

Parsimony vs predictive and functional performance of three stomatal optimization principles in a big-leaf framework

Maoya Bassiouni  and Giulia Vico 

Department of Crop Production Ecology, Swedish University of Agricultural Sciences (SLU), Uppsala 750 07, Sweden

Maoya Bassiouni
Email: maoya.bassiouni@slu.se

Received: 23 December 2020
Accepted: 2 April 2021

New Phytologist (2021) **231**: 586–600
doi: 10.1111/nph.17392

Key words: information theory, model performance, optimality theory, photosynthesis, stomatal conductance, transpiration, water use efficiency, xylem vulnerability.

Summary

- Stomatal optimization models can improve estimates of water and carbon fluxes with relatively low complexity, yet there is no consensus on which formulations are most appropriate for ecosystem-scale applications. We implemented three existing analytical equations for stomatal conductance, based on different water penalty functions, in a big-leaf comparison framework, and determined which optimization principles were most consistent with flux tower observations from different biomes.
- We used information theory to dissect controls of soil water supply and atmospheric demand on evapotranspiration in wet to dry conditions and to quantify missing or inadequate information in model variants. We ranked stomatal optimization principles based on parameter uncertainty, parsimony, predictive accuracy, and functional accuracy of the interactions between soil moisture, vapor pressure deficit, and evapotranspiration.
- Performance was high for all model variants. Water penalty functions with explicit representation of plant hydraulics did not substantially improve predictive or functional accuracy of ecosystem-scale evapotranspiration estimates, and parameterizations were more uncertain, despite having physiological underpinnings at the plant level.
- Stomatal optimization based on water use efficiency thus provided more information about ecosystem-scale evapotranspiration compared to those based on xylem vulnerability and proved more useful in improving ecosystem-scale models with less complexity.

Introduction

Stomata, pores on the surface of leaves, control the exchange of water and carbon dioxide (CO₂), thus playing a key role in regulating terrestrial water and carbon fluxes. Globally, plants recycle over 60% of terrestrial water from the land to the atmosphere through transpiration (Good *et al.*, 2015) and sequester 30% of annual anthropogenic CO₂ emissions through photosynthesis (Friedlingstein *et al.*, 2019). Natural resource management adapted to increasingly variable climatic conditions requires robust modelling of water and carbon cycles and hence of stomatal conductance. Improving our ability to characterize the terrestrial biosphere response to climate – more specifically, accurately translating mechanistic stomatal behavior to larger scales – is essential to reduce uncertainty in water and carbon fluxes (Rogers *et al.*, 2017) and climate projections (Friedlingstein *et al.*, 2014). Under climate change, empirical approaches have limited skill in estimating stomatal conductance (De Kauwe *et al.*, 2013; Liu *et al.*, 2020a). Mechanistic models are needed (Buckley, 2017), which accurately account for hydraulic system feedbacks (Fatichi *et al.*, 2016) and leaf-to-landscape scaling of soil-plant-water transport (Mencuccini *et al.*, 2019a). Complex models, however, have high data requirements and are computationally demanding.

Among mechanistic approaches, stomatal optimization has the potential to supersede empirical models with low complexity (Buckley, 2017). Optimality theory synthesizes stomatal response to environmental change in terms of plants' tradeoffs in carbon assimilation (Cowan & Farquhar, 1977), offering analytical expressions of optimal stomatal conductance. Earth System Models (ESMs) can use these analytical expressions instead of canonical empirical equations for stomatal conductance (Ball *et al.*, 1987; Leuning *et al.*, 1995) because they are functionally similar, with the added benefit of physically meaningful parameters describing ecophysiological tradeoffs (Medlyn *et al.*, 2011). Existing stomatal optimization models differ in how they formulate the cost or penalty of opening stomata, for example water loss vs xylem damage through transpiration (Wang *et al.*, 2020). Global variability of water cost per carbon gained is explained by biophysical characteristics that can be generalized by biome (Lin *et al.*, 2015) and in principle determined as a function of a system's boundary conditions (Cowan & Farquhar, 1977; Manzoni *et al.*, 2013a). A carbon cost associated with xylem damage directly couples stomatal conductance to hydraulic mechanisms (Wolf *et al.*, 2016) and requires no fitting parameters if plant hydraulic traits are known (Sperry *et al.*, 2017; Eller *et al.*, 2018).

The growing number of stomatal optimization approaches and emergent interest in their performance leads to debates over

which is most appropriate (Buckley *et al.*, 2017; Dewar *et al.*, 2018; Wang *et al.*, 2020). There is no consensus on the most realistic carbon cost or water penalty formulations and how optimality criteria vary with soil moisture. Further, optimality criteria are often based on instantaneous costs and benefits, while accounting for longer-term opportunity costs is more consistent with observations (Lu *et al.*, 2020). Finally, the functional accuracy of both leaf- and xylem-level optimization theories, including their parameters, in representing ecosystem-scale behavior and in capturing short- and long-term feedbacks between water consumption, plant status, and soil-water depletion is unknown.

Stomatal optimization models have been reviewed and compared theoretically and empirically at the leaf level (Buckley *et al.*, 2017; Anderegg *et al.*, 2018b; Dewar *et al.*, 2018; Wang *et al.*, 2020) and their predictive performance at large scales compared to empirical approaches, specifically during drought (Eller *et al.*, 2020; Liu *et al.*, 2020a; Sabot *et al.*, 2020). However, ecosystem-scale comparisons of diverse optimality principles and clarity on which formulations improve ESMs are lacking. To this end, robust performance diagnostics compared to ecosystem-scale observations are necessary.

Plant transpiration is simultaneously constrained by soil water supply and atmospheric water demand. Disentangling the effects of these co-varying variables on evapotranspiration and accurately representing their combined effects in models is challenging (Novick *et al.*, 2016; Gentine *et al.*, 2019). Stomatal conductance models should have high *predictive* performance, that is, should accurately reproduce values of a target observation such as transpiration. Stomatal conductance models should also have adequate *functional* performance, that is, should accurately reproduce the sensitivity of transpiration to both soil moisture and vapor pressure deficit as well as the sensitivity of their interactions in varying conditions and ecosystems. Tradeoffs between predictive and functional performance can indicate that a model provides accurate estimates for the wrong reasons due to model structural errors (Ruddell *et al.*, 2019), and thus may not capture the correct response of stomatal conductance to environmental conditions. *Parameterization* performance ensures that data, at the scale of interests, encodes enough information to characterize model relational assumptions, that is, model complexity matches data availability (Feng, 2020).

Information theory can quantify directional and nonlinear functional relations to characterize joint causal interactions in complex ecohydrological systems (Ruddell & Kumar, 2009; Goodwell *et al.*, 2020), such as those between evapotranspiration, soil moisture, and vapor pressure deficit. Information theory also provides performance diagnostics that quantify how much information models extract from observations in a nonparametric way, and can attribute observed errors to specific processes (Weijs *et al.*, 2010; Ruddell *et al.*, 2019). Information quantification is emerging as a powerful technique for model evaluation and hypothesis testing (Nearing *et al.*, 2020). It has been applied to the benchmarking of ESMs (Nearing *et al.*, 2018) but has yet to be used for ecophysiological hypothesis testing.

In addition, quantifying stomatal optimization parameters and characteristics of the soil–plant–atmosphere continuum for

ESMs that preserve key ecological mechanisms of an ecosystem is challenging because such parameters cannot be measured directly at large scales (Mencuccini *et al.*, 2019a). Comparisons of observed vs modeled or leaf vs ecosystem water use efficiency reveal uncertainties in upscaling ecophysiology (Medlyn *et al.*, 2017; Lavergne *et al.*, 2019). Water and carbon fluxes are sensitive to plant trait diversity within an ecosystem. The challenge of summarizing complex interactions resulting from species coexistence using single parameter values remains unresolved (Pappas *et al.*, 2016; Anderegg *et al.*, 2018a). It is thus important to determine whether any added uncertainty in stomatal optimization parameterizations is justified in terms of information gained about water and carbon dynamics at ecosystem scales and apply ‘Occam’s razor’, that is, the principle that improved predictions should outweigh increased complexity (Weijs & Ruddell, 2020).

We evaluate three analytical equations for optimal stomatal conductance, based on different water penalty functions, using flux tower observations and information theory diagnostics. Our goal is to rank optimality principles based on their ability to provide information about transpiration and its interactions with soil moisture and vapor pressure deficit at ecosystem scales. The ideal optimization principle should minimize information loss in terms of uncertainty in model structure, including uncertainty in transpiration estimates, relations with driving variables, and complexity; that is, it should have high *predictive* and high *functional* performance across different plant functional types (PFT) and in wet to dry conditions; and adequate *parameterization*, with parsimonious and well-constrained calibrated parameters that reflect realistic traits of different PFTs.

Materials and Methods

Data

We selected 30 sites (Supporting Information Table S1) from the FLUXNET2015 dataset (Pastorello *et al.*, 2020) that met the following criteria: dominance of a single PFT; observations of model variables for at least 150 (nonconsecutive) growing season days to ensure analysis robustness; soil moisture measurements between 10 and 32 cm to represent soil water content at the average rooting depth; and minimum growing season soil water potential (ψ_s) below -1 MPa to include water-limiting conditions. Selected sites comprise six boreal, 17 temperate, and seven Mediterranean climates; and nine needleleaf, eight broadleaf, nine C3 grass, and four C3 crop PFTs (Table S1).

This analysis required half-hourly data for latent heat flux (LE), gross primary production (GPP, based on nighttime respiration partitioning), atmospheric CO₂ concentration (C_a), soil water content (θ), air vapor pressure deficit (D), air temperature, air pressure, wind speed, friction velocity, incoming photosynthetically active radiation, and sensible heat flux. We identified sensor positions, average canopy height, and soil texture from FLUXNET2015 metadata and references (Table S1). We determined soil hydraulic parameters from soil texture and converted θ to ψ_s (Brooks & Corey, 1964).

We only analyzed data flagged as observed. We excluded outliers and focused on growing season observations during which transpiration dominates LE, following criteria commonly used to estimate water-use efficiency from flux tower data (Knauer *et al.*, 2018; see Methods S1 for details). The remaining data comprise 151 to 1299 d of observations per site, with on average 15 half-hourly observations per day (Table S1).

Optimal stomatal conductance

The generalized optimality theory for stomatal conductance states that plants seek to maximize net carbon gain, that is, photosynthesis (A , mol m⁻² s⁻¹) minus a penalty (Θ , mol m⁻² s⁻¹) (Wolf *et al.*, 2016). Thus, optimal stomatal conductance ($g_{s,opt}$, mol m⁻² s⁻¹) varies with environmental conditions so that the marginal photosynthetic gain of transpiration ($\partial A/\partial T$) is equal to the marginal penalty of transpiration or marginal cost of water per carbon gained ($\partial \Theta/\partial T$):

$$\max(A - \Theta) \equiv \frac{\partial A}{\partial T} = \frac{\partial \Theta}{\partial T} \quad \text{Eqn 1}$$

Although there are a number of existing conceptually different optimality principles, Eqn 1 can be solved using a common mathematical framework. A study by Dewar *et al.* (2018) provides a generic derivation for $g_{s,opt}$ for any choice of penalty function Θ and leaf photosynthesis model A . Neglecting nonstomatal reductions in photosynthesis and assuming $\alpha D(\frac{\partial \Theta}{\partial T}) \ll C_a - \Gamma^*$ for simplicity, the generic analytical solution for $g_{s,opt}$ is

$$g_{s,opt} = \frac{f_0}{\gamma + \Gamma^*} \frac{x}{(1-x) \left(\frac{C_a - \Gamma^*}{\gamma + \Gamma^*} x + 1 \right)} \quad \text{Eqn 2}$$

where $x = \left(1 + \sqrt{\alpha D(\partial \Theta/\partial T)/(\gamma + \Gamma^*)} \right)^{-1}$; C_a is atmospheric CO₂ concentration (mol mol⁻¹); D is the difference in water

vapor concentration between the atmosphere and saturated air space in the leaf (mol mol⁻¹); Γ^* is the CO₂ photorespiratory compensation point (mol mol⁻¹); α is the relative diffusivity of water vapor with respect to CO₂ (mol mol⁻¹); f_0 is the CO₂-saturated rate of photosynthesis (mol m⁻² s⁻¹); and γ is a kinetic parameter (mol mol⁻¹). We represented A by a continuous function embedding limitations from both CO₂ transport (Rubisco activity) and incident light (RuBP regeneration) to avoid *a priori* selection of light- vs Rubisco-limited conditions and streamline model calculation (Vico *et al.*, 2013). We defined this function and associated f_0 and γ in Methods S2.

