Quantitative comparison of canopy conductance models using a Bayesian approach

S. Samanta,[1] M. K. Clayton,[2] D. S. Mackay,[3] E. L. Kruger,[4] and B. E. Ewers[5]

Received 14 December 2007; revised 16 June 2008; accepted 8 July 2008; published 23 September 2008.

[1] A quantitative model comparison methodology based on deviance information criterion, a Bayesian measure of the trade-off between model complexity and goodness of fit, is developed and demonstrated by comparing semiempirical transpiration models. This methodology accounts for parameter and prediction uncertainties associated with such models and facilitates objective selection of the simplest model, out of available alternatives, which does not significantly compromise the ability to accurately model observations. We use this methodology to compare various Jarvis canopy conductance model configurations, embedded within a larger transpiration model, against canopy transpiration measured by sap flux. The results indicate that descriptions of the dependence of stomatal conductance on vapor pressure deficit, photosynthetic radiation, and temperature, as well as the gradual variation in canopy conductance through the season are essential in the transpiration model. Use of soil moisture was moderately significant, but only when used with a hyperbolic vapor pressure deficit relationship. Subtle differences in model quality could be clearly associated with small structural changes through the use of this methodology. The results also indicate that increments in model complexity are not always accompanied by improvements in model quality and that such improvements are conditional on model structure. Possible application of this methodology to compare complex semiempirical models of natural systems in general is also discussed.

Citation: Samanta, S., M. K. Clayton, D. S. Mackay, E. L. Kruger, and B. E. Ewers (2008), Quantitative comparison of canopy conductance models using a Bayesian approach, *Water Resour. Res.*, *44*, W09431, doi:10.1029/2007WR006761.

1. Introduction

[2] Numerical models are often used to estimate or predict responses of complex Earth systems under various conditions, and to identify functional relationships between measurable quantities that describe these conditions and system response [*Wainwright and Mulligan*, 2004]. However, the interacting entities, processes, parameters, and boundary conditions for most Earth systems are not completely known [*Oreskes et al.*, 1994]. Moreover, such systems are usually too complex to specify completely within a mathematically and computationally tractable model simply by including functional descriptions of all the interacting entities. Therefore, in practice, realistic descriptions of a system are often substituted with simplified and approximate descriptions, where some of the parameters and mathematical functions in the model represent integrated effects of highly heterogeneous or nonlinear underlying details [*Beven*, 1995; *Grayson and Blöschl*, 2001]. While such semiempirical models (subsequently referred to simply as models in this paper) are not completely mechanistic, they are also not purely empirical, because the model development process is guided by fundamental theory, usually employing a combination of theoretical, heuristic, and empirical approaches [*Swartzman and Kaluzny*, 1987]. The process-based structure common in such models facilitates the integration of knowledge and information from diverse areas of research, which is essential for modeling complex Earth systems, and also makes such models potentially more useful for making predictions or testing process-based hypotheses compared to purely empirical models [*Box et al.*, 1978; *Beven*, 1989]. Hydrological components of Earth system models often use this semiempirical approach, especially at large spatial scales, because of the above advantages [e.g., *Sellers et al.*, 1986; *Band et al.*, 1991].

[3] Usually, the problem of modeling any particular complex system may be addressed by using several different semiempirical conceptualizations, each resulting in a different model structure for the same system. These model structures differ in terms of scientific hypotheses, assumptions regarding system functions, mathematical representation of interacting entities and processes, approximation and aggregation techniques, etc. Consequently, different models provide different estimates of system response, which may lead to different predictions and inferences regarding system functions [e.g., *Pan et al.*, 1998; *Cramer et al.*, 1999;

[1] Woods Hole Research Center, Falmouth, Massachusetts, USA.
[2] Departments of Statistics and Plant Pathology, University of Wisconsin, Madison, Madison, Wisconsin, USA.
[3] Department of Geography, State University of New York at Buffalo, Buffalo, New York, USA.
[4] Department of Forest Ecology and Management, University of Wisconsin, Madison, Madison, Wisconsin, USA.
[5] Department of Botany, University of Wyoming, Laramie, Wyoming, USA.

0043-1397/08/2007WR006761$09.00

Luckai and Larocque, 2002; *Adams et al.*, 2004]. The inability to identify a unique model structure (i.e., structural uncertainty) out of the various possibilities is often taken into account in model prediction and the estimate of prediction uncertainty by using multiple models in an ensemble in methodologies such as Bayesian Model Averaging (BMA) [e.g., *Neuman*, 2003; *Ye et al.*, 2004; *Raftery et al.*, 2005; *Ajami et al.*, 2007; *Vrugt and Robinson*, 2007]. However, an important objective of using semiempirical models in the analysis of complex Earth systems is to be able to make inferences about system processes. Such inferences may also be affected by structural uncertainty, but would not benefit from an ensemble technique when alternative model structures represent alternative process representations or conflicting hypotheses. In this study, we explore the possibility of addressing this problem through a methodology of quantitative model comparison based on observed data. This methodology uses Markov chain Monte Carlo (MCMC) simulations for parameter estimation and deviance information criterion (DIC) [*Spiegelhalter et al.*, 2002] as the model comparison metric. DIC is a statistical model comparison metric, which provides a balance between model complexity and its fit to observed data within a Bayesian framework [*Bayes*, 1763; *Gelman et al.*, 1995]. DIC has been widely used to compare statistical models in many different fields, including ecology and Earth sciences [e.g., *Cowles and Zimmerman*, 2003; *Cam et al.*, 2004; *Helser and Lai*, 2004; *Thogmartin et al.*, 2004; *Manda and Meyer*, 2005]. However, the models compared in the existing studies are stochastic, while the models compared here are originally formulated as deterministic models with multiple embedded components. Therefore, so that the effectiveness of this method might be better understood before it is applied in a larger and more complex setting, the scope of this analysis was restricted to the problem of modeling canopy conductance, which is well studied but not completely resolved in terms of structure [*Baldocchi et al.*, 1991; *Leuning*, 1995; *Monteith*, 1995] and parameterization [*Dekker et al.*, 2001; *Ewers et al.*, 2001; *Komatsu et al.*, 2007]. Specifically, we compare multiple canopy conductance models with respect to their ability to model the rate of transpiration measured by sap flux. The compared models employ different configurations of the stomatal conductance model proposed by *Jarvis* [1976] (subsequently referred to as the Jarvis model), but employ the same transpiration model structure in other respects. The two primary objectives of this analysis are (1) to quantitatively evaluate various commonly used functional components of the Jarvis model against transpiration data and (2) to demonstrate and discuss the use of DIC for comparing complex semiempirical models.

2. Overview of Model Comparison and Selection Approaches: Rationale for Choosing DIC

[4] Two different but complementary approaches are generally used for model comparison. The first approach is to determine the level of compatibility between the available knowledge of the system and the mathematical structures of different models through a detailed theoretical analysis. The second approach is to fit different models to observed data and compare their performances quantitatively. The two approaches provide different kinds of information, both useful for the subsequent step of model selection.

[5] In the first approach, the mathematical and algorithmic structures of the models are analyzed in detail to identify differences in terms of representation of entities (e.g., levels of detail, aggregation) and processes (e.g., mathematical descriptions of functional dependencies). These differences are then evaluated with respect to theoretical knowledge of the system to determine whether one model should be considered a more accurate representation of the system compared to the others. Sensitivity of system responses to the identified differences may also be evaluated by comparing outputs from different models. Examples of this approach are widely available in the literature [e.g., *Perruchoud and Fischlin*, 1995; *Tiktak and van Grinsven*, 1995; *Pan et al.*, 1998; *Cramer et al.*, 1999; *Adams et al.*, 2004]. However, the success of this approach depends largely on experience and knowledge of the system and the models. Such a detailed structural analysis is often impractical because of incomplete process knowledge at the modeled scale, complexity of the models, and undocumented programming details [*Perruchoud and Fischlin*, 1995; *Tiktak and van Grinsven*, 1995; *Adams et al.*, 2004]. Therefore, for most model comparison and selection problems involving complex models, there is a need to augment the above approach with the more straightforward second approach, where a suitable metric is used to compare the performance of models against observed data [*Kros and Warfvinge*, 1995].

[6] The convenience of this second approach is that the models may be treated as black boxes for model selection purposes. The comparison metric provides a measure of how well the observed data support various model configurations, and therefore, also provides an indirect evaluation of different underlying hypotheses where available theoretical knowledge is not sufficient to conclusively determine the superiority of any particular model through structural analysis alone. In this manner, the two approaches may be combined within an objective development procedure for complex process-based models. However, the success of the second approach depends critically on the choice of the model comparison metric. Often the comparison metric is purely a quantitative measure of the goodness of fit, such as the coefficient of determination, R^2, calculated through linear regression between model output and observed data [e.g., *Katul et al.*, 2000; *van Wijk et al.*, 2000; *Moriana et al.*, 2002; *Misson et al.*, 2004]. However, like the transpiration models used here, semiempirical models are often nonlinear, and all of the models included in a comparison may not use the same predictor variables, therefore, the most appropriate way to calculate R^2 uniformly for all the models is not always clear [*Healy*, 1984; *Kvålseth*, 1985; *Anderson-Sprecher*, 1994; *Mitchell*, 1997]. Moreover, a goodness of fit measure is affected by errors from several sources other than inaccuracies in model structure, such as inappropriate parameter values, measurement errors, and natural variability in the data. Major differences in the goodness of fit for different models are often due to differences in the accuracy of parameter values among models [*van Grinsven et al.*, 1995]. In order to minimize parameterization errors, optimized goodness of fit values are often used for comparing models, where the parameter values for

each model have been calibrated to available observations, instead of using the same value for a conceptually related parameter in all models.

[7] The difficulty of using the optimized goodness of fit values for model comparison is that these values might overestimate model performance. One of the reasons for the overestimation is that the calibrated parameter values can become configured to compensate for structural inadequacies in the model or to explain part of the noise (e.g., measurement errors, natural variability) present in the calibration data [*Akaike*, 1974; *Forster*, 2000; *Gaganis and Smith*, 2001], a condition also known as ''overfitting.'' A complex model with a larger number of calibrated parameters is generally more susceptible to overfitting, compared to a simpler one, because the additional parameters provide more ways of adjusting the model output to match the data. Consequently, model selection guided solely by goodness of fit is expected to favor complex models over simple ones [*Akaike*, 1974]. Another reason for the overestimation is that the optimized goodness of fit does not explicitly account for the uncertainties in calibrated parameter values, as it is calculated at the optimized parameter values [*Marshall et al.*, 2005]. Many recent studies show that the uncertainties in calibrated parameter values are considerable for semiempirical hydrologic models [e.g., *Beven and Binley*, 1992; *Gupta et al.*, 1998; *Kuczera and Parent*, 1998; *Krzysztofowicz*, 1999; *Bates and Campbell*, 2001; *Thiemann et al.*, 2001; *Samanta and Mackay*, 2003; *Vrugt et al.*, 2003b; *Montanari and Brath*, 2004; *Samanta et al.*, 2007], and therefore, indicate the need to account for its effect on model comparison.

[8] Because of the above reasons, a methodology adopted for comparative analysis models should provide a balance between model performance and model complexity, while accounting for the uncertainties, so that reliable inferences may be obtained. Therefore, selection of an appropriate comparison metric is a complex issue that requires careful consideration of characteristics of the models being compared, goal of the comparison, as well as its consistency with the method used for parameter estimation [*Ward*, 2008]. Several methodologies have recently been proposed for comparing semiempirical hydrologic models, some of which are based on classical statistics [e.g., *Foglia et al.*, 2007], others on Bayesian statistics [e.g., *van der Perk*, 1997; *Marshall et al.*, 2005; *Vrugt et al.*, 2003a]. These methodologies use diverse techniques, e.g., cross validation and MCMC, and metrics, e.g., Akaike Information Criterion (AIC) [*Akaike*, 1974], Bayes Information Criterion (BIC) [*Schwarz*, 1978], and Bayes factor [*Kass and Raftery*, 1995]. For this study, we adopt a Bayesian modeling framework based on MCMC (Section 3.3). The primary motivations for adopting a Bayesian approach, over other important approaches used for uncertainty estimation in hydrologic modeling, such as, the Generalized Likelihood Uncertainty Estimation procedure (GLUE) [*Beven and Binley*, 1992], and the Pareto optimality approach [*Gupta et al.*, 1998], are explained in detail by *Samanta et al.* [2007], the most important reasons being probabilistic interpretation of the uncertainties, and availability of powerful computational techniques that makes it possible to analyze very complex models.

