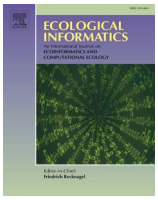
Ecological Informatics 82 (2024) 102697

Contents lists available at ScienceDirect

Ecological Informatics

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



Ensemble machine learning for interpretable soil heat flux estimation

James F. Cross [a,b], Darren T. Drewry [a,b,c,d,*]

[a] Department of Food, Agricultural and Biological Engineering, Ohio State University, Columbus, OH 43210, USA
[b] Environmental Sciences Graduate Program, Ohio State University, Columbus, OH 43210, USA
[c] Department of Horticulture and Crop Science, Ohio State University, Columbus, OH 43210, USA
[d] Translational Data Analytics Institute, Ohio State University, Columbus, OH 43210, USA

ARTICLE INFO

Keywords:
Soil heat flux
Surface energy balance
Explainable machine learning
Remote sensing
Land surface temperature

ABSTRACT

Soil heat flux (SHF) is a key component of the surface energy balance and driver of soil physiochemical and biological processes. Accurate estimation of soil heat flux is challenging due to variations in soil composition, overlying vegetation density and phenology, and highly variable environmental forcings. Existing SHF process-based and data-driven estimation methods have focused on midday landscape scale estimates that correspond to satellite acquisitions, despite the high variability that SHF displays at diurnal scales and throughout the growing season. Recently developed data-driven techniques have emphasized pre-determined predictor variables, with model complexity and predictor selection not carefully evaluated, leaving a gap in our understanding of what information is required to accurately predict SHF. Here we developed and evaluated a suite of ensemble machine learning (ML) models to quantify the ability of meteorological and remote sensing data to predict SHF variability at half-hourly temporal resolution throughout a growing season, producing a comprehensive evaluation of the importance of predictor set composition for SHF estimation at high temporal resolution. We compared this suite of machine learning models to six semi-empirical models and found that the machine learning models broadly outperformed the existing models in capturing diurnal variability across the growing season for four agro-ecosystems (soybean, corn, sorghum, and miscanthus). Crop-specific ML models were able to capture over 86% of the variability in SHF using only two predictor variables, pointing to the need for careful evaluation of predictor sets to identify synergistic combinations. ML models developed using pooled data across all crops captured almost 80% of SHF variability using three predictors, demonstrating the power and generalizability of these methods independent of crop type. Shapley additive explanations (SHAP) were used to examine model interpretability, providing insights into the typically opaque ML modelling process and interaction of predictor variables. Models trained with fewer input variables tended to display more linear and interpretable feature attribution, motivating the use of interpretability as an important consideration in parsimonious model selection. These results provide a robust demonstration of the ability of ML to capture the variability in SHF at sub-hourly resolution across growing seasons spanning a wide range of phenological variation for unique agricultural systems. This study provides a comprehensive evaluation of predictor requirements for model performance, guiding future applications that will take advantage of the next generation of satellite-based observing systems, or in-situ proximal observations of vegetation status and meteorological conditions.

1. Introduction

Soil heat flux (SHF) is the thermal energy exchange between the environment immediately above the surface and the topmost surface soil layer. SHF is a crucial component of the surface energy balance which determines energy transfer into the soil system. The net radiation of the terrestrial land surface (shortwave and longwave inputs minus reflected and emitted radiation) is utilized for metabolic processes, latent energy flux (LE), sensible heat flux (H) and soil heat flux (Allen, 1998; Drewry et al., 2010a, 2010b; Priestley and Taylor, 1972). Accurate modelling of surface energy conditions is critical for a range of applications including climate modelling (Franzke et al., 2020), crop modelling (Gupta et al., 2022; Oliver et al., 1987) and water management (Mandal et al., 2020; Yao et al., 2010). Improving soil heat flux characterization is therefore required to fully understand surface energy balance through energy balance closure approaches and conditions of the surface environment

* Corresponding author at: 590 Woody Hayes Drive, Columbus, OH 43201, USA.
E-mail address: drewry.19@osu.edu (D.T. Drewry).

https://doi.org/10.1016/j.ecoinf.2024.102697
Received 3 July 2023; Received in revised form 19 June 2024; Accepted 19 June 2024
Available online 21 June 2024

(Bai et al., 2022).

Quantifying soil heat flux is essential for predicting the dynamics of soil temperature, which controls a wide range of processes that are critical for understanding the soil microclimate (Sauer and Peng, 2020), soil biogeochemical cycles (Pregitzer and King, 2005) and plant growth (Brown, 2018). Soil heat flux is linearly proportional to the change in soil temperature with time, alongside changes in soil characteristics. Soil temperature regulates mineralization, a primary source of nutrients for plants through the decomposition of organic matter (Burke et al., 1989; Schimel et al., 1994). In agricultural soils, increases in nitrogen uptake and short-term nitrogen availability have been linked with increased soil temperatures (Dessureault-Rompré et al., 2010; Gavito et al., 2001; Miller and Geisseler, 2018; Pregitzer and King, 2005). Increased soil temperature has also been linked with changes in root morphology through specific root length and branching angle that drives an increase in nutrient accessibility (Pregitzer and King, 2005). Beyond nutrient availability, soil temperature controls soil microbe communities and broader soil biodiversity (Bodelier, 2011). Globally, soil acts as a carbon reservoir, the extent and stability of which is principally governed by soil temperature (Ciais et al., 2014; Schlesinger, 1997). Ongoing efforts to sequester carbon through agriculture and forestry management practices depend on the capacity to model soil energy fluxes across a range of land-use conditions (Pechanec et al., 2018; Zhai et al., 2024).

Despite the central role of soil heat flux in the surface energy balance and soil biogeochemical cycling, estimating this important land surface energy flux remains a challenging task. Direct measurement is labor-intensive and often requires disturbing the soil column to place instrumentation which provides limited spatial scope. Remote sensing has been demonstrated to accurately retrieve biophysical parameters at moderate spatial scales (Garcia-Santos et al., 2022; Kustas et al., 1993; Liu, 2022; Zheng and Jia, 2022), but retrieval of SHF is particularly challenging due to the complexity of formulating a biophysical model for a flux with many potentially important predictive factors such as near-surface climate (Santanello and Friedl, 2003), vegetative cover and associated phenological changes (Clothier et al., 1986; Kustas et al., 1993) and soil moisture content (Idso et al., 1975).

Prediction of soil heat flux can be done through thermal diffusion calculations, which are limited by a dependence on estimates of soil characteristics and soil temperature (Evett et al., 2012; Lettau, 1954; Wang and Bras, 1999). Several studies have focused on the development of semi-empirical models that estimate soil heat flux through observations that tend to be widely available (Bastiaanssen et al., 1998; Choudhury, 1999; Gao et al., 2017; Kustas et al., 1993; Santanello and Friedl, 2003) such as incident radiation, soil temperature, and plant cover. A recent review of these methods found limited temporal and spatial extensibility leading to the development of data-driven approaches for regional and global SHF estimation (Purdy et al., 2016).

Machine learning methods provide a powerful basis for developing flexible, non-parametric models of complex phenomena and have demonstrated strong predictive capabilities across many areas of the physical sciences generally and the earth sciences specifically (Bai et al., 2021; Gewali et al., 2019; Kamilaris and Prenafeta-Boldú, 2018; Reichstein et al., 2019). Several studies have successfully estimated soil heat flux for a range of environmental and spatial contexts (Bonsoms and Boulet, 2022; Canelón and Chávez, 2011; de Andrade et al., 2021; Zheng and Jia, 2022). de Andrade et al. (2021) and Bonsoms and Boulet (2022) targeted a mixture of field meteorological observations of net radiation and surface temperature, and satellite observations to train neural network and random forest architectures for regional-scale soil heat flux estimation across South America and Africa, respectively. Zheng and Jia (2022) used daily observations from 144 FluxNet sites and satellite observations of leaf area index (LAI) and albedo to produce a globally applicable soil heat flux model. A key limitation in these studies is their limitation to midday observations, defined by satellite overpass times. Santanello and Friedl (Santanello and Friedl, 2003) recognized significant variation in the relationship between SHF and net radiation across the day resulting in asymmetric diurnal SHF patterns which may be lost when only midday forcings are considered.

While machine learning methods have been demonstrated to be flexible predictive tools capable of integrating diverse data streams, they present significant challenges in terms of interpretability and generalizability (Reichstein et al., 2019). This is especially true in the context of ecohydrological or biophysical systems, where the objective is often to develop a better understanding of the underlying system rather than exclusively improve predictive performance (Drewry et al., 2010a, 2010b). There is a growing recognition that interpretability, physical consistency, and data complexity are key challenges in the successful adoption of machine learning methodologies (Baptista et al., 2022; Hu et al., 2021). One proposed approach is the concept of parsimony which seeks to use the simplest set of explanatory variables to reach a target performance threshold (de Silva et al., 2020). Parsimony is founded on concepts from dominant balance physics, a modelling approach that approximates complex physical systems through relatively few and simple dominant physical processes which can improve model generalizability, interpretability, and physical consistency (Callaham et al., 2021; Kaiser et al., 2022; Kutz and Brunton, 2022). Here we refer to parsimony as a minimal or reduced predictor set (i.e. the number of observed physical variables used to predict SHF) that provides near-maximum (but likely somewhat below the absolute maximum) model performance achieved across all models.

The main objective of this research is to quantify the limits of predictive performance for soil heat flux using widely available and non-invasive environmental and remote sensing observations at sub-hourly temporal resolution. We leverage the power of machine learning, with rigorous cross-validation, to explore the information content in each potential predictor variable and their combinations to understand which combinations of variables provide optimal predictive power (i.e. minimize redundancy and maximize synergy). We conduct this work at a unique agricultural study site in which homogeneous fields of four contrasting agricultural systems (soybean, corn, miscanthus and sorghum) are instrumented with a wide array of sensors including eddy covariance flux systems, net radiometers, meteorological stations and proximal remote sensing observations of reflectance indices and radiometric surface temperature.

2. Methods

2.1. SABR experimental study site

This study was conducted at the Iowa State University Sustainable Advanced Bioeconomy Research Farm (SABR) outside of Ames, Iowa (Aslan-Sungur et al., 2023; Bendorf et al., 2022). The 2021 growing season around Ames was characterized by higher temperatures and less precipitation on average, particularly during the critical period of crop development in June (Thessen et al., 2022).

SABR consists of four independently monitored plots to examine the performance of standard Midwest commodity crops (corn, soybean) relative to potential bioenergy crops (energy sorghum, miscanthus) in the same climate and soil system. At the center of each plot is a tower with an eddy covariance (EC) system to continuously monitor the exchange of carbon dioxide, water and energy between each crop and the atmosphere, as well as a full suite of meteorological sensors at several meters above each land cover type (Campbell Scientific HMP-155), downwelling shortwave (SW) and longwave (LW) radiation (Kipp & Zonen CNR4 net radiometer), and sensors to monitor soil temperature and soil heat flux at a depth of 10 cm (Hukseflux HFP01SC soil heat flux plates). All flux and environmental variables were averaged to 30-min intervals, and independently measured at each of the four SABR towers. More information on the sensors and equipment at the site can be found on the SABR website (https://sabr.shinyapps.io/appSABR/).

Additionally, at each tower a set of in-situ proximal remote sensing instruments were deployed over the growing season to continuously

monitor land surface temperature (LST), the normalized difference vegetation index (NDVI) (Tucker, 1979) and the photochemical reflectance index (PRI) (Gamon et al., 1997), providing high temporal resolution observations of these remote sensing variables allowing us to examine their importance in diurnal SHF estimation. NDVI and PRI were collected using two-band radiometers (SRS sensors from METER Inc., Washington, USA) affixed to each of the four towers. These sensors were positioned such that they viewed a homogenous region of the canopy to the south of each tower. LST measurements were collected using infrared thermometers (SI-421, SI-431, SI-4H1 infrared thermometers, Apogee Instruments, Utah, USA) deployed on each tower looking down on the land surface / canopy from an approximate height of 3–4 m above the surface. These in-situ remote sensors collected observations at 5-min intervals which were then averaged to 30-min intervals to correspond with the averaging period of the other tower observations.

