

LETTER • OPEN ACCESS

## Hybrid modeling of evapotranspiration: inferring stomatal and aerodynamic resistances using combined physics-based and machine learning

To cite this article: RedaReda ElGhawi *et al* 2023 *Environ. Res. Lett.* **18** 034039

View the [article online](#) for updates and enhancements.

You may also like

- [Shift in controlling factors of carbon stocks across biomes on the Qinghai-Tibetan Plateau](#)  
Daorui Han, Zhongmin Hu, Xuhui Wang et al.
- [Responses of land evapotranspiration to Earth's greening in CMIP5 Earth System Models](#)  
Zhenzhong Zeng, Zaichun Zhu, Xu Lian et al.
- [Increasing tropical cyclone intensity and potential intensity in the subtropical Atlantic around Bermuda from an ocean heat content perspective 1955–2019](#)  
Samantha Hallam, Mark Guishard, Simon A Josey et al.



## Breath Biopsy<sup>®</sup> OMNI<sup>®</sup>

The most advanced, complete solution for global breath biomarker analysis

TRANSFORM YOUR RESEARCH WORKFLOW



Expert Study Design & Management



Robust Breath Collection



Reliable Sample Processing & Analysis



In-depth Data Analysis



Specialist Data Interpretation

ENVIRONMENTAL RESEARCH  
LETTERS

## LETTER

## Hybrid modeling of evapotranspiration: inferring stomatal and aerodynamic resistances using combined physics-based and machine learning

## OPEN ACCESS

RECEIVED  
14 September 2022REVISED  
23 January 2023ACCEPTED FOR PUBLICATION  
14 February 2023PUBLISHED  
7 March 2023

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

RedaReda ElGhawi<sup>1,2,3,\*</sup> , Basil Kraft<sup>1</sup>, Christian Reimers<sup>1</sup>, Markus Reichstein<sup>1,4</sup> , Marco Körner<sup>3</sup> , Pierre Gentine<sup>5</sup> and Alexander J WinklerWinkler<sup>1,\*</sup> <sup>1</sup> Max Planck Institute for Biogeochemistry, Biogeochemical Integration, Jena, Germany<sup>2</sup> International Max Planck Research School for Global Biogeochemical Cycles, Max Planck Institute for Biogeochemistry, Jena, Germany<sup>3</sup> Department of Aerospace and Geodesy, School of Engineering and Design, Technical University of Munich, Munich, Germany<sup>4</sup> ELLIS Unit Jena, Michael-Stifel-Center, University of Jena, Jena, Germany<sup>5</sup> Department of Earth and Environmental Engineering, Columbia University, NY, New York, 10027, United States of America

\* Authors to whom any correspondence should be addressed.

E-mail: [relghawi@bgc-jena.mpg.de](mailto:relghawi@bgc-jena.mpg.de) and [awinkler@bgc-jena.mpg.de](mailto:awinkler@bgc-jena.mpg.de)**Keywords:** hybrid modeling, physics-constrained, machine learning, multi-task learning, evapotranspiration, surface resistance, aerodynamic resistanceSupplementary material for this article is available [online](#)**Abstract**

The process of evapotranspiration transfers liquid water from vegetation and soil surfaces to the atmosphere, the so-called latent heat flux ( $Q_{LE}$ ), and modulates the Earth's energy, water, and carbon cycle. Vegetation controls  $Q_{LE}$  by regulating leaf stomata opening (surface resistance  $r_s$  in the Big Leaf approach) and by altering surface roughness (aerodynamic resistance  $r_a$ ). Estimating  $r_s$  and  $r_a$  across different vegetation types is a key challenge in predicting  $Q_{LE}$ . We propose a hybrid approach that combines mechanistic modeling and machine learning for modeling  $Q_{LE}$ . The hybrid model combines a feed-forward neural network which estimates the resistances from observations as intermediate variables and a mechanistic model in an end-to-end setting. In the hybrid modeling setup, we make use of the Penman–Monteith equation in conjunction with multi-year flux measurements across different forest and grassland sites from the FLUXNET database. This hybrid model setup is successful in predicting  $Q_{LE}$ , however, this approach leads to equifinal solutions in terms of estimated physical parameters. We follow two different strategies to constrain the hybrid model and therefore control for the equifinality that arises when the two resistances are estimated simultaneously. One strategy is to impose an *a priori* constraint on  $r_a$  based on mechanistic assumptions (theory-driven strategy), while the other strategy makes use of more observational data and adds a constraint in predicting  $r_a$  through multi-task learning of both latent and sensible heat flux ( $Q_H$ ; data-driven strategy) together. Our results show that all hybrid models predict the target variables with a high degree of success, with  $R^2 = 0.82$ – $0.89$  for grasslands and  $R^2 = 0.70$ – $0.80$  for forest sites at the mean diurnal scale. The predicted  $r_s$  and  $r_a$  show strong physical consistency across the two regularized hybrid models, but are physically implausible in the under-constrained hybrid model. The hybrid models are robust in reproducing consistent results for energy fluxes and resistances across different scales (diurnal, seasonal, and interannual), reflecting their ability to learn the physical dependence of the target variables on the meteorological inputs. As a next step, we propose to test these heavily observation-informed parameterizations derived through hybrid modeling as a substitute for ad hoc formulations in Earth system models.

**1. Introduction**

Evapotranspiration, i.e. surface latent heat flux ( $Q_{LE}$ ), plays a key role in driving Earth's energy, water, and

carbon cycles, and is primarily controlled by dynamic meteorological conditions and soil water conditions, as well as static properties such as soil characteristics and plant traits (Jung *et al* 2010, Dou and Yang

2018, Ajami 2021). Plants critically influence  $Q_{LE}$  mainly through their direct control of transpiration, but also through shaping aerodynamic surface properties (i.e. roughness). They use their leaf stomata to dynamically regulate the water loss to the atmosphere, which depends not only on atmospheric water demand, but also on soil water availability (Damour *et al* 2010, Kennedy *et al* 2019, Carminati and Javaux 2020). While the physical drivers that cause water to evaporate are well described and understood, the influence of biological control on  $Q_{LE}$ , mainly the transpirative water flux, is more difficult to assess. As a consequence, empirical formulations, especially for surface ( $r_s$ ) and aerodynamic resistance ( $r_a$ ), remain used in process-based models, which can lead to large uncertainties in predicting  $Q_{LE}$  (Polhamus *et al* 2013). Most formulations of  $r_s$  are empirical or rely on optimality concepts, such as minimizing water loss while maximizing carbon assimilation (e.g. Tan *et al* 2021). As such, these concepts do not take into account the active transpiration mechanism that some plants use to down-regulate leaf temperature through evaporative cooling to prevent leaf overheating at high irradiance and air temperature (Lin *et al* 2017, Drake *et al* 2018). Other empirical approaches, e.g. the Jarvis–Stewart formulation, Ball–Berry model, and Leuning model aim to derive parametrizations based on statistical correlations between  $r_s$  (or canopy resistance) and the key environmental variables (Jarvis 1976, Stewart 1988, Leuning *et al* 1991, Leuning 1995).

These ad hoc formulations have several drawbacks, e.g. they are considered too rigid, especially when evaluated in a coupled system of atmosphere–biosphere feedbacks where some of the environmental variables are actually also a function of  $r_s$  (Ronda *et al* 2001). Overall, these empirical representations for  $r_s$  and  $r_a$  in deterministic models for  $Q_{LE}$  obey physical laws and phenomenological behavior (Krasnopol'sky 2013, de Bezenac *et al* 2017). Yet, they exhibit limited capability to adapt to other or changing vegetation composition or long-term climatic conditions, especially with respect to soil moisture (Damour *et al* 2010, Medlyn *et al* 2011, Kennedy *et al* 2019).

Statistical models have been proposed as alternative approaches to reliably estimate  $Q_{LE}$  due to their data-adaptiveness (Tramontana *et al* 2016, Dou and Yang 2018, Carter and Liang 2019). In particular, approaches that use machine learning (ML) techniques are gaining traction because they can implicitly learn unknown latent processes and constitute a more complete statistical representation of the processes that influence  $Q_{LE}$  at different scales in space and time (Jung *et al* 2009, 2020, Dou and Yang 2018). However, these data-driven models have several drawbacks, such as the need for large amounts of high-quality data, their limited

physical consistency, and their lack of mechanistic interpretability (Karpatne *et al* 2017a, 2017b).

Accordingly, physics-based models are restricted by the ad hoc assumptions of the system, and ML models are limited by their inability to produce physically interpretable and consistent predictions. Therefore, the combination of mechanistic and ML modeling promises physically interpretable performance of predicting and inferring intermediate (or latent) states and variables by merging the advantages of the causal understanding of physics-based models and the predictive power of ML. Different approaches have been proposed to circumvent the issues originating from using pure physics-based and ML models. They combine the complementary strengths of both techniques, which enables ML models to capture dynamic patterns and improve the accuracy and physical interpretability of predictions. These methods include a form of physics-guided ML techniques, where the neural network (NN) is constrained by different means to produce predictions that mirror realistic climate conditions and fluxes. The physics-guided ML approaches can be generally subdivided into physics-guided loss functions, initialization, architecture design, and hybrid modeling (Karpatne *et al* 2017a, Reichstein *et al* 2019, 2022, Jia *et al* 2020a, Willard *et al* 2020). The combination of ML and mechanistic modeling, here denoted hybrid modeling, makes it possible to combine the strengths of both techniques: ensure physical consistency while efficiently harvesting the growing resource of observational data (Reichstein *et al* 2019, 2022). Therefore, the synergy of both techniques offers promising solutions to the shortcomings encountered in using both techniques separately. Several studies have successfully applied hybrid modeling in hydrological applications, such as the characterization of the different known and unknown variables governing the global water cycle (Kraft *et al* 2020, 2022), simulation of lake temperature dynamics (Jia *et al* 2020b), and the modeling of global extreme flooding events (Yang *et al* 2019). Other studies focusing on land–atmosphere interactions of ecosystem fluxes, such as  $Q_{LE}$  (Zhao *et al* 2019), showed that these hybrid approaches allow for better extrapolation and generalization capabilities during extreme conditions.

