Hybrid Modeling of Evapotranspiration: Inferring Stomatal and Aerodynamic Resistances Using Combined Physics-Based and Machine Learning

Reda ElGhawi*[1,2,3], Basil Kraft[1], Christian Reimers[1], Markus Reichstein[1], Marco Körner[3], Pierre Gentine[4] and Alexander J. Winkler*[1]

[1] Max Planck Institute for Biogeochemistry, Biogeochemical Integration, Jena, Germany

[2] International Max Planck Research School for Global Biogeochemical Cycles, Max Planck Institute for Biogeochemistry, Jena, Germany

[3] Technical University of Munich, TUM School of Engineering and Design, Department of Aerospace and Geodesy, Munich, Germany

[4] Department of Earth and Environmental Engineering, Columbia University, New York, NY, 10027, USA

* Authors to whom any correspondence should be addressed

E-mail: {relghawi, awinkler}@bgc-jena.mpg.de

Received xxxxxx; Accepted for publication xxxxxx; Published xxxxxx

Abstract

The process of evapotranspiration transfers water vapour from vegetation and soil surfaces to the atmosphere, the so-called latent heat flux (Q_{LE}), and thus crucially modulates Earth's energy, water, and carbon cycles. Vegetation controls Q_{LE} through regulating the leaf stomata (i.e., surface resistance r_s) and through altering surface roughness (aerodynamic resistance r_a). Estimating r_s and r_a across different vegetation types proves to be a key challenge in predicting Q_{LE}. Here, we propose a hybrid modeling approach (i.e., combining mechanistic modeling and machine learning) for Q_{LE} where neural networks independently learn the resistances from observations as intermediate variables. In our hybrid modeling setup, we make use of the Penman-Monteith equation based on the Big Leaf theory in conjunction with multi-year flux measurements across different forest and grassland sites from the FLUXNET database. We follow two conceptually different strategies to constrain the hybrid model to control for equifinality arising when estimating the two resistances simultaneously. One strategy is to impose an *a priori* constraint on r_a based on our mechanistic understanding (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 the latent as well as the sensible heat flux (Q_{H}; data-driven strategy). Our

results show that all hybrid models exhibit a fairly high predictive skill for the target variables with R^2= 0.82-0.89 for grasslands and R^2= 0.70-0.80 for forests sites at the mean diurnal scale. The predictions of r_s and r_a show 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, 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 overly simple *ad hoc* formulations in Earth system models.

Keywords: Hybrid modeling, physics-constrained, machine learning, multi-task learning, evapotranspiration, surface conductance, aerodynamic conductance

1. Introduction

Evapotranspiration, *i.e.* the 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 as well as rather static soil properties and plant traits (Jung et al., 2010; Dou & Yang, 2018; Ajami, 2021). The characterization of Q_{LE}, however, remains challenging as our understanding of the underlying processes that control the exchange flux of water between land and atmosphere is still limited (Friedl, 1996; Sellers et al., 1997; Wang & Dickinson, 2012; Chen et al., 2014). While the physical drivers that cause water to evaporate are well described and understood, the influence of the biological control on Q_{LE}, mainly the transpirative water flux, is more difficult to assess. The key problem is that we cannot formulate universally valid mechanistic laws to describe plant behavior in their interactions between changing atmospheric and soil conditions. As a consequence, empirical formulations, especially for surface and aerodynamic resistance (Polhamus et al., 2013), remain in process-based models, which can lead to large uncertainties in predicting Q_{LE}. In this study, we propose a hybrid modeling approach that allows inference of these biological controls from observational data of Q_{LE} across ecosystems using machine learning, while adhering to known physical laws (Reichstein et al., 2022).

Plants critically influence Q_{LE} mainly through their direct control of transpiration, but also through shaping aerodynamic surface properties. Plants use their leaf stomata to dynamically regulate the water loss to the atmosphere, which not only depends on the atmospheric water demand, but also on soil water availability (Damour et al., 2010; Kennedy et al., 2019; Carminati & Javaux, 2020). Simultaneously, plants use the stomata to resorb atmospheric CO_2 as the central ingredient in the photosynthetic process (Schulze, 1986; Chaves et al., 2016). Consequently, most formulations of stomatal conductance (or the inverse, stomatal resistance r_s) are empirical and rely on optimality concepts: minimizing the water loss while maximizing carbon assimilation (e.g. Tan et al., 2021). However, 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). Other empirical approaches, e.g. the Jarvis–Stewart formulation, aim to derive parametrizations based on statistical correlations between r_s (or canopy resistance) and the key environmental variables (Jarvis, 1976; Stewart, 1988). 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).

Formulations of how plants affect Q_{LE} by shaping surface roughness and associated aerodynamic properties are considered less uncertain, but vary considerably among vegetation types (Shaw & Pereira, 1982; Nakai et al., 2008; Maurer et al., 2015). Generally, near-surface wind enhances turbulent mixing and thus the exchange of mass and heat between the surface and the atmosphere. The surface roughness lengths — critically determined by plant physiology — influence the mechanical turbulence as well as the near-surface atmospheric thermal structure (Vila-Guerau de Arellano et al., 2015). These relationships are formulated in the aerodynamic resistance r_a, which is conventionally assumed to scale inversely (hyperbola-type function) with wind speed, frictional velocity and atmospheric instability based on the diagnostic empirical Monin–Obukhov similarity theory (Knauer et al., 2018). Several studies (Chehbouni et al., 1996; Liu et al., 2006; Su et al., 2021; Trebs et al., 2021) demonstrated that these parameterizations might work under controlled settings in the laboratory, however, show large discrepancies when applied to other terrain and vegetation types. Overall, such empirical representations for r_s and r_a in deterministic models for Q_{LE}, although generally obeying physical laws and phenomenological behaviour (Krasnopolsky, 2013; de Bezenac et al., 2017), exhibit limited capability to adapt to other or changing vegetation composition and/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 due to their data-adaptive nature have been proposed as alternative approaches to reliably estimate $\boldsymbol{Q_{\mathrm{LE}}}$ (Dou & Yang, 2018; Carter & Liang, 2019). In particular, approaches that use machine learning (ML) techniques are gaining traction because they potentially can reveal unknown latent processes or constitute a more complete statistical representation of the processes that influence $\boldsymbol{Q_{\mathrm{LE}}}$ at different scales in space and time (Dou & Yang, 2018; Jung et al., 2009, 2020). However, these data-driven models also are subject to several drawbacks, such as the need for large amounts of high-quality data, their usually limited, poor out-of-sample generalizability, and their lack of mechanistic interpretability (Karpatne, Atluri, et al., 2017; Karpatne, Watkins, et al., 2017).

By combining machine-learning and mechanistic modeling, here denoted hybrid modeling, allows us to combine the strengths of both techniques: ensure physical consistency while efficiently harvesting the growing resource of observational data (Reichstein et al., 2019, 2022). 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., 2020), 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 capabilities during extreme conditions.

In this study we adapt the hybrid modeling approach and develop different models of Q_{LE} using the Penman-Monteith equation (Penman, 1948; Monteith, 1965) and eddy covariance flux measurements from several grassland as well as forest sites (Baldocchi et al., 2001; Li et al., 2018). Our hybrid models not only seek to yield accurate predictions of Q_{LE}, but more importantly enable to learn the functioning and influence of biological processes on Q_{LE}, expressed as the surface (r_s) and aeordynamic resistances (r_a). We present and explore the problem of equifinality in our setting (i.e. many combinations of r_a and r_s can result in the same Q_{LE}) and propose two conceptually different solutions (theory- versus data-driven). We evaluate the predictions of our various hybrid models for Q_{LE}, r_a and r_s against purely statistical models as well as against established mechanistic models. We conclude with the lessons learned.

2. Methodology

2.1 FLUXNET 2015 Data

The global flux network (FLUXNET; https://fluxnet.fluxdata.org), a global network of eddy covariance (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. We select only measured data and omit gap-filled data (Reichstein et al., 2005). 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 / systematic bias in our results (Wilson et al., 2002). Daytime values are selected based on a threshold of sensible heat flux $Q_{\mathrm{H}} > 5\ \mathrm{Wm}^{-2}$ and incoming short-wave radiation $\mathrm{SW}_{\mathrm{in}} > 50\ \mathrm{Wm}^{-2}$ to avoid stable boundary layer conditions (Lin et al., 2018; 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 (Zhou et al., 2016), and winter months (between the 10$^{\mathrm{th}}$ and 3$^{\mathrm{rd}}$ months of the year) are excluded to focus on surface heat fluxes when the vegetation is active (Zhao et al., 2019). The FLUXNET sites chosen include 3 forests and 3 grasslands with varying climates, site properties and long-term site year data (Table1).

2.2 The physically-based component: Penman-Monteith equation

Various process-based models exist for the estimation of Q_{LE}. They can be subdivided into energy, mass transfer-based, water balance methods, and aerodynamic methods (Brutsaert, 2005; Zhao et al., 2013). One prominent example is the Penman-Monteith (PM) equation (Penman, 1948; Monteith, 1965) that provides the theoretical basis for determining Q_{LE}, and for showing how Q_{LE} responds to changing climate and vegetation conditions (Monteith & 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, which was then extended to vegetative surfaces (Monteith, 1985; Monteith & Unsworth, 2013; Vialet-Chabrand & Lawson, 2019). The PM equation

$$Q_{\mathrm{LE}} = \frac{s_{\mathrm{c}}(R_{\mathrm{n}} - Q_{\mathrm{G}}) + \frac{\rho_a c_{\mathrm{p}}(e_{\mathrm{s}} - e_{\mathrm{a}})}{r_a}}{s_{\mathrm{c}} + \gamma(1 + \frac{r_s}{r_a})}\ , \tag{1}$$

describes the latent heat flux representing Q_{LE} fraction (MJ m^{-2} d^{-1}). Where R_n is the net radiation (MJ m^{-2} d^{-1}), Q_G is the soil heat flux (MJ m^{-2} d^{-1}), s_c is the slope of the saturation vapor pressure-temperature relationship (kPa C^{-1}), $e_s - e_a$ is the vapor pressure deficit (VPD) of air (kPa), ρ_a is the mean air density at constant pressure (kg m^{-3}), c_p is the specific heat of dry air at constant pressure (1004.834 J kg^{-1} C^{-1}), γ is the psychrometric constant (kPa C^{-1}), r_s is the bulk surface resistance (sm^{-1}), and r_a is the aerodynamic resistance (sm^{-1}).