Proximal remote sensing observations were made from June 9th to October 1st. This produced approximately 5500 half-hourly observation periods for each of the crops. Corn, sorghum, and soybean were planted on April 28th, June 2nd, and May 12th in 2021, respectively. The perennial Miscanthus crop was originally planted on May 3rd, 2019.

2.2. Data post-processing

2.2.1. Soil heat flux

Soil heat flux was measured at each of the four SABR EC towers using three heat flux plates in the proximity of each tower. All SHF measurements were made at a depth of 10 cm. To avoid introducing additional sources of uncertainty that could be induced by empirical approaches to transform 10 cm observations to the heat flux at the soil surface, all machine learning models were developed here using the observed SHF at 10 cm (Gao et al., 2017). Section 2.4 discusses corrections to estimate the surface soil heat flux. No gap filling was performed on soil heat flux observations. In this study, we use environmental and proximal remote sensing observations as predictors in SHF estimation. Below we briefly describe these two classes of predictor variables.

2.2.2. Environmental data

The environmental predictors are based on meteorological measurements from the four SABR fields containing instrumented towers. Where meteorological data is missing for one of the fields, it is first filled from a nearby meteorological station at the Ames Municipal Airport, located 6.9 km from the site, as part of the standard processing methods applied to SABR data. Further details are available at the SABR webpage (https://sabr.shinyapps.io/appSABR/). Remaining missing data points are filled using a linear transformation from one of the SABR towers where the missing variables are available.

2.2.3. Remote sensing data

The remote sensing predictors for this study were collected at 5-min intervals at each of the four SABR towers. These remote sensing observations included: NDVI (Tucker, 1979), PRI (Gamon et al., 1997), and LST collected using calibrated infrared thermometers. Each remote sensing variable was averaged to 30-min intervals to be consistent with the temporal resolution of SABR tower observations. NDVI and PRI were each calculated from two-band sensors centered at 650 nm and 810 nm, with 50 nm and 40 nm full width half maximum (FWHM) band widths respectively, for NDVI, and 532 nm and 570 nm with 10 nm FWHM band widths for PRI. To reduce noise and variability of these indices, a 4-h window around local solar noon was selected and averaged for each day to produce daily NDVI and PRI values. This provided dynamic estimates of these indices for each crop throughout the growing season while minimizing sensor noise and eliminating periods during which these calculations are invalid (i.e. nighttime and periods of low radiation forcing). This averaging makes NDVI and PRI effectively daily observations, as opposed to LST and the environmental predictor variables that were all used at the 30-min resolution for which SHF was measured and modeled.

It should be noted that in the literature there is some inconsistency regarding the definition of surface temperature, which can refer to either the soil surface (T_S) or the canopy surface (T_C). Throughout this paper we use radiometric observations from a height above each crop canopy as the source of surface temperature observations and will use LST to refer to these remote sensing observations as they may partially observe the soil surface or the canopy depending on canopy density and coverage.

2.3. Study crops

Corn, miscanthus and sorghum are C4 grasses that have potential to be used for future biofuel production (Carpita and McCann, 2008). Miscanthus is a perennial crop known for its high water and nutrient use efficiency (Vermerris, 2008). Sorghum, an annual crop, demonstrates drought tolerance and high rates of biomass production and so is also considered a promising bioenergy feedstock (Carpita and McCann, 2008). Soybeans are the only C3 crop considered in this study.

Variations in plant physiology, morphology, and phenology influence interactions throughout the plant-soil system (Moore et al., 2021; van der Putten et al., 2013). The state of canopy closure and canopy density (i.e. leaf area index) control the amount of incident shortwave radiation reaching the soil surface as well as the emission and propagation of longwave radiation throughout the canopy (Drewry et al., 2010a, 2010b).

Fig. S1 presents mean diurnal soil heat flux and soil temperature (T_S) at a 10 cm depth as well as downwelling shortwave radiation across the growing season for each of the four crops. These diurnal averages show that SHF peaks in the late afternoon / evening, preceding the peak in soil temperature at the 10 cm measurement depth. Both SHF and soil temperature show a lag relative to incident shortwave radiation.

Increased NDVI values (an indicator of a dense canopy) lead to a decrease in measured soil temperature at 10 cm for each of the four crops studied here (Fig. S2). Miscanthus is a perennial crop, and NDVI for it was relatively constant during the study period and does not show the negative trend in T_S seen for the other three annual crops.

2.4. Empirical soil heat flux models

Six empirical models were used to compute SHF for each of the four crops to provide benchmarks for assessing the performance of the machine learning models developed in this study (Bastiaanssen et al., 1998; Boegh et al., 2004; Choudhury, 1999; Gao et al., 2017; Kustas et al., 1993; Santanello and Friedl, 2003). These empirical models often include a dependence on a variable related to canopy structure (i.e. NDVI, LAI – four of the six empirical models evaluated here) and net radiation (all six empirical models), with land surface temperature also included in some instances as a forcing variable (see Table S1 in the Supplementary Material). We note that as described in the section below, we do not use an aggregate radiation variable such as the net radiation flux (R_n) as a predictor variable in this study as R_n requires information on both downwelling and upwelling, shortwave and longwave radiative fluxes which require specific sensors (net radiometers) that are not as widely available as other weather observations and would limit the applicability of the models developed here. In this way many of the ML models developed in this study are more widely applicable than these semi-empirical models evaluated here, as they only require standard meteorological data as predictors. In addition we evaluate models that include remote sensing data products (reflectance indices and land surface temperature) that are now widely available from satellite sources and proximal sensing methods.

Each of the six empirical models is presented in Table S1. As soil heat flux is typically estimated at the soil surface, a correction factor is necessary to compensate for the energy stored between the soil surface

and the heat flux plate at 10 cm (Sauer and Horton, 2005). Here we apply the calorimetric method (Fuchs, 1986) to perform the transformation between soil heat flux observations at 10 cm and soil surface heat flux estimates to evaluate the performance of the empirical SHF models. The calorimetric method is the most common approach to this transformation and calculates a soil heat storage term to describe the amount of energy stored in the soil between the surface and a given depth based on soil temperature, soil moisture and soil composition (Kustas and Daughtry, 1990; Liebethal et al., 2005; Liu et al., 2017).

Despite the popularity of the calorimetric method, several studies have examined the limitations of the method to accurately compensate for stored heat (Massman, 1992; Sauer and Horton, 2005). With this in mind, we chose to use the heat flux measurement at 10 cm depth for the development and evaluation of all ML models here. This prevents introducing additional biases or errors due to the choice of a specific correction methodology into the ML models (Gao et al., 2017), and provides a test of the ML models to accurately estimate the observed physical variable. We did test a selection of ML model formulations that were trained against surface soil heat flux estimates using the calorimetric method. These results are presented in Table S2.

In order to fairly evaluate the six empirical benchmark models, which were developed to estimate the surface soil heat flux, we produce independent parameter optimizations and evaluations of them both directly against SHF observations (Fig. S3) as well as evaluations against surface soil heat flux estimates developed by correcting SHF using the calorimetric method (Fig. S4). We found that performing a simple calibration for each of the six empirical benchmark models improved results for this SABR dataset. A nested, repeated k-fold cross-validation (described in Section 2.7) is used to parameterize each of these six empirical models to provide a set of optimized benchmarks against which to evaluate the performance of the machine learning models. The MATLAB Optimization Toolbox was used to calibrate model-specific parameters from the training set independently for each empirical model and crop combination.

2.5. Machine learning methods

Recent studies have found that using observations of land surface temperature, NDVI, and net radiation to train machine learning models can improve the performance of soil heat flux predictions. Bonsoms and Boulet (2022) analyzed sites throughout Europe and Africa and found that their models outperformed existing empirical models. However, this study was limited in temporal resolution due to the use of sun-synchronous satellite observations. Zheng and Jia (2022) used observations of surface albedo and LAI from the Global LAnd Surface Satellite (GLASS) with measurements of soil temperature, net radiation, and soil heat flux from 144 FluxNet sites to compare empirical methods to three machine learning models at a global scale. They found that all three machine learning methods tested outperformed empirical methods, with as much as 79% of observed variability captured using a random forest model. de Andrade et al. (2021) developed a neural network for soil heat flux estimation using MODIS observations across 23 flux towers in South America and found that adding a land cover classification to a dataset of surface temperature, albedo, net radiation and the enhanced vegetation index (EVI) improved predictions particularly during mid-morning. Each of these three studies focused on soil heat flux estimates at spatial resolutions of approximately $1km^2$ due to dependencies on satellite data, and likewise focused on daily estimations near midday due to typical satellite overpass times. This current study provides a significantly more robust evaluation of the ability of ML to capture the variability in SHF at sub-hourly resolution across a growing season spanning a wide range of phenological variation in four unique agricultural systems. We provide a comprehensive evaluation of predictor requirements for model performance, guiding future applications that will take advantage of the next generation of satellite-based observing systems, or in-situ proximal observations of vegetation status and meteorological conditions.

Among machine learning techniques, random forest models have been found to be particularly effective and are used here due to their demonstrated robust performance for a wide range of contexts. The use of random forests in a variety of areas of remote sensing is well-documented (Baret and Buis, 2008; Bonsoms and Boulet, 2022; Chan and Paelinckx, 2008; Maxwell et al., 2018; Zheng and Jia, 2022) and motivated by flexibility in hyperparameter selection, input dimensionality, and data volume. Random forest models are a class of machine learning methods referred to as ensemble learning, an architecture that operates by generating many randomized decision trees based on subsets of the input features (Breiman, 2001). By averaging the results from each tree, the model produces a more reliable result than any one tree could produce. The number of trees, leaves, and clustering methods are the primary hyperparameters required for application of the random forest method. Here, hyperparameter optimization was performed using Bayesian Optimization to minimize a 5 k-fold cross-validation loss during model development (Section 2.7).

2.6. Selection of predictor variables

Soil heat flux has been shown to correlate strongly with a range of environmental forcing variables such as downwelling shortwave and longwave radiation, soil moisture content and land surface temperature (Liebethal and Foken, 2007; Santanello and Friedl, 2003; Venegas et al., 2013). Likewise the density of vegetation above the surface plays an important role in controlling environmental conditions at the soil surface including the magnitudes of radiative fluxes incident on the surface (Bastiaanssen et al., 1998; de Andrade et al., 2021).

In many SHF empirical models the net radiation flux at the surface is used as a predictor variable. Here we use the individual downwelling components of R_n, SW and LW, and do not use observations of the upwelling radiation components that contain information on canopy structure and temperature. This provides a more rigorous test of SHF estimation that does not rely on 4-band net radiometer observations or other methods for estimating the two upwelling radiation fluxes. In this way the ML models developed in this study are more widely applicable than the empirical models evaluated here which all require R_n as an input variable.

We also include air temperature (T_a) and LST as independent predictor variables. While these two variables will be correlated throughout the growing season they represent unique states (atmospheric vs surface) that have the potential to influence SHF predictions in independent ways.

In total, seven predictor variables were explored in this study with each of these variables collected uniquely for each of the four agroecosystems (i.e. each of the four SABR measurement towers). These variables are: downwelling shortwave and longwave radiation, air temperature, LST, NDVI and PRI. We also include as a predictor variable solar zenith angle, which is calculated based on location, day of year and time of day.