In this study, we propose a hybrid modeling approach that allows the inference of these biophysical controls based on observational data of  $Q_{LE}$  across ecosystems, while adhering to known physical laws (Reichstein *et al* 2022). The aim of this study is to offer guidance on how to infer the hidden controls of land–atmosphere coupling from observational data using hybrid learning rather than from ad hoc assumptions with rigid parametrizations. The hybrid modeling approach illustrates the ability to provide physically interpretable and accurate predictions against observations at different temporal scales and ecosystems.

The obtained observation-informed parametrizations for  $r_s$  and  $r_a$  reveal variability across different vegetation canopy structures, which is unaccounted for by conventional parameterizations.

In the methods in sections 2.1 and 2.2, we describe the hybrid modeling approach and introduce different models of  $Q_{LE}$  using the Penman–Monteith (PM) equation (Penman 1948, Monteith 1965) and eddy covariance (EC) flux measurements from several grassland and forest sites (Baldocchi *et al* 2001, Li *et al* 2018). Our hybrid models should not only yield accurate predictions of  $Q_{LE}$ , but also enable us to better understand the functioning and influence of biophysical processes on  $Q_{LE}$  expressed through the surface and aerodynamic resistances. We present and explore the problem of equifinality in our setting (section 2.3.2) (i.e. different combinations of  $r_a$  and  $r_s$  may result in the same  $Q_{LE}$ ) and propose two conceptually different solutions (theory- versus data-driven) to this problem (section 2.3.3). We evaluate the predictions of our hybrid models for  $Q_{LE}$ ,  $r_a$  and  $r_s$  against purely statistical models as well as against established mechanistic models in section 3.

## 2. Methodology

In this section we describe the data pre-processing methods and different model setups used. Section 2.1 describes the data and processing. Section 2.2 defines the physics-based component of the hybrid model, and section 2.3 provides an overview of all models.

### 2.1. FLUXNET 2015 data

The flux network (FLUXNET; <https://fluxnet.org>), a global network of EC towers, provides estimates of energy, water, and carbon fluxes at the land surfaces across climate regimes and plant functional types (Baldocchi *et al* 2001, Li *et al* 2018). The measurements in the FLUXNET 2015 Tier 1 dataset are resolved at a half-hourly frequency. Following Reichstein *et al* (2005), we select only measured data and omit gap-filled data. Further, we restrict our analysis to energy-balance-corrected measurements, because the EC data do not satisfy the energy balance budget closure, which potentially introduces high uncertainty and systematic bias in our results (Wilson *et al* 2002). Daytime values are selected based on a threshold of sensible heat flux  $Q_H > 5 \text{ Wm}^{-2}$  and incoming short-wave radiation  $SW_{in} > 50 \text{ Wm}^{-2}$  to avoid stable boundary layer conditions following Lin *et al* (2018) and Li *et al* (2019). Only positive values are selected for the latent heat flux ( $Q_{LE}$ ), net radiation ( $R_n$ ), soil heat flux ( $Q_G$ ), and vapor pressure deficit (VPD) for daylight data according to Zhou *et al* (2016). Winter months between October and March are excluded to focus on surface heat fluxes when the vegetation is active (Zhao *et al* 2019). The FLUXNET

sites chosen include three forest and three grassland sites with varying climatic conditions and site characteristics (see table 1 in supplementary information).

### 2.2. The physically-based component: PM equation

Various process-based models exist for the estimation of  $Q_{LE}$ . They can be subdivided into energy, mass transfer-based methods, water balance methods, and aerodynamic methods (Brutsaert 2005, Zhao *et al* 2013). One prominent example is the PM equation (Penman 1948, Monteith 1965) that provides the theoretical basis for determining  $Q_{LE}$  and its response to changing climate and vegetation conditions (Monteith and Unsworth 2013). The estimation of  $Q_{LE}$  can be traced back to the model proposed by Penman (1948), which combines the energy balance and mass transfer approaches to estimate evaporation from open water surfaces. The model was later extended to vegetative surfaces (Monteith 1985, Monteith and Unsworth 2013, Viallet-Chabrand and Lawson 2019). The PM equation

$$Q_{LE} = \frac{s_c (R_n - Q_G) + \frac{\rho_a c_p (e_s - e_a)}{r_a}}{s_c + \gamma \left(1 + \frac{r_s}{r_a}\right)}, \quad (1)$$

describes the latent heat flux  $Q_{LE}$  ( $\text{Wm}^{-2}$ ), where  $R_n$  and  $Q_G$  are measured in ( $\text{Wm}^{-2}$ ),  $r_s$  and  $r_a$  are estimated in ( $\text{sm}^{-1}$ ),  $s_c$  is the slope of the saturation vapor pressure–temperature relationship ( $\text{kPa C}^{-1}$ ),  $e_s - e_a$  is the VPD of air ( $\text{kPa}$ ),  $\rho_a$  is the mean air density at constant pressure ( $\text{kg m}^{-3}$ ),  $c_p$  is the specific heat of dry air at constant pressure ( $1004.834 \text{ Jkg}^{-1} \text{ C}^{-1}$ ), and  $\gamma$  is the psychrometric constant ( $\text{kPa C}^{-1}$ ).

### 2.3. Overview of models

The following subsections present the different models used, which differ in their approach towards being more data- or theory-driven. Each subsection describes in detail the structure of and differences between the models. All models were randomly initialized and drawn from a uniform distribution.

#### 2.3.1. Inverted PM and pure ML model

The PM equation is considered to be physics-based, since core physiological and aerodynamic factors describe the evaporative process (Jain *et al* 2008). The equation highlights the relationship between evapotranspiration and surface conductance, which is regulated by the leaf stomata to minimize the water loss to the atmosphere (Hetherington and Woodward 2003, Damour *et al* 2010, Gerosa *et al* 2012). Different approaches model surface conductance at the leaf level with various success. The determination of surface conductance at the canopy scale, however, is even more challenging due to canopy heterogeneity and variability in microclimate within the canopy (Bonan *et al* 2011, Lin *et al* 2018). A common approach is to

invert the PM equation for  $r_s$  to obtain the bulk surface resistance and understand its variations

$$r_s = \frac{r_a s_c (R_n - Q_G) + \rho_a c_p (e_s - e_a) - r_a Q_{LE} (s_c + \gamma)}{\gamma Q_{LE}}, \quad (2)$$

assuming that aerodynamic resistance  $r_a$  is known; a strong assumption as we will revisit later. The inverted PM equation (PM Inv) is used to quantify canopy parameters and expresses the relative significance of advective and radiative energy for  $Q_{LE}$  as a function of the ratio of surface to aerodynamic resistance (Kelliher et al 1992, Köstner et al 1992, Zeppel and Eamus 2008, Zhang et al 2016).

The inversion of the PM equation, leads to highly unstable estimates of the resistances. Therefore, we restrict surface and aerodynamic resistance values derived using PM inversion and empirical formulations (Knauer et al 2018) based on intervals that are physically realistic ( $0-2000 \text{ sm}^{-1}$  and  $0-500 \text{ sm}^{-1}$ , respectively).

The estimates for  $r_s$  from equation (2) derived through inverting the PM equation are referred to here as the PM Inv model. Values for  $r_a$  are estimated using the Big Leaf formulation from Knauer et al (2018), which calculates  $r_a$  as the sum of aerodynamic resistance for momentum ( $r_{am}$ ) and canopy boundary layer resistance for heat ( $r_{bh}$ )

$$r_{am} = WS/U^{*2}, \quad (3)$$

$$r_{bh} = 6.2U^{*-0.667}, \quad (4)$$

and

$$r_a = r_{am} + r_{bh}, \quad (5)$$

where  $WS$  is wind speed ( $\text{ms}^{-1}$ ) and  $U^*$  is friction velocity ( $\text{ms}^{-1}$ ). The PM Inv model represents a baseline physical model for comparison against pure data-driven models for  $Q_{LE}$ . The pure ML model for  $Q_{LE}$  is set up to evaluate predictions against hybrid models. The pure ML model consists of a feed-forward NN (FNN) (figure 1), and details about the hyperparameters of the model are found in table 2 of the supplementary information. The  $r_s$  is calculated from  $Q_{LE}$  predictions from the pure ML model by using PM Inv, and  $r_a$  is estimated using the ad hoc formulation (equation (5)) approach. This model is purely data-driven and does not contain any physical constraint regarding  $Q_{LE}$ .

### 2.3.2. Under-constrained hybrid model

The hybrid model estimates  $Q_{LE}$  using the PM equation (equation (1)), where the two intermediate variables  $r_s$  and  $r_a$  are estimated by two FNNs (figure 1). The variables used for predicting  $r_s$  are air temperature (TA), water availability index (WAI),

incoming shortwave radiation ( $SW_{in}$ ), mean incoming shortwave potential ( $SW_{pot \text{ sm}}$ ), VPD, and  $R_n$ . The WAI is calculated as the annual cumulative difference between  $Q_{LE}$  and precipitation ( $P$ ). The WAI at time  $t$  ( $WAI_t$ ) is calculated from the difference between  $Q_{LE_t}$  and  $P_t$  added to WAI at the previous time step ( $WAI_{t-1}$ )

$$WAI_t = P_t - Q_{LE_t} + WAI_{t-1}. \quad (6)$$

The variables for predicting  $r_a$  are  $WS$  and  $U^*$ . The input variables chosen for the latent variables  $r_s$  and  $r_a$  were selected based on variables included in the Big Leaf and PM equations and physical intuition and interpretability through manual tuning of parameters. The predictors are normalized using the mean and standard deviation of the training dataset. Thus, the hybrid model first predicts the intermediate (or latent) variables  $r_s$  and  $r_a$  and uses them to estimate  $Q_{LE}$  based on the PM equation. The hybrid model predicts  $Q_{LE}$  that exist between the initial input and resulting output phases in one step (Reimers and Requena-Mesa 2020). The loss function minimizes the difference between predicted and observed  $Q_{LE}$  and is defined as the mean absolute difference between the model predictions and observations with  $n$  sample size, and parameters  $\theta$  for  $r_s$  and  $r_a$

