Journal of Hydrology 600 (2021) 126579

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Evaluation of prediction and forecasting models for evapotranspiration of agricultural lands in the Midwest U.S

Ammara Talib [a,*], Ankur R. Desai [b,1], Jingyi Huang [c,2], Tim J. Griffis [d,3], David E. Reed [e,4], Jiquan Chen [f]

[a] Dept of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA
[b] Dept of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA
[c] Dept of Soil Science, University of Wisconsin-Madison, Madison, WI 53706, USA
[d] Dept. of Soil, Water, and Climate, University of Minnesota Saint Paul, MN 55108, USA
[e] University of Science and Arts of Oklahoma, Environmental Science, Chickasha, OK 73018, USA
[f] Michigan Technological University, School of Forestry and Wood Products, Houghton, MI 49931, USA

ARTICLE INFO

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Qiuhua Liang, Associate Editor

Keywords:
Evapotranspiration
Machine learning
Agriculture
Drought
Irrigation
Forecasting

ABSTRACT

Evapotranspiration (ET) prediction and forecasting play a vital role in improving water use in agriculturally intensive areas. Metrological and biophysical predictors that drive ET in managed landscapes have complex nonlinear relationships. Deep learning and data-driven methods have shown promising performance for identifying the dependencies among variables. Here, we evaluated the potentials of random forest (RF) and long short-term memory (LSTM) neural networks to estimate and forecast daily ET for corn, soybeans, and potatoes in diverse agricultural farms during 2003–2019. The modeling framework was applied for nineteen fields where eddy covariance ET and meteorological observations in the Midwest USA for growing season (April-October) is available. In this study, we applied data-driven models (RF and LSTM) with 3 sets of predictors (5, 11, and 16 predictors). Results show that a 16 predictor RF model (RF_16 R^2 = 0.7, Willmott's skill score = 0.90) outperformed a process-based land surface model (LSM R^2 = 0.57, Willmott's skill score = 0.86) for predicting daily ET, while LSTM performance was lower (LSTM_16 R^2 = 0.65, Willmott's skill score = 0.89 and LSTM_11 R^2 = 0.62, Willmott's skill score = 0.86) than RF using the same sets of predictors. Vapor pressure and crop coefficients were identified as the most important predictors for irrigated crops, while short wave radiation and enhanced vegetation index were key predictors for non-irrigated crops. For certain crop types, such as corn and soybeans on fine-grained soils (silt loam), a simpler version RF, using only 11 drivers, can provide comparable results (R^2 = 0.70 vs 0.69 and Willmott's skill score = 0.90 vs 0.88). For short-term 3-day ET forecasting, LSTM is more sensitive to uncertainty in ensemble forecast meteorology than RF. ET forecasts were strongly sensitive to forecast uncertainty of vapor pressure. The proposed modeling architecture provides a field-scale, locally calibrated tool for accurate prediction and short-term forecasting of daily ET in areas where in situ ET, metrological, and biophysical data are lacking.

1. Introduction

Terrestrial water in the biosphere and atmosphere is linked through evapotranspiration (ET) (Donohue et al., 2010; Priestley and Taylor, 1972; Wei et al., 2017). ET is the second-largest term in the global land surface water budget (Barr et al., 2014; Narasimhan and Srinivasan, 2005; Trenberth et al., 2007; Wang and Dickinson, 2012). In order to understand terrestrial ecosystem processes in a changing climate such as flash droughts (Kim et al., 2019; Otkin et al., 2016), water resource management (e.g., irrigation efficiency), it is important to accurately

* Corresponding author.
E-mail address: talib@wisc.edu (A. Talib).
[1] ORCID ID: 0000-0002-5226-6041.
[2] ORCID ID: 0000-0002-1209-9699.
[3] ORCID ID: 0000-0002-2111-5144.
[4] ORCID ID: 0000-0002-8892-1423.

https://doi.org/10.1016/j.jhydrol.2021.126579
Received 3 January 2021; Received in revised form 3 April 2021; Accepted 14 June 2021
Available online 18 June 2021
0022-1694/Published by Elsevier B.V.

estimate and forecast ET (Allen et al., 1998; Anderson et al., 2011; Shugart, 1998). Hydrological applications geared towards conservation of water resources especially for irrigation require prediction and forecasting of ET as a fundamental component. Hence for sustainable agriculture, an ET prediction and forecasting tool can be useful for farmers and water managers to handle water resource challenges (Djaman et al., 2020; Moratiel et al., 2020; Payero and Irmak, 2013; Perera et al., 2014). Actual ET can be measured directly using eddy covariance (EC) towers (Baldocchi et al., 2001; Barr et al., 2012; Wilson et al., 2001) but costs, logistics, and measurement scale inhibit regional and long-term studies such as EC and Bowen ratio methods (Rosenberry et al., 2007). Further, ET needs to be assessed across a range of crop varieties and soil/climate types that influence it, requiring many observation sites. Hence there is a need for models that are based on more readily available drivers to predict and forecast ET for broader applications.