[9] For this analysis, MCMC simulation technique, which is a Bayesian method, was used for parameter estimation. Therefore, commonly used model comparison metrics that use maximum likelihood estimates for parameters, e.g., AIC and BIC, were not considered to be appropriate, because the maximum likelihood values necessary for calculating such metrics may not be correctly and consistently identified for all models through MCMC. Bayes factor represents an attractive choice for comparing models within a Bayesian framework [*Marshall et al.*, 2005; *Ward*, 2008]. However, the use of Bayes factor requires that all of the candidate models be specified at the start of the comparison with their prior probabilities assigned, which is difficult because of the numerous modeling possibilities available for any complex natural system [*Marshall et al.*, 2005]. This is not necessary for using DIC, and therefore, was not done for this analysis. DIC offers an attractive alternative because it is not only consistent with a Bayesian approach, but also simple to calculate along with the usual MCMC steps without increasing the computational burden significantly. Unlike Bayes factor, which assumes that a true model exists and it is included in the compared model set [*Burnham and Anderson*, 2002], DIC only requires that the models be reasonable approximations of the true model. These transpiration models are expected to be reasonable for these data on the basis of the results obtained by *Samanta et al.* [2007], but none of them may be considered as true models because of inherent simplifications and approximations. DIC was selected for comparing the models in this analysis on the basis of the above considerations. However, properties of model comparison metrics and their applicability in various situations are subjects of active research, and therefore, our choice is not intended to imply that DIC is the only metric that is suitable for comparing complex semiempirical models. A brief description of DIC and the methodology for its computation are described in Sections 3.1 and 3.3, respectively.

3. Methodology

3.1. Deviance Information Criterion

[10] DIC was developed as a generalization of AIC within a Bayesian framework for comparing models of arbitrary complexity, all of which need not be structurally related [*Spiegelhalter et al.*, 2002]. We briefly describe DIC below, focusing on the aspects necessary for its application in this analysis, while referring to the work by *Spiegelhalter et al.* [2002] for its derivation and various statistical properties. DIC is based on a deviance statistic, $D(\theta)$, which is related to the residual information in the data [*Kullback and Leibler*, 1951], and defined as:

$$D(\theta) = -2\log[p(Y|\theta)] + 2\log[f(Y)], \quad (1)$$

where $\boldsymbol{Y}$ is the observed data, θ is the parameter vector for the model, $p(Y|\theta)$ is the likelihood function defined by the model, and $f(\boldsymbol{Y})$ is a standardizing term. The value of $f(\boldsymbol{Y})$ is the likelihood of the saturated model, which may be calculated by defining a saturated model for the data as shown through examples by *Spiegelhalter et al.* [2002]. Because the value of $f(\boldsymbol{Y})$ depends only on $\boldsymbol{Y}$, it remains the same for all the compared models. Therefore, the $f(\boldsymbol{Y})$ terms

cancel out when calculating the DIC difference between two models, and therefore, its value does not influence the model comparison, which is based only on differences in DIC. For computational simplicity, we assume $f(\boldsymbol{Y})$ to be equal to one for this analysis as recommended by *Spiegelhalter et al.* [2002]. DIC accounts for the posterior distribution of $D(\theta)$ by using the posterior mean deviance, $\overline{D(\theta)}$, as the Bayesian measure of fit or "adequacy" of the model, instead of using the optimized measure of fit. *Spiegelhalter et al.* [2002] provide an information theoretic argument in favor of using the difference between the posterior mean deviance and the deviance at the posterior parameter estimates as an estimate of the reduction in uncertainty due to parameter estimation, i.e., the degree of overfitting, in a Bayesian context. This quantity, termed p_D, is considered in DIC to be the measure of model complexity or the effective number of parameters. *Spiegelhalter et al.* [2002] also analytically investigate the formal properties of p_D for various statistical models with useful examples. However, its algebraic forms are not necessary for its use, as p_D can be obtained directly using samples from an MCMC sequence (see Section 3.3 for details) by using the equation:

$$p_D = \overline{D(\theta)} - D\left(\tilde{\theta}\right), \tag{2}$$

where $\tilde{\theta}$ is the posterior estimate of θ (e.g., posterior mean, median, or mode), and $D(\tilde{\theta})$ is the deviance at $\tilde{\theta}$. From the above definitions of fit and complexity, $\overline{D(\theta)}$ is penalized by p_D to define DIC as:

$$DIC = \overline{D(\theta)} + p_D = 2\overline{D(\theta)} - D\left(\tilde{\theta}\right). \tag{3}$$

[11] For model selection, DIC is minimized across candidate models, as a lower DIC value identifies a better model. However, the significance of a specific magnitude of DIC difference is difficult to determine analytically [*Spiegelhalter et al.*, 2002]. For the purposes of the present analysis, we consider that a DIC difference of less than three is not significant, between three and seven is moderately significant, and greater than seven is highly significant. These significance levels are based on the levels of empirical support in AIC differences recommended by *Burnham and Anderson* [2002]. These are also recommended by *Spiegelhalter et al.* [2002] for use with DIC on the basis of their experience and conceptual similarities between AIC and DIC.

3.2. Transpiration Model and Canopy Conductance Submodel Configurations

[12] The transpiration models used in this study have identical structures in all respects other than the use of different canopy conductance submodels, thereby providing a stable model structural context within which the canopy conductance submodels could be evaluated. The transpiration models are based on the Penman-Monteith equation [*Monteith*, 1965], which uses canopy conductance as a parameter, and calculates the rate of transpiration per unit ground area on the basis of energy balance and mass transfer. The basic structure of the transpiration model used here is described in more detail by *Samanta et al.* [2007]. Each transpiration model uses a different submodel embedded within it for calculating the canopy conductance, which differ in terms of structure, parameters, and variables, as described below in more detail.

[13] The canopy conductance submodels use various configurations of the Jarvis model to calculate stomatal conductance. The leaf area index, necessary to convert stomatal conductance to canopy conductance, is either held constant over time or adjusted from day to day using a semiempirical function described later in this section. The Jarvis model provides a general functional form describing the relationship between canopy surface conductance per unit leaf area and measurable environmental variables (e.g., vapor pressure deficit, incident photosynthetically active radiation) that affect this conductance. The effect of any particular environmental variable on the conductance is taken to be independent of the others, which provides a simple yet flexible method for modeling the response of canopy conductance to environmental conditions. The stomatal conductance per unit leaf area, g_S, is calculated as a fraction of the highest possible conductance for fully developed leaves per unit leaf area, $g_{S\max}$, with the fraction representing the overall constraining effect of environmental conditions on $g_{S\max}$. Under the assumption that the effect of each environmental variable on stomatal conductance is independent of the others, the effect of an environmental variable, z_j, can be described as a function, $f_j(z_j)$, of z_j alone. The functions, $f_j(z_j)$, also called constraint functions, are defined in such a way that $f_j(z_j)$ assumes a dimensionless value between zero and one for all reasonable z_j values. Therefore, the combined effect of m different environmental variables on g_S is expressed in the general form of the Jarvis model as:

$$g_S = g_{S\max} \prod_{j=1}^{m} f_j\left(z_j\right), \tag{4}$$

with its specific configuration depending on the set of environmental variables used and the mathematical formulations of the corresponding constraint functions.

[14] The Jarvis model configurations compared in this study use constraint functions for vapor pressure deficit within the canopy, D_c (kPa), average photosynthetic photon flux density, Q_p (μmol m^{-2} s^{-1}), air temperature within the canopy, T_c (°C), and soil water potential, Ψ_s (MPa), in various combinations to calculate g_S by scaling $g_{S\max}$, which is a calibrated parameter for all the models used here. The constraint functions and the associated parameters are shown in Table 1, along with the range of values for each variable in the data used. Two alternative forms of constraint functions are used to describe the effect of D_c on g_S, one is a linear function [*Jarvis*, 1976] and the other is a hyperbolic function [*Lohammar et al.*, 1980], with sensitivity parameters δ and δ_h, respectively. The constraint functions for Q_p, T_c, and Ψ_s are based on functions commonly used for such purposes [*Jarvis*, 1976; *Lhomme*, 2001; *Nishida et al.*, 2003]. The light response of g_S is determined by the value of the calibrated parameter A. The temperature response of g_S is determined by three calibrated parameters, T_{opt}, T_{lo}, and T_{hi}, representing the optimal, lowest, and highest temperatures for transpiration, respectively. The response of g_S to Ψ_s is determined by the calibrated parameter Ψ_0, which represents the minimum soil water potential necessary for transpiration.

Table 1. Environmental Variables, Vapor Pressure Deficit Within Canopy (D_c), Average Photosynthetic Photon Flux Density (Q_p), Air Temperature Within Canopy (T_c), and Soil Water Potential (Ψ_s)[a]

Variable	Range of Values	Constraint Function	Parameter
D_c (kPa)	0.60 to 2.45	$1 - \delta D_c$	δ (kPa^{-1})
D_c (kPa)	As above	$(1 + \delta_h D_c)^{-1}$	δ_h (kPa^{-1})
Q_p (μmol m^{-2} s^{-1})	202.78 to 1981.70	$\frac{Q_p}{(Q_p+A)}$	A (μmol m^{-2} s^{-1})
T_c (°C)	13.07 to 32.70	$\frac{(T_c-T_{lo})}{(T_{opt}-T_{lo})}\left[\frac{(T_{hi}-T_c)}{(T_{hi}-T_{opt})}\right]^{\frac{(T_{hi}-T_{opt})}{(T_{opt}-T_{lo})}}$	T_{opt}, T_{lo}, and T_{hi} (°C)
Ψ_s (MPa)	-0.0554 to -0.0029	$1 - \frac{\psi_S}{\psi_0}$	Ψ_0 (MPa)

[a]The environmental variables, vapor pressure deficit within canopy (D_c), average photosynthetic photon flux density (Q_p), air temperature within canopy (T_c), and soil water potential (Ψ_s), were used to describe the response of stomatal conductance to the environment with the Jarvis model constraint functions shown in Table 1. Different model configurations used different combinations of variables and constraint functions as described in the text. Note that only one of the two constraint functions for D_c can be used in one model. Two more parameters, g_{Smax}, the highest possible conductance for fully developed leaves per unit leaf area, and σ^2, the error variance, were also estimated from data for all models, in addition to the parameters shown in the fourth column.

[15] In all the models, the canopy is subdivided into two classes of leaves, one sunlit and the other shaded. The surface conductance for each class is calculated as the product of the corresponding values of g_S and single sided leaf area index for the class, L_{cdoy}. The two conductances are then combined in parallel to obtain the total surface vapor conductance of the canopy. The applicable values of Q_p used for calculating g_S for the two classes of leaves are modeled from observed incident radiation using the approach described by *Spitters et al.* [1986]. L_{cdoy} values are calculated using the following equation, which assumes a spherical distribution of leaves within the canopy:

$$L_{cdoy} = \begin{cases} \frac{1-e^{-K_b L_{doy}}}{K_b}, & for \quad sunlit \quad class \\ L_{doy} - \frac{1-e^{-K_b L_{doy}}}{K_b}, & for \quad shaded \quad class \end{cases}, \qquad (5)$$

where K_b is the canopy extinction coefficient for beam radiation for the current zenith angle, and L_{doy} is the current leaf area index for the entire canopy [*Campbell and Norman*, 1998]. In some of the models, L_{doy} is held constant at an observed single sided leaf area index, L_{obs}, and in the others, L_{doy} is adjusted on the basis of the ordinal day of year (DOY). This adjustment function is used as a simple semiempirical technique of accounting for the combined effect of leaf area dynamics and many other processes that affect the maximum canopy conductance gradually through the growing season (e.g., leaf aging, and changes in leaf, sapwood, and root areas). As identified through a detailed analysis of a transpiration model of the type used here, such a function is necessary to account for the gradual change in transpiration through the season, which is not otherwise accounted for in the model [*Samanta et al.*, 2007]. The usefulness of including this function is further tested through model comparison in the present analysis. The L_{doy} adjustment function is based on fitted transpiration values, lf_{doy}, obtained by fitting a second-degree polynomial function of DOY to a large set of observed transpiration data described in Section 4. The adjusted value of L_{doy} is calculated from lf_{doy} values by:

$$L_{doy} = L_{obs}\left[1 - lf_{scl}\left(1 - \frac{lf_{doy}}{lf_{\max}}\right)\right], \qquad (6)$$

where $lf_{\max}$ is the maximum value of lf_{doy} during the entire period, and lf_{scl} is a leaf area scale factor estimated from data, which determines the rate of change of L_{doy} with respect to lf_{doy}.