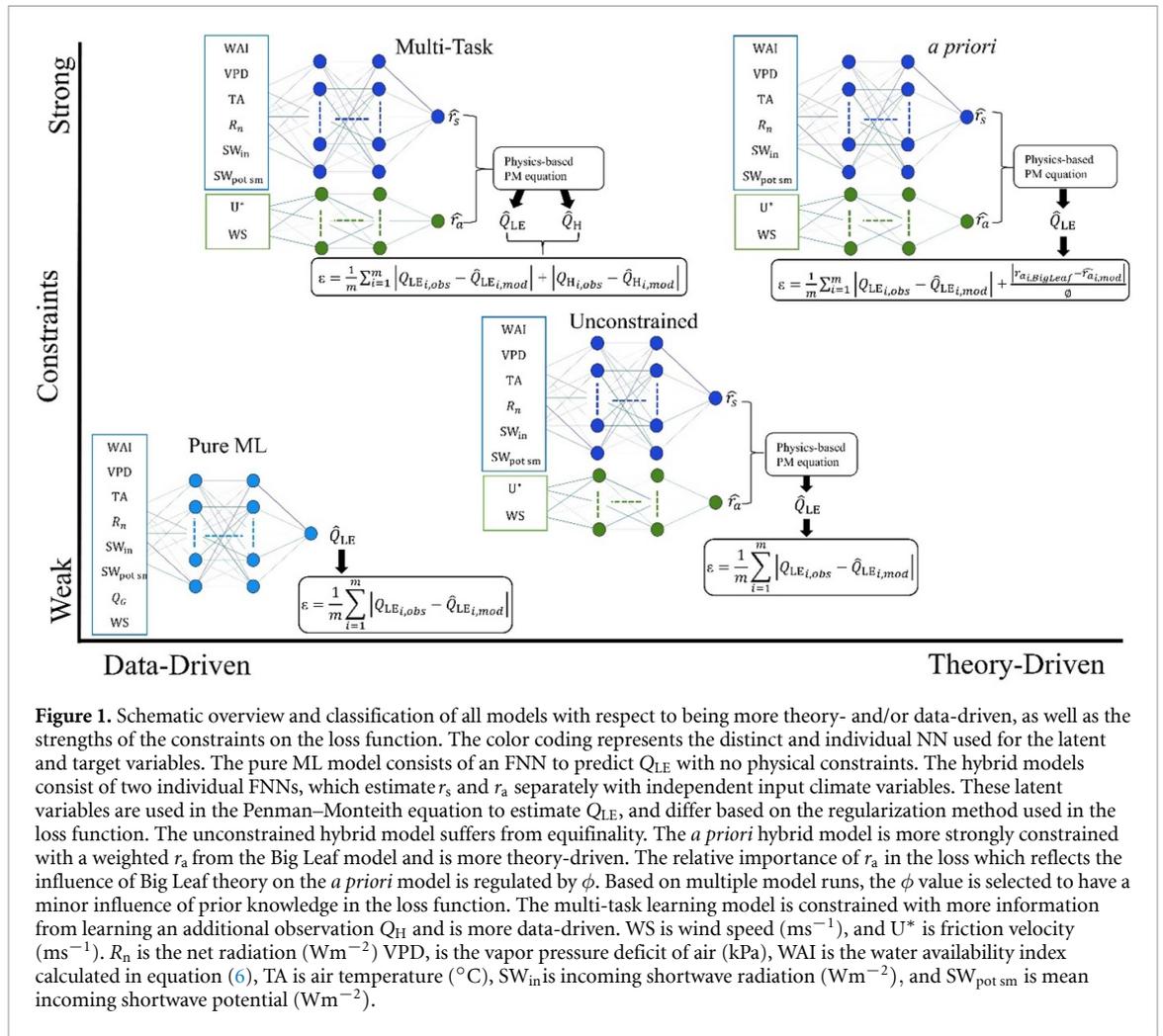
$$\min_{\theta_{r_a}, \theta_{r_s}} \sum_{i=1}^n \left| \hat{Q}_{LE_i} - Q_{LE_i} \right|. \quad (7)$$

We use the mean absolute error as it is less sensitive to outliers than the mean squared error. The pure ML model and the hybrid model both optimize against  $Q_{LE}$  as highlighted in the loss function (equation (7)), however, the hybrid model estimates  $r_s$  and  $r_a$  as intermediate variables and uses the PM equation to estimate  $Q_{LE}$ . While the pure ML model directly predicts  $Q_{LE}$  without the physical constraints imposed by the PM equation.

Although the two FNNs for  $r_a$  and  $r_s$  take different predictor variables, the hybrid model is under-constrained when simultaneously estimating the two intermediate variables using only one target  $Q_{LE}$ . The proposed hybrid model thus suffers from an equifinality problem. The issue of equifinality, or non-uniqueness, occurs when different model parametrization or structures result in equivalent representations of the system (Beven 2006, Schmidt et al 2020). Thus, many different combinations of  $r_s$  and  $r_a$  can result in the same  $Q_{LE}$  value (figure 2).

### 2.3.3. Constrained hybrid models: a priori and multi-task learning models

The identification and elimination of equifinality, in the physics-based component is one of the key challenges in hybrid modeling (Kraft et al 2022). One way to reduce equifinality is to restrict the parameter space



**Figure 1.** Schematic overview and classification of all models with respect to being more theory- and/or data-driven, as well as the strengths of the constraints on the loss function. The color coding represents the distinct and individual NN used for the latent and target variables. The pure ML model consists of an FNN to predict  $Q_{LE}$  with no physical constraints. The hybrid models consist of two individual FNNs, which estimate  $r_s$  and  $r_a$  separately with independent input climate variables. These latent variables are used in the Penman–Monteith equation to estimate  $Q_{LE}$ , and differ based on the regularization method used in the loss function. The unconstrained hybrid model suffers from equifinality. The *a priori* hybrid model is more strongly constrained with a weighted  $r_a$  from the Big Leaf model and is more theory-driven. The relative importance of  $r_a$  in the loss which reflects the influence of Big Leaf theory on the *a priori* model is regulated by  $\phi$ . Based on multiple model runs, the  $\phi$  value is selected to have a minor influence of prior knowledge in the loss function. The multi-task learning model is constrained with more information from learning an additional observation  $Q_H$  and is more data-driven. WS is wind speed ( $\text{ms}^{-1}$ ), and  $U^*$  is friction velocity ( $\text{ms}^{-1}$ ).  $R_n$  is the net radiation ( $\text{Wm}^{-2}$ ) VPD, is the vapor pressure deficit of air (kPa), WAI is the water availability index calculated in equation (6), TA is air temperature ( $^{\circ}\text{C}$ ),  $SW_{in}$  is incoming shortwave radiation ( $\text{Wm}^{-2}$ ), and  $SW_{pot\_sm}$  is mean incoming shortwave potential ( $\text{Wm}^{-2}$ ).

through model regularization. This can be achieved through two approaches; by including either additional theory or data via additional loss terms. The integration of *a priori* knowledge in the loss function (i.e. a regularization) induces an *a priori* constraint on  $r_a$  in the hybrid model (figure 1) based on the empirical formulation presented in equation (5), as the formulation for  $r_a$  is considered to be more robust than for  $r_s$ . To do so we regularize the loss function by adding a constraint on the loss minimizing aerodynamic resistance  $\text{Loss}(r_a, \hat{r}_a)/\phi$ . The relative importance of  $r_a$  in the new loss is regulated by  $\phi$ , which is varied between the high influence and low influence of the constraint. Based on multiple model runs with varying values for  $\phi$ , we select a value for  $\phi$  to only impose a low influence in the overall loss function.

Another way of restricting the parameter space is by extending the framework to model auxiliary target variables, whereby the auxiliary tasks help to regularize the problem objective (Liebel and Körner 2018). Since the sensible heat flux ( $Q_H$ ) is also dependent on the aerodynamic resistance  $r_a$ , we explore a multi-task learning approach by restricting the parameter

space through modeling  $Q_H$  and  $Q_{LE}$  simultaneously (figure 1). The estimation of  $Q_H$  is based on the resistance formulation

$$Q_H = \frac{\rho_a c_p (T_S - T_A)}{r_a}, \tag{8}$$

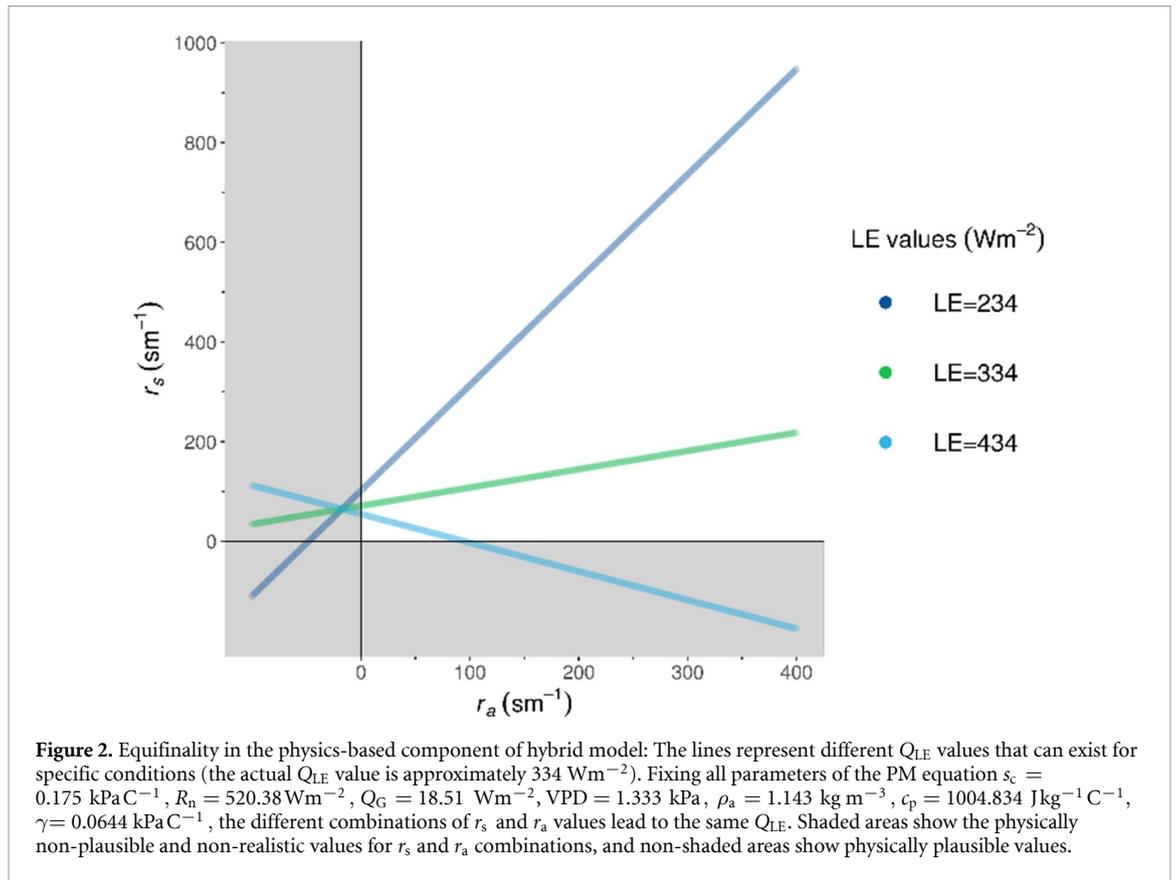
where  $T_S$  and  $T_A$  are surface and air temperature respectively. The  $T_S$  is estimated using the Stefan–Boltzmann equation