Data from satellite sensors have been used in earlier studies to estimate ET over domains of different regional scales such as watershed or continent (Anderson et al., 2021; Crosbie et al., 2015; Filgueiras et al., 2020; Fisher et al., 2020; Scott et al., 2008; Yao et al., 2013), though satellites are hampered by tradeoff in spatial resolution and revisit frequency, cloud cover, and model assumptions used in linking observations of surface reflectance or brightness temperatures to ET. In addition, data assimilation methods (Meng et al., 2009; Xu et al., 2018; Zou et al., 2017) as well as land surface models (Lian et al., 2018; Vinukollu et al, 2012) have been used. However, the relative error range for ET estimates compared with ground measurements is from 14% to 44% (Long et al., 2014; Velpuri et al., 2013) due to factors such as spatial variation, heterogeneity, model parametrization, and unconstrained water balance. In addition, while there are many studies to estimate or predict and forecast reference ET in different climatic conditions (e.g., Fang et al., 2018; Li et al., 2016; Yang et al., 2006) there are not many studies for forecasting actual ET in intensively irrigated and non-irrigated areas.

Field-scale crop models are another avenue for predicting ET. Current crop models that are designed to simulate agricultural practices such as soil composition, nutrients, tilling practices, and irrigation scheduling can be coupled with computational hydrologic and land–atmosphere models (Pauwels et al., 2007). The development of these physically-based and spatially explicit representation of land surface interaction and agricultural processes at the farm scale have high computational costs (Chaney et al., 2016; Clark et al., 2017), which requires significant parameterization and tuning, subject to collection of a myriad of trait and driver datasets. Even though those models accurately simulate hydrological processes, challenges in calibrating these biophysically-based models make accurate physical process simulations at individual fields challenging. In addition, the available data for calibration and validation of these models, e.g., three-dimensional information about sub-surface heterogeneity (such as soil texture, moisture, and groundwater flow) limit the application of those models for larger areas with intensive agriculture. However, these models are useful for small-scale regional studies.

In addition to process-based hydrological models, empirical models based on statistical correlations of potential evapotranspiration with meteorological parameters have also been used (Valipour et al., 2017). Often, variables like canopy cover is used in these methods to convert potential evapotranspiration to actual evapotranspiration. The problem with such an approach is that performance may significantly depend on the estimate of canopy cover. An alternate approach to existing empirical and physical based methods is to use data-driven methods to estimate actual evapotranspiration.

A variety of data-driven models have been used in ET simulation studies (Deo and Şahin, 2015; Fang et al., 2018; Izadifar and Elshorbagy, 2010; Pandey et al., 2017). It is efficient to combine information from readily available predictors from remote sensing along with ground observation by applying machine learning (ML) methods that may be able to predict and forecast ET based on relationships between input predictors without utilizing field-based physical parameters. ML algorithms extract non-linear relationships hidden in time series or spatial data and then apply those patterns to estimate and forecast future scenarios. For example, Yang et al. (2006) and Tabari et al. (2012) used a support vector machine (SVM) approach to estimate eight-day averaged ET and reference ET respectively using ground observation and remote sensing predictors. Landeras et al. (2009) used autoregressive models to forecast weekly reference ET and Bodesheim et al. (2018) applied a regression trees based random forest (RF) approach for ET estimation. Without explicit training, RF can manage high dimensional regression problems and extract the interaction among model predictors (Auret and Aldrich, 2012; te Beest et al., 2017). Shiri (2018) used a coupled wavelet-random forest model for estimating reference ET and showed the potential of a tree-based model in terms of the accuracy of the reference ET model. The use of ensemble trees and randomization makes this approach more flexible, simple, robust and avoids overfitting by making the best use of limited data and reliable performance on both training and test data (Zhang et al., 2017; Chen et al., 2020a, 2020b).

In addition to ensemble trees algorithms, the artificial neural network (ANN) approaches have been used for both reference and actual ET prediction (Abdullah et al., 2015; Cobaner, 2011; Feng et al., 2017; Ferreira et al., 2019; Jung et al., 2011; Kisi and Alizamir, 2018). Most of these ANN approaches such as convoluted neural network (CNN) for ET modeling are based on a feed-forward neural network approach where the algorithm is introduced for a single layer (Tavares et al., 2015; Yassin et al., 2016). However, for time series analysis, one of the drawbacks of feed-forward ANNs is that any information about the sequence of inputs is lost and data pre-processing for singular spectrum analysis of time series in these models require complicated procedures (Sahoo et al., 2017). In addition, traditional ANNs also have a problem of exploding or vanishing gradient (Rangapuram et al., 2018). Hence a special type of neural network architecture, recurrent neural networks (RNNs) is designed where input is processed in its sequential order to understand temporal dynamics (Carriere et al., 1996). For problems such as ET prediction and time series forecasting, for which order of the input variables is important, a specific kind of RNN is Long Short-Term Memory (LSTM) that can solve the problem of vanishing gradient. Since our study focuses on time series prediction and forecasting, RNN such as LSTM along with ensemble trees algorithm such as RF is a suitable choice.