In addition, Eqn 2 is often recast as a generic relation between $g_{s,opt}$ and A (Medlyn *et al.*, 2011; Wolf *et al.*, 2016; Dewar *et al.*, 2018), which is functionally analogous to commonly used empirical models (Ball *et al.*, 1987; Leuning *et al.*, 1995):

$$g_{s,opt} = \left(1 + \sqrt{\frac{\gamma + \Gamma^*}{\alpha D \frac{\partial \Theta}{\partial T}}} \right) \frac{A}{C_a - \Gamma^*} \quad \text{Eqn 3}$$

Selected penalty functions

We selected three optimality principles based on different formulations of Θ , which provide varying $\partial \Theta/\partial T$ functional forms (Table 1) and cover broad categories of existing optimization models (Wang *et al.*, 2020). The water use efficiency principle (WUE) defines the penalty function as water loss through transpiration (Katul *et al.*, 2009; Medlyn *et al.*, 2011); the key optimization parameter represents marginal water use efficiency. The carbon maximization (CM) and xylem vulnerability (SOX) principles define the penalty function as hydraulic damage (Wolf *et al.*, 2016; Anderegg *et al.*, 2018b) or risk (Eller *et al.*, 2018), respectively; the key parameters reflect plant hydraulic behavior.

For WUE, we represented $\partial \Theta/\partial T$ by an empirical exponential function, emerging from a leaf-level meta-analysis (Manzoni

Table 1 Definition of selected stomatal optimization principles.

Name (abbreviation)	Water use efficiency (WUE)	Carbon maximization (CM)	Xylem vulnerability (SOX)
Optimization principle	Maximum difference between gains and costs of water losses from transpiration	Maximum difference between gains and costs of hydraulic damage	Maximum difference between gains and hydraulic risk
Water penalty (Θ)	$\lambda T(g_s)$	$b_2 \psi(T)^2 + b_1 \psi(T) + b_0$	$A \left(1 - \frac{K(\psi(T))}{K_{\max}} \right)$
Marginal water penalty ($\partial \Theta/\partial T$)	$\lambda_{ww} e^{\beta_0 \psi }$	$\frac{b'_2 \psi + b'_1}{K'(\psi)}$	$-\frac{\partial K}{\partial T} \frac{A}{K_{\max} K'(\psi)}$
Parameters	$\lambda_{ww}; \beta_0$	$b'_1; b'_2; \psi_{50}; a$	$K_{\max}; \psi_{50}; a$
References	Katul <i>et al.</i> (2009), Manzoni <i>et al.</i> (2011), Medlyn <i>et al.</i> (2011)	Wolf <i>et al.</i> (2016), Anderegg <i>et al.</i> (2018b)	Eller <i>et al.</i> (2018)

A (mol m⁻² s⁻¹), photosynthesis; g_s (mol m⁻² s⁻¹), stomatal conductance; T (mol m⁻² s⁻¹), transpiration; ψ (MPa), plant water potential; $K = K_{\max} [1 + (\psi/\psi_{50})^a]^{-1}$ (mol m⁻² s⁻¹ MPa⁻¹), soil-to-canopy conductance as an empirical function of ψ (Manzoni *et al.*, 2013a), K_{\max} (mol m⁻² s⁻¹ MPa⁻¹), maximum soil-to-canopy conductance; $K' = K(\psi)/K_{\max}$ (unitless), normalized soil-to-canopy conductance describing hydraulic vulnerability; λ (mol mol⁻¹), marginal water use efficiency; λ_{ww} (mol mol⁻¹), well-watered λ ; β_0 (MPa⁻¹), slope of the relation between $\log(\lambda)$ and ψ ; b'_1 (mol mol⁻¹), intercept of marginal xylem tension efficiency normalized by K_{\max} ; b'_2 (mol mol⁻¹ MPa⁻¹), slope of marginal xylem tension efficiency normalized by $2K_{\max}$; ψ_{50} (MPa), water potential at 50% conductivity loss; a (unitless), shape parameter representing the sensitivity of xylem conductance to changes in water potential; $\partial K/\partial T = (a/\psi)(\psi/\psi_{50})^a K'$. Model variants using literature values for the hydraulic vulnerability curve instead of calibrated values are evaluated in Supporting Information Fig. S6.

et al., 2011) and reflecting solutions from Cowan (1982). This function requires two parameters: marginal water use efficiency under well-watered conditions (λ_{ww} , mol mol⁻¹) and a fitting parameter (β_0 , MPa⁻¹). For CM, $\partial\Theta/\partial T$ requires four parameters: intercept (b'_1 , mol mol⁻¹) and slope (b'_2 , mol mol⁻¹ MPa⁻¹) of an empirical marginal xylem tension efficiency (normalized by maximum soil-to-canopy conductance (K_{max} , mol m⁻² s⁻¹ MPa⁻¹)); water potential at 50% loss of xylem conductance (ψ_{50} , MPa); and a shape parameter that controls the sensitivity of xylem conductance to changes in plant water potential (a , unitless). For SOX, $\partial\Theta/\partial T$ requires three parameters of the hydraulic vulnerability curve: K_{max} , ψ_{50} and a .

Several other optimality models exist in the literature, including numerical or dynamic stomatal optimization models, which were not evaluated in this study (see reviews by Buckley *et al.*, 2017; Dewar *et al.*, 2018; Mencuccini *et al.*, 2019; Wang *et al.*, 2020). We focus on analytical expressions of $g_{\text{s,opt}}$ resulting from WUE, CM, and SOX because they can be applied without numerical optimization. Using each principle, $\partial\Theta/\partial T$ connects stomatal behavior to the hydraulic system in a more physically meaningful way than commonly used water-limitation constraints (Feddes *et al.*, 1978). Our goal was to evaluate which of the selected optimality principles has the most potential to robustly replace empirical relations between soil moisture, vapor pressure deficit, and stomatal conductance in ESMs without increasing computational costs.

Big-leaf comparison framework

We developed a generic model that provides a level playing field to focus on the comparison of different $\partial\Theta/\partial T$ formulations without confounding factors. We used a big-leaf framework (Bonan, 2019) to scale $g_{\text{s,opt}}$ per unit leaf area, determined based on mechanistic descriptions of leaf-level stomatal behavior, to total surface conductance per unit ground area (G_{surf} , mol m⁻² s⁻¹), representing the flux tower scale. We implemented G_{surf} in the Penman–Monteith equation (Penman & Keen, 1948; Monteith, 1965) using forcing data from FLUXNET2015 and calculated ecosystem-scale LE. Model variants refer to the implementation of each of the stomatal optimization principles (WUE, CM, and SOX) in the comparison framework.

We defined surface conductance to water vapor, G_{surf} , as the sum of soil (G_{soil}) and canopy (G_{canopy}) components (Li *et al.*, 2019). We expressed G_{soil} as a linear function of soil moisture, bound by a maximum value ($G_{\text{soil,sat}}$) at soil saturation (θ_s), that is, $G_{\text{soil}} = G_{\text{soil,sat}}\theta/\theta_s$. This simple assumption is reasonable because we considered only days with no rain and with no rain during the preceding day, which generally follow water-limited soil water evaporation (Or *et al.*, 2013). The big-leaf framework assumes all leaves in the canopy respond to the same environmental conditions measured by the flux tower. We defined G_{canopy} within the big-leaf framework by recasting the relation between $g_{\text{s,opt}}$ and A (Eqn 3) at the ecosystem scale (Lin *et al.*, 2018), assuming air vapor pressure deficit (normalized by air pressure) represents D and substituting A (CO₂ flux per unit leaf area) with GPP (CO₂ flux per unit ground area):

$$G_{\text{canopy}} = \alpha \left(1 + \sqrt{\frac{\gamma + \Gamma^*}{\alpha D \frac{\partial\Theta}{\partial T}}} \right) \frac{\text{GPP}}{C_a - \Gamma^*} \quad \text{Eqn 4}$$

For all optimality principles, we calculated $\partial\Theta/\partial T$ as a function of concurrent ψ_s converted to predawn ψ , reflecting a daily optimization timescale. We thus mitigated uncertainties in the instantaneous estimation of canopy and xylem water potentials and avoided iterative calculations, which did not improve the performance of our comparison framework (Fig. S1).

While the big-leaf comparison framework did not explicitly model certain processes, they were included in its structure to some extent by exploiting the availability of GPP data. Our comparison framework was devised to minimize structural errors that are beyond the optimization principles and may affect model comparisons, including uncertainty in scaling from leaf to surface conductance and in estimating photosynthesis. The framework must match information in the data of both LE observations, as the target, in terms of predictions, and the functional relation with GPP. The substitution of A with GPP measurements, despite observational uncertainties, provides constraints to partition soil water evaporation; and scaling factors representing effective ecosystem leaf area index; while accounting for all environmental factors affecting photosynthesis, including irradiance, leaf temperature, and leaf water potential (Figs S1–S4; Methods S2). We could thus focus on comparing selected stomatal optimization principles and how $\partial\Theta/\partial T$ varied with soil moisture through predawn ψ .

Parameterization

We evaluated the adequacy of the optimization principles for generalized applications, that is, when site-specific or species-level information is not available and parameter values are based on PFTs. We calibrated the parameters in two steps: we first estimated site-specific $G_{\text{soil,sat}}$ values, then PFT-level parameter values to express $\partial\Theta/\partial T$ (Table 1). This two-step parameterization ensures that soil water evaporation is equal in each model variant. We could thus assume that comparisons of model diagnostics mainly reflected differences in $\partial\Theta/\partial T$ formulations rather than uncertainty in evapotranspiration partitioning resulting from varying $G_{\text{soil,sat}}$ calibration in model variants.

Best-fit parameters minimized least square errors between observed LE and modeled LE (Penman–Monteith implemented with $G_{\text{surf}} = G_{\text{soil}} + G_{\text{canopy}}$), using ‘Non-Linear Least-Square Minimization and Curve-Fitting for Python (v.0.8.0)’ (Newville *et al.*, 2014). We constrained parameters to broad but realistic ranges, based on the literature (Manzoni *et al.*, 2011; Lin *et al.*, 2015; Anderegg *et al.*, 2018b; Eller *et al.*, 2020; Sabot *et al.*, 2020).

We performed the least-square minimization for each model variant (WUE, CM, and SOX) on the full record of each site and averaged the resulting $G_{\text{soil,sat}}$ estimates. We then implemented the average site-specific $G_{\text{soil,sat}}$ and performed a bootstrapping calibration process for the remaining model-specific parameters (λ_{ww} , β_0 , b'_1 , b'_2 , K_{max} , a , ψ_{50}). For each site and model variant,

we repeated the least-square minimization 200 times on randomly selected data subsamples and initial value guesses. A different 50% of the record in each subsample was randomly split in half for calibration (25%) and validation (25%) to obtain a distribution of 200 parameter estimates. For each site, we determined leave-one-site-out parameters to express $\partial\Theta/\partial T$ for WUE, CM, and SOX as the median of parameter estimates from all other sites of the same PFT.