2.3 Overview of models

2.3.1 Inverted Penman-Monteith and Pure ML model

The PM equation is considered to be physically-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 & Woodward, 2003; Damour et al., 2010; Gerosa et al., 2012). Extensive approaches exist to model surface conductance at the leaf level. 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). A common approach thus is to simply to solve the Penman-Monteith equation for r_s to obtain the bulk surface resistance

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

assuming that the aerodynamic resistance r_a is known. 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 & Eamus, 2008; Zhang et al., 2016). We restrict surface and aerodynamic resistance values derived using Penman-Monteith inversion and empirical formulations (Knauer et al., 2018) based on intervals that are physically realistic (0-2000 sm^{-1} and 0-500 sm^{-1}, respectively).

The estimates for r_s from Eq. 2 derived through inverting the PM equation are referred to as the PM Inv model. Values for r_a are estimated using the Big Leaf package (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})

$$r_{\mathrm{am}} = \mathrm{WS}/\mathrm{U}^{*2} \tag{3}$$

$$r_{\mathrm{bh}} = 6.2\ \mathrm{U}^{*-0.667} \tag{4}$$

and

$$r_{\mathrm{a}} = r_{\mathrm{am}} + r_{\mathrm{bh}}\ , \tag{5}$$

where WS is wind speed (ms^{-1}) and U^* is friction velocity (ms^{-1}). The PM Inv model represents a baseline physical model for comparison against data-driven models for Q_{LE}. The pure ML model for Q_{LE} is set up to evaluate predictions against hybrid models. 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 (Eq. 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 framework calculates Q_{LE} using the PM equation (ref to equation), where the two intermediate variables r_s and r_a are estimated by two feed-forward neural networks (FNN)(Fig. 1). The predictors used for predicting r_{s} consist of air temperature (TA), water availability index (WAI), incoming shortwave radiation ($\mathrm{SW_{in}}$), mean incoming shortwave potential ($\mathrm{SW}_{pot\ sm}$), vapor pressure deficit (VPD), net radiation (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 and added WAI at previous time step ($\mathrm{WAI}_{t\text{-}1}$)

$$\mathrm{WA}I_t = \mathrm{P}_t - Q_{\mathrm{LE}_t} + \mathrm{WAI}_{t\text{-}1}\ . \tag{6}$$

The predictors used for predicting r_a include wind speed (WS) and is friction velocity (U^*). The predictors are normalized using the mean and standard deviation of the training dataset. Thus, the hybrid model predicts first the intermediate (also called latent) variables r_s and r_a and uses them to estimate

Q_{LE} based on the PM equation. The loss function is hence defined as the difference of mean absolute Q_{LE} errors between the model predictions and observations with n sample size, and parameters θ for r_s and r_a

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

Although the two FNN for r_a and r_s take different predictor variables, the hybrid model is characterized as under-constrained when simultaneously estimating the two intermediate variables using only one target Q_{LE}. The proposed hybrid model thus suffers from equifinality in this framework, so that many different combinations of r_s and r_a can result in the same Q_{LE} value (Fig. 2). The issue of equifinality, or non-uniqueness, is when different model parametrization and/or structures result in equivalent representations of the system (Beven, 2006; Schmidt et al., 2020).

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

The identification and elimination of equifinality due to the non-uniqueness 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 through model regularization (Fig. 3). This may be achieved by including additional theory, for example by the integration of a priori knowledge in the loss function (i.e. a regularization). Here, we choose to induce an a priori constraint on r_a in the hybrid model based on empirical formulation presented in Eq. 5 as the formulation for r_a are considered to be more robust than for r_s. To do so we regularize the loss function by adding a constraint on aerodynamic resistance r_a / ϕ. The relative importance of r_a in the new loss is regulated by ϕ, which is varied between 0.01 (high influence) to 2 (little influence). Based on multiple model runs, with different ϕ values, the value of $\phi = 2$ was selected for the results presented, thus only imposing a minor influence based on prior knowledge in the loss function. Another way of restricting the parameter space is by extending the framework to model auxiliary target variables in a multi-task setting. Since the sensible heat flux (Q_{H}) is also dependent on the aerodynamic resistance r_a, we explore multi-task learning approach (Liebel & Körner, 2018) on the intermediate variable regularization by adding Q_{H} as an auxiliary target variable in addition to Q_{LE} (Fig. 3). The estimation of Q_{H} is based on the resistance formulation

$$Q_{\mathrm{H}} = \frac{\rho_a c_{\mathrm{p}}(\mathrm{TS}-\mathrm{TA})}{r_a}, \tag{8}$$

where TS and TA are surface and air temperature respectively. The TS is estimated using the Stefan-Boltzmann equation

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

Where $Q_{\mathrm{LW_{out}}}$ is the outgoing longwave radiation ($\mathrm{Wm^{-2}}$), σ is the Stefan-Boltzmann constant (5.789×10^{-8} $\mathrm{Wm^{-2}K^{-4}}$) and ϵ is emissivity (dimensionless). The emissivity ranges between 0-1, and the values chosen were based on a sensitivity analyses for each model with the highest predictive accuracy.

2.4 Evaluation

In total, four models are constructed, three of which are hybrid models, consisting of one under-constrained and two constrained hybrid models, where the latter consists of either an *a priori* constraint on r_a or using a multi-task learning approach. For a baseline comparison, we use a pure ML model 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). We select the hyperparameters based on manual optimization and tuning. Evaluation metrics such as the root mean square error (RMSE) and 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 at the half-hourly and 7-day mean aggregated scale. 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 Statistical performance and mechanistic plausibility of the models

We evaluate predicted Q_{LE} ($\hat{Q}_{\mathrm{LE}}$) from all the hybrid models and the pure ML model against observed Q_{LE} ($Q_{\mathrm{LE}_{\mathrm{obs}}}$) at half-hourly scale and at 7-day mean aggregates (mean diurnal) for forest (Fig. 4) and grassland (Fig. 5) sites. All models reproduce similar Q_{LE} patterns compared to observations with minor differences in performance. For forests (Fig. 4), the more flexible models, the under-constrained hybrid model and pure ML model, exhibit a slightly higher performance (R^2 = 0.49) in comparison to 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 higher 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 ML model (R^2 = 0.74-0.75) (Fig. 5). This finding could indicate that our theory-based constraint for r_a might be too rigid and is not supported by observations of the system. Overall, the RMSE ranges from 70-73 Wm^{-2} for forests and 60-71 Wm^{-2} for grasslands at a half-hourly scale for all models. Moreover, the MAE at half-hourly measurements range between 50-53 Wm^{-2} for forests and 43-48 Wm^{-2} for grasslands for all models. The multi-task learning model provides predictions for Q_{H} ($\hat{Q}_{\mathrm{H}}$) (Fig. 6) of similar accuracy compared to the Q_{LE} predictions for all sites (Fig. 4-5), reaching R^2= 0.53 for forests and R^2= 0.68 for grasslands sites at half-hourly scale.

Our results at half-hourly scale are largely impacted by random measurement noise in EC data. To reduce the effect of this instrumental noise source, we aggregate half-hourly predictions in a 7-day window and calculate the mean diurnal cycle. The results presented in this noise-corrected manner demonstrate an even higher fit between $Q_{\mathrm{LE}_{\mathrm{obs}}}$ versus $\hat{Q}_{\mathrm{LE}}$ (Fig. 4-5) and $Q_{\mathrm{H}_{\mathrm{obs}}}$ versus $\hat{Q}_{\mathrm{H}}$ (Fig. 6) for forests and grasslands. The R^2 coefficient increases across all models by 53-70% for forests and 15-25% for grasslands sites based on the aggregated mean diurnal predictions. Further, the RMSE drops by 47-52% for forests, and by 43-48% for grasslands, while MAE also decreases by 47-52% for forests and 42-46% for grasslands. Controlling for noise in $\hat{Q}_{\mathrm{H}}$ in the same manner also increases R^2 from 0.68 to 0.87 for grasslands, and R^2 from 0.53 to 0.69 for forests (Fig. 6).

To assess the physical plausibility of the presented models, we evaluate their predictions of $\hat{Q}_{\mathrm{LE}}$ against the key predictor for atmospheric dryness, VPD. In all models, $\hat{Q}_{\mathrm{LE}}$ increases sharply at relatively low values of VPD (0-1 kPa), but starts to stabilize and eventually decreases for higher values of VPD (> 1 kPa; Fig 7). This behavior of the models aligns well with other studies that have shown that the transpiration rate increases with increasing VPD at the low and medium range, but starts to decrease again when VPD reaches high values (Buckley, 2005; Massmann et al., 2019; Monteith, 1995; Mott &

Peak, 2013). This plant response could reflect their ability to downregulate stomatal conductance as a preemptive measure to decrease water losses and to circumvent damages arising from intense dehydration of the canopy when the lower atmosphere becomes too dry (Farquhar, 1978; Massmann et al., 2019; Vico et al., 2013). Generally, grasslands sites reach higher $\hat{Q}_{\mathrm{LE}}$ values than forest sites for the same VPD range. Again, this result could be related to the different plant responses to VPD, since grasses are assumed to exhibit higher surface conductance (lower surface resistance r_s, respectively) compared to forests, resulting in higher transpiration rates (Garratt, 1992; Jarvis & Stewart, 1979). This aspect is discussed further in the next section when evaluating the learned resistances, r_s and r_a.

We next evaluate the hybrid models' consistency with respect to interannual variability of Q_{LE} for the different sites. The interannual anomalies are calculated as the difference between the average annual estimates of $Q_{\mathrm{LE}_{\mathrm{obs}}}$ in the training dataset and the annual estimates of $Q_{\mathrm{LE}_{\mathrm{obs}}}$ and $\hat{Q}_{\mathrm{LE}}$ 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 4 and 5 show the overall fit and performance of the models in predicting interannual anomalies of $\hat{Q}_{\mathrm{LE}}$ compared to observed anomalies of $Q_{\mathrm{LE}_{\mathrm{obs}}}$. The values of R^2 range between 0.47-0.49 for the interannual $\hat{Q}_{\mathrm{LE}}$ anomalies for forests and thus exhibit a comparable performance as at the half-hourly frequency (R^2 ranges between 0.46-0.49) (Fig. 4). We observe a similar behavior at grassland site: R^2 ranges between 0.65-0.75 at the half-hourly scale and between 0.62-0.74 for the interannual Q_{LE} anomalies (Fig. 5). 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 subtler year-to-year anomalies. The presented consistency reflects that the models learn the physically correct dependence of the meteorological predictor variables controlling Q_{LE}.

