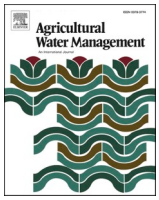
Agricultural Water Management 317 (2025) 109643

Contents lists available at ScienceDirect

Agricultural Water Management

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



Explainable machine learning to quantify the value of proximal remote sensing in latent energy flux estimation

Srishti Gaur [a], Guler Aslan-Sungur (Rojda) [b], Andy VanLoocke [b], Darren T. Drewry [a,c,d,*]

[a] Department of Food, Agricultural and Biological Engineering, Ohio State University, Columbus, OH, USA
[b] Department of Agronomy, Iowa State University, Ames, IA, USA
[c] Department of Horticulture and Crop Science, Ohio State University, Columbus, OH, USA
[d] Translational Data Analytics Institute, Ohio State University, Columbus, OH, USA

ARTICLE INFO

Keywords:
Proximal remote sensing
Explainable machine learning
Latent energy flux
Evapotranspiration
Surface energy balance
Vegetation biophysics

ABSTRACT

Proximal remote sensing has the potential to provide critical information on vegetation biophysical factors that can predict land-atmosphere exchange of water and energy. Latent energy (LE) flux is traditionally estimated using process-based models which rely on vegetation parameters that change during the growing season. Data-driven models have the potential to address these issues by offering flexible predictor selection and more efficient utilization of the information in predictor sets. These models require careful choice of predictors to avoid redundancy and allow robust cross-validation. In this study we present a systematic and comprehensive evaluation of machine learning (ML) models to assess the capability of meteorological and proximal sensing data for predicting LE at a half-hourly temporal resolution across multiple growing seasons for an agricultural system. The results presented here demonstrate that a model using four environmental predictors in combination with two proximal sensing variables can capture 88 % of the variability in LE. ML models using only three predictors (one meteorological and two proximal remote sensing) captured 81 % of LE variability, offering the best trade-off between performance and complexity. An ML model utilizing only two predictors, one proximal remote sensing variable and downwelling radiation, captured 77 % of LE variability. These results demonstrate the power of proximal remote sensing and meteorological observations to estimate land-atmosphere water vapor exchange, providing a solution where more direct methods such as eddy covariance are not available and for evaluations of agronomic management and genotypic variations.

1. Introduction

Latent energy flux (LE) is a land surface process that plays a central role in the surface energy balance (Mallick et al., 2013; Anderson et al., 2011; Mallick et al., 2016). The physiological processes that control LE link the terrestrial water, carbon, and energy budgets (Drewry et al., 2010; Drewry et al., 2010; Mallick et al., 2013; Anderson et al., 2011; Mallick et al., 2016). LE is the energy flux equivalent of evapotranspiration (ET) and therefore defines crop water consumption in agricultural systems, making LE estimation essential for agricultural water management broadly. Example applications of LE estimation in managed systems include irrigation system design (Brombacher et al., 2022; Dela Cruz et al., 2020; Arif et al., 2022), crop water requirement monitoring (Li et al., 2010; Zhang et al., 2023), climate change impact assessments on agriculture (Castellví and Snyder, 2010; Denich and Bradford, 2010; Gaur et al., 2022; Payero and Irmak, 2008) and characterizing genotypic and agronomic management variations in crop growth (Bai et al., 2024).

LE can be measured using direct methods such as field lysimeters that use changes in above-ground mass to estimate water use (Castellví and Snyder, 2010; Denich and Bradford, 2010; Payero and Irmak, 2008) and eddy covariance (EC) that uses high frequency observations of water vapor concentration and wind fluctuations to estimate net canopy-atmosphere exchange of latent energy (Harder et al., 2023; Baldocchi, 2020). Remote sensing (RS) techniques are widely used to estimate LE at larger scales, and at locations where more direct but less spatially extensive field methods such as EC are not available (Kustas and Norman, 1996; Mallick et al., 2007; Anderson et al., 2011). RS-based models rely on an understanding of the surface energy balance and utilize observations of variables related to vegetation structure or temperature to constrain the energy balance (Huang et al., 2019; Zhang

* Corresponding author at: Department of Food, Agricultural, and Biological Engineering, Ohio State University, Columbus, OH, USA.
E-mail address: drewry.19@osu.edu (D.T. Drewry).

https://doi.org/10.1016/j.agwat.2025.109643
Received 10 November 2024; Received in revised form 8 June 2025; Accepted 27 June 2025
Available online 17 July 2025

et al., 2023). Unlike reflectance measurements of vegetation in the visible through shortwave infrared (i.e. solar spectrum) which take several days to show signs of stress, remote sensing of LE and associated physiological changes offers the potential for earlier detection of ecosystem stress (Wang et al., 2019).

RS models have often relied on either vegetation indices to integrate observations of canopy development and phenology (Liu et al., 2022) and productivity (Dold et al., 2019), or land surface temperature (LST) to provide information on the physiological status of the vegetation (Bhattarai et al., 2019; Anderson et al., 2012). Vegetation index-based models derive ET based on crop coefficients as a function of spectral reflectance values. In contrast, LST-based models utilize thermal infrared remote sensing observations to calculate LST to close the energy balance (Chen et al., 2023; Chen and Liu, 2020). In both cases, RS-based models often rely on models that are developed around biophysical principles, with varying levels of sophistication. These models have demonstrated the utility of both reflectance-based vegetation indices and land surface temperature to accurately estimate the land-atmosphere exchange of water vapor. However, these models rely on accurate specification of parameter values that often change with crop growth stage and environmental conditions, providing challenges to utilizing these formulations in highly dynamic agricultural systems. Proximal remote sensing offers a solution by providing high-frequency, real-time data directly from the field (e.g., continuous NDVI and LST measurements at sub-hourly intervals). This allows for much more dynamic and responsive modeling, as real-time measurements can capture rapid changes in crop physiology and environmental conditions that are missed by traditional satellite-based RS models (Yi et al., 2024). Another advantage of proximal remote sensing is the ability to estimate LE for a variety of treatments such as different cropping systems or crop genotypes within a narrow range of soil and weather conditions (Bai et al., 2024). This approach could be a promising way to guide the improvement of water-related crop traits, such as water use efficiency, when coupled with machine learning (Ferguson et al., 2021).

Machine learning (ML) models offer a powerful new approach for directly modeling LE or physiological variables that control surface water flux (Abdullah et al., 2015; Amani and Shafizadeh-Moghadam, 2023; Gaur and Drewry, 2024; Bai et al., 2021). The non-parametric nature of ML, flexibility in predictor selection, and robust predictive capabilities make it particularly compelling for predictions in earth sciences in general and hydrology and climate and agricultural sciences in particular (Cross and Drewry, 2024; Gaur and Drewry, 2024; Reichstein et al., 2019; Vidyarthi and Jain, 2023; Vidyarthi and Jain, 2020; Zhao et al., 2019; Bai et al., 2021). Numerous studies have employed a range of meteorological datasets for LE estimation. Granata (2019) applied multiple machine-learning algorithms with different combinations of meteorological predictors for LE prediction. Yamaç and Todorovic (2020) evaluated the performance of four scenarios involving different combinations of available meteorological variables for predicting crop evapotranspiration using ML algorithms. Their findings indicated that the model incorporating the full set of meteorological predictors outperformed the other ML models. Bai et al. (2021) emphasized the utility of multi-ensemble machine ML models in predicting crop evapotranspiration across a wide range of environmental conditions. Their study showed that ML-based ensemble models outperformed conventional ensemble models in ET predictions. These studies that utilized machine learning for LE estimation have relied primarily on weather data, demonstrating the power of these methods in combination with widely available in-situ observations, but leaving open questions related to the value of information in remote sensing observations from proximal, airborne or orbital systems. Recent advances in ML have largely focused on utilizing satellite-based RS observations and ambient environmental variables (Jung et al., 2019) to estimate LE, but these approaches lack information on the rapid changes and fine-scale variability inherent in dynamic terrestrial ecosystems, where crop growth and environmental conditions can rapidly change (Amani and Shafizadeh-Moghadam, 2023). In contrast, proximal RS can address these shortcomings by providing high-frequency observations for a more accurate representation of land surface conditions. While ML models have shown remarkable predictive capabilities and are increasingly applied in water and energy flux estimation (Lucarini et al., 2024), understanding how different predictor combinations influence predictions is often lacking, hindering parsimonious model selection. Interpretability tools, such as Shapley Additive Explanations (SHAP), provide techniques to uncover the contribution of each predictor variable, making it easier to understand how input variables interact to influence the target predictions. This is especially important for understanding model outcomes across different crop phenological stages and environmental conditions (Hu et al., 2022; Baptista et al., 2022).

In this study we present a systematic evaluation of the use of weather and proximal remote sensing observations for the estimation of LE in an agricultural system. Eddy covariance observations of LE provide validation data spanning two soybean growing seasons at a site outside of Ames, IA, USA. Proximal remotely sensed estimates of normalized difference vegetation index and LST allow us to examine how high frequency (sub-hourly) information on canopy structure and environmental and physiological controls on surface temperature can be leveraged for LE estimation using a flexible deep learning neural network (DLNN). The DLNN is a multi-layer neural network that extracts higher-level features from input datasets through representation learning (LeCun et al., 2015). This study provides guidelines for predictor combinations that provide near-optimal / parsimonious predictive performance, leveraging both widely available weather and remote sensing variables across a wide range of phenological and climatic variability. In addition, explainable machine learning is applied to understand under what conditions across a growing season specific predictors are required, mitigating many challenges associated with the interpretability and physical consistency of ML approaches for environmental prediction.

2. Methodology

In the sub-sections below we describe the SABR study site and array of flux and environmental / meteorological variables collected at the site that are utilized here. This is followed by a description of the two in-situ proximal remote sensing observations that, in combination with environmental variables, are evaluated as predictors in machine learning models of latent energy flux. Following this we describe our approach to systematically evaluate a wide range of ML models spanning one to six predictor variables, allowing us to identify the most synergistic and parsimonious sets of predictors, and particularly the value added by proximal remote sensing of biophysical variables related to vegetation phenology and physiological response to environment. We conclude this section with descriptions of our methods for model development and validation and the application of post-hoc ML interpretability methods.

2.1. SABR study site

The data used in this study was collected at the Iowa State University Sustainable Advanced Bioeconomy Research Farm (SABR) in Boone County, IA, near Ames (42.00°N, 93.70°W) (Bendorf et al., 2022). SABR features four independently monitored plots to assess the performance of soybean, corn, sorghum and miscanthus under identical climatological and soil conditions. Each plot contains an eddy covariance tower installed in the center of the plots to measure the exchange of mass and energy between each crop canopy and the atmosphere. The soil at the SABR site is primarily composed of Canisteo clay loam (59 %), with smaller proportions of Clarion loam (19.3 %) and Webster clay loam (18.1 %). A detailed description of the sensors and data protocols utilized at SABR can be found at: https://sabr.shinyapps.io/appSABR/.