Information theory diagnostics

Information theory can measure directional and nonlinear functional relations using nonparametric techniques, thus overcoming limitations of commonly used correlation measures in the analysis of complex environmental processes (Cover & Thomas, 2012). The analysis of information flows is an established technique that quantifies shared information between variables in environmental systems and characterizes functional couplings and joint causal interactions (Ruddell & Kumar, 2009; Goodwell *et al.*, 2020). Information theory is particularly well suited to comparing stomatal optimization principles because metrics only extract information from data, without imposing potentially incorrect assumptions that can lead to misinterpretation of incomplete information (Weijs *et al.*, 2010). Missing or bad information can result from model structural errors (Nearing & Gupta, 2015). An information-theoretical approach especially benefits ecosystem-scale analyses because their observations and processes are often uncertain. We exploited information flows to dissect instantaneous controls on latent heat (LE) by two of its key drivers, soil moisture, θ , and vapor pressure deficit, D ; quantify how much information about LE is missing, given a stomatal optimization principle; and evaluate whether model variants under- or over-estimate the interactions between LE, θ , and D .

Measurements of θ and D provide information about LE, that is, knowing θ and/or D reduces LE uncertainty. Information theory quantifies uncertainty in a variable (observed or modeled), by Shannon's entropy and shared information with mutual information, both in units of bits (Cover & Thomas, 2012). Mutual information between LE and one of its key drivers ($I(\theta, \text{LE})$ or $I(D, \text{LE})$, Fig. 1a) includes unique (U_θ or U_D) information about LE, provided by θ or D separately and by themselves, and redundant (R) information, provided by knowing either θ or D . Conditional mutual information ($I(\theta; \text{LE}|D)$, Fig. 1b) quantifies the reduction in LE uncertainty given knowledge of θ beyond the information provided by D ; it includes U_θ and synergistic (S) information, provided by knowing both θ and D because they interact in defining LE. Finally, total multi-variate mutual information ($I(\theta, D; \text{LE})$, Fig. 1c,d), the total reduction in uncertainty of LE given knowledge of variables θ and D together, is the sum of four nonnegative partitions $U_\theta + U_D + S + R$ (Williams & Beer, 2010; Goodwell & Kumar, 2017). Three-dimensional information flows thus reveal aspects of nonlinear dependencies of LE on θ and D that are undetectable when considering relations in pairs or with correlation measures. Relative magnitudes of information partitions reflect causal interactions between variables in the system (Goodwell *et al.*, 2018) and are relevant to

understanding complex functional ecophysiological relations and accurately representing them in models.

We calculated information flows from half-hourly observed and modeled LE independently, following established methods: we estimated entropy by binning normalized variables and tested statistical significance with shuffled surrogates (Methods S3, S4). We focused on instantaneous controls of θ and D on LE at the half-hourly timescale. We considered $I(\theta, D; \text{LE})$ a meaningful measure of functional relations between these variables because it is statistically significant (> 99% confidence level) at all study sites. We normalized the information partitions U_θ , U_D , S and R by the total information $I(\theta, D; \text{LE})$ to measure relative controls of θ and D on LE in units of bits per bit.

Parameterization, predictive and functional performance

Performance diagnostics based on information theory (Methods S5) represent missing or inadequate information in a model compared to observations (Ruddell *et al.*, 2019; Nearing *et al.*, 2020).

Parameterization performance is a measure of how much information a model variant can extract from data about its parameters and model relational assumptions. An overly complex model may require more data than is available to constrain the values of its parameters. We quantified parameter uncertainty or information missing about each parameter given a model variant as the information entropy of the 200 parameter estimates for each site. We evaluated tradeoffs between goodness-of-fit and complexity by comparing the model variants based on their Akaike information criterion (AIC), which is a measure of information loss, taking into account the number of model parameters (Akaike, 1974).

Predictive performance (A_p) is a measure of a model's ability to represent the target variable (LE), similarly to commonly used goodness-of-fit metrics. We calculated A_p as the relative fraction of LE uncertainty, observed in the flux tower dataset, which is explained by a model, that is, the relative difference between the entropy of LE observations and mutual information between observed and modeled LE. A value of A_p equal to zero indicates that there is no missing information or no uncertainty in the model about LE and all the variability in observed LE is captured by modeled LE; while $A_p = 1$ indicates that observed and modeled LE are independent.

Functional performance (A_f) and its components are a measure of a model's ability to represent the correct strength of the relations between the target variable and its drivers, that is, the total information from θ and D about LE and its partitioning (U_θ , U_D , S and R) in the model should match that in the observations. We calculated six A_f components as the relative difference between observed and modeled information flows: total multi-variate mutual information ($A_{f,T}$); its partitioning into unique ($A_{f,\theta}$, $A_{f,D}$), synergistic ($A_{f,S}$), and redundant ($A_{f,R}$) information; and the sum of these four partitioning accuracy absolute values ($A_{f,p}$). Positive/negative $A_{f,T}$ values indicate that the strength of the functional relations between θ and D and LE in the model are over/underestimated compared to observations. Overestimation of $A_{f,T}$ indicates an overly deterministic model. Positive/

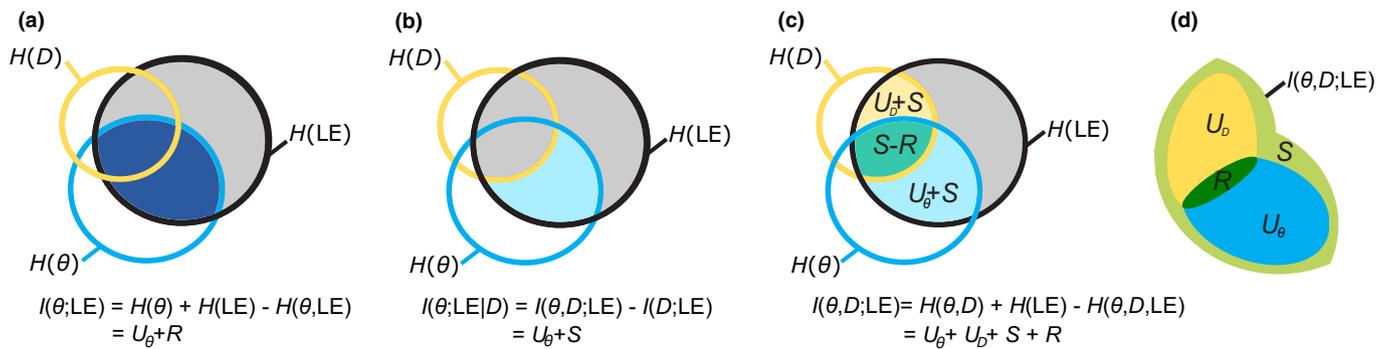


Fig. 1 Illustration of information theory metrics (bits). Shannon's entropy ($H(X)$) quantifies uncertainty in a variable X , such as latent heat flux (LE, black circles), soil moisture (θ , blue circles), and vapor pressure deficit (D , yellow circles). Colored overlapping areas represent shared information or a reduction in LE uncertainty given knowledge of other variables; grey areas represent remaining uncertainty about LE. (a) Mutual information ($I(\theta;LE)$, dark blue) is the reduction in uncertainty about LE given knowledge of θ . (b) Conditional mutual information, ($I(\theta;LE|D)$, light blue) is the reduction in uncertainty about LE given knowledge of θ , beyond information already provided by D ($I(D;LE)$). (c) Multi-variate mutual information ($I(\theta, D;LE)$, multiple colors) is the total reduction in uncertainty about LE given knowledge of θ and D together, and is composed of four nonnegative components. (d) Unique information from θ (U_{θ} , blue); unique information from D (U_D , yellow); synergistic information (S , light green), and redundant information (R , dark green). Interaction information ($S - R$), emerging from dependencies between θ and D , is positive if S dominates and negative if R dominates. Supporting Information Methods S4 and S5 further define information theory metrics and provide technical details for their calculation.

negative $A_{f,\theta}$ (or $A_{f,D}$) values indicate that specifically unique information θ (or D) provided about LE in the model is over/underestimated compared to observations, thus attributing errors to the strength of a particular functional relation. Positive/negative $A_{f,S}$ and $A_{f,R}$ values indicate that controls from θ and D together on LE are inaccurate, thus attributing errors to interactive effects between θ and D represented in the model. $A_{f,P}$ summarizes overall functional accuracy in reproducing the sensitivity of LE to θ and D together, which also takes into account dependencies between the driving variables.

Model ranking and performance scores

We used five criteria to capture a broad range of tradeoffs in performance for each model variant and at each study site: (a) calibration/validation goodness-of-fit, measured by the relative difference between validation and calibration Nash–Sutcliffe efficiencies (NSE) and median absolute percentage errors (MAPE) resulting from the 200 subsamples; (b) low parameter uncertainty measured by the coefficient of variation and the entropy of the 200 parameter estimates; (c) adequate parsimony measured by AIC; (d) predictive accuracy in estimating evapotranspiration measured by A_p ; (e) functional accuracy in reproducing the sensitivity of evapotranspiration to θ and D measured by $A_{f,T}$ and $A_{f,P}$. We quantified and ranked model performance for the full record and for dry, mesic, and wet conditions separately, by calculating AIC, A_p , and the six A_f components from data below, between, and above the interquartile range of θ observed in each site's record.

We ranked the model variants at each site according to diagnostics for each criterion. We then averaged site-specific and diagnostic-specific ranks and converted them to overall performance scores (0, 1) for each criterion. The performance score translates to the fraction of sites for which (or probability that) the model variant in question is the most desirable. A score > 0.5 indicates that a criterion is statistically better for the considered

model variant than a randomly selected variant. A score > 0.67 (or < 0.33) indicates that a model variant is statistically the most (or least) desirable among the three considered. The ranking approach was necessary to summarize results because diagnostic values from sites with heterogeneous record lengths and characteristics are not comparable and absolute values are less meaningful.

Results

We implemented three stomatal optimization principles (WUE, CM, and SOX), based on different water penalty functions in a big-leaf comparison framework to estimate evapotranspiration, and quantified missing or inadequate information in the three model variants using flux tower data. Model ranking according to the five evaluation criteria was largely consistent across biomes (Fig. 2) and soil moisture conditions (Fig. S5).

Calibration/validation goodness-of-fit

The bootstrapping calibration was good for all model variants (Figs 3, 4). SOX scored highest (0.58) and CM lowest (0.38) in terms of calibration/validation goodness-of-fit (Fig. 2a). The median validation NSE and MAPE were 85 and 16%, and the relative difference in NSE and MAPE between calibration and validation subsamples was largely under 5%, for all variants and sites combined (200 per site).