[16] Each canopy conductance submodel used in this analysis represent a specific combination of the above functions, and therefore, the comparison might also be considered as indicative of the relative importance of including each environmental variable in the Jarvis model using a specific functional form, within the context of the overall transpiration model and the data used in the analysis. For ease of reference, the transpiration models are designated by strings of letters that indicate the functions present in their canopy conductance submodels, using the letters L, H, Q, T, W, and P. The letter L indicates the linear form of the constraint function for D_c is in use, while the letter H indicates the use of the hyperbolic form. The letters Q, T, and W indicate the uses of constraint functions for Q_p, T_c, and Ψ_s, respectively. The letter P indicates the use of adjusted L_{doy}. When a particular function is not used, the corresponding letter is dropped from the designation. This comparison includes the two most complex models possible within the above family of models, viz., LQTWP and HQTWP, and nine other models derived from these two by removing one constraint function at a time. Therefore, the contribution of each of the above functions could be evaluated by comparing specific model pairs. The model LQP, which was used previously for a detailed analysis of the same data [*Samanta et al.*, 2007], and its hyperbolic analog, HQP, are also included in this comparison as convenient reference models.

3.3. MCMC Simulation, Estimation of Deviance, and DIC Calculation

[17] The transpiration models described in the previous section are deterministic, where each model uses an input sequence of environmental measurements, $\boldsymbol{x}$, and parameter vector, β, to calculate an output sequence of canopy transpiration rates per unit ground area, $\boldsymbol{E}_{canopy}(\boldsymbol{x}, \beta)$. The elements of β are parameters directly associated with the canopy conductance submodel used, which are estimated from a sequence of measured half-hourly canopy transpiration rate, $\boldsymbol{E}$, through MCMC simulations. Other parameters

associated with the transpiration model were held constant at reasonable values described by *Samanta et al.* [2007], and therefore, could be regarded as integral parts of the invariant component of the transpiration model for the purposes of this analysis. The sequences $\boldsymbol{x}$ and $\boldsymbol{E}$ have equal number of elements, n, and the corresponding elements, $\boldsymbol{x}_i$ and E_i, respectively, are temporally synchronized with subscript i (1 to n) in the temporal order of acquisition. These data are described in more detail in Section 4.

[18] A detailed analysis the model LQP [*Samanta et al.*, 2007] shows that the residuals for models of the type used in this study are approximately normal, without signs of heteroscedasticity, but not strictly independent. However, for semiempirical model building, we prefer to address this lack of independence in the residuals by progressively refining the models through the use of semiempirical components. Within this context, the normal error model can serve as a consistent basis for evaluating such models under fairly general conditions [*Kuczera*, 1983]. Although the use of a different error model would be warranted in case a gross violation of the normal error assumption is noticed or required because of process-based reasons. On the basis of the evidence and the principle described above, the likelihood functions necessary for a Bayesian analysis of the deterministic transpiration models are formulated here with the assumption that the errors are independent and normally distributed with a constant but unknown variance, σ^2. The generic form of the likelihood function in this case is expressed by:

$$p(E|\beta,\sigma^2) \propto \sigma^{-n}\prod_{i=1}^{n}\exp\left\{-\frac{1}{2\sigma^2}\left[E_i - E_{canopy}(x_i,\beta)\right]^2\right\}, \quad (7)$$

where $\boldsymbol{E}_{canopy}(\boldsymbol{x}_i, \beta)$ is the modeled transpiration rate corresponding to E_i, calculated by the deterministic transpiration model with input $\boldsymbol{x}_i$, and parameter β. Substituting $p(Y|\theta)$ in equation 1 with $p(E|\beta, \sigma^2)$ above, $D(\theta)$ for these models may be expressed as:

$$\begin{aligned} D(\theta) &= -2\log\left[p(E|\beta,\sigma^2)\right] \\ &= 2n\log(\sigma) + \sum_{i=1}^{n}\frac{\left[E_i - E_{canopy}(x_i,\beta)\right]^2}{\sigma^2}, \end{aligned} \quad (8)$$

assuming the value of $f(\boldsymbol{Y})$ to be one (i.e., 2 log $[f(Y)] = 0$) for model comparison purposes [*Spiegelhalter et al.*, 2002]. The joint posterior distributions for the models are derived from the likelihood function (equation 7) using noninformative prior distributions of the form:

$$p(\beta,\sigma^2) \propto \frac{1}{\sigma^2}, \quad (9)$$

which assumes that β is distributed uniformly within a specified multidimensional interval, and $\log(\sigma)$ is distributed uniformly [*Box and Tiao*, 1973; *Gelman et al.*, 1995]. The prior intervals for the parameters are generally dictated by the requirement that g_S should be nonnegative for all reasonable input values [*Samanta et al.*, 2007]. Applying Bayes' rule with the above forms of prior distributions and likelihood functions, the joint posterior distributions for the models are of the form:

$$p(\beta,\sigma^2|E) \propto \sigma^{-(n+2)}\prod_{i=1}^{n}\exp\left\{-\frac{1}{2\sigma^2}\left[E_i - E_{canopy}(x_i,\beta)\right]^2\right\}. \quad (10)$$

A Markov chain Monte Carlo (MCMC) simulation method based on the Metropolis algorithm [*Metropolis and Ulam*, 1949] was used, with the additional calculation of $D(\theta)$ at each iterative step, to simulate parameter samples from the joint posterior distribution. Detailed description of the algorithm may be found in texts on Bayesian statistics [e.g., *Gelman et al.*, 1995], as well as in the hydrologic literature [e.g., *Kuczera and Parent*, 1998; *Vrugt et al.*, 2003b; *Samanta et al.*, 2007]. The methodology is briefly described below, with an emphasis on the details specific to the present application.

[19] In MCMC simulation, draws from the joint posterior distribution are iteratively simulated in the form of a chain or sequence. First a candidate parameter value is randomly generated from a distribution called the proposal distribution or the candidate-generating density. Next, the decision to accept or reject the candidate parameter value as a sample in the sequence is made on the basis of the ratio of posterior density at the candidate parameter value to that at the current parameter value. This process is carried out long enough for the chain to converge to a stationary distribution and provide a reasonable number, typically hundreds or even thousands, of samples from the posterior distribution after the chain has converged.

[20] The multivariate normal proposal distributions used in this study used an adaptive approach, similar to that described by *Gelman et al.* [1995], in order to achieve reasonable acceptance rates, defined as the proportion of generated candidate parameter values that are accepted. The variance matrices of the proposal distributions were initialized on the basis of samples from preliminary simulations. During the first half of an MCMC run, the variance matrix was updated on the basis of recent iterations, if the acceptance rate since the last update was not between 0.2 and 0.3. The variance matrix was not adjusted further during the second half. Chain convergence was assessed quantitatively by using the potential scale reduction factor, $\sqrt{\hat{R}}$ [*Gelman and Rubin*, 1992; *Gelman et al.*, 1995], and visually by plotting traces of sampled parameter values against iterations [*Kass et al.*, 1998]. In order to use $\sqrt{\hat{R}}$ as a convergence diagnostic, four independent MCMC sequences were run for each model. Each sequence started at different randomly generated values for β and σ. The adopted technique of running each chain for 200,000 iterations, and discarding the first 100,000 iterations as the "burn-in" period, resulted in $\sqrt{\hat{R}}$ values of 1.003 or less for all models. These values are considerably lower than the value of 1.2 suggested by *Gelman et al.* [1995] as evidence of convergence. The final 100,000 iterations, constituting the second half of each sequence, were subsampled systematically at a gap of 50 iterations to obtain the posterior samples of β, σ^2, and $D(\theta)$, used in this analysis. The subsampling is not strictly necessary, as discussed by *Geyer* [1992], but was done here in order to reduce autocorrelation among the samples [*Samanta et al.*, 2007].

[21] For DIC calculations, $\overline{D(\theta)}$ was calculated by taking the average of the posterior samples of $D(\theta)$, calculated at each iteration during the MCMC simulations. For each model, $\overline{D(\theta)}$ values from individual chains corresponding to that model were compared to confirm that they do not differ by more than 0.5, so that the model comparison is not greatly affected by random Monte Carlo sampling errors [*Spiegelhalter et al.*, 2002]. The posterior parameter estimate, $\tilde{\theta}$, required for calculating $D(\tilde{\theta})$ and p_D, may be obtained using a standard estimator, such as the mean, median, or mode of the posterior samples of β and σ^2. However, as discussed by *Spiegelhalter et al.* [2002], use of different estimators may sometimes lead to different p_D values for the same model, and a negative p_D value either indicates that the estimator used is not appropriate or that there is a conflict between the prior and the data. *Spiegelhalter et al.* [2002] also show that, in some cases, use of median as the estimator, in comparison to the mean, may lead to a closer approximation to an invariant p_D value, which is less influenced by the prior data conflict. Although the posterior mean is the most commonly used estimator, it did not result in positive p_D values for all of the models in this case, possibly because of the high asymmetry in the posterior distributions of certain parameters and the strong interdependence between certain parameters (see Section 5). However, the use of posterior median resulted in positive p_D values for all the models. Therefore, to ensure consistency across models, the posterior median values were used to calculate the DIC values reported here. For the remainder of this paper, $\tilde{\theta}$ and $\tilde{\beta}$ refer to parameter vectors consisting of the medians of posterior samples. The values of $D(\tilde{\theta})$ for each model are calculated through equation 8 by setting θ equal to the corresponding $\tilde{\theta}$ value.

4. Data

[22] The data used for this study were collected as part of the Chequamegon Ecosystem-Atmosphere Study (ChEAS) [*Bakwin et al.*, 1998; *Davis et al.*, 2003], a collaborative research effort that maintains multiple data collection sites located in and around the Chequamegon-Nicolet National Forest in Northern Wisconsin, USA. Nearly all of the data, with the exceptions noted below, are from two of the above sites, viz., the Willow Creek site [*Cook et al.*, 2004; *Desai et al.*, 2005] and the Hay Creek site [*Ewers et al.*, 2002, 2007a, 2007b]. These two sites are a little over 21 km apart and have similar sandy loam soils. The forests at both sites consist of upland hardwoods dominated by sugar maple (*Acer saccharum* Marsh.) and basswood (*Tilia americana* L.).

[23] The modeled transpiration rates (mm s^{-1}) are average half-hourly values for eight sugar maple trees at the Hay Creek site, obtained between 9:00 A.M. to 6:00 P.M. Central Standard Time (CST) from 5 May to 19 September 2001, DOY 125 and 262, respectively. The transpiration values were calculated from sap flux and sapwood area per unit ground area measurements using methodologies described by *Oren et al.* [1998] and *Ewers et al.* [2002]. The value of L_{obs} is 4.6, calculated from litter-fall data collected in 2001 [*Ewers et al.*, 2007a]. Each element of the model input vector, $\boldsymbol{x}_i$, consists of half-hourly measurements of incident photosynthetic photon flux density, Q_p (μmol m^{-2} s^{-1}), wind speed above canopy (m s^{-1}), air temperature above canopy (°C), vapor pressure deficit above canopy (kPa), atmospheric pressure (kPa), ground heat flux at 7.5 cm soil depth (W m^{-2}), volumetric soil moisture content 10 cm below soil surface, ω, midcanopy air temperature, T_c (°C), and midcanopy vapor pressure deficit, D_c (kPa). The average canopy height of the sugar maple trees at this site was 18.6 m. The above-canopy measurements were made at 29.6 m elevation and the midcanopy measurements were made at two thirds of canopy height. Out of the above input variables, D_c and T_c were measured at the Hay Creek site, and the others were measured at the Willow Creek site. About 10% of the final input data were obtained from measurements made at the WLEF TV eddy flux tower in Chequamegon National Forest, less than 10 km away from the measurement sites, so that large gaps in the input data, possibly due to equipment malfunctions in the field, may be avoided. The measured ω values were used to calculate Ψ_s using the following equation:

$$\psi_s = a\omega^{-b}, \tag{11}$$

where a and b are experimentally determined parameters, which were fixed at the values of -3.6189×10^{-5} MPa and 4.9, respectively, for the sandy loam soils at the sites [*Clapp and Hornberger*, 1978]. Available volumetric soil moisture measurements at depths other than 10 cm (viz., 5 cm, 20 cm, 50 cm, and 1 m) show strong linear correlation to ω, and therefore, the use of additional measurements or the average is not expected to significantly affect the fit of the model to transpiration data and alter the inference of this model comparison. However, the assumption that the soil moisture at Hay Creek is the same as that at Willow Creek introduces additional uncertainty in this model input, which could not be avoided because of nonavailability of soil moisture data at the Hay Creek site.