$$T_S = \sqrt[4]{\frac{Q_{LW_{out}}}{\sigma \epsilon}}, \tag{9}$$

where  $Q_{LW_{out}}$  is the outgoing longwave radiation ( $\text{Wm}^{-2}$ ),  $\sigma$  is the Stefan–Boltzmann constant ( $5.789 \times 10^{-8} \text{Wm}^{-2}\text{K}^{-4}$ ) and  $\epsilon$  is emissivity (dimensionless). The emissivity ranges from 0 to 1, and the values chosen were based on selecting models with the highest predictive accuracy.

### 2.4. Evaluation

We evaluate four models, i.e. one pure ML model, one under-constrained hybrid model (i.e. with no strategy

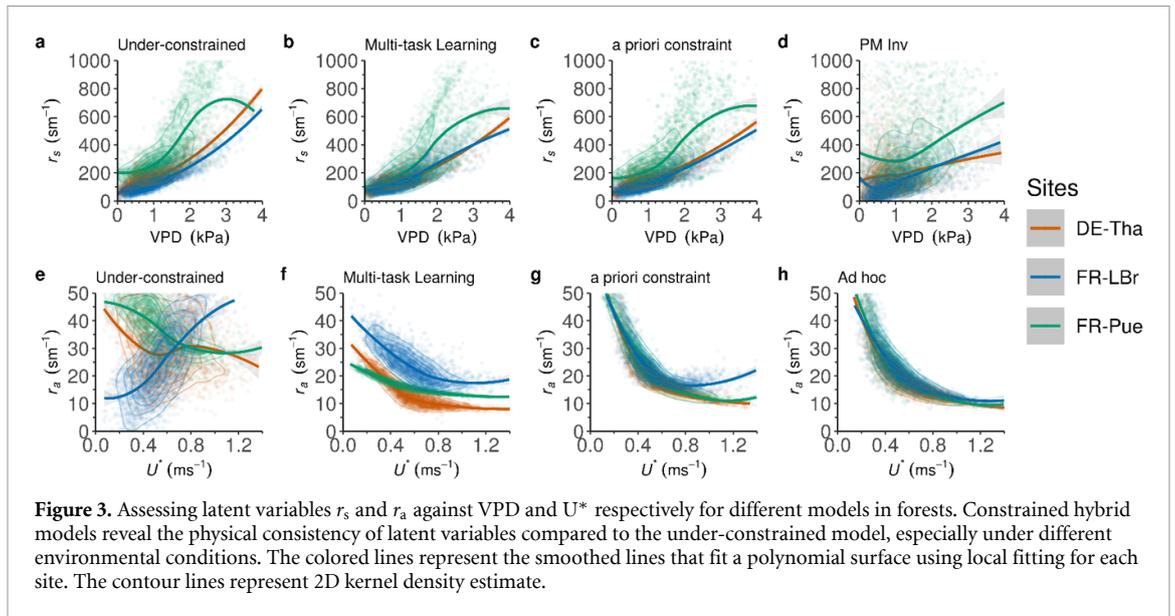


to decouple  $r_a$  and  $r_s$ ), and two constrained hybrid models. The constrained hybrid models either consist of an *a priori* constraint on  $r_a$  or use a multi-task learning approach. For a baseline comparison, we use a pure ML model predicting  $Q_{LE}$  directly without intermediate resistances and the estimation of the inverted PM equation to evaluate the predictions of the hybrid models. The network architectures and hyperparameters used are similar for the different models (table 2 in the supplementary information) for a fair comparison. Evaluation metrics such as the root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination ( $R^2$ ) are used to evaluate the model predictions. To highlight the impact of noise on model performance, we evaluate the model predictions on half-hourly and 7 d mean aggregated scales. The intermediate variables are assessed against the key meteorological predictor variables to scrutinize physical consistency and plausibility. The target variables are assessed against observations as well as the key meteorological predictor variables to estimate model performance and interpretability. We conduct five model runs with random initializations for each of the hybrid models and for one forest site (DE-Tha) as well as, one grassland site (DE-Gri) to evaluate model robustness at the mean diurnal scale. More information can be found in table 3 of the supplementary information.

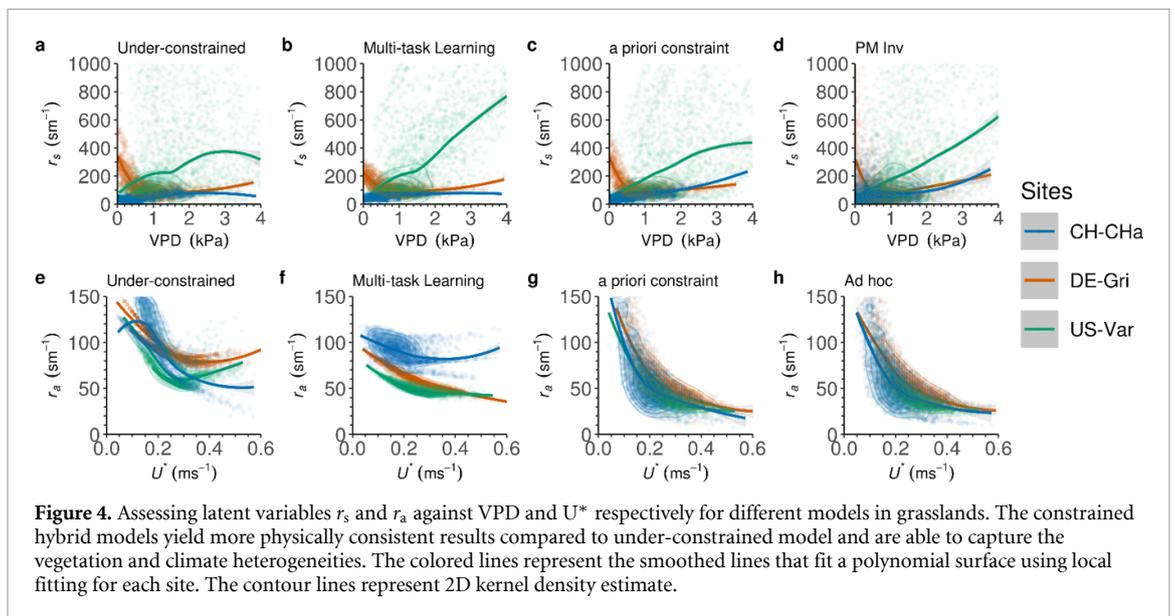
### 3. Results and discussion

#### 3.1. Evaluation of the learned latent variables $\hat{r}_s$ and $\hat{r}_a$

We evaluate the impact of the  $Q_{LE}$ -controlling resistances  $\hat{r}_s$  and  $\hat{r}_a$  which are treated as intermediate variables in our hybrid approach. Note that the models are driven by all relevant predictor variables for  $r_s$  and  $r_a$ , respectively (cf figure 1), and only for evaluation and interpretability purposes do we plot predictions against individual predictor variables. Based on the strongest dependencies discovered by our methods, we first plot the inferred estimates of  $\hat{r}_s$  and  $\hat{r}_a$  against the key meteorological drivers, namely VPD and the frictional velocity  $U^*$ , respectively (figures 3 and 4). The behavior of  $\hat{r}_s$  against VPD is consistent across all the models and reflects a similar behavior as presented for  $\hat{Q}_{LE}$  (figure 5). The predicted  $\hat{r}_s$  shows a subtle increase at lower ranges of VPD, reflecting that stomata are still open for gas exchange with the atmosphere. However, as VPD increases, the stomata start to close and thus surface resistance increases sharply (Massmann *et al* 2019). Further, we find that  $\hat{r}_s$  is generally lower for grasslands, which explains the generally higher estimates of  $Q_{LE}$  in comparison to forests (figure 5). Another striking finding is that the models seem to be able to identify differences in the physiological functioning across different plant types in controlling  $\hat{r}_s$ . For instance, the inferred relationship of  $\hat{r}_s$



**Figure 3.** Assessing latent variables  $r_s$  and  $r_a$  against VPD and  $U^*$  respectively for different models in forests. Constrained hybrid models reveal the physical consistency of latent variables compared to the under-constrained model, especially under different environmental conditions. The colored lines represent the smoothed lines that fit a polynomial surface using local fitting for each site. The contour lines represent 2D kernel density estimate.