Parameter uncertainty

WUE scored highest (0.65) and CM lowest (0.35) in terms of least parameter uncertainty (Fig. 2b). Parameter coefficients of variation resulting from the bootstrapping calibration, for all sites combined, was < 0.1 for λ_{ww} , β_0 and a (in SOX only), 0.2 for β'_1 , and > 0.5 for other parameters (Fig. 5a). Entropy of the 200 parameter estimates per site ranged from 0 to 6 bits (Fig. 5b),

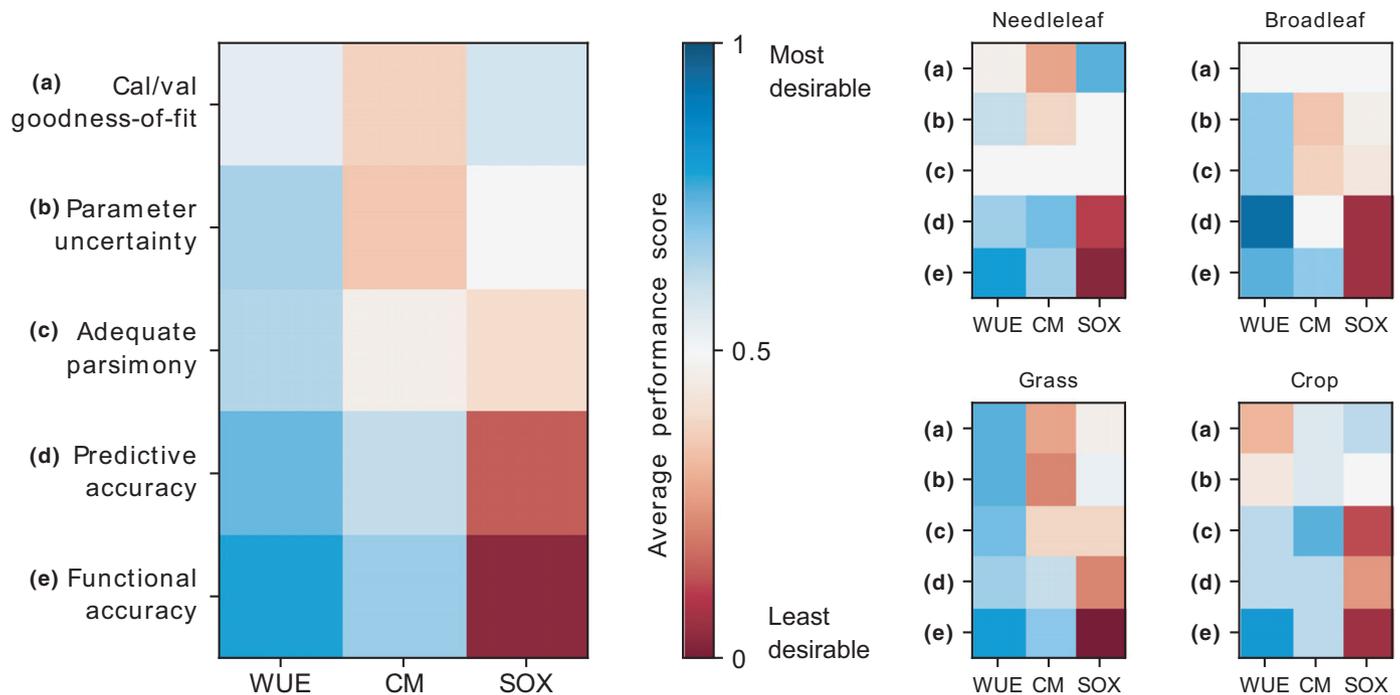


Fig. 2 Average performance scores of model variants using water use efficiency (WUE), carbon maximization (CM), and xylem vulnerability (SOX) stomatal optimization principles (Table 1) according to five criteria: (a) calibration/validation goodness-of-fit, measured by differences in Nash–Sutcliffe efficiency and median absolute percentage error; (b) parameter uncertainty, measured by the coefficient of variation and entropy of calibrated parameter values; (c) adequate parsimony, measured by the Akaike Information Criterion; (d) predictive accuracy, measured by missing information about latent heat flux (LE) in the model (A_p); and (e) functional accuracy, measured by the accuracy of multi-variate mutual information from soil moisture (θ) and vapor pressure deficit (D) about LE ($A_{f,T}$) and its partitioning into unique, synergistic, and redundant components ($A_{f,P} = |A_{f,\theta}| + |A_{f,D}| + |A_{f,S}| + |A_{f,R}|$). We rank model variants within each site using full records and average the 30 site-specific (left) and the vegetation-specific site scores (small plots on the right) for each diagnostic and model variant. We provide details of individual diagnostic scores in each criterion, calculated for the full record, wet, mesic, and dry conditions, in Supporting Information Fig. S5.

which translates to the average number of binary questions that need to be asked to determine the parameter value at a site.

The range of parameter estimates within each PFT was large (Fig. 6). The relatively uniform parameter distributions for CM, especially ψ_{50} and a , indicated that the values were difficult to constrain and that, using CM, $\partial\Theta/\partial T$ had a low sensitivity to its parameterization. Correlations between parameters were statistically significant ($P < 0.01$) for all optimality principles. Risks of equifinality appeared highest for SOX and CM, for which Spearman correlations between ψ_{50} and other parameters were highest and ranged from 0.43 with b'_1 to as high as -0.89 with K_{\max} .

Adequate parsimony

The AIC scores indicated that WUE (0.63) was more desirable than CM (0.47) and SOX (0.40) at a majority of sites in terms of prediction vs parsimony, especially in dry conditions (Figs 7a, 2c).

Predictive accuracy

Ranges of predictive performance diagnostics among the sites were similarly good for all model variants (Fig. 7b). WUE

(0.73) and CM (0.62) scored highest, while SOX (0.15) was the statistically least desirable principle in terms of predictive performance (Fig. 2d). Within a site, predictive performance always decreased in dry conditions, compared to mesic and wet conditions, as reflected by a higher A_p value. Missing information was at most 25% greater in dry vs wet conditions, for all sites and models combined. WUE was more robust in dry conditions than CM at a majority of sites (Fig. S5). The increase in missing information resulting from the use of leave-one-site-out parameters instead of median site-specific parameters was $< 5\%$, for all sites and models combined (Fig. 7c).

Functional accuracy

Total information from θ and D about LE was overestimated (positive $A_{f,T}$) at most sites for all model variants; this overestimation was overall lower under wet conditions compared to mesic and dry conditions (Fig. 8a). Functional performance measured by information partitioning (Fig. 8b–e) indicated that for all optimality principles θ controls on LE were slightly underestimated (negative $A_{f,\theta}$) in mesic and wet conditions; D controls on LE were overestimated (positive $A_{f,D}$); synergistic contributions of θ and D information on LE were

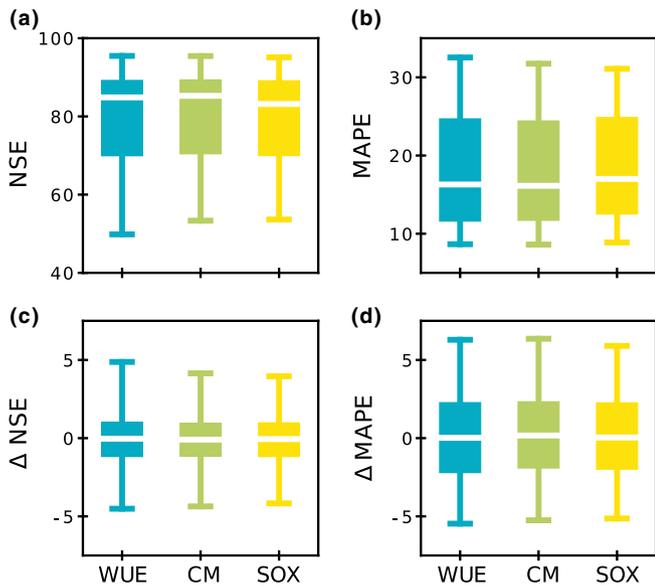


Fig. 3 Goodness-of-fit of the bootstrapping calibration approach using water use efficiency (WUE), carbon maximization (CM), and xylem vulnerability (SOX) stomatal optimization principles (Table 1) to estimate latent heat flux (LE): (a) Nash–Sutcliffe efficiency (NSE, %); (b) median percentage error (MAPE, %) of validation subsets; and percentage differences between (c) NSE and (d) MAPE validation and calibration subsets. Boxes represent the interquartile range for the 30 study sites, whiskers represent the 5th and 95th percentiles, and white lines represent medians.

underestimated (negative $A_{f,S}$); and redundant contributions were accurately represented ($A_{f,R}$ close to zero). WUE (0.79) and CM (0.68) scored highest, while SOX (0.03) was the statistically least desirable principle in terms of functional performance (Fig. 2e).

Predictive and functional performance tradeoffs

Sites with high predictive performance generally also had the highest functional performance (Fig. 9). The fact that the relation between A_p and $A_{f,P}$ (Fig. 9a) was stronger than that of A_p and $|A_{f,T}|$ (Fig. 9b) indicated that $A_{f,T}$ is insufficient to quantify functional performance. Predictive performance was higher when information from θ and D about LE in the models matched observed information partitions (U_θ , U_D , S and R) rather than total information amount ($I(\theta, D; LE)$). Tradeoffs between the accuracy of total information and its partitioning (Fig. 9c) indicated that total information accuracy can occur at the expense of more inaccurate model representations of θ and D controls on LE. For example, if a model underestimates U_θ and overestimates U_D by the same amount of information, performance based on $A_{f,T}$ can be close to zero (modeled and observed multi-variate mutual information match), while $A_{f,P}$, the sum of partitioning components, will be high (modeled and observed relations between θ , D , and LE do not match). Performance was

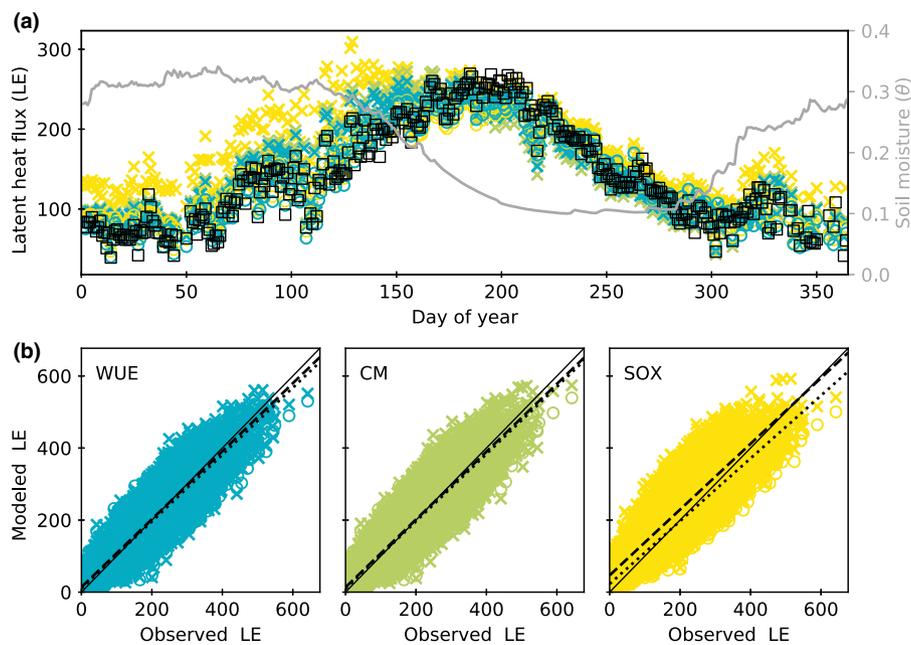


Fig. 4 Goodness-of-fit between observed and modeled latent heat flux (LE, $W m^{-2}$) at an example site (US-Blo, a seasonally dry evergreen needleleaf forest (Goldstein *et al.*, 2000)), using water use efficiency (WUE, blue), carbon maximization (CM, green), and xylem vulnerability (SOX, yellow) stomatal optimization principles (Table 1). (a) Observed (black squares) and modeled LE, averaged by day of year, using site-specific (colored circles) and leave-one-site-out (colored crosses) calibrated parameters. Soil moisture (θ , $cm^3 cm^{-3}$), averaged by day of year in grey to illustrate seasonal soil-water availability. (b) Scatter plots of half-hourly observed vs modeled LE. Black lines represent 1 : 1 (solid) and goodness-of-fit regressions using site-specific (dotted) and leave-one-site-out (dashed) calibrated parameters. The R^2 (and bias) values using leave-one-site-out parameters are 0.92 (0.08), 0.90 (0.08) and 0.87 (0.28) for WUE, CM, and SOX, respectively.