[24] The sequence of available transpiration data, consisting of 2579 half-hourly values, is shown in Figure 1 (both filled and open circles). This entire sequence was used to obtain the lf_{doy} values through local polynomial regression (loess) fit (the superimposed line in Figure 1) following the technique described by *Samanta et al.* [2007]. However, input data were available only for 1899 transpiration values in the above sequence. Out of these 1899 measurements, 708 data points corresponding to above-canopy vapor pressure deficit less than 0.6 kPa were discarded because of the potential for large errors in the transpiration estimates [*Ewers and Oren*, 2000]. Moreover, the process of evaporation of intercepted precipitation or dew at the leaf surface was not incorporated in the model for the sake of simplicity. Therefore, 182 additional data points, coincident with precipitation or temperature inversions, were removed from the data, as the omitted evaporation process could be important at these points. Finally, an additional 37 data points corresponding to very low incident photosynthetic photon flux density, less than 200 μmol m^{-2} s^{-1}, were removed from the data. The data eliminated by this last condition are measurements made on a few of the observed days after 4:00 P.M. The number of elements, n, remaining in the sequences $\boldsymbol{E}$ and $\boldsymbol{x}$ is 972 after the above elimination process (the filled circles in Figure 1). The transpiration values remaining in $\boldsymbol{E}$ range between 3.973×10^{-7} and 5.971×10^{-5} mm s^{-1}. The ranges of values for the

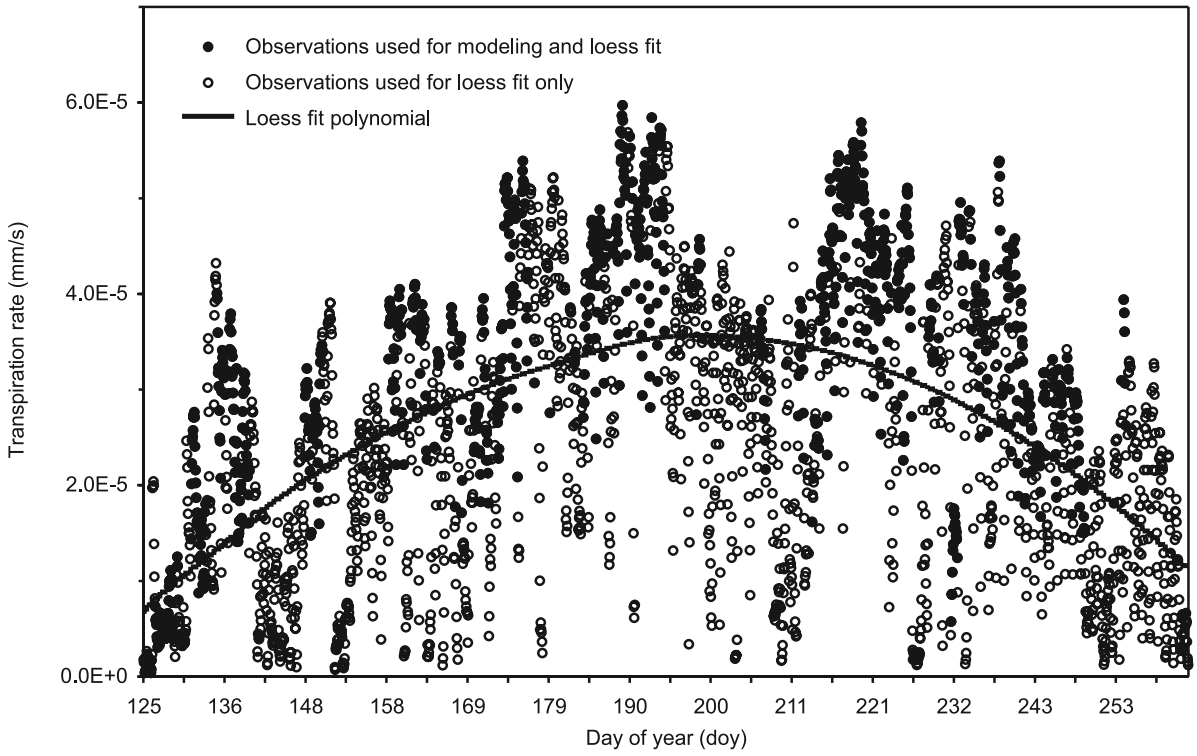

Figure 1. Plot of half-hourly transpiration rates measured by sap flux, superimposed with the local polynomial regression (loess) fit of transpiration rates to the ordinal DOY described in the text.

variables used in the constraint function are shown in Table 1. Among these variables, the values for D_c, Q_p, and T_c, cover relatively wide ranges. However, the minimum value of Ψ_s in these data, -0.0554 MPa, is considerably higher than the wilting point (approximately -1.5 MPa), indicating that moisture stress was unlikely to have been encountered during the data collection period. Note that the measurement errors associated with $\boldsymbol{x}$ are not explicitly modeled in this analysis. However, as the measurements were conducted over a relatively small area, we expect them to be small compared to the prediction errors, and therefore, neglect them to avoid additional analytical complexities.

5. Results

[25] The model comparison results are summarized in Table 2, showing DIC and p_D values for all the models, listed in the order of increasing DIC. The model HQTWP has the best DIC value of -21783.1, which is 4.1 lower than that of HQTP, the model with the next lowest DIC value. This DIC difference is only moderately significant according to the criterion used, and therefore, the evidence of superiority of HQTWP over HQTP is not overwhelming in these data. However, most other models are well separated from HQTWP in terms of DIC, enabling clear-cut inferences regarding the effects of individual canopy conductance model functions on the quality of the transpiration model.

[26] The degree to which DIC is affected by each of the functional modifications (Section 3.2) changes significantly from function to function, implying that these functional components significantly differ in terms of their contributions toward modeling these data. The impact of P (temporal trend in effective leaf area index) on DIC is the largest. When the model HQTWP is simplified to HQTW, and LQTWP is simplified to LQTW, DIC increases by 516.7, and 533.0, respectively. The second largest degradation, a DIC increment of 357.8, occurs when H, the hyperbolic dependence of g_S on D_c, is removed from the model

Table 2. Summary of Model Comparison Results[a]

Model Designation[b]	p_D	DIC	DIC Change With Respect to HQTWP
HQTWP	5.6	−21783.1	-
HQTP	5.6	−21779.0	4.1
HQP	5.0	−21710.8	72.3
HQWP	5.0	−21710.1	73.0
LQTP	5.9	−21699.0	84.1
LQTWP	6.2	−21698.4	84.7
HTWP	4.6	−21685.4	97.7
LQP	5.1	−21673.2	109.9
LQWP	5.3	−21671.6	111.5
LTWP	4.9	−21593.9	189.2
QTWP	5.6	−21425.3	357.8
HQTW	0.5	−21266.4	516.7
LQTW	3.5	−21165.4	617.7

[a]Models are listed in the order of descending acceptability according to deviance information criterion (DIC), the model comparison metric. The effective number of parameters, p_D, and DIC increment with respect to HQTWP for each model are shown in the second and fourth columns, respectively.

[b]Indicates functions included in the model with L for linear constraint for vapor pressure deficit within the canopy (D_c), H for hyperbolic constraint for D_c, Q for constraint for average photosynthetic photon flux density (Q_p), T for constraint for air temperature within the canopy (T_c), W for constraint for soil water potential (Ψ_s), and P for day-to-day leaf area adjustment.

HQTWP, in contrast, the corresponding DIC increment due to the removal of the linear dependence, L, from LQTWP is only 273.1. Therefore, in terms of DIC, H is a significantly more useful function than L. Removing Q (the relationship between photosynthetic radiation and canopy conductance) increases DIC by 104.5 (with H in the model) or 97.7 (with L in the model). The contributions from T (the relationship between temperature and canopy conductance) are also highly significant, but in all cases lower than the contributions from L, H, Q, and P. The biggest contribution from W (the relationship between soil water potential and canopy conductance) is only moderately significant (a DIC reduction of 4.1) and occurs with the model HQTWP.

[27] From the results, the degree to which a particular functional modification affects DIC appears to be conditional on the overall structure of the model. For example, the removal of W from LQTWP decreases DIC by 0.6, a very insignificant amount, while the same modification to HQTWP reduces DIC by 4.1. Therefore, the inference regarding W is sensitive to the form of the dependence of g_S on D_c used in the model. Although the inferences regarding the other functions do not change here depending on model structure, the DIC differences are highly significant for other functions also. For example, the removal of T increases DIC by 73.0 or by 26.8, respectively, depending on whether the hyperbolic or the linear relationship between g_S and D_c is used. DIC changes associated with Q also show significant differences depending on the model used. The functions H, T, and W also appear to reinforce the contributions of each other. Each of these changes made individually to LQP results in DIC changes of −37.6, −25.8, and +1.6, respectively, while making all of them together results in a change of −109.9, a significantly larger gain than the sum of the individual changes. Possible reasons and implications of this dependence on model development are discussed further in Section 6.

[28] Except for the models LQP and HQP, the effective number of parameters (Table 2), p_D, is less than the number of estimated parameters, which is often used as the measure of model complexity [*Akaike*, 1974]. The two-dimensional contour plots for the posterior distribution of HTQP parameters (Figure 2) show considerable interdependence among parameters. In general, similar patterns of parameter interdependence are noticeable for all the models, with small variations from model to model. The parameter interdependence common in these models might be the reason behind the reduction in effective dimensionality [*Spiegelhalter et al.*, 2002]. Parameter independence in semiempirical models may be difficult to achieve through parameter transformations without prior knowledge about such dependencies [*Raftery et al.*, 1995]. Regardless, the results indicate the importance of accounting for these dependencies in the inferences, particularly where the $\overline{D(\theta)}$ difference between models is small. The use of p_D provides a simple method of doing so.

[29] As can be seen in the posterior histograms for models HQTWP, LQTWP, and HQWP (Figures 3, 4, and 5, respectively), many model parameters have nearly symmetric posterior distributions. However, T_{opt}, T_{lo}, T_{hi}, and Ψ_0 have highly asymmetric posterior distributions, indicating that the posterior mean may not be a good estimator for these parameters. This asymmetry in the posterior distributions is likely to be responsible for the negative p_D values for a few models using these parameters, when they are calculated using the posterior mean. For our analysis, use of the posterior median values led to reasonably consistent p_D values for the compared models, with the possible exceptions of HQTW and LQTW (Table 2). However, the apparent discrepancies are inconsequential for model selection in this case, because of the extremely large differences between the DIC values for either of these models and those of the others. Considering these facts, DIC appears to provide a valuable tool for making inferences among these models.

[30] As can be seen from the examples of posterior parameter histograms (Figures 3, 4, and 5), and the examples of $\tilde{\beta}$ with the 95% posterior intervals (Table 3), considerable uncertainties are associated with the parameter estimates, particularly with those of T_{opt}, T_{lo}, and T_{hi}, and Ψ_0. Moreover, the posterior parameter distributions and the uncertainties appear to be sensitive to model structure. For example, the distribution for Ψ_0 is highly asymmetric for HQTWP, but less so for HQWP, and the distribution for $g_{S\max}$ for LQTWP shows less variability compared to that for HQTWP. This evidence of high parameter uncertainty, and its sensitivity to model structure, supports the need to account for parameter uncertainties in model comparison, as done here.

[31] A few general observations about the parameter estimates are as follows. The estimates of $g_{S\max}$ for different models show an increasing tendency with the number of included constraint functions, which is expected from the Jarvis model structure of multiplicative constraint functions. Uses of H are associated with high estimates of $g_{S\max}$ and T_{lo}. The models that include T have slightly lower lf_{scl} estimates than those without T, but the estimates of lf_{scl} are significantly different from zero, in fact, close to one for all models. Estimates of some of the parameters (e.g., $g_{S\max}$, δ, δ_h, T_{lo}) change appreciably from one model structure to another, while the estimates of others (e.g., lf_{scl}, A) are similar across models. Therefore, a conceptually equivalent parameter, such as $g_{S\max}$, is not numerically equivalent in all the models. Implications of the above results for model parameterization are discussed in Section 6.