3.2 Evaluation of the learned intermediate variables $\widehat{r_s}$ and $\widehat{r_a}$

Next, we evaluate the inference of the Q_{LE}-controlling resistances $\widehat{r_s}$ and $\widehat{r_a}$ which are treated as intermediate variables in our hybrid approach. First, we plot the inferred estimates of $\widehat{r_s}$ and $\widehat{r_a}$ against the key meteorological drivers, namely VPD and the frictional velocity U^*, respectively (Fig. 8-9). The behavior of $\widehat{r_s}$ against VPD is consistent across all the models and reflects a somewhat similar behavior as presented for $\hat{Q}_{\mathrm{LE}}$. The predicted $\widehat{r_s}$ shows a gentle increase at lower ranges of VPD, so the stomata are still open for gas exchange with the atmosphere. However, as VPD increases to higher values, the stomata start to close and thus the surface resistance increases sharply. Further, we find that $\hat{r}_s$ is generally lower for grasslands, which explains the generally higher estimates of Q_{LE} compared to forests, as discussed above (Fig. 7). 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 $\hat{r}_s$ and VPD is very similar for the two forest sites DE-Tha and FR-LBr, which are dominated by evergreen needle-leaf trees, however, is quite different for the more arid site FR-Pue, which is dominated by evergreen broad-leaf trees (Fig. 8 a-c). There, the hybrid models show that on average r_s rises more steeply with increasing VPD but flattens at very high VPD (compare fit lines in Fig. 8 a-c). Future research will need 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 (Fig. 8 a-c) is much more precise than its conventional derivation by inverting the Penman-Monteith equation while making assumptions for r_a (Fig. 8d). 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 between 0-30% (Wehr & 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 (Fig. 8f) and *a priori* constraint (Fig. 8g), consistently reflect the expected negative logarithmic relationship of $\hat{r}_a$ against U^*(Fig. 8-9). 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 (Fig. 8f, g and Fig. 9f, g). Thus, they are capable of capturing the canopy heterogeneity across the sites and are more flexible than the conventional rigid parameterizations shown in Fig 8h (forests) and Fig. 9h (grasslands), where r_a is a homogenous function of U^* across the different sites.

The under-constrained hybrid model (Fig. 8e), however, illustrates the risk of equifinality and physics-violating results in this approach. In other words, $\hat{r}_a$ exhibits physically inconsistent relationships in the under-constrained model across the sites (Fig. 8e), while the predicted $\hat{r}_s$ and $\hat{Q}_{\mathrm{LE}}$ retain physically plausible estimates (Fig. 8a and Fig. 7 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 (Fig 8e and 9e) (Chen & Liu, 2020).

The 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}. There uncertainties in Q_{LE} and Q_{H} mainly arise from the uncertainty in estimating in r_a for both dense and sparse canopy, and particularly for arid and semi-arid conditions (Trebs et al., 2021). Our multi-task learning hybrid model, however, is able to provide a 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} (Fig. 5) and Q_{H} (Fig. 6c, d) at the US-Var grassland site, characterized by a dry Mediterranean-type climate (Xu & Baldocchi, 2004; De Kauwe et al., 2017), are fairly accurate and relate to physically consistent r_a predictions.

To get an estimate of uncertainty for the inferred relationships for r_s and r_a, we train each model five times with random initializations (ref to methods). 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 uninterpretable r_a for all initializations, especially for forests while the constrained hybrid models are able to reproduce consistently the physically plausible relationships for r_s and r_a. Hence, we show that our hybrid modeling approach yields robust predictions, yet, we stress the caveats related to equifinality in under-constrained model setups.

Lastly, we compare the behavior of surface conductance (g_s) against $Q_{\mathrm{LE}_{\mathrm{obs}}}$ with varying VPD at the mean diurnal scale for the multi-task learning model, the most promising approach, and the conventionally analyzed inverted PM equation for selected sites (Fig. 10). Both agree on a quasi-linear relationship between $\mathrm{g_s}$ and $Q_{\mathrm{LE}_{\mathrm{obs}}}$ with a gradient in g_{s} (y direction) with changing VPD. So, as VPD increases, the g_s decreases for the same level of evapotranspiration. This is consistent with the findings of Monteith (1995) whereby model estimates reflect the surface feedback response where a decrease in g_s as VPD increases is a result of a direct increase in transpiration lowering leaf water potential (Streck, 2003; Mallick et al., 2013, 2016). The general behavior of g_s is similar between the multi-task learning (Fig. 10b, d) model and the PM Inv model (Fig. 10a, c), however, the estimation of g_s alongside changing $Q_{\mathrm{LE}_{\mathrm{obs}}}$ in the multi-task learning model is less sensitive to noise at low $Q_{\mathrm{LE}_{\mathrm{obs}}}$ compared to the PM Inv. Overall, g_s based on the inverted PM equation is considerably higher than based on the hybrid modeling approach. The higher estimation could constitute a systematic bias in g_s rooted in the inversion of PM. In particular, for dense canopies, the overestimation could be related to the non-linear relationship of the stomata to light, as is the case for the DE-Tha forest (Fig. 10a) (Campbell, G. S., & Norman, 1998; Irmak, S. et al., 2008). In grasslands, like DE-Gri (Fig. 10c), the overestimation could be attributed to the propagation of measurement error in deriving the energy balance (Wohlfahrt et al., 2009; Knauer et al., 2018). In summary, the multi-task learning model not only provides more confined but also lower estimates for g_s in contrast to widely used inversion method.

4 Conclusions

We present approaches of 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 compared to measurements of latent heat fluxes at different scales, ranging from daily to seasonal to interannual variations. The 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 our approach are data-driven parameterizations for r_s and r_a jointly estimated by two separate neural networks. We show that the neural networks 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 neural networks. 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.

In the determination of r_a, we find considerable variation between sites compared to the very uniform empirical formulations conventionally 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 heterogeneity caused by the specific vegetation canopy structure. Also in the determination of r_s, the parameterizations derived from hybrid modeling show differences between sites, highlighting in particular the different physiological functions of the different plant types. 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 Penman-Monteith 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 (Winkler, et al., 2019 a,b). 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 biogeochemical processes in land-atmosphere coupling.

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).

Acknowledgements

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. All authors contributed ideas and to the interpretation of the results. R.E.G. and A.J.W. drafted the manuscript with inputs from all authors.

References

Ajami, H. (2021). Geohydrology: Global Hydrological Cycle. In *Encyclopedia of Geology* (pp. 393–398). Elsevier. https://doi.org/10.1016/b978-0-12-409548-9.12387-5

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., … Wofsy, S. (2001). FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bulletin of the American Meteorological Society*, *82*(11), 2415–2434. https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2

Besnard, S., Carvalhais, N., Arain, M. A., Black, A., Brede, B., Buchmann, N., Chen, J., Clevers, J. G. P. W., Dutrieux, L. P., Gans, F., Herold, M., Jung, M., Kosugi, Y., Knohl, A., Law, B. E., Paul-Limoges, E., Lohila, A., Merbold, L., Roupsard, O., … Reichstein, M. (2019). Memory effects of climate and vegetation affecting net ecosystem CO2 fluxes in global forests. *PLOS ONE*, *14*(2), e0211510. https://doi.org/10.1371/journal.pone.0211510

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1–2), 18–36. https://doi.org/10.1016/J.JHYDROL.2005.07.007

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., Lawrence, D. M., & Swenson, S. C. (2011). Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET dataJournal of Geophysical Research: Biogeosciences 116, no. G2 (2011). *Journal of Geophysical Research: Biogeosciences*, *116*(G2). https://doi.org/10.1029/2010JG001593

Brutsaert, W. (2005). *Hydrology: an introduction*. Cambridge University Press.

Buckley, T. N. (2005). The control of stomata by water balance. *New Phytologist*, *168*(2), 275–292. https://doi.org/10.1111/J.1469-8137.2005.01543.X

Campbell, G. S., & Norman, J. M. (1998). *An Introduction to Environmental Biophysics*. Springer-Verlag.

Carminati, A., & Javaux, M. (2020). Soil Rather Than Xylem Vulnerability Controls Stomatal Response to Drought. *Trends in Plant Science*, *25*(9), 868–880. https://doi.org/10.1016/J.TPLANTS.2020.04.003

Carter, C., & Liang, S. (2019). Evaluation of ten machine learning methods for estimating terrestrial evapotranspiration from remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *78*, 86–92. https://doi.org/10.1016/J.JAG.2019.01.020

Chaves, M. M., Costa, J. M., Zarrouk, O., Pinheiro, C., Lopes, C. M., & Pereira, J. S. (2016). Controlling stomatal aperture in semi-arid regions—The dilemma of saving water or being cool? *Plant Science*, *251*, 54–64. https://doi.org/10.1016/J.PLANTSCI.2016.06.015

Chehbouni, A., Lo Seen, D., Njoku, E. G., & Monteny, B. M. (1996). Examination of the difference between radiative and aerodynamic surface temperatures over sparsely vegetated surfaces. *Remote Sensing of Environment*, *58*(2), 177–186. https://doi.org/10.1016/S0034-4257(96)00037-5

Chen, J. M., & Liu, J. (2020). Evolution of evapotranspiration models using thermal and shortwave remote sensing data. *Remote Sensing of Environment*, *237*, 111594. https://doi.org/10.1016/J.RSE.2019.111594

Chen, Y., Xia, J., Liang, S., Feng, J., Fisher, J. B., Li, X., Li, X., Liu, S., Ma, Z., Miyata, A., Mu, Q., Sun, L., Tang, J., Wang, K., Wen, J., Xue, Y., Yu, G., Zha, T., Zhang, L., … Yuan, W. (2014). Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. *Remote Sensing of Environment*, *140*, 279–293. https://doi.org/10.1016/j.rse.2013.08.045

Damour, G., Simonneau, T., Cochard, H., & Urban, L. (2010). An overview of models of stomatal conductance at the leaf level. In *Plant, Cell and Environment* (Vol. 33, Issue 9, pp. 1419–1438). Plant Cell Environ. https://doi.org/10.1111/j.1365-3040.2010.02181.x

de Bezenac, E., Pajot, A., & Gallinari, P. (2017). Deep Learning for Physical Processes: Incorporating Prior Scientific Knowledge. *ArXiv*. http://arxiv.org/abs/1711.07970

De Kauwe, M. G., Medlyn, B. E., Knauer, J., & Williams, C. A. (2017). Ideas and perspectives: How coupled is the vegetation to the boundary layer? *Biogeosciences*, *14*(19), 4435–4453. https://doi.org/10.5194/BG-14-4435-2017

Dou, X., & Yang, Y. (2018). Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. *Computers and Electronics in Agriculture*, *148*, 95–106. https://doi.org/10.1016/j.compag.2018.03.010

Farquhar, G. D. (1978). Feedforward Responses of Stomata to Humidity. *Functional Plant Biology*, *5*(6), 787–800. https://doi.org/10.1071/PP9780787

Friedl, M. A. (1996). Relationships among remotely sensed data, surface energy balance, and area-averaged fluxes over partially vegetated land surfaces. *Journal of Applied Meteorology and Climatology*, *35*(11), 2091–2103. https://doi.org/10.1175/1520-0450(1996)035<2091:RARSDS>2.0.CO;2

Garratt, J. R. (1992). *The Atmospheric Boundary Layer*. Cambridge University Press.