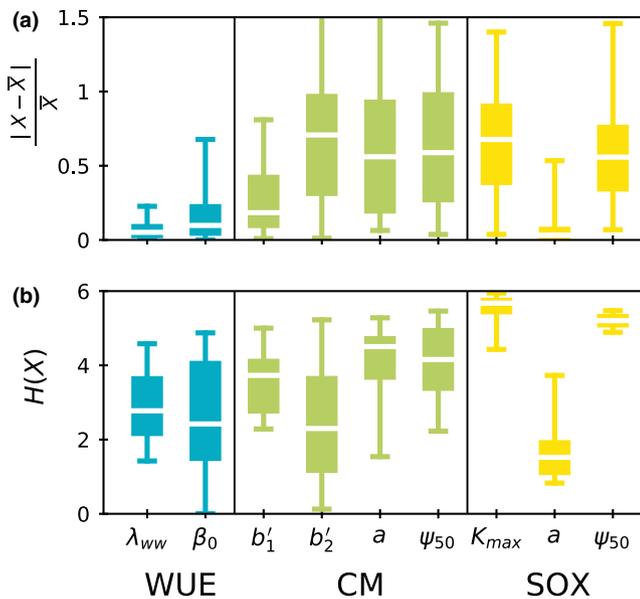


Fig. 5 Parameter uncertainty resulting from the bootstrapping model calibration approach. (a) Distribution of parameter estimates (200 per site) normalized as the relative absolute difference from the site-specific mean value. (b) Distribution of parameter entropy ($H(X)$, bits) at each site. Water use efficiency (WUE), carbon maximization (CM), and xylem vulnerability (SOX) stomatal optimization principles and respective model parameters are defined in Table 1. Boxes represent the interquartile range for the 30 study sites, whiskers represent the 5th and 95th percentiles, and white lines represent medians.

higher at sites with herbaceous vegetation compared to woody vegetation, and performance tradeoffs were less present.

Discussion

We evaluated three existing stomatal optimization principles beyond predictive performance using a novel data-driven multi-criteria method. We also quantified the extent to which optimization principles were robustly parameterized and accurately captured stomatal response to environmental conditions – in other words, whether they were ‘right for the right reasons’. We used a big-leaf framework, taking advantage of flux tower LE and GPP measurements to minimize model structural errors, and created a level playing field to focus on the comparison of stomatal optimization principles. We were thus able to attribute differences in performance metrics between model variants using WUE, CM, and SOX principles to the differences in their respective $\partial\Theta/\partial T$, as this was the only component of the comparison framework that changed in each variant. We used information theory to diagnose the multiple aspects of model performance, which proved a simple and effective technique to disentangle nonlinear ecohydrological relations and overcome challenges of ecosystem-scale analysis when data and processes are uncertain. Our approach was thus able to rank the ecosystem-scale performance of selected optimality principles in terms of prediction, function and parameterization, and reveal which principle can be most structurally adequate for ecosystem-scale applications beyond the comparison framework.

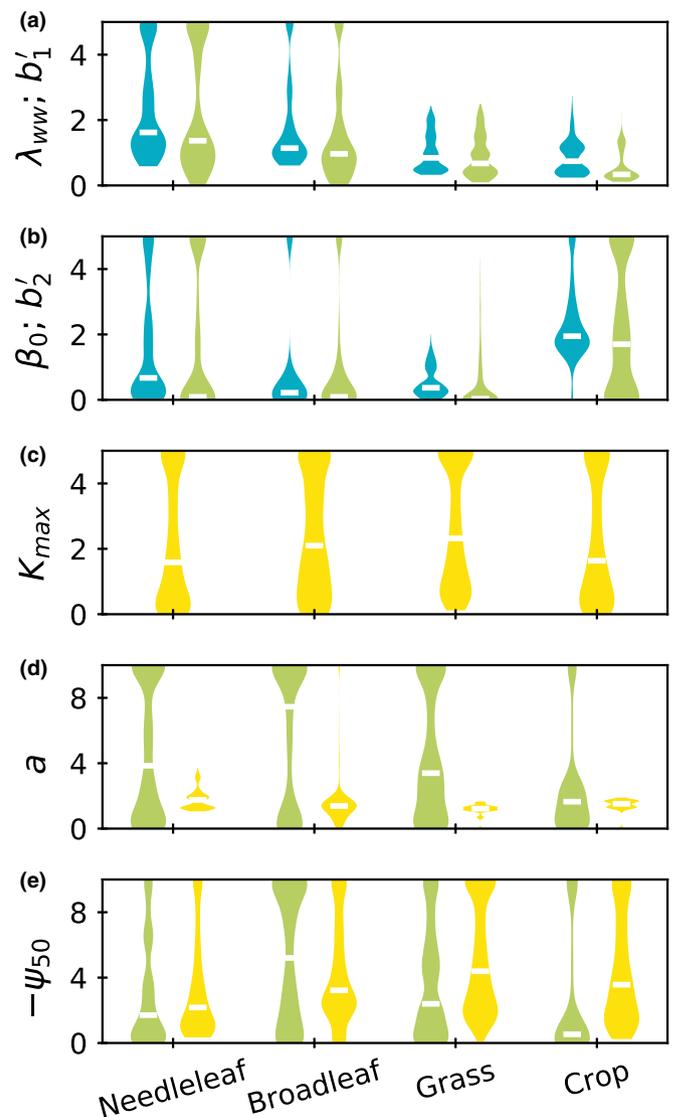
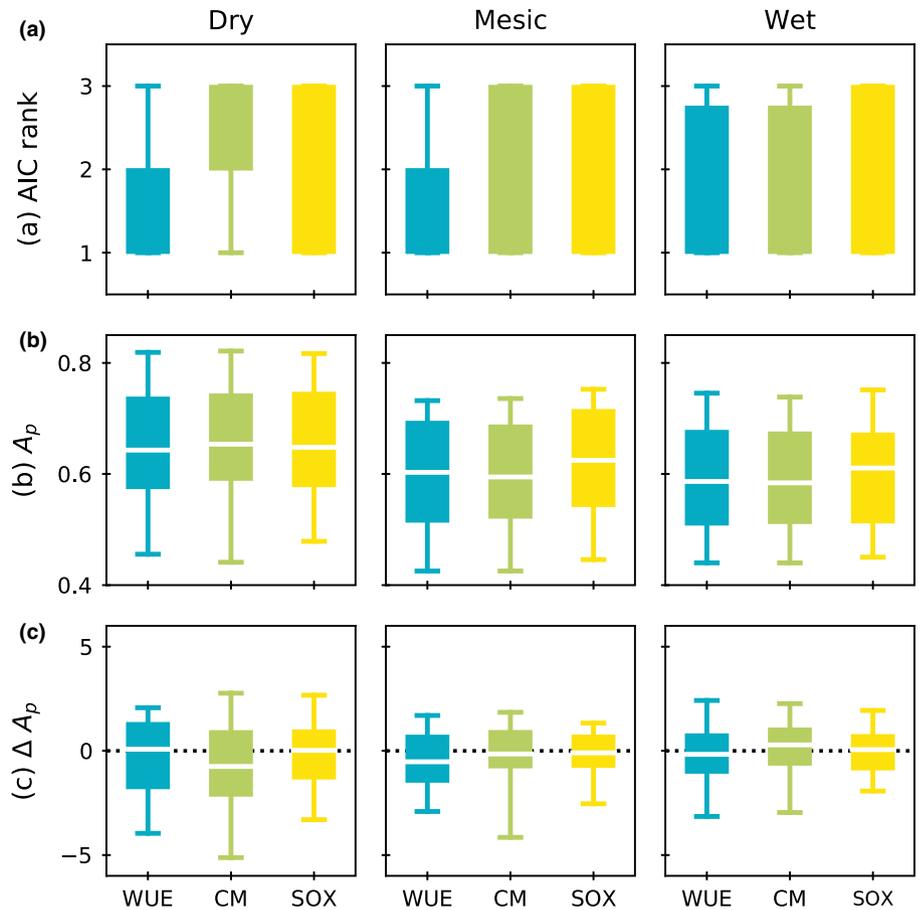


Fig. 6 Distribution of model parameter estimates resulting from the bootstrapping calibration approach for the 30 study sites (200 estimates per site) categorized according to needleleaf ($n = 8$), broadleaf ($n = 8$), grass ($n = 9$), and crop ($n = 4$) plant functional types, including: (a) λ_{ww} ($10^{-3} \text{ mol mol}^{-1}$), well-watered marginal water use efficiency; b'_1 ($10^{-3} \text{ mol mol}^{-1}$), intercept of marginal xylem tension efficiency normalized by K_{max} ; (b) β_0 (MPa^{-1}), slope of the relation between the log of marginal water use efficiency and water potential; b'_2 ($10^{-3} \text{ mol mol}^{-1} \text{ MPa}^{-1}$), slope of marginal xylem tension efficiency normalized by $2K_{max}$; (c) K_{max} ($10^{-3} \text{ mol m}^{-2} \text{ s}^{-1} \text{ MPa}^{-1}$), maximum soil-to-canopy conductance; (d) a (unitless), shape parameter representing the sensitivity of xylem conductance to changes in water potential; (e) ψ_{50} (MPa), water potential at 50% conductivity loss. Water use efficiency (WUE, blue), carbon maximization (CM, green), and xylem vulnerability (SOX, yellow) stomatal optimization principles are defined in Table 1. The y-axis limits of each subplot represent the ranges of each parameter considered in the calibration process.

Parameterization performance for WUE and SOX was superior to CM

Ranges of calibrated parameter values were larger within than between PFTs (Fig. 6), which is a common issue for ecosystem

Fig. 7 Predictive model performance for dry, mesic, and wet conditions, using water use efficiency (WUE), carbon maximization (CM), and xylem vulnerability (SOX) stomatal optimization principles (Table 1) to estimate latent heat flux (LE). (a) The Akaike Information Criterion (AIC) rank between model variants characterizes model tradeoffs between prediction and the number of calibrated model parameters. Within a site, the model with the lowest AIC is the most desirable (rank = 1). (b) Predictive performance (A_p , bits bit⁻¹) quantifies the relative fraction of information about LE missing from the model, compared to observed LE. (c) The relative difference in predictive performance (ΔA_p , %) using site-specific parameters vs out-of-sample parameters indicates low information loss from parameters generalized at the level of plant functional types. We define dry, mesic, and wet as data points below, between and above the interquartile range of the observed soil moisture for each site. Boxes represent the interquartile range for the 30 study sites, whiskers represent the 5th and 95th percentiles, and white lines represent medians.



estimates (Medlyn *et al.*, 2017; Liu *et al.*, 2020b). Stomatal optimization models were nevertheless adequately generalized at the PFT level (Figs 4, 7c), suggesting that all optimization principles are robust and applicable beyond site-specific calibration. Ecosystem-scale parameters were, however, not equally well constrained. For all PFTs, WUE and SOX parameters were easier to calibrate, in relative terms, while CM parameters were more uncertain (Fig. 5). Hydraulic parameters were also more difficult to constrain than marginal water use efficiency in other ecosystem-scale model calibrations (Liu *et al.*, 2020b). Model variants using literature values for a and ψ_{50} instead of calibrated values reduced the uncertainty of parameterizations and lowered risks of equifinality; these variants were found to be less desirable using SOX and more desirable using CM in terms of prediction vs complexity tradeoffs; but the variants still did not outperform WUE overall (Fig. S6).