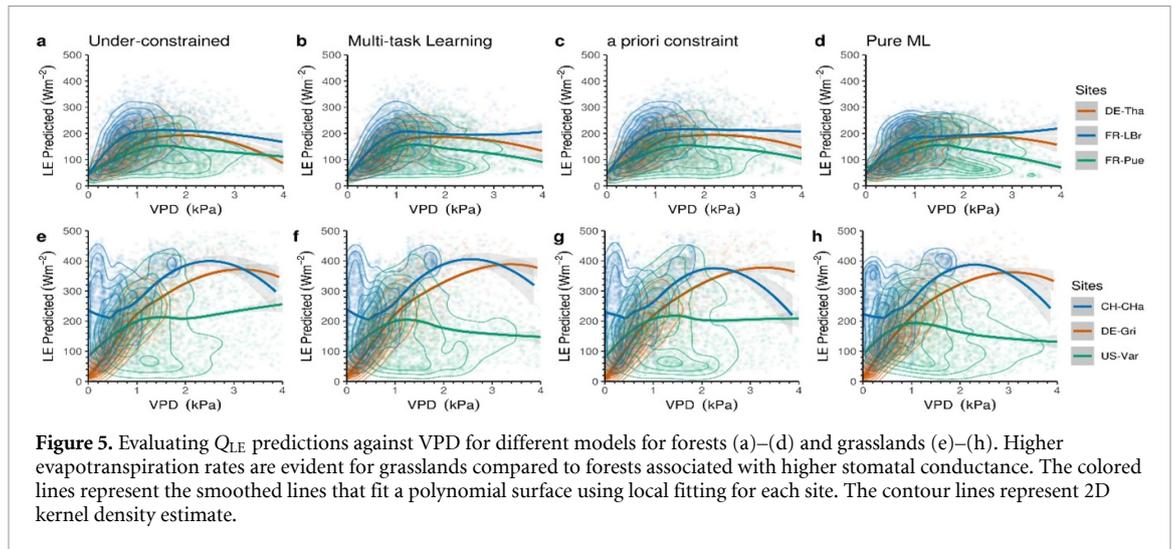


**Figure 4.** Assessing latent variables  $r_s$  and  $r_a$  against VPD and  $U^*$  respectively for different models in grasslands. The constrained hybrid models yield more physically consistent results compared to under-constrained model and are able to capture the vegetation and climate heterogeneities. The colored lines represent the smoothed lines that fit a polynomial surface using local fitting for each site. The contour lines represent 2D kernel density estimate.

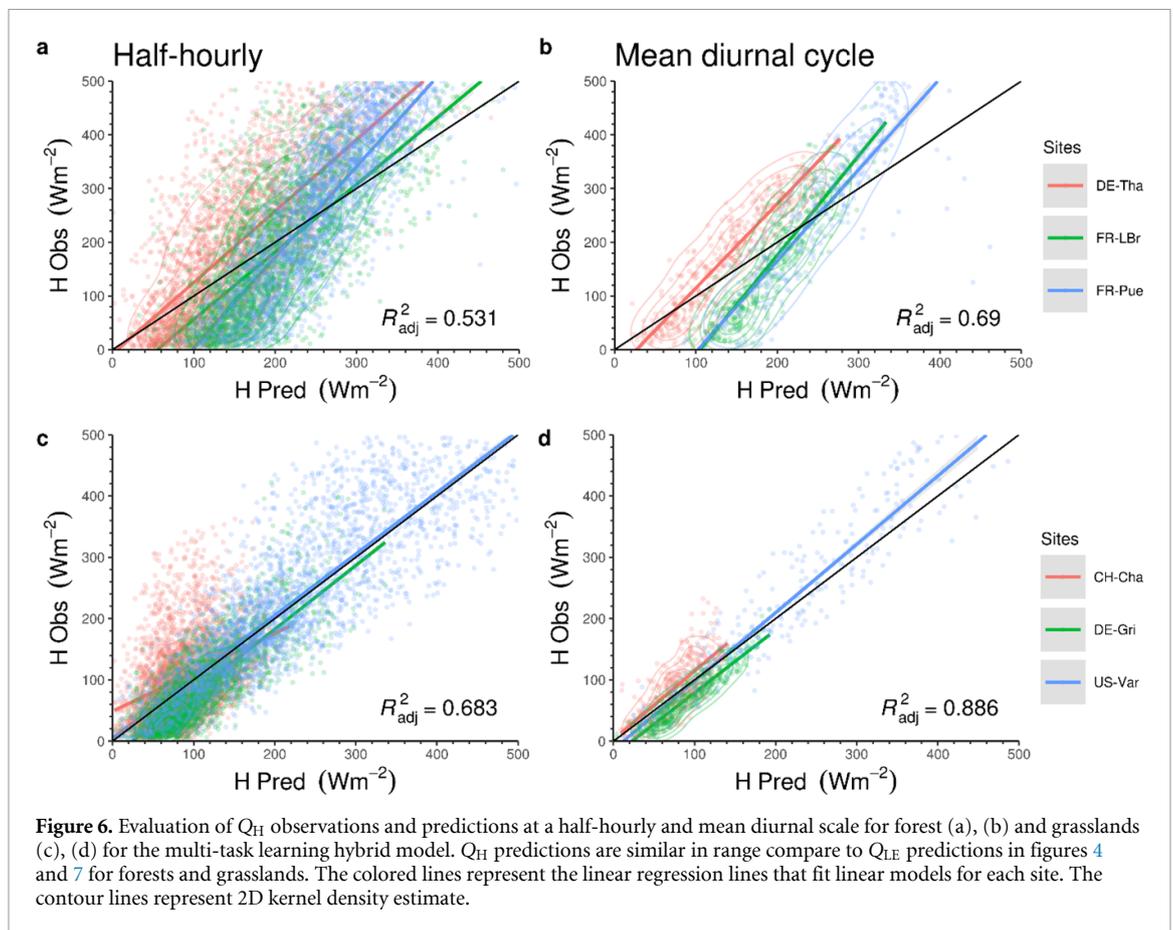
and VPD is very similar for the two forest sites DE-Tha and FR-LBr, which are dominated by evergreen needle-leaf trees, but it is quite different for the more arid site FR-Pue, which is dominated by evergreen broad-leaf trees (figures 3(a)–(c)). There, the hybrid models show that on average  $r_s$  rises more steeply with increasing VPD but flattens out at very high VPD (compare fit lines in figures 3(a)–(c)). Future research is needed to determine whether this behavior actually reflects the plants' mechanism for preventing leaf overheating by maintaining some evaporative cooling through the stomata (Lin *et al* 2017), or whether it is just an artifact of too sparse data at high VPD. Overall, the inferred  $\hat{r}_s$  through hybrid modeling (figures 3(a)–(c)) is much more precise than its conventional derivation by inverting the PM equation while making assumptions for  $r_a$  (figure 3(d)). This aspect constitutes a key advantage of our hybrid

approach as opposed to the inversion method, where artificial noise in the flux measurements directly propagates into the inverted estimates of  $\hat{r}_s$  resulting in high artificial variability and a bias in  $\hat{r}_s$  ranging from 0% to 30% (Wehr and Saleska 2021).

The inferred relationship for  $\hat{r}_a$  against its key driver  $U^*$  is not consistent across the hybrid models. The two constrained hybrid models, i.e. multi-task learning (figure 3(f)) and *a priori* constraint (figure 3(g)), consistently reflect the expected negative logarithmic relationship of  $\hat{r}_a$  against  $U^*$  (figures 3 and 4). In particular, in the case of the hybrid multi-tasking model, this result is promising because the relationship emerges from the observational data alone, without inducing any prespecified knowledge. Furthermore, the two constrained hybrid models show variations of the  $\hat{r}_a$  relationship across the sites (figures 3(f), (g) and 4(f), (g)). Thus, they are capable



**Figure 5.** Evaluating  $Q_{LE}$  predictions against VPD for different models for forests (a)–(d) and grasslands (e)–(h). Higher evapotranspiration rates are evident for grasslands compared to forests associated with higher stomatal conductance. The colored lines represent the smoothed lines that fit a polynomial surface using local fitting for each site. The contour lines represent 2D kernel density estimate.



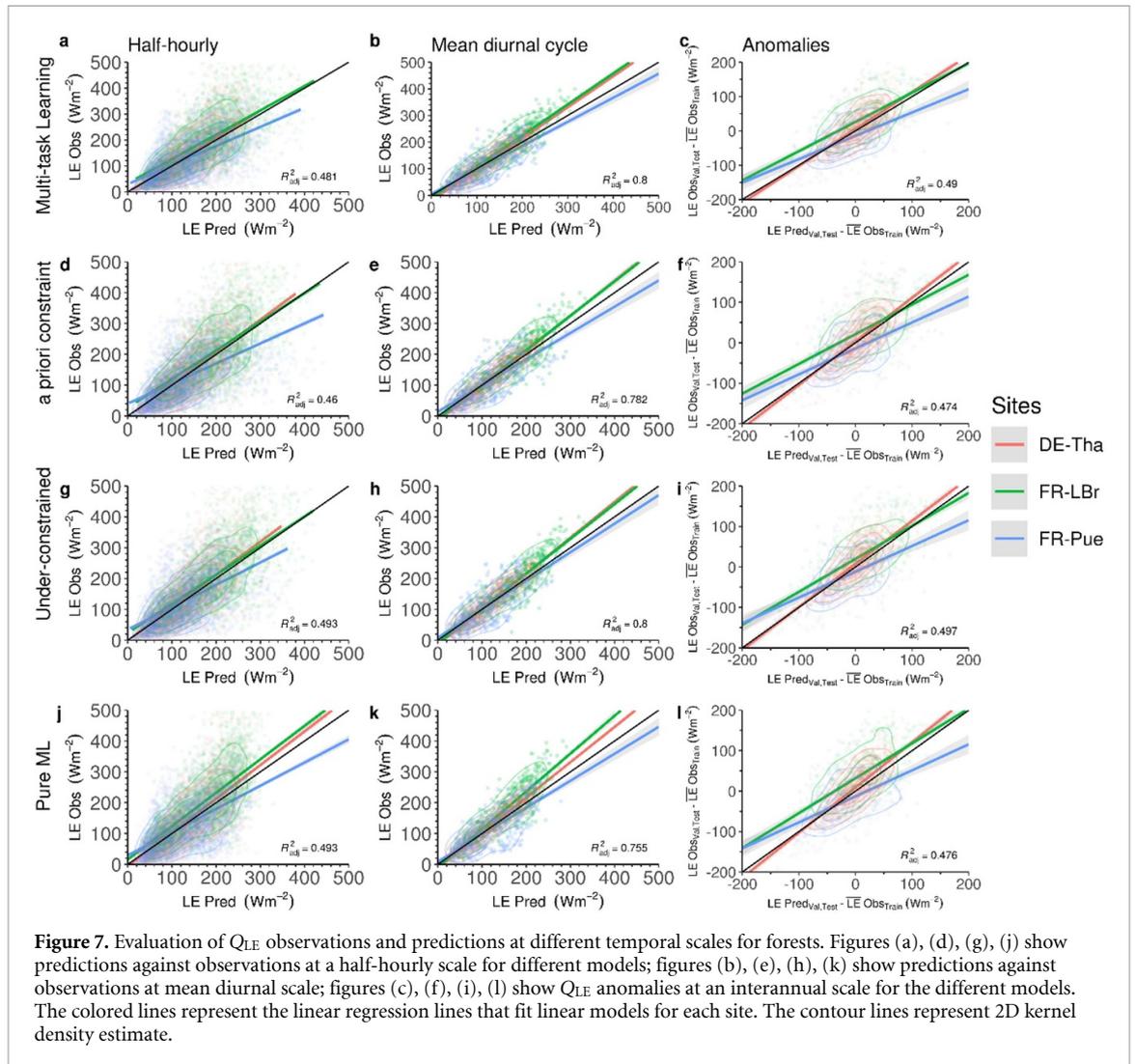
**Figure 6.** Evaluation of  $Q_H$  observations and predictions at a half-hourly and mean diurnal scale for forest (a), (b) and grasslands (c), (d) for the multi-task learning hybrid model.  $Q_H$  predictions are similar in range compare to  $Q_{LE}$  predictions in figures 4 and 7 for forests and grasslands. The colored lines represent the linear regression lines that fit linear models for each site. The contour lines represent 2D kernel density estimate.

of capturing the canopy heterogeneity across sites and are more flexible than the conventional rigid parameterizations shown in figure 3(h) (forests) and figure 4(h) (grasslands), where  $r_a$  is a homogenous function of  $U^*$  across the different sites.