Here we explore three categories of predictor variable sets for ML model training and evaluation. The first category examines the extent to which simple models, developed using sets of one, two or three predictor variables, can be used to accurately predict soil heat flux. The exploration of models in this category provides insights into the ability of models with simpler data requirements, and lower computational costs, to capture the variability in SHF. The second model category considers model formulations that utilize exclusively either meteorological observations (LW, SW, and T_a) or remote sensing (RS) observations that can now be made in-situ at high temporal resolution (PRI, NDVI, and LST). Due to the widespread availability of meteorological stations we also evaluated how the addition of a single remote sensing observation to the three meteorological predictors used here could improve soil heat flux estimation. The final model category we explored is one developed using the full set of predictor variables considered in this study,

spanning meteorological and remote sensing variables. We developed and evaluated these three categories of models for each agro-ecosystem represented at the SABR experimental facility (soybean, corn, sorghum, miscanthus) and also evaluate models developed using the combined dataset spanning all four systems (i.e. models that can be applied generally across these four agro-ecosystems).

Across these three categories we evaluated 65 predictor sets for each of the four agro-ecosystems, totaling 260 unique models. All predictors used here were averaged to 30-min resolution spanning June 9th to October 1st of the 2021 growing season.

2.7. Machine learning training methodology

To ensure accuracy and reliability of results, a repeated nested k-fold cross-validation approach was employed for all ML models developed here. An outer 10 k-fold cross-validation provided an initial division of the data into holdout data and data for model development (Molinaro et al., 2005). An inner 5 k-fold cross-validation was then used to minimize the bias caused by tuning of hyperparameters on training results, as demonstrated in previous studies (Cawley and Talbot, 2010; Yates et al., 2022). The training performance was calculated based on the inner cross-validated ensemble, while the testing performance was determined by selecting the optimal set of hyperparameters from the entire training set to predict on the holdout data. This process was repeated 15 times for each unique model to reduce the potential bias and variance related to random sampling (Kim, 2009).

This cross-validation methodology also allows for the assessment of uncertainty in the model performance and post-hoc explainability metrics, such as SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017). Typically SHAP explanations are based on evaluation of a single partitioning of a model and its set of hyperparameters. Scheda and Diciotti (2022) propose a cross-validation method for SHAP values, which we use here to assess the consistency of learned relationships.

Squared Pearson correlation values (R^2) are used to evaluate model performance due to the theoretical monotonic increase that it displays as additional predictors are added to model development (Chicco et al., 2021). In practice, an excessively large predictor set may reduce model performance (Banks and Fienberg, 2003), but was not found to be the case in this study. Further, R^2 is more interpretable than MSE and RMSE, allowing for a simpler model comparison (Chicco et al., 2021).

2.8. Machine learning interpretability

SHapley Additive exPlanations is a widely used technique for interpreting machine learning models after they have been trained (Lundberg and Lee, 2017). The method calculates a feature attribution score for each input feature (predictor) that provides an estimate of the contribution of that feature to the model prediction. Unlike other methods that are specific to certain model architectures, SHAP is model-agnostic and uses local information from the input data to provide explanations. We employed the built-in Shapley methods of MATLAB for this calculation (Lundberg and Lee, 2017). Following Scheda and Diciotti (2022) we averaged SHAP feature attributions across repeated k-fold cross-validations to reduce variance resulting from data partitioning.

We use SHAP values to analyze how different input sets affect the explainability and performance of the models developed here. These values provide detailed information on how individual predictors and their interactions contribute to model predictions, which aids in understanding the physical meaning of the models and provides insight into synergisms and redundancies among predictor variables. The motivation for evaluating such a wide range of models is to quantify how predictive performance and model complexity (number of predictor variables used) are related, and to provide interpretability to the interactions of predictor variables to inform model selection when considering complex goals such as parsimony.

3. Results and discussion

Summary performance statistics for each of the model categories (various combinations of predictors) are presented in Table 1. Across all model categories at least one machine learning model outperformed the existing empirical methods reviewed in Section 2.4 (see Figs. S3 and S4), demonstrating the flexibility of machine learning to optimize the use of information in even simple, but relevant, predictor sets to quantify SHF at diurnal to seasonal temporal scales.

In the first predictor sets we assess the capabilities of models requiring few input predictors and the importance of each predictor: the One Predictor, Two Predictor, and Three Predictor categories. The R^2 performance and variables in the optimal predictor set of each category and agro-ecosystem grouping are listed in the first three rows of Table 1.

The strongest single predictors for soil heat flux estimation were air temperature (corn and miscanthus) and LST (sorghum and soybean) among the seven predictors evaluated here. Across the Two Predictor models a combination of one of the two reflectance indices (NDVI and PRI) and one of the two temperature variables (T_a and LST) provided the best performance for all four systems. Two Predictor models that include both T_a and LST do not significantly improve predictive performance relative to the optimal One Predictor model due to the high correlation and redundancy between these two temperature variables (see Table S3 providing performance results for all Two Predictor models), pointing to the need to build parsimonious ML models using synergistic and non-redundant sources of information. Adding NDVI or PRI to one of the temperature variables adds independent information about canopy cover, and therefore the ability of incident radiation to reach the soil surface. These findings are consistent with those of Bonsoms and Boulet (2022) who found that both empirical and machine learning soil heat flux models benefit from inclusion of vegetation metrics and radiometric observations. The Three Predictor model class evaluations demonstrated that a combination of the reflectance and temperature variables

Table 1
R^2 validation performance averaged across ML models trained and evaluated respectively on each of the four crop systems: corn, miscanthus, sorghum, and soybean. Model input features are included for each result. In the One Predictor, Two Predictor, and Three Predictor model categories, only the best performing model (set of predictors) for each agro-ecosystem is displayed.

Model Category	Corn	Miscanthus	Sorghum	Soybean
One Predictor	0.790 ± 0.007 [T_a]	0.845 ± 0.008 [T_a]	0.686 ± 0.009 [LST]	0.758 ± 0.015 [LST]
Two Predictor	0.873 ± 0.010 [NDVI, T_a]	0.887 ± 0.006 [PRI, T_a]	0.875 ± 0.014 [NDVI, LST]	0.866 ± 0.013 [NDVI, LST]
Three Predictor	0.922 ± 0.015 [NDVI, PRI, T_a]	0.927 ± 0.015 [NDVI, PRI, T_a]	0.909 ± 0.028 [NDVI, PRI, LST]	0.895 ± 0.015 [NDVI, LST, T_a]
NDVI + LST	0.858 ± 0.009 [NDVI, LST]	0.852 ± 0.005 [NDVI, LST]	0.875 ± 0.014 [NDVI, LST]	0.866 ± 0.013 [NDVI, LST]
Remote Sensing	0.911 ± 0.015 [NDVI, PRI, LST]	0.909 ± 0.012 [NDVI, PRI, LST]	0.909 ± 0.028 [NDVI, PRI, LST]	0.903 ± 0.020 [NDVI, PRI, LST]
Meteorology	0.836 ± 0.005 [LW, SW, T_a]	0.897 ± 0.007 [LW, SW, T_a]	0.658 ± 0.021 [LW, SW, T_a]	0.764 ± 0.018 [LW, SW, T_a]
Meteorology + NDVI	0.906 ± 0.010 [LW, SW, T_a, NDVI]	0.922 ± 0.010 [LW, SW, T_a, NDVI]	0.890 ± 0.024 [LW, SW, T_a, NDVI]	0.911 ± 0.020 [LW, SW, T_a, NDVI]
Seven Predictor (Complete Set)	0.934 ± 0.018	0.954 ± 0.012	0.930 ± 0.017	0.948 ± 0.006

produced the optimal predictor set for all crops.

The next sets of model categories we evaluated are composed solely of either remote sensing or meteorological predictor variables. Remote sensing observations are particularly valuable for SHF prediction, with at least one of the three remote sensing variables (NDVI and LST) frequently selected as an optimal predictor in the One, Two, and Three Predictor categories. The purely remote sensing model using only the three remote sensing predictors [NDVI, PRI, LST], performed extremely well across all four agro-ecosystems ($R^2 > 0.9$ in all cases). The model based on only [NDVI, LST], two widely available remote sensing observables, likewise performed extremely well with >85% of the variability in SHF captured for all four systems.

Model performance for predictor sets restricted solely to meteorological variables was mixed across the four agro-ecosystems, with R^2 values ranging from approximately 0.65–0.9. The meteorology-only models underperformed the remote sensing model in all cases, most significantly in sorghum and soybean. The mixed performance of the meteorology-only models is indicative of a varied degree of decoupling between the environments above and within the canopy, potentially due in part to variations in canopy structure across the four ecosystems that could result in differences in canopy roughness and radiation transfer.

The combination of all seven predictors provided minor predictive improvements over the Three Predictor or Remote Sensing models (Table 1), suggesting that redundancy among some of the variables limits their utility in SHF estimation, and pointing to a much greater degree of parsimony in carefully selected 2- and 3-predictor model formulations.

In the subsequent sections, we assess model performance in the context of model complexity and explainability, crop specific training against generalized training, and contrast remote sensing and meteorological measurements to provide predictive performance for several potential use cases.

3.1. Model complexity

The management of model complexity is an important consideration in data-driven model development (Jolliffe, 1986; Wang et al., 2017; Yates et al., 2021). In this study we define model complexity as the number of variables included in the predictor set. Proper selection of variables, i.e. feature selection, can minimize training costs, measurement requirements, and decrease the likelihood of overfitting (Banks and Fienberg, 2003; Cawley and Talbot, 2010).

Here we examine the performance of models developed individually across the four agro-ecosystems at SABR. Fig. 1 presents model performance as a function of number of input variables, with the performance presented being that of the best performing model with the specified number of predictor variables. The optimal variable combinations for the first three sets for each ecosystem are included in Table 1. Extending the size of the predictor set to allow an additional variable produces at least equivalent predictive power to the previous in addition to potentially new information, as indicated by the R^2 metric (Chicco et al., 2021). The performance of all Two Predictor model permutations is provided in Table S3 of the Appendix for reference.

For each of the four agro-ecosystems, R^2 predictive performance reaches an approximate asymptote for three-predictor models, representing close to optimal performance. Between the best performing model trained on two predictors and the model trained on all seven predictors, there was an R^2 performance increase of 6.0% for corn, 5.8% for miscanthus, 3.5% for sorghum, and 8.9% for soybean (Fig. 1), indicating that in general Two or Three Predictor models, with appropriate predictor selection, can be considered parsimonious for SHF estimation in agricultural systems.

Performance of multi-variable models are dependent on the relevance of each variable to the predicted value, the independent information contained within each, and added information that may arise from the interaction between variables, which we refer to here as synergy. Variables such as incident shortwave radiation and zenith angle provide potentially redundant information. Additional variables may also act to fill any weaknesses left in a simpler model. For instance, soil heat flux is strongly correlated with LST, though this relationship is strongly influenced by the presence, or absence, of a canopy. A model lacking an indicator for canopy density, such as NDVI, would fail to capture this detail. In Section 3.5 we evaluate the impact of feature selection on model explainability in further detail.

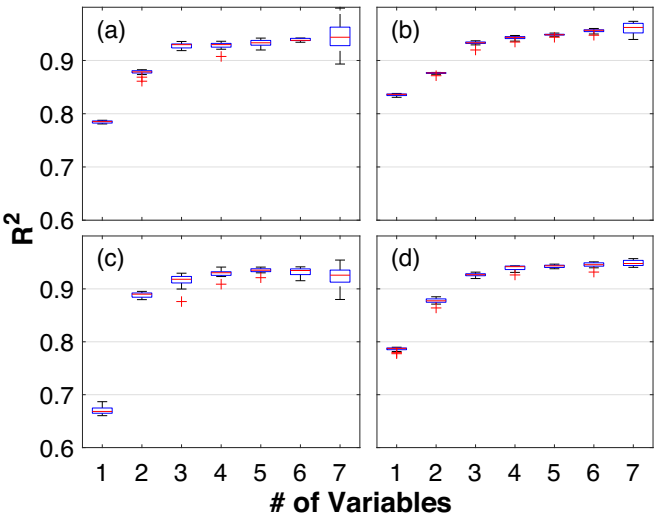


Fig. 1. R^2 validation performance for models trained and evaluated respectively on each of the four agricultural systems: (a) corn, (b) miscanthus, (c) sorghum, and (d) soybean. Results are the aggregate of 15 cross-validated iterations of the highest performing model utilizing the respective number of input predictors.