Here we focus on data collected for the SABR soybean plots for the 2021 and 2022 growing seasons. Relative to 2022, the 2021 growing

season had higher temperatures and reduced precipitation, particularly during the critical crop development phase that began in June (Cross and Drewry, 2024). The data used in this study was collected from each of the two study years between June 2 to September 30, providing approximately 5500 half-hourly observations.

2.1.1. Flux and environmental data

The environmental predictors utilized in this study were collected on the sampling tower in the SABR soybean fields in 2021 and 2022, co-located with the eddy covariance sampling system used to measure latent energy flux. These environmental predictors included air temperature (T_a, °C) and relative humidity collected with a Campbell Scientific HMP-155. Vapor pressure deficit (VPD, kPa) was computed from these variables using the Tetens equation (Campbell and Norman, 2000). Photosynthetically active radiation (PAR, μmol/m^2/s) was collected with a LI-190 quantum sensor (Licor Biosciences, Nebraska, USA). Wind speed (U, m/s) was collected with a 3D sonic anemometer (Gill WindMaster, Hampshire, UK). A Kipp & Zonen CNR4 net radiometer provided downwelling and upwelling shortwave and longwave radiation, and sensors measured soil temperature and soil heat flux (Hukseflux HFP01SC soil heat flux plates) to provide necessary information for surface energy budget closure.

All environmental variables were averaged to 30-minute intervals. Missing meteorological data was first filled using data from a nearby meteorological station at the Ames Municipal Airport as part of the standard processing methods applied to SABR data (Cross and Drewry, 2024). Any remaining missing data points were filled using a linear interpolation from one of the neighboring SABR towers where the missing variables were available.

LE measurements were obtained from the eddy covariance instrumentation on the soybean tower and averaged to 30-minute resolution. No gap-filling was applied to the LE observations, ensuring that model development and validation conducted in this study was performed using observations and not modeled estimates. Bowen ratio correction (Dugas et al., 1991) was applied to the LE data before its use in the model development and evaluation exercises presented here to ensure that energy balance closure was maintained for each observation period.

2.1.2. Proximal remote sensing data

In-situ proximal remote sensing instruments were deployed on the soybean EC tower throughout the study period to continuously monitor LST and the normalized difference vegetation index (NDVI) (Tucker,

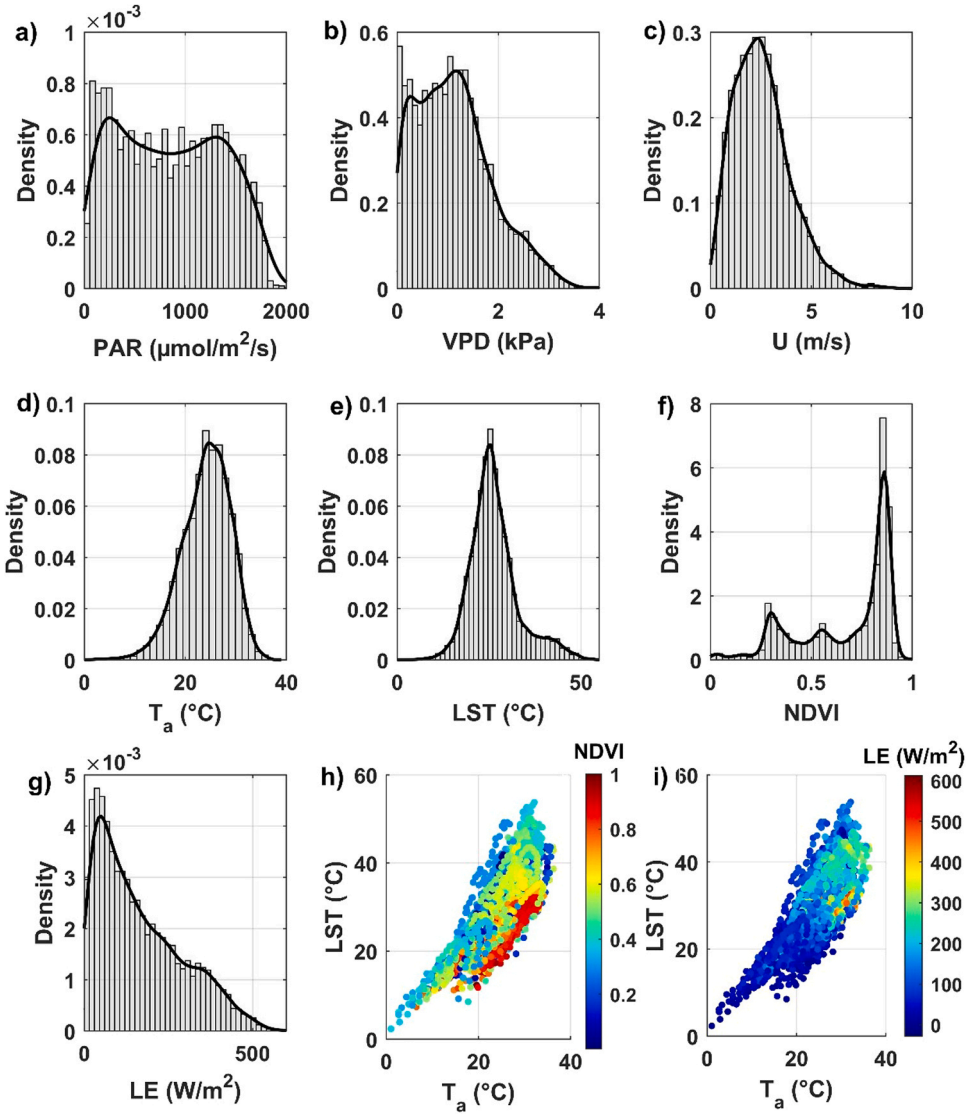


Fig. 1. Histograms and density plots of (a-d) environmental variables used as predictors, (e-f) proximal remote-sensing variables used as predictors, and (g) latent energy flux observations used for model development and validation. (h-i) LST plotted against T_a with data points colored according to the NDVI and LE values of each observation, respectively. All data was collected at the SABR soybean site during the 2021 and 2022 growing seasons.

1979). These instruments provided high temporal (5-minute) resolution observations of these remote sensing variables, which were averaged to 30-minute periods to coincide with the flux and environmental data. NDVI was collected using two-band radiometers (SRS sensors from METER Inc., WA, USA) affixed to each of the four towers. The sensors were positioned to view a homogeneous region of the canopy to the south of each tower. To reduce noise and variability, a 4-hour window around local solar noon was selected and averaged each day to produce a daily NDVI value. This was done as daily NDVI observations are sufficient to characterize variability in canopy development relevant to our modeling efforts here. LST measurements were collected using infrared thermometers (Apogee Instruments, Utah, USA) deployed on each tower, oriented to look down onto the land surface/canopy from an approximate height of 3–4 m above the surface.

Fig. 1 presents histograms and density plots for the environmental and proximal remote sensing predictors utilized in this study. PAR exhibits a fairly uniform distribution as observations were collected across each study day. VPD and U show a wide range of variability, with the majority of observations occurring during fairly low VPD and U conditions. A clear peak in the NDVI distribution (values greater than 0.7) shows that most observations throughout these two growing seasons were made when the soybean canopy was dense. Although T_a and LST both represent temperature observations, their distributions differ as LST is in part controlled by ambient temperature, but also by physiological controls and associated energy balance processes. Plotting LST as a function of T_a and coloring data points according to values of NDVI and LE (Figs. 1h and 1i, respectively) shows a high degree of correlation between the two variables, with a correlation coefficient of 0.75 across the two growing seasons. These temperature variables deviate from a strong linear correlation as T_a increases above 18 °C. As NDVI increases to its maximum values T_a increases as the growing season peaks, but LST demonstrates the impact of evaporative cooling to lower surface temperatures relative to maximum LST seen when the canopy is not closed (lower NDVI).

In general, all predictors (PAR, VPD, T_a, U, NDVI, LST) display a wide range of variability across these two growing seasons, providing a strong test of the ability of machine learning to accurately predict LE.

2.2. Machine learning model development for latent energy flux estimation

In this study we develop and apply a large set of deep learning neural networks to explore the ability of various predictor sets to accurately predict latent energy flux across growing seasons as crop phenology and ambient environment show dramatic variations. Fig. 2 presents a schematic of the methodology we apply here. Latent energy flux is often predicted using crop or biophysical models driven by weather data in combination with an understanding of the canopy state (Drewry et al., 2010; Drewry et al., 2010; Le et al., 2012; Anderson et al., 2011; Mallick et al., 2016). Here we utilized six predictor variables spanning both meteorological (PAR, VPD, T_a, U) and proximal-remote sensing (NDVI, LST) to predict LE using DLNNs (Fig. 2a). We retain both T_a and LST as predictors in the models we develop here to evaluate the unique information each may have toward LE prediction, and the extent to which their combination provides value (i.e. through the difference in air and surface temperature) in LE prediction.

2.2.1. Categorization of ML models for LE predictions

Here we utilize meteorological and proximal-remote sensing predictors for LE predictions as discussed in Section 2.2. We organize this large set of ML models into four categories (Fig. 2b) to allow us to examine the impact of proximal remote sensing information on model performance. Within each of these four categories, we further consider the models in terms of their complexity, defined here simply as the number of predictors used by each model. This provides a framework for understanding model parsimony, i.e. how predictive performance varies as a function of model complexity, and within each complexity level (i.e. for all models developed using the same number of predictor variables). This allows us to determine which predictor sets perform best and how proximal sensing information improves predictive performance for LE. Additionally, this approach can provide insights into the biophysical controls on LE and enhance understanding of the functioning of the cropping system. Overall, we develop and evaluate 64 ML models developed using unique sets of predictor variables to predict LE:

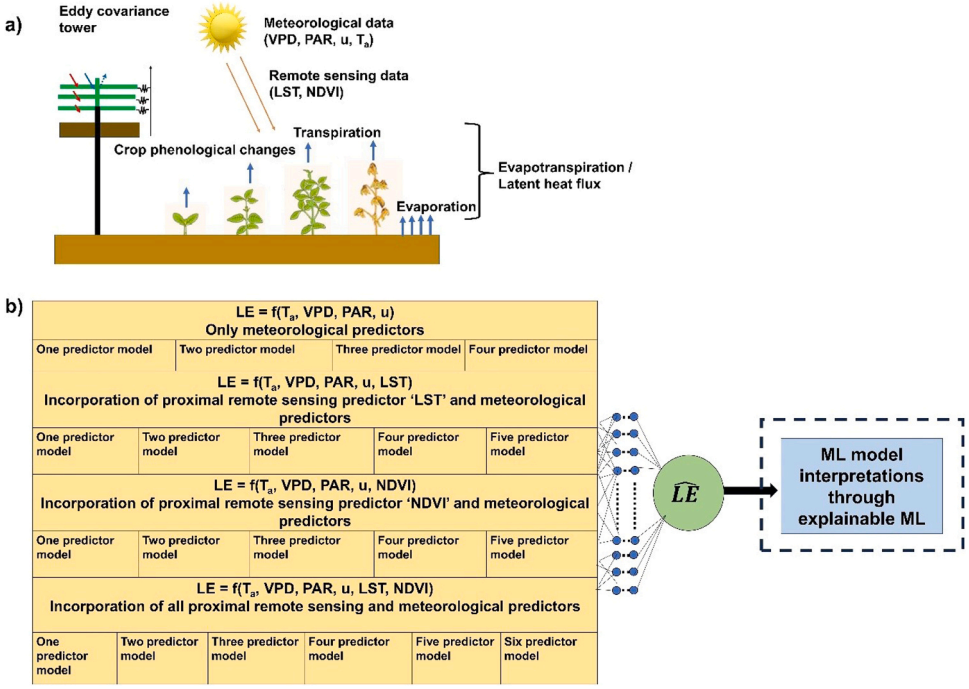


Fig. 2. Schematic of the methodology used in this study, including (a) a diagram depicting the field observations used in this study, and (b) the variable combinations used to formulate the four primary model categories evaluated here for the prediction of LE ($\widehat{LE}$).