The under-constrained hybrid model (figure 3(e)), however, illustrates the risk of equifinality and the physics-violating behavior of this approach. In other words,  $\hat{r}_a$  exhibits physically inconsistent relationships in the under-constrained model across the sites (figure 3(e)), while the

predicted  $\hat{r}_s$  and  $\hat{Q}_{LE}$  retain physically plausible estimates (figures 3(a) and 5(g)–(i), respectively). The issue of equifinality is more prominent in forests than in grasslands, likely because aerodynamic resistance is less dominant in controlling  $Q_{LE}$  in forests (figures 3(e) and 4(e); Chen and Liu 2020).

Aerodynamic resistance  $r_a$  constitutes a critical link in the surface energy balance, especially under different environmental and stability conditions, as it has a bearing on both,  $Q_{LE}$  and  $Q_H$ . Uncertainties in  $Q_{LE}$  and  $Q_H$  mainly arise from the uncertainty



**Figure 7.** Evaluation of  $Q_{LE}$  observations and predictions at different temporal scales for forests. Figures (a), (d), (g), (j) show predictions against observations at a half-hourly scale for different models; figures (b), (e), (h), (k) show predictions against observations at mean diurnal scale; figures (c), (f), (i), (l) show  $Q_{LE}$  anomalies at an interannual scale for the different models. The colored lines represent the linear regression lines that fit linear models for each site. The contour lines represent 2D kernel density estimate.

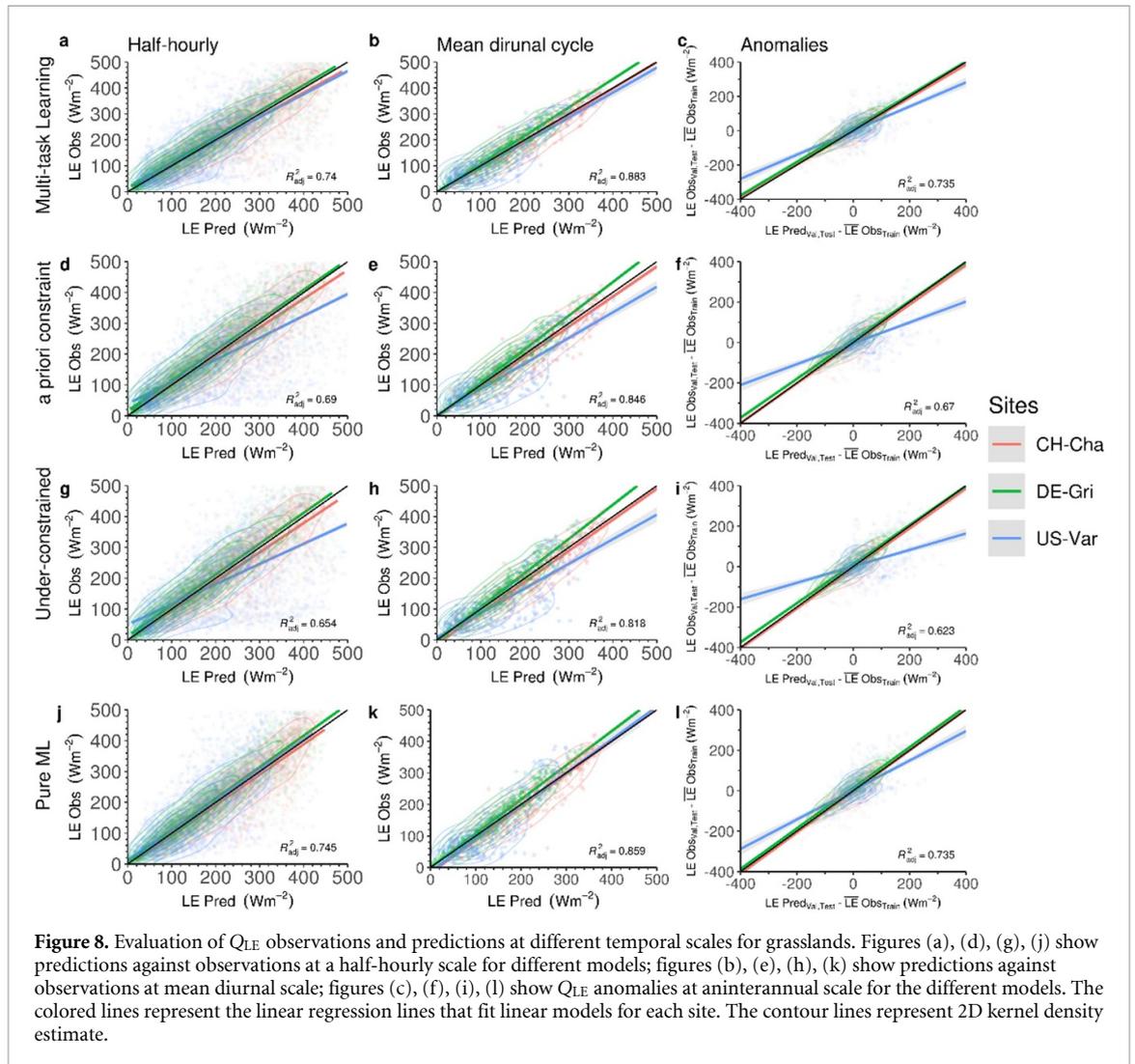
in estimating  $r_a$  for both dense and sparse canopy, and particularly under arid and semi-arid conditions (Trebs *et al* 2021). Our multi-task learning hybrid model, however, is able to provide fairly high accuracy for  $Q_{LE}$  and  $Q_H$  predictions for grasslands under unstable and semi-arid conditions without overestimating  $r_a$ , which has been proven difficult in other modeling efforts (Trebs *et al* 2021). For example, the predictions for  $Q_{LE}$  (figure 5) and  $Q_H$  (figures 6(c) and (d)) at the US-Var grassland site, characterized by a dry Mediterranean-type climate (Xu and Baldocchi 2004, De Kauwe *et al* 2017), are fairly accurate and relate to physically consistent  $r_a$  predictions.

To get an estimate of the structural (epistemic) uncertainty for the inferred relationships for  $r_s$  and  $r_a$ , we train each model five times with random initializations (refer to section 2.3). The hybrid models show consistent predictions for the relationships for  $r_s$  and  $r_a$  at mean diurnal scale across the model runs with different initializations. The under-constrained hybrid model is consistent in producing physically non-interpretable  $r_a$  for all initializations. The constrained hybrid models, on the other hand, are

able to consistently reproduce the physically plausible relationships for  $r_s$  and  $r_a$ , especially at forest sites. Hence, our hybrid modeling approach yields robust predictions, yet we stress the caveats related to equifinality in these under-constrained model setups.

### 3.2. Evaluation of the target variables $\hat{Q}_{LE}$ and $\hat{Q}_H$

We evaluate the predicted  $Q_{LE}$  ( $\hat{Q}_{LE}$ ) from all the hybrid models and the pure ML model against observed  $Q_{LE}$  ( $Q_{LE,obs}$ ) at a half-hourly scale and at 7 d mean aggregates (mean diurnal) for forest (figure 7) and grassland (figure 8) sites. All models produce similar  $Q_{LE}$  patterns compared to observations with minor differences in performance. For forests (figure 7), the more flexible models, i.e. the under-constrained hybrid model and pure ML model, perform slightly better ( $R^2 = 0.49$ ) than do the multi-task learning model ( $R^2 = 0.48$ ) and the *a priori* constraint model ( $R^2 = 0.46$ ). For grasslands, the performance of all models is generally better than for forests. We find that the performance of the multi-task learning model exceeds the performance of the *a priori* constraint model and is similar to the pure



**Figure 8.** Evaluation of  $Q_{LE}$  observations and predictions at different temporal scales for grasslands. Figures (a), (d), (g), (j) show predictions against observations at a half-hourly scale for different models; figures (b), (e), (h), (k) show predictions against observations at mean diurnal scale; figures (c), (f), (i), (l) show  $Q_{LE}$  anomalies at interannual scale for the different models. The colored lines represent the linear regression lines that fit linear models for each site. The contour lines represent 2D kernel density estimate.

ML model ( $R^2 = 0.74$ – $0.75$ ) (figure 8). This finding could indicate that our theory-based constraint for  $r_a$  might be too rigid and is not supported by the flux observations. Overall, the RMSE ranges from 70 to 73  $Wm^{-2}$  for forests and 60–71  $Wm^{-2}$  for grasslands at a half-hourly scale for all models. The MAE at half-hourly measurements range from 50 to 53  $Wm^{-2}$  for forests and from 43 to 48  $Wm^{-2}$  for grasslands for all models. The multi-task learning model provides predictions for  $Q_H$  ( $\hat{Q}_H$ ) (figure 6) of similar accuracy compared to the  $Q_{LE}$  predictions for all sites (figures 7 and 8), reaching  $R^2 = 0.53$  for forests and  $R^2 = 0.68$  for grasslands sites at a half-hourly scale. Overall, the pure ML model slightly outperforms the hybrid models for both forest and grassland sites, owing to its flexibility and non-parametric attributes as it only minimizes the loss of  $Q_{LE}$ , without being constrained by the PM equation. More information on the physical interpretability of  $\hat{Q}_{LE}$  assessed against meteorological variables can be found in the Suppl. Info. Sec. 3.3.

