982). Unexpected significance of a parameter or lack of significance can lead to questioning of our mechanistic understanding and the relative role of different processes. Sensitivity and uncertainty analyses are also useful for guiding research priorities. Knowing which parameters are most influential in a model, or under what circumstances simulated fluxes are most uncertain, can help focus research efforts in those areas where more attention would bring proportionately more benefit in terms of predictive accuracy (Hornberger and Spear, 1981; Spear and Hornberger, 1980).

## 2. Methods

### 2.1. Site description

The WLEF-TV tower (90.28°W, 45.95°N) is located within the Chequamegon Nicolet National Forest in the northern highlands region of Wisconsin, approximately 100 km south of Lake Superior. Micrometeorological, eddy covariance and $CO_2$ concentration measurements have been made since 1995 on this tall tower (Bakwin et al., 1998; Berger et al., 2001; Davis et al., 2003). The vegetation of the region is composed of mixed northern hardwoods, upland Jack and Red pine, lowland conifers, aspen and wetlands (Burrows et al., 2002; Wisconsin Department of Natural Resources (WiDNR), 1998). Much of the area was logged from 1860 to 1920 and has since re-grown (USDAFS, 2001). The topography is mild, with a local relief of up to 45 m, and elevations reaching to approximately 450 m above sea level (Burrows et al., 2002; Davis et al., 2003; Martin, 1965). The climate of the area is cool continental with a 30-year (1971–2000) annual average mean temperature of 4.9 °C (−12.3 °C average in January; 19.6 °C average in July) and annual precipitation averaging 815 mm year$^{-1}$ (MRCC, 2002).

### 2.2. Meteorological and flux data

Wind speed and direction, air temperature and relative humidity are measured on the tall tower at 30, 122, and 396 m while atmospheric pressure, precipitation, photosynthetically active radiation (PAR) and net radiation are measured at the surface in the clearing surrounding the tower (Bakwin et al., 1998; Berger et al., 2001; Davis et al., 2003). These measurements were averaged at hourly intervals. Incoming long-wave radiation was not measured, so the method of Idso (1981) was used to estimate long-wave radiation from air temperature and humidity. A hierarchical sequence of methods was used to fill gaps in the meteorological data due to intermittent instrument or data logger failures. Baker et al. (2003) give a thorough description of this process. The filled meteorological variables used in this paper are: air temperature, dew point temperature and wind speed at 122 m, atmospheric pressure, incoming long-wave radiation, incoming short wave radiation and precipitation at the surface.

High frequency flux observations of latent (LE) and sensible heat (H) and the net ecosystem exchange of $CO_2$ (NEE) are also made on the tall tower at three heights (30, 122 and 396 m) (Bakwin et al., 1998; Berger et al., 2001; Davis et al., 2003). At each of these heights three-dimensional wind speeds, sonic temperature and the mixing ratios of $CO_2$ and $H_2O$ are sampled at 5 Hz and fluxes computed at hourly intervals. In addition, the mixing ratio of $CO_2$ and $H_2O$ are measured at three other heights (11, 76 and 244 m). Lag and frequency response corrections were applied to the flux estimates, and the mixing ratio profiles were used to estimate storage terms below each measurement height. 'Preferred' estimates of terrestrial NEE, LE and H were then selected from among the three measurement heights based on atmospheric conditions (Davis et al., 2003). The 'preferred' values used in the following analyses are considered to best represent the heterogeneous landscape in the area by maintaining a large, consistent flux footprint while lessening the influence of the cleared patch (radius 200 m) underneath the tower (Davis et al., 2003).

For this paper, flux and meteorological data from the year 1997 were used. Year 1997 was chosen because conditions were close to the 30-year average for temperature and precipitation and because it was the first full year of flux data collected at WLEF. The data were also thoroughly analyzed in a paper by Davis et al. (2003). Overall, 90% of NEE observations and 85% of the possible observations of H and LE were available.

### 2.3. Model description

The Simple Biosphere Model version 2 (SiB2; Sellers et al., 1996b) is a single canopy-layer scheme describing the transfers of heat, water, momentum and carbon in the soil-vegetation atmosphere continuum. SiB2's physical model of radiation balance, heat and water transport conserves mass and energy. SiB2 incorporates leaf-level physiology controlling photosynthesis (Collatz et al., 1992, 1991; Farquhar et al., 1980) and the Ball–Berry (Ball et al., 1987) description of stomatal behavior which links water loss and carbon assimilation. Canopy scale calculations of carbon and water fluxes are estimated from leaf-level physiology and exchange using a canopy integration factor ($\Pi$) that is related to the extinction of photosynthetically active radiation (PAR) through the canopy (Sellers et al., 1992). $\Pi$ reflects the assumption of continuously varying stomatal conductance, carbon assimilation, and

rubisco velocity vertically in the canopy in response to the time-mean distribution of light, with nitrogen allocation responding accordingly (Field and Mooney, 1986). The integration factor provides a scaling relationship between canopy scale fluxes and larger spatial scales. More importantly, $\Pi$ can be estimated over large areas because it is proportional to the fraction of absorbed photosynthetically active radiation ($f_{PAR}$) which in turn is near linearly related to remote sensing estimates of the normalized difference vegetation index (NDVI) (Sellers, 1985, 1987).

In SiB2 the annual integral of ecosystem respiration is assumed to be equal to the annual integral of gross primary production. Respiratory losses are parameterized diurnally and seasonally by partitioning respiration between the six soil layers and the surface and regulating the respiration rate through temperature, moisture and soil texture (Denning et al., 1996b; Raich and Nadelhoffer, 1989). When run at an annual time scale with this parameterization there is a net zero balance of carbon into and out of the system. At WLEF in 1997 a net zero balance of carbon into and out of the system is a reasonable assumption, measured NEE was $16 \pm 19$ g C m$^{-2}$ year$^{-1}$ (Davis et al., 2003).

The most recent version of the model, SiB2.5 (Baker et al., 2003), includes prognostic equations describing transfers of heat, water and carbon that also account for canopy air-space storage. Also incorporated into SiB2.5 is a six-layer soil temperature model, an update from the three-layer model in previous versions. Flux output from SiB2.5 was averaged from the 10-min model time step to hourly output, to correspond with the observational time step. Baker et al. (2003) have previously demonstrated the ability of SiB2.5 to simulate fluxes of heat, water and $CO_2$ at the WLEF tall-tower site.

### 2.4. Generalized likelihood uncertainty estimations (GLUE)

The GLUE methodology can be thought of as a model analysis framework that includes the following basic components (Beven and Binley, 1992): (1) a formal definition of a likelihood measure, (2) an appropriate initial distribution of parameter values, (3) a procedure for using likelihood weights in uncertainty estimation, (4) a procedure for updating likelihood weights and (5) a procedure for evaluating uncertainty and the value of additional data. The GLUE methodology utilizes probabilistic likelihood measures to weight model predictions and calculate uncertainty bounds (Beven and Binley, 1992). Bayes theorem is used in GLUE applications to calculate weights from likelihoods and to update likelihood weights with new information as it becomes available.

### 2.5. Parameters

The parameters for SiB2.5 that were evaluated in this paper are those that are determined external to the model and which characterize land surface conditions using a combination of land cover type, monthly maximum NDVI and soil properties. They can be further subdivided into time invariant parameters (40; Table 1) and time-varying parameters (8; Table 2). Time invariant parameters for the purpose of this paper are those that are set at the beginning of a simulation and fixed for the length of the simulation, but which vary in space based on biome, soil type and geographical location. Time-varying parameters are those that change at intervals shorter than the time period of the simulation. Typically these parameters change once a month to represent changes in vegetation phenology throughout the year. The majority of parameters in Table 1 are defined from the literature and are assigned by biome, of which there are 9 distinct types in SiB2.5 (Sellers et al., 1996a). For WLEF, the vegetation was classified as mixed deciduous and coniferous forest. The soil hydraulic and thermal properties are calculated from the percent sand and percent clay of the soil using the equations of Cosby et al. (1984) while the parameters used in calculations of soil heterotrophic respiration are estimated from the percent clay of the soil using curves fit to data in Raich et al. (1991) (Schaeffer, Colorado State University, personal communication). The soil at WLEF was defined as consisting of 37% sand and 15% clay, which was determined from the STATSGO soil database (Staff, 1994). The time-varying parameters listed in Table 2 are estimated either directly or indirectly from NDVI data using the methods of Los et al. (2000) and Sellers et al. (1996a,b). Here they were calculated from an annual series of monthly mean NDVI from 8 km Advanced Very High Resolution Radiometer (AVHRR) data for 1997 (Tucker et al., 2005). GMUDMU, the time-mean leaf projection, is not estimated from NDVI but varies at a monthly time step depending on canopy type, solar declination and latitude. Standard parameters for 1997 for the WLEF site are given in Tables 1 and 2.

### 2.6. Parameter set generation

For this experiment 46 parameters were randomly varied within physically reasonable ranges, assuming a uniform distribution for all parameters (Tables 1 and 2). Such a large number of parameters were varied in order to allow for unexpected interactions. Uniform distributions of parameters were chosen because there was no reliable information from which to calculate other statistical distributions. While choosing uniform distributions might make the results more sensitive to the extreme parts of the parameter range, there is evidence that if parameter ranges are sufficiently large yet also reasonable, any negative effects of a uniform distribution are greatly diminished (O'Neill et al., 1982).

Time invariant parameters were for the most part independently randomized. However, in a few situations it was not physically realistic to vary parameters independently. For example, canopy base height ($Z_1$) cannot be higher than canopy top height ($Z_2$). In six cases parameters were made partially dependent on other parameters. For five of the parameters ($Z_1$, ROOTD, TRDM, HHTI, SOREF2) a randomization interval was selected (see Table 1) and the random value ($R$) within this range was multiplied by the independent parameter ($I$) to give the new parameter value ($P$):

$$P = IR \tag{1}$$

In this way, parameter $Z_1$ (canopy base height) is partially dependent on $Z_2$ (canopy top height), ROOTD (rooting depth) is partially dependent on SODEP (soil depth), TRDM (respiration inhibition 1/2-point temperature) is partially dependent on TROP (respiration optimum temperature),

**Table 1 – List of parameters varied in this experiment**

| Parameter | Definition | Minimum | Maximum | WLEF |
|---|---|---|---|---|
| $Z_2$ | Canopy top height (m) | 10.0 | 30.0 | 20.0 |
| $Z_1$ | Canopy base height (m) | 0.2[a] | 0.8[a] | 10.0 |
| ZC | Canopy inflection height (m) | 0.3[a] | 0.8[a] | 15.0 |
| VCOVER | Vegetation cover fraction | 0.5 | 1.0 | 0.9875 |
| CHIL | Leaf angle distribution factor | −0.5 | 0.5 | 0.125 |
| ROOTD | Rooting depth (m) | 0.2[a] | 0.9[a] | 1.5 |
| PH | 1/2 critical leaf water potential (m) | −450.0 | −50.0 | −200.0 |
| TRAN11 | Green leaf transmittance (PAR) | 0.0 | 0.1 | 0.05 |
| TRAN21 | Green leaf transmittance (NIR) | 0.05 | 0.3 | 0.15 |
| TRAN12 | Brown leaf transmittance (PAR) | 0.0 | 0.1 | 0.001 |
| TRAN22 | Brown leaf transmittance (NIR) | 0.0 | 0.1 | 0.001 |
| REF11 | Green leaf reflectance (PAR) | 0.02 | 0.2 | 0.07 |
| REF21 | Green leaf reflectance (NIR) | 0.2 | 0.5 | 0.38 |
| REF12 | Brown leaf reflectance (PAR) | 0.05 | 0.25 | 0.16 |
| REF22 | Brown leaf reflectance (NIR) | 0.2 | 0.5 | 0.42 |
| VMAX0 | Rubisco velocity of sun leaf (mol $m^{-2}$ $s^{-1}$) | 2.5E−5 | 15E−5 | 7.5E−5 |
| EFFCON | Quantum efficiency (mol $mol^{-1}$) | 0.03 | 0.13 | 0.08 |
| GRADM | Conductance–photosynthesis slope parameter | 3.0 | 18.0 | 9.0 |
| BINTER | Minimum stomatal conductance (mol $m^{-2}$ $s^{-1}$) | 0.0 | 0.02 | 0.01 |
| ATHETA | Light and rubisco coupling parameter | 0.5 | 1.0 | 0.98 |
| BTHETA | Light, rubisco and CHO sink parameter | 0.5 | 1.0 | 0.95 |
| TRDA | Respiration temperature response ($K^{-1}$) | 0.1 | 1.5 | 1.3 |
| TRDM | Respiration inhibition 1/2-point temperature (K) | 1.04[a] | 1.1[a] | 328.16 |
| TROP | Respiration optimum temperature (K) | 283.0 | 308.0 | 298.16 |
| RESPCP | Leaf respiration fraction of VMAX | 0.01 | 0.1 | 0.015 |
| SLTI | Photosynthesis low temperature response ($K^{-1}$) | 0.1 | 1.5 | 0.2 |
| SHTI | Photosynthesis high temperature response ($K^{-1}$) | 0.1 | 1.5 | 0.3 |
| HLTI | Photosynthesis low temperature inhibition 1/2-point (K) | 270.0 | 290.0 | 280.66 |
| HHTI | Photosynthesis high temperature inhibition 1/2-point (K) | 1.04[a] | 1.1[a] | 307.16 |
| BEE | Soil wetness exponent | 4.0 | 8.5 | 5.29 |
| PHSAT | Soil water potential at saturation (m) | −0.05 | −0.35 | −0.25 |
| SATCO | Saturated hydraulic conductivity | 2.5E−6 | 100E−6 | 3.4E−6 |
| POROS | Soil porosity | 0.3 | 0.5 | 0.44 |
| SLOPE | Cosine of mean terrain slope | 0.1 | 0.25 | 0.1736 |
| WOPT | Optimal percent of soil saturation for respiration | 30.0 | 80.0 | 62.12 |
| ZM | Skewness exponent of respiration vs. soil water | −0.2 | 0.5 | 0.362 |
| WSAT | Respiration rate at soil water saturation coefficient | 0.5 | 0.8 | 0.537 |
| SODEP | Soil depth (m) | 0.5 | 4.0 | 2.0 |
| SOREF1 | Soil reflectance (PAR) | 0.01 | 0.4 | 0.11 |
| SOREF2 | Soil Reflectance (NIR) | 1.1[a] | 1.5[a] | 0.22 |
| LWIDTH | Leaf width (m) | 0.003 | 0.1 | 0.04 |
| LLENGTH | Leaf length (m) | 0.03 | 0.4 | 0.1 |
| LTMAX | Maximum leaf area index | 4.0 | 9.0 | 7.5 |
| STEM | Stem area index | 0.0 | 0.25 | 0.08 |
| ND98 | 98th percentile NDVI (over 12 months, by biome) | 0.5 | 1.0 | 0.686 |
| ND02 | 2nd percentile NDVI (over 12 months, by biome) | 0.0 | 0.1 | 0.034 |

The minimum and maximum values of the uniform distribution used to assign random parameters values are given. The WLEF site value typically used for point simulations is also given.