In LSTM, connections between units and cells allow data to move in a forward and backward direction within the model framework. This method helps to overcome the problem of learning lagged dependencies found in traditional RNN. In the case of the water cycle, such an approach allows the model to preserve previous information for future uses such as water storage effects (e.g. snow) or shallow groundwater-driven systems. Kao et al. (2020) used an LSTM model to forecast floods in inundation-prone areas and found that LSTM can be used to link the sequence of rainfall with a sequence of runoff. In addition, Kratzert et al. (2019) applied process-based constraints on an LSTM modeling framework to simulate runoff for a variety of watersheds and found that LSTM outperformed benchmark physically-based coupled models.

As noted above, challenges in existing methods for predicting and forecasting actual ET are the need for extensive parametrization, lack of relevant data drivers, the computational cost of process-based models, and lack of direct estimate of actual ET from empirical models. Knowledge of the performance of data-driven models in different types of irrigated and non– irrigated crops under different soil types is still partial and fragmented. In addition, models in existing studies have only been applied on limited test data sets. Few studies have evaluated the relative contributions of the different input datasets (predictors) to the accuracy and uncertainty of the actual ET models in agricultural fields, particularly across different management (irrigated vs. rain-fed), crop types, and soil textures.

Here, we ask 1) *how well can empirical ML models predict and forecast*

ET 3 days in advance in irrigated and rain-fed agricultural lands across the Midwest US? 2) *what are important drivers for predicting and forecasting ET 3 days in advance in irrigated and non-irrigated areas*? We evaluate two different ML models, RF and LSTM, with differing numbers of predictors (5, 11, 16) across a range of crop and soil texture types where eddy covariance observations were available between 2003 and 2019. The results of this evaluation allow us to better understand the predictors of accuracy and uncertainty in the ET models and propose a multistep prediction and forecast agricultural ET model that can be applied to locations with limited in situ observations. Since there is no clear understanding of minimum required predictors for accurate estimates of ET, our models with different sets of predictors (5,11, 16) can help to understand the need for important or minimum drivers for different crop fields on various soil textures in areas with scarce data.

2. Methods

In this paper prediction and forecasting models based on RF and LSTM framework are proposed. For ET prediction, RF and LSTM model with 5, 11, and 16 predictors are proposed. For all model experiments, simulations are based on data from 2003 to 2019.

2.1. Data description

The proposed model performance was assessed by using the observed ET data obtained from the AmeriFlux database or site investigators (Table 1) for 19 sites located in the agricultural areas of the US Midwest in states of Iowa, Illinois, Michigan, Minnesota, Nebraska, Ohio and Wisconsin (Fig. 1). Out of those 19 sites, five are irrigated and 14 are rainfed (Table 1). Study sites were all located in a temperate climate with cool to cold winters and hot, humid summers. The dominant crops in those regions are soybeans, potatoes, and corn with coarse-grained (sandy loam, loamy sand, loam) and fine-grained (silt Loam and silt clay) soils. The data duration used during this study ranged from 2003 to 2019 with a daily time step for continuous variables. After removing outliers, only months with less than 3 days gap were used and years with more than one month of missing data were removed. Data gaps for quality-controlled half-hourly ET observations were filled with post-processing software REddyProc (Wutzler et al. 2018). REddyProc method uses co-variation and temporal autocorrelation of turbulent fluxes and gaps are filled based on available information about air temperature, incoming solar radiation, and vapor pressure deficient based on marginal distribution sampling. Additional meteorological data were obtained from Daymet (Thornton et al., 2014) and North American Land Data Assimilation System (NLDAS) Land Surface Model (LSM) (Xia et al., 2012). In addition, MODIS (Aqua MODIS MYD09GA. Aqua MODIS MYD09GA) satellite data (Vermote, 2015) was also used for enhanced vegetation index (EVI), albedo, and solar zenith angle. Table 1 describes the study site locations, duration of measurements, and ancillary information. Summary statistics such as mean, maximum, and variance of ET across different observation sites is included in Table 2.

The selection of model input predictors was due to their influence on ET and their availability for agricultural sites (Fig. 2). Sixteen model predictors used on daily time stamp for model predictions include moving average precipitation for 7 days (Prcp7), 15 days (Prcp15), and

Table 1
Description of agricultural flux towers study sites located in Midwest USA. Sites names are based on AmeriFlux ID.