Simplicity is preferred where an acceptable trade-off for accuracy can be made. Simpler models are simultaneously easier to explain and more robust to noise than more complex models (Diaz-Ramirez et al., 2021; Kutz and Brunton, 2022; Yates et al., 2021). From the perspective of integrating machine learning into physical observing systems, a robust and computationally inexpensive model presents opportunities to extend measurement capabilities. Further, instrumentation and measurements carry costs, so simple models may lower observational costs (Yates et al., 2021) and provide greater resilience to instrument failure than models using more predictors. The reduction of input variables shifts the focus to data volume, which is more easily achievable with fewer meteorological and/or remote sensing instruments.

3.2. Impact of crop type

Among the seven candidate predictors of soil heat flux, surface temperature and air temperature achieve the highest individual R^2 predictive performance, indicating a strong physical relationship to soil heat flux variation (see Table S3). Other predictors such as NDVI and PRI, while significant in multi-predictor models (Table 1), do not perform well individually due to low variability from day to day, and the fact that they do not vary diurnally as do the radiation fluxes that provide the source of energy for soil heat fluxes. LST is a stronger predictor than air temperature for sorghum and soybean, while air temperature performs marginally better in corn and miscanthus models.

Miscanthus models consistently produce the highest predictive performance across the four crops. This is due in part to the dense canopy (high NDVI) throughout our observation period (see Fig. 2). Miscanthus is a perennial crop and had developed a closed canopy by the start of the observation period (discussed below). This is the reason that NDVI alone has the lowest predictive performance for miscanthus. The three other agro-ecosystems experience a sharp increase in growth during the early portion of the observing period, and decline at the end of the season,

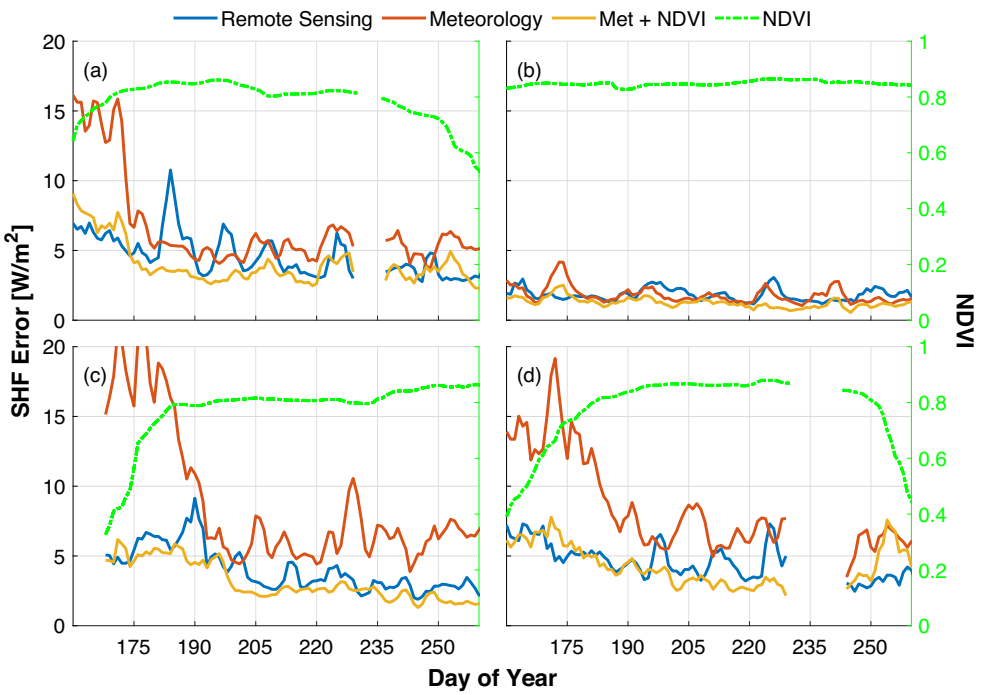


Fig. 2. Daily averaged absolute prediction error for (a) corn, (b) miscanthus, (c) sorghum, and (d) soybean across the growing season. Blue depicts the model trained in the remote sensing scenario. Orange depicts the model trained in the meteorology-only scenario. Yellow depicts the meteorology model with the addition of NDVI. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

which allows NDVI to provide important information on canopy closure and density variation for SHF estimation. See Tables S2 and S3 for further details of the relative importance of NDVI for SHF predictions across these four crops.

On average, sorghum models underperform the other crop models that utilize a single predictor. The daily soil heat flux response for sorghum does not differ significantly from corn or soybean (see Fig. S1). It may be that the soil heat flux response for sorghum is more complex, thus the relative improvement in performance with more predictors, or contains dependencies that are not captured in the predictor sets evaluated here. The sorghum and soybean systems show a significant performance increase with LST as a predictor, relative to T_a. This is not the case for the corn and miscanthus systems and may indicate that there are crop-specific traits (i.e. related to canopy structure and not solely canopy density) that impact the utility of LST in soil energy balance estimation. In Section 3.5 we examine the ability of a single ML formulation to predict SHF across all agro-ecosystems examined here (i.e. a generalized model of SHF).

3.3. Assessment of predictor source

Here we assess the performance attainable by using only measurements from one of two sources: remote sensing measurements (NDVI, PRI, and LST) and meteorological observations (LW, SW, and T_a). In addition, we highlight the potential benefit of introducing remote sensing observations into models developed using meteorological observations. In Fig. 2 NDVI is displayed as a green line and can be seen to rise and fall at the shoulders of the growing season of the three annual crops (soybean, sorghum, and corn), representing seasonal plant growth / phenology. The remote sensing model that incorporates NDVI has lower SHF prediction errors relative to the meteorology-only model throughout the season for the three annual crops. The addition of NDVI to the meteorology-only model provides a significant improvement in R^2 predictive performance in corn (0.84 → 0.90), sorghum (0.66 → 0.89), and soybean (0.76 → 0.86), and a modest performance increase in miscanthus (0.90 → 0.92) which was mature during the data collection period.

In the span of a day, soil heat flux experiences significant diurnal variation. All three models demonstrate an overestimation in the period before noon, and an underestimation in the period after noon (Fig. 3). This behavior is more prominent in the remote sensing model, for which the only source of information on diurnal variations is contained in LST, as NDVI and PRI are considered here as daily averages. Due to the proximity of the four towers, each station recorded similar meteorological observations. However, the meteorology-only models show different patterns of error due to different learned patterns for each system, and possibly subtle variations in meteorological variables directly over each canopy or differences in sensor noise across observation sites. These differences may also be due to variations in canopy structure and density that impact the relationships between canopy-top meteorology and SHF. The meteorology model for sorghum (Fig. 3c) shows a sharp transition from overestimation to underestimation near 14:00, which becomes less significant in the remote sensing model or the meteorology model once NDVI is included as a predictor.

Meteorological variables alone are much less capable of capturing variation in the canopy structure above the soil throughout the season. This in part explains the reliance of the semi-empirical SHF models on R_n rather than the downwelling radiation fluxes alone, as upwelling shortwave reflected and emitted thermal radiation fluxes carry some information on the land surface cover related to vegetation cover and phenology. As a result of the lack of phenological information in the meteorological variable set, the meteorology-only models tend to show an underestimation early in the growing season prior to canopy maturity and an overestimation after the canopy has developed.

The one-to-one modeled vs observed SHF performance for meteorology-only [SW, LW, T_a] models developed for each crop individually is presented in Fig. 4. The models for each crop are seen to perform well, with the greatest error occurring in the soybean and sorghum model predictions early in the growing season. The diurnal error trend is apparent in corn, soybean, and sorghum particularly early in the growing season when underestimates at high observed SHF values (midday) are balanced by overestimates for low SHF conditions in the late

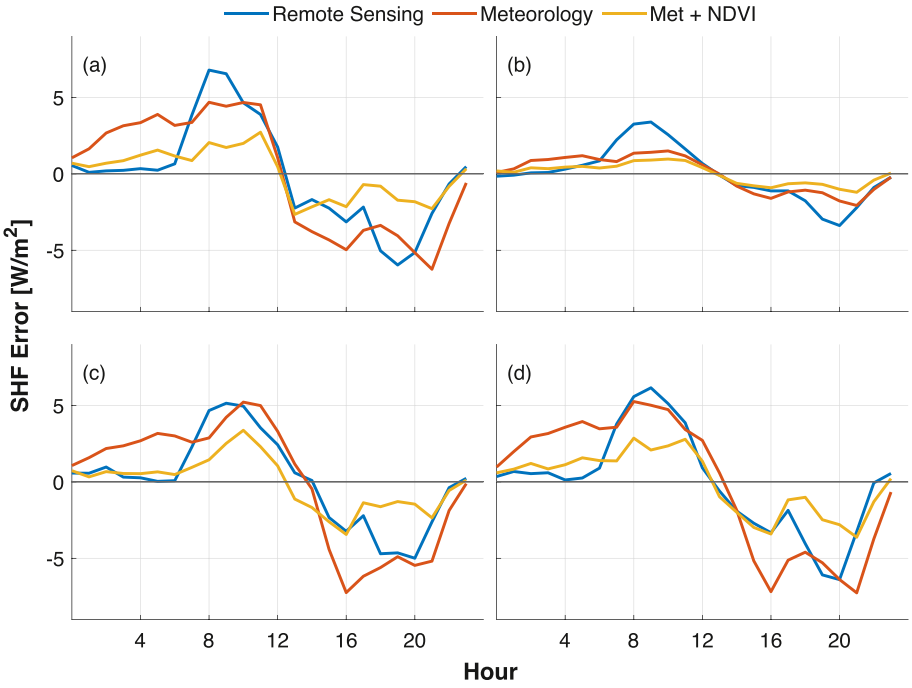


Fig. 3. Diurnal prediction error across the entire observation season for (a) corn, (b) miscanthus, (c) sorghum and (d) soybean. Blue depicts the model trained using remote sensing predictors [LST, NDVI, PRI], orange the model trained on meteorology data [SW, LW, T_a] and yellow the meteorology predictor model that includes NDVI [NDVI, SW, LW, T_a]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

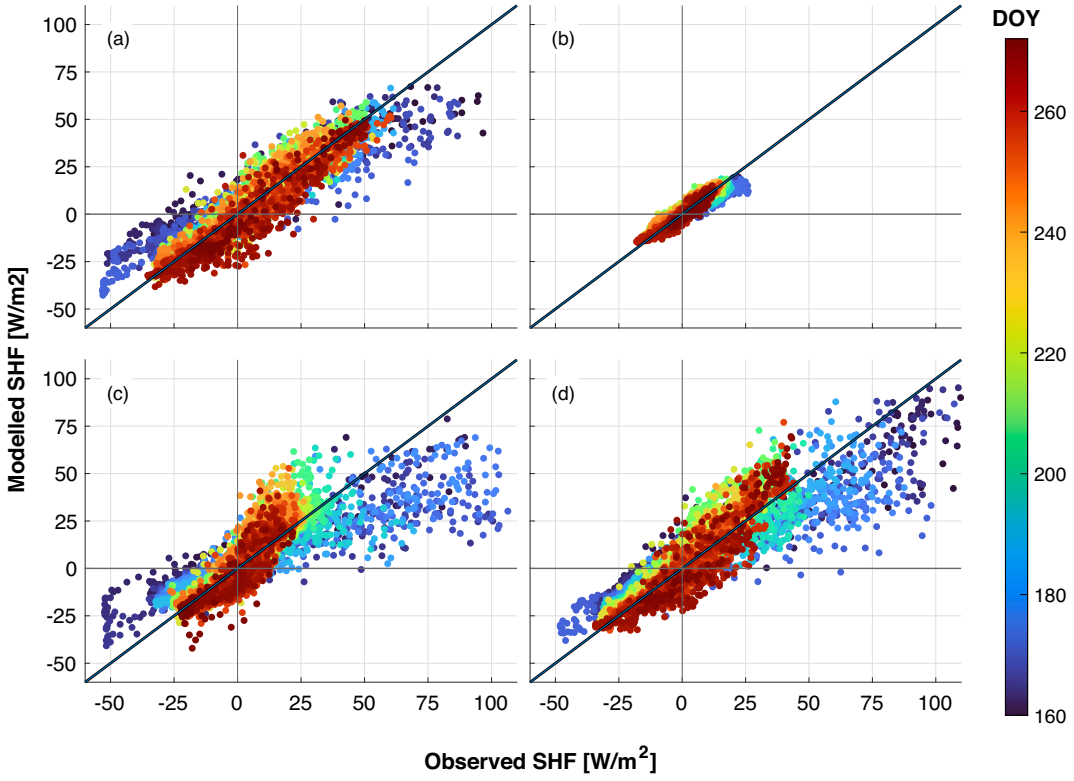


Fig. 4. Validation performance for models trained individually for (a) corn, (b) miscanthus, (c) sorghum and (d) soybean on the three meteorological predictors [T_a, SW, LW]. Each point is the average half-hourly prediction from 15 training iterations. Color indicates the day of year.

afternoon and nighttime. The small range of variability of the miscanthus SHF values are due to the dense canopy cover of this system throughout the measurement period, resulting in strong model performance as the phenological information carried by NDVI is not important in this context.