[a] Min/max values are multipliers for parameters that depend on others (see text Eqs. (1) and (2)).

HHTI (photosynthesis high temperature inhibition 1/2-point) is partially dependent on HLTI (photosynthesis low temperature inhibition 1/2-point) and SOREF2 (soil near infrared reflectance) is partially dependent on SOREF1 (soil visible reflectance). In the sixth case, canopy inflection height (ZC) was made partially dependent on both canopy top height ($Z_2$) and canopy base height ($Z_1$) as follows:

$$ZC = (Z_2 - Z_1)R + Z_1 \quad (2)$$

Six parameters that are used to calculate the time-varying parameters were also varied, they are: LWIDTH (leaf width), LLENGTH (leaf length), LTMAX (maximum leaf area index), STEM (stem area index), ND98 (98th percentile NDVI by biome) and ND02 (2nd percentile NDVI by biome). These six parameters are used with the seasonally varying NDVI to calculate the following time-varying parameters: fractional absorbed photosynthetically active radiation ($f_{PAR}$), leaf area index (ZLT), greenness fraction (GREEN), roughness length (Z0D), displacement height (ZP_DISP), a coefficient for boundary-layer resistance (CC1) and a coefficient for canopy-layer resistance (CC2). These parameters were varied, instead of independently adjusting the time-varying parameters, to preserve the gross seasonal changes in the NDVI data while

**Table 2 – Standard time-varying parameters for the WLEF site, 1997**

| Parameter | Definition | January | February | March | April | May | June | July | August | September | October | November | December |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| APAR | Fraction of absorbed photosynthetically active radiation | 0.42 | 0.3 | 0.2 | 0.47 | 0.69 | 0.94 | 0.95 | 0.95 | 0.95 | 0.89 | 0.63 | 0.51 |
| ZLT | Leaf area index | 0.58 | 0.39 | 0.29 | 0.78 | 1.53 | 4.04 | 7.42 | 6.7 | 4.97 | 2.77 | 1.12 | 0.71 |
| GREEN | Greenness fraction | 0.65 | 0.65 | 0.54 | 0.9 | 0.95 | 0.98 | 0.99 | 0.88 | 0.78 | 0.54 | 0.53 | 0.89 |
| Z0D | Surface roughness length | 0.5 | 0.39 | 0.32 | 0.59 | 0.81 | 1.02 | 1.02 | 1.03 | 1.03 | 0.96 | 0.71 | 0.56 |
| ZP_DISP | Zero plane displacement height | 15.3 | 15.1 | 14.9 | 15.4 | 15.7 | 16.2 | 16.6 | 16.6 | 16.4 | 16.0 | 15.6 | 15.4 |
| CC1 | Boundary-layer resistance coefficient | 41.6 | 58.3 | 78.6 | 32.3 | 18.7 | 9.2 | 6.4 | 6.8 | 8.1 | 11.9 | 24.0 | 35.1 |
| CC2 | Canopy-layer resistance coefficient | 281.0 | 269.6 | 261.6 | 292.2 | 331.2 | 462.8 | 665.9 | 619.5 | 515.8 | 395.0 | 310.3 | 288.2 |
| GMUDMU | Time-mean leaf projection | 1.10 | 1.03 | 0.94 | 0.88 | 0.84 | 0.83 | 0.84 | 0.87 | 0.92 | 1.01 | 1.09 | 1.13 |

still altering the range of calculated values. In other words, the land surface will still have more active vegetation in the summer months than in the winter, but the timing of leaf-on and leaf-off as well as the magnitude of the vegetation changes will be altered.

The model was run with the same forcing data for six annual cycles to equilibrium to define soil moisture and soil temperature profiles for each set of parameters. The assumption is that the resultant soil temperature and moisture profiles should approach some equilibrium between the weather forcing and the model dynamics for the study area. Without a spin up, modeling results are sensitive to the transient. That is, if initial conditions are poorly specified for a particular model configuration, the effect will be anomalous results for an undetermined period of time as the model adjusts to the initial conditions. In this study the focus is on the parameterization of the model as opposed to dependence on initial conditions and so spin up of the model was considered beneficial. Further, by spinning up for each parameter set using the same climate data, common properties of the simulation are preserved and sensitivities to the parameters are more straightforward to interpret.

Simulations were carried out using 20,000 sets of randomly generated parameters, which give a small, but important, range of variation in combinations of parameters. The number of simulations directly impacts the robustness of both parameter sensitivities and uncertainty estimates, more simulations would increase the power of the results. This is particularly important in cases where optimization is the goal. However, in this paper we are exploring the more general case of model-data mismatch and so we tried to balance computational cost and representation of variability, since it would be infeasible to simulate all possible parameter combinations.

### 2.7. *Goodness of fit criteria*

Goodness of fit (or likelihood) measures used to rank the success of a simulation can be based on a single objective criterion or on multi-objective criteria that combine several comparisons between observations and simulations (Bastidas et al., 1999; Franks, 1998; Franks et al., 1999; Gupta et al., 1998, 1999; Meixner et al., 1999; Schulz et al., 1999, 2001). Both are useful for analyzing model sensitivity, however multi-objective criteria add additional constraints on model performance and parameter interactions. On the other hand, in multi-objective approaches there is generally no unique, best solution for all criteria (Beven and Binley, 1992; Gupta et al., 1999; Yapo et al., 1998).

Both single and multi-objective goodness of fit statistics were calculated for each of the 20,000 runs, comparing predicted hourly fluxes with observed fluxes within each month and for the full year. For the single objective criteria the root mean square difference (RMSD) between the observed and estimated values of LE, H and NEE was calculated, giving a goodness of fit statistic in units of flux. For the multi-objective criteria a likelihood function developed by Franks (1998) was used (Eq. (3)). This likelihood function utilizes the variance of errors normalized by the minimum variance of errors and thus permits the multi-objective criteria to include variables with

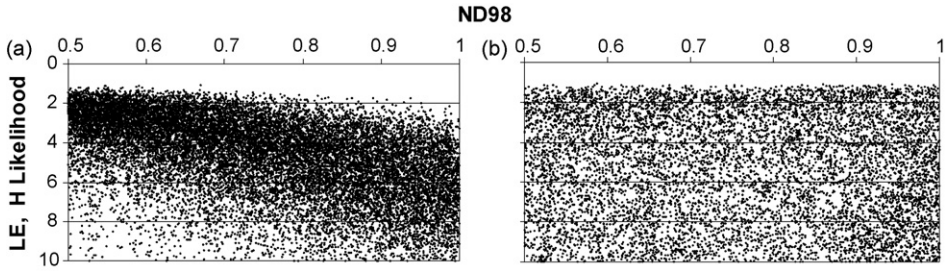



**Fig. 1 – Plots showing the LE, H likelihood statistic against the ND98 parameter for (a) April and (b) July 1997. Smaller likelihood values represent better simulations. The range of likelihood values (y-axis) is the same on both plots for comparison purposes. Simulations with computed likelihoods >10 are therefore not shown. For April this range of likelihood values includes approximately 94% of the 20,000 simulations, but only 43% of the simulations in July. The distribution of the full dataset for July is similarly uniform.**

disparate units (in this case Watts and micromoles) and different dynamic ranges.

$$L(\theta|\underline{Y}, \sigma_\alpha^2, \sigma_\beta^2, \ldots) = \left( \left( \frac{\sigma_\alpha^2}{\hat{\sigma}_\alpha^2} \right) \left( \frac{\sigma_\beta^2}{\hat{\sigma}_\beta^2} \right) \cdots \right)^{-N} \quad (3)$$

Where $\theta$ is the parameter set, $\underline{Y}$ the forcing data, $\sigma_\alpha^2$ the variance of errors for a particular variable (in this paper we evaluate NEE, LE and H), and $\hat{\sigma}_\alpha^2$ is the minimum of the variance of errors for the time period (in this paper either a single month or the annual cycle). Combined likelihood statistics were calculated for LE and H and for LE, H and NEE, giving two unitless goodness of fit statistics. The minimum possible likelihood value for the multi-objective criteria is one and likelihood values close to one represent model runs with the least amount of error in predicted versus observed fluxes.

The scaling factor $N$ in Eq. (3) is used to weight higher performing parameter sets more heavily when the likelihood value is used as a cut-off to distinguish good simulations from bad (Franks, 1998). However, in this paper we chose to use a percentage of runs for further analysis because it is difficult to determine what an acceptable fit for a model is given that errors in the observations are not easily quantified. A perfect fit is not a reasonable assumption. Thus, we have set the scaling factor $N$ to negative one such that the likelihoods were not rescaled.

Plots of the likelihood statistic versus parameter value show patterns in sensitivity or equifinality for each parameter/month combination (Beven and Freer, 2001; Franks et al., 1997; Mo and Beven, 2004; Schulz and Beven, 2003; Schulz et al., 2001). Fig. 1 shows an example of the parameter ND98 for April and July. ND98 is the biome type specific 98th percentile value of NDVI used to calculate $f_{PAR}$. In April there is a higher density of points with lower likelihood values (better simulations), particularly toward the lower range of ND98. This suggests that in April, lower values of the parameter ND98 are important for successful model runs. In contrast, the scatter of points is more uniform in July, indicating less sensitivity and a tendency towards equifinality (i.e. equally good simulations can be found across the full range of ND98 values).

### 2.8. *Parameter sensitivity*

The model runs were sorted based on the likelihood estimate for each month and for the annual cycle and the top 10% of the runs in each case were retained for further evaluation. Within the top 10% of runs, the cumulative frequency distribution for each parameter, multi-objective criteria, month and the annual cycle were calculated. Cumulative frequency gives an indication of parameter sensitivity when it deviates from an expected distribution (Bastidas et al., 1999; Franks et al., 1997; Hornberger and Spear, 1981; Meixner et al., 1999; Schulz et al., 1999; Spear and Hornberger, 1980). Fig. 2 shows an example for ND98 for April and July. In July the retained runs collapse toward the 1:1 line (equivalent to a uniform distribution), indicating a lack of sensitivity to the parameter across the full range of ND98. In April there is clear separation between the retained runs and the uniform distribution. The area of steepest slope indicates where the majority of

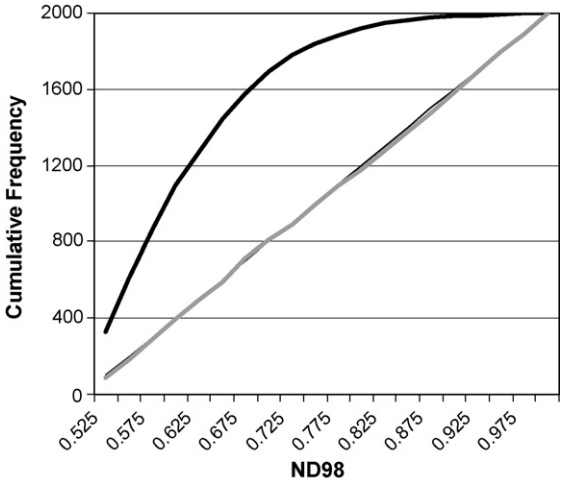



**Fig. 2 – Cumulative frequency distribution of parameter ND98 for the LE, H multi-objective criterion for the months of April (thick black) and July (thick grey), 1997. The uniform distribution (thin black) shows the expected relationship for 'non-sensitive' parameters. July and the uniform distribution are almost indistinguishable.**

parameter values fall and provides a first estimate of optimal parameter values in cases where the model is sensitive to that parameter.