State	Site ID	Lat.	Long.	Duration	Soil Type	Rain-Fed/ Irrigated	Crop types	Doi
IA	US-Br1	41.97	−93.69	2005–2011	Loam	Rain-Fed	Corn in odd years, Soy in even years and 2011	Prueger and Parkin (2001a) https://doi.org/10.17190/AMF/1246038
IA	US-Br3	41.97	−93.69	2005–2011	Clay Loam	Rain-Fed	Corn in even years, Soy in odd years	Prueger and Parkin (2001b) https://doi.org/10.17190/AMF/1246039
IL	US-IB1	41.86	−88.22	2006–2017	Silt Loam	Rain-Fed	Corn in even years and 2013, 2017, Soy in odd years and 2014	Matamala (2005) https://doi.org/10.17190/AMF/1246065
IL	US-Bo2	40.01	−88.29	2004–2007	Silty Clay	Rain-Fed	Corn in even years, Soy in odd years	Bernacchi (2004–2008) https://doi.org/10.17190/AMF/1246037
IL	US-Bo1	40.01	−88.29	2005–2016 except 2008–2009	Silt Loam	Rain-Fed	Corn in odd years, Soy in even years	Meyers (1996) https://doi.org/10.17190/AMF/1246036
MI	US-KL1	42.48	−85.44	2009–2018	Sandy Loam	Rain-Fed	Soy in 2009, Corn in 2010–2018	Chen (2009–2018) https://ameriflux.lbl.gov/sites
MI	US-JCK	42.21	−84.85	2018	Sandy Loam	Rain-Fed	Soy	Chen (2018a) https://ameriflux.lbl.gov/sites
MI	US-KM1	42.44	−85.33	2014–2018	Sandy Loam	Rain-Fed	Corn	Chen (2018b)
MI	Jackson 1	42.26	−84.84	2018	Loamy Sand	Irrigated	Corn	Chen (2018c)
MN	US-Ro1	44.71	−93.09	2004–2016	Silt Loam	Rain-Fed	Corn in odd years, Soy in even years	Baker and Griffis (2003-2017a) https://doi.org/10.17190/AMF/1246092
MN	US-Ro2	44.73	−93.09	2008, 2011, 2012, 2016	Silt Loam	Rain-Fed	Soy in 2012, Corn in 2008, 2011, 2016	Baker and Griffis (2003-2017b) https://doi.org/10.17190/AMF/1418683
MN	US-Ro3	44.72	−93.09	2004–2007	Silt Loam	Rain-Fed	Corn in odd years, Soy in even years	Baker and Griffis (2003–2010) https://doi.org/10.17190/AMF/1246093
MN	US-Ro5	44.69	−93.06	2017–2018	Silt Loam	Rain-Fed	Soy in 2017, Corn in 2018	Baker and Griffis (2017) https://doi.org/10.17190/AMF/1419508
NE	US-Ne2	41.16	−96.47	2003–2013	Silt Loam	Irrigated	Soy in 2004, 2006 and 2008. Corn in other years	Suyker (2001a) https://doi.org/10.17190/AMF/1246085
NE	US-Ne3	41.18	−96.44	2003–2013	Silt Loam	Rain-Fed	Corn in odd years, Soy in even years	Suyker (2001b) https://doi.org/10.17190/AMF/1246086
NE	US-Ne1	41.17	−96.48	2003–2012	Silty Clay Loam	Irrigated	Corn	Suyker (2001c) https://doi.org/10.17190/AMF/1246084
OH	US-CRT	41.63	−83.35	2011–2012	Silt Loam	Rain-Fed	Soy	Chen and Chu (2011–2013) https://doi.org/10.17190/AMF/1246156
WI	US-CS1	44.10	−89.54	2018–2019	Loamy Sand	Irrigated	Potatoes	Desai (2018–2019) https://doi.org/10.17190/AMF/1617710
WI	US-CS3	44.14	−89.57	2019	Loamy Sand	Irrigated	Potatoes	Desai (2019–2020) https://doi.org/10.17190/AMF/1617713