Calibrated WUE parameters generally reflected plant-level patterns (Manzoni *et al.*, 2011; Lin *et al.*, 2015): λ_{ww} was higher for needleleaf than broadleaf trees and for grasses than crops; and marginal water use efficiency was more sensitive to decreasing soil water for needleleaf than for broadleaf trees, and for crops than for grasses (higher β). Calibrated CM parameters were less physiologically meaningful; CM had low sensitivity to a and ψ_{50} , and consistency in best-fit b'_1 and b'_2 with previous leaf-level parameterizations was uncertain (Anderegg *et al.*, 2018b). Calibrated b'_1 had similar values to λ_{ww} , and b'_2 (the key parameter that

distinguishes CM from WUE) often approached 0 and had similar variability to β_0 . This suggests that WUE is structurally more consistent with flux tower data than CM, and marginal xylem tension efficiency at ecosystem scales is not sensitive to plant water potential. Equifinality of CM parameters is consistent with leaf-level studies (Zenes *et al.*, 2020). Calibrated values for a in SOX were generally consistent with previous ecosystem-scale applications (Eller *et al.*, 2020) and within broad ranges of measured hydraulic traits (Manzoni *et al.*, 2013a; Mencuccini *et al.*, 2019b): calibrated a was higher (more nonlinear behavior) for needleleaf, closest to 1 for grasses, and overall was smaller than values fit to species-level vulnerability curves. K_{max} was greater for broadleaf than needleleaf trees. Patterns of variation in calibrated K_{max} and ψ_{50} values were not always consistent with those of measured traits due to tradeoffs between parameters that lead to similar model fits. While the correlations between hydraulic parameters have ecological significance, equifinality is a challenge for determining generalized parameters for ecosystem-scale models. This equifinality points to a need to parameterize K_{max} independently, for example using long-term climate data (Sabot *et al.*, 2020).

Predictive and functional performance was high overall and inferior for SOX

Predictive and functional performance of all stomatal optimization principles was high (Figs 7, 8), which is expected as model

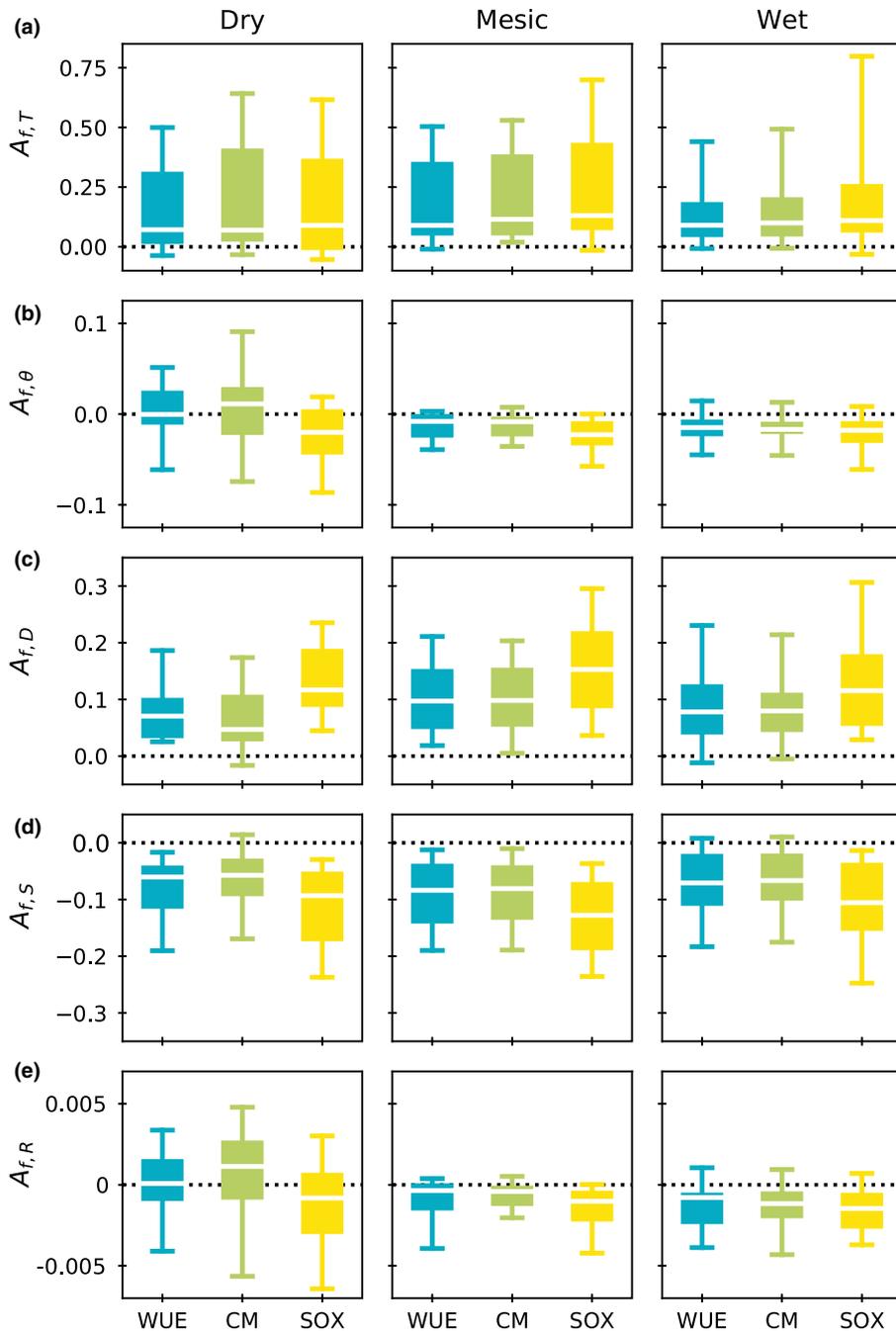
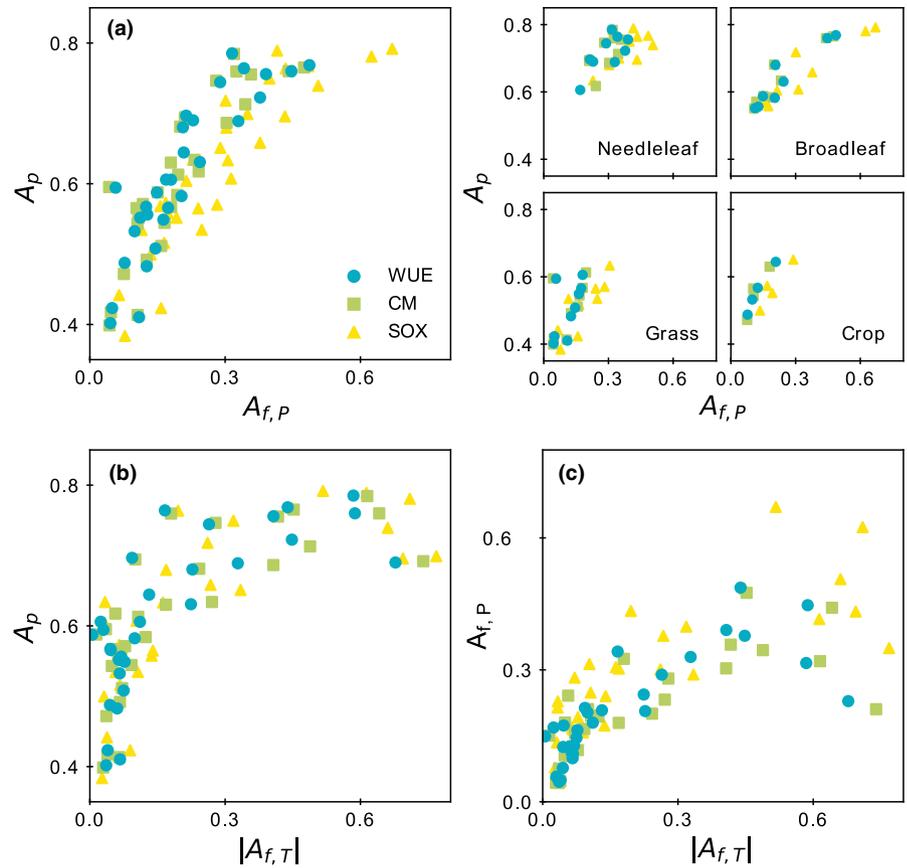


Fig. 8 Functional model performance for dry, mesic, and wet conditions, using water use efficiency (WUE), carbon maximization (CM), and xylem vulnerability (SOX) stomatal optimization principles (Table 1) to estimate latent heat flux (LE). Functional performance is the relative difference between observed and modeled (a) total multi-variate mutual information from soil moisture (θ) and vapor pressure deficit (D) about LE ($A_{f,T}$, bits bit^{-1}) and its partitioning into (b) unique from soil moisture ($A_{f,\theta}$, bits bit^{-1}), (c) unique from atmospheric demand ($A_{f,D}$, bits bit^{-1}), (d) synergistic ($A_{f,S}$, bits bit^{-1}), and (e) redundant ($A_{f,R}$, bits bit^{-1}) information. We define dry, mesic, and wet as data points below, between and above the interquartile range of the observed soil moisture for each site. Boxes represent the interquartile range for the 30 study sites, whiskers represent 5th and 95th percentiles, and white lines represent medians.

variants were derived from a common eco-evolutionary theory of optimal plant water–carbon tradeoffs and reflect coordination between stomatal behavior and plant hydraulic strategies (Wang *et al.*, 2020). Functional performance indicated that models were generally overly deterministic, in particular in the sensitivity of LE to D , slightly underestimated θ controls on $\partial\Theta/\partial T$, and underestimated synergistic information from D and θ . The positive correlation between predictive and functional performance (Fig. 9) indicated that optimality principles can generally be calibrated to ecosystem-scale observations at the level of PFT without overfitting parameters for highest prediction accuracy at the expense of poor functional accuracy.

Model performance at sites with herbaceous vegetation was higher, in terms of less missing information and more accurate functional representation of θ and D controls on LE, than at forested sites. Forests create and respond to more heterogeneous environmental conditions (soil moisture at different depths, sunlit vs shaded leaves, aerodynamic resistance, vapor pressure deficit within vs above the canopy), which cannot be represented in a big-leaf framework and can be a source of functional inaccuracy. Nevertheless, these errors affected all variants in the same way. Further comparison framework assumptions, including optimization timescale, evapotranspiration partitioning, leaf-to-canopy scaling, and photosynthesis (Figs S1–S4; Methods S2), do not

Fig. 9 Relations and tradeoffs between performance diagnostics based on information theory. (a) Predictive accuracy, measured in terms of missing information about latent heat flux (LE) in model variants (A_p , bits bit^{-1}), vs functional accuracy, measured in terms of the accuracy of multi-variate mutual information from soil moisture (θ) and vapor pressure deficit (D) about LE partitioned into unique, redundant and synergistic components ($A_{f,P} = |A_{f,\theta}| + |A_{f,D}| + |A_{f,S}| + |A_{f,R}|$, bits bit^{-1}). (b) Predictive accuracy (A_p , bits bit^{-1}) vs functional accuracy, measured in terms of the accuracy of total multi-variate mutual information from θ and D about LE ($|A_{f,T}|$, bits bit^{-1}). (c) Tradeoff between the two functional accuracies ($A_{f,T}$ vs $A_{f,P}$). The absolute value of functional accuracy is taken because A_f components can be positive or negative, where a value close to zero indicates the best match with observations. Markers represent the performance of a study site and are colored according to water use efficiency (WUE, blue circles), carbon maximization (CM, green squares), and xylem vulnerability (SOX, yellow triangles) stomatal optimization principles (Table 1). Small subplots on the right in (a) separate study sites by plant functional type.



affect ranks because inputs, uncertainties and errors were also shared by all model variants. We do not expect that model performance was affected by the use of predawn vs instantaneous water potential because predictive and functional accuracy were inferior when implementing canopy water potential estimates, and the use of predawn water potential led to consistent model ranking (Fig. S1). Additionally, there were no clear discrepancies in functional performance between broadleaf vs needleleaf trees (Fig. 9a), which generally have different stringencies of stomatal control (iso/anisohydry).