A Kolmogorov–Smirnov (K–S) test (Sokal and Rohlf, 1981) was performed to assess the difference between the observed and expected (uniform) distribution and determine the significance. The model was considered sensitive to a parameter if the K–S test statistic was significant at the $p = 0.01$ level. However, the higher the K–S value, the more sensitive the model is to the specification of that parameter, thus the K–S test statistic can be used to rank the parameters (Hornberger and Spear, 1981). The K–S test statistic has been used successfully in previous sensitivity tests (Bastidas et al., 1999; Hornberger and Spear, 1981; Meixner et al., 1999; Spear and Hornberger, 1980). The total number of sensitive parameters was calculated in this way for both multi-objective criteria, for all time period/parameter combinations, and the parameters were ranked by sensitivity.

### 2.9. *Uncertainty estimation*

The simulations of H, LE and NEE associated with the retained parameter sets for each month and the full year were used to estimate uncertainty bounds attributable to parameter uncertainty in each time period. In this case the likelihood values for each retained parameter set were recalculated using the shaping factor N in Eq. (3) set to 1. Including N does not change the rank order of the results but does weight better simulations proportionately more (Franks et al., 1999). Eq. (4), from Beven and Freer (2001), illustrates how the posterior likelihoods for each of the retained simulations are calculated.

$$L[M(\theta_i)] = \frac{L_o[M(\theta_i)]L_T[M(\theta_i|Y_T,Z_T)]}{C} \quad (4)$$

Where $L_o[M(\theta_i)]$ is the prior likelihood for the vector of predicted fluxes using the SiB model with the ith set of parameters; $L_T[M(\theta_i|Y_T, Z_T]$ the likelihood measure calculated for the model with parameter set $\theta_i$ over time period T conditioned on the forcing data $Y_T$ and observed fluxes $Z_T$, $L[M(\theta_i)]$ the posterior likelihood and C is a scaling constant such that the cumulative sum of the posterior likelihoods for the retained runs equals one. Because we used non-informative priors (uniform distributions for each parameter), $L_o[M(\theta_i)]$ is constant.

Eq. (5) (Beven and Freer, 2001) illustrates how the prediction quantiles ($P(\hat{Z}_t < z)$; uncertainty bounds) are then calculated from the retained posterior likelihoods.

$$P(\hat{Z}_t < z) = \sum_{i=1}^{n} L[M(\theta_i)|\hat{Z}_{t,i} < z] \quad (5)$$

Where $\hat{Z}_{t,i}$ is the value of variable Z at time t simulated by model $M(\theta_i)$ with parameter set $\theta_i$. The computation of 90% uncertainty bounds for *each hourly time step* proceeds in this way: (1) predictions of LE, H and NEE from the retained runs are ranked from low to high flux, $\hat{Z}_{t,i}$, (2) likelihood values associated with each parameter set are maintained along with the fluxes as they are ranked, $L[M(\theta_i)|\hat{Z}_{t,i} < z]$, (3) the likelihood values are then summed until 0.05 and 0.95 are reached $\sum_{i=1}^{n} L[M(\theta_i)|\hat{Z}_{t,i} < z]$ and (4) the flux at these points (z; 0.05 and 0.95) are the 5th and 95th prediction quantiles/uncertainty bounds for that time step. Uncertainty bounds such as these are probabilistic and represent the predictive uncertainty of the SiB2.5 model as conditioned on the input data, the observations, the parameter sets and the chosen likelihood measure (Beven and Freer, 2001).

**Table 3 – Minimum root mean square difference between observations and simulations across all 20,000 simulations**

| Month | NEE ($\mu$mol m$^{-2}$ s$^{-1}$) | LE (W m$^{-2}$) | H (W m$^{-2}$) |
|---|---|---|---|
| All | 2.68 | 42.04 | 36.60 |
| January | 0.66 | 8.05 | 41.25 |
| February | 0.85 | 15.00 | 30.92 |
| March | 0.65 | 19.81 | 31.08 |
| April | 1.27 | 23.20 | 39.36 |
| May | 1.73 | 37.85 | 42.69 |
| June | 4.56 | 62.42 | 35.30 |
| July | 4.44 | 70.38 | 28.82 |
| August | 4.59 | 64.21 | 26.76 |
| September | 3.09 | 52.50 | 26.32 |
| October | 1.93 | 32.58 | 34.16 |
| November | 1.19 | 14.23 | 24.76 |
| December | 0.46 | 11.00 | 17.82 |

## 3. Results

### 3.1. *Minimum RMSD errors*

There is a minimum level of error between observations and simulations that could not be surpassed despite the many parameter combinations and thousands of simulations (Table 3). The magnitude of this error, however, varies throughout the year. For NEE and LE the largest errors occur during the growing season: June, July, August and September. For H, the largest errors occur during the shoulder seasons: April, May and October as well as in January. In general, the largest errors occur in times of highest flux.

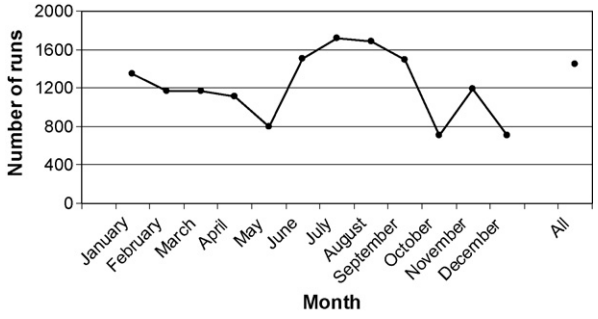



**Fig. 3 – Number of parameter sets common to both LE, H and NEE, LE, H multi-objective criteria. Common parameter sets were evaluated in the 2000 best-performing (smallest likelihood value) parameter sets for each multi-objective criterion. A larger proportion of common runs indicates a stronger degree of coupling between energy balance and net ecosystem exchange.**

### 3.2. *Parameter rankings*

Parameter rankings are given in Tables 4 and 5. Many of the parameters the model is most sensitive to are shared by both criteria and are therefore the most important to define well. Considering only the 10 most influential parameters in each month, three parameters stand out as present in this group in each month and for both criteria, these are: ROOTD, $Z_1$ and HHTI. Other parameters in this most influential group that the two criteria share at least half the time (e.g. $\geq$6 months) are: ZC, TRDM, SOREF2 and $Z_2$. HHTI and TRDM, which are both photosynthesis parameters, are highly influential for both criteria, indicating correct simulation of $CO_2$ fluxes is important in constraining the energy as well as the carbon fluxes.

The total number of influential parameters varies through time for the two multi-objective criteria, from a low of 28% of the parameters (LE, H criterion in January) to a high of 63% of the parameters (NEE, LE, H criterion in May). At any given time, model results are relatively insensitive to nearly half the input parameters. A greater proportion of common parameter sets indicate that the processes controlling NEE, LE and H are more coupled during those time periods. The highest number of parameter sets common to both criteria occurs in mid-growing season and in the winter, the lowest number of common

**Table 4 – Model parameters from the top performance class ranked for each month by the Kolmogorov–Smirnov coefficient for the LE, H multi-objective criteria**

| Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | All |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| z2 | z2 | rootd | nd98 | soref2 | hhti | hhti | hhti | hhti | zc | z2 | z2 | hhti |
| zc | zc | nd98 | rootd | soref1 | hlti | gradm | gradm | rootd | z2 | zc | zc | hlti |
| z1 | z1 | z2 | gradm | z1 | rootd | hlti | hlti | gradm | nd98 | z1 | z1 | rootd |
| rootd | rootd | zc | soref1 | zc | z1 | rootd | rootd | z1 | z1 | soref2 | soref2 | gradm |
| hhti | hhti | z1 | ltmax | rootd | zc | vmax0 | z1 | vmax0 | rootd | chil | soref1 | z1 |
| trdm | soref2 | stem | z1 | z2 | gradm | z1 | vmax0 | hlti | chil | soref1 | sodep | zc |
| chil | trdm | nd02 | soref2 | hhti | soref2 | respcp | zc | zc | soref2 | sodep | rootd | z2 |
| soref2 | nd98 | chil | zc | lwidth | trdm | zc | respcp | vcover | hhti | bee | hhti | vmax0 |
| vcover | stem | hhti | vmax0 | trdm | vcover | trdm | z2 | nd98 | trdm | rootd | trdm | soref2 |
| bee | nd02 | trdm | hhti | vcover | binter | vcover | trdm | z2 | ltmax | hhti | bee | trdm |
| nd98 | chil | soref2 | trdm | gradm | respcp | soref2 | vcover | respcp | lwidth | trdm | nd98 | respcp |
| poros | sodep | ltmax | bee | ref21 | vmax0 | binter | lwidth | lwidth | sodep | vcover | poros | vcover |
| sodep | ltmax | vcover | chil | chil | nd98 | effcon | nd98 | trdm | soref1 | satco | satco | binter |
| lwidth | lwidth | tran21 | stem | hlti | trop | trop | soref2 | soref2 | bee | lwidth | tran21 | trop |
| satco | tran21 | ref22 | nd02 | nd98 | z2 | z2 | binter | binter | ref21 | gradm | stem | lwidth |
| tran21 | bee | ref21 | binter | tran21 | poros | soref1 | effcon | ref21 | gradm | llength | ltmax | effcon |
| ref12 | ref21 | sodep | sodep | bee | effcon | ref21 | trop | effcon | tran21 | poros | ref21 | chil |
| llength | llength | tran22 | ref21 | vmax0 | lwidth | btheta | ref21 | btheta | binter | phsat | trop | nd98 |
| effcon | slti | hlti | respcp | stem | shti | atheta | soref1 | soref1 | satco | tran21 | tran11 | stem |
| phsat | shti | btheta | satco | satco | sodep | tran21 | ltmax | ltmax | poros | vmax0 | ref11 | sodep |
| binter | binter | bee | tran21 | binter | ltmax | ph | ph | tran21 | ref11 | trop | slope | soref1 |
| ref22 | respcp | gradm | vcover | ref11 | satco | lwidth | btheta | ref11 | nd02 | nd98 | lwidth | ltmax |
| soref1 | ref12 | tran11 | trop | sodep | soref1 | ref11 | tran21 | ph | ref22 | ref21 | zm | nd02 |
| ltmax | ref22 | poros | hlti | effcon | atheta | bee | atheta | atheta | trop | wopt | respcp | satco |
| respcp | soref1 | ref12 | lwidth | trop | ph | chil | chil | stem | llength | nd02 | binter | ph |
| tran12 | satco | llength | llength | atheta | chil | poros | stem | trop | tran22 | slti | btheta | atheta |
| stem | poros | lwidth | poros | respcp | btheta | stem | ref11 | bee | btheta | ref12 | chil | poros |
| slope | wopt | soref1 | z2 | shti | tran22 | ref22 | bee | sodep | ref12 | binter | vcover | tran11 |
| trop | tran22 | tran12 | ph | slti | slti | sodep | poros | chil | vcover | atheta | ref22 | ref22 |
| btheta | effcon | effcon | phsat | btheta | tran21 | trda | nd02 | shti | respcp | ph | effcon | bee |
| ref11 | atheta | vmax0 | atheta | tran11 | phsat | ltmax | shti | nd02 | stem | ltmax | hlti | btheta |
| wopt | trop | respcp | tran11 | poros | slope | satco | satco | satco | slope | slope | tran22 | ref12 |
| gradm | hlti | wsat | wsat | zm | bee | tran11 | ref12 | ref22 | shti | respcp | phsat | tran21 |
| nd02 | slope | slope | effcon | phsat | llength | wsat | trda | llength | hlti | wsat | vmax0 | tran12 |
| wsat | ph | ref11 | shti | slope | wsat | slti | llength | tran11 | slti | trda | nd02 | tran22 |
| vmax0 | tran12 | satco | wopt | trda | tran11 | ref12 | tran12 | poros | wopt | stem | wopt | phsat |
| trda | ref11 | atheta | ref11 | tran22 | ref21 | nd98 | ref22 | wsat | zm | tran22 | gradm | llength |
| shti | gradm | phsat | trda | llength | nd02 | phsat | sodep | tran22 | atheta | tran12 | shti | wsat |
| zm | btheta | wopt | slope | ref12 | ref22 | slope | slope | slti | tran12 | shti | slti | zm |
| slti | vmax0 | zm | btheta | ph | tran12 | wopt | wopt | trda | vmax0 | effcon | ph | trda |
| tran22 | phsat | binter | zm | tran12 | stem | nd02 | phsat | phsat | phsat | tran11 | tran12 | wopt |
| hlti | wsat | shti | ref22 | wopt | wopt | shti | zm | tran12 | ph | ref11 | ref12 | slti |
| atheta | vcover | slti | slti | ltmax | ref12 | tran12 | slti | slope | wsat | hlti | atheta | ref21 |
| tran11 | trda | ph | tran12 | nd02 | trda | tran22 | wsat | wopt | effcon | zm | trda | ref11 |
| ph | zm | trop | tran22 | wsat | ref11 | llength | tran11 | zm | tran11 | ref22 | llength | slope |
| ref21 | tran11 | trda | ref12 | ref22 | zm | zm | tran22 | ref12 | trda | btheta | wsat | shti |

Parameters are ranked in descending order of influence. Shaded parameters are significant at the $p = 0.01$ level.