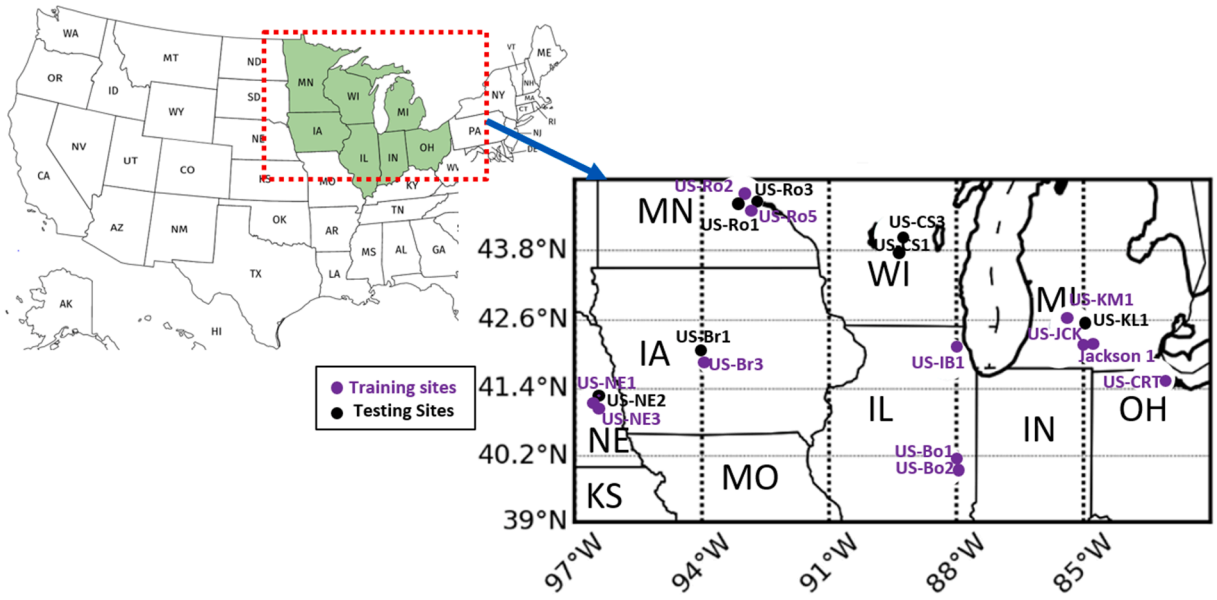

Fig. 1. Study sites and locations for calibration and evaluation data. AmeriFlux site ID were used to identify locations. Thirteen sites were used for training with some part of data (2009–2016) for calibration and remaining data (2017–2018) from the same sites for testing. Hence thirteen sites (n = 18481) were used for training and seven independent sites (n = 7850) were used in testing model performance.

Table 2
Descriptive statistics for agricultural flux towers study sites located in Midwest USA. Mean, standard deviation, sample variance, skewness, minimum and maximum daily ET was calculated to show ET variability across sites.

State	Site ID	Mean	Standard Deviation	Sample Variance	Skewness	Minimum daily ET	Maximum daily ET	Rain-Fed/Irrigated
IA	US-Br1	2.56	1.64	2.70	0.64	0.00	7.56	Rain-Fed
IA	US-Br3	2.50	1.57	2.46	0.63	0.00	9.27	Rain-Fed
IL	US-IB1	2.54	1.38	1.91	0.63	0.19	8.12	Rain-Fed
IL	US-Bo2	1.96	1.61	2.60	0.64	0.00	8.23	Rain-Fed
IL	US-Bo1	1.60	1.26	1.60	0.93	0.00	6.81	Rain-Fed
MI	US-KL1	2.22	1.13	1.27	0.81	0.11	6.86	Rain-Fed
MI	US-JCK	1.75	0.77	0.60	0.45	0.30	3.85	Rain-Fed
MI	US-KM1	2.15	1.13	1.28	0.81	0.14	7.04	Rain-Fed
MI	Jackson 1	2.10	1.30	1.68	1.93	0.00	7.99	Irrigated
MN	US-Ro1	2.32	1.51	2.29	0.88	0.00	10.51	Rain-Fed
MN	US-Ro2	2.55	1.25	1.57	0.27	0.08	6.97	Rain-Fed
MN	US-Ro3	2.13	1.30	1.69	0.86	0.00	7.19	Rain-Fed
MN	US-Ro5	2.06	1.24	1.54	0.79	0.28	6.64	Rain-Fed
NE	US-Ne2	2.75	1.77	3.13	0.60	0.08	9.52	Irrigated
NE	US-Ne3	2.46	1.62	2.62	0.57	0.09	6.85	Rain-Fed
NE	US-Ne1	2.86	1.85	3.42	0.53	0.00	9.77	Irrigated
OH	US-CRT	2.83	1.60	2.55	0.93	0.51	8.87	Rain-Fed
WI	US-CS1	1.82	1.29	1.66	0.90	0.30	5.70	Irrigated
WI	US-CS3	1.70	1.25	1.63	0.92	0.28	5.30	Irrigated

30 days (Prcp30), as proxies for soil moisture (because direct soil moisture data was not present at all sites), maximum air temperature (Tmax), long-wave radiation (LW), incoming short-wave radiation (SW), solar zenith angle (SolarZenith), albedo (Albedo), enhanced vegetation index (EVI), soil texture (Soil), irrigated versus non irrigated proxy (Irr_nonirr), crop cover (Crop_cover), crop coefficient (Crop coeff), cumulative growing degree days (CumGDD), wind speed (Wind) and vapor pressure (VP). For RF_5 and LSTM_5 daily air temperature (Tavg) was used while for RF_11, RF_16 and LSTM_11 and LSTM_16 maximum air temperature (Tmax) was used. Since RF_5 and LSTM_5 were based on drivers from Priestley Taylor equation, Tavg was used instead of Tmax or Tmin for simpler models. These predictors were chosen because of their ability to explain physical processes (Cobaner, 2011; FAO, 2015; Feng et al, 2017) of ET as well as easy availability in most regions. The data source for 16 model predictors along with different commination for predictors for various model versions is included in Table 1 and Table 3.

Cumulative growing degree days (CumGDD) are associated with different phases of plant development (Cleland et al., 2007) and calculated for all growing seasons based on the method described in Anandhi (2016). Crop coefficients were calculated based on the Food and Agriculture Organization of the United Nations (the FAO-56 method) first proposed by Allen et al. (1998). FAO-56 method provides both transpiration and evaporation from soil and reference ET is calculated based on Penman–Monteith equation. Based on the related version of FAO-56 method (Allen et al., 1998), adjustments were made according to local crop physical condition.