Gerosa, G., Mereu, S., Finco, A., & Marzuoli, R. (2012). Stomatal Conductance Modeling to Estimate the Evapotranspiration of Natural and Agricultural Ecosystems. In *Evapotranspiration - Remote Sensing and Modeling*. InTech. https://doi.org/10.5772/21825

Halladay, K., & Good, P. (2017). Non-linear interactions between CO 2 radiative and physiological effects on Amazonian evapotranspiration in an Earth system model. *Climate Dynamics*, *49*(7–8), 2471–2490. https://doi.org/10.1007/s00382-016-3449-0

Hetherington, A. M., & Woodward, F. I. (2003). The role of stomata in sensing and driving environmental change. *Nature*, *424*(6951), 901–908. https://doi.org/10.1038/nature01843

Irmak, S., M., Irmak, A., Arkebauer, T. J., Weiss, A., Martin, D. L., & Eisenhauer, D. E. (2008). On the scaling up leaf stomatal resistance to canopy resistance using photosynthetic photon flux density. *Agricultural and Forest Meteorology*, *148*(6–7), 1034–1044.

Jain, S. K., Nayak, P. C., & Sudheer, K. P. (2008). Models for estimating evapotranspiration using artificial neural networks, and their physical interpretation. *Hydrological Processes*, *22*(13), 2225–2234. https://doi.org/10.1002/hyp.6819

Jarvis, P. G. (1976). The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *273*(927), 593–610. https://doi.org/10.1098/RSTB.1976.0035

Jarvis, P. G., & Stewart, J. B. (1979). Evaporation of water from plantation forest. *D. Ford, D.C. Malcolm, J. Atterson (Eds.), The Ecology of Even-Aged Forest Plantations, Institute of Terrestrial Ecology, Cambridge*, 327–350.

Jia, X., Zwart, J., Sadler, J., Appling, A., Oliver, S., Markstrom, S., Willard, J., Xu, S., Steinbach, M., Read, J., & Kumar, V. (2020). Physics-Guided Recurrent Graph Model for Predicting Flow and Temperature in River Networks. *ArXiv Preprint*.

Jung, M., Reichstein, M., & 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*(10), 2001–2013. https://doi.org/10.5194/BG-6-2001-2009

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., De Jeu, R., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., … Zhang, K. (2010). Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, *467*(7318), 951–954. https://doi.org/10.1038/nature09396

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., S Goll, D., Haverd, V., Köhler, P., Ichii, K., K Jain, A., Liu, J., Lombardozzi, D., … Reichstein, M. (2020). Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach. *Biogeosciences*, *17*(5), 1343–1365. https://doi.org/10.5194/bg-17-1343-2020

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., & Kumar, V. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, *29*(10), 2318–2.

Karpatne, A., Watkins, W., Read, J., & Kumar, V. (2017). Physics-guided Neural Networks (PGNN): An Application in Lake Temperature Modeling. *ArXiv*. http://arxiv.org/abs/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., & Schulze, E. D. (1992). Evaporation, xylem sap flow, and tree transpiration in a New Zealand broad-leaved forest. *Agricultural and Forest Meteorology*, *62*(1–2), 53–73. https://doi.org/10.1016/0168-1923(92)90005-O

Kennedy, D., Swenson, S., Oleson, K. W., Lawrence, D. M., Fisher, R., Lola da Costa, A. C., & Gentine, P. (2019). Implementing Plant Hydraulics in the Community Land Model, Version 5. *Journal of Advances in Modeling Earth Systems*, *11*(2), 485–513. https://doi.org/10.1029/2018MS001500

Knauer, J., El-Madany, T. S., Zaehle, S., & Migliavacca, M. (2018). Bigleaf—An R package for the calculation of physical and physiological ecosystem properties from eddy covariance data. *PLOS ONE*, *13*(8), e0201114. https://doi.org/10.1371/JOURNAL.PONE.0201114

Köstner, B. M. M., Schulze, E. D., Kelliher, F. M., Hollinger, D. Y., Byers, J. N., Hunt, J. E., McSeveny, T. M., Meserth, R., & 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*(3), 350–359. https://doi.org/10.1007/BF00317623

Kraft, B., Jung, M., Körner, M., Koirala, S., & Reichstein, M. (2022). Towards hybrid modeling of the global hydrological cycle. *Hydrology and Earth System Sciences*, *26*(6), 1579–1614. https://doi.org/10.5194/HESS-26-1579-2022

Kraft, B., Jung, M., Körner, M., & Reichstein, M. (2020). Hybrid modeling: Fusion of a deep approach and physics-based model for global hydrological modeling. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *43*, 1537–1544. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1537-2020

Krasnopolsky, V. M. (2013). *The application of neural networks in the Earth system sciences*.

Li, L., Wang, Y., Arora, V. K., Eamus, D., Shi, H., Li, J., Cheng, L., Cleverly, J., Hajima, T., Ji, D., Jones, C., Kawamiya, T., Li, W., Tjiputra, J., Wiltshire, A., Zhang, L., & Yu, Q. (2018). Evaluating global land surface models in CMIP5: Analysis of ecosystem water- and light-use efficiencies and rainfall partitioning. *Journal of Climate*, *31*(8), 2995–3008. https://doi.org/10.1175/JCLI-D-16-0177.1

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. https://doi.org/10.1016/J.AGRFORMET.2018.11.017

Liebel, L., & Körner, M. (2018). *Auxiliary Tasks in Multi-task Learning*. https://doi.org/10.48550/arxiv.1805.06334

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. https://doi.org/10.1016/j.agrformet.2017.12.078

Lin, H., Chen, Y., Zhang, H., Fu, P., & 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. *Functional Ecology*, *31*(12), 2202–2211. https://doi.org/10.1111/1365-2435.12923/SUPPINFO

Liu, S., Mao, D., & Lu, L. (2006). Measurement and estimation of the aerodynamic resistance. *European Geosciences Union*, *3*(3), 681–705. www.hydrol-earth-syst-sci-discuss.net/3/681/2006/

Mallick, K., Jarvis, A., Fisher, J. B., Tu, K. P., Boegh, E., & Niyogi, D. (2013). Latent Heat Flux and Canopy Conductance Based on Penman–Monteith, Priestley–Taylor Equation, and Bouchet's Complementary Hypothesis. *Journal of Hydrometeorology*, *14*(2), 419–442. https://doi.org/10.1175/JHM-D-12-0117.1

Mallick, K., Trebs, I., Boegh, E., Giustarini, L., Schlerf, M., Drewry, D. T., Hoffmann, L., Von Randow, C., Kruijt, B., Araùjo, A., Saleska, S., Ehleringer, J. R., Domingues, T. F., Ometto, J. P. H. B., Nobre, A. D., Luiz Leal De Moraes, O., Hayek, M., William Munger, J., & Wofsy, S. C. (2016). Canopy-scale biophysical controls of transpiration and evaporation in the Amazon Basin. *Hydrology and Earth System Sciences*, *20*(10), 4237–4264. https://doi.org/10.5194/HESS-20-4237-2016

Massmann, A., Gentine, P., & Lin, C. (2019). When Does Vapor Pressure Deficit Drive or Reduce Evapotranspiration? *Journal of Advances in Modeling Earth Systems*, *11*(10), 3305–3320. https://doi.org/10.1029/2019MS001790

Maurer, K. D., Bohrer, G., Kenny, W. T., & Ivanov, V. Y. (2015). Large-eddy simulations of surface roughness parameter sensitivity to canopy-structure characteristics. *Biogeosciences*, *12*(8), 2533–2548. https://doi.org/10.5194/BG-12-2533-2015

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., & Wingate, L. (2011). Reconciling the optimal and empirical approaches to modelling stomatal conductance. *Global Change Biology*, *17*(6), 2134–2144. https://doi.org/10.1111/J.1365-2486.2010.02375.X

Monteith, J. L. (1995). A reinterpretation of stomatal responses to humidity. *Plant, Cell & Environment*, *18*(4), 357–364. https://doi.org/10.1111/J.1365-3040.1995.TB00371.X

Monteith, J. L. (1965). Evaporation and environment. *Fogg (Ed.), Symposium of the Society for Experimental Biology, The State and Movement of Water in Living Organisms, Vol. 19*, 205–234.

Monteith, J. L. (1985). Evaporation from land surfaces: progress in analysis and prediction since 1948. *In: Advances in Evapotranspiration, Proc. National Conference on Advances in Evapotranspiration. Am.: Soc. Agric. Eng., St. Joseph, MI, Chicago, IL*, 4–12. https://agris.fao.org/agris-search/search.do?recordID=US8644525

Monteith, J. L., & Unsworth, M. (2013). *Principles of environmental physics: plants, animals, and the atmosphere* (Fourth). Elsevier Ltd.

Mott, K. A., & Peak, D. (2013). Testing a vapour-phase model of stomatal responses to humidity. *Plant, Cell & Environment*, *36*(5), 936–944. https://doi.org/10.1111/PCE.12026

Mystakidis, S., Davin, E. L., Gruber, N., & Seneviratne, S. I. (2016). Constraining future terrestrial carbon cycle projections using observation-based water and carbon flux estimates. *Global Change Biology*, *22*(6), 2198–2215. https://doi.org/10.1111/gcb.13217

Nakai, T., Sumida, A., Daikoku, K., Matsumoto, K., van der Molen, M. K., Kodama, Y., Kononov, A. V., Maximov, T. C., Dolman, A. J., Yabuki, H., Hara, T., & Ohta, T. (2008). Parameterisation of aerodynamic roughness over boreal, cool- and warm-temperate forests. *Agricultural and Forest Meteorology*, *148*(12), 1916–1925. https://doi.org/10.1016/J.AGRFORMET.2008.03.009

Penman, H. L. (1948). Natural evaporation from open water, hare soil and grass. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *193*(1032), 120–145. https://doi.org/10.1098/rspa.1948.0037

Polhamus, A., Fisher, J. B., & Tu, K. P. (2013). What controls the error structure in evapotranspiration models? *Agricultural and Forest Meteorology*, *169*, 12–24. https://doi.org/10.1016/j.agrformet.2012.10.002

Reichstein, M., Ahrens, B., Kraft, B., Camps-Valls, G., Carvalhais, N., Gans, F., Gentine, P., & Winkler, A. J. (2022). Combining System Modeling and Machine Learning into Hybrid Ecosystem Modeling. *Knowledge-Guided Machine Learning*, 327–352. https://doi.org/10.1201/9781003143376-14

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grunwald, T., Havrankova, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., … Valentini, R. (2005). On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology*, *11*(9), 1424–1439. https://doi.org/10.1111/j.1365-2486.2005.001002.x

Ronda, R. J., De Bruin, H. A. R., & Holtslag, A. A. M. (2001). Representation of the canopy conductance in modeling the surface energy budget for low vegetation. *Journal of Applied Meteorology*, *40*(8), 1431–1444.