In contrast, the remote sensing predictor model [NDVI, PRI, LST] results (Fig. 5) demonstrate the importance of capturing canopy cover variations throughout the season. While this dataset in general has less

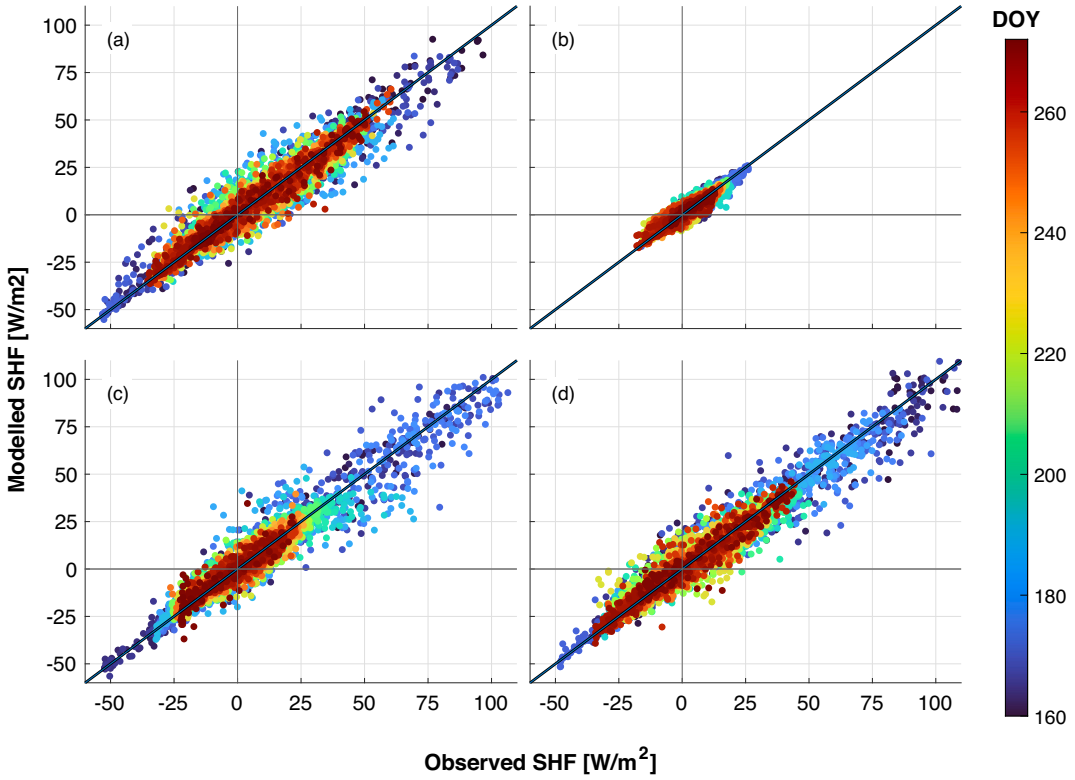


Fig. 5. Validation performance for (a) corn, (b) miscanthus, (c) sorghum and (d) soybean models trained using remote sensing predictors [LST, NDVI, PRI]. Each point is the average half-hourly prediction from 15 training iterations. Color indicates the day of year. Measurements were made from a set of radiometric sensors placed in-situ on each observing tower.

information about short-term changes in the environment (i.e. due to clouds, weather systems, etc.), the information on diurnal variability in the important drivers of SHF variability is contained in LST observations, allowing for strong predictive performance across the four agro-ecosystems. These models demonstrate unbiased results throughout the season with very high predictive performance using only three predictor variables, and only modestly under-performing models using all seven predictor variables examined here (see Table 1).

In studying soil heat flux across various time scales, it's evident that there is significant hourly, daily, and monthly variation that requires multi-source data integration to overcome. Quantifying soil heat flux at high resolution requires observations of local environmental conditions, canopy characteristics, and radiative forcings. For regional or global applications, vegetation indices acquired at a daily or weekly scale may be sufficient as they primarily govern seasonal changes in SHF (Liang et al., 2022). However, hourly or sub-hourly estimates of local conditions and forcings are necessary to minimize error within a day. In applications of evapotranspiration (ET) modelling, instantaneous midday estimates are upscaled to daily values to compensate for once-a-day satellite overpass times (Cammalleri et al., 2014). These upscaling techniques introduce several biases, in part due to improper assumptions or modelling of SHF. With the models developed here, such as the Meteorology + NDVI model, meteorological forcings in conjunction with infrequent remote sensing observations can be used to parameterize SHF models more accurately.

3.4. Model Explainability

Data-driven modelling exercises are often motivated to achieve maximum predictive performance using all available predictor variables rather than focusing on understanding the value of each predictor variable toward developing more parsimonious models. Feature selection is often performed prior to model development to reduce training costs, and data requirements (McLachlan, 2005). Model explainability is seldom considered as a trade-off for feature selection despite the diverse array of tools available. Shapley additive explanations is a post-hoc feature attribution method used to assess the relative impact of predictors on model predictions. SHAP is uniquely suited for model explainability and physical consistency, leveraging local explanations (between data points) to build a complete global explanation (between predictors) (Lundberg et al., 2018; Lundberg and Lee, 2017). SHAP values only capture the relationship between predictors and the model output and should not be used to assess causality without careful consideration of uncaptured latent variables and redundancy (Baptista et al., 2022). The analysis presented here utilizes domain knowledge to examine the principles that have been learned by the models, while keeping in mind the limitations of this technique.

Beeswarm plots (sometimes called waterfall plots) are a commonly used tool to visualize the relative importance of predictors as determined from the magnitude of their SHAP attribution (Hu et al., 2021; Zhang et al., 2023). These plots, as seen in Fig. 6 (e)-(h), show the distribution of attribution scores (x-axis) for the variables present in the model (y-axis). Vertical offsets and coloration are added to signify the density of points in the region and the relative value of each predictor respectively.

Beeswarm plots are limited by the lack of information they contain on the interactions occurring between variables that inform a given model prediction. To address this need partial dependence plots have been developed which display the impacts of multiple predictors on the model's outputs (Lundberg et al., 2020). In Fig. 6a-d we explore SHAP as a tool for visualizing the synergistic effects of NDVI with LST through a partial dependence plot of LST and NDVI to the SHAP attribution for LST. Each of the four subplot pairs (a and e; b and f; c and g; d and h) represents a single crop system trained on predictor sets of NDVI, LST, and PRI. Despite the fact that these models had similar aggregate performance (R^2 of 0.91, 0.91, 0.91, and 0.90 for corn, miscanthus,

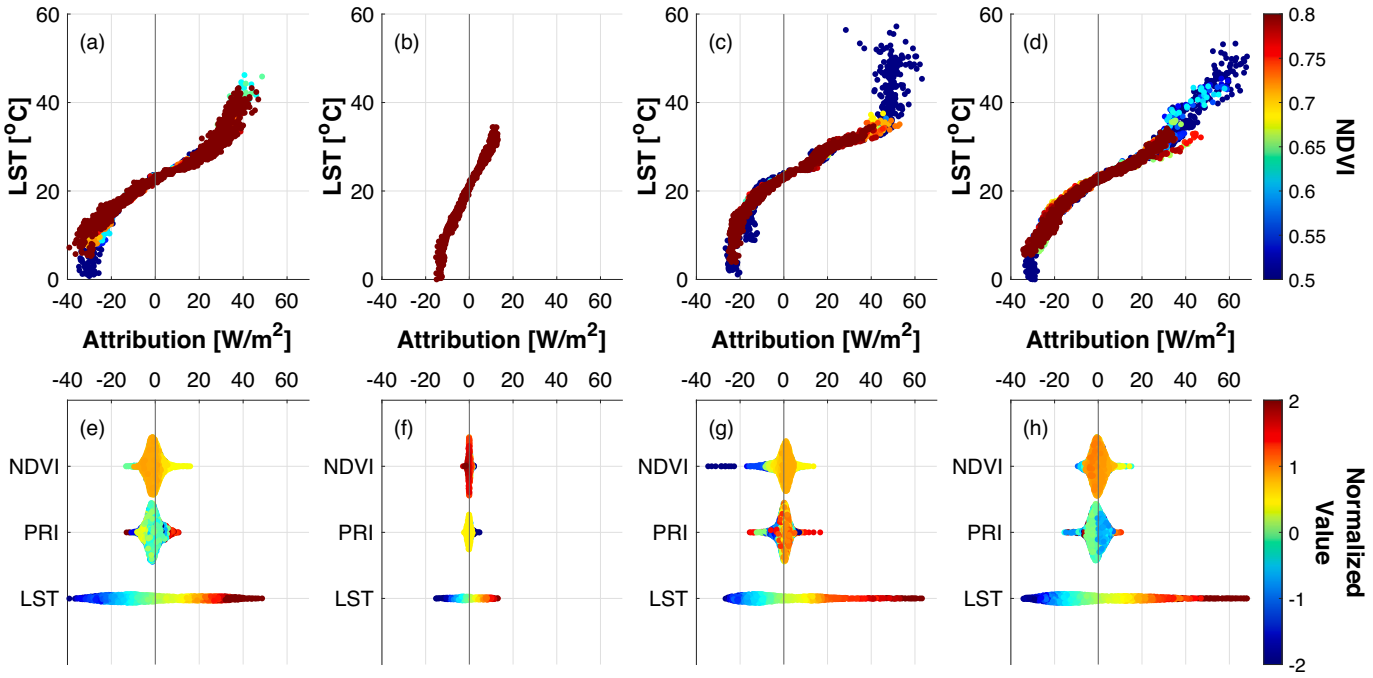


Fig. 6. (a-d) Scatter plots of LST vs SHAP attribution for LST, colored by NDVI value. (e-h) Swarm plots of SHAP attribution for LST, NDVI, and PRI from models trained individually on (a.e) corn, (b,f) miscanthus, (c,g) sorghum and (d,h) soybean.

sorghum, and soybean, respectively), the SHAP values show that the models have learned significantly different relationships between predictors and model outputs.