2.2. *Random forest model framework*

RF is an ensemble of different trees where trees are built to explain the variability of the output by grouping data in homogenous sets. Unique trees are built by data splitting in random sets with replacement like bootstrapping as well as by random subsets of predictors, which

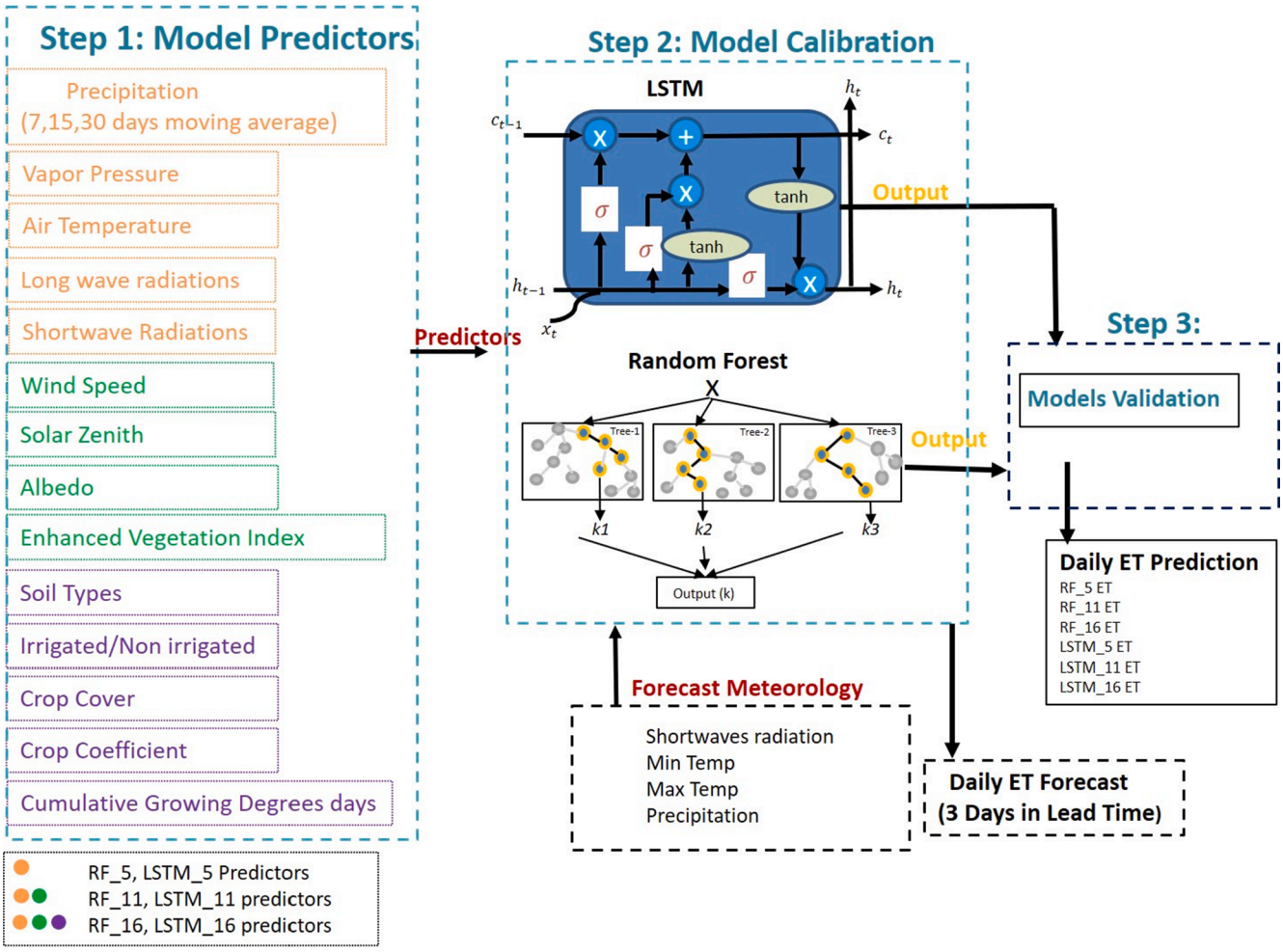

Fig. 2. Framework of key steps for proposed daily ET prediction and forecast models. RF_5, LSTM_5 model predictors are in orange color, RF_11, LSTM_11 model predictors are in orange and green color, RF_16, LSTM_16 predictors are in orange, green and purple color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

helps to increase diversity among trees (Breiman, 2001).

$$\{h(x, \acute{O}_t), t = 1, 2, 3, 4 \cdots\cdots T\}$$

where daily ET (independent variable) is represented by $\times$, T is the number of distinct regression trees and predicted value of regression tree in form of ET is represented by $h\ (x, \theta_t)$. Hence random forest builds a large forest where each tree predicts a value for ET. In this study regression, RF of daily ET is affected by different predictors and the average of all those values is the final prediction of RF.