The least desirable principle in terms of ecosystem-scale predictive and functional performance was SOX (Fig. 2d,e), despite having a more physiologically grounded penalty function based only on measurable plant parameters (Wang *et al.*, 2020). The SOX optimality principle resulted in the most overly deterministic model variant, which was the least functionally accurate in terms of information partitioning, and it was the only principle to consistently underestimate unique information from θ in dry conditions. Structural differences in SOX lead to lower $\partial\Theta/\partial T$ values throughout the time series at a majority of sites compared to WUE, and these values were less consistent with observations. Equations for WUE and CM were structurally very similar and emerged as the more robust model variants at the flux tower scale. Predictive and functional performance was higher at more sites using WUE than CM (Fig. 2d,e), although they were not substantially different, in contrast to previous results based on leaf-level gas exchange measurements (Anderegg *et al.*, 2018b; Wang *et al.*, 2020).

Stomatal optimization based on WUE provided more information about evapotranspiration than CM or SOX

The ideal stomatal optimization principle should provide maximum information about ecosystem evapotranspiration and thus minimize uncertainty in both evapotranspiration estimates and model structure. The simplest optimization principle, WUE, was the most adequate for ecosystem applications (Fig. 2), with better-constrained *parameterization* that characterize realistic behavior of herbaceous and woody PFTs, and higher *predictive* and *functional* performance in reproducing the sensitivity of LE to both θ and D as well as the sensitivity of these interactions in different biomes and in wet to dry conditions. In addition, WUE was the principle with the best tradeoff between prediction and parsimony (lowest AIC) and was the most robust in dry conditions (Fig. S5).

Water penalty functions based on xylem vulnerability did not substantially improve predictive or functional accuracy compared to water use efficiency and are also potentially more difficult to parameterize and apply at ecosystem scales. WUE, despite lacking an explicit representation of plant hydraulics, provided physiologically meaningful parameterizations with lower uncertainty than marginal xylem tension efficiency in CM and higher predictive and functional performance than SOX. WUE, including the exponential dependency of $\partial\Theta/\partial T$ on predawn ψ , is mathematically more consistent with maximization of cumulative carbon over a certain time period (Cowan, 1982), while CM and SOX assume instantaneous

carbon maximization. The higher functional accuracy of WUE may indicate that the associated $\partial\Theta/\partial T$ functional form represents some longer-term ecosystem optimization reflected in observations (Lu *et al.*, 2020), which are not captured by CM and SOX.

As such, stomatal optimization based on xylem vulnerability had limited utility in improving modeled interactions between θ , D , and LE at ecosystem scales. Nevertheless, explicit representation of xylem vulnerability provides information for other research goals, including modeling ecosystem adaptation and mortality (Sperry *et al.*, 2019).

Insights for ecosystem-scale stomatal optimization modelling

Our results complement those of previous empirical, numerical, and theoretical analyses of stomatal optimization models, providing insight into which is most robust for ESMs. While all $\partial\Theta/\partial T$ formulations performed well, our results raised caution towards formulations based on plant hydraulic traits vs water use efficiency in generalized and larger-scale estimation of evapotranspiration. The physiologically grounded functional relations based on measurable plant-scale parameters can lose their relevance at larger scales; thus, their utility in formulating water penalty functions for ESMs remains empirical and uncertain.

Cowan & Farquhar's (1977) original optimality theory, in which $\partial\Theta/\partial T$ should be solved dynamically from boundary conditions without fitting parameters (Cowan, 1982; Manzoni *et al.*, 2013b; Lu *et al.*, 2016; Mrad *et al.*, 2019), is challenging to approximate into practical analytical equations for stomatal conductance (Buckley *et al.*, 2017). Hence, the ability of $g_{s,opt}$ to capture these long-term dynamics depends on the adequacy of the empirical dependence of $\partial\Theta/\partial T$ on ψ . Our results highlight that this problem, at ecosystem scales, remains unsolved by water penalty functions, requiring an increasing number of parameters to characterize plant hydraulic strategies with increasing detail.

Occam's razor is particularly relevant to ecosystem-scale modelling of evapotranspiration, when parameters cannot be measured directly and may be less physically meaningful when aggregated at large scales. While there is a demand for increasingly complex models to assess climate change effects on Earth systems, parameterization and model evaluation is limited by a lack of adequate data (Feng, 2020; Paschalis *et al.*, 2020). An adequate degree of parsimony is required to exploit satellite data and extract ecosystem-scale ecohydrological parameters through inverse modeling (Bassiouni *et al.*, 2020), which may provide further opportunities to parameterize stomatal optimization models at large scales and improve ESMs.

The fundamental optimality theory, whether based on water use efficiency or xylem vulnerability, is consistent with data at all scales. Yet, until our approximations (parametrizations) of the theory are accurately translated from leaf or plant to ecosystems, simpler formulations with equivalent performance are preferred according to the principle of Occam's razor.

Acknowledgements

We thank four anonymous reviewers and Stefano Manzoni for feedback on the manuscript. We thank Grey Nearing for discussion about model inter-comparisons using information theory. We thank the European Commission and the Swedish Research Council for Sustainable Development FORMAS (grant no. 2018-02787) for funding in the frame of the collaborative international-consortium iAquaduct financed under the 2018 Joint call of the WaterWorks 2017 ERA-NET Cofund. This ERA-NET is an integral part of the activities developed by the Water JPI. This study was also supported by funding from FORMAS (grant nos. 2018-01820 and 2018-00968) and the Swedish government through the project Trees and Crops for the Future (TC4F). This work used data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia and USCCC. The FLUXNET eddy covariance data processing and harmonization were carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC and the ICOS Ecosystem Thematic Center and the OzFlux, ChinaFlux and AsiaFlux offices.

Author contributions

MB led the study conception, modeling, and data analysis, and wrote the manuscript with input at all stages from GV.

ORCID

Maoya Bassiouni  <https://orcid.org/0000-0001-5795-9894>
Giulia Vico  <https://orcid.org/0000-0002-7849-2653>

Data availability

Code and data that support the findings in this study are openly available. The scripts in the PYTHON computing language necessary to repeat all analysis can be found in Bassiouni (2020); <http://doi.org/10.5281/zenodo.4386353>. The FLUXNET2015 data products can be found at <https://fluxnet.org/data/fluxnet2015-dataset/>.

References

- Akaike H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–723.
- Anderegg WRL, Konings AG, Trugman AT, Yu K, Bowling DR, Gabbitas R, Karp DS, Pacala S, Sperry JS, Sulman BN *et al.* 2018a. Hydraulic diversity of forests regulates ecosystem resilience during drought. *Nature* 561: 538–541.
- Anderegg WRL, Wolf A, Arango-Velez A, Choat B, Chmura DJ, Jansen S, Kolb T, Li S, Meinzer FC, Pita P *et al.* 2018b. Woody plants optimise stomatal behaviour relative to hydraulic risk. *Ecology Letters* 21: 968–977.
- Ball JT, Woodrow IE, Berry JA. 1987. A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions. In: Biggins J, ed. *Progress in photosynthesis research: vol. 4. Proceedings of the VIIth international congress on photosynthesis*.