In the sorghum system (Fig. 6c,g) LST is the dominant predictor and NDVI can be seen as a factor that modifies the relationship of LST to SHAP attribution and thus SHF. Over a fully developed canopy, with (red) saturated NDVI values, the slope of the LST to attribution curve steepens and the distribution of points narrows, indicating greater certainty in the attribution to LST. When the canopy is less dense (low NDVI values, blue dots) the attribution varies more and deviates from the typical relationship at extreme LST values. These values correspond with the beginning of the growing season before the canopy was well

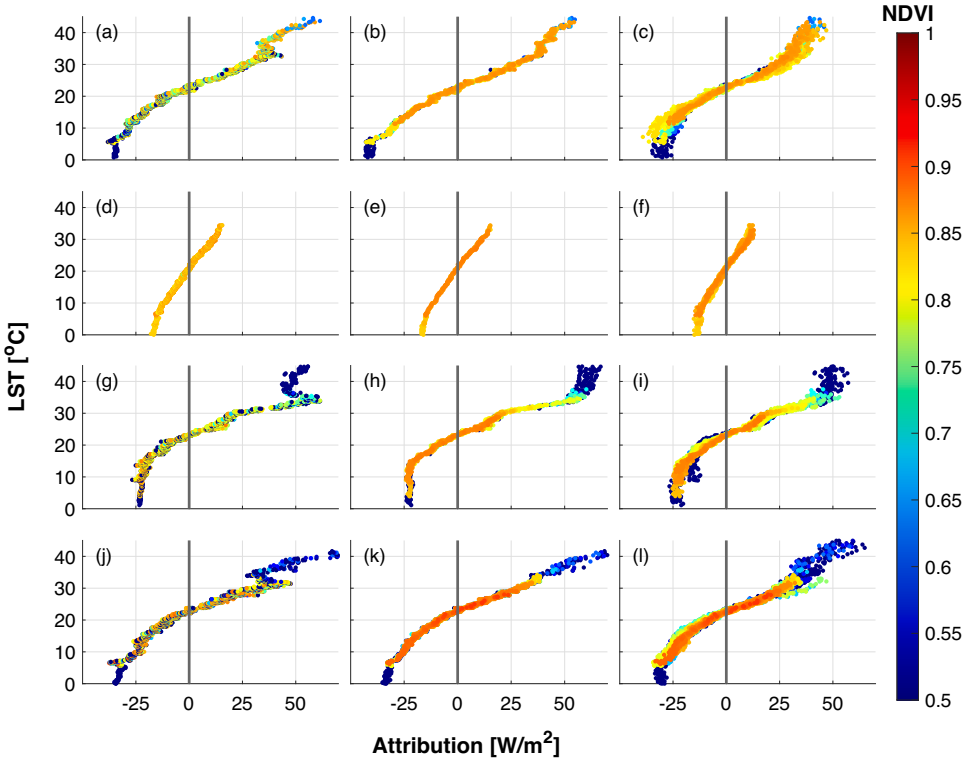


Fig. 7. SHAP attribution scatter plots for models trained on LST (a,d,g,j); NDVI and LST (b,e,h,k); NDVI, PRI, and LST (c,f,i,l). Points are positioned by measured LST value (x-axis) against SHAP model attribution to LST (y-axis), colored by NDVI value. Rows correspond to (a-c) corn, (d-f) miscanthus, (g-i) sorghum and (j-l) soybean.

developed and represent a small portion of the dataset. Fig. 6g shows that there is a minimal direct SHAP contribution for most values of NDVI and PRI. Despite this, their addition in the model increases model performance significantly through modification of the LST attribution curve, resulting in predictive performance increases of: R^2 of 0.69 → 0.88 → 0.91 for [NDVI]; [NDVI, LST]; [NDVI, PRI, LST] respectively (see Fig. 7 for a presentation of these attribution curves). Through analyzing these interactions, it becomes possible to assess model sensitivity to specific conditions and the nature of synergistic predictor effects.

The relationship between NDVI and LST differs strongly for the miscanthus SHF model relative to the other three agricultural systems. Miscanthus had developed a mature canopy by the start of data collection and therefore showed very little variation in NDVI throughout the observation period. This produced very little data from which the model could learn interaction effects between LST and NDVI, effectively making all but the highest values of NDVI unseen for the miscanthus model. The three other crops have LST attribution curves that share a sigmoidal shape spanning the range of LST values that is influenced by the information on canopy density. LST values for miscanthus do not reach the magnitudes seen in the other three crops, and the shape of the LST attribution curve is steeper, more linear and shows less spread. Fig. 6(f) shows that both NDVI and PRI have low attribution and impact on miscanthus model predictions, which is supported in the attribution curves of the reduced order miscanthus models in Fig. 7, which show little variation as NDVI and PRI are added to LST as model predictors. While the [LST, NDVI, PRI] models for the four crops have very similar performance (R^2 between 0.90 and 0.91), the data in Fig. 6 show contrasting model attributions to these predictors, indicating that it may be challenging to accurately model all four systems simultaneously with a single (generalized) model developed using combined data from all four systems and these three predictors (evaluated further in Section 3.5).

As model complexity (number of predictor variables) increases so does the potential for redundancy, whereby adding predictors may primarily increase model complexity without improvement in performance. This has the potential to lead the model to exploit non-physical patterns, such as sensor noise, reducing interpretability. The ability to infer relationships between input variables also becomes significantly more difficult as models become more complex. Importantly our results demonstrate that strong SHF predictive performance is possible with only one or two carefully chosen predictor variables (see Table 1 and Table S2).

3.5. Generalizing model findings

The models and findings discussed thus far have focused on models developed for individual agro-ecosystems that have unique phenological and canopy structural properties. If models like this are to be applied at large scale, data on crop type may not be available, making it difficult to choose a crop-specific model to apply. Here we examine the extent to which models such as these can be developed to predict SHF across multiple distinct agricultural systems, given widely available remote sensing observations and environmental variables. Here we have repeated the training procedure described in Section 2.7 for the combined data from all four study agro-ecosystems to quantify how predictive performance is impacted by exposing the training data to a mixture of agricultural systems. The results of these "generalized" models, determined by evaluation on withheld data for each crop, are presented in Table 2.

We note that for all ML models using between one and seven predictors the performance is always better than that found for any of the six empirical models examined here (see Table S1 and Figs. S3 and S4). We observed that the generalized models achieved a similar performance overall, but that performance on individual crop datasets is reduced relative to models trained on each agricultural system individually (see Fig. 8 and Fig. S5). This reduction in performance is an expected consequence of the development of a model using data across multiple systems that may have fundamentally different relationships between predictor variables and soil heat flux, and for which data was collected during different growth / phenological stages.

Table 2
R^2 validation performance from models trained on the combined dataset spanning all four agro-ecosystems. Results are averaged across 15 randomized trials of a 10 k-fold cross-validation. For the One and Two Predictor categories the predictor sets were chosen as the ones that demonstrated the highest average performance across the four agro-ecosystems.

Model Category	Corn	Miscanthus	Sorghum	Soybean
One Predictor [LST]	0.732 ± 0.007	0.664 ± 0.007	0.656 ± 0.003	0.732 ± 0.005
Two Predictor [PRI, LST]	0.786 ± 0.031	0.524 ± 0.020	0.707 ± 0.019	0.731 ± 0.020
NDVI + LST [NDVI, LST]	0.780 ± 0.023	0.587 ± 0.022	0.743 ± 0.012	0.762 ± 0.021
Remote Sensing [NDVI, PRI, LST]	0.845 ± 0.034	0.556 ± 0.032	0.764 ± 0.016	0.751 ± 0.020
Meteorology [LW, SW, T_a]	0.844 ± 0.009	0.700 ± 0.006	0.713 ± 0.027	0.786 ± 0.017
Meteorology + NDVI [LW, SW, T_a, NDVI]	0.847 ± 0.031	0.724 ± 0.009	0.805 ± 0.021	0.853 ± 0.022
Seven Predictor (Complete Set)	0.861 ± 0.016	0.660 ± 0.046	0.780 ± 0.016	0.817 ± 0.016

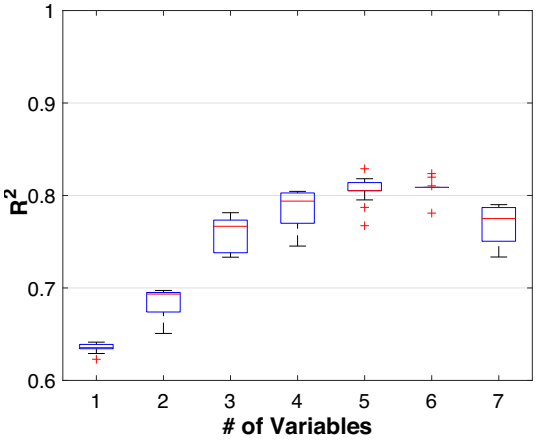


Fig. 8. R^2 validation performance for "generalized" models trained and evaluated on a combined dataset from corn, miscanthus, sorghum, and soybean. Results are the aggregate of a set of cross-validated iterations of the highest performing model utilizing the specified number of input predictors.

Among the four systems, corn and soybean, which are similar in terms of diurnal variability in SHF (see Fig. S1), exhibit the highest predictive performance for the generalized models. This difference between the very strong performance of the generalized models for corn and soybean, and somewhat reduced performance for sorghum and miscanthus, may be due to similarities between corn and soybean SHF and meteorological and canopy dynamics. These similarities could effectively bias the training dataset toward these two similar crops, reducing performance for the other two crops which diverge in these relationships. The generalized models perform least well for miscanthus, which is unique relative to the other crops examined here as it is perennial and data collection began after the canopy had closed resulting in a diminished importance for variables such as NDVI relative to the other crops. Less than 9% of SHF variability is captured by NDVI for miscanthus, relative to 20–37% for the other crops (see Table S4 which presents performance of all One Predictor models developed for each crop).

In the Meteorology + NDVI set of models, the discrepancy in Miscanthus performance relative to the same set of predictors trained individually on Miscanthus indicates limitations of NDVI as a predictor.

It is well known that NDVI becomes less sensitive to changes in leaf area as the canopy becomes more dense (Bajocco et al., 2022) in addition to changes due to different stages of growth (Qiao et al., 2019). As a result, differing canopies can present similar NDVI values while expressing different controls on SHF. Future development of mechanistic SHF models should consider approaches to disentangle biophysical traits contained in NDVI or similar metrics characterizing canopy structural traits.

While performance generally increases with the number of predictor variables, a degradation of performance is seen when using the complete predictor set. Although the complete predictor set scores the highest training performance, it begins to suffer performance degradation on individual crop types providing further motivation for selecting a reduced number of parsimonious predictors over absolute predictive performance. Examining the performance comparisons in Table 2, the 4-predictor (Meteorology + NDVI) model may be an optimal choice. This model is able to capture 70–85% of the SHF variability for the four agro-ecosystems, representing a parsimonious choice balancing strong predictive performance with a small number of widely available predictors. This 4-predictor model yields the highest R^2 performance for three out of the four systems, higher than the model developed with all seven predictor variables. The predictor set includes only NDVI as a crop-specific observation, which encourages the model to develop general relationships between overlying meteorological conditions (T_a, LW, and SW) and SHF.

4. Conclusions

Accurate soil heat flux estimation is vital to improving our understanding and predictive ability for subsurface energy transport impacting soil energy storage, temperature and biogeochemical cycling, as well as being an important component of the surface energy balance. Here we demonstrate the power of machine learning to produce accurate non-parametric predictions of SHF for multiple agro-ecosystems at timescales spanning sub-hourly to seasonal. These models demonstrated significant improvements over standard empirical formulations that have been developed around widely available satellite data. This remained true when ML models were generalized across all four agro-ecosystems (miscanthus, energy sorghum, soybean, and corn), such that a single model was used to predict SHF for all crop types, demonstrating the flexibility and robustness of well-validated ML approaches.