$$h(x) = \frac{1}{T}\sum_{t=1}^{T}(h(x, \acute{O}_t))$$

Out-of-bag (oob) sampling is used for RF internal validation. In addition, the importance of each predictor can be determined by holding some predictors constant, while permuting each predictor at a time and then comparing the oob error. The paraments that are tuned during RF calibration include n_estimator (number of trees in the forest), and min_samples_split (minimum number of samples required to split an internal node), and min_samples_leaf (minimum number of samples required to be at a leaf node). The mean of yearly and monthly observed ET, precipitation, and air temperature was computed across various sites and then sites were split between training and testing dataset such a way that each data set has dry, wet, and average years for representation of site conditions. Three RF models RF_16, RF_11, and RF_5 were built with 16, 11, and 5 predictors respectively (Table 3) with 70% of the data were used for training and 30% of the remaining data were used for evaluation/validation based on the hold-out method.

2.3. Long Short-Term memory network (LSTM)

LSTM is a special kind of RNN, without the limitation to learn time series dependencies between input and output features. One limitation of traditional RNN is the inability to "remember" a sequence with long lengths (e.g., >10) (Bengio et al., 1994). However, the LSTM framework retains memory about the previous timestamp which can help to model lags in energy balance fluxes. The information about long-term memory for each time step is contained in cell state or cell memory c_t of LSTM and sequence of inputs (model predictors) as $\times$ is presented in the model and output (predicted or forecast ET) is obtained as h while six parameters show in equations below are updated at each time step in each cell.

Feed-Farward ANNs such as CNN does not store information in memory. We compared LSTM with CNN and chose LSTM algorithm for our prediction and forecasting based on performance . All LSTM models outperformed CNN models. For example, NSE and Willmott's skill score for LSTM_16 was 0.65 and 0.88 respectively while NSE and Willmott's skill score for CNN_16 was 0.53 and 0.84 (Fig. S1 and Table S1 in Supplementary Materials).

In LSTM model a sigmoid function is computed by a forget gate (f_t) on new input x_t and previous result h_{t-1}. The sigmoid function is a smooth, differentiable nonlinear function that produces non-binary activation where weights can be updated with every data point. The

Table 3
Description of model inputs and predictors used for different versions of models are included. Number at end of each model name shows the number of predictors used to build model. e.g RF_16 is RF model with 16 predictors and LSTM_5 is LTM model with five predictors. Data sources are included.

Model	Driver	Abbreviation	Source
RF_5	Precipitation	Prcp7, Prcp15, Prcp30	Daymet
RF_11	Vapor pressure	VP	Daymet
RF_16	Air Temperature	Tmax, Tavg	Daymet
LSTM_5	Long wave radiation	LW	NLDAS_Forcing LSM
LSTM_11 LSTM_16	Shortwave radiation	SW	NLDAS_Forcing LSM
RF_11	Wind Speed	Wind	Rain-Fed
RF_16	Solar Zenith	SolarZenith	Aqua MODIS MYD09GA
LSTM_11	Albedo	Albedo	Aqua MODIS MYDTBGA
LSTM_16	Enhanced vegetation Index	EVI	Aqua MODIS MYD09GA
RF_16 LSTM_16	Soil types	Soil	Soil Survey Geographic Database (SSURGO)
	Irrigated-non irrigated	Irr_nonirr	Ameri flux sites
	Crop Cover	Crop_cover	Ameri flux sites
	Crop Coefficient	Crop coeff	Computed as function of growing degrees days
	Cumulative growing degrees days	CumGDD	Computed by empirical formula based on temperature

differentiable activation function is necessary because it can compute the gradient which is required for training via backpropagation. In addition, it can be derived from a maximum entropy model. The sigmoid function helps the forget gate to decide what information needs to be remembered and what information can be discarded from memory. The sigmoid function is also provided with adjustable weights (W) and biases (b) in each LSTM cell. The new information that is going to be remembered is placed in a cell state with the help of the input gate (it), which is also calculated by a sigmoid function. A tanh function is used to calculate a new cell state (c_t). The output gate regulates the information of the state of cell c_t using a sigmoid function.

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (1)$$

$$c_t = tanh(W_c x_t + U_c h_{t-1} + b_c \quad (2)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (3)$$

$$C_t = f_i{}^* c_{t-1} + i_i c_t \quad (4)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_0) \quad (5)$$

$$h_t = \tanh(c_t)^* o_t \quad (6)$$

where matrices of weights from the input, forget, and output gates to the input are denoted by Wi, Wf, and Wo, respectively. The bias vectors for input, forget, and output gates are shown by bi, bf, bo, respectively. The hidden layer matrix of weights from the input, forget, and output gates are represented by Ui, Uf, and Uo, respectively. Logistic sigmoid σ is an element-wise non-linear activation function and the element-wise multiplication of two vectors is denoted with *.