- **Category 1:** ML models formulated using only meteorological predictors using all possible combinations of meteorological predictors: PAR, VPD, T_a, and U. This includes the formulation of fifteen models in total: one-predictor models ([PAR], [VPD], [T_a], [U]); two-predictor models ([PAR, VPD], [PAR, T_a], [PAR, U], [VPD, T_a], [VPD, U], [T_a, U]); three-predictor models ([PAR, VPD, T_a], [PAR, VPD, U], [VPD, T_a, U], [PAR, T_a, U]); and a four-predictor model ([PAR, VPD, T_a, U])
- **Category 2:** ML models formulated using meteorological predictors in combination with proximal sensing LST. In this category a total of sixteen models are formulated that include: one-predictor model ([LST]); two-predictor models ([T_a, LST], [VPD, LST], [U, LST], [PAR, LST]); three-predictor models ([PAR, T_a, LST], [VPD, T_a, LST], [T_a, U, LST], [PAR, VPD, LST], [VPD, U, LST], [PAR, U, LST]); four-predictor models ([PAR, VPD, T_a, LST], [VPD, T_a, U, LST], [PAR, T_a, U, LST], [PAR, VPD, U, LST]); and a five-predictor model ([PAR, VPD, U, T_a, LST]).
- **Category 3:** ML models formulated using meteorological predictors in combination with proximal sensing NDVI. The developed models include sixteen models in total: one-predictor model **([NDVI]);** two-predictor models **([PAR, NDVI], [VPD, NDVI], [T_a, NDVI], [U, NDVI]);** three-predictor models **([PAR, VPD, NDVI], [PAR, T_a, NDVI], [T_a, U, NDVI], [VPD, T_a, NDVI], [PAR, U, NDVI], [VPD, U, NDVI]);** four-predictor models **([PAR, VPD, T_a, NDVI], [VPD, T_a, U, NDVI], [PAR, T_a, U, NDVI], [PAR, VPD, U, NDVI]);** and **a** five-predictor model **([PAR, VPD, U, T_a, NDVI]).**
- **Category 4:** ML models formulated using meteorological predictors in combination with both proximal sensing variables. These predictor combinations resulted in the formulation of seventeen models: a two-predictor model ([LST, NDVI]); three-predictor models ([U, LST, NDVI], [PAR, LST, NDVI], [VPD, LST, NDVI], [T_a, LST, NDVI]); four-predictor models ([T_a, U, LST, NDVI], [PAR, U, LST, NDVI], [VPD, U, LST, NDVI], [PAR, VPD, LST, NDVI], [PAR, T_a, LST, NDVI], [PAR, U, LST, NDVI], [VPD, T_a, LST, NDVI]); five-predictor models ([VPD, T_a, U, LST, NDVI], [PAR, T_a, U, LST, NDVI], [PAR, VPD, U, LST, NDVI], [PAR, VPD, T_a, LST, NDVI]); and a six-predictor model ([PAR, VPD, U, T_a, LST, NDVI]).

2.2.2. Machine learning model training and evaluation

All ML models were robustly cross validated following the procedures of Cross and Drewry (2024) and Gaur and Drewry (2024). The cross-validation scheme used in this study involves iterative training and validation across 100 random splits of the data. Model training (calibration) is conducted by identifying the optimal set of hyperparameters based on these 100 random splits, each of which is used to train a separate model. Thus, this robust cross-validation framework effectively incorporates both the calibration (training) and validation of the machine learning models. A detailed description of the cross validation procedure we used follows.

The data was randomly split into training (80 %) and validation (20 %) sets using an iterative procedure in which the dataset was randomly split and unique models developed for each of 100 data splits. All ML model development presented here was performed using the Python programming language. All predictors were normalized to have a zero mean and one standard deviation prior to model development. Hyperparameter tuning (HPT) using randomized search was performed on each randomly split dataset, resulting in 100 sets of best-performing hyperparameters. Here HPT was performed on three hyperparameters: (a) the number of hidden layers which was varied from 4 to 12; (b) the number of neurons which was allowed to take on one of the following values: 16, 32, 64, 128, 256, 518; and (c) the learning rate which was allowed to be one of the following values: 0.001, 0.0001, 0.0002, 0.0003, 0.0004, 0.0005. A rectified linear unit (ReLU) was used for the activation function of the neural networks developed here.

Early stopping was performed to prevent model overfitting (Zhao et al., 2019). Early stopping evaluates the performance of the ML model on the validation set and terminates the training process when validation loss improvement plateaus (Vilares Ferro et al., 2023; Zhao et al., 2019). Here, the early stopping criteria was set to 50 epochs. Each ML model was individually trained on each random split of data and the final performance of the model was obtained by averaging the performance metrics for each data point on validation sets. The post model performance was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE) and Akaike information criterion (AIC).

2.3. Posthoc interpretability of ML predictions

While ML techniques are highly flexible modeling tools, they lack the interpretability of process-based models (Reichstein et al., 2022; Reichstein et al., 2019). Explainable machine learning is a set of techniques that provides insights into the value of predictor variables to be derived, illuminating aspects of model structure in what are often considered to be 'black-box' modeling approaches (Lundberg et al., 2020; Lundberg and Lee, 2017; Zhi et al., 2024). Here we utilized the SHapley Additive exPlanations algorithm (Lundberg and Lee, 2017) to interpret the outcomes of DLNNs used for predicting LE. We used SHAP to understand how different input predictors affect the performance of the models developed here. SHAP values provide information about how an individual data point in a predictor contributes to the model predictions in terms of local explanation (Lundberg et al., 2020). This can provide insights into the synergisms and redundancies inherent between predictors and can help to physically interpret model performance when evaluated against environmental or phenological conditions (Cross and Drewry, 2024). We use beeswarm plots to visualize instance-based local explanations in terms of magnitude, prevalence, and direction of the predictor's effect on LE estimates. These plots show the distribution of individual continuous variables (predictors in our case), with each data point plotted separately. They offer deeper insights, especially when comparing or visualizing multiple variables. Further details about the SHAP algorithm can be found in Lundberg and Lee (2017) and Lundberg et al. (2020).

3. Results and discussion

3.1. Latent energy flux prediction with model complexity

The average performance of the ML models on the validation fraction across one hundred model iterations are presented in Fig. 3 as a function of model complexity (number of predictor variables used). The best sets of hyperparameters obtained from hyperparameter tuning over one hundred iterations fall within the following ranges: number of layers (3–10), number of neurons per layer (32–518), and learning rate (0.0001–0.001). Figures S1-S4 illustrate the distribution of these parameters across all iterations for each model using box plots. The one-predictor models vary across a wide range of predictive performance with R^2 values ranging from 0.05 for U to 0.56 for the PAR model (Tables 1–3). PAR is critical for LE predictions as it is a shortwave radiation variable associated with the primary source of energy absorbed by leaves during the day to drive photosynthesis and energy exchange.

The performance of the two-predictor models ranges from $R^2 = 0.26$ to $R^2 = 0.77$ (Tables 1–4). Among the models driven solely by meteorological predictors [PAR, VPD] and [PAR, T_a] capture 61 % and 60 % of the variability in LE predictions (Table 1), respectively. Of the two-predictor models, the best performing combination includes downwelling PAR and NDVI [PAR, NDVI] (Table 3) which captures 77 % of the variability in LE over the two growing seasons examined here. This finding is consistent with those of Wang et al. (2021), Wu et al. (2020) and Yebra et al. (2015) who found that NDVI adds critical information on seasonal phenological changes in vegetation that substantially improves LE predictions. The two-predictor model that combines downwelling PAR with proximal LST also performs quite well ($R^2 = 0.68$)

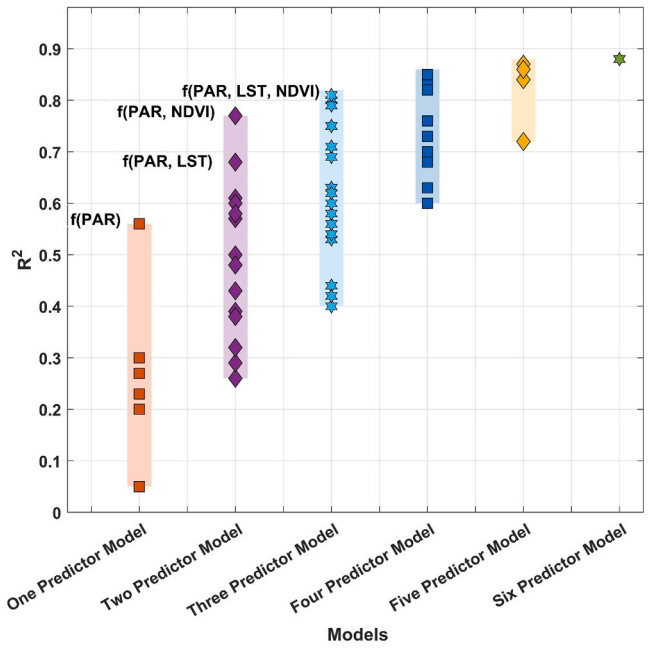


Fig. 3. Performance of ML models during validation. The shaded regions represent the range of R^2 performance for all models with a specific number of predictors. Symbols within those shaded regions show the R^2 performance of specific models with that number of predictors. Tables 1–4 provide statistical performance metrics for all models developed in this study.

Table 1
Performance of ML models formulated using only meteorological predictors for LE estimation.