**Table 5 – Model parameters from the top performance class ranked for each month by the Kolmogorov–Smirnov coefficient for the LE, H, NEE multi-objective criteria**

| Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | All |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| z2 | z2 | rootd | vmax0 | soref2 | hhti | hhti | hhti | hhti | hhti | zc | hlti | hhti |
| zc | zc | hhti | nd98 | vmax0 | hlti | hlti | hlti | vmax0 | z1 | z2 | hhti | hlti |
| z1 | hhti | stem | rootd | soref1 | rootd | gradm | gradm | rootd | zc | z1 | z1 | rootd |
| rootd | z1 | hlti | gradm | rootd | z1 | vmax0 | vmax0 | gradm | z2 | hhti | zc | vmax0 |
| hhti | rootd | zc | hhti | z1 | vmax0 | rootd | rootd | hlti | rootd | soref2 | soref2 | z1 |
| trdm | hlti | z2 | z1 | nd98 | zc | z1 | z1 | z1 | nd98 | chil | z2 | gradm |
| respcp | respcp | z1 | soref1 | hhti | gradm | respcp | zc | zc | vmax0 | sodep | rootd | zc |
| soref2 | trdm | nd98 | trop | zc | soref2 | zc | respcp | vcover | respcp | trdm | soref1 | trdm |
| chil | trop | trdm | hlti | trdm | trdm | trdm | trdm | trdm | soref2 | rootd | sodep | soref2 |
| vcover | nd98 | vmax0 | trdm | hlti | nd98 | soref2 | soref2 | respcp | trdm | respcp | trdm | respcp |
| trop | soref2 | trop | zc | effcon | binter | effcon | z2 | soref2 | chil | soref1 | trop | trop |
| nd98 | stem | soref2 | stem | slti | respcp | binter | lwidth | z2 | hlti | bee | vmax0 | binter |
| hlti | chil | chil | ltmax | gradm | trop | trop | vcover | effcon | slti | trop | respcp | z2 |
| bee | effcon | respcp | soref2 | z2 | vcover | vcover | effcon | lwidth | lwidth | vcover | stem | effcon |
| sodep | ltmax | nd02 | binter | ref21 | effcon | btheta | binter | binter | bee | hlti | nd98 | vcover |
| poros | btheta | ltmax | bee | btheta | ph | ph | trop | btheta | sodep | slti | bee | chil |
| vmax0 | nd02 | vcover | sodep | trop | poros | atheta | btheta | ref21 | ltmax | satco | vcover | nd98 |
| soref1 | bee | gradm | ref21 | ltmax | sodep | ref21 | nd98 | ph | vcover | lwidth | ref11 | ph |
| btheta | lwidth | tran21 | chil | shti | tran22 | soref1 | ph | nd98 | effcon | nd98 | ref21 | lwidth |
| wsat | tran21 | ref22 | ph | respcp | shti | nd98 | atheta | soref1 | trop | vmax0 | ltmax | sodep |
| lwidth | tran11 | sodep | nd02 | atheta | btheta | tran21 | soref1 | trop | binter | llength | slti | atheta |
| ltmax | ref22 | ref21 | lwidth | vcover | chil | z2 | ref21 | atheta | shti | gradm | chil | poros |
| effcon | atheta | effcon | vcover | lwidth | soref1 | ref11 | chil | ref11 | satco | phsat | poros | satco |
| llength | gradm | ref12 | btheta | stem | atheta | bee | ltmax | tran21 | gradm | wsat | ph | tran21 |
| stem | vmax0 | ph | tran21 | tran21 | ltmax | lwidth | tran21 | stem | stem | effcon | effcon | btheta |
| tran11 | wsat | tran11 | satco | satco | satco | poros | poros | bee | poros | binter | slope | slti |
| gradm | ref21 | tran12 | shti | ph | slope | ref12 | ref11 | ltmax | ph | poros | tran21 | ref12 |
| ref11 | satco | ref11 | wsat | bee | lwidth | chil | stem | zm | btheta | wopt | wsat | stem |
| tran12 | slope | binter | tran11 | chil | ref12 | ltmax | bee | chil | ref21 | shti | binter | ref22 |
| binter | binter | trda | slti | llength | tran21 | ref22 | phsat | sodep | tran21 | ref21 | satco | soref1 |
| phsat | vcover | soref1 | atheta | ref11 | bee | stem | trda | tran11 | ref12 | nd02 | gradm | ltmax |
| satco | tran12 | slope | poros | nd02 | trda | trda | ref12 | poros | ref22 | tran21 | trda | ref21 |
| wopt | phsat | bee | z2 | sodep | llength | satco | shti | tran12 | soref1 | ltmax | shti | tran22 |
| zm | zm | tran22 | wopt | slope | ref11 | tran22 | tran11 | shti | wsat | tran12 | phsat | nd02 |
| slti | sodep | wsat | ref11 | tran12 | zm | sodep | zm | wsat | atheta | slope | lwidth | slope |
| tran22 | slti | llength | trda | wsat | ref21 | llength | wsat | slti | wopt | tran11 | tran11 | trda |
| slope | llength | lwidth | phsat | poros | stem | wsat | wopt | slope | trda | tran22 | nd02 | bee |
| ref21 | ref11 | poros | effcon | tran11 | phsat | shti | sodep | ref12 | llength | trda | llength | wsat |
| shti | wopt | atheta | llength | binter | wopt | zm | tran12 | satco | tran11 | atheta | ref22 | wopt |
| ref22 | soref1 | wopt | slope | tran22 | z2 | phsat | llength | nd02 | tran22 | ref12 | atheta | llength |
| tran21 | trda | btheta | respcp | wopt | nd02 | tran11 | nd02 | tran22 | slope | stem | zm | ref11 |
| ph | ph | phsat | zm | zm | slti | nd02 | satco | ref22 | phsat | zm | tran12 | phsat |
| ref12 | ref12 | satco | tran12 | ref22 | wsat | tran12 | slope | trda | zm | ref11 | btheta | shti |
| trda | tran22 | shti | ref12 | ref12 | tran12 | slope | slti | wopt | nd02 | ref22 | ref12 | tran11 |
| atheta | shti | zm | ref22 | trda | ref22 | wopt | tran22 | llength | tran12 | btheta | tran22 | tran12 |
| nd02 | poros | slti | tran22 | phsat | tran11 | slti | ref22 | phsat | ref11 | ph | wopt | zm |

Parameters are ranked in descending order of influence. Shaded parameters are significant at the $p = 0.01$ level.

parameter sets occurs in the shoulder seasons of spring and fall (Fig. 3).

### 3.3. *Posterior distributions*

The posterior distributions of parameters from the retained parameter sets contain information on the appropriateness of the parameter range used in the analysis and information on optimum parameter values (Braswell et al., 2005). Analyses of posterior distributions are given below for parameters used to calculate aerodynamic resistance, $f_{PAR}$, energy balance and photosynthesis. Though an analysis of all the parameters is not given here, some general insights can be drawn from these examples for LSMs.

The model is sensitive to the specification of canopy base height ($Z_1$) and canopy inflection height (ZC) for all months, and canopy top height ($Z_2$) in nearly all months for both criteria. These parameters are used in SiB to determine aerodynamic resistance and canopy snow in winter. An examination of the frequency histograms for the parameters $Z_1$ and $Z_2$ using the LE, H criterion shows that in the winter months, parameter sets which produce the best runs have both shorter canopies and smaller crown depths ($Z_2 - Z_1$) (Fig. 4a and b). Clearly, it is not reasonable to expect canopy dimensions to change during the year and short canopy heights are not appropriate.

The parameters ND98 (98th percentile NDVI), ND02 (2nd percentile NDVI), LTMAX (maximum leaf area) and STEM (stem area index) are used in the calculation of available energy and net assimilation of carbon via $f_{PAR}$ and leaf area index. Lower values of ND98 increase estimates of leaf area and $f_{PAR}$. Higher values of ND98 decrease estimates of leaf area

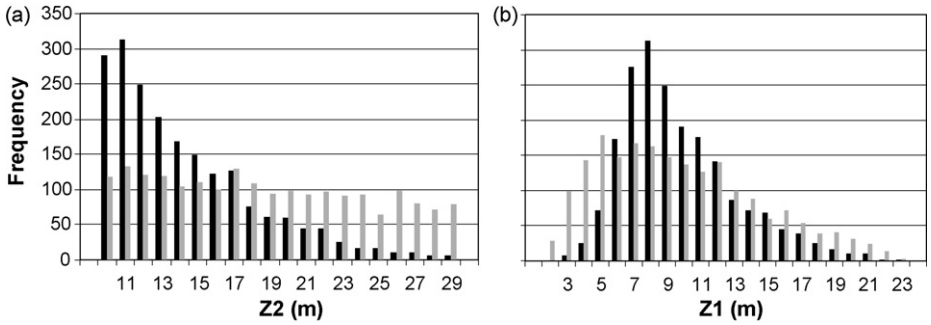



**Fig. 4 – Frequency histograms of: (a) Canopy top height ($Z_2$) and (b) Canopy bottom height ($Z_1$) for the LE, H criterion in January (black) and July (grey) 1997. January is representative of the winter months, July is representative of the summer months.**

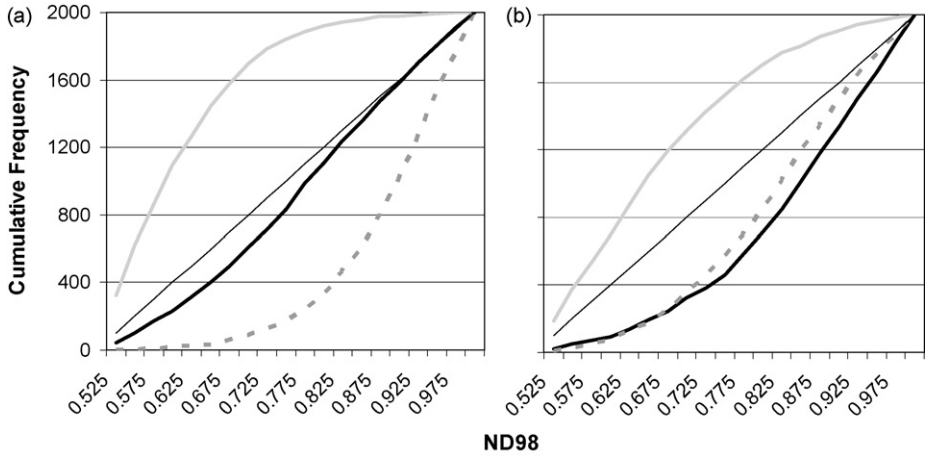



**Fig. 5 – Cumulative frequency diagrams for the parameter ND98 for the months April (thick grey), May (thick black) and October (dashed grey) and the uniform distribution (thin black). (a) LE, H criterion and (b) NEE, LE, H criterion.**

and $f_{PAR}$. SiB2.5 shows the greatest sensitivity to ND98 in the shoulder seasons (Fig. 5). For the LE, H criterion and the parameter ND98, the best-performing values in April are in the lower part of its range while the best-performing values in October are in the upper part of its range (Fig. 5a). For the NEE, LE, H criterion, the situation is similar to the LE, H criterion, though less extreme, in April and October (Fig. 5b). In May however, unlike in the LE, H criterion, the best-performing values for the parameter ND98 are also in the upper part of its range.

Soil reflectance in SiB2.5 is used in the calculation of surface albedo and consequently available energy. In standard SiB2.5, soil reflectance is assigned by biome, meaning that all soils for a particular biome have the same reflectance properties that do not change with soil texture, season or the moisture content of the soil. Sensitivity to these parameters is exhibited year round with the LE, H criterion (Fig. 6) and in particular in May, November and December for both criteria.

The parameter HHTI, the photosynthesis high temperature inhibition 1/2-point, consistently displayed a normal distribution with a peak between 300 and 310 K (e.g. Fig. 7a). The stability of the distribution through time suggests that there is a relatively fixed optimal value for this parameter. The posterior distribution for the photosynthesis low temperature inhibition 1/2-point (HLTI) favored values that were in the upper half of its range, particularly in summer (Fig. 7b). TRDM, the autotrophic respiration inhibition 1/2-point temperature, was sensitive across all months with a normal distribution centered consistently on 321 K, about seven degrees lower than the default value.