Schmidt, L., Heße, F., Attinger, S., & Kumar, R. (2020). Challenges in Applying Machine Learning Models for Hydrological Inference: A Case Study for Flooding Events Across Germany. *Water Resources Research*, *56*(5), e2019WR025924. https://doi.org/10.1029/2019WR025924

Schulze, E. D. (1986). Carbon dioxide and water vapor exchange in response to drought in the atmosphere and in the soil. *Annu. Rev. Plant Physiol.; (United States)*, *37*(1), 247–274. https://doi.org/10.1146/ANNUREV.PP.37.060186.001335

Sellers, P. J., Dickinson, R. E., Randall, D. A., Betts, A. K., Hall, F. G., Berry, J. A., Collatz, G. J., Denning, A. S., Mooney, H. A., Nobre, C. A., Sato, N., Field, C. B., & Henderson-Sellers, A. (1997). Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. In *Science* (Vol. 275, Issue 5299, pp. 502–509). American Association for the Advancement of Science. https://doi.org/10.1126/science.275.5299.502

Shaw, R. H., & Pereira, A. R. (1982). Aerodynamic roughness of a plant canopy: A numerical experiment. *Agricultural Meteorology*, *26*(1), 51–65. https://doi.org/10.1016/0002-1571(82)90057-7

Stewart, J. B. (1988). Modelling surface conductance of pine forest. *Agricultural and Forest Meteorology*, *43*(1), 19–35. https://doi.org/10.1016/0168-1923(88)90003-2

Streck, N. A. (2003). Stomatal response to water vapor pressure deficit: an unsolved issue. *Current Agricultural Science and Technology*, *9*(4), 317–322.

Su, Y., Zhang, C., Chen, X., Liu, L., Ciais, P., Peng, J., Wu, S., Wu, J., Shang, J., Wang, Y., Yuan, W., Yang, Y., Wu, Z., & Lafortezza, R. (2021). Aerodynamic resistance and Bowen ratio explain the biophysical effects of forest cover on understory air and soil temperatures at the global scale. *Agricultural and Forest Meteorology*, *308–309*, 108615. https://doi.org/10.1016/J.AGRFORMET.2021.108615

Tan, S., Wang, H., Prentice, I. C., & Yang, K. (2021). Land-surface evapotranspiration derived from a first-principles primary production model. *Environmental Research Letters*, *16*(10), 104047. https://doi.org/10.1088/1748-9326/AC29EB

Trebs, I., Mallick, K., Bhattarai, N., Sulis, M., Cleverly, J., Woodgate, W., Silberstein, R., Hinko-Najera, N., Beringer, J., Meyer, W. S., Su, Z., & Boulet, G. (2021). The role of aerodynamic resistance in thermal remote sensing-based evapotranspiration models. *Remote Sensing of Environment*, *264*, 112602. https://doi.org/10.1016/J.RSE.2021.112602

Vialet-Chabrand, S., & Lawson, T. (2019). Dynamic leaf energy balance: deriving stomatal conductance from thermal imaging in a dynamic environment. *Journal of Experimental Botany*, *70*(10), 2839. https://doi.org/10.1093/JXB/ERZ068

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. https://doi.org/10.1016/J.AGRFORMET.2013.07.005

Vila-Guerau de Arellano, J., C. van Heerwaarden, C., J. H. van Stratum, B., & van den Dries, K. (2015). *Atmospheric Boundary Layer*. Cambridge University Press. https://doi.org/10.1017/CBO9781316117422

Wang, K., & Dickinson, R. E. (2012). A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Reviews of Geophysics*, *50*(2). https://doi.org/10.1029/2011RG000373

Wehr, R., & Saleska, S. R. (2021). Calculating canopy stomatal conductance from eddy covariance measurements, in light of the energy budget closure problem. *Biogeosciences*, *18*(1), 13–24. https://doi.org/10.5194/BG-18-13-2021

Winkler, A. J., Myneni, R. B., Alexandrov, G. A., & Brovkin, V. (2019). Earth system models underestimate carbon fixation by plants in the high latitudes. *Nature Communications 2019 10:1*, *10*(1), 1–8. https://doi.org/10.1038/s41467-019-08633-z

Winkler, A. J., Myneni, R. B., & Brovkin, V. (2019). Investigating the applicability of emergent constraints. *Earth System Dynamics*, *10*(3), 501–523. https://doi.org/10.5194/ESD-10-501-2019

Wohlfahrt, G., Haslwanter, A., Hörtnagl, L., Jasoni, R. L., Fenstermaker, L. F., Arnone, J. A., & Hammerle, A. (2009). On the consequences of the energy imbalance for calculating surface conductance to water vapour. *Agricultural and Forest Meteorology*, *149*(9), 1556–1559. https://doi.org/10.1016/J.AGRFORMET.2009.03.015

Xu, L., & Baldocchi, D. D. (2004). Seasonal variation in carbon dioxide exchange over a Mediterranean annual grassland in California. *Agricultural and Forest Meteorology*, *123*(1–2), 79–96. https://doi.org/10.1016/J.AGRFORMET.2003.10.004

Yang, T., Sun, F., Gentine, P., Liu, W., Wang, H., Yin, J., Du, M., & Liu, C. (2019). Evaluation and machine learning improvement of global hydrological model-based flood simulations. *Environmental Research Letters*, *14*(11), 114027. https://doi.org/10.1088/1748-9326/ab4d5e

Zeppel, M., & 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. *Australian Journal of Botany*, *56*(2), 97–108. https://doi.org/10.1071/BT07091

Zhang, Z. Z., Zhao, P., McCarthy, H. R., Zhao, X. H., Niu, J. F., Zhu, L. W., Ni, G. Y., Ouyang, L., & Huang,

Y. Q. (2016). Influence of the decoupling degree on the estimation of canopy stomatal conductance for two broadleaf tree species. *Agricultural and Forest Meteorology*, *221*, 230–241. https://doi.org/10.1016/J.AGRFORMET.2016.02.018

Zhao, L., Xia, J., Xu, C. yu, Wang, Z., Sobkowiak, L., & Long, C. (2013). Evapotranspiration estimation methods in hydrological models. *Journal of Geographical Sciences 2013 23:2*, *23*(2), 359–369. https://doi.org/10.1007/S11442-013-1015-9

Zhao, W. L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., & Qiu, G. Y. (2019). Physics-Constrained Machine Learning of Evapotranspiration. *Geophysical Research Letters*, *46*(24), 14496–14507. https://doi.org/10.1029/2019GL085291

Zhou, S., Yu, B., Zhang, Y., Huang, Y., & Wang, G. (2016). Partitioning evapotranspiration based on the concept of underlying water use efficiency. *Water Resources Research*, *52*(2), 1160–1175. https://doi.org/10.1002/2015WR017766

Table 1: Detailed description of each site used derived from the FLUXNET 2015 Tier 1 data.

Site ID	IGBP	Elevation (m)	Mean Annual Temperature (°C)	Mean Annual Precipitation (mm)	Data Availability	DOI
DE-Tha	ENF[1]	385	8.2	843	19 years (1996 - 2014)	Christian Bernhofer, Thomas Grünwald, Uta Moderow, Markus Hehn, Uwe Eichelmann, Heiko Prasse, Udo Postel (1996-2014) FLUXNET2015 DE-Tha Tharandt, Dataset. https://doi.org/10.18140/FLX/1440152
FR-Pue	EBF[2]	270	13.5	883	15 years (2000 - 2014)	Jean-Marc Ourcival, Karim Piquemal, Richard Joffre, Limousin Jean-Marc (2000-2014) FLUXNET2015 FR-Pue Puechabon, Dataset. https://doi.org/10.18140/FLX/1440164
FR-LBr	ENF[1]	61	13.6	900	12 years (1996 - 2008)	Paul Berbigier, Jean Bonnefond, Alexandre Bosc, Pierre Trichet, Denis Loustau (1996-2008) FLUXNET2015 FR-LBr Le Bray, Dataset. https://doi.org/10.18140/FLX/1440163
CH-Cha	GRA[3]	393	9.5	1136	10 years (2005 - 2014)	Lutz Merbold, Kathrin Fuchs, Nina Buchmann, Lukas Hörtnagl (2012-2016) FLUXNET-CH4 CH-Cha Chamau, Dataset. https://doi.org/10.18140/FLX/1669629
DE-Gri	GRA[3]	385	7.8	901	11 years (2004 - 2014)	Christian Bernhofer, Thomas Grünwald, Uta Moderow, Markus Hehn, Uwe Eichelmann, Heiko Prasse, Udo Postel () FLUXNET2015 DE-Gri , Dataset. https://doi.org/10.18140/FLX/1440147
US-Var	GRA[3]	129	15.8	559	15 years (2000 - 2014)	(2000-2014) FLUXNET2015 US-Var Vaira Ranch- Ione, Dataset. https://doi.org/10.18140/FLX/1440094

1. ENF (Evergreen Needleleaf Forests: Lands dominated by woody vegetation with a percent cover >60% and height exceeding 2 meters. Almost all trees remain green all year. Canopy is never without green foliage).
2. EBF (Evergreen Broadleaf Forests: Lands dominated by woody vegetation with a percent cover >60% and height exceeding 2 meters. Almost all trees and shrubs remain green year-round. Canopy is never without green foliage).
3. GRA (Grasslands: Lands with herbaceous types of cover. Tree and shrub cover is less than 10%. Permanent wetlands lands with a permanent mixture of water and herbaceous or woody vegetation. The vegetation can be present in either salt, brackish, or fresh water.)