Predictors	R^2	MAE (W/m^2)	RMSE (W/m^2)	AIC
[PAR]	0.56	60	80.7	8.7
[VPD]	0.23	84.6	106.35	9.3
[T_a]	0.27	81.3	103.5	9.2
[U]	0.05	98.3	118.4	9.5
[PAR, VPD]	0.61	54.6	75.3	8.6
[PAR, T_a]	0.60	55.1	76.2	8.6
[PAR, U]	0.57	58.4	79.8	8.7
[VPD, T_a]	0.39	73.4	95	9.1
[VPD, U]	0.26	82.9	104.5	9.2
[T_a, U]	0.29	80.3	102.3	9.2
[PAR, VPD, T_a]	0.75	42.2	60.2	8.2
[PAR, VPD, U]	0.63	52.8	73.8	8.6
[VPD, T_a, U]	0.42	70.9	91.9	9.0
[PAR, T_a, U]	0.62	53.8	75.05	8.6
[PAR, VPD, T_a, U]	0.76	41.2	59.6	8.1

(Table 2). This is likely the case as LST contains information about canopy phenology, as the land surface tends to cool as it transitions from bare soil to fully vegetated due to the effect of evaporative cooling from evapotranspiration (Feldman et al., 2023) It is notable that multiple two-predictor models are capable of capturing greater than 60 % of the variability in LE across multiple growing seasons.

The predictive performance of the three-predictor models ranges from $R^2 = 0.40$ to $R^2 = 0.81$ (Tables 1–4). Among these, the models incorporating a combination of environmental and remote sensing predictors provide the best performance compared to those driven solely by environmental predictors. Downwelling PAR in addition to the two remote sensing variables ([PAR, LST, NDVI]) produces a model that captures 81 % of the observed LE variability ($R^2 = 0.81$, Table 4), as does [PAR, T_a, LST] ($R^2 = 0.81$, Table 2). While the performance of these three-predictor models is notable, it is only slightly better than the best-performing two-predictor model, demonstrating the importance of knowledge of energy input into the system and vegetation structure (i.e. phenology).

Table 2
Performance of ML models formulated using meteorological predictors and proximal sensing LST for LE estimation.

Predictors	R^2	MAE (W/m^2)	RMSE (W/m^2)	AIC
[LST]	0.30	78.8	101.2	9.2
[T_a, LST]	0.43	69.7	91.3	9.0
[VPD, LST]	0.38	73.2	95.7	9.1
[U, LST]	0.32	77.2	99.4	92
[PAR, LST]	0.68	49	68.1	8.4
[PAR, T_a, LST]	0.81	37.7	52.2	7.9
[VPD, T_a, LST]	0.58	58.1	79.2	8.7
[T_a, U, LST]	0.44	68.7	90.3	9.0
[PAR, VPD, LST]	0.69	47.9	69.4	8.4
[VPD, U, LST]	0.40	71.8	94.0	9.0
[PAR, U, LST]	0.71	47.7	65.7	8.3
[PAR, VPD, T_a, LST]	0.85	33.4	46.9	7.6
[VPD, T_a, U, LST]	0.60	57.4	77.5	8.7
[PAR, T_a, U, LST]	0.82	36.4	50.8	7.8
[PAR, VPD, U, LST]	0.73	45.2	63.9	8.3
[PAR, VPD, U, T_a, LST]	0.86	30.7	45.0	7.6

Table 3
Performance of ML models formulated using meteorological predictors and proximal sensing NDVI for LE estimation.

Predictors	R^2	MAE (W/m^2)	RMSE (W/m^2)	AIC
[NDVI]	0.20	87.2	109.1	9.3
[PAR, NDVI]	0.77	40.2	57.4	8.1
[VPD, NDVI]	0.50	64.4	85.9	8.9
[T_a, NDVI]	0.48	65.8	87.1	8.9
[U, NDVI]	0.32	79.6	99.8	9.2
[PAR, VPD, NDVI]	0.80	37.2	54.2	7.9
[PAR, T_a, NDVI]	0.81	36.3	52.6	7.9
[T_a, U, NDVI]	0.53	63.1	82.8	8.8
[VPD, T_a, NDVI]	0.54	61.4	82.2	8.8
[PAR, U, NDVI]	0.79	37.4	55	8.0
[VPD, U, NDVI]	0.56	60.1	80.1	8.7
[PAR, VPD, T_a, NDVI]	0.69	50.6	68.8	8.4
[VPD, T_a, U, NDVI]	0.84	34	48.1	7.7
[PAR, T_a, U, NDVI]	0.82	34.4	50.3	7.8
[PAR, VPD, U, NDVI]	0.82	34.6	82.2	7.8
[PAR, VPD, U, T_a, NDVI]	0.86	31.2	45.2	7.6

Table 4
Performance of ML models formulated using meteorological predictors and both proximal sensing variables (LST, NDVI) for LE estimation.

Predictors	R^2	MAE (W/m^2)	RMSE (W/m^2)	AIC
[LST, NDVI]	0.58	57.9	78.2	8.7
[U, LST, NDVI]	0.62	55.2	74.5	8.6
[PAR, LST, NDVI]	0.81	37.4	52.9	7.9
[VPD, LST, NDVI]	0.63	54.1	73.2	8.6
[T_a, LST, NDVI]	0.60	56.2	75.9	8.6
[T_a, U, LST, NDVI]	0.63	53.2	72.9	8.5
[PAR, U, LST, NDVI]	0.68	50.7	68.1	8.4
[VPD, U, LST, NDVI]	0.68	50.0	68.6	8.4
[PAR, VPD, LST, NDVI]	0.82	35.9	51.0	7.8
[PAR, T_a, LST, NDVI]	0.85	33.5	47.6	7.7
[PAR, U, LST, NDVI]	0.85	33.4	86.7	7.8
[VPD, T_a, LST, NDVI]	0.70	48.5	66.4	8.3
[VPD, T_a, U, LST, NDVI]	0.72	52.6	67	8.2
[PAR, T_a, U, LST, NDVI]	0.86	31.6	45.5	7.6
[PAR, VPD, U, LST, NDVI]	0.84	33.1	48.1	7.7
[PAR, VPD, T_a, LST, NDVI]	0.87	31.1	44	7.5
[PAR, VPD, U, T_a, LST, NDVI]	0.88	29.1	42	7.4

The performance of the four-predictor models ranges from $R^2 = 0.60$ to $R^2 = 0.85$ (Tables 1–4). Generally, the models incorporating both meteorological and remote sensing predictors demonstrate better predictive performance than those that only use meteorological predictors. The five-predictor models have R^2 validation performance values

ranging from 0.71 to 0.87 (Tables 2–4). The model incorporating all six predictor variables achieved the strongest performance with $R^2 = 0.88$ (Table 4).

These results demonstrate the power of even a two-predictor model of LE when the predictor variables are carefully selected. Predictive performance increases with model complexity in general, but careful selection of predictor sets can provide near-optimal performance using only three or four predictors, demonstrating the value of an exercise such as this to carefully evaluate the performance of various predictor combinations (Cross and Drewry, 2024; Gaur and Drewry, 2024).

3.2. Value of remote sensing for latent energy flux prediction

Model performance for a set of models that span three levels of complexity and increasingly incorporate remote sensing observations is presented in Fig. 4. These models all include downwelling photosynthetically active radiation and progressively include the two proximal remote sensing variables individually and together: [PAR], [PAR, LST], [PAR, NDVI] and [PAR, LST, NDVI]. The data points in Fig. 4 are colored according to the day of the year (DOY) when they were collected to provide insights into the influence of the time within the growing season on model performance. As discussed in Section 3.1, incorporating remote sensing predictors such as LST and NDVI adds information about crop phenological responses. This is illustrated in Fig. 4, where the best performing one predictor model [PAR] does not account for crop phenology. Consequently, it underestimates LE during the mid-growing season (DOY 200 to DOY 240) as depicted in Fig. 4a. Early and late-season predictions improve when remote sensing predictors are added (Figs. 4b-4d). Adding both remote sensing predictors (LST and NDVI) to the meteorological predictor (PAR) substantially enhances LE prediction. The [PAR] model has a clear maximum LE prediction limit under 350 W/m^2. This constrained prediction level is also seen in the [PAR, LST] model, but with the limit increased to slightly over 400 W/m^2, likely due to the added degree of freedom provided to the model. These model prediction plateaus are not apparent when NDVI, the predictor most closely associated with crop phenology, is included.

This suggests that the inclusion of proximal remote sensing variables like NDVI and LST improves LE predictions across a range of canopy states by including information on crop phenology, which standalone meteorological variables do not contain. This reduces errors, particularly during key growth stages and periods of high variability in conditions such as elevated VPD and temperature. Consequently, proximal remote sensing significantly enhances the model's ability to capture dynamic changes in agricultural ecosystems (Amani and Shafizadeh-Moghadam, 2023). Fig. 5 presents the variation in half-hourly absolute LE prediction error (absolute difference between measured LE and predicted LE by ML models) with NDVI (Fig. 5a), LST (Fig. 5b) and VPD (Fig. 5c). These prediction error values are averaged over bins across the full observation ranges of NDVI, LST and VPD.

The highest error values occur between NDVI values of 0.2–0.5 for the [PAR] and [PAR, LST] models (Fig. 5a). The distributions in Fig. 1f show that this range of NDVI is less well represented in the complete dataset, with most of the data corresponding to a closed, dense soybean canopy with NDVI values greater than 0.7. ML models developed using only meteorological variables are not able to infer phenological changes and the associated impacts on LE and perform best if they assume canopy closure which is representative of most of the observations. When RS observations of canopy density are incorporated into ML models

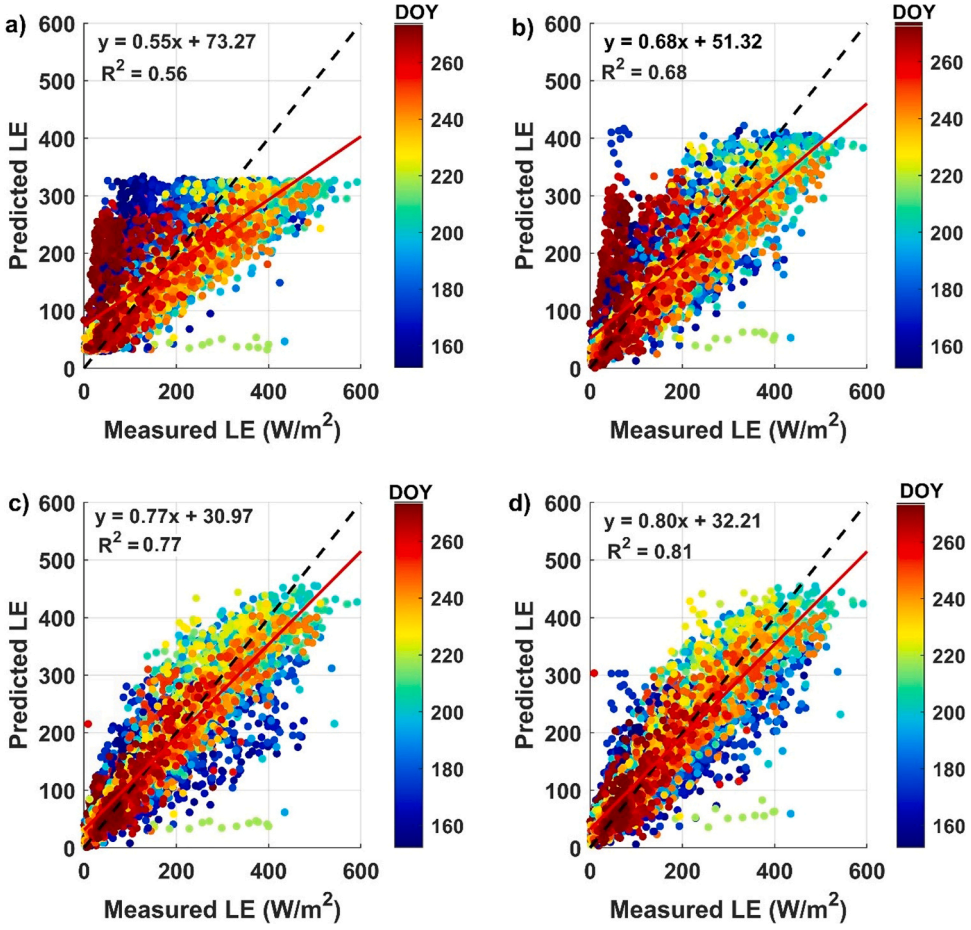


Fig. 4. Predicted versus observed LE for half-hourly observations over two growing seasons for four different ML models: (a) [PAR]; (b) [PAR, LST]; (c) [PAR, NDVI]; (d) [PAR, LST, NDVI]. Data points are colored according to the day of year that they were collected to provide information on seasonality in model performance.