- providence, Rhode Island, USA, 10–15 August 1986. Dordrecht, the Netherlands: Springer, 221–224.
- Bassiouni M. 2020. Diagnostics of ecosystem-scale stomatal optimization models. *Zenodo*. doi: 10.5281/zenodo.4386353.
- Bassiouni M, Good SP, Still CJ, Higgins CW. 2020. Plant water uptake thresholds inferred from satellite soil moisture. *Geophysical Research Letters* 47: e2020GL087077.
- Bonan G. 2019. Climate change and terrestrial ecosystem modeling. Cambridge, UK: Cambridge University Press, 354.
- Brooks RH, Corey AT. 1964. Hydraulic properties of porous media. Fort Collins, CO, USA: Colorado State University.
- Buckley TN. 2017. Modeling stomatal conductance. *Plant Physiology* 174: 572–582.
- Buckley TN, Sack L, Farquhar GD. 2017. Optimal plant water economy. *Plant, Cell & Environment* 40: 881–896.
- Cover TM, Thomas JA. 2012. *Elements of information theory*. New York, NY, USA: John Wiley & Sons.
- Cowan IR. 1982. Regulation of water use in relation to carbon gain in higher plants. In: Lange OL, Nobel PS, Osmond CB, Ziegler H, ed. *Encyclopedia of plant physiology. Physiological plant ecology II: Water relations and carbon assimilation*. Berlin/Heidelberg: Springer, 589–613.
- Cowan IR, Farquhar GD. 1977. Stomatal function in relation to leaf metabolism and environment. *Symposia of the Society for Experimental Biology* 31: 471–505.
- De Kauwe MG, Medlyn BE, Zaehle S, Walker AP, Dietze MC, Hickler T, Jain AK, Luo Y, Parton WJ, Prentice IC *et al.* 2013. Forest water use and water use efficiency at elevated CO₂: a model-data intercomparison at two contrasting temperate forest FACE sites. *Global Change Biology* 19: 1759–1779.
- Dewar R, Mauranen A, Mäkelä A, Hölttä T, Medlyn B, Vesala T. 2018. New insights into the covariation of stomatal, mesophyll and hydraulic conductances from optimization models incorporating nonstomatal limitations to photosynthesis. *New Phytologist* 217: 571–585.
- Eller CB, Rowland L, Oliveira RS, Bittencourt PRL, Barros FV, da Costa ACL, Meir P, Friend AD, Mencuccini M, Sitch S *et al.* 2018. Modelling tropical forest responses to drought and El Niño with a stomatal optimization model based on xylem hydraulics. *hydraulic traits retrieved using London Series B: Biological Sciences* 373: 20170315.
- Eller CB, Rowland L, Mencuccini M, Rosas T, Williams K, Harper A, Medlyn BE, Wagner Y, Klein T, Teodoro GS *et al.* 2020. Stomatal optimization based on xylem hydraulics (SOX) improves land surface model simulation of vegetation responses to climate. *New Phytologist* 226: 1622–1637.
- Fatichi S, Pappas C, Ivanov VY. 2016. Modeling plant–water interactions: an ecohydrological overview from the cell to the global scale. *WIREs Water* 3: 327–368.
- Feddes RA, Kowalik PJ, Zaradny H. 1978. *Simulation of field water use and crop yield*. Simulation monographs. Wageningen, the Netherlands: Halsted Press.
- Feng X. 2020. Marching in step: The importance of matching model complexity to data availability in terrestrial biosphere models. *Global Change Biology* 26: 3190–3192.
- Friedlingstein P, Jones MW, O'Sullivan M, Andrew RM, Hauck J, Peters GP, Peters W, Pongratz J, Sitch S, Quéré CL *et al.* 2019. Global carbon budget 2019. *Earth System Science Data* 11: 1783–1838.
- Friedlingstein P, Meinshausen M, Arora VK, Jones CD, Anav A, Liddicoat SK, Knutti R. 2014. Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate* 27: 511–526.
- Gentine P, Green JK, Guérin M, Humphrey V, Seneviratne SI, Zhang Y, Zhou S. 2019. Coupling between the terrestrial carbon and water cycles—a review. *Environmental Research Letters* 14: 83003.
- Goldstein AH, Hultman NE, Fracheboud JM, Bauer MR, Panek JA, Xu M, Qi Y, Guenther AB, Baugh W. 2000. Effects of climate variability on the carbon dioxide, water, and sensible heat fluxes above a ponderosa pine plantation in the Sierra Nevada (CA). *Agricultural and Forest Meteorology* 101: 113–129.
- Good SP, Noone D, Bowen G. 2015. Hydrologic connectivity constrains partitioning of global terrestrial water fluxes. *Science* 349: 175–177.
- Goodwell AE, Jiang P, Ruddell BL, Kumar P. 2020. Debates—Does information theory provide a new paradigm for earth science? Causality, interaction, and feedback. *Water Resources Research* 56: e2019WR024940.
- Goodwell AE, Kumar P. 2017. Temporal information partitioning: characterizing synergy, uniqueness, and redundancy in interacting environmental variables. *Water Resources Research* 53: 5920–5942.
- Goodwell AE, Kumar P, Fellows AW, Flerchinger GN. 2018. Dynamic process connectivity explains ecohydrologic responses to rainfall pulses and drought. *Proceedings of the National Academy of Sciences, USA* 115: E8604–E8613.
- Katul GG, Palmroth S, Oren R. 2009. Leaf stomatal responses to vapour pressure deficit under current and CO₂-enriched atmosphere explained by the economics of gas exchange. *Plant, Cell & Environment* 32: 968–979.
- Knauer J, Zaehle S, Medlyn BE, Reichstein M, Williams CA, Migliavacca M, De Kauwe MG, Werner C, Keitel C, Kolari P *et al.* 2018. Towards physiologically meaningful water-use efficiency estimates from eddy covariance data. *Global Change Biology* 24: 694–710.
- Lavergne A, Graven H, Kauwe MGD, Keenan TF, Medlyn BE, Prentice IC. 2019. Observed and modelled historical trends in the water-use efficiency of plants and ecosystems. *Global Change Biology* 25: 2242–2257.
- Leuning R, Kelliher FM, Pury DGG, Schulze E-D. 1995. Leaf nitrogen, photosynthesis, conductance and transpiration: scaling from leaves to canopies. *Plant, Cell & Environment* 18: 1183–1200.
- Li X, Gentine P, Lin C, Zhou S, Sun Z, Zheng Y, Liu J, Zheng C. 2019. A simple and objective method to partition evapotranspiration into transpiration and evaporation at eddy-covariance sites. *Agricultural and Forest Meteorology* 265: 171–182.
- Lin C, Gentine P, Huang Y, Guan K, Kimm H, Zhou S. 2018. Diel ecosystem conductance response to vapor pressure deficit is suboptimal and independent of soil moisture. *Agricultural and Forest Meteorology* 250–251: 24–34.
- Lin Y-S, Medlyn BE, Duursma RA, Prentice IC, Wang H, Baig S, Eamus D, de Dios V, Mitchell P, Ellsworth DS *et al.* 2015. Optimal stomatal behaviour around the world. *Nature Climate Change* 5: 459–464.
- Liu Y, Kumar M, Katul GG, Feng X, Konings AG. 2020a. Plant hydraulics accentuates the effect of atmospheric moisture stress on transpiration. *Nature Climate Change* 10: 691–695.
- Liu Y, Holtzman NM, Konings AG. 2020b. Global ecosystem-scale plant hydraulic traits retrieved using model-data fusion. *Hydrology and Earth System Sciences Discussions*. doi: 10.5194/hess-2020-649.
- Lu Y, Duursma RA, Medlyn BE. 2016. Optimal stomatal behaviour under stochastic rainfall. *Journal of Theoretical Biology* 394: 160–171.
- Lu Y, Duursma RA, Farrior CE, Medlyn BE, Feng X. 2020. Optimal stomatal drought response shaped by competition for water and hydraulic risk can explain plant trait covariation. *New Phytologist* 225: 1206–1217.
- Manzoni S, Vico G, Katul G, Fay PA, Polley W, Palmroth S, Porporato A. 2011. Optimizing stomatal conductance for maximum carbon gain under water stress: a meta-analysis across plant functional types and climates: optimal leaf gas exchange under water stress. *Functional Ecology* 25: 456–467.
- Manzoni S, Vico G, Katul G, Palmroth S, Jackson RB, Porporato A. 2013a. Hydraulic limits on maximum plant transpiration and the emergence of the safety-efficiency trade-off. *New Phytologist* 198: 169–178.
- Manzoni S, Vico G, Palmroth S, Porporato A, Katul G. 2013b. Optimization of stomatal conductance for maximum carbon gain under dynamic soil moisture. *Advances in Water Resources* 62: 90–105.
- Medlyn BE, De Kauwe MG, Lin Y-S, Knauer J, Duursma RA, Williams CA, Arneth A, Clement R, Isaac P, Limousin J-M *et al.* 2017. How do leaf and ecosystem measures of water-use efficiency compare? *New Phytologist* 216: 758–770.
- Medlyn BE, Duursma RA, Eamus D, Ellsworth DS, Prentice IC, Barton CVM, Crous KY, Angelis PD, Freeman M, Wingate L. 2011. Reconciling the optimal and empirical approaches to modelling stomatal conductance. *Global Change Biology* 17: 2134–2144.
- Mencuccini M, Manzoni S, Christoffersen B. 2019a. Modelling water fluxes in plants: from tissues to biosphere. *New Phytologist* 222: 1207–1222.
- Mencuccini M, Rosas T, Rowland L, Choat B, Cornelissen H, Jansen S, Kramer K, Lapenis A, Manzoni S, Niinemets Ü *et al.* 2019b. Leaf economics and plant hydraulics drive leaf : wood area ratios. *New Phytologist* 224: 1544–1556.
- Monteith JL. 1965. Evaporation and environment. *Symposia of the Society for Experimental Biology* 19: 205–234.

- Mrad A, Sevanto S, Domec J-C, Liu Y, Nakad M, Katul G. 2019. A dynamic optimality principle for water use strategies explains isohydric to anisohydric plant responses to drought. *Frontiers in Forests and Global Change* 2: 49.
- Nearing GS, Gupta HV. 2015. The quantity and quality of information in hydrologic models. *Water Resources Research* 51: 524–538.
- Nearing GS, Ruddell BL, Clark MP, Nijssen B, Peters-Lidard C. 2018. Benchmarking and process diagnostics of land models. *Journal of Hydrometeorology* 19: 1835–1852.
- Nearing GS, Ruddell BL, Bennett AR, Prieto C, Gupta HV. 2020. Does information theory provide a new paradigm for earth science? Hypothesis testing. *Water Resources Research* 56: e2019WR024918.
- Newville M, Stensitzki T, Allen DB, Ingargiola A. 2014. LMFIT: non-linear least-square minimization and curve-fitting for python. *Zenodo*. doi: 10.5281/zenodo.598352.
- Novick KA, Ficklin DL, Stoy PC, Williams CA, Bohrer G, Oishi A, Papuga SA, Blanken PD, Noormets A, Sulman BN *et al.* 2016. The increasing importance of atmospheric demand for ecosystem water and carbon fluxes. *Nature Climate Change* 6: 1023–1027.
- Or D, Lehmann P, Shahraeeni E, Shokri N. 2013. Advances in soil evaporation physics—a review. *Vadose Zone Journal* 12: vzj2012.0163.
- Pappas C, Fatichi S, Burlando P. 2016. Modeling terrestrial carbon and water dynamics across climatic gradients: does plant trait diversity matter? *New Phytologist* 209: 137–151.
- Paschalis A, Fatichi S, Zscheischler J, Ciais P, Bahn M, Boysen L, Chang J, De Kauwe M, Estiarte M, Goll D *et al.* 2020. Rainfall manipulation experiments as simulated by terrestrial biosphere models: Where do we stand? *Global Change Biology* 26: 3336–3355.
- Pastorello G, Trotta C, Canfora E, Chu H, Christianson D, Cheah Y-W, Poindexter C, Chen J, Elbashandy A, Humphrey M *et al.* 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data* 7: 225.
- Penman HL, Keen BA. 1948. Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London Series A: Mathematical and Physical Sciences* 193: 120–145.
- Rogers A, Medlyn BE, Dukes JS, Bonan G, von Caemmerer S, Dietze MC, Kattge J, Leakey ADB, Mercado LM, Niinemets Ü *et al.* 2017. A roadmap for improving the representation of photosynthesis in Earth system models. *New Phytologist* 213: 22–42.
- Ruddell BL, Drewry DT, Nearing GS. 2019. Information theory for model diagnostics: structural error is indicated by trade-off between functional and predictive performance. *Water Resources Research* 55: 6534–6554.
- Ruddell BL, Kumar P. 2009. Ecohydrologic process networks: 2. Analysis and characterization. *Water Resources Research* 45: 3.
- Sabot MEB, Kauwe MGD, Pitman AJ, Medlyn BE, Verhoef A, Ukkola AM, Abramowitz G. 2020. Plant profit maximization improves predictions of European forest responses to drought. *New Phytologist* 226: 1638–1655.
- Sperry JS, Venturas MD, Anderegg WRL, Mencuccini M, Mackay DS, Wang Y, Love DM. 2017. Predicting stomatal responses to the environment from the optimization of photosynthetic gain and hydraulic cost. *Plant, Cell & Environment* 40: 816–830.
- Sperry JS, Venturas MD, Todd HN, Trugman AT, Anderegg WRL, Wang Y, Tai X. 2019. The impact of rising CO₂ and acclimation on the response of US forests to global warming. *Proceedings of the National Academy of Sciences, USA* 116: 25734–25744.
- Vico G, Manzoni S, Palmroth S, Weih M, Katul G. 2013. A perspective on optimal leaf stomatal conductance under CO₂ and light co-limitations. *Agricultural and Forest Meteorology* 182–183: 191–199.
- Wang Y, Sperry JS, Anderegg WRL, Venturas MD, Trugman AT. 2020. A theoretical and empirical assessment of stomatal optimization modeling. *New Phytologist* 227: 311–325.
- Weijs SV, Ruddell BL. 2020. Debates: does information theory provide a new paradigm for earth science? Sharper predictions using Occam's digital razor. *Water Resources Research* 56: e2019WR026471.
- Weijs SV, Schoups G, van de Giesen N. 2010. Why hydrological predictions should be evaluated using information theory. *Hydrology and Earth System Sciences* 14: 2545–2558.
- Williams PL, Beer RD. 2010. *Nonnegative decomposition of multivariate information*. <https://arxiv.org/abs/1102.1507v1>. [accessed 1 January 2020].
- Wolf A, Anderegg WRL, Pacala SW. 2016. Optimal stomatal behavior with competition for water and risk of hydraulic impairment. *Proceedings of the National Academy of Sciences, USA* 113: E7222–E7230.
- Zenes N, Kerr KL, Trugman AT, Anderegg WRL. 2020. Competition and drought alter optimal stomatal strategy in tree seedlings. *Frontiers in Plant Science* 11: 478.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Performance is higher for model variants using predawn vs canopy water potential in the big-leaf comparison framework.

Fig. S2 Average soil-to-surface conductance ratios for the study sites.

Fig. S3 Canopy-to-leaf conductance ratios averaged by day of year at the study sites.

Fig. S4 Performance is higher for model variants using photosynthesis parameters at 25°C vs those with temperature dependency in the big-leaf comparison framework.

Fig. S5 Model ranking and performance scores for all diagnostics and soil moisture conditions.

Fig. S6 Performance scores for five stomatal optimization model variants.

Methods S1 Data selection criteria.

Methods S2 Photosynthesis model and comparison framework assumptions.

Methods S3 Information theory metrics.

Methods S4 Information partitioning.

Methods S5 Model performance diagnostics based on information theory.

Table S1 Selected FLUXNET sites and data characteristics.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.