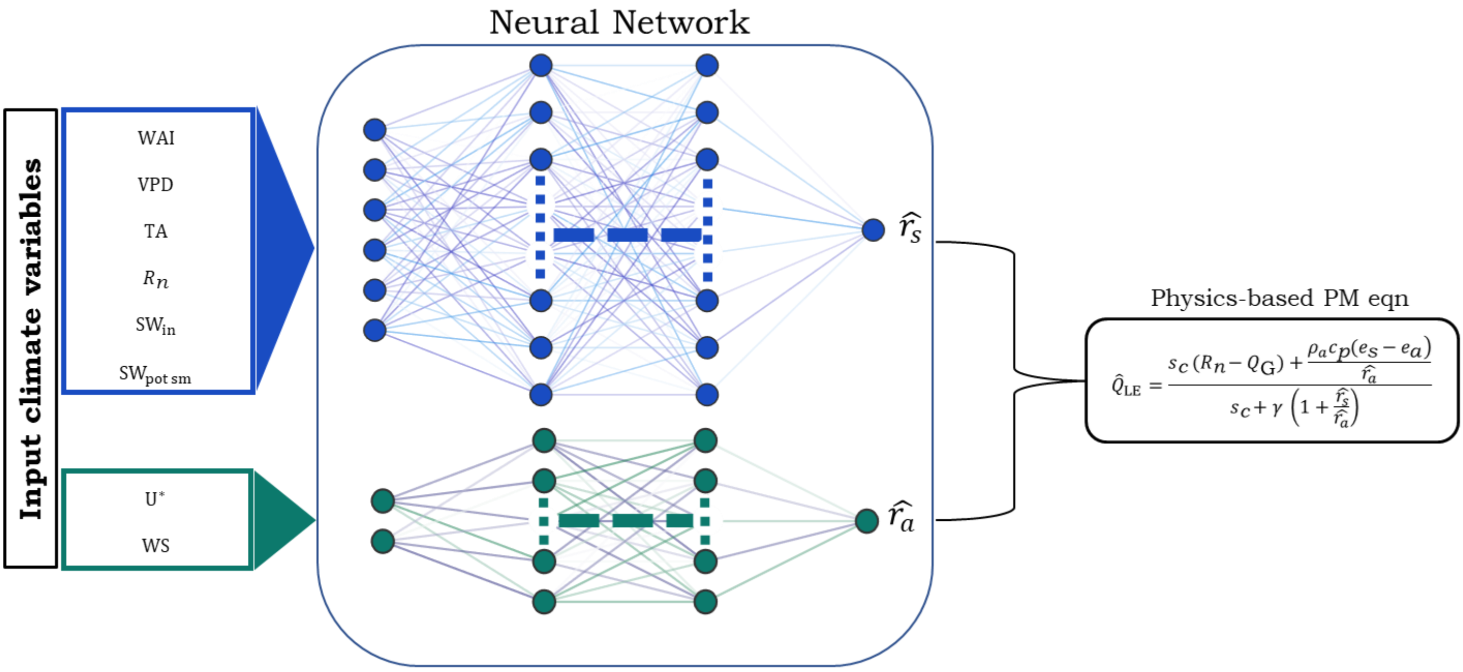


Figure 1: Architecture of the basic hybrid model consists of two neural networks, which estimate r_s and r_a individually with independent input climate variables. The latent variables are used in the Penman-Monteith equation to estimate the latent heat flux (Q_{LE}), and the objective function minimizes losses for Q_{LE}.

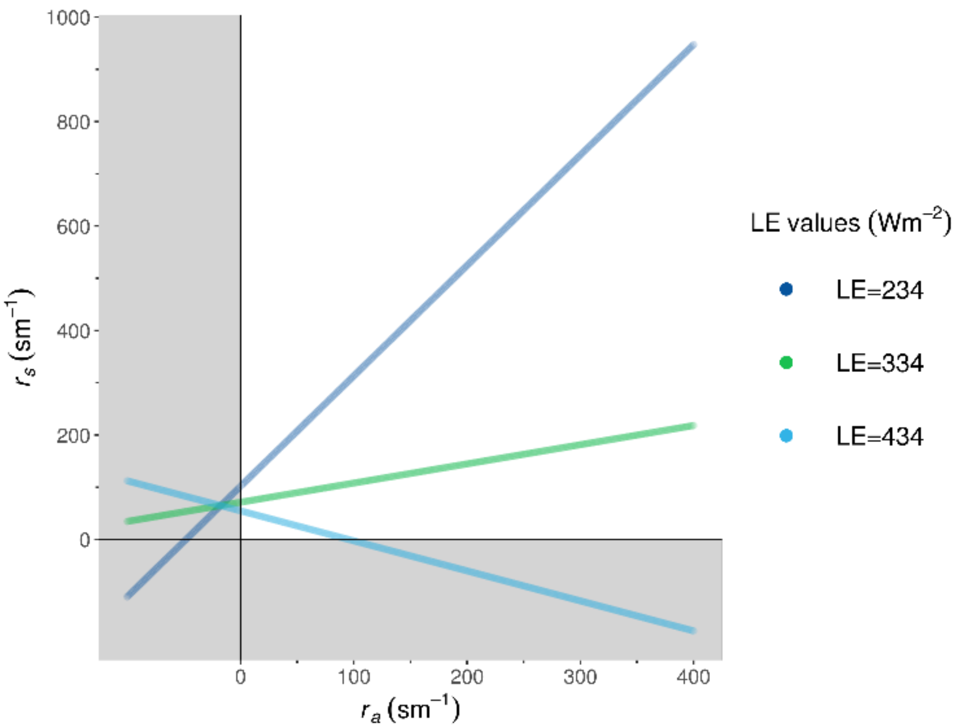
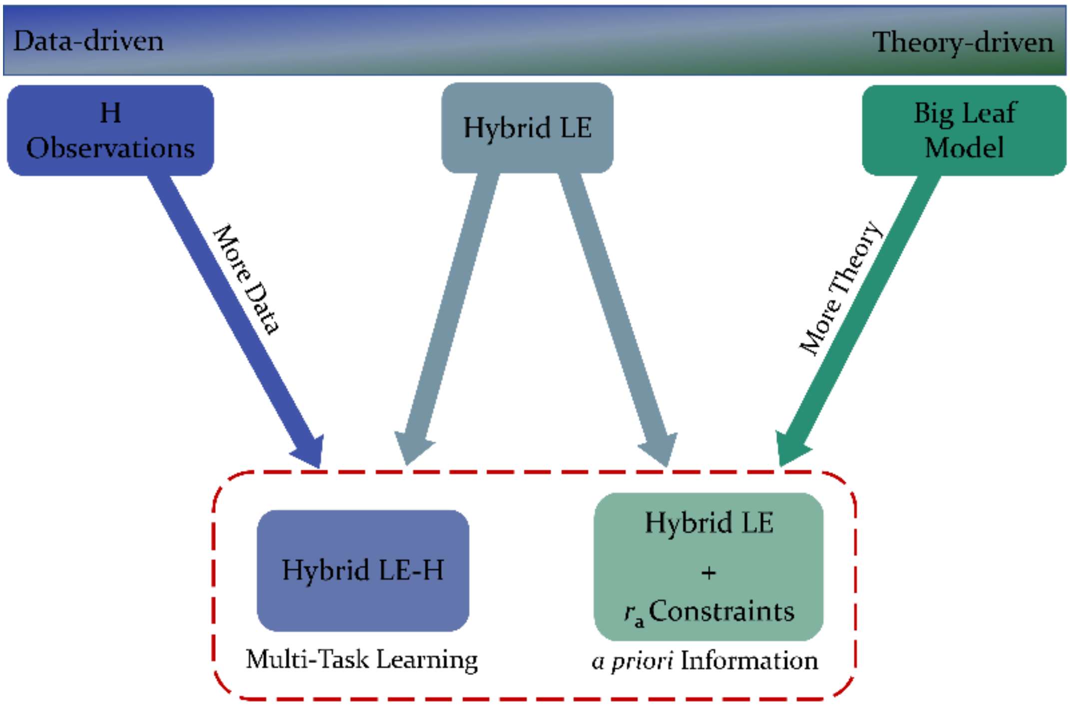


Figure 2: Equifinality in the physics-based component of hybrid model: The lines represent different Q_{LE} values that can exist for specific conditions (the actual Q_{LE} value is approximately 334 Wm^{-2}). Fixing all parameters of the PM equation s_c = 0.175 kPaC^{-1}, R_n= 520.38 Wm^{-2}, Q_G = 18.51 Wm^{-2}, VPD = 1.333 kPa, ρ_a = 1.143 kg m^{-3}, c_p =1004.834 J kg^{-1} C^{-1}, γ= 0.0644 kPaC^{-1}, the different combinations of r_s and r_a values lead to the same Q_{LE}. Shaded areas show the physically non-plausible and non-realistic values for r_s and r_a combinations, and non-shaded areas show physically plausible values.



730

Figure 3: Proposed methods for constraining the hybrid model: Right-side shows the theory-driven hybrid model with *a priori* constraint for r_a from the Big Leaf model. Left-side shows data-driven hybrid model with more information from learning an additional target variable Q_H through multi-task learning.