6. Discussion of Results

[32] The results show that improvements in the quality of the transpiration model obtained by the inclusions of constraint functions with vapor pressure deficit (D_c), photosynthetic radiation (Q_p), and air temperature within the canopy (T_c) in the Jarvis model are highly significant. Of these three variables, D_c has the biggest impact on model performance and T_c the least. Evidence in favor of using the constraint function for soil water potential (Ψ_s) in these data is moderately significant at best. The relative importance of the environmental variables, determined here through the use of DIC, is consistent with widely held concepts regarding the use of environmental variables for modeling of stomatal conductance, on the basis of theoretical, experimental, and parameter sensitivity analyses [e.g., *Baldocchi et al.*, 1991; *Leuning*, 1995; *Monteith*, 1995; *Lhomme*, 2001], as well as data analyses from ChEAS sites [*Ewers et al.*, 2002; *Cook et al.*, 2004; *Desai et al.*, 2005;

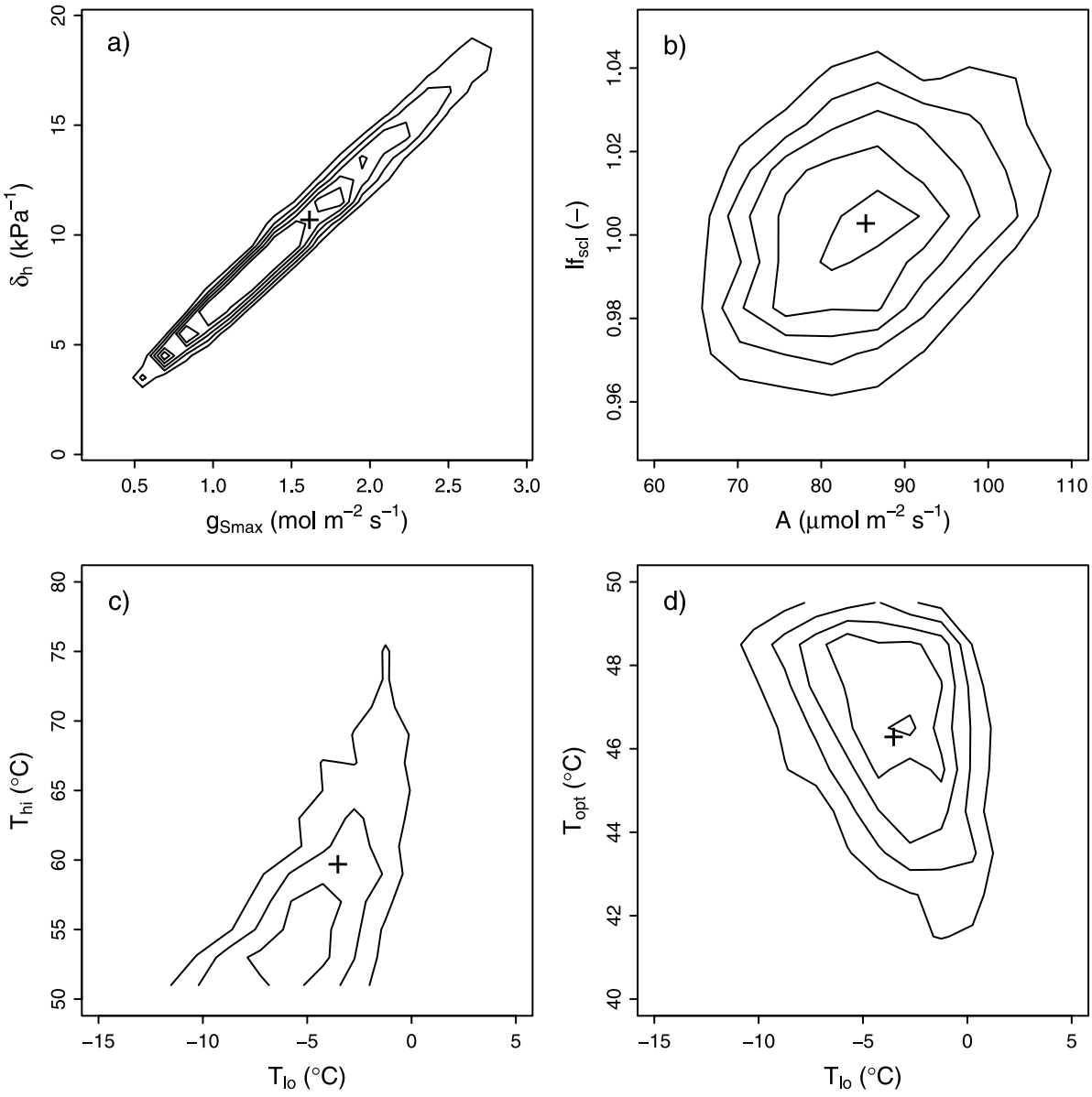

Figure 2. Two-dimensional contour plots drawn from the posterior parameter samples for model HQTP showing (a) δ_h, the parameter determining the dependence of stomatal conductance on vapor pressure deficit within the canopy (D_c) in the hyperbolic form of the constraint function, against g_{Smax}, the highest possible conductance for fully developed leaves per unit leaf area; (b) lf_{scl}, the leaf area scale parameter, against A, the parameter determining the light response of stomatal conductance; (c) T_{hi} against T_{lo}; and (d) T_{opt} against T_{lo}, where T_{opt}, T_{lo}, and T_{hi} are the optimal, lowest, and highest temperatures for transpiration, respectively. HQTP includes hyperbolic constraint function for D_c, constraint functions for average photosynthetic photon flux density (Q_p), air temperature within the canopy (T_c), and day-to-day leaf area adjustment. The contours are equidistant in terms of frequency count differences and connect points of equal frequency counts summed over quadrats. The plus symbols indicate median values.

Ewers et al., 2007a; 2007b; *Mackay et al.*, 2007]. Therefore, DIC appears to be a reasonable metric for comparing conceptual transpiration models of the form used here, and provides the additional benefit of quantitative estimates of changes in model quality associated with specific changes in model structure. The additional information obtained through such an analysis is not only useful for clearly identifying nonessential model components, but also for identifying possible avenues for improving the model structure and the estimates of the parameters, when considered together with the characteristics of the data set used in the evaluation.

[33] Very large reductions in DIC with the inclusions of P indicate that a functional description of the gradual trend in the transpiration data (Figure 1) is essential in the model. Because the estimates of lf_{scl} are close to one and significantly different from zero for all compared models, the gradual variations in the environmental variables used here for modeling g_S are not sufficient for modeling this trend without P. Although inclusions of H and T are

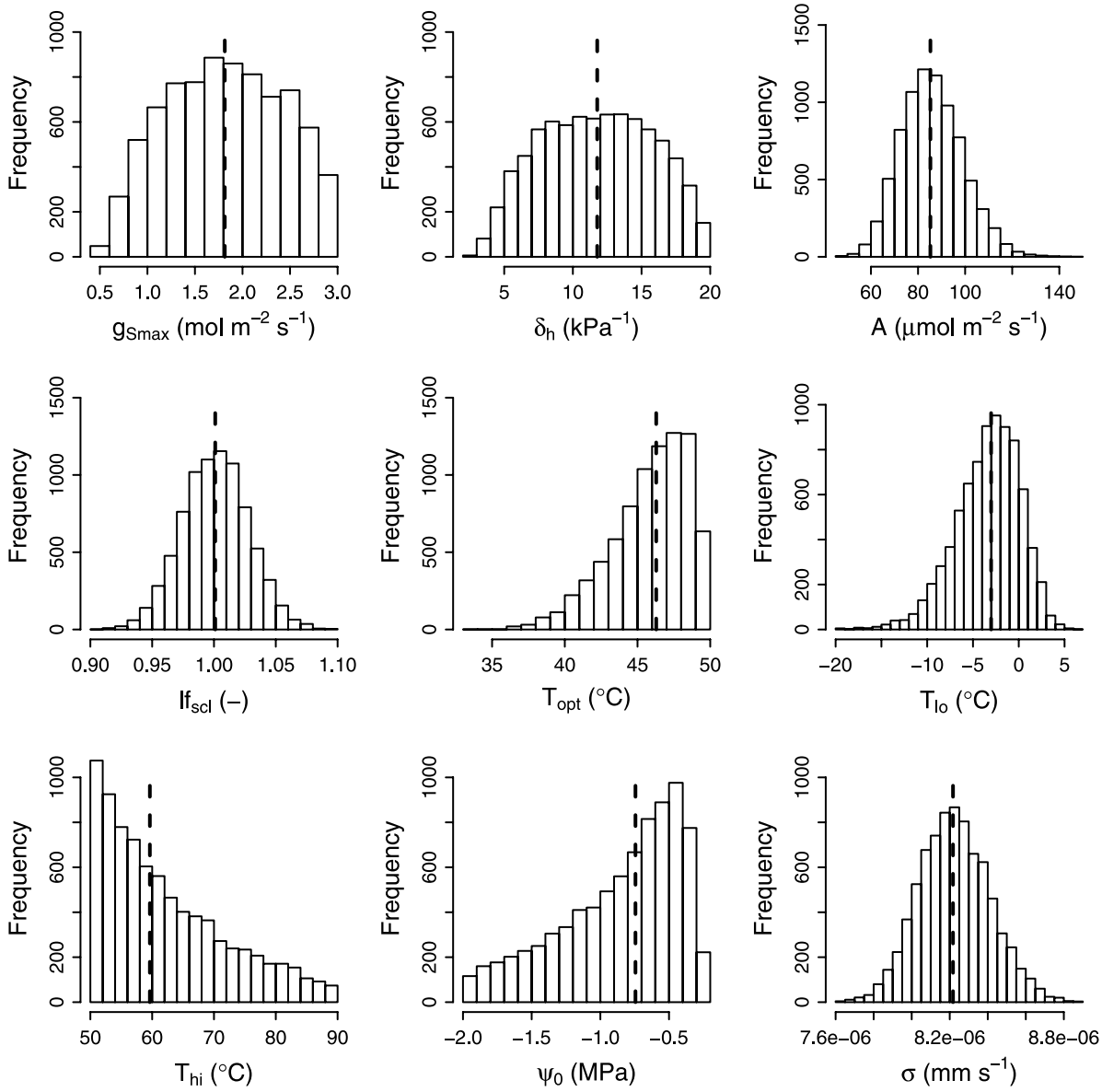

Figure 3. Histograms of the marginal posterior parameter distributions for HQTWP, which includes the hyperbolic constraint function for vapor pressure deficit within the canopy (D_c), constraint functions for average photosynthetic photon flux density (Q_p), air temperature within the canopy (T_c), and soil water potential (Ψ_s), and the day-to-day leaf area adjustment function. The parameter g_{Smax} is the highest possible conductance for fully developed leaves per unit leaf area; δ_h describes the dependence of stomatal conductance on D_c; A determines the light response of stomatal conductance; T_{opt}, T_{lo}, and T_{hi} are the optimal, lowest, and highest temperatures for transpiration, respectively; Ψ_0 is the minimum soil water potential for transpiration; and lf_{scl} is leaf area scale parameter. Dashed lines indicate posterior median values.

generally associated with slightly lower estimates of lf_{scl}, which may suggest that a small part of this trend might be explained by gradual variations in D_c and T_c through the season, considering the 95% posterior intervals of the lf_{scl} estimates (Table 3), the differences do not appear to be significant. While P provides a useful semiempirical method for modeling the transpiration trend through day-to-day adjustment of leaf area, the method requires transpiration data to be available for its use. Moreover, this trend in transpiration is likely to be the combined effect of gradual changes in both leaf area and stomatal conductance over a growing season, processes that are not physically described in P. Therefore, replacing P with functions that utilize conceptual or physically based approaches to model leaf phenology and other causes of long-term variation in g_{vc} [e.g., *Kikuzawa*, 1995; *Wilson et al.*, 2000; *Gratani and Ghia*, 2002; *Brodribb and Holbrook*, 2003] would improve the general applicability, and perhaps also the quality, of the transpiration model. However, inclusion of the above processes would increase the complexity of the transpiration