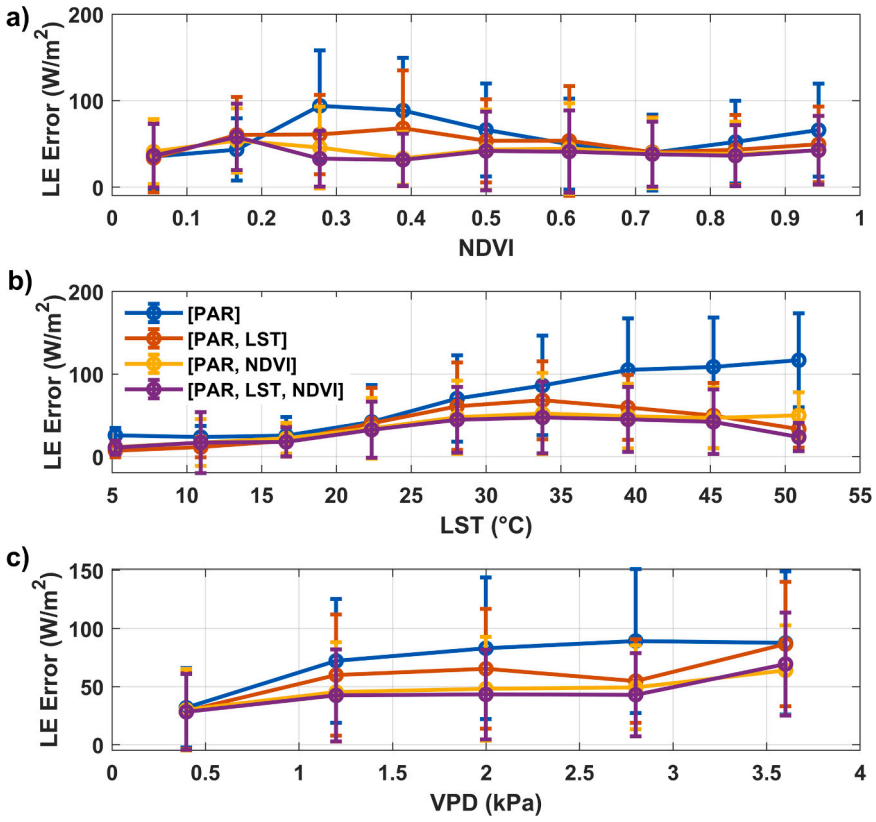


Fig. 5. Variation of half-hourly averaged absolute LE prediction errors for [PAR], [PAR, LST], [PAR, NDVI], and [PAR, LST, NDVI] as a function of: (a) NDVI; (b) LST; (c) VPD.

([PAR, NDVI], [PAR, LST, NDVI]) these errors are corrected as the ML models now have the information they need to estimate LE across the full range of canopy structural states spanning growing seasons.

In Fig. 5b the inclusion of either LST or NDVI with PAR allows the ML models to accurately model LE, relative to the PAR-only model that has minimal information on seasonality. In Fig. 5c it is clear that models that include NDVI show significant improvement as VPD increases, as these environmental conditions coincide with the phenological conditions captured by NDVI data.

Fig. 6 presents the variation of LE with VPD and T_a. The relationship between LE and VPD is inherently nonlinear, with a general increase in LE as VPD increases up to approximately 2 kPa. Increases in VPD result in greater surface water loss driven by increased atmospheric demand. These increases continue until a VPD level is reached that triggers stomatal closure to conserve water as atmospheric demand further increases, resulting in a decrease in LE (Wang et al., 2019). This effect is seen here as VPD increases beyond 2 kPa (Fig. 6a). Models may struggle to accurately capture this nonlinear behavior, leading to greater discrepancies between predicted and observed LE flux (Eamus et al., 2013) (Fig. 6a). At higher temperatures the interplay between PAR and LE

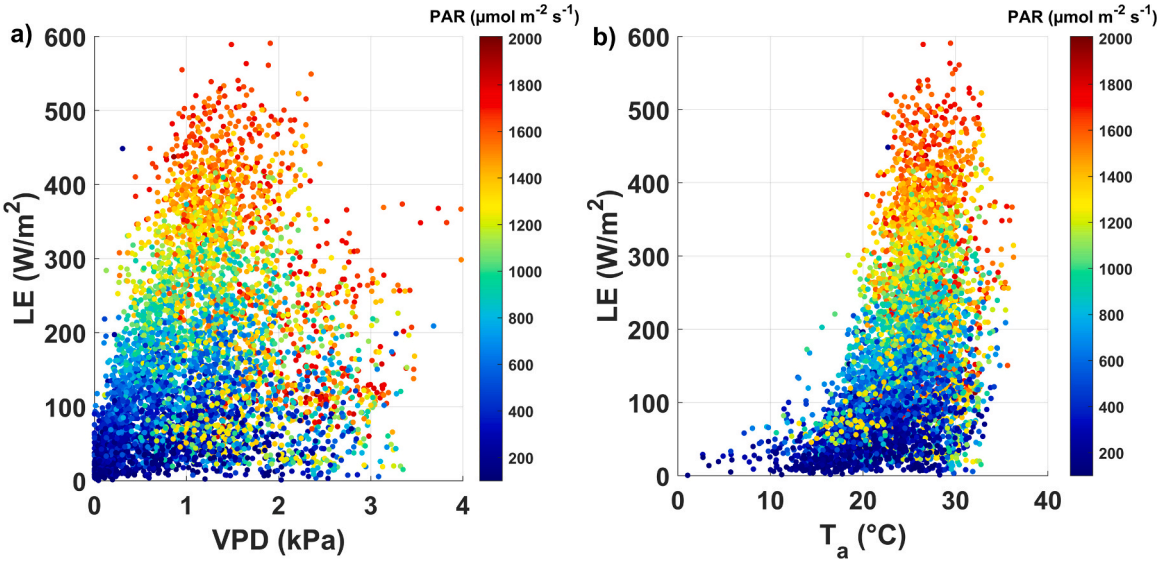


Fig. 6. (a) Variation of LE with VPD and (b) variation of LE with T_a. The data points are colored according to the observed PAR values at each time point.

becomes more complex due to the increased influence of temperature on stomatal behavior (Fig. 6b). These influences of temperature, through both leaf surface temperature and VPD, can influence stomatal conductance and LE, complicating the relationship with PAR and potentially contributing to increased errors in LE estimation under high-temperature conditions (Oren et al., 1999).

3.3. Model interpretation

The post hoc interpretability of outcomes from the LE models was performed using SHAP. Fig. 7 presents SHAP violin summary plots for four models that increasingly utilize proximal remote sensing observations: [PAR], [PAR, LST], [PAR, NDVI], and [PAR, LST, NDVI]. The violin summary plots show the distribution and attribution of each predictor present in each model (Lundberg et al., 2020; Lundberg and Lee, 2017; Mehdiyev et al., 2024). Each violin plot shows predictor variables ordered by importance, with the most important predictor at the top. Wider bands indicate higher data density. The color variation from blue to red in Fig. 7 represents an increase in the magnitude of the predictor; for instance, as the magnitude of PAR increases the SHAP values also increase demonstrating the positive influence of PAR on LE prediction estimates, consistent across all four models (Fig. 7a-d). Likewise, as NDVI increases it has larger positive impacts on model predictions (Fig. 7c,d). Higher NDVI values indicate greater biomass and leaf area coverage by the canopy, increasing transpiration and therefore LE (Cihlar et al., 1991). These summary plots also present the relative importance of predictors; for example, the longer tail of PAR indicates that higher PAR values are more important for predicting LE than higher magnitudes of NDVI.

The SHAP summary plots further explain the predictive performance of these four models as shown in Fig. 4. The SHAP values for the machine learning models [PAR, LST] reach larger magnitudes compared to those of the [PAR] model, indicating that the inclusion of LST provides the model a greater ability to attribute higher PAR values to larger LE fluxes, improving some of the overestimation in LE made by the [PAR] model for the data points early and late in the growing season when the canopy was not yet mature or partially senesced. These errors are greatly reduced when NDVI is included in either model (Fig. 4c,d), as NDVI is able to modulate how PAR is utilized in LE predictions as a function of canopy maturity.

To better understand the synergistic and redundant effects of predictor interactions we produced interaction plots to understand the impacts of combining proximal sensing predictors within a model (Lundberg et al., 2020). Fig. 8 presents SHAP interaction plots for models [PAR], [PAR, LST], [PAR, NDVI], and [PAR, LST, NDVI]. In Fig. 8, SHAP attributions to PAR ($SHAP_{PAR}$) are plotted against PAR values for each data point of the two seasons analyzed here. Each subplot shows how these attributions change as model composition (predictor set) is modified. In addition, coloring data points by their respective NDVI and LST values provides insights into how canopy phenological state and canopy temperature / physiological controls impact these attributions.

Fig. 8a shows that when only PAR is considered in the modeling process, $SHAP_{PAR}$ increases approximately linearly with increasing PAR across a wide range of PAR values. At the highest PAR values the attributions show non-linear shifts due to the inability of PAR alone to capture surface energy partitioning and water use across the wide range of phenological variability evaluated here (see Fig. 1f). This demonstrates the limitation of PAR as a single predictor used to estimate LE, despite the importance of solar energy input to surface energy exchange.