731

732

733

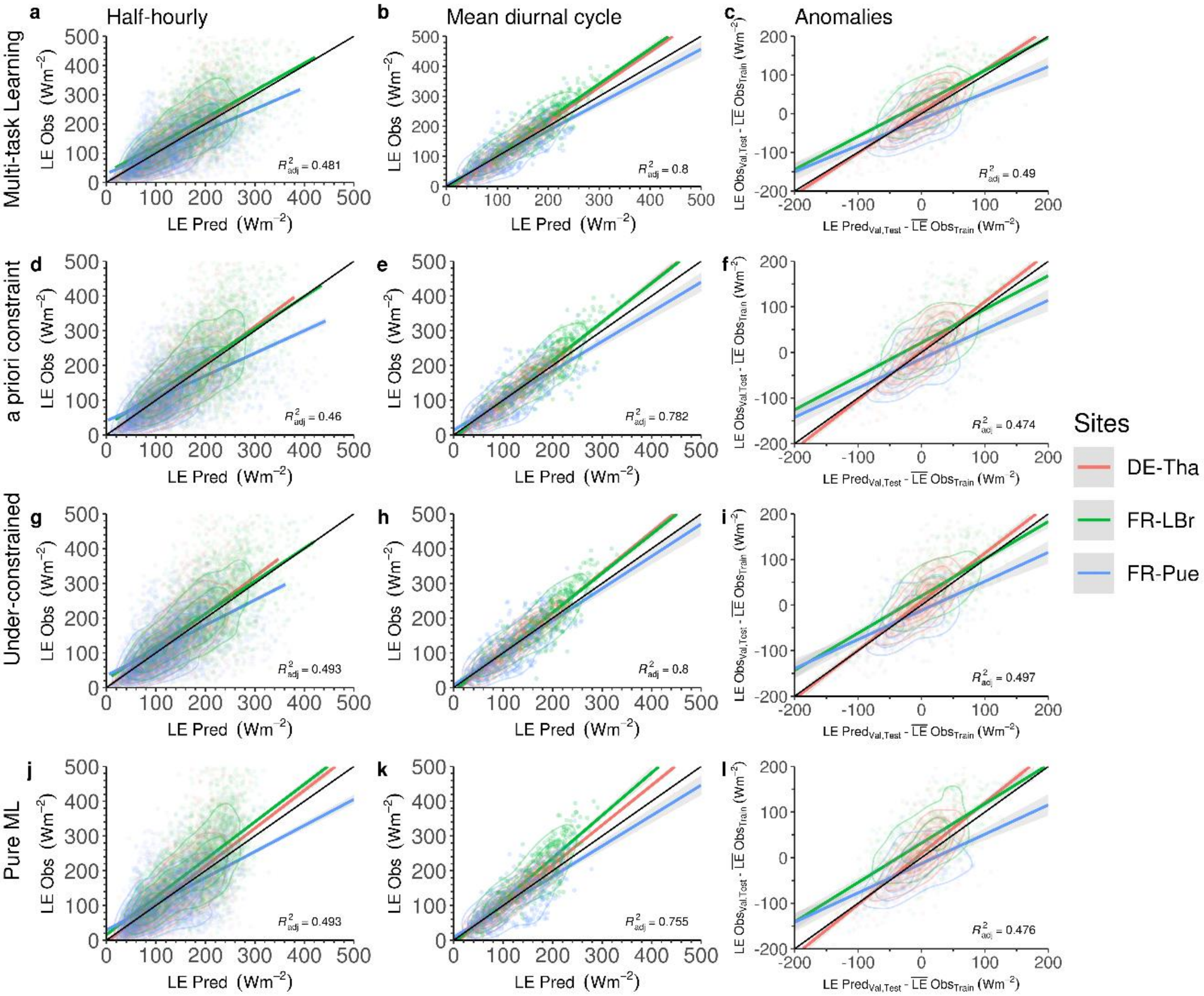


Figure 4: Evaluation of Q_{LE} observations and predictions at different temporal scales for forests. a,d,g,j show predictions against observations at a half-hourly scale for different models; b,e,h,k show predictions against observations at mean diurnal scale; c,f,i,l show Q_{LE} anomalies at interannual scale for the different models.

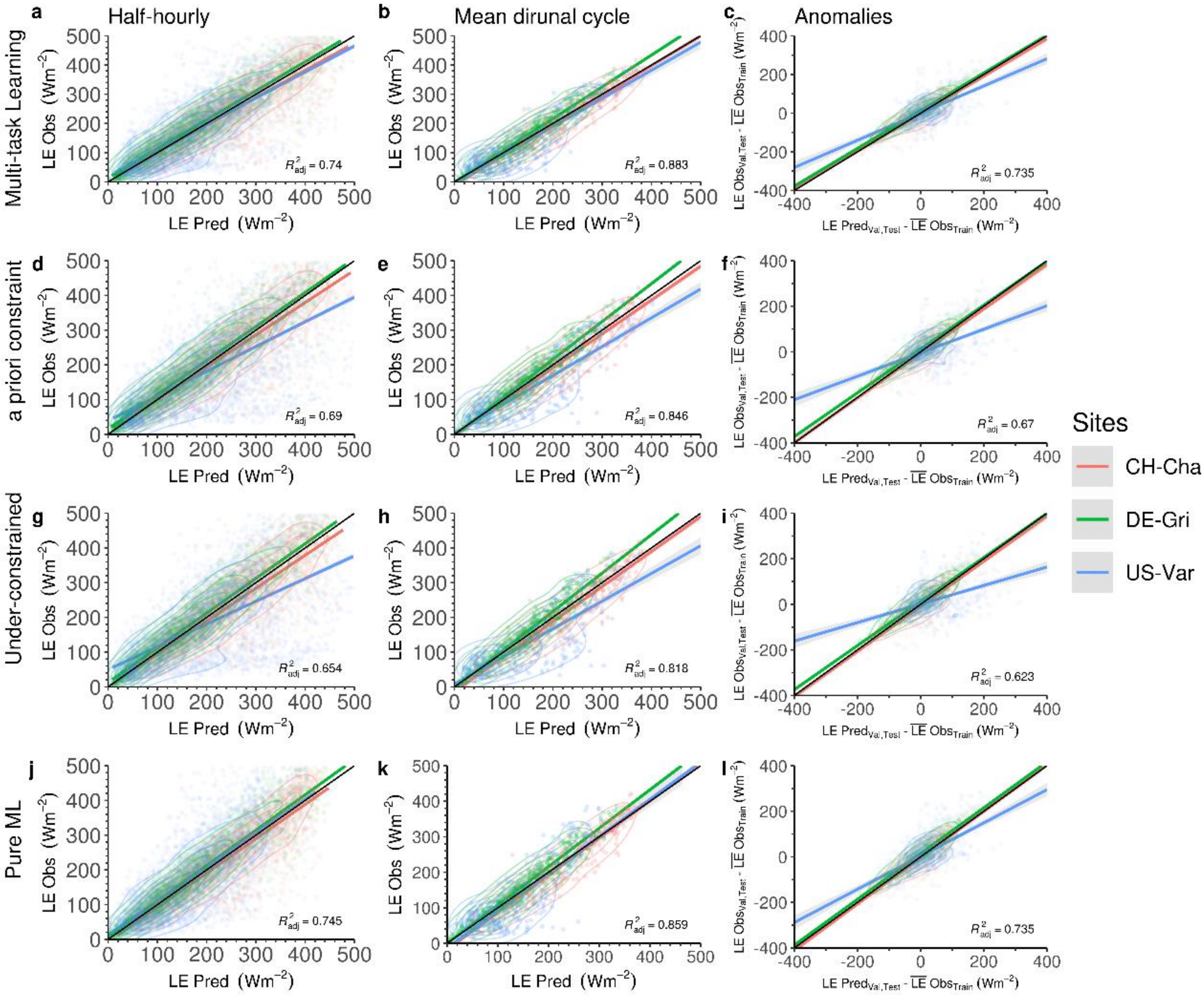


Figure 5: Evaluation of Q_{LE} observations and predictions at different temporal scales for grasslands. a,d,g,j show predictions against observations at a half-hourly scale for different models. b,e,h,k show predictions against observations at mean diurnal scale. c,f,i,l show Q_{LE} anomalies at interannual scale for the different models.

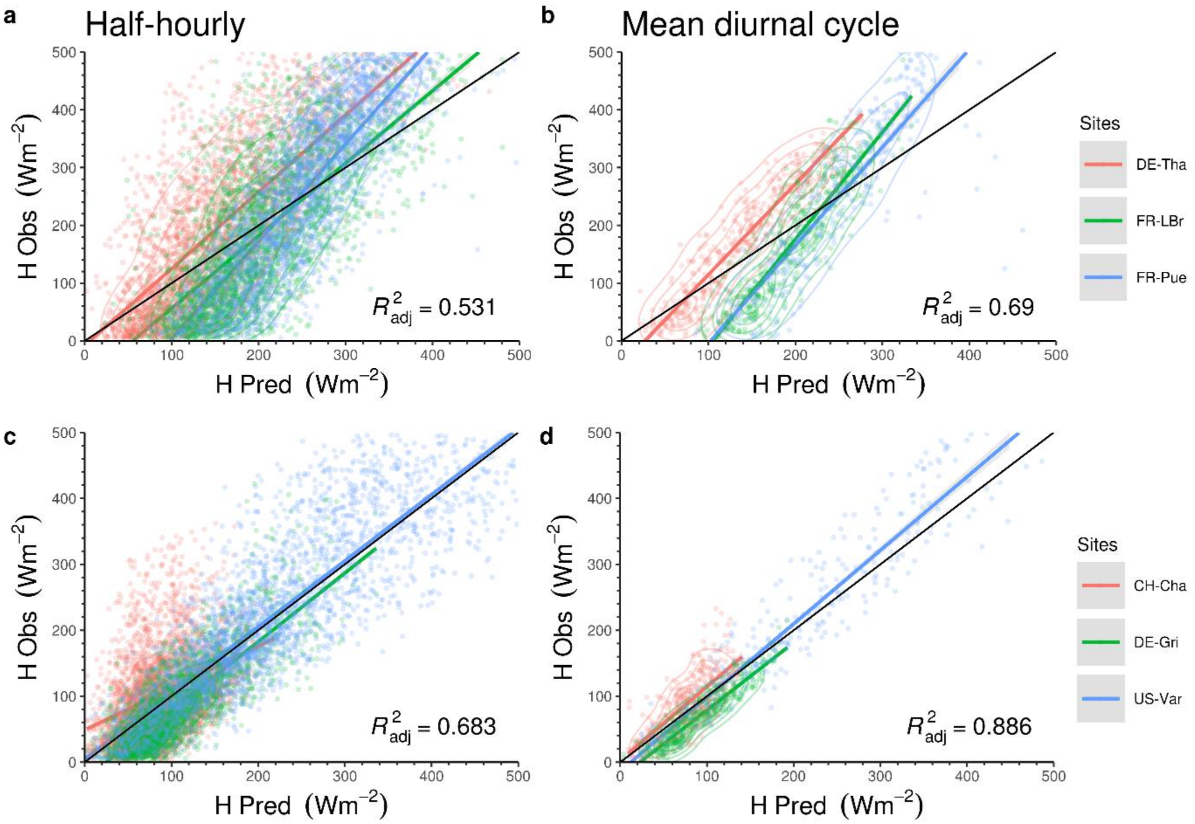


Figure 6: Evaluation of Q_{H} observations and predictions at half-hourly, and mean diurnal scale for forest (a,b) and grasslands (c,d) for multi-task learning hybrid model. Q_{H} predictions are similar in range compare to Q_{LE} predictions in figures 4-5 for forests and grasslands.