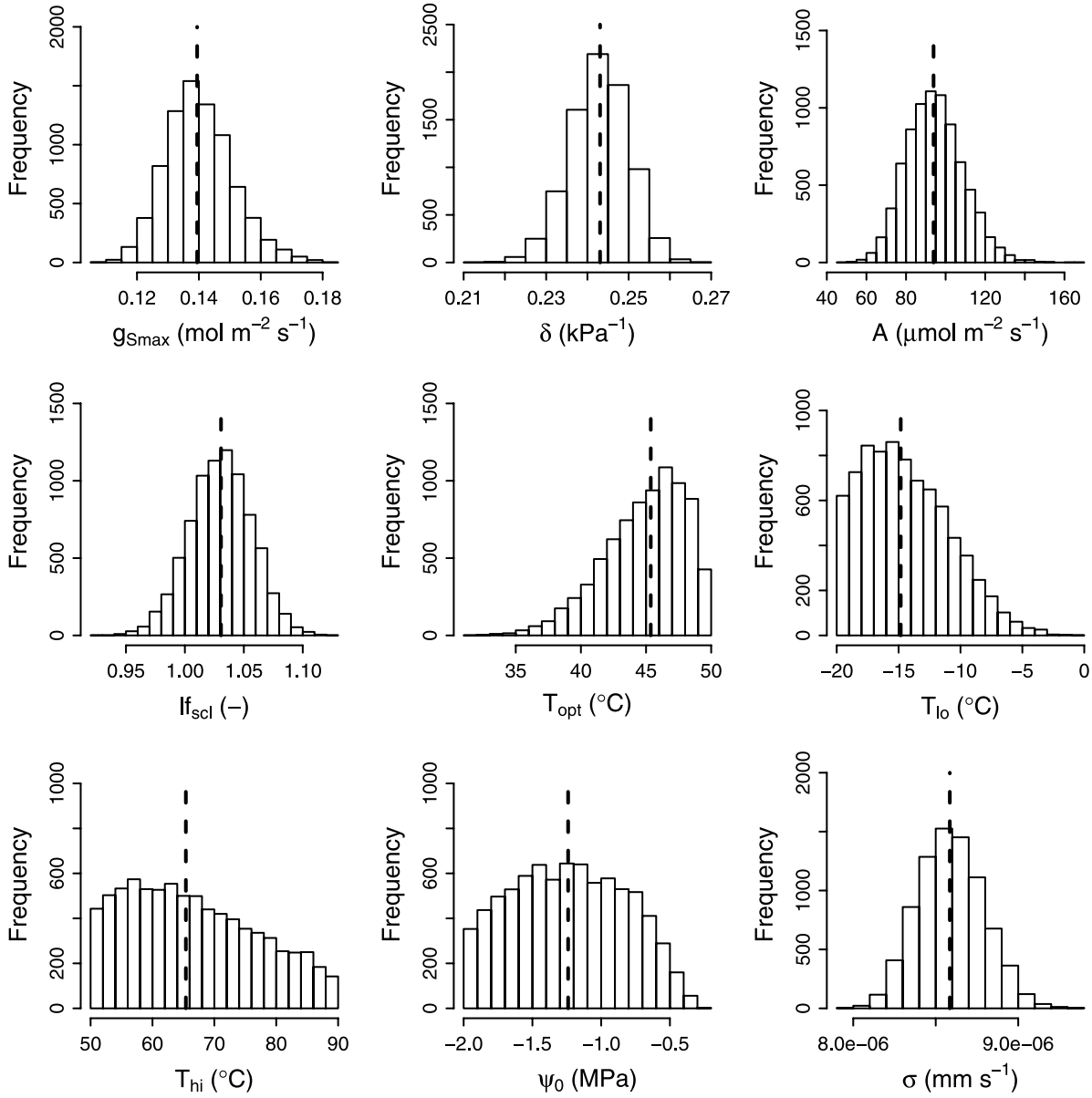

Figure 4. Histograms of the marginal posterior distributions of the parameters for LQTWP, which includes the linear constraint function for vapor pressure deficit within the canopy (D_c), constraint functions for average photosynthetic photon flux density (Q_p), air temperature within the canopy (T_c), and soil water potential (Ψ_s), and the day-to-day leaf area adjustment function. The parameter g_{Smax} is the highest possible conductance for fully developed leaves per unit leaf area; δ describes the dependence of stomatal conductance on D_c; A determines the light response of stomatal conductance; T_{opt}, T_{lo}, and T_{hi} are the optimal, lowest, and highest temperatures for transpiration, respectively; Ψ_0 is the minimum soil water potential for transpiration; and lf_{scl} is leaf area scale parameter. Dashed lines indicate posterior median values.

model, as well as its data requirement. Therefore, the benefit of including such processes would need to be balanced with the associated costs.

[34] The low to moderate significance of W, and the lower significance of T, compared to those of H, L, and Q, appear to be related to the nature of the data used in this evaluation, besides their values in modeling canopy conductance in general. In the case of W, the small effect on DIC may be partially explained by the fact that the values of ω are not less than 20% in these data. Therefore, conditions where soil water availability would limit transpiration were not likely to have been encountered at the site during this period [*Mackay et al.*, 2007]. Similarly, the values of T_c in the data are between 13°C and 33°C, with the majority above 20°C, and therefore, do not represent conditions that impose severe temperature limitations on transpiration, because the temperature response of g_S is fairly broad near the optimum temperature [*Leuning*, 1995]. The above results are also reasonably consistent with the lack of relationship between surface fluxes and soil moisture or

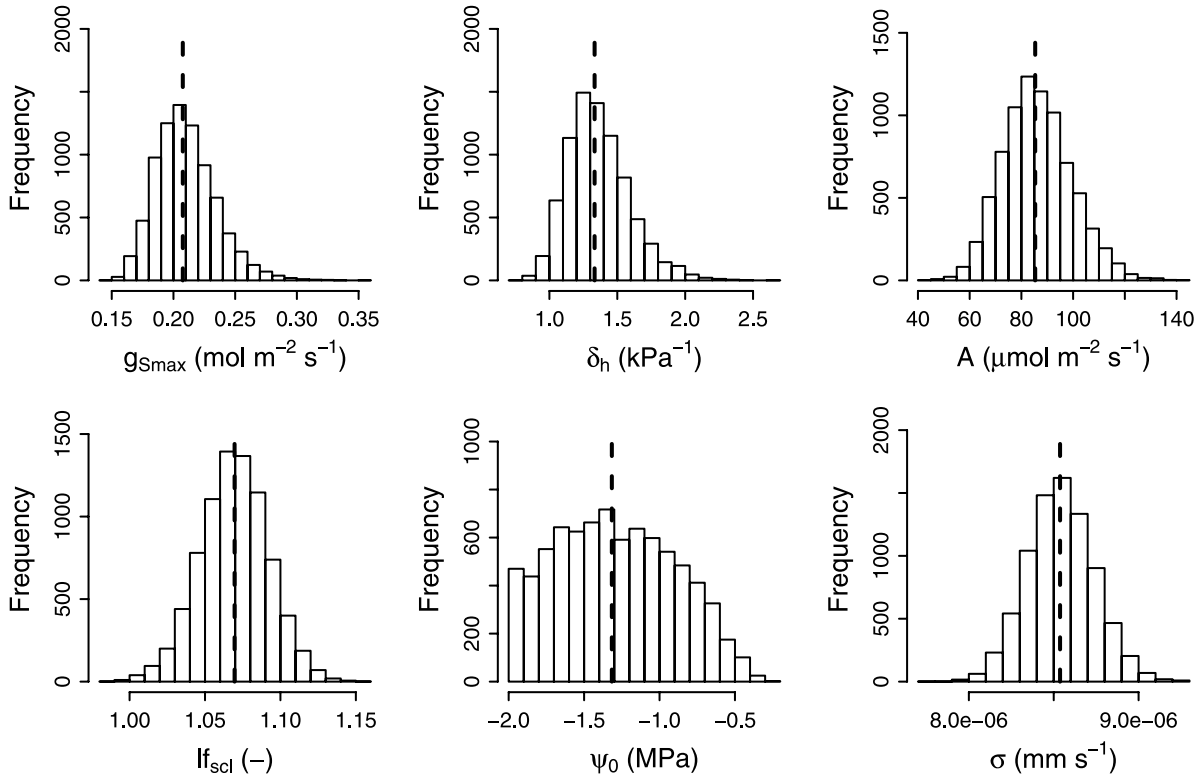

Figure 5. Histograms of the marginal posterior distributions of the parameters for HQWP, which includes the hyperbolic constraint function for vapor pressure deficit within the canopy (D_c), constraint functions for average photosynthetic photon flux density (Q_p), and soil water potential (Ψ_s), and the day-to-day leaf area adjustment function. The parameter g_{Smax} is the highest possible conductance for fully developed leaves per unit leaf area, δ_h describes the dependence of stomatal conductance on D_c, A determines the light response of stomatal conductance, Ψ_0 is the minimum soil water potential for transpiration, and lf_{scl} is leaf area scale parameter. Dashed lines indicate posterior median values.

temperature found by other analyses of surface flux data from the ChEAS sites, except under relatively rare drought conditions in this region [*Ewers et al.*, 2002; *Cook et al.*, 2004; *Desai et al.*, 2005; *Mackay et al.*, 2007]. Therefore, the estimates of contribution of the processes obtained through this DIC-based model comparison appear to reflect well the site conditions and characteristics embedded in the data. However, the environmental variables, and consequently the corresponding constraint functions, might be expected to assume different levels of significance under different conditions or at different sites (e.g., W or T might become more valuable for modeling transpiration during a

Table 3. Posterior Estimates of Canopy Conductance Submodel Parameters for a Few of the Compared Models[a]

Parameter (Units)	HQTWP[b]	HQTP[b]	HQP[b]	LQTWP[b]	LQP[b]
g_{Smax} (mol m^{-2} s^{-1})	1.81 (0.74, 2.89)	1.61 (0.63, 2.81)	0.20 (0.17, 0.26)	0.14 (0.12, 0.16)	0.10 (0.09, 0.11)
δ or δ_h (kPa^{-1})	11.77 (4.58, 18.84)	10.69 (4.02, 18.91)	1.32 (0.97, 1.86)	0.24 (0.23, 0.26)	0.22 (0.20, 0.23)
A (μmol m^{-2} s^{-1})	85.3 (62.7, 113.2)	85.3 (62.7, 113.5)	85.6 (62.4, 113.8)	93.9 (68.7, 124.5)	91.3 (65.9, 121.3)
T_{opt} (°C)	46.3 (39.7, 49.3)	46.3 (39.4, 49.3)	-	45.4 (37.9, 49.3)	-
T_{lo} (°C)	−3.0 (−11.05, 2.4)	−3.5 (−12.3, 2.1)	-	−14.8 (−19.7, −6.7)	-
T_{hi} (°C)	59.6 (50.4, 85.3)	59.7 (50.4, 85.9)	-	65.4 (51.0, 87.2)	-
Ψ_0 (MPa)	−0.74 (−1.84, −0.29)	-	-	−1.24 (−1.94, −0.49)	-
lf_{scl} (-)	1.00 (0.95, 1.05)	1.00 (0.95, 1.05)	1.07 (1.02, 1.11)	1.03 (0.98, 1.08)	1.07 (1.02, 1.11)

[a]Median values are shown to illustrate the sensitivity of parameter values to model structure discussed in the text. The 95% posterior intervals are shown within parenthesis. The values shown are rounded off to facilitate presentation. The parameter g_{Smax} is the highest possible conductance for fully developed leaves per unit leaf area; δ and δ_h describe the dependence of stomatal conductance on vapor pressure deficit within the canopy (D_c) in the linear and hyperbolic forms of the constraint function, respectively; A determines the light response of stomatal conductance; T_{opt}, T_{lo}, and T_{hi} are the optimal, lowest, and highest temperatures for transpiration, respectively; Ψ_0 is the minimum soil water potential for transpiration; and lf_{scl} is leaf area scale parameter for day-to-day leaf area adjustment.

[b]Indicates functions included in the model with L for linear constraint for D_c, H for hyperbolic constraint for D_c, Q for constraint for average photosynthetic photon flux density (Q_p), T for constraint for air temperature within the canopy (T_c), W for constraint for soil water potential (Ψ_s), and P for day-to-day leaf area adjustment.

severe drought, or during freezing conditions, respectively). Therefore, analyses of multiple data sets following a similar methodology would be useful to estimate the significance of such differences on the modeling of transpiration. However, the pattern of diminishing improvements in model quality with the addition of constraint functions suggests that further extensions or modifications to the Jarvis model (e. g., through the uses of alternative forms of constraint functions, alternative environmental variables, such as leaf water potential instead of Ψ_s, or additional environmental variables, such as ambient CO_2 concentration) may be of small benefit within the context of transpiration models of the form used in this analysis.

[35] The results also show that the contributions of constraint functions may, in some cases, be conditional on the overall structure of the model, in addition to data and system characteristics. This dependence is most apparent for the inclusions of W and T, and strong enough to change the inference regarding the value of W (Section 5). Possible reasons for the dependence here are the negative correlation between T_c and ω, and the high positive correlation between T_c and D_c. However, such correlations between environmental variables are to be expected, and therefore, the possibility that the contribution of an individual model component is dependent on the rest of the model structure should be taken into account for model comparison.