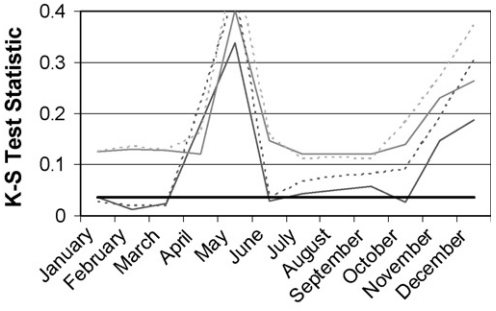



**Fig. 6 – Plot of the Kolmogorov–Smirnov test statistic for visible (black) and near infrared (grey) soil reflectance for both multi-objective criteria. Thin solid lines represent the NEE, LE, H criterion, dashed lines indicate the LE, H criterion. Sensitivity is indicated if points fall above the thick black line ($p = 0.01$).**

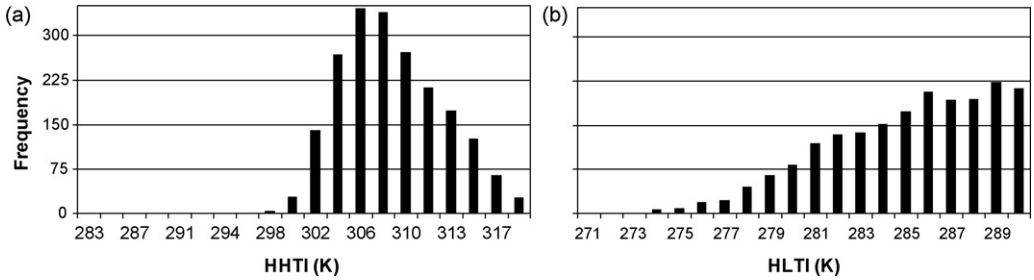



**Fig. 7 – Frequency histograms for (a) photosynthesis high temperature inhibition 1/2-point (HHTI) and (b) photosynthesis low temperature inhibition 1/2-point (HLTI) for the month of July using the NEE, LE, H criterion.**

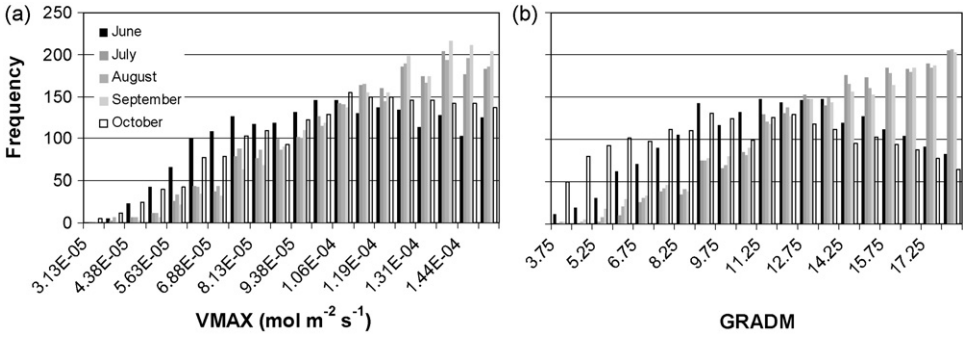



**Fig. 8 – Frequency histograms for the parameters (a) rubisco velocity of a leaf in full sun (VMAX) and (b) the slope of the Ball–Berry photosynthesis–conductance relationship (GRADM) for the growing season using the NEE, LE, H criterion. The black and white bars show lower optimal values at the beginning (June; white) and end (October; black) of the growing season, respectively, the gray bars show higher optimal values during the height of the growing season (July, August, September).**

Three parameters that control the rates and efficiency of photosynthesis show similar patterns: the rubisco velocity of a leaf in full sun (VMAX), the slope of the Ball–Berry photosynthesis–conductance relationship (GRADM) and the minimum stomatal conductance (BINTER). They are all highly sensitive during the growing season but also show a peak in sensitivity in April. During the growing season, the parameters VMAX and GRADM undergo shifts in their ranges that appear to be ecologically meaningful. At the beginning and the end of the growing season, the best-performing parameter sets have lower values (indicating reduced photosynthetic capacity and reduced conductance) than during the height of the growing season (Fig. 8). This is consistent with results from a recent model-data comparison with SiB2 in an agricultural–prairie system indicating that photosynthetic capacity was overestimated by SiB2 at the beginning and end of the growing season and that both young and senescing leaves might have less photosynthetic capacity than high-season leaves (Hanan et al., 2005).

### 3.4. *Uncertainty estimation*

The expectation, given a realistic model of the system, is that the predictive uncertainty bounds will be wide enough to encompass the observations and that the observations will fall in the center of the bounds (Beven and Freer, 2001). The parameters were constrained at monthly and annual time scales and so systematic errors at hourly and daily time scales can indicate an inability of the model to capture some relatively high frequency process or environmental condition, or noise in the observations.

Fig. 9 shows example diurnal cycles of NEE, LE and H and the uncertainty arising from parameter choice for 2 days in 1997, July 29 and October 23. The predictive uncertainty bounds shown were constrained monthly with the NEE, LE, H criterion. The observations of NEE are well centered within the uncertainty bounds in both time periods (Fig. 9a and d). Sensible heat flux observations (Fig. 9c and f), are also quite well centered within the uncertainty bounds in July, but in October the observations fall well above the uncertainty bounds during the daytime. For LE there is more variability during both time periods (Fig. 9b and e). In both July and October, the model is unable to simulate the relatively higher afternoon latent heat flux. The magnitude of the uncertainty bounds also varies with time of day (or magnitude of flux).

There are patterns in both the observations and the hourly predictive uncertainty bounds that occur within a month (Fig. 10). At the WLEF site, leaf-off occurred in the days around October 10th (Davis et al., 2003). After this time, both the observations and uncertainty bounds for NEE change in magnitude, reflecting the loss of photosynthesizing vegetation (Fig. 10b). During the time period just before and after leaf-off, observations of nighttime net ecosystem exchange fall well below the predictive uncertainty bounds, indicating systematic under prediction of nighttime NEE during this time period. In other months, such as July, the relationship between

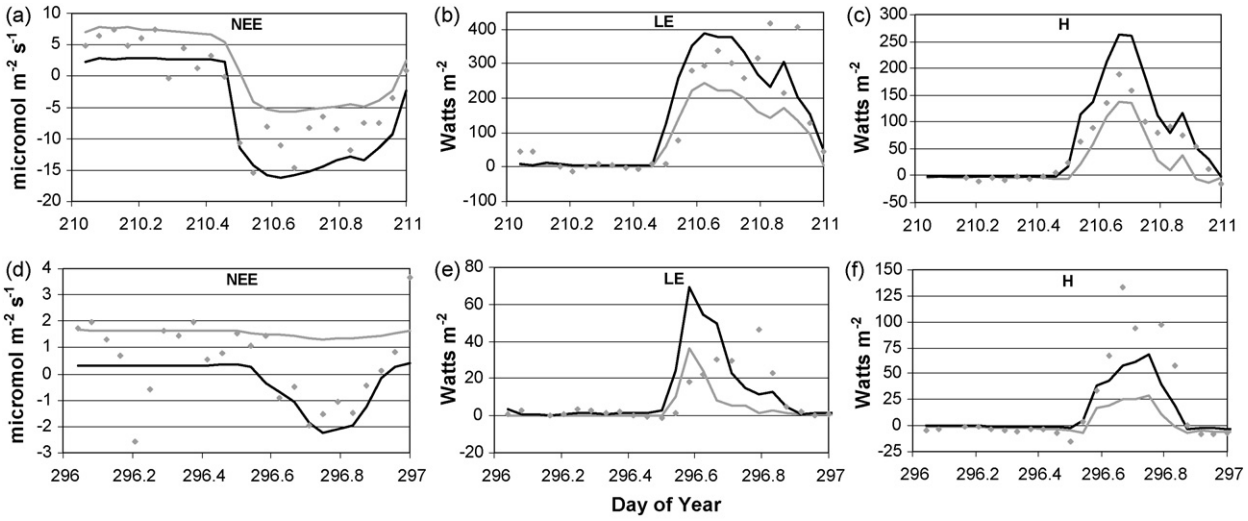



**Fig. 9 – Diurnal cycles of observations (symbol) and predictive uncertainty bounds (5th percentile in grey, 95th percentile in black) for July 29 (a–c) and October 23 (d–f), 1997 for NEE (a and d), LE (b and e) and H (c and f). Predictive uncertainty bounds were constrained with the NEE, LE, H criterion and parameter sets were conditioned monthly.**

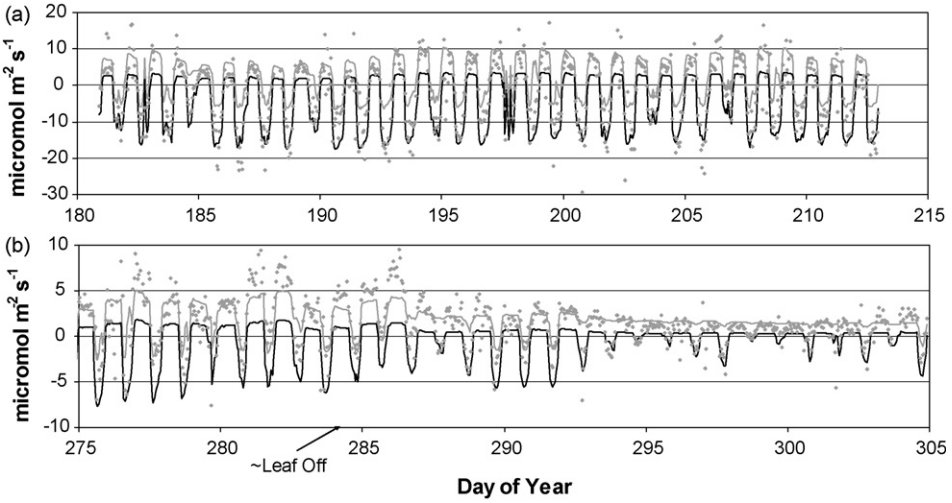



**Fig. 10 – Hourly predictive uncertainty bounds for NEE constrained monthly on the NEE, LE, H criterion (a) July, 1997 and (b) October, 1997. Observations are symbols, 5th percentile uncertainty bound is in grey, 95th percentile uncertainty bound is in black.**

**Table 6 – Percentage of observations above, within and below the 5th and 95th percentile predictive uncertainty bounds for latent heat flux**

| | LE, H | | | NEE, LE, H | | |
|---|---|---|---|---|---|---|
| | %Above | %Within | %Below | %Above | %Within | %Below |
| January | 43 | 19 | 38 | 42 | 19 | 39 |
| February | 53 | 11 | 36 | 52 | 12 | 36 |
| March | 44 | 13 | 43 | 45 | 13 | 42 |
| April | 21 | 19 | 60 | 20 | 19 | 61 |
| May | 40 | 33 | 27 | 43 | 30 | 27 |
| June | 29 | 35 | 36 | 28 | 36 | 36 |
| July | 28 | 36 | 36 | 28 | 36 | 36 |
| August | 32 | 33 | 35 | 33 | 32 | 35 |
| September | 26 | 29 | 45 | 26 | 28 | 46 |
| October | 32 | 29 | 39 | 30 | 31 | 39 |
| November | 44 | 20 | 36 | 44 | 21 | 35 |
| December | 54 | 15 | 31 | 55 | 16 | 29 |

LE, H and NEE, LE, H constraints, conditioned monthly.

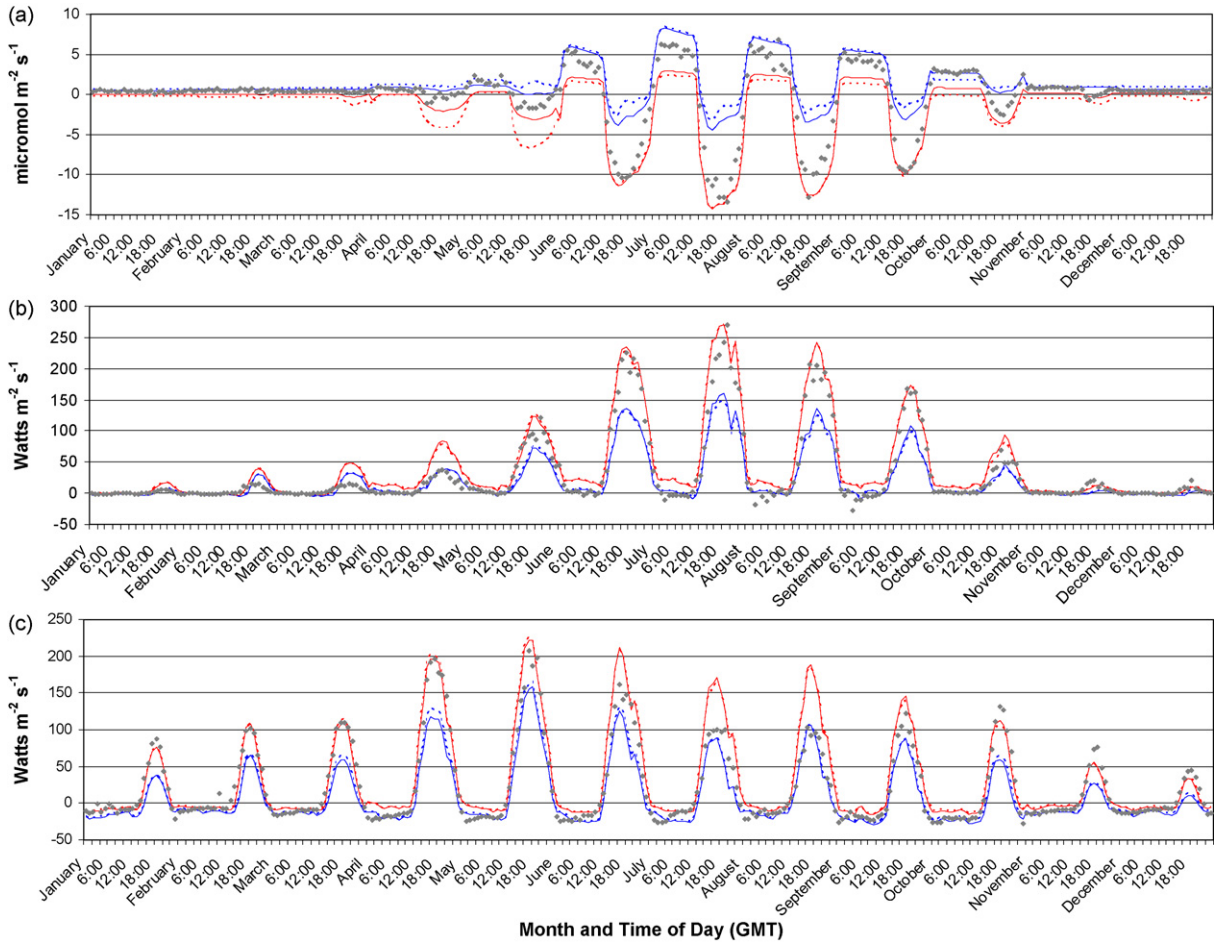



**Fig. 11 – Monthly mean diurnal predictive uncertainty constrained monthly on LE, H (dashes) and NEE, LE, H (solid line) for (a) NEE, (b) LE and (c) H. Observations are symbols, 5th percentile uncertainty bound are in blue, 95th percentile uncertainty bound are in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)**

observations and the predictive uncertainty bounds is more stable (Fig. 10a).