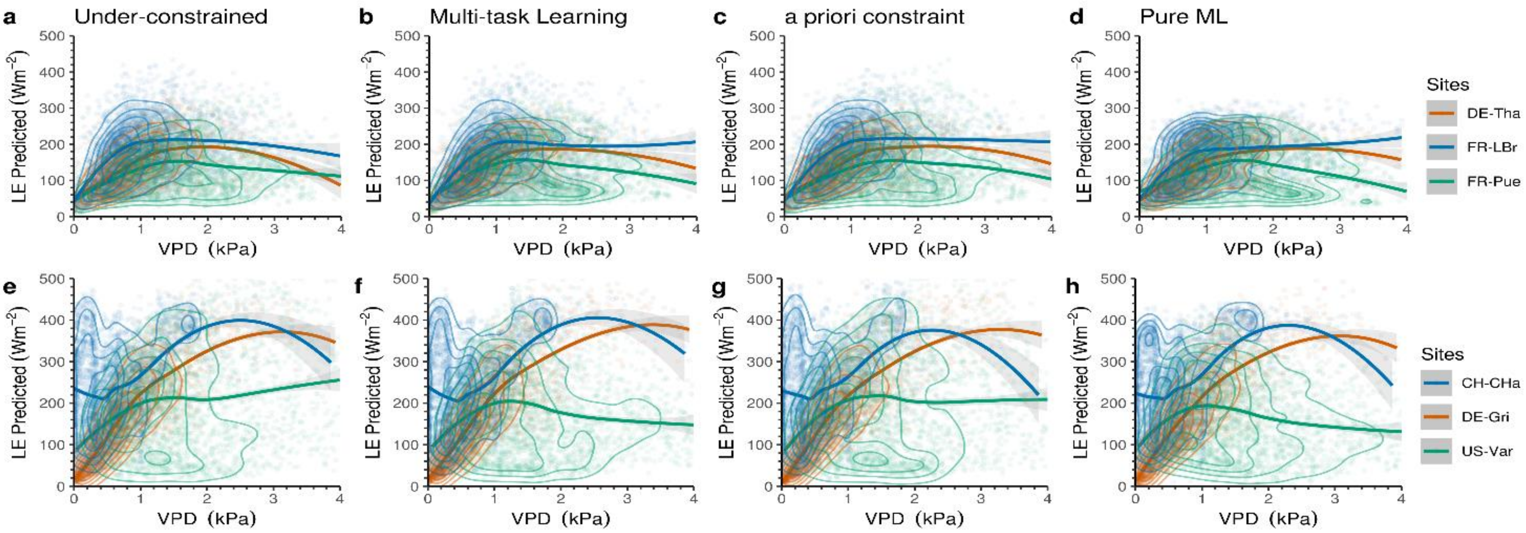


Figure 7: Evaluating Q_{LE} predictions against VPD for different models for forests (a-d) and grasslands (e-h). Higher evapotranspiration rates evident for grasslands compared to forests associated with higher stomatal conductance.

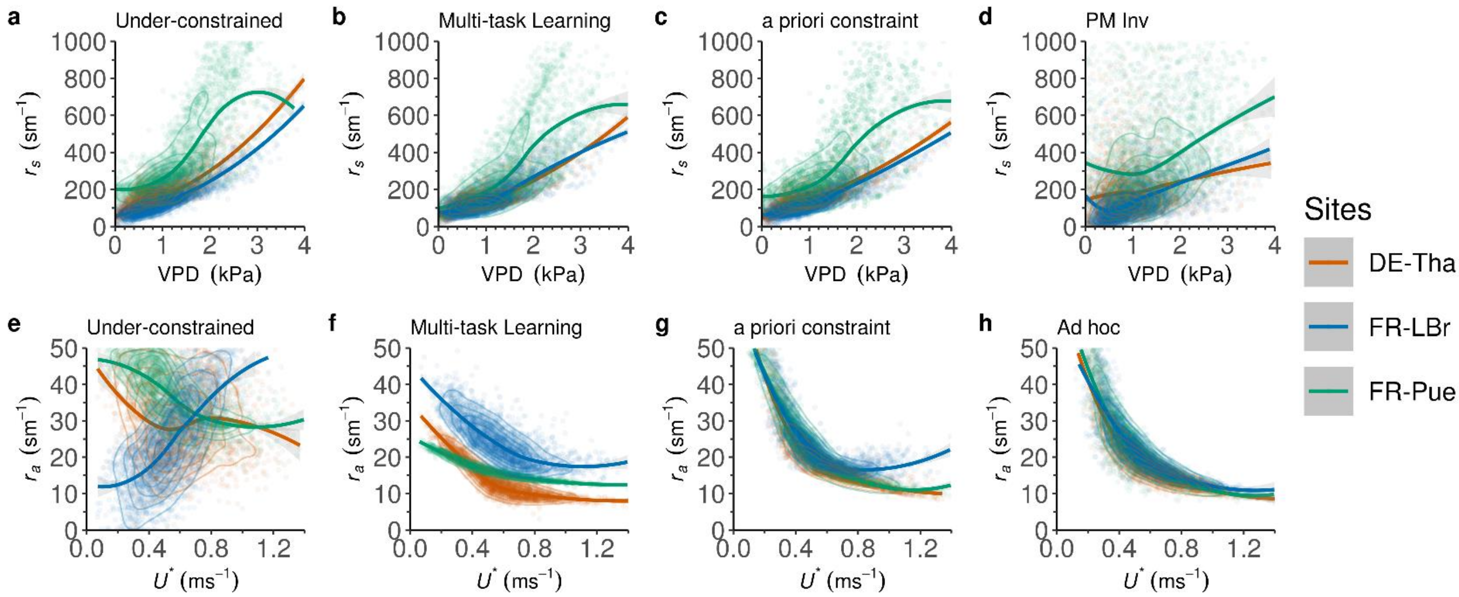


Figure 8: Assessing latent variables r_s and r_a against VPD and U* respectively for different models in forests. Constrained hybrid models reveal physical consistency of latent variables compared to under-constrained model, especially under different environmental conditions.

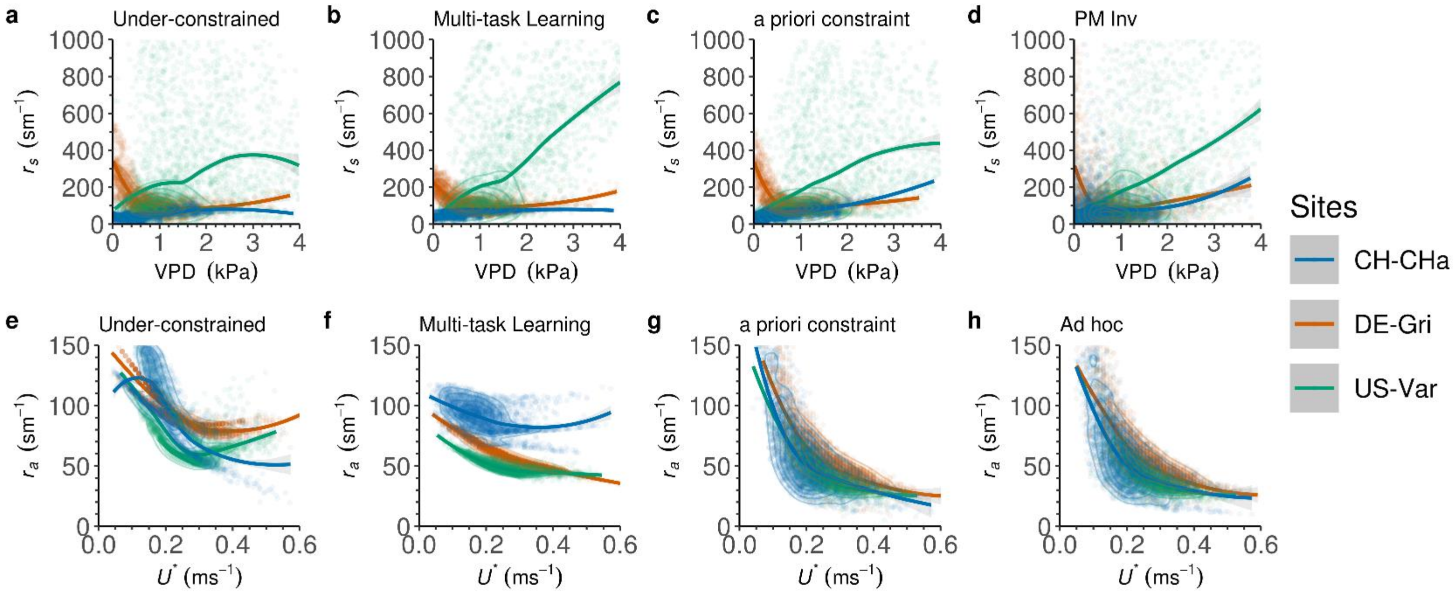


Figure 9: Assessing latent variables r_s and r_a against VPD and U* respectively for different models in grasslands. The constrained hybrid models yield more physically consistent results compared to under-constrained model, and able to capture the vegetation and climate heterogeneities.

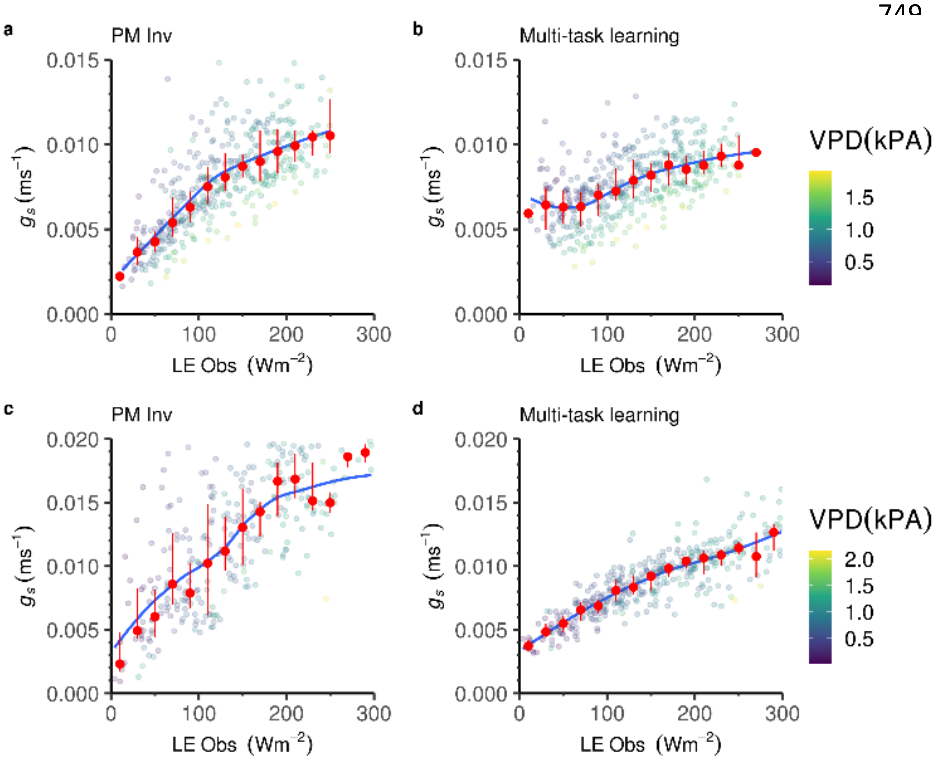


Figure 10: Physical consistency of g_s and $Q_{\mathrm{LE_{obs}}}$ with VPD at mean diurnal scale of DE-Tha forest (a,b) and DE-Gri grassland (c,d). The multi-task learning model is able to capture the same patterns as shown by Penman-Monteith, while being more resistant to noise in the data which may cause overestimation of surface conductance due to the instability of the inversion.