Air temperature and land surface temperature were the two strongest individual predictors of soil heat flux. NDVI was the most significant complementary predictor to the temperature variables, providing performance improvements in both air temperature and surface temperature models by providing information on canopy structure and phenology not available in the temperature variables. Analysis of SHAP values revealed that models developed with NDVI, PRI, and LST were capable of accurately estimating soil heat flux and describing physically meaningful relationships between model predictors and outputs. The addition of other predictor variables produced a diminished return on predictive performance while increasing data requirements and loss of model transparency and interpretability.

Predictive performance across the four crop types was comparable overall, with some specific exceptions noted in Section 3.1. In relation to previous modelling work, the utilization of simple site-level remote-sensing observations proved to be extremely useful. The three remote sensing observations [NDVI, PRI, and LST] provided a powerful predictor set in the evaluations presented here. NDVI and LST, which have a long-standing history in remote sensing, are remarkably capable of estimating soil heat flux across the four crop systems. The sub-hourly observations used in this study allowed us to develop and test models of soil heat flux at much higher temporal resolution than is typically done when the focus is on satellite observations with typical repeat frequencies of several days to weeks. These demonstrations of strong predictive performance throughout the diurnal cycle for a complete growing season across four agro-ecosystems provides confidence that these same tools and formulations will transfer to satellite applications where these predictor variables are available.

CRediT authorship contribution statement

James F. Cross: Formal analysis, Investigation, Methodology, Validation, Writing – original draft. **Darren T. Drewry:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

None.

Data availability

SABR field data can be made available by contacting SABR PIs. Machine learning models developed here can be made available by contacting the communicating author.

Acknowledgements

DTD and JFC acknowledge support from the US National Science Foundation (Awards #2239877 and 1954556). DTD also acknowledges support from the National Aeronautics and Space Administration (Award #80NSSC20K1789), the United States Department of Agriculture (Award #2023-67013-39619), the College of Food, Agricultural and Environmental Sciences and the Translational Data Analytics Institute at Ohio State University. JFC also acknowledges support from the Environmental Sciences Graduate Program 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). The authors thank Dr. Andy VanLoocke and Dr. Guler (Rojda) Aslan-Sungur for making SABR data available for this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2024.102697.

References

Allen, R.G. (Ed.), 1998. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. Food and Agriculture Organization of the United Nations.

Aslan-Sungur, G., Moore, C.E., Bernacchi, C.J., Heaton, E., VanLoocke, A., 2023. Artificial Neural Networks Estimate Evapotranspiration for Miscanthus × Giganteus as Effectively as Empirical Model But with Fewer Inputs. doi:10.2139/ssrn.4576599.

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.

Bai, Y., Bhattarai, N., Mallick, K., Zhang, S., Hu, T., Zhang, J., 2022. Thermally derived evapotranspiration from the surface temperature initiated closure (STIC) model improves cropland GPP estimates under dry conditions. Remote Sens. Environ. 271, 112901 https://doi.org/10.1016/j.rse.2022.112901.

Bajocco, S., Ginaldi, F., Savian, F., Morelli, D., Scaglione, M., Fanchini, D., et al., 2022. On the Use of NDVI to Estimate LAI in Field Crops: Implementing a Conversion Equation Library. Remote Sensing 14, 3554. https://doi.org/10.3390/rs14153554.

Banks, D.L., Fienberg, S.E., 2003. Data mining, statistics. In: Meyers, R.A. (Ed.), Encyclopedia of Physical Science and Technology, 3rd ed. Academic Press, pp. 247–261. https://doi.org/10.1016/B0-12-227410-5/00164-2.

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.

Baret, F., Buis, S., 2008. Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems. In: Liang, S. (Ed.), Advances in Land Remote Sensing. Springer, Netherlands, pp. 173–201. https://doi.org/10.1007/978-1-4020-6450-0_7.

Bastiaanssen, W.G.M., Menenti, M., Feddes, R.A., Holtslag, A.A.M., 1998. A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. J. Hydrol. 212–213, 198–212. https://doi.org/10.1016/S0022-1694(98)00253-4.

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

Bodelier, P.L.E., 2011. Toward understanding, managing, and protecting microbial ecosystems. Front. Microbiol. 2, 80. https://doi.org/10.3389/fmicb.2011.00080.

Boegh, E., Soegaard, H., Christensen, J.H., Hasager, C.B., Jensen, N.O., Nielsen, N.W., Rasmussen, M.S., 2004. Combining weather prediction and remote sensing data for the calculation of evapotranspiration rates: application to Denmark. Int. J. Remote Sens. 25 (13), 2553–2574. https://doi.org/10.1080/01431160310001647984.

Bonsoms, J., Boulet, G., 2022. Ensemble machine learning outperforms empirical equations for the ground heat flux estimation with remote sensing data. Remote Sens. 14 (8) https://doi.org/10.3390/rs14081788. Article 8.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Brown, M., 2018. Effects of soil temperature on some soil properties and plant growth. Adv. Plants Agricult. Res. 8 (1) https://doi.org/10.15406/apar.2018.08.00288.

Burke, I., Yonker, C., Parton, W., Cole, C., Flach, K., Schimel, D., 1989. Texture, climate, and cultivation effects on soil organic matter content in U.S. grassland soils. Soil Sci. Soc. Am. J. 53, 800–805. https://doi.org/10.2136/sssaj1989.03615995005300030029x.

Callaham, J.L., Koch, J.V., Brunton, B.W., Kutz, J.N., Brunton, S.L., 2021. Learning dominant physical processes with data-driven balance models. Nat. Commun. 12 (1), 1016. https://doi.org/10.1038/s41467-021-21331-z.

Cammalleri, C., Anderson, M.C., Kustas, W.P., 2014. Upscaling of evapotranspiration fluxes from instantaneous to daytime scales for thermal remote sensing applications. Hydrology and Earth System Sciences 18, 1885–1894. https://doi.org/10.5194/hess-18-1885-2014.

Canelón, D.J., Chávez, J.L., 2011. Soil heat flux modeling using artificial neural networks and multispectral airborne remote sensing imagery. Remote Sens. 3 (8) https://doi.org/10.3390/rs3081627. Article 8.

Carpita, N.C., McCann, M.C., 2008. Maize and sorghum: genetic resources for bioenergy grasses. Trends Plant Sci. 13 (8), 415–420. https://doi.org/10.1016/j.tplants.2008.06.002.

Cawley, G., Talbot, N., 2010. On over-fitting in model selection and subsequent selection Bias in performance evaluation. J. Mach. Learn. Res. 11, 2079–2107.

Chan, J.C.-W., Paelinckx, D., 2008. Evaluation of random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sens. Environ. 112 (6), 2999–3011. https://doi.org/10.1016/j.rse.2008.02.011.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comp. Sci. 7, e623 https://doi.org/10.7717/peerj-cs.623.

Choudhury, BhaskarJ, 1999. Evaluation of an empirical equation for annual evaporation using field observations and results from a biophysical model. J. Hydrol. 216 (1), 99–110. https://doi.org/10.1016/S0022-1694(98)00293-5.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Quéré, C.L., Myneni, R., Piao, S., Thornton, P., Metzl, N., Wania, R., 2014. Carbon and Other Biogeochemical Cycles. Cambridge University Press, Cambridge, United Kingdom. https://doi.org/10.1017/CBO9781107415324.015.

Clothier, B.E., Clawson, K.L., Pinter, P.J., Moran, M.S., Reginato, R.J., Jackson, R.D., 1986. Estimation of soil heat flux from net radiation during the growth of alfalfa. Agric. For. Meteorol. 37 (4), 319–329. https://doi.org/10.1016/0168-1923(86)90069-9.

de Andrade, B.C.C., Pedrollo, O.C., Ruhoff, A., Moreira, A.A., Laipelt, L., Kayser, R.B., Biudes, M.S., dos Santos, C.A.C., Roberti, D.R., Machado, N.G., Dalmagro, H.J., Antonino, A.C.D., de Lima, J.R.S., de Souza, E.S., Souza, R., 2021. Artificial neural network model of soil heat flux over multiple land covers in South America. Remote Sens. 13 (12), 2337. https://doi.org/10.3390/rs13122337.

de Silva, B.M., Higdon, D.M., Brunton, S.L., Kutz, J.N., 2020. Discovery of physics from data: universal Laws and Discrepancies. Front. Artif. Intell. 3, 25. https://doi.org/10.3389/frai.2020.00025.

Dessureault-Rompré, J., Zebarth, B.J., Georgallas, A., Burton, D.L., Grant, C.A., Drury, C. F., 2010. Temperature dependence of soil nitrogen mineralization rate: comparison of mathematical models, reference temperatures and origin of the soils. Geoderma 157 (3), 97–108. https://doi.org/10.1016/j.geoderma.2010.04.001.

Diaz-Ramirez, L.G., Lee, S.J., Smith, A.K., Gan, S., Boscardin, W.J., 2021. A novel method for identifying a parsimonious and accurate predictive model for multiple clinical outcomes. Comput. Methods Prog. Biomed. 204, 106073 https://doi.org/10.1016/j.cmpb.2021.106073.

Drewry, D.T., Kumar, P., Long, S., Bernacchi, C., Liang, X.-Z., Sivapalan, M., 2010a. Ecohydrological responses of dense canopies to environmental variability: 1. Interplay between vertical structure and photosynthetic pathway. J. Geophys. Res. 115 (G4), G04022. https://doi.org/10.1029/2010JG001340.

Drewry, D.T., Kumar, P., Long, S., Bernacchi, C., Liang, X.-Z., Sivapalan, M., 2010b. Ecohydrological responses of dense canopies to environmental variability: 2. Role of acclimation under elevated CO_2. J. Geophys. Res. 115 (G4), G04023. https://doi.org/10.1029/2010JG001341.

Evett, S.R., Agam, N., Kustas, W.P., Colaizzi, P.D., Schwartz, R.C., 2012. Soil profile method for soil thermal diffusivity, conductivity and heat flux: comparison to soil heat flux plates. Adv. Water Resour. 50, 41–54. https://doi.org/10.1016/j.advwatres.2012.04.012.

Franzke, C.L.E., Barbosa, S., Blender, R., Fredriksen, H.-B., Laepple, T., Lambert, F., Nilsen, T., Rypdal, K., Rypdal, M., Scotto, M.G., Vannitsem, S., Watkins, N.W., Yang, L., Yuan, N., 2020. The structure of climate variability across scales. Rev. Geophys. 58 (2), e2019RG000657 https://doi.org/10.1029/2019RG000657.

Fuchs, M., 1986. Heat Flux," in *Methods of Soil Analysis. Part 1. Physical and mineralogical me*. John Wiley & Sons, Ltd, Madison, WI, pp. 957–968. https://doi.org/10.2136/sssabookser5.1.2ed.c40.

Gamon, J., Serrano, L., Surfus, J.S., 1997. The photochemical reflectance index: an optical Indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. Oecologia 112, 492–501. https://doi.org/10.1007/s004420050337.

Gao, Z., Russell, E.S., Missik, J.E.C., Huang, M., Chen, X., Strickland, C.E., Clayton, R., Arntzen, E., Ma, Y., Liu, H., 2017. A novel approach to evaluate soil heat flux calculation: an analytical review of nine methods. J. Geophys. Res. Atmos. 122 (13), 6934–6949. https://doi.org/10.1002/2017JD027160.

Garcia-Santos, V., Sánchez-Tomás, J., Cuxart, J., 2022. Evapotranspiration acquired with remote sensing thermal-based algorithms: a state-of-the-art review. Remote Sens. 14, 3440. https://doi.org/10.3390/rs14143440.

Gavito, M.E., Curtis, P.S., Mikkelsen, T.N., Jakobsen, I., 2001. Interactive effects of soil temperature, atmospheric carbon dioxide and soil N on root development, biomass and nutrient uptake of winter wheat during vegetative growth. J. Exp. Bot. 52 (362), 1913–1923. https://doi.org/10.1093/jexbot/52.362.1913.