Figs. 8b and 8c show the same $SHAP_{PAR}$ vs PAR dynamic, but for models that add one of the proximal sensing predictors, LST (Fig. 8b) and NDVI (Fig. 8c). In each case the data points are colored according to the value of the proximal remote sensing variable used in the model. Both of these models show an increase in $SHAP_{PAR}$ with PAR as would be expected with any ML model containing PAR as a predictor. The [PAR, NDVI] model (Fig. 8c) shows a wider range of variability in SHAP values for PAR, with that variability under both low and high PAR values clearly driven by changes in surface vegetation cover (coloration). The widest range of $SHAP_{PAR}$ variability is seen at the largest PAR values, with these SHAP attributions increasing with increasing NDVI. In general the slope of the $SHAP_{PAR}$ vs PAR relationship increases as NDVI increases.

For the [PAR, LST] model (Fig. 8b) little variability is seen in PAR values less than 1000 [$\mu mol/m^2/s$] which most commonly correspond to either early or late day time periods, or cloudy conditions. Under conditions associated with the peak of the day when the canopy is strongly illuminated by solar radiation the addition of LST to the model allows for two distinct dynamics, one in which very high LST conditions cause a reduction in $SHAP_{PAR}$, with surface temperatures less than approximately 35 °C resulting in the highest $SHAP_{PAR}$ attributions. This strong deviation in SHAP attributions to PAR as LST approaches the highest values seen in the dataset implies that physiological control imposed during conditions of canopy water stress result in energy partitioning by the canopy (increasing Bowen ratio) when energy inputs are highest.

Adding both proximal sensing variables to the predictor set with PAR results in a model that performs better than any two-predictor model, capturing 81 % of the variability in LE, relative to 78 % for [PAR, NDVI] and 68 % for [PAR, LST] (see Figs. 3,4 and Tables 2–4). The [PAR, NDVI, LST] model exhibits large variability in $SHAP_{PAR}$ at both low and high values of PAR, with clear broad patterns imposed by the phenological information contained in NDVI, and regions of sharp reduction in SHAP attribution in PAR for LE prediction when the land surface is extremely warm (i.e. $> 35°C$).

In order to better understand the role of remotely sensed LST on model structure and predictions, Fig. 9 provides an analogous figure to

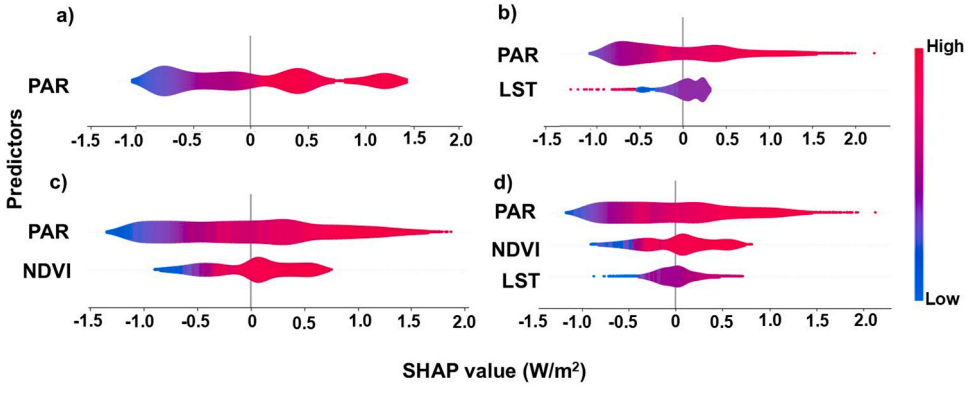


Fig. 7. SHAP violin summary plots for ML models: (a) [PAR]; (b) [PAR, LST]; (c) [PAR, NDVI]; and (d) [PAR, LST, NDVI]. Coloration from blue to red indicates increasing values of the predictor variables.

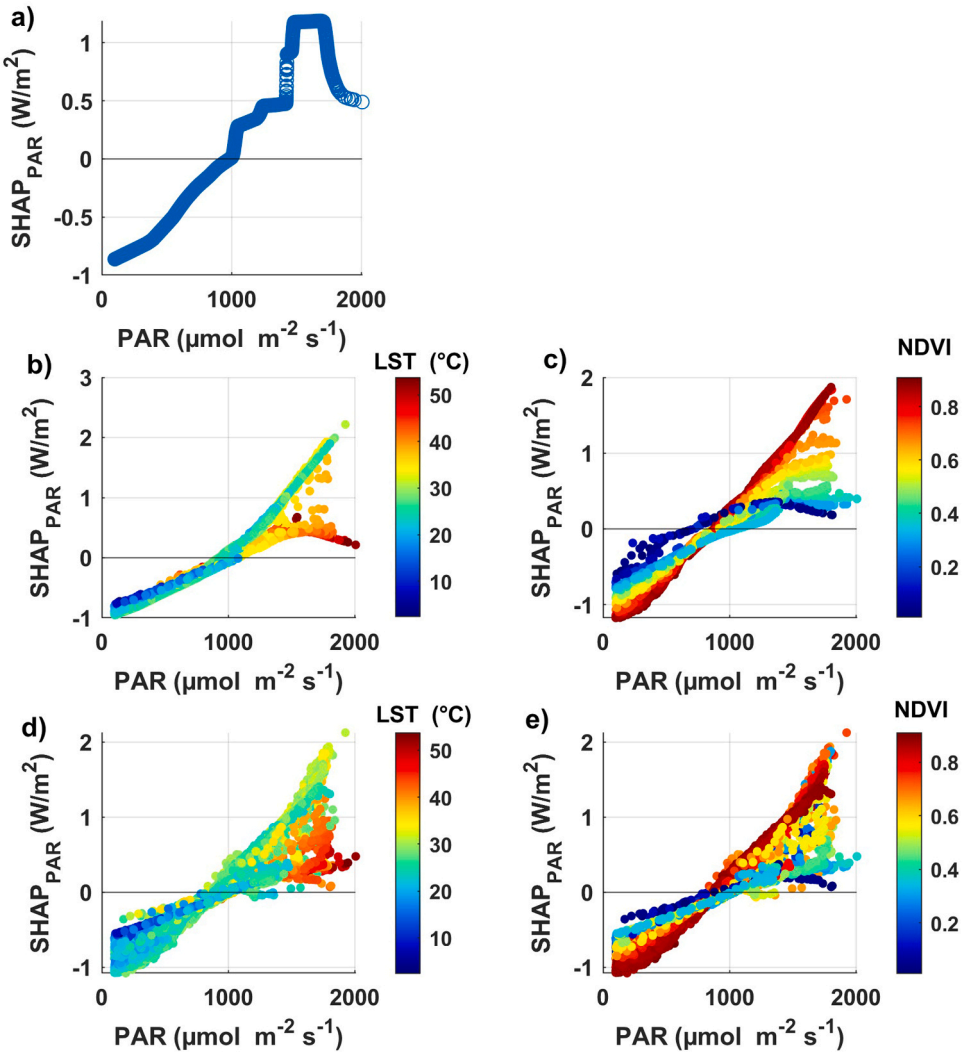


Fig. 8. SHAP values for PAR ($SHAP_{PAR}$) plotted against the associated PAR values for: (a) the ML model developed using only PAR ([PAR]); (b) the [PAR, LST] model with data values colored according to their respective LST values; (c) the [PAR, NDVI] model with data values colored according to their respective NDVI values; (d) the [PAR, LST, NDVI] model with data values colored according to their respective LST values; (e) the [PAR, LST, NDVI] model with data values colored according to their respective NDVI values.

that of Fig. 8, but examining the SHAP attributions to LST ($SHAP_{LST}$) as a function of LST. Fig. 9a shows an LST-only model ([LST]), with the [PAR, LST] model presented with data points colored by NDVI value in Fig. 9b and colored by VPD value in Fig. 9c. The three-predictor model [PAR, LST, NDVI] attributions are presented with data colored by NDVI value in Fig. 9d and colored by VPD value in Fig. 9e.

Across all three levels of model complexity positive model attributions to LST are found within an LST range that spans 20–40 °C, when positive temperatures are strongly associated with enhanced canopy transpiration and photosynthesis.

$SHAP_{LST}$ becomes negative at lower temperatures for [PAR, LST] (Fig. 9b,c), between 30 and 40 °C, relative to the [PAR, LST, NDVI] model (Fig. 9d,e) which maintains positive LST attributions until 40 °C. Providing the model with the information on canopy development and density contained in NDVI allows the model to utilize LST more effectively for the range of contrasting conditions spanning bare soil to dense closed canopy. Adding NDVI as a predictor also results in larger positive attributions to LST, particularly when NDVI is indicative of a closed canopy (i.e. > 0.6) and LST is indicative of the canopy receiving greater energy to drive LE. The model that includes NDVI produces negative LST attributions at LST values approximately greater than 40 °C, where VPD values reach levels likely to cause stomatal closure and reductions in LE (Fig. 9e). Adding NDVI as a predictor allows the model freedom to use a wider range of attribution magnitudes to LST dependent on the status of the land surface, from bare soil to densely vegetated.

Intermediate NDVI values in the 3-predictor model have primarily positive $SHAP_{LST}$ values when NDVI is included (Fig. 9d), in contrast to the [PAR, LST] model in which these canopy states, which may allow a partial view of the soil and therefore have higher LST, are confused with closed-canopy stress states. Likewise, without having knowledge of canopy phenology the [PAR, LST] model produces negative $SHAP_{LST}$ for intermediate to high VPD values greater than 2.5 kPa. When NDVI is included in the model many of these higher VPD conditions result in positive attributions as they allow for higher LE when the canopy is not stressed. Figure S5 presents SHAP attributions to NDVI ($SHAP_{NDVI}$) plotted against NDVI values for each data point for [NDVI], [PAR, NDVI] and [PAR, LST, NDVI] models.