We evaluate the hybrid models' consistency with respect to the interannual variability of  $Q_{LE}$  for the

different sites. The interannual anomalies are calculated as the difference between the average annual estimates of  $Q_{LE,obs}$  in the training dataset and the annual estimates of  $Q_{LE,obs}$  and  $\hat{Q}_{LE}$  in the validation and test dataset for the EC data and models, respectively, to evaluate the predictive capacity of the different models (Jung *et al* 2009, Besnard *et al* 2019). Figures 7 and 8 show the overall fit and performance of the models in predicting interannual anomalies of  $\hat{Q}_{LE}$  compared to observed anomalies of  $Q_{LE,obs}$ . The values of  $R^2$  range between 0.47 and 0.49 for the interannual  $\hat{Q}_{LE}$  anomalies for forests and thus exhibit a comparable performance at half-hourly frequency ( $R^2$  ranges between 0.46 and 0.49) (figure 7). We observe a similar behavior at grassland sites:  $R^2$  ranges between 0.65 and 0.75 at the half-hourly scale and between 0.62 and 0.74 for the interannual  $Q_{LE}$  anomalies (figure 8). Overall, the evaluation of the models at multiple temporal scales shows that the models are capable of learning not only the predominant structure of the diurnal and seasonal cycle, but also the more subtle year-to-year anomalies. The presented consistency reflects that the models learn the

physically correct dependence of the meteorological predictor variables controlling  $Q_{LE}$  (figure 5).

#### 4. Conclusions

We present a new approach to end-to-end hybrid modeling of latent heat fluxes that can simultaneously retrieve the two controlling intermediate variables—the surface ( $r_s$ ) and aerodynamic resistance ( $r_a$ )—while maintaining physical consistency across different vegetation types. The hybrid models provide reliable predictions against measurements of latent heat fluxes at different time scales, ranging from daily to seasonal to interannual variations. This cross-scale consistency shows that our model framework is able to learn the physically consistent dependencies between the meteorological input variables and the target fluxes, rather than just the dominant structure of diurnal and seasonal cycles.

The main novelty and outcome of this study are the data-driven parameterizations for  $r_s$  and  $r_a$  jointly estimated by two separate NNs, which can lead new insights on biophysical regulation of surface evaporation. We show that the NNs together can provide many solutions (non-uniqueness) and lead to physically plausible predictions for  $Q_{LE}$  fluxes, while presenting physically implausible relationships to the predictors. This non-uniqueness can be mitigated by introducing either more data or theory into the loss function of the hybrid model. Specifically, we make use of two different approaches (*a priori* constraint and multi-task learning) to regularize the parameter space for the NNs. The constrained hybrid models in general yield accurate and physically interpretable predictions, with the multi-task learning model estimating the target and latent variable predictions with accuracy similar to the pure ML model, but with constraints that respect the surface energy budget. Therefore, by incorporating additional observation-based information, the multi-task learning model is the optimal hybrid model for the problem at hand. This architecture makes it possible to reduce equifinality and enables the model to extract underlying information from observations rather than ad hoc assumptions, while allowing the NN enough flexibility within the limits of physical interpretability of the surface energy balance.

When using the hybrid models to determine  $r_s$  and  $r_a$ , we find substantial differences between sites compared to the very uniform empirical formulations commonly used. This inter-site spread in the observation-based parameterizations suggests that the conventional empirical formulations are too rigid and do not account for the variability caused by the vegetation canopy structure. The hybrid models show differences among sites, highlighting in particular the different physiological functions of the different plant types, in comparison to the PM equation under the Big Leaf assumption. The resulting relationships for

$r_s$  and  $r_a$  not only show physically consistent behavior across scales, but also reveal new insights into how the varying resistances control surface energy fluxes. By evaluating the relationship of  $r_s$  and  $r_a$  to the driving meteorological variables, we are able to identify the effect of structural differences between forests and grasslands. The general response of stomatal conductance to VPD and photosynthesis in forest and grassland ecosystems is more aligned with the optimality theory as it considers the interactions between transpiration and carbon assimilation. However, grasslands tend to show a weaker dependence of stomatal conductance on photosynthesis and VPD that can be attributed to structural vegetation differences of the leaf area index and significantly larger weight and impact of  $r_a$  on  $r_s$ .  $r_a$  is generally higher for grassland sites than for forests which is attributed to the surface roughness of leaves.  $r_s$  is higher for forest sites compared to grasslands owing to the different atmospheric demand of the canopy and water uptake through roots that highlight the functional balance between shoots and roots under water-stressed conditions. In addition, we detect that these learned parameterizations in the hybrid models exhibit lower stomatal conductance, suggesting that the  $r_s$  values usually obtained by inversion of the PM equation may be systematically overestimated.

Several approaches have already been proposed to use the growing number of observations to constrain uncertainty in mechanistic model simulations, especially for key unknown plant behavior in the coupled Earth system (Lian *et al* 2018, Winkler *et al* 2019a, 2019b, Varney *et al* 2020). As a next step, we propose to derive parameterizations directly from observations using hybrid modeling, as presented in this study, to replace these ad hoc formulations in Earth system models. This approach will not only help reduce uncertainty, but also advance significantly the understanding of biogeophysical and biogeochemical processes in land–atmosphere coupling.

#### Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: [10.5281/zenodo.7078501](https://doi.org/10.5281/zenodo.7078501). Data will be available from 14 March 2023.

#### Code and data availability

All data used in this study are available from public databases or the literature, which can be found with the references provided in the respective ‘Data and methods’ subsection. Processed data and analysis scripts are available from the corresponding author upon request, and the repository will be published together with this article. Correspondence and requests for materials should be addressed to Reda ElGhawi (relghawi@bgc-jena.mpg.de).

## Acknowledgments

This research was funded by the European Research Council (ERC) Synergy Grant ‘Understanding and modeling the Earth System with Machine Learning (USMILE)’ under the Horizon 2020 research and innovation programme (Grant Agreement No. 855187). Gentine acknowledges funding from the National Science Foundation grant, Learning the Earth with Artificial intelligence and Physics (LEAP).

## Author contributions

R E G, A J W and M R designed the study. R E G conducted the analysis. B K provided technical support in setting up the hybrid modelling framework. C R, M K and B K contributed to the conceptual and technical machine learning aspect of the study. All authors contributed ideas and to the interpretation of the results. R E G and A J W drafted the manuscript with input from all authors.

## ORCID iDs

RedaReda ElGhawi  <https://orcid.org/0000-0003-2930-4537>

Markus Reichstein  <https://orcid.org/0000-0001-5736-1112>

Marco Körner  <https://orcid.org/0000-0002-9186-4175>

Pierre Gentine  <https://orcid.org/0000-0002-0845-8345>

Alexander J Winkler  <https://orcid.org/0000-0001-6574-4471>

## References

- Ajami H 2021 Geohydrology: global hydrological cycle hydrological cycle *Encyclopedia of Geology* (Cambridge: Academic Press) pp 393–8
- Baldocchi D *et al* 2001 FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities *Bull. Am. Meteorol. Soc.* **82** 2415–34
- Besnard S *et al* 2019 Memory effects of climate and vegetation affecting net ecosystem CO<sub>2</sub> fluxes in global forests *PLoS One* **14** e0211510
- Beven K 2006 A manifesto for the equifinality thesis *J. Hydrol.* **320** 18–36
- Bonan G B, Lawrence P J, Oleson K W, Levis S, Jung M, Reichstein M, Lawrence D M and Swenson S C 2011 Improving canopy processes in the community land model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data *J. Geophys. Res. Biogeosci.* **116**
- Brutsaert W 2005 *Hydrology: An Introduction* (Cambridge: Cambridge University Press)
- Carminati A and Javaux M 2020 Soil rather than xylem vulnerability controls stomatal response to drought *Trends Plant Sci.* **25** 868–80
- Carter C and Liang S 2019 Evaluation of ten machine learning methods for estimating terrestrial evapotranspiration from remote sensing *Int. J. Appl. Earth Obs. Geoinf.* **78** 86–92
- Chen J M and Liu J 2020 Evolution of evapotranspiration models using thermal and shortwave remote sensing data *Remote Sens. Environ.* **237** 111594
- Damour G, Simonneau T, Cochard H and Urban L 2010 An overview of models of stomatal conductance at the leaf level *Plant Cell Environ.* **33** 1419–38
- de Bezenac E, Pajot A and Gallinari P 2017 Deep learning for physical processes: incorporating prior scientific knowledge (arxiv:1711.07970)
- De Kauwe M G, Medlyn B E, Knauer J and Williams C A 2017 Ideas and perspectives: how coupled is the vegetation to the boundary layer? *Biogeosciences* **14** 4435–53
- Dou X and Yang Y 2018 Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems *Comput. Electron. Agric.* **148** 95–106
- Drake J E *et al* 2018 Trees tolerate an extreme heatwave via sustained transpirational cooling and increased leaf thermal tolerance *Glob. Change Biol.* **24** 2390–402
- Gerosa G, Mereu S, Finco A and Marzuoli R 2012 Stomatal conductance modeling to estimate the evapotranspiration of natural and agricultural ecosystems *Evapotranspiration—Remote Sensing and Modeling* (Rijeka: InTech) (<https://doi.org/10.5772/21825>)
- Hetherington A M and Woodward F I 2003 The role of stomata in sensing and driving environmental change *Nature* **424** 901–8
- Jain S K, Nayak P C and Sudheer K P 2008 Models for estimating evapotranspiration using artificial neural networks, and their physical interpretation *Hydrol. Process.* **22** 2225–34
- Jarvis P G 1976 The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field *Phil. Trans. R. Soc. B* **273** 593–610
- Jia X *et al* 2020b Physics-guided recurrent graph model for predicting flow and temperature in river networks (arxiv:2009.12575v2)
- Jia X, Willard J, Karpatne A, Read J S, Zwart J A, Steinbach M and Kumar V 2020a Physics-guided machine learning for scientific discovery: an application in simulating lake temperature profiles (arxiv:2001.11086)
- Jung M *et al* 2010 Recent decline in the global land evapotranspiration trend due to limited moisture supply *Nature* **467** 951–4
- Jung M *et al* 2020 Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach *Biogeosciences* **17** 1343–65
- Jung M, Reichstein M and Bondeau A 2009 Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model *Biogeosciences* **6** 2001–13
- Karpatne A, Atluri G, Faghmous J H, Steinbach M, Banerjee A, Ganguly A, Shekhar S, Samatova N and Kumar V 2017a Theory-guided data science: a new paradigm for scientific discovery from data *IEEE Trans. Knowl. Data Eng.* **29** 2318–2
- Karpatne A, Watkins W, Read J and Kumar V 2017b Physics-guided neural networks (PGNN): an application in lake temperature modeling (arXiv:1710.11431)
- Kelliher F M, Köstner B M M, Hollinger D Y, Byers J N, Hunt J E, McSeveny T M, Meserth R, Weir P L and Schulze E D 1992 Evaporation, xylem sap flow, and tree transpiration in a New Zealand broad-leaved forest *Agric. For. Meteorol.* **62** 53–73
- Kennedy D, Swenson S, Oleson K W, Lawrence D M, Fisher R, Lola da Costa A C and Gentine P 2019 Implementing plant hydraulics in the community land model version 5 *J. Adv. Model. Earth Syst.* **11** 485–513
- Knauer J, El-Madany T S, Zaehle S and Migliavacca M 2018 Bigleaf—an R package for the calculation of physical and physiological ecosystem properties from eddy covariance data *PLoS One* **13** e0201114
- Köstner B M M, Schulze E D, Kelliher F M, Hollinger D Y, Byers J N, Hunt J E, McSeveny T M, Meserth R and Weir P L 1992 Transpiration and canopy conductance in a pristine broad-leaved forest of Nothofagus: an analysis of xylem sap flow and eddy correlation measurements *Oecologia* **91** 350–9