### 3.5. *Predictive uncertainty with additional constraints*

Monthly mean diurnal cycles provide a useful summary of the results. Fig. 11 presents the monthly mean diurnal uncertainty in simulations of NEE, LE and H constrained monthly on LE, H and NEE, LE, H. The two constraints, LE, H and NEE, LE, H, resulted in very similar predictive uncertainty bounds for both LE and H (Fig. 11b and c; Tables 6 and 7). The optimized model frequently exhibits a systematic bias in H and LE which cannot be corrected by adjusting parameters (Fig. 11b and c). For both LE and H, less than 50% of the observations fall within the uncertainty bounds, with better performance in summer. In many months, where one heat flux shows systematic over-prediction, the other shows systematic under-prediction. Simulated NEE shows much less systematic bias relative to its uncertainty than LE and H (Fig. 11a, Table 8).

There are more obvious differences between the predictive uncertainty bounds for the two constraints (LE, H and NEE, LE, H) with respect to observations of NEE than there were for observations of LE and H. April, May and October show the most difference, with the predictive uncertainty bounds being narrower in April and May for the NEE, LE, H constraint and the uncertainty bound for October being narrower in the nighttime for the LE, H constraint. The summer months, which were time periods shown to have a large number of common simulations between the two constraints, show the least difference between the two constraints.

### 3.6. *Annual versus monthly constraints*

The observational constraint on model parameterization, and resulting estimation uncertainty, was evaluated at two different time scales: monthly and for the annual cycle (Fig. 12). For NEE, daytime fluxes were better captured by the monthly constraint while nighttime fluxes were better captured by the annual constraint, reflecting the time scales on which these two processes operate. Ecosystem respiration in SiB2.5 is scaled to match the annual assimilation of carbon

**Table 7 – Percentage of observations above, within and below the 5th and 95th percentile predictive uncertainty bounds for sensible heat flux**

| | LE, H | | | NEE, LE, H | | |
|---|---|---|---|---|---|---|
| | %Above | %Within | %Below | %Above | %Within | %Below |
| January | 48 | 15 | 37 | 48 | 16 | 36 |
| February | 48 | 18 | 34 | 48 | 18 | 34 |
| March | 48 | 18 | 34 | 49 | 18 | 33 |
| April | 29 | 31 | 40 | 29 | 31 | 40 |
| May | 29 | 28 | 43 | 30 | 28 | 42 |
| June | 23 | 39 | 38 | 23 | 39 | 38 |
| July | 20 | 44 | 36 | 21 | 44 | 35 |
| August | 25 | 33 | 42 | 25 | 34 | 41 |
| September | 19 | 39 | 42 | 20 | 40 | 40 |
| October | 38 | 22 | 40 | 38 | 27 | 35 |
| November | 43 | 20 | 37 | 44 | 20 | 36 |
| December | 50 | 18 | 32 | 51 | 21 | 28 |

LE, H and NEE, LE, H constraints conditioned monthly.

by the vegetation. Conditioning on monthly measurements was more effective for LE and H. Predictive uncertainty for daytime NEE was reduced in April, May and June for the annual constraint, whereas predictive uncertainty for nighttime NEE was reduced in the summer months. For LE, the monthly constrained predictive uncertainty was less than the annually constrained predictive uncertainty in all months except July and August.

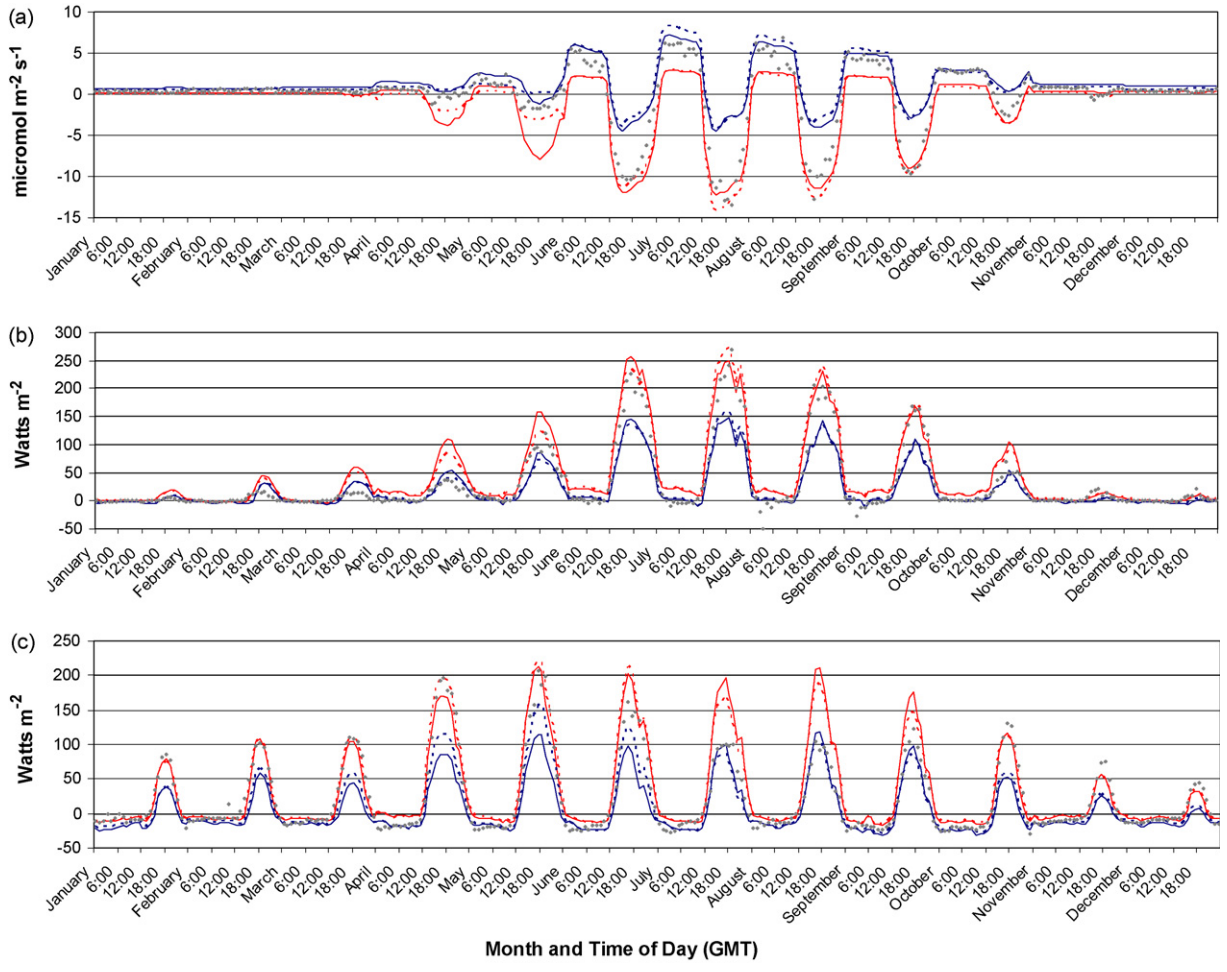



**Fig. 12 – Monthly mean diurnal predictive uncertainty constrained annually (solid line) and monthly (dashed line) on NEE, LE, H for (a) NEE, (b) LE and (c) H. Observations are symbols, 5th percentile uncertainty bound are in blue, 95th percentile uncertainty bound are in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)**

**Table 8 – Percentage of observations above, within and below the 5th and 95th percentile predictive uncertainty bounds for net ecosystem exchange**

| | LE, H | | | NEE, LE, H | | |
|---|---|---|---|---|---|---|
| | %Above | %Within | %Below | %Above | %Within | %Below |
| January | 8 | 56 | 36 | 22 | 39 | 39 |
| February | 6 | 57 | 37 | 18 | 46 | 36 |
| March | 4 | 76 | 20 | 21 | 51 | 28 |
| April | 7 | 74 | 19 | 13 | 58 | 29 |
| May | 6 | 74 | 20 | 10 | 48 | 42 |
| June | 20 | 63 | 17 | 21 | 59 | 20 |
| July | 18 | 71 | 11 | 19 | 68 | 13 |
| August | 18 | 61 | 21 | 20 | 57 | 23 |
| September | 15 | 66 | 19 | 18 | 60 | 22 |
| October | 5 | 40 | 55 | 9 | 51 | 40 |
| November | 3 | 74 | 23 | 17 | 56 | 27 |
| December | 3 | 87 | 10 | 29 | 39 | 32 |

LE, H and NEE, LE, H constraints conditioned monthly.

## 4. Discussion and conclusions

In conducting this analysis, in which tens of thousands of simulations were run with randomly varying but realistic parameter sets, it was found that there was an irreducible level of mismatch between the simulated and observed fluxes. It was not possible to predict fluxes of latent and sensible heat and net ecosystem exchange more precisely than these limits, which varied by month and by flux (Table 3). This minimum possible error reflects a level of variability (or noise) in the observations, structural problems with the model and/or mismatch between the model and observations that cannot be overcome by varying parameters. The variability in the observations may be related to measurement errors; surveys of FLUXNET data suggest there is a mean energy imbalance in the measurements of approximately 20% (Wilson et al., 2002) and other research suggests that errors in NEE are around 7–12% (Baldocchi, 2003; Law et al., 2002). It is also possible that the variability in the measurements stems from the heterogeneous nature of the surface and the changing footprint of the measurements, though the data processing techniques were designed to minimize this. The minimum possible error is a measure of uncertainty that reflects the expected degree of hourly model-data mismatch, which in this analysis was for periods of 1 month or 1 year. In the context of regional atmospheric inversions, the minimum error can be used to estimate prior model uncertainty where modeled fluxes are used, or provide an estimate of uncertainty for assimilation of measured fluxes.

Patterns in the sensitivity of the parameterization were related to many things: the ecology of the system, the specification of input data and how they were used, and weakness in the model structure itself. As a result, it is difficult to determine in many cases whether better specification of a parameter or better representation of a process is more essential to improving model results. However, some parameters, such as leaf temperature sensitivities HHTI and TRDM, were remarkably consistent and simple adjustment of the parameter values themselves is all that seems necessary for improvement.

Unlike many sensitivity analyses, which have focused on time invariant parameterizations, we investigated how sensitivities change through time. In simulating land surface atmosphere exchanges in SiB2.5 a combination of time varying and time invariant parameters are utilized. The time-varying parameters represent seasonally changing vegetation properties evaluated using the normalized difference vegetation index. The results presented here suggest that a much larger set of parameters may need to be treated as seasonally varying.

Parameters related to photosynthesis were more influential in the growing season and parameters related to energy balance more influential at other times of the year. The greatest sensitivity to parameters, both in the magnitude of the sensitivity as well as in the number of parameters, occurred in the shoulder seasons, those periods of rapid land surface change surrounding leaf-growth in the spring and leaf-fall in the autumn. During these periods a combination of parameters reflecting energy balance, the temperature response of photosynthesis and the aerodynamic resistance of the land surface were most influential.

The variability of the top parameter sets within seasons provided additional useful information and suggests that any optimization on tower flux data should account for and allow for variability at sub-annual time scales in order to capture the most information. The variability of the photosynthesis parameters VMAX, GRADM and HLTI within the growing season suggests that changing the photosynthetic capacity of leaves during the growing season might be required. Similarly, the high variability and sensitivity of the model to the soil reflectance parameters could represent true temporal variability in soil properties that is not represented using fixed values.