[36] On the basis of the foregoing results and discussions, increased model complexity is not always accompanied by a significant improvement in model quality and the extent of improvement achieved by additional functional complexity is not easily predictable. These improvements may also be subtle and detectable only when the rest of the model structure is of requisite quality, which is also an unknown. Therefore, a modification or an increment in complexity that appears to be of insignificant benefit with one model structure may assume a higher significance with another model with structural differences that are not obviously related to the modification being tested. For example, it appears from the results that modeling soil moisture availability in greater detail [e.g., *Noilhan and Planton*, 1989] would not be of value in the model presented here, but the possibility that such a component could become valuable, when commensurate details are added to other parts of the model (e. g., accounting for storage effects in the sap flow through the stem [e.g., *Kumagai*, 2001], representing the plant canopy in greater detail [e.g., *Raulier et al.*, 1999], modeling canopy conductance in terms of photosynthesis [e.g., *Collatz et al.*, 1991]), cannot be ruled out. In other words, inferences regarding the value of a model component, which in the case of semiempirical models can also be regarded as a test for the hypotheses underlying the model component being tested, are applicable only within the larger context of the knowledge and hypotheses already embedded within the model. An interesting consequence of the above is that an advancement of process knowledge in one discipline may not only be useful by itself, but may also render knowledge or hypotheses in other disciplines more useful for modeling aggregated responses of Earth systems. However, it may be impossible to determine a priori whether a model structure has the requisite quality for testing a particular hypothesis. Therefore, the balanced modeling approach recommended by *Grayson and Blöschl* [2001], where different model components are developed and evaluated in concert, should be preferred over the approach of describing only a few model components in as much detail as possible, particularly where the models are to be used to evaluate the significance of individual processes in modeling natural systems responses at large scales. The iterative model building approach suggested by *Box* [2001], along with the collection of appropriate data for objective evaluation of the benefits of increasing model complexity, would be useful for the development of such balanced models. On the basis of the results obtained here, DIC appears to serve well as a useful metric for the evaluation step in such model development cycles because of its abilities to deal with complex model structure, account for parameter uncertainty and interdependence common in Earth system models, and detect subtle changes in model quality.

[37] In addition to the selection of an appropriate model structure, estimation of parameters is also important for using a model to estimate or predict system response. The results indicate that sufficient information regarding all the parameters may not be available in the data being modeled, even when the model components associated with such parameters contribute positively to model quality. This lack of information could be addressed within the Bayesian framework through the use of informative priors, which incorporate information regarding parameters from various independent sources. For conceptual models, informative priors might be specified with the additional requirements of not contradicting known physical relationships, where such knowledge is available [*Raftery et al.*, 1995]. Direct measurements, exploiting the physical relationships embedded in such models, might also be useful for specifying informative priors, although the practical difficulties of doing so are evident from the sensitivity of parameter estimates to model structure seen in the results. As suggested by *Monteith* [1995], this difficulty may be alleviated if measurements intended for determining parameter values are made following experimental protocols designed with their future use in models into consideration. How such measurements and process knowledge might be used together in this way for improving model quality are interesting subjects for further research.

7. Conclusions

[38] According to the results of this model comparison study, DIC is capable of detecting subtle model improvements that result from small structural changes in complex semiempirical models. The need to account for uncertainties in goodness of fit, parameter uncertainty, and the effect of parameter interdependence on the estimate of model complexity, is also clearly indicated. While the specific inferences of this analysis regarding the use of the Jarvis model may not be universally applicable, this study provides an example of how this quantitative model comparison procedure can be used, along with theoretical considerations, to identify structural modifications that are valuable and supported by available observations. In conclusion, this methodology provides a promising technique, which is generally applicable to deterministic Earth system models composed of multiple process-based components that can lead to

significant improvements in modeling Earth system responses with further experimentation and research.

[39] **Acknowledgments.** Funding for this research was provided by the graduate school and the department of forest ecology and management at the University of Wisconsin, Madison, NASA Land Surface Hydrology grant (NAG5–8554) to D. S. Mackay, and USDA Hatch formula funds to D. S. Mackay. Additional support was provided from NSF Hydrological Sciences grants, EAR-0405306 to D. S. Mackay, EAR-0405318 to E. L. Kruger, and EAR-0405381 to B. E. Ewers. Critical reviews by A. J. Guswa and three anonymous reviewers contributed greatly to the quality of the paper. These contributions are gratefully acknowledged.

References

Adams, B., A. White, and T. M. Lenton (2004), An analysis of some diverse approaches to modelling terrestrial net primary productivity, *Ecol. Modell.*, *177*(3–4), 353–391, doi:10.1016/j.ecolmodel.2004.03.014.

Ajami, N. K., Q. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resour. Res.*, *43*, W01403, doi:10.1029/2005WR004745.

Akaike, H. (1974), A new look at the statistical model identification, *IEEE Trans. Autom. Control*, *19*(6), 716–723, doi:10.1109/TAC.1974.1100705.

Anderson-Sprecher, R. (1994), Model comparisons and R2, *Am. Stat.*, *48*(2), 113–117, doi:10.2307/2684259.

Bakwin, P. S., P. P. Tans, D. F. Hurst, and C. Zhao (1998), Measurements of carbon dioxide on very tall towers: Results of the NOAA/CMDL program, *Tellus, Ser. B*, *50*(5), 401–415, doi:10.1034/j.1600-0889.1998.t01-4-00001.x.

Baldocchi, D. D., R. J. Luxmoore, and J. L. Hatfield (1991), Discerning the forest from the trees: An essay on scaling canopy stomatal conductance, *Agric. For. Meteorol.*, *54*(2–4), 197–226, doi:10.1016/0168-1923(91)90006-C.

Band, L., D. Peterson, S. Running, J. Coughlan, R. Lammers, J. Dungan, and R. Nemani (1991), Forest ecosystem processes at the watershed scale: Basis for distributed simulation, *Ecol. Modell.*, *56*(1–4), 171–196, doi:10.1016/0304-3800(91)90199-B.

Bates, B. C., and E. P. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resour. Res.*, *37*, 937–947, doi:10.1029/2000WR900363.

Bayes, T. (1763), An essay towards solving a problem in the doctrine of chances. by the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, *Philos. Trans. R. Soc. London*, *53*, 370–418, doi:10.1098/rstl.1763.0053.

Beven, K. (1989), Changing ideas in hydrology—The case of physically-based models, *J. Hydrol. Amsterdam*, *105*(1–2), 157–172, doi:10.1016/0022-1694(89)90101-7.

Beven, K. (1995), Linking parameters across scales: Subgrid parameterizations and scale dependent hydrological models, *Hydrol. Process.*, *9*(5–6), 507–525, doi:10.1002/hyp.3360090504.

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, *6*(3), 279–298, doi:10.1002/hyp.3360060305.

Box, G. E. P. (2001), Statistics for discovery, *J. Appl. Stat.*, *28*(3–4), 285–299, doi:10.1080/02664760120034036.

Box, G. E. P., and G. C. Tiao (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Boston, Mass.

Box, G. E. P., W. G. Hunter, and J. S. Hunter (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, Wiley Series in Probability and Mathematical Statistics*, John Wiley, New York.

Brodribb, T. J., and N. M. Holbrook (2003), Changes in leaf hydraulic conductance during leaf shedding in seasonally dry tropical forest, *New Phytol.*, *158*, 295–303, doi:10.1046/j.1469-8137.2003.00736.x.

Burnham, K. P., and D. R. Anderson (2002), *Model Selection and Multi-Model Inference*, 3rd ed., Springer, New York.

Cam, E., J. Y. Monnat, and J. A. Royle (2004), Dispersal and individual quality in a long lived species, *Oikos*, *106*(2), 386–398, doi:10.1111/j.0030-1299.2003.13097.x.

Campbell, G. S., and J. M. Norman (1998), *An Introduction to Environmental Biophysics*, Springer, New York.

Clapp, R. B., and G. M. Hornberger (1978), Empirical equations for some soil hydraulic properties, *Water Resour. Res.*, *14*, 601–604, doi:10.1029/WR014i004p00601.

Collatz, G. J., J. T. Ball, C. Grivet, and J. A. Berry (1991), Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: A model that includes a laminar boundary layer, *Agric. For. Meteorol.*, *54*(2–4), 107–136, doi:10.1016/0168-1923(91)90002-8.

Cook, B. D., et al. (2004), Carbon exchange and venting anomalies in an upland deciduous forest in northern Wisconsin, USA, *Agric. For. Meteorol.*, *126*(3–4), 271–295, doi:10.1016/j.agrformet.2004.06.008.

Cowles, M. K., and D. L. Zimmerman (2003), A Bayesian space-time analysis of acid deposition data combined from two monitoring networks, *J. Geophys. Res.*, *108*(D24), 9006, doi:10.1029/2003JD004001.

Cramer, W., D. W. Kicklighter, A. Bondeau, B. Moore III, G. Churkina, B. Nemry, A. Ruimy, A. L. Schloss, and the Participants of the Potsdam NPP Model Intercomparison (1999), Comparing global models of terrestrial net primary productivity (NPP): Overview and key results, *Global Change Biol.*, *5*(1), 1–15, doi:10.1046/j.1365-2486.1999.00009.x.

Davis, K. J., P. S. Bakwin, C. X. Yi, B. W. Berger, C. L. Zhao, R. M. Teclaw, and J. G. Isebrands (2003), The annual cycles of CO_2 and H_2O exchange over a northern mixed forest as observed from a very tall tower, *Global Change Biol.*, *9*(9), 1278–1293, doi:10.1046/j.1365-2486.2003.00672.x.

Dekker, S. C., W. Bouten, and F. C. Bosveld (2001), On the information content of forest transpiration measurements for identifying canopy conductance model parameters, *Hydrol. Process.*, *15*(14), 2821–2832, doi:10.1002/hyp.270.

Desai, A. R., P. V. Bolstad, B. D. Cook, K. J. Davis, and E. V. Carey (2005), Comparing net ecosystem exchange of carbon dioxide between an old-growth and mature forest in the upper Midwest, USA, *Agric. For. Meteorol.*, *128*(1–2), 33–55, doi:10.1016/j.agrformet.2004.09.005.

Ewers, B. E., and R. Oren (2000), Analyses of assumptions and errors in the calculation of stomatal conductance from sap flux measurements, *Tree Physiol.*, *20*(9), 579–589.

Ewers, B. E., R. Oren, K. H. Johnsen, and J. J. Landsberg (2001), Estimating maximum mean canopy stomatal conductance for use in models, *Can. J. For. Res.*, *31*(2), 198–207, doi:10.1139/cjfr-31-2-198.

Ewers, B. E., D. S. Mackay, S. T. Gower, D. E. Ahl, S. N. Burrows, and S. Samanta (2002), Tree species effects on stand transpiration in northern Wisconsin, *Water Resour. Res.*, *38*(7), 1103, doi:10.1029/2001WR000830.

Ewers, B. E., D. S. Mackay, and S. Samanta (2007a), Interannual consistency in canopy stomatal conductance control of leaf water potential across seven tree species, *Tree Physiol.*, *27*(1), 11–24.

Ewers, B. E., D. S. Mackay, J. Tang, P. Bolstad, and S. Samanta (2007b), Intercomparison of sugar maple (Acer saccharum Marsh.) stand transpiration responses to environmental conditions from the western Great Lakes region of the United States, *Agric. For. Meteorol.*, *148*(2), 231–246, doi:10.1016/j.agrformet.2007.08.003.

Foglia, L., S. W. Mehl, M. C. Hill, P. Perona, and P. Burlando (2007), Testing alternative ground water models using cross-validation and other methods, *Ground Water*, *45*(5), 627–641, doi:10.1111/j.1745-6584.2007.00341.x.

Forster, M. R. (2000), Key concepts in model selection: Performance and generalizability, *J. Math. Psychol.*, *44*, 205–231, doi:10.1006/jmps.1999.1284.

Gaganis, P., and L. Smith (2001), A Bayesian approach to the quantification of the effect of model error on the predictions of groundwater models, *Water Resour. Res.*, *37*(9), 2309–2322, doi:10.1029/2000WR000001.

Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Stat. Sci.*, *7*(4), 457–472, doi:10.1214/ss/1177011136.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995), *Bayesian Data Analysis, Texts in Statistical Science Series*, edited by C. Chatfield and J. V. Zidek, CRC Press, London.

Geyer, C. J. (1992), Practical Markov chain Monte Carlo, *Stat. Sci.*, *7*(4), 473–483, doi:10.1214/ss/1177011137.

Gratani, L., and E. Ghia (2002), Changes in morphological and physiological traits during leaf expansion of *Arbutus unedo*, *Environ. Exp. Bot.*, *48*(1), 51–60, doi:10.1016/S0098-8472(02)00010-2.

Grayson, R., and G. Blöschl (2001), Spatial modeling of catchment dynamics, in *Spatial Patterns in Catchment Hydrology: Observations and Modeling*, edited by R. Grayson and G. Blöschl, pp. 51–81, Cambridge Univ. Press, Cambridge, U.K.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*, 751–763, doi:10.1029/97WR03495.

Healy, M. J. R. (1984), The use of R^2 as a measure of goodness of fit, *J. R. Stat. Soc., Ser. A*, *147*, 608–609, doi:10.2307/2981848.

Helser, T. E., and H. L. Lai (2004), A Bayesian hierarchical meta-analysis of fish growth: With an example for North American largemouth bass, *Micropterus salmoides*, *Ecol. Modell.*, *178*(3–4), 399–416, doi:10.1016/j.ecolmodel.2004.02.013.