In summary, the SHAP-based interpretability analysis presented here highlights the value of continuous proximal remote sensing variables like NDVI and LST in improving ML-based LE predictions. While PAR remains the single-most important predictor, NDVI and LST play crucial roles allowing information on downwelling radiation to be utilized appropriately throughout the vegetation growth cycle and during conditions when plant physiology constrains gas exchange. This analysis

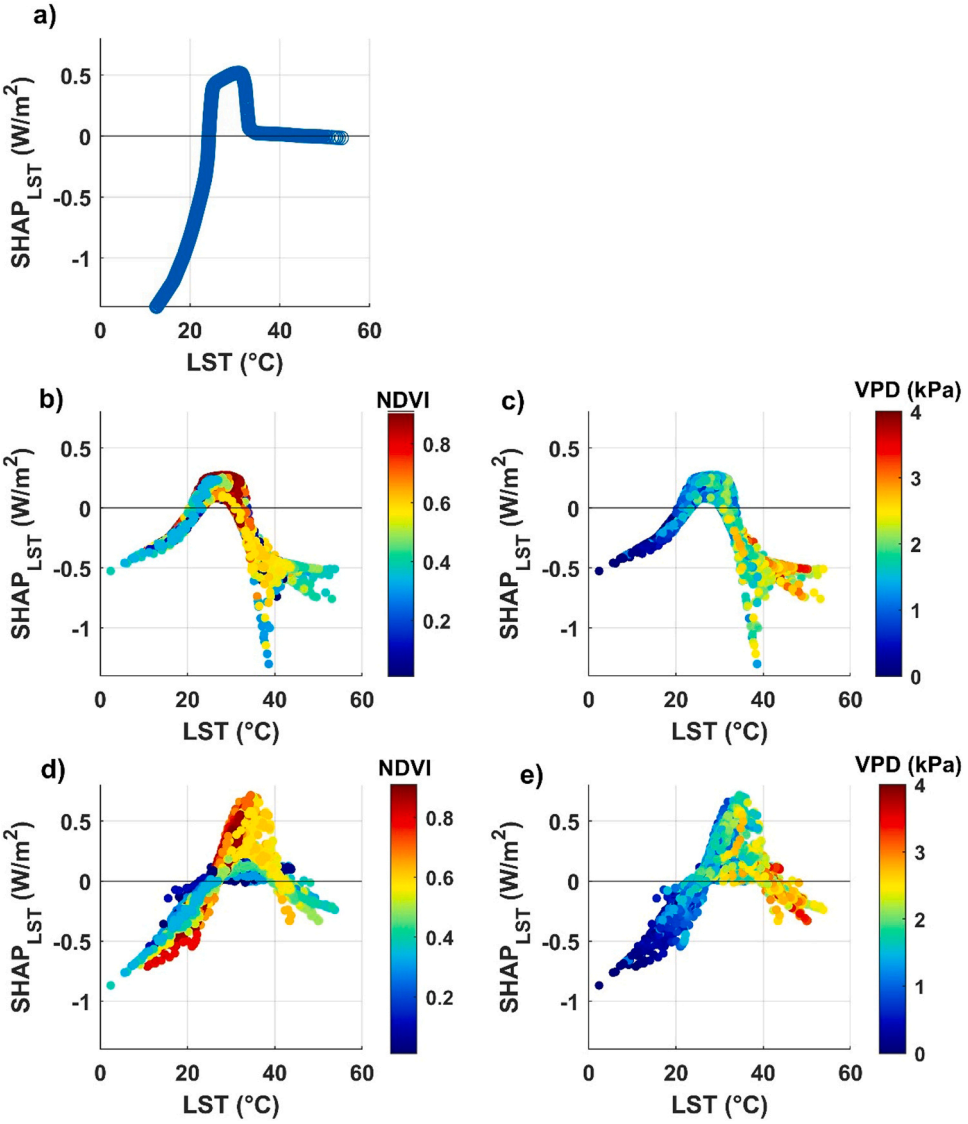


Fig. 9. SHAP values of LST plotted against the associated LST values for: (a) the ML model developed using only LST ([LST]); (b) the [PAR, LST] model with data values colored according to their respective NDVI values; (c) the [PAR, LST] model with data values colored according to their respective VPD values; (d) the [PAR, LST, NDVI] model with data values colored according to their respective NDVI values; (e) the [PAR, LST, NDVI] model with data values colored according to their respective VPD values.

underscores the importance of proximal remote sensing variables in addressing the limitations of traditional meteorological models by enhancing the model's sensitivity to rapid changes in crop phenology and surface temperature.

4. Conclusions

We present a systematic evaluation of the utility of machine learning to capture the variability in latent energy flux in managed agricultural systems across multiple growing seasons utilizing both meteorological and in-situ proximal remote sensing of NDVI and LST. We developed and rigorously cross-validated 64 sets of ML models that span predictor sets utilizing all combinations of one to six variables, including four environmental variables and two proximal remote sensing variables (NDVI and LST).

The findings of the study emphasize the importance of carefully selecting predictors or sets of predictors for surface energy balance estimation. For example, the performance of two-predictor models, such as [PAR, LST] and [PAR, NDVI], substantially improved performance over single-predictor models, and performed almost as well as the best three- and four-predictor models. While predictive performance generally increases with model complexity, careful selection of predictor sets can result in parsimonious models that leverage synergistic information for predictions of land-atmosphere interactions.

This study highlights the utility of proximal remote sensing predictors for estimating LE, particularly for models that include only two or three predictors: [PAR, LST], [PAR, NDVI], [PAR, LST, NDVI]. These relatively simple models utilizing only 2–3 predictors capture between 68 % and 81 % of the variability in half-hourly latent energy flux across two growing seasons of a Midwest US soybean system. These predictors provide the machine learning models with information about canopy phenology and plant physiological status that is not available in meteorological observations alone. An explainability analysis provided insight into how the ML models utilize the information in remote sensing observations, allowing the models to use available information in biophysically meaningful ways. This work points to the need to evaluate ML models developed with weather and proximal sensing information to estimate crop water use in other crop species and climate contexts.

Author contributions: CRediT

All authors contributed to conceptualization, investigation, reviews and editing of manuscript drafts. DTD provided supervision, project administration and procured funds. SG conducted the analysis, visualization and wrote the original draft. ASG and AVL oversaw data curation and field resources and field experiment investigation.

CRediT authorship contribution statement

Drewry Darren: Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Srishti Gaur:** Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Andy VanLoocke:** Writing – review & editing, Validation, Resources, Investigation, Data curation. **Aslan Sungur Guler:** Writing – review & editing, Resources, Investigation, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

DTD and SG acknowledge support from the National Science Foundation Cyber-Physical Systems Program (Award 1954556). DTD also acknowledges support from the National Aeronautics and Space Administration (Award #80NSSC20K1789), the College of Food, Agricultural and Environmental Sciences and the Translational Data Analytics Institute at Ohio State University. This work was supported, in part, by Hatch funds from the USDA National Institute of Food and Agriculture (Hatch Project OHO01509) and by the DOE Center for Advanced Bioenergy and Bioproducts Innovation, U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under Award Number DE-SC0018420.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.agwat.2025.109643.

Data availability

The data used in this study will be made available through contacting the SABR site principle investigator.

References

Abdullah, S.S., Malek, M.A., Abdullah, N.S., Kisi, O., Yap, K.S., 2015. Extreme learning machines: a new approach for prediction of reference evapotranspiration. J. Hydrol. 527, 184–195. https://doi.org/10.1016/j.jhydrol.2015.04.073.

Amani, S., Shafizadeh-Moghadam, H., 2023. A review of machine learning models and influential factors for estimating evapotranspiration using remote sensing and ground-based data. Agric. Water Manag. 284, 108324. https://doi.org/10.1016/j.agwat.2023.108324.

Anderson, M.C., Kustas, W.P., Norman, J.M., Hain, C.R., Mecikalski, J.R., Schultz, L., González-Dugo, M.P., Cammalleri, C., d'Urso, G., Pimstein, A., Gao, F., 2011. Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. Hydrol. Earth Syst. Sci. 15, 223–239. https://doi.org/10.5194/hess-15-223-2011.

Anderson, M.C., Allen, R.G., Morse, A., Kustas, W.P., 2012. Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources. Remote Sens. Environ. 122, 50–65. https://doi.org/10.1016/j.rse.2011.08.025.

Arif, C., Saptomo, S.K., Setiawan, B.I., Taufik, M., Suwarno, W.B., Mizoguchi, M., 2022. A Model of evapotranspirative irrigation to manage the various water levels in the system of rice intensification (SRI) and its effect on crop and water productivities. Water 14, 170. https://doi.org/10.3390/w14020170.

Bai, G., Barker, B., Scoby, D., Irmak, S., Luck, J.D., Neale, C.M.U., Schnable, J.C., Awada, T., Kustas, W.P., Ge, Y., 2024. High-throughput physiological phenotyping of crop evapotranspiration at the plot scale. Field Crops Res. 316, 109507. https://doi.org/10.1016/j.fcr.2024.109507.

Bai, Y., Zhang, S., Bhattarai, N., Mallick, K., Liu, Q., Tang, L., Im, J., Guo, L., Zhang, J., 2021. On the use of machine learning based ensemble approaches to improve evapotranspiration estimates from croplands across a wide environmental gradient. Agric. For. Meteorol. 298–299, 108308. https://doi.org/10.1016/j.agrformet.2020.108308.

Baldocchi, D.D., 2020. How eddy covariance flux measurements have contributed to our understanding of *Global Change Biology*. Glob. Change Biol. 26, 242–260. https://doi.org/10.1111/gcb.14807.

Baptista, M.L., Goebel, K., Henriques, E.M.P., 2022. Relation between prognostics predictor evaluation metrics and local interpretability SHAP values. Artif. Intell. 306, 103667. https://doi.org/10.1016/j.artint.2022.103667.

Bendorf, J., Heaton, E., Hartman, T., Aslan-Sungur, G. (Rojda), VanLoocke, A., 2022. Agroecosystem model simulations reveal spatial variability in relative productivity in biomass sorghum and maize in Iowa, USA. GCB Bioenergy 14, 1336–1360. https://doi.org/10.1111/gcbb.13004.

Bhattarai, N., Mallick, K., Stuart, J., Vishwakarma, B.D., Niraula, R., Sen, S., Jain, M., 2019. An automated multi-model evapotranspiration mapping framework using remotely sensed and reanalysis data. Remote Sens. Environ. 229, 69–92. https://doi.org/10.1016/j.rse.2019.04.026.

Brombacher, J., Silva, I.R.D.O., Degen, J., Pelgrum, H., 2022. A novel evapotranspiration based irrigation quantification method using the hydrological similar pixels algorithm. Agric. Water Manag. 267, 107602. https://doi.org/10.1016/j.agwat.2022.107602.

Campbell, G.S., Norman, J.M., 2000. *An introduction to environmental biophysics*. Springer Science & Business Media.

Castellví, F., Snyder, R.L., 2010. A comparison between latent heat fluxes over grass using a weighing lysimeter and surface renewal analysis. J. Hydrol. 381, 213–220. https://doi.org/10.1016/j.jhydrol.2009.11.043.

Chen, H., Ghani Razaqpur, A., Wei, Y., Huang, J.J., Li, H., McBean, E., 2023. Estimation of global land surface evapotranspiration and its trend using a surface energy balance constrained deep learning model. J. Hydrol. 627, 130224. https://doi.org/10.1016/j.jhydrol.2023.130224.

Chen, J.M., Liu, J., 2020. Evolution of evapotranspiration models using thermal and shortwave remote sensing data. Remote Sens. Environ. 237, 111594. https://doi.org/10.1016/j.rse.2019.111594.

Cihlar, J., St.-Laurent, L., Dyer, J.A., 1991. Relation between rainfall and the normalized difference vegetation index and ecological variables. Remote Sens. Environ. 35 (2–3), 279–298.

Cross, J.F., Drewry, D.T., 2024. Ensemble machine learning for interpretable soil heat flux estimation. Ecol. Inform. 82, 102697. https://doi.org/10.1016/j.ecoinf.2024.102697.