Gewali, U.B., Monteiro, S.T., Saber, E., 2019. Machine Learning Based Hyperspectral Image Analysis: A Survey arXiv:1802.08701 [Cs, Eess]. Retrieved from http://arxiv.org/abs/1802.08701.

Gupta, D., Gujre, N., Singha, S., Mitra, S., 2022. Role of existing and emerging technologies in advancing climate-smart agriculture through modeling: a review. Eco. Inform. 71, 101805 https://doi.org/10.1016/j.ecoinf.2022.101805.

Hu, X., Shi, L., Lin, G., Lin, L., 2021. Comparison of physical-based, data-driven and hybrid modeling approaches for evapotranspiration estimation. J. Hydrol. 601, 126592 https://doi.org/10.1016/j.jhydrol.2021.126592.

Idso, S.B., Aase, J.K., Jackson, R.D., 1975. Net radiation—soil heat flux relations as influenced by soil water content variations. Bound.-Layer Meteorol. 9 (1), 113–122. https://doi.org/10.1007/BF00232257.

Jolliffe, I.T., 1986. Principal Component Analysis. Springer-Verlag. http://www.gbv.de/dms/hbz/toc/ht002702574.pdf.

Kaiser, B.E., Saenz, J.A., Sonnewald, M., Livescu, D., 2022. Automated identification of dominant physical processes. Eng. Appl. Artif. Intell. 116, 105496 https://doi.org/10.1016/j.engappai.2022.105496.

Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: a survey. Comput. Electron. Agric. 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016.

Kim, J.-H., 2009. Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap. Comp. Stat. Data Anal. 53 (11), 3735–3745. https://doi.org/10.1016/j.csda.2009.04.009.

Kustas, W.P., Daughtry, 1990. Estimation of the soil heat flux/net radiation ratio from spectral data. Agricultural and Forest Meteorology 49, 205–223. https://doi.org/10.1016/0168-1923(90)90033-3.

Kustas, W.P., Daughtry, C.S.T., Van Oevelen, P.J., 1993. Analytical treatment of the relationships between soil heat flux/net radiation ratio and vegetation indices. Remote Sens. Environ. 46 (3), 319–330. https://doi.org/10.1016/0034-4257(93)90052-Y.

Kutz, J.N., Brunton, S.L., 2022. Parsimony as the ultimate regularizer for physics-informed machine learning. Nonlinear Dynam. 107 (3), 1801–1817. https://doi.org/10.1007/s11071-021-07118-3.

Lettau, H., 1954. Improved models of thermal diffusion in the soil. EOS Trans. Am. Geophys. Union 35 (1), 121–132. https://doi.org/10.1029/TR035i001p00121.

Liang, A., Xie, C., Wang, J., Che, S., 2022. Daily Dynamics of Soil Heat Flux and Its Relationship with Net Radiation in Different Urban Riparian Woodlands. Forests 13, 2062. https://doi.org/10.3390/f13122062.

Liebethal, C., Foken, T., 2007. Evaluation of six parameterization approaches for the ground heat flux. Theor. Appl. Climatol. 88 (1), 43–56. https://doi.org/10.1007/s00704-005-0234-0.

Liebethal, C., Huwe, B., Foken, T., 2005. Sensitivity analysis for two ground heat flux calculation approaches. Agricultural and Forest Meteorology 132, 253–262. https://doi.org/10.1016/j.agrformet.2005.08.001.

Liu, Z., 2022. Accuracy of five ground heat flux empirical simulation methods in the surface-energy-balance-based remote-sensing evapotranspiration models. Hydrol. Earth Syst. Sci. 26, 6207–6226. https://doi.org/10.5194/hess-26-6207-2022.

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, pp. 4768–4777.

Lundberg, S., Erion, G., Lee, S.-I., 2018. Consistent Individualized Feature Attribution for Tree Ensembles.

Liu, X., Yang, S., Xu, J., Zhang, J., Liu, J., 2017. Effects of soil heat storage and phase shift correction on energy balance closure of paddy fields. Atmósfera 30, 39–52. https://doi.org/10.20937/ATM.2017.30.01.04.

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. *Nature*. Machine Intell. 2(1), Article 1 https://doi.org/10.1038/s42256-019-0138-9.

Mandal, S., Vema, V.K., Kurian, C., Sudheer, K., 2020. Improving the crop productivity in rainfed areas with water harvesting structures and deficit irrigation strategies. J. Hydrol. 586, 124818 https://doi.org/10.1016/j.jhydrol.2020.124818.

Massman, W.J., 1992. Correcting errors associated with soil heat flux measurements and estimating soil thermal properties from soil temperature and heat flux plate data. Agric. For. Meteorol. 59, 249–266. https://doi.org/10.1016/0168-1923(92)90096-M.

Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. Int. J. Remote Sens. 39 (9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343.
McLachlan, G.J., 2005. Discriminant Analysis and Statistical Pattern Recognition. John Wiley & Sons.
Miller, K.S., Geisseler, D., 2018. Temperature sensitivity of nitrogen mineralization in agricultural soils. Biol. Fertil. Soils 54 (7), 853–860. https://doi.org/10.1007/s00374-018-1309-2.
Molinaro, A.M., Simon, R., Pfeiffer, R.M., 2005. Prediction error estimation: a comparison of resampling methods. Bioinformatics 21 (15), 3301–3307. https://doi.org/10.1093/bioinformatics/bti499.
Moore, C.E., von Haden, A.C., Burnham, M.B., Kantola, I.B., Gibson, C.D., Blakely, B.J., Dracup, E.C., Masters, M.D., Yang, W.H., DeLucia, E.H., Bernacchi, C.J., 2021. Ecosystem-scale biogeochemical fluxes from three bioenergy crop candidates: how energy sorghum compares to maize and miscanthus. GCB Bioenergy 13 (3), 445–458. https://doi.org/10.1111/gcbb.12788.
Oliver, S.A., Oliver, H.R., Wallace, J.S., Roberts, A.M., 1987. Soil heat flux and temperature variation with vegetation, soil type and climate. Agric. For. Meteorol. 39 (2), 257–269. https://doi.org/10.1016/0168-1923(87)90042-6.
Pechanec, V., Purkyt, J., Benc, A., Nwaogu, C., Štěrbová, L., Cudlín, P., 2018. Modelling of the carbon sequestration and its prediction under climate change. Eco. Inform. 47, 50–54. https://doi.org/10.1016/j.ecoinf.2017.08.006.
Pregitzer, K.S., King, J.S., 2005. Effects of soil temperature on nutrient uptake. In: BassiriRad, H. (Ed.), Nutrient Acquisition by Plants: An Ecological Perspective. Springer, pp. 277–310. https://doi.org/10.1007/3-540-27675-0_10.
Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Weather Rev. 100 (2), 81–92. https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2.
Purdy, A., Fisher, J., Goulden, M., Famiglietti, J., 2016. Ground heat flux: an analytical review of 6 models evaluated at 88 sites and globally. J. Geophys. Res. Biogeosci. 121 https://doi.org/10.1002/2016JG003591.
Qiao, K., Zhu, W., Xie, Z., Li, P., 2019. Estimating the Seasonal Dynamics of the Leaf Area Index Using Piecewise LAI-VI Relationships Based on Phenophases. Remote Sensing 11, 689. https://doi.org/10.3390/rs11060689.
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.
Santanello, J.A., Friedl, M.A., 2003. Diurnal covariation in soil heat flux and net radiation. J. Appl. Meteorol. Climatol. 42 (6), 851–862. https://doi.org/10.1175/1520-0450(2003)042<0851:DCISHF>2.0.CO;2.
Sauer, T.J., Horton, R., 2005. Soil Heat Flux. Micrometeorology in Agricultural Systems, (American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America) 131–154. https://doi.org/10.2134/agronmonogr47.c7.
Sauer, T.J., Peng, X., 2020. Soil temperature and heat flux. In: Agroclimatology. John Wiley & Sons, Ltd., pp. 73–93. https://doi.org/10.2134/agronmonogr60.2016.0024
Scheda, R., Diciotti, S., 2022. Explanations of machine learning models in repeated nested cross-validation: an application in age prediction using brain complexity features. Appl. Sci. 12 (13), Article 13. https://doi.org/10.3390/app12136681.
Schimel, D., Braswell, B., Holland, E., McKeown, R., Ojima, D., Painter, T., Parton, W., Townsend, A., 1994. Climatic, edaphic, and biotic controls over storage and turnover of carbon in soils. Global Biogeochem. Cycles. 8, 279–294. https://doi.org/10.1029/94GB00993.
Schlesinger, W.H., 1997. Biogeochemistry: An Analysis of Global Change, 1st Edition. Academic Press. https://doi.org/10.1016/B978-0-12-625157-9.50002-2.
Thessen, G., Maliszewski, S., Rothstein, C., 2022, October. 2022-Iowa-Annual-Bulletin.pdf. In: 2022 Iowa Agricultural Statistics. https://www.nass.usda.gov/Statistics_by_State/Iowa/Publications/Annual_Statistical_Bulletin/2022-Iowa-Annual-Bulletin.pdf.
Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8 (2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0.
van der Putten, W.H., Bardgett, R.D., Bever, J.D., Bezemer, T.M., Casper, B.B., Fukami, T., Kardol, P., Klironomos, J.N., Kulmatiski, A., Schweitzer, J.A., Suding, K.N., Van de Voorde, T.F.J., Wardle, D.A., 2013. Plant–soil feedbacks: the past, the present and future challenges. J. Ecol. 101 (2), 265–276. https://doi.org/10.1111/1365-2745.12054.
Venegas, P., Grandón, A., Jara, J., Paredes, J., 2013. Hourly estimation of soil heat flux density at the soil surface with three models and two field methods. Theor. Appl. Climatol. 112 (1), 45–59. https://doi.org/10.1007/s00704-012-0705-z.
Vermerris, W. (Ed.), 2008. Genetic Improvement of Bioenergy Crops. Springer, New York. https://doi.org/10.1007/978-0-387-70805-8.
Wang, J., Bras, R.L., 1999. Ground heat flux estimated from surface soil temperature. J. Hydrol. 216 (3), 214–226. https://doi.org/10.1016/S0022-1694(99)00008-6.
Wang, J., Wei, J.-M., Yang, Z., Wang, S.-Q., 2017. Feature selection by maximizing independent classification information. IEEE Trans. Knowl. Data Eng. 29 (4), 828–841. https://doi.org/10.1109/TKDE.2017.2650906.
Yao, W., Han, M., Xu, S., 2010. Estimating the regional evapotranspiration in Zhalong wetland with the two-source energy balance (TSEB) model and Landsat7/ETM+ images. Eco. Inform. 5 (5), 348–358. https://doi.org/10.1016/j.ecoinf.2010.06.002.
Yates, L.A., Richards, S.A., Brook, B.W., 2021. Parsimonious model selection using information theory: a modified selection rule. Ecology 102 (10), 3475. https://doi.org/10.1002/ecy.3475.
Yates, L.A., Aandahl, Z., Richards, S.A., Brook, B.W., 2022. Cross validation for model selection: a review with examples from ecology. Ecol. Monogr. 1557 https://doi.org/10.1002/ecm.1557.
Zhai, Y., Zhai, G., Chen, Y., Liu, J., 2024. Research on regional terrestrial carbon storage based on the pattern-process-function. Eco. Inform. 80, 102523 https://doi.org/10.1016/j.ecoinf.2024.102523.
Zhang, Y., Li, C., Duan, H., Yan, K., Wang, J., Wang, W., 2023. Deep learning based data-driven model for detecting time-delay water quality indicators of wastewater treatment plant influent. Chem. Eng. J. 467, 143483 https://doi.org/10.1016/j.cej.2023.143483.
Zheng, C., Jia, L., 2022. Evaluation of different methods for soil heat flux estimation at large scales using remote sensing observations. In: IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, pp. 6081–6084. https://doi.org/10.1109/IGARSS46834.2022.9883851.