Jarvis, P. G. (1976), The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field, *Philos. Trans. R. Soc. London, Ser. B.*, *273*(927), 593–610, doi:10.1098/rstb.1976.0035.

Kass, R. E., and A. E. Raftery (1995), Bayes factors, *J. Am. Stat. Assoc.*, *90*(430), 773–795, doi:10.2307/2291091.

Kass, R. E., B. P. Carlin, A. Gelman, and R. M. Neal (1998), Markov chain Monte Carlo in practice: A roundtable discussion, *Am. Stat.*, *52*(2), 93–100, doi:10.2307/2685466.

Katul, G. G., D. S. Ellsworth, and C.-T. Lai (2000), Modelling assimilation and intercellular CO_2 from measured conductance: A synthesis of approaches, *Plant Cell Environ.*, *23*(12), 1313–1328, doi:10.1046/j.1365-3040.2000.00641.x.

Kikuzawa, K. (1995), Leaf phenology as an optimal strategy for carbon gain in plants, *Can. J. Bot.*, *73*(2), 158–163, doi:10.1139/b95-019.

Komatsu, H., N. Hotta, and T. Kume (2007), What is the best way to represent surface conductance for a range of vegetated sites?, *Hydrol. Process.*, *21*(9), 1142–1147, doi:10.1002/hyp.6674.

Kros, H., and P. Warfvinge (1995), Evaluation of model behaviour with respect to the biogeochemistry at the Solling spruce site, *Ecol. Modell.*, *83*(1–2), 255–262, doi:10.1016/0304-3800(95)00103-3.

Krzysztofowicz, R. (1999), Bayesian theory of probabilistic forecasting via deterministic hydrologic model, *Water Resour. Res.*, *35*, 2739–2750, doi:10.1029/1999WR900099.

Kuczera, G. (1983), Improved parameter inference in catchment models 1. Evaluating parameter uncertainty, *Water Resour. Res.*, *19*, 1151–1162.

Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The metropolis algorithm, *J. Hydrol. Amsterdam*, *211*(1–4), 69–85, doi:10.1016/S0022-1694(98)00198-X.

Kullback, S., and R. A. Leibler (1951), On information and sufficiency, *Ann. Math. Stat.*, *22*, 79–86, doi:10.1214/aoms/1177729694.

Kumagai, T. (2001), Modeling water transportation and storage in sapwood—Model development and validation, *Agric. For. Meteorol.*, *109*(2), 105–115, doi:10.1016/S0168-1923(01)00261-1.

Kvålseth, T. O. (1985), Cautionary note about R2, *Am. Stat.*, *39*(4), 279–285.

Leuning, R. (1995), A critical appraisal of a combined stomatal-photosynthesis model for C3 plants, *Plant Cell Environ.*, *18*(4), 339–355, doi:10.1111/j.1365-3040.1995.tb00370.x.

Lhomme, J.-P. (2001), Stomatal control of transpiration: Examination of the Jarvis-type representation of canopy resistance in relation to humidity, *Water Resour. Res.*, *37*, 689–700, doi:10.1029/2000WR900324.

Lohammar, T., S. Larsson, S. Linder, and S. O. Falk (1980), FAST–simulation models of gaseous exchange in Scots pine, in *Structure and Function of Northern Coniferous Forests—An Ecosystem Study*, *Ecol. Bull.*, vol. 32, edited by T. Persson, pp. 505–523, Swed. Natl. Sci. Res. Counc., Stockholm.

Luckai, N., and G. R. Larocque (2002), Challenges in the application of existing process-based models to predict the effects of climate change on C pools in forest ecosystems, *Clim. Change*, *55*(1–2), 39–60, doi:10.1023/A:1020269007045.

Mackay, D. S., B. E. Ewers, B. D. Cook, and K. J. Davis (2007), Environmental drivers of evapotranspiration in a shrub wetland and an upland forest in northern Wisconsin, *Water Resour. Res.*, *43*, W03442, doi:10.1029/2006WR005149.

Manda, S., and R. Meyer (2005), Age at first marriage in Malawi: A Bayesian multilevel analysis using a discrete time-to-event model, *J. R. Stat. Soc., Ser. A*, *168*, 439–455, doi:10.1111/j.1467-985X.2005.00357.x.

Marshall, L., D. Nott, and A. Sharma (2005), Hydrological model selection: A Bayesian alternative, *Water Resour. Res.*, *41*, W10422, doi:10.1029/2004WR003719.

Metropolis, N., and S. Ulam (1949), The Monte Carlo method, *J. Am. Stat. Assoc.*, *44*(247), 335–341, doi:10.2307/2280232.

Misson, L., J. A. Panek, and A. H. Goldstein (2004), A comparison of three approaches to modeling leaf gas exchange in annually drought-stressed ponderosa pine forests, *Tree Physiol.*, *24*(5), 529–541.

Mitchell, P. L. (1997), Misuse of regression for empirical validation of models, *Agric. Syst.*, *54*(3), 313–326, doi:10.1016/S0308-521X(96)00077-7.

Montanari, A., and A. Brath (2004), A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, *40*, W01106, doi:10.1029/2003WR002540.

Monteith, J. L. (1965), Evaporation and environment, in *Proceedings of the 19th Symposium of the Society for Experimental Biology*, pp. 205–233, Cambridge Univ. Press, New York.

Monteith, J. L. (1995), A reinterpretation of stomatal responses to humidity, *Plant Cell Environ.*, *18*(4), 357–364, doi:10.1111/j.1365-3040.1995.tb00371.x.

Moriana, A., F. J. Villalobos, and E. Fereres (2002), Stomatal and photosynthetic responses of olive (Olea europaea L.) leaves to water deficits, *Plant Cell Environ.*, *25*(3), 395–405, doi:10.1046/j.0016-8025.2001.00822.x.

Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environ. Res. Risk Assess*, *17*(5), 291–305, doi:10.1007/s00477-003-0151-7.

Nishida, K., R. R. Nemani, S. W. Running, and J. M. Glassy (2003), An operational remote sensing algorithm of land surface evaporation, *J. Geophys. Res.*, *108*(D9), 4270, doi:10.1029/2002JD002062.

Noilhan, J., and S. Planton (1989), A simple parameterization of land surface processes for meteorological models, *Mon. Weather Rev.*, *117*, 536–549, doi:10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2.

Oren, R., N. Phillips, G. G. Katul, B. E. Ewers, and D. E. Pataki (1998), Scaling xylem sap flux and soil water balance and calculating variance: A method for partitioning water flux in forests, *Ann. Sci. For.*, *55*(1–2), 191–216, doi:10.1051/forest:19980112.

Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994), Verification, validation, and confirmation of numerical-models in the earth-sciences, *Science*, *263*(5147), 641–646, doi:10.1126/science.263.5147.641.

Pan, Y., et al. (1998), Modeled responses of terrestrial ecosystems to elevated atmospheric CO_2: A comparison of simulations by the biogeochemistry models of the Vegetation/Ecosystem Modeling and Analysis Project (VEMAP), *Oecologia*, *114*(3), 389–404, doi:10.1007/s004420050462.

Perruchoud, D. O., and A. Fischlin (1995), The response of the carbon cycle in undisturbed forest ecosystems to climate change: A review of plant-soil models, *J. Biogeogr.*, *22*(4–5), 759–774, doi:10.2307/2845978.

Raftery, A. E., G. H. Givens, and J. E. Zeh (1995), Inference from a deterministic population dynamics model for bowhead whales, *J. Am. Stat. Assoc.*, *90*(430), 402–416, doi:10.2307/2291048.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005), Using Bayesian model averaging to calibrate forecast ensembles, *Mon. Weather Rev.*, *133*, 1155–1174, doi:10.1175/MWR2906.1.

Raulier, F., P. Y. Bernier, and C.-H. Ung (1999), Canopy photosynthesis of sugar maple (Acer saccharum): Comparing big-leaf and multilayer extrapolations of leaf-level measurements, *Tree Physiol.*, *19*(7), 407–420.

Samanta, S., and D. S. Mackay (2003), Flexible automated parameterization of hydrologic models using fuzzy logic, *Water Resour. Res.*, *39*(1), 1009, doi:10.1029/2002WR001349.

Samanta, S., D. S. Mackay, M. K. Clayton, E. L. Kruger, and B. E. Ewers (2007), Bayesian analysis for uncertainty estimation of a canopy transpiration model, *Water Resour. Res.*, *43*, W04424, doi:10.1029/2006WR005028.

Schwarz, G. (1978), Estimating the dimension of a model, *Ann. Stat.*, *6*(2), 461–464, doi:10.1214/aos/1176344136.

Sellers, P. J., Y. Mintz, Y. C. Sud, and A. Dalcher (1986), A simple biosphere model (SiB) for use within general-circulation models, *J. Atmos. Sci.*, *43*(6), 505–531, doi:10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002), Bayesian measures of model complexity and fit (with discussions), *J. R. Stat. Soc., Ser. B*, *64*, 583–639, doi:10.1111/1467-9868.00353.

Spitters, C. J. T., H. A. J. M. Toussaint, and J. Goudriaan (1986), Separating the diffuse and direct component of global radiation and its implications for modeling canopy photosynthesis. 1. Components of incoming radiation, *Agric. For. Meteorol.*, *38*(1–3), 217–229, doi:10.1016/0168-1923(86)90060-2.

Swartzman, G. L., and S. P. Kaluzny (1987), *Ecological Simulation Primer*, Macmillan, New York.

Thiemann, M., M. Trosset, H. V. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, *37*, 2521–2535, doi:10.1029/2000WR900405.

Thogmartin, W. E., J. R. Sauer, and M. G. Knutson (2004), A hierarchical spatial model of avian abundance with application to cerulean warblers, *Ecol. Appl.*, *14*(6), 1766–1779, doi:10.1890/03-5247.

Tiktak, A., and H. J. M. van Grinsven (1995), Review of sixteen forest-soil-atmosphere models, *Ecol. Modell.*, *83*(1–2), 35–53, doi:10.1016/0304-3800(95)00081-6.

van der Perk, M. (1997), Effect of model structure on the accuracy and uncertainty of results from water quality models, *Hydrol. Process.*, *11*(3), 227–239, doi:10.1002/(SICI)1099-1085(19970315)11:3<227::AID-HYP440>3.0.CO;2-#.

van Grinsven, H. J. M., C. T. Driscoll, and A. Tiktak (1995), Workshop on comparison of forest-soil-atmosphere models: Preface, *Ecol. Modell.*, *83*(1–2), 1–6.

van Wijk, M. T., S. C. Dekker, W. Bouten, F. C. Bosveld, W. Kohsiek, K. Kramer, and G. M. J. Mohren (2000), Modeling daily gas exchange of a Douglas-fir forest: Comparison of three stomatal conductance models with and without a soil water stress function, *Tree Physiol.*, *20*(2), 115–122.

Vrugt, J. A., and B. A. Robinson (2007), Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging, *Water Resour. Res.*, *43*, W01411, doi:10.1029/2005WR004838.

Vrugt, J. A., W. Bouten, H. V. Gupta, and J. W. Hopmans (2003a), Toward improved identifiability of soil hydraulic parameters: On the selection of a suitable parametric model, *Vadose Zone J.*, *2*, 98–113.

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003b), A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, *39*(8), 1201, doi:10.1029/2002WR001642.

Wainwright, J., and M. Mulligan (2004), *Introduction in Environmental Modelling: Finding Simplicity in Complexity*, John Wiley, Hoboken, N. J.

Ward, E. J. (2008), A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools, *Ecol. Modell.*, *211*(1–2), 1–10, doi:10.1016/j.ecolmodel.2007.10.030.

Wilson, K. B., D. D. Baldocchi, and P. J. Hanson (2000), Quantifying stomatal and non-stomatal limitations to carbon assimilation resulting from leaf aging and drought in mature deciduous tree species, *Tree Physiol.*, *20*(12), 787–797.

Ye, M., S. P. Neuman, and P. D. Meyer (2004), Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff, *Water Resour. Res.*, *40*, W05113, doi:10.1029/2003WR002557.

M. K. Clayton, Departments of Statistics and Plant Pathology, University of Wisconsin, Madison, Madison, WI 53706, USA.

B. E. Ewers, Department of Botany, University of Wyoming, Laramie, WY 82071, USA.

E. L. Kruger, Department of Forest Ecology and Management, University of Wisconsin, Madison, Madison, WI 53706, USA.

D. S. Mackay, Department of Geography, State University of New York at Buffalo, Buffalo, NY 14261, USA.

S. Samanta, Woods Hole Research Center, 149 Woods Hole Road, Falmouth, MA 02540, USA. (sudeep@uwalumni.com)