Given the number of parameters investigated, it was not surprising to find that SiB2.5 exhibited signs of equifinality. At any given time, half of the parameters were found to be non-influential. It is encouraging that reasonable estimates of these parameters are likely sufficient to obtain a good fit to observations, but their values cannot therefore be estimated from the data used here. The influential and non-influential parameters varied by month and would likely also vary

between vegetation types and depending on different environmental conditions. Further, "non-influential" does not mean that parameters are unimportant. If the non-influential parameters contribute to the mechanistic underpinnings of the model, even if only in a minor way, they are still important, particularly for model applications to new ecosystems or novel environmental conditions.

The results of the sensitivity analysis also emphasized the risk involved in undirected model optimization. Parameter interactions in complex models such as SiB2.5 are very nonlinear and compensation among parameters led to many parameter sets that did an equally good job of predicting fluxes. This is a serious consideration for optimization studies, suggesting that uncertainty resulting from the parameterization of the model should be incorporated in any model-data fusion exercise. Besides compensating for one another, the parameters also compensate for imperfect driver data and model structural problems. In some instances, the parameter distributions resulting from the sensitivity analysis were not biologically realistic with respect to observed conditions (e.g. $Z_1$ and $Z_2$), though they produced the best fluxes. Thus, too highly automated optimization procedures should be viewed with caution, especially since it can be difficult to determine whether better parameter values or better model structure is needed. Sensitivity analyses such as conducted here can help identify important parameters for optimization, avoid obtaining a good fit to data for the wrong reasons and thus maintain predictive power in model simulations.

The addition of NEE changes the ranking and number of influential parameters (Tables 4 and 5), and better constrains model parameterizations within physically or biologically reasonable ranges (e.g. Fig. 5). As such, when using the LE, H criterion alone, the highest ranking (optimal) parameter sets may provide good simulations of energy fluxes for the wrong reasons, resulting in parameterizations that have poor predictive power.

Even using optimized parameters, biases in the simulated fluxes as compared to the observed fluxes were observed. Because the parameters were given a large amount of freedom to vary, systematic errors indicate either an error in the structure or logic of the model that precludes the representation of a process or systematic errors in the observations themselves. The high percentage of observations that fell within the predictive uncertainty bounds for net ecosystem exchange indicates that the structure and logic of the model for this flux is quite good, at least at diurnal time scales. Over and under estimation of daytime net ecosystem exchange in times of rapid landscape change (spring and fall) was associated with the lack of high temporal precision in the time-varying parameterization of the model, especially leaf area index and $f_{PAR}$. Instead of linearly ramping these quantities from the center points of each month, shorter compositing periods of NDVI could be used that would more directly relate to the rapidly changing surface properties at such sensitive times of the year. Biases in the nighttime net ecosystem exchange, which were observed in some months, could be related to the simplified respiration model that responds to temperature, moisture and the annual integral of GPP, but does not respond to recent photosynthetic activity or recent input of litter.

Far fewer of the observations of sensible and latent heat flux fell within the predictive uncertainty bounds and biases were more obvious and persistent than for net ecosystem exchange. There are several possible reasons why heat fluxes are more problematic then net ecosystem exchange for SiB2.5. As mentioned previously, energy balance is enforced in most single point land surface-atmosphere models, however observations of the components of net radiation made in the field rarely close the surface energy budget (Wilson et al., 2002). Further, despite careful rules to ensure that the flux measurements are representative of the heterogeneous landscape at WLEF, turbulence and landscape variability that is not characterized in SiB2.5 could be playing a significant role in the variability of the fluxes. For example, Ewers et al. (2002) found large differences in the transpiration of four different forest types (northern hardwoods, conifer, aspen/fir, forested wetlands) in the WLEF area; in the model, a single ecosystem classification ("mixed forest") represents the WLEF area. Shifts in wind direction at certain times of day or days of year could produce fluxes dominated by one forest type or another that SiB2.5 is not intended to reproduce. It is possible that predicting net ecosystem exchange is less problematic because it is less variable across the different vegetation assemblages at WLEF than are the energy fluxes and therefore mean-field parameterizations and functional groups are adequate representations.

While adding NEE increased model sensitivity and helped constrain model parameterization within reasonable ranges (and thus should reduce potential for model bias) it did not reduce predictive uncertainty bounds for LE and H. Given the coupling of transpiration and carbon assimilation in SiB2.5, this is somewhat surprising. It is possible that the greater variability of the heat fluxes mentioned above prevented the additional constraint from having much of an effect. As expected, the additional NEE constraint did generally reduce the predictive uncertainty of net ecosystem exchange. The least amount of reduction in predictive uncertainty occurred in the summer months. The largest reduction occurred during the spring and fall months, when rapid changes in the landscape (thaw/freeze/leaf-on/leaf-off) and shifts from coupled energy, water and carbon fluxes to a less coupled state increased the power of an additional constraint.

Conventional wisdom suggests that the more data a model is conditioned and constrained with, and the longer the period of time, the better and more robust the results will be. Depending on the temporal scale of variability of the quantity being predicted and the ability of the model to match that scale, this may or may not be true. The results presented here on constraining predictions annually versus monthly show that some quantities, such as nighttime net ecosystem exchange, were on average better constrained annually, whereas other quantities that show more variability, such as daytime NEE, latent and sensible heat flux, were better constrained monthly. The results are partially dependent on the scale of variability represented in SiB2.5. For example, ecosystem respiration is an annually bounded flux in SiB2.5 that varies only with soil temperature and moisture. The assimilation of carbon and latent heat flux are more dependent on shorter-term environmental forcing and temporally varying physiology and thus the monthly constraint

was more appropriate. Constraining on shorter time periods may compensate for a lack of flexibility and variability in the model through adjustments in the parameterization. Understanding these compensations is important for the formulation of robust predictive models, because while some will be ecologically meaningful, others will not. Further, these results also suggest that rather than overwhelm the problem of optimization with observations, there are ways to maximize the value of the data, such as breaking it down into shorter time periods for some processes and not for others.

The evaluation of predictive uncertainty in land surface models can be used to investigate the sources of observed bias and correct the model and/or observations for deficiencies in structure or logic. Another use would be to incorporate the uncertainty into the modeling itself. In the analysis presented here there was clearly variability in the real world that was not captured by the model and may in fact not be easily calculated or parameterized. The addition of variability into the model predictions based on the level of observed uncertainty might be possible, particularly when ensemble runs such as these are impractical. Additionally, larger-scale atmospheric models often must parameterize the uncertainty due to the variability in land surface fluxes. An analysis such as this provides estimates of predictive uncertainty as it varies through time and may be used to quantify the expected model-data mismatch for inversion studies.

The results of model analyses such as this one are partially dependent on the site that is being modeled and the weather conditions during the simulation, because the strengths and weaknesses of any model vary with both of these conditions. Drought or flood conditions that produce anomalous environmental conditions might not provide good predictors of how well a model does at characterizing a site in more average years or how sensitive a model is generally to its parameterization. On the other hand, an anomalous year might help to fine tune parameterizations and/or reveal other deficiencies in model structure and logic not previously identified (Franks et al., 1997).

Given the complexity and large number of parameters of a LSM like SiB2.5, use of 20,000 parameter sets may not be sufficient to fully explore model behavior in all possible parameter configurations. Thus, the parameter sensitivity and uncertainty estimates reported in this paper are based on a relatively small, but important, range of variation in combinations of parameters. The number of simulations directly impacts the robustness of both parameter sensitivities and uncertainty estimates: more simulations would increase the power of the results. Because the full parameter space could not be explored, the sensitivities and uncertainties reported in this paper should be considered only as they relate to the analysis presented here. To refine these estimates, a further analysis with both a reduced parameter space and improved parameters and model formulation based on the results presented here could be done. The GLUE methodology is designed such that additional and/or improved data can be easily incorporated and the value assessed.

The research presented in this paper addresses some of the key issues that carbon cycle science is currently facing and that will become more important within the framework of continental scaling efforts using model-data fusion such as the North American Carbon Program. In particular it addresses the use of carbon cycle data assimilation for the analysis of model sensitivity to parameterization and model behavior. Reducing uncertainty in surface flux estimates from LSMs is critical to regional and large-scale efforts to characterize the carbon cycle. This requires both a reduction in uncertainty in the parameterization of models and a strengthening of the mechanistic understanding on which the models are built. Model-data fusion is non-trivial, requiring the synergistic development of multiple models along with improved observations and significant amounts of computing power. Sensitivity and uncertainty analyses such as the ones presented here help to define the areas where the most benefit can be gained.

## Acknowledgements

This work was supported by a NASA Earth System Science Fellowship—NGT5-30174 and NGT5-30182, National Science Foundation Grant DEB-9977066 and National Oceanic and Atmospheric Administration contract NA17RJ1228. Flux measurements at the WLEF tower were supported in part by the Office of Science (BER), U.S. Department of Energy, through both grant number DE-FG02-03ER63681, and through the Midwestern Regional Center of the National Institute for Global Environmental Change under Cooperative Agreement No. DE-FC03-90ER61010. We thank Stewart Franks for early discussions on these analyses and three anonymous reviewers for constructive comments.

## REFERENCES

Baker, I., Denning, A.S., Hanan, N.P., Prihodko, L., Uliasz, M., Vidale, P.-L., Davis, K.J., Bakwin, P.S., 2003. Simulated and observed fluxes of sensible and latent heat and $CO_2$ at the WLEF-TV tower using SiB2.5. Global Change Biol. 9, 1262–1277.

Bakwin, P.S., Tans, P.P., Hurst, D.F., Zhao, C.L., 1998. Measurements of carbon dioxide on very tall towers: results of the NOAA/CMDL program. Tellus Ser. B 50 (5), 401–415.

Baldocchi, D.D., 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. Global Change Biol. 9 (4), 479–492.

Ball, J.T., Woodrow, I.E., Berry, J.A., 1987. A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions. In: Biggins, J. (Ed.), Progress in Photosynthesis Research. Proceedings of the 8th International Photosynthesis Congress, Martinus Nijhoff, Dordrecht, pp. 221–224.

Bastidas, L.A., Gupta, H.V., Sorooshian, S., Shuttleworth, W.J., Yang, Z.L., 1999. Sensitivity analysis of a land surface scheme using multicriteria methods. J. Geophys. Res. 104 (D16), 19481–19490.

Berger, B.W., Davis, K.J., Yi, C.X., Bakwin, P.S., Zhao, C.L., 2001. Long-term carbon dioxide fluxes from a very tall tower in a northern forest: flux measurement methodology. J. Atmos. Ocean. Technol. 18 (4), 529–542.

Beven, K.J., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298.

Beven, K.J., Freer, J., 2001. Equifinality, data assimilation and uncertainty estimation in mechanistic modelling of