Dela Cruz, J.C., Caya, M.V.C., Ballado, A.H., Aggabao, M.C.R., Bacolor, E.I., Riego, H.A. G.G., Vergara, M.E.M., 2020. Evapotranspiration-based irrigation system for mustard green crop cultivation using public weather forecast. In: 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC). Presented at the 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC). IEEE, Shah Alam, Malaysia, pp. 203–208. https://doi.org/10.1109/ICSGRC49013.2020.9232454.

Denich, C., Bradford, A., 2010. Estimation of evapotranspiration from bioretention areas using weighing lysimeters. J. Hydrol. Eng. 15, 522–530. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000134.

Dold, C., Hatfield, J.L., Prueger, J.H., Moorman, T.B., Sauer, T.J., Cosh, M.H., Drewry, D. T., Wacha, K.M., 2019. Upscaling gross primary production in corn-soybean rotation systems in the midwest. Remote Sens. 11. https://doi.org/10.3390/rs11141688.

Drewry, D.T., Kumar, P., Long, S., Bernacchi, C., Liang, X.Z., Sivapalan, M., 2010. Ecohydrological responses of dense canopies to environmental variability: 2. Role of acclimation under elevated CO2. J. Geophys. Res. 115, G04023. https://doi.org/10.1029/2010jg001341.

Drewry, D.T., Kumar, P., Long, S., Bernacchi, C., Liang, X.Z., Sivapalan, M., 2010. Ecohydrological responses of dense canopies to environmental variability: 1. Interplay between vertical structure and photosynthetic pathway. J. Geophys. Res. 115, G04022. https://doi.org/10.1029/2010jg001340.

Dugas, W.A., Fritschen, L.J., Gay, L.W., Held, A.A., Matthias, A.D., Reicosky, D.C., Steduto, P., Steiner, J.L., 1991. Bowen ratio, eddy correlation, and portable chamber measurements of sensible and latent heat flux over irrigated spring wheat. Agric. For. Meteorol. 56, 1–20. https://doi.org/10.1016/0168-1923(91)90101-U.

Eamus, D., Cleverly, J., Boulain, N., Grant, N., Faux, R., Villalobos-Vega, R., 2013. Carbon and water fluxes in an arid-zone acacia savanna woodland: an analyses of seasonal patterns and responses to rainfall events. Agric. For. Meteorol. 182–183, 225–238. https://doi.org/10.1016/j.agrformet.2013.04.020.

Feldman, A.F., Short Gianotti, D.J., Dong, J., Trigo, I.F., Salvucci, G.D., Entekhabi, D., 2023. Tropical surface temperature response to vegetation cover changes and the role of drylands. Glob. Change Biol. 29 (1), 110–125.

Ferguson, J.N., Fernandes, S.B., Monier, B., Miller, N.D., Allen, D., Dmitrieva, A., Schmuker, P., Lozano, R., Valluru, R., Buckler, E.S., Gore, M.A., Brown, P.J., Spalding, E.P., Leakey, A.D.B., 2021. Machine learning-enabled phenotyping for GWAS and TWAS of WUE traits in 869 field-grown sorghum accessions. Plant Physiol. 187, 1481–1500. https://doi.org/10.1093/plphys/kiab346.

Gaur, S., Drewry, D.T., 2024. Explainable machine learning for predicting stomatal conductance across multiple plant functional types. Agric. For. Meteorol. 350, 109955. https://doi.org/10.1016/j.agrformet.2024.109955.

Gaur, S., Singh, B., Bandyopadhyay, A., Stisen, S., Singh, R., 2022. Spatial pattern-based performance evaluation and uncertainty analysis of a distributed hydrological model. Hydrol. Process. 36, e14586. https://doi.org/10.1002/hyp.14586.

Granata, F., 2019. Evapotranspiration evaluation models based on machine learning algorithms—a comparative study. Agric. Water Manag. 217, 303–315. https://doi.org/10.1016/j.agwat.2019.03.015.

Harder, P., Helgason, W.D., Johnson, B., Pomeroy, J.W., 2023. Crop water use efficiency from eddy covariance methods in cold water-limited regions. Agric. For. Meteorol. 341, 109657. https://doi.org/10.1016/j.agrformet.2023.109657.

Hu, C., Li, L., Li, Y., Wang, F., Hu, B., Peng, Z., 2022. Explainable machine-learning model for prediction of in-hospital mortality in septic patients requiring intensive care unit readmission. Infect. Dis. Ther. 11, 1695–1713. https://doi.org/10.1007/s40121-022-00671-3.

Huang, L., Cai, J., Zhang, B., Chen, H., Bai, L., Wei, Z., Peng, Z., 2019. Estimation of evapotranspiration using the crop canopy temperature at field to regional scales in large irrigation district. Agric. For. Meteorol. 269–270, 305–322. https://doi.org/10.1016/j.agrformet.2019.02.024.

Jung, H.C., Getirana, A., Arsenault, K.R., Holmes, T.R., McNally, A., 2019. Uncertainties in evapotranspiration estimates over West Africa. Remote Sens. 11 (8), 892.

Kustas, W.P., Norman, J.M., 1996. Use of remote sensing for evapotranspiration monitoring over land surfaces. Hydrol. Sci. J. 41, 495–516. https://doi.org/10.1080/02626669609491522.

Le, P.V.V., Kumar, P., Drewry, D.T., Quijano, J.C., 2012. A graphical user interface for numerical modeling of acclimation responses of vegetation to climate change. Comput. Geosci. 49, 91–101. https://doi.org/10.1016/j.cageo.2012.07.007.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Li, L., Nielsen, D.C., Yu, Q., Ma, L., Ahuja, L.R., 2010. Evaluating the crop water stress Index and its correlation with latent heat and CO2 fluxes over winter wheat and maize in the North China plain. Agric. Water Manag. 97, 1146–1155. https://doi.org/10.1016/j.agwat.2008.09.015.

Liu, L., Cao, R., Chen, J., Shen, M., Wang, S., Zhou, J., He, B., 2022. Detecting crop phenology from vegetation index time-series data by improved shape model fitting in each phenological stage. Remote Sens. Environ. 277, 113060. https://doi.org/10.1016/j.rse.2022.113060.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17. Curran Associates Inc., Red Hook, NY, USA, pp. 4768–4777.

Lucarini, A., Cascio, M.L., Marras, S., Sirca, C., Spano, D., 2024. Artificial intelligence and Eddy covariance: a review. Sci. Total Environ., 175406

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9.

Mallick, K., Bhattacharya, B.K., Chaurasia, S., Dutta, S., Nigam, R., Mukherjee, J., Banerjee, S., Kar, G., Rao, V.U.M., Gadgil, A.S., Parihar, J.S., 2007. Evapotranspiration using MODIS data and limited ground observations over selected agroecosystems in India. Int. J. Remote Sens. 28, 2091–2110. https://doi.org/10.1080/01431160600935620.

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. J. Hydrometeorol. 14, 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., De Moraes, O.L.L., Hayek, M., Munger, J.W., Wofsy, S.C., 2016. Canopy-scale biophysical controls of transpiration and evaporation in the Amazon Basin. Hydrol. Earth Syst. Sci. 20, 4237–4264. https://doi.org/10.5194/hess-20-4237-2016.

Mehdiyev, N., Majlatow, M., Fettke, P., 2024. Quantifying and explaining machine learning uncertainty in predictive process monitoring: an operations research perspective. Ann. Oper. Res. 1–40.

Oren, R., Sperry, J.S., Katul, G.G., Pataki, D.E., Ewers, B.E., Phillips, N., Schäfer, K.V.R., 1999. Survey and synthesis of intra- and interspecific variation in stomatal sensitivity to vapour pressure deficit. Plant Cell Environ. 22, 1515–1526. https://doi.org/10.1046/j.1365-3040.1999.00513.x.

Payero, J.O., Irmak, S., 2008. Construction, installation, and performance of two repacked weighing lysimeters. Irrig. Sci. 26, 191–202. https://doi.org/10.1007/s00271-007-0085-9.

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. in: Knowledge-Guided Machine Learning. Chapman and Hall/CRC, Boca Raton, pp. 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, 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150.

Vidyarthi, V.K., Jain, A., 2020. Knowledge extraction from trained ANN drought classification model. J. Hydrol. 585, 124804. https://doi.org/10.1016/j.jhydrol.2020.124804.

Vidyarthi, V.K., Jain, A., 2023. Advanced rule-based system for rainfall occurrence forecasting by integrating machine learning techniques. J. Water Resour. Plan. Manag. 149, 04022072. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001631.

Vilares Ferro, M., Doval Mosquera, Y., Ribadas Pena, F.J., Darriba Bilbao, V.M., 2023. Early stopping by correlating online indicators in neural networks. Neural Netw. 159, 109–124. https://doi.org/10.1016/j.neunet.2022.11.035.

Wang, P., Li, D., Liao, W., Rigden, A., Wang, W., 2019. Contrasting evaporative responses of ecosystems to heatwaves traced to the opposing roles of vapor pressure deficit and surface resistance. Water Resour. Res. 55, 4550–4563. https://doi.org/10.1029/2019WR024771.

Wang, L., Zhang, Y., Yao, Y., Xiao, Z., Shang, K., Guo, X., Yang, J., Xue, S., Wang, J., 2021. GBRT-based estimation of terrestrial latent heat flux in the haihe river basin from satellite and reanalysis datasets. Remote Sens. 13 (6), 1054.

Wu, B., Zhu, W., Yan, N., Xing, Q., Xu, J., Ma, Z., Wang, L., 2020. Regional actual evapotranspiration estimation with land and meteorological variables derived from multi-source satellite data. Remote Sens. 12 (2), 332.

Yamaç, S.S., Todorovic, M., 2020. Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data. Agric. Water Manag. 228, 105875. https://doi.org/10.1016/j.agwat.2019.105875.

Yebra, M., Van Dijk, A.I., Leuning, R., Guerschman, J.P., 2015. Global vegetation gross primary production estimation using satellite-derived light-use efficiency and canopy conductance. Remote Sens. Environ. 163, 206–216.

Yi, K., Senay, G.B., Fisher, J.B., Wang, L., Suvočarev, K., Chu, H., Moore, G.W., Novick, K. A., Barnes, M.L., Keenan, T.F., Mallick, K., 2024. Challenges and future directions in quantifying terrestrial evapotranspiration. Water Resour. Res. 60 (10), e2024WR037622.

Zhang, T., Qiu, R., Ding, R., Wu, J., Clothier, B., 2023. Multi-scale spectral characteristics of latent heat flux over flooded rice and winter wheat rotation system. Agric. Water Manag. 288, 108471. https://doi.org/10.1016/j.agwat.2023.108471.

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. Geophys. Res. Lett. 46, 14496–14507. https://doi.org/10.1029/2019GL085291.

Zhi, W., Appling, A.P., Golden, H.E., Podgorski, J., Li, L., 2024. Deep learning for water quality. Nat. Water 2, 228–241. https://doi.org/10.1038/s44221-024-00202-z.