- Kraft B, Jung M, Körner M, Koirala S and Reichstein M 2022 Towards hybrid modeling of the global hydrological cycle *Hydrol. Earth Syst. Sci.* **26** 1579–614
- Kraft B, Jung M, Körner M and Reichstein M 2020 Hybrid modeling: fusion of a deep approach and physics-based model for global hydrological modeling *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **43** 1537–44
- Krasnopolsky V M 2013 *The Application of Neural Networks in the Earth System Sciences* (Dordrecht: Springer) (<https://doi.org/10.1007/978-94-007-6073-8>)
- Leuning R 1995 A critical appraisal of a combined stomatal-photosynthesis model for C3 plants *Plant Cell Environ.* **18** 339–55
- Leuning R, Kriedemann P E and McMurtrie R E 1991 Simulation of evapotranspiration by trees *Agric. Water Manag.* **19** 205–21
- Li L et al 2018 Evaluating global land surface models in CMIP5: analysis of ecosystem water- and light-use efficiencies and rainfall partitioning *J. Clim.* **31** 2995–3008
- Li X, Gentine P, Lin C, Zhou S, Sun Z, Zheng Y, Liu J and Zheng C 2019 A simple and objective method to partition eddy-covariance sites *Agric. For. Meteorol.* **265** 171–82
- Lian X et al 2018 Partitioning global land evapotranspiration using CMIP5 models constrained by observations *Nat. Clim. Change* **8** 640–6
- Liebel L and Körner M 2018 Auxiliary tasks in multi-task learning (arXiv:1805.06334)
- Lin C, Gentine P, Huang Y, Guan K, Kimm H and Zhou S 2018 Diel ecosystem conductance response to vapor pressure deficit is suboptimal and independent of soil moisture *Agric. For. Meteorol.* **250–251** 24–34
- Lin H, Chen Y, Zhang H, Fu P and Fan Z 2017 Stronger cooling effects of transpiration and leaf physical traits of plants from a hot dry habitat than from a hot wet habitat *Funct. Ecol.* **31** 2202–11
- Massmann A, Gentine P and Lin C 2019 When does vapor pressure deficit drive or reduce evapotranspiration? *J. Adv. Model. Earth Syst.* **11** 3305–20
- Medlyn B E, Duursma R A, Eamus D, Ellsworth D S, Prentice I C, Barton C V M, Crous K Y, De Angelis P, Freeman M and Wingate L 2011 Reconciling the optimal and empirical approaches to modelling stomatal conductance *Glob. Change Biol.* **17** 2134–44
- Monteith J L 1985 Evaporation from land surfaces: progress in analysis and prediction since 1948 *Advances in Evapotranspiration, Proc. National Conf. on Advances in Evapotranspiration* (Joseph, MI: American Society of Agricultural Engineers) pp 4–12 (available at: <https://agris.fao.org/agris-search/search.do?recordID=US8644525>)
- Monteith J L 1965 Evaporation and environment *Symp. Society for Experimental Biology, the State and Movement of Water in Living Organisms* vol 19, ed G E Fogg pp 205–34
- Monteith J L and Unsworth M 2013 *Principles of Environmental Physics: Plants, Animals, and the Atmosphere* 4th edn (Cambridge: Academic Press) (<https://doi.org/10.1016/C2010-0-66393-0>)
- Penman H L 1948 Natural evaporation from open water, bare soil and grass *Proc. R. Soc. A* **193** 120–45
- Polhamus A, Fisher J B and Tu K P 2013 What controls the error structure in evapotranspiration models? *Agric. For. Meteorol.* **169** 12–24
- Reichstein M et al 2005 On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm *Glob. Change Biol.* **11** 1424–39
- Reichstein M, Ahrens B, Kraft B, Camps-Valls G, Carvalhais N, Gans F, Gentine P and Winkler A J 2022 Combining system modeling and machine learning into hybrid ecosystem modeling *Knowledge-Guided Machine Learning* (New York: Chapman and Hall/CRC) pp 327–52
- Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J and Carvalhais N 2019 Deep learning and process understanding for data-driven Earth system science *Nature* **566** 195–204
- Reimers C and Requena-Mesa C 2020 Deep learning—an opportunity and a challenge for geo- and astrophysics *Knowledge Discovery in Big Data from Astronomy and Earth Observation: Astroinformatics* (Amsterdam: Elsevier) pp 251–65
- Ronda R J, De Bruin H A R and Holtslag A A M 2001 Representation of the canopy conductance in modeling the surface energy budget for low vegetation *J. Appl. Meteorol.* **40** 1431–44
- Schmidt L, Heße F, Attinger S and Kumar R 2020 Challenges in applying machine learning models for hydrological inference: a case study for flooding events across Germany *Water Resour. Res.* **56** e2019WR025924
- Stewart J B 1988 Modelling surface conductance of pine forest *Agric. For. Meteorol.* **43** 19–35
- Tan S, Wang H, Prentice I C and Yang K 2021 Land-surface evapotranspiration derived from a first-principles primary production model *Environ. Res. Lett.* **16** 104047
- Tramontana G et al 2016 Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms *Biogeosciences* **13** 4291–313
- Trebs I et al 2021 The role of aerodynamic resistance in thermal remote sensing-based evapotranspiration models *Remote Sens. Environ.* **264** 112602
- Varney R M, Chadburn S E, Friedlingstein P, Burke E J, Koven C D, Hugelius G and Cox P M 2020 A spatial emergent constraint on the sensitivity of soil carbon turnover to global warming. *Nat. Commun.* **11** 1–8
- Violet-Chabrand S and Lawson T 2019 Dynamic leaf energy balance: deriving stomatal conductance from thermal imaging in a dynamic environment *J. Exp. Bot.* **70** 2839
- Wehr R and Saleska S R 2021 Calculating canopy stomatal conductance from eddy covariance measurements, in light of the energy budget closure problem *Biogeosciences* **18** 13–24
- Willard J, Jia X, Xu S, Steinbach M and Kumar V 2020 Integrating physics-based modeling with machine learning: a survey 1 (arXiv:2003.04919 [physics.comp-ph]) p 34
- Wilson K et al 2002 Energy balance closure at FLUXNET sites *Agric. For. Meteorol.* **113** 223–43
- Winkler A J, Myneni R B, Alexandrov G A and Brovkin V 2019a Earth system models underestimate carbon fixation by plants in the high latitudes *Nat. Commun.* **10** 1–8
- Winkler A J, Myneni R B and Brovkin V 2019b Investigating the applicability of emergent constraints *Earth Syst. Dyn.* **10** 501–23
- Xu L and Baldocchi D D 2004 Seasonal variation in carbon dioxide exchange over a Mediterranean annual grassland in California *Agric. For. Meteorol.* **123** 79–96
- Yang T, Sun F, Gentine P, Liu W, Wang H, Yin J, Du M and Liu C 2019 Evaluation and machine learning improvement of global hydrological model-based flood simulations *Environ. Res. Lett.* **14** 114027
- Zeppel M and Eamus D 2008 Coordination of leaf area, sapwood area and canopy conductance leads to species convergence of tree water use in a remnant evergreen woodland *Aust. J. Bot.* **56** 97–108
- Zhang Z Z, Zhao P, McCarthy H R, Zhao X H, Niu J F, Zhu L W, Ni G Y, Ouyang L and Huang Y Q 2016 Influence of the decoupling degree on the estimation of canopy stomatal conductance for two broadleaf tree species *Agric. For. Meteorol.* **221** 230–41
- Zhao L, Xia J, Xu C, Yu Wang Z, Sobkowiak L and Long C 2013 Evapotranspiration estimation methods in hydrological models *J. Geogr. Sci.* **23** 359–69
- Zhao W L, Gentine P, Reichstein M, Zhang Y, Zhou S, Wen Y, Lin C, Li X and Qiu G Y 2019 Physics-constrained machine learning of evapotranspiration *Geophys. Res. Lett.* **46** 14496–507
- Zhou S, Yu B, Zhang Y, Huang Y and Wang G 2016 Partitioning evapotranspiration based on the concept of underlying water use efficiency *Water Resour. Res.* **52** 1160–75