complex environmental systems using the GLUE methodology. J. Hydrol. 249, 11–29.
Bounoua, L., DeFries, R., Collatz, G.J., Sellers, P.J., Khan, H., 2002. Effects of land cover conversion on surface climate. Clim. Change 52 (1–2), 29–64.
Braswell, B.H., Sacks, W.J., Linder, E., Schimel, D., 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. Global Change Biol. 11, 335–355.
Brazier, R.E., Beven, K.J., Freer, J., Rowan, J.S., 2000. Equifinality and uncertainty in physically based soil erosion models: application of the GLUE methodology to WEPP-the water erosion prediction project-for sites in the UK and USA. Earth Surf. Proc. Land. 25, 825–845.
Brazier, R.E., Beven, K.J., Nathony, S.G., Rowan, J.S., 2001. Implications of model uncertainty for the mapping of hillslope-scale soil erosion predictions. Earth Surf. Proc. Land. 26, 1333–1352.
Burrows, S.N., Gower, S.T., Clayton, M.K., Mackay, D.S., Ahl, D.E., Norman, J.M., Diak, G., 2002. Application of geostatistics to characterize LAI for flux towers to landscapes. Ecosystems 5, 667–679.
Colello, G.D., Grivet, C., Sellers, P.J., Berry, J.A., 1998. Modeling of energy, water, and $CO_2$ flux in a temperate grassland ecosystem with SiB2: May–October 1987. J. Atmos. Sci. 55 (7), 1141–1169.
Collatz, G.J., Ribascarbo, M., Berry, J.A., 1992. Coupled photosynthesis-stomatal conductance model for leaves of C4 plants. Aust. J. Plant Physiol. 19 (5), 519–538.
Collatz, G.J., Ball, J.T., Grivet, C., Berry, J.A., 1991. Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration—a model that includes a laminar boundary-layer. Agric. For. Meteorol. 54 (2–4), 107–136.
Collatz, G.J., Bounoua, L., Los, S.O., Randall, D.A., Fung, I.Y., Sellers, P.J., 2000. A mechanism for the influence of vegetation on the response of the diurnal temperature range to changing climate. Geophys. Res. Lett. 27 (20), 3381–3384.
Cosby, B.J., Hornberger, G.M., Clapp, R.B., Ginn, T.R., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water Resour. Res. 20 (6), 682–690.
Cox, P.M., Betts, R.A., Jones, C.D., Spall, S.A., Toterdell, I.J., 2000. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. Nature 408, 184–187.
Davis, K.J., Bakwin, P.S., Yi, C., Berger, B.W., Zhao, C., Teclaw, R.M., Isebrands, J.G., 2003. The annual cycles of $CO_2$ and $H_2O$ exchange over a northern mixed forest as observed form a very tall tower. Global Change Biol. 9, 1278–1293.
Denning, A.S., Fung, I.Y., Randall, D., 1995. Latitudinal gradient of atmospheric $CO_2$ due to seasonal exchange with land biota. Nature 376, 240–243.
Denning, A.S., Randall, D.A., Collatz, G.J., Sellers, P.J., 1996a. Simulations of terrestrial carbon metabolism and atmospheric $CO_2$ in a general circulation model. 2. Simulated $CO_2$ concentrations. Tellus Ser. B 48 (4), 543–567.
Denning, A.S., Nicholls, M., Prihodko, L., Baker, I., Vidale, P.-L., Davis, K.J., Bakwin, P.S., 2003. Simulated variations in atmospheric $CO_2$ over a Wisconsin forest using a coupled ecosystem–atmosphere model. Global Change Biol. 9, 1241–1250.
Denning, A.S., Collatz, G.J., Zhang, C.G., Randall, D.A., Berry, J.A., Sellers, P.J., Colello, G.D., Dazlich, D.A., 1996b. Simulations of terrestrial carbon metabolism and atmospheric $CO_2$ in a general circulation model. 1. Surface carbon fluxes. Tellus Ser. B 48 (4), 521–542.
El Maayar, M., Price, D.T., Black, T.A., Humphreys, E.R., Jork, E.-M., 2002. Sensitivity tests of the integrated biosphere simulator to soil and vegetation characteristics in a pacific coastal coniferous forest. Atmos.-Ocean 40 (3), 313–332.
Ewers, B.E., Mackay, D.S., Gower, S.T., Ahl, D.E., Burrows, S.N., Samanta, S.S., 2002. Tree species effects on stand transpiration in northern Wisconsin. Water Resour. Res. 38 (7) Art. No. 1103.
Farquhar, G.D., Caemmerer, S.V., Berry, J.A., 1980. A biochemical model of photosynthetic $CO_2$ assimilation in leaves of C-3 species. Planta 149 (1), 78–90.
Fedra, K., Van Straten, G., Beck, M.B., 1981. Uncertainty and arbitrariness in ecosystems modelling: a lake modelling example. Ecol. Model. 13, 87–110.
Field, C., Mooney, H.A., 1986. The photosynthesis-nitrogen relationship in wild plants. In: Givnish, T.J. (Ed.), On the Economy of Plant Form and Function. Cambridge University Press, Cambridge, pp. 25–55.
Franks, S.W., 1998. An evaluation of single and multiple objective SVAT model conditioning schemes: parameteric, predictive and extrapolative uncertainty. Report: 167.09.1998, The University of Newcastle, Department of Civil, Surveying and Environmental Engineering, Callagghan, New South Wales.
Franks, S.W., Beven, K.J., 1999. Conditioning a multiple-patch SVAT model using uncertain time-space estimates of latent heat fluxes as inferred from remotely sensed data. Water Resour. Res. 35 (9), 2751–2761.
Franks, S.W., Beven, K.J., Gash, J.H.C., 1999. Multi-objective conditioning of a simple SVAT model. Hydrol. Earth. Syst. Sci. 3 (4), 477–489.
Franks, S.W., Beven, K.J., Quinn, P.F., Wright, I.R., 1997. On the sensitivity of soil-vegetation-atmosphere transfer (SVAT) schemes: equifinality and the problem of robust calibration. Agric. For. Meteorol. 86, 63–75.
Franks, S.W., Gineste, P., Beven, K.J., Merot, P., 1998. On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process. Water Resour. Res. 34 (4), 787–797.
Freer, J., Beven, K.J., 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. Water Resour. Res. 32 (4), 2161–2173.
Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. Water Resour. Res. 34 (4), 751–763.
Gupta, H.V., Bastidas, L., Sorooshian, S., Shuttleworth, W.J., Yang, Z.-L., 1999. Parameter estimation of a land surface scheme using multicriteria methods. J. Geophys. Res. 104 (D16), 19491–19503.
Hanan, N.P., Berry, J.A., Verma, S., Walter-Shea, E.A., Suyker, A.E., Burba, G.G., Denning, A.S., 2005. Model analyses of biosphere-atmosphere exchange of $CO_2$, water and energy in Great plains tall grass prairie and wheat ecosystems. Agric. For. Meteorol. 131 (3–4), 162–179.
Henderson-Sellers, A., 1993. A factorial assessment of the sensitivity of the BATS land-surface parameterization scheme. J. Clim. 6, 227–247.
Hornberger, G.M., Spear, R.C., 1981. An approach to the preliminary analysis of environmental systems. J. Environ. Manage. 12, 7–18.
Idso, S.B., 1981. A set of equations for full spectrum and 8mm to 14 mm and 10.5 mm to 12.5 mm thermal radiation for cloudless skies. Water Resour. Res. 17, 295–304.
Kaminski, T., Knorr, W., Rayner, P.J., Heimann, M., 2002. Assimilating atmospheric data into a terrestrial biosphere model: a case study of the seasonal cycle. Global Biogeochem. Cy. 16 (4).

Kaminski, T., Giering, R., Scholze, M., Rayner, P., Knorr, W., 2003. An example of an automatic differentiation-based modelling system. In: Computational Science and Its Applications—Iccsa 2003, Pt 2, Proceedings. Lecture Notes in Computer Science. pp. 95–104.

Knorr, W., Kattge, J., 2005. Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. Global Change Biol. 11, 1333–1351.

Law, B.E., Falge, E., Gu, L., Baldocchi, D.D., Bakwin, P., Berbigier, P., Davis, K., Dolman, A.J., Falk, M., Fuentes, J.D., Goldstein, A., Granier, A., Grelle, A., Hollinger, D., Janssens, I.A., Jarvis, P., Jensen, N.O., Katul, G., Mahli, Y., Matteucci, G., Meyers, T., Monson, R., Munger, W., Oechel, W., Olson, R., Pilegaard, K., Paw, K.T., Thorgeirsson, H., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2002. Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation. Agric. For. Meteorol. 113 (1–4), 97–120.

Los, S.O., Collatz, G.J., Sellers, P.J., Malmstrom, C.M., Pollack, N.H., DeFries, R.S., Bounoua, L., Parris, M.T., Tucker, C.J., Dazlich, D.A., 2000. A global 9-yr biophysical land surface dataset from NOAA AVHRR data. J. Hydrometeorol. 1 (2), 183–199.

Martin, L., 1965. The Physical Geography of Wisconsin. The University of Wisconsin Press, Madison and Milwaukee, p. 608.

Meixner, T., Gupta, H.V., Bastidas, L., Bales, R., 1999. Sensitivity analysis using mass flux and concentration. Hydrol. Process. 13 (2233–2244).

Mo, X., Beven, K.J., 2004. Multi-objective parameter conditioning of a three-source wheat canopy model. Agric. For. Meteorol. 122, 39–63.

MRCC, 2002. Historical Climate Summaries. Midwestern Regional Climate Center.

Nicholls, M.E., Denning, A.S., Prihodko, L., Vidale, P.-L., Baker, I., Davis, K.J., Bakwin, P.S., 2004. A Multiple-scale simulation of variations in atmospheric carbon dioxide using a coupled biosphere-atmospheric model. J. Geophys. Res. 109 (D18117), 18.

O'Neill, R.V., Gardner, R.H., Carney, J.H., 1982. Parameter constraints in a stream ecosystem model: incorporation of a priori information in Monte Carlo error analysis. Ecol. Model. 16, 51–65.

Pielke, R.A., Avissar, R., Raupach, M., Dolman, A.J., Zeng, X.B., Denning, A.S., 1998. Interactions between the atmosphere and terrestrial ecosystems: influence on weather and climate. Global Change Biol. 4 (5), 461–475.

Potter, C.S., Wang, S.S., Nikolov, N.T., McGuire, A.D., Liu, J., King, A.W., Kimball, J.S., Grant, R.F., Frolking, S.E., Clein, J.S., Chen, J.M., Amthor, J.S., 2001. Comparison of boreal ecosystem model sensitivity to variability in climate and forest site parameters. J. Geophys. Res. -Atmos. 106 (D24), 33671–33687.

Raich, J.W., Nadelhoffer, K.J., 1989. Belowground carbon allocation in forest ecosystems. Global Trends. Ecol. 70 (5), 1346–1354.

Raich, J.W., Rastetter, E.B., Melillo, J.M., Kicklighter, D.W., Steudler, P.A., Peterson, B.J., Grace, A.L., Moore III, B., Vorosmarty, C.J., 1991. Potential net primary productivity in South America: application of a global model. Ecol. Appl. 1 (4), 399–429.

Randall, D.A., Dazlich, D.A., Zhang, C., Denning, A.S., Sellers, P.J., Tucker, C.J., Bounoua, L., Los, S.O., Justice, C.O., Fung, I., 1996. A revised land surface parameterization (SiB2) for GCMs. 3. The greening of the Colorado State University general circulation model. J. Clim. 9 (4), 738–763.

Rayner, P., Scholze, M., Knorr, W., Kaminski, T., Giering, R., Widmann, H., 2005. Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system. Global Biogeochem. Cy. 19, GB2026, doi:10.1029/2004GB002254.

Romanowicz, R., Beven, K.J., Tawn, J.A., 1994. Evaluation of predictive uncertainty in nonlinear hydrological models using a Bayesian approach. In: Barnett, V., Turkman, K.F. (Eds.), Statistics for the Environment, vol. 2. John Wiley and Sons, Chichester, pp. 297–318.

Schulz, K., Beven, K.J., 2003. Data-supported robust parameterisations in land surface-atmosphere flux predictions: towards a top–down approach. Hydrol. Process. 17, 2259–2277.

Schulz, K., Beven, K.J., Huwe, B., 1999. Equifinality and the problem of robust calibration in Nitrogen budget simulations. Soil Sci. Soc. Am. J. 63, 1934–1941.

Schulz, K., Jarvis, A., Beven, K.J., Soegaard, H., 2001. The predictive uncertainty of land surface fluxes in response to increasing ambient carbon dioxide. J. Clim. 14, 2551–2562.

Sellers, P.J., 1985. Canopy reflectance, photosynthesis and transpiration. Int. J. Rem. Sens. 6 (8), 1335–1371.

Sellers, P.J., 1987. Canopy reflectance, photosynthesis and transpiration. II. The role of biophysics in the linearity of their independence. Remote Sens. Environ. 21, 143–183.

Sellers, P.J., Berry, J.A., Collatz, G.J., Field, C.B., Hall, F.G., 1992. Canopy reflectance, photosynthesis, and transpiration. III. A reanalysis using improved leaf models and a new canopy integration scheme. Remote Sens. Environ. 42, 187–216.

Sellers, P.J., Los, S.O., Tucker, C.J., Justice, C.O., Dazlich, D.A., Collatz, G.J., Randall, D.A., 1996a. A revised land surface parameterization (SiB2) for atmospheric GCMs. 2. The generation of global fields of terrestrial biophysical parameters from satellite data. J. Clim. 9 (4), 706–737.

Sellers, P.J., Randall, D.A., Collatz, G.J., Berry, J.A., Field, C.B., Dazlich, D.A., Zhang, C., Collelo, G.D., Bounoua, L., 1996b. A revised land surface parameterization (SiB2) for atmospheric GCMs. 1. Model formulation. J. Clim. 9 (4), 676–705.

Sokal, R.R., Rohlf, F.J., 1981. Biometry. W.H. Freeman and Company, New York, p. 859.

Spear, R.C., Hornberger, G.M., 1980. Eutrophication in peel inlet. 2. Identification of critical uncertainties via generalized sensitivity analysis. Water Res. 14 (1), 43–49.

Spear, R.C., Grieb, T.M., Shang, N., 1994. Parameter uncertainty and interaction in complex environmental models. Water Resour. Res. 30 (11), 3159–3169.

Staff, S.S., 1994. State Soil Geographic Database (STATSGO). Report, U.S. Department of Agriculture, National Conservation Service, Fort Worth.

Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D., Pak, E.W., Mahoney, R., Vermote, E., El Saleous, N., 2005. An extended AVHRR 8-km NDVI data set compatible with MODIS and SPOT vegetation NDVI data. Int. J. Rem. Sens. 26 (20), 4485–4498.

USDAFS, 2001. Chequamegon-Nicolet National Forest Area History. USDA Forest Service Chequamegon-Nicolet National Forest.

Wang, Y.P., Leuning, R., Cleugh, H.A., Coppin, P.A., 2001. Parameter estimation in surface exchange models using nonlinear inversion: how many parameters can we estimate and which measurements are most useful? Global Change Biol. 7, 495–510.

WiDNR, 1998. WISCLAND Land Cover (WLCGW930). Wisconsin Department of Natural Resources (WiDNR), Madison, Wisconsin.

Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, P., Bernhofer, C., Ceulemans, R., Dolman, H., Field, C., Grelle, A., Ibrom, A., Law, B.E., Kowalski, A., Meyers, T., Moncrieff, J., Monson, R., Oechel, W., Tenhunen, J., Valentini, R., Verma, S., 2002. Energy balance closure at FLUXNET sites. Agric. For. Meteorol. 113 (1–4), 223–243.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. J. Hydrol. 204 (1–4), 83–97.