In this study, three LSTM models LSTM_16, LSTM_11, LSTM_5 were built with 16, 11 and 5 predictors respectively. Here the model with 16 predictors is assumed to account for more variability than a simple model of 5 predictors because it includes predictors related to both meteorological and biophysical processes. The model with 5 predictors was built to make a model based on inputs from the common physical ET model of Priestley-Taylor, while the model with 11 predictors was used as an intermediate framework between the complex and simple model. Each model had two fully connected layers. The first layer is called the encoder layer with 50 neurons and that layer is responsible for reading and interpreting the input sequence. Initially, the model was run with 25, 50, 100, and 200 neurons, and an optimal number of 50 neurons was selected for layer 1 based on lower ubRMSE and high erWillmott's index (Fig. S2 in Supplementary Materials).

In order to combat the problem of overfitting, a regularization method of "dropout" was applied after the first layer where the dropout value is a percentage between 0 (no dropout) and 1 (no connection) for LSTM units (Kratzert et al., 2019). Models were tested using different values for dropout and evaluation statistics were calculated to find the optimal number of neurons. In addition, training and testing data performance was compared to avoid an overfitting or underfitting problem (Fig. S2 in Supplementary Materials). A dropout value of 0.10 was applied in LSTM_16, and a dropout value of 0.25 was applied for LSTM_11 and LSTM_5 (Table 4). After the dropout function, a decoder layer was applied which used the output of the encoder (first layer) as an input. A second LSTM layer that comes after the encoder had 25, 50, and 100 neurons for LSTM_5, 11, and 16 respectively (Table 4). The optimal number of neurons was obtained by using different combinations of neurons and dropout factors until reduced ubRMSE was obtained without overfitting or undefining model.

Lastly, two dense layers were applied. The model was calibrated using ADAM (*adaptive moment estimation)* optimizer and mean squared error loss function. A moderate rate of 0.001 is used for the ADAM optimizer for learning. During the calibration process, it was observed that a high learning rate of 0.1 missed the optimal point ($R^2 > 0.6$) and a smaller learning rate of 10^{-6} led to a longer convergence time for the model (Zhang et al., 2018).

Randomly selected 70% of raw data were used for calibration and 30% of the remaining data were used for evaluation using the hold-out method. During the training and optimization of the learning algorithm, a loss function was used to estimate the error of the current state of the model. The purpose of this loss function is to reduce the loss of the next evaluation by updating weights (Kratzert et al., 2019). During training initial loss function was 0.59, 0.72 and 0.70 for LSTM_16, LSTM_11 and LSTM_5 respectively that was reduced to 0.30, 0.51 and 0.60 for LSTM_16, LSTM_11 and LSTM_5 respectively by end of training.

2.4. Land surface model

We benchmarked our empirical models against output from process-based model ET from the North American Land Data Assimilation System (NLDAS) version 2 LSM model (Xia et al., 2012). Daily ET data were downloaded from Land Data Assimilation System (LDAS) (https://ldas.gsfc.nasa.gov/nldas/). Penman-Monteith equation is used in NLDAS-Noah LSM energy balance for latent heat flux here ET is based on evaporation, and plant transpiration is driven by soil moisture stress on the top layer of the soil profile (Chen et al., 1996). Hence under wet conditions, ET is equal to potential evapotranspiration. Richards (Richards, 1931) equation is used in this model to simulate soil moisture dynamics. Root zone plant transpiration is driven by canopy interception and canopy resistance that is parameterized by solar radiation, air temperature, vapor pressure, and soil moisture (Koster and Suarez,

Table 4
Parameters for different version of RF and LSTM prediction models. For RF version, parameters of n_estimators, min_samples_leaf included. LSTM model versions are calibrated using layer 1 dropout, layer 2 neurons, and epoch.

Model	n_estimator	min_samples_leaf	min_samples_split = 8
RF_5	100	5	8
RF_11	100	5	8
RF_16	150	5	6
Model	Layer 1 dropout	Layer 2 Neurons	Epoch
LSTM_5	0.25	25	100
LSTM_11	0.25	50	100
LSTM_16	0.1	100	65

1996). Surface albedo is simulated based on diurnal variations and simulated LAI (500 m resolution) varies seasonally as well as spatially and the minimum stomatal resistance parameters are based on vegetation types. In addition, surface runoff is calculated based on the Simple Water Balance (SWB) model, and baseflow is represented by gravity drainage (Chen and Dudhia, 2001).

2.5. Significant predictors

The significance of each predictor variable with respect to its effect on the RF model is displayed by predictor importance. The RF model algorithm calculates predictor importance internally to account for bias in test data (Liaw and Wiener, 2002). In decision trees, the node uses predictors to split values of output (ET) and similar values of the output (ET) end up in the same set after the split. Predictor importance is measured by measuring how much each predictor contributes to decreasing the variance. In order words. importance of predictor is based on the frequency of its inclusion in the sample by all trees and it is a measure of how much removing a predictor decreases accuracy (Breiman, 2002; Pedregosa et al., 2011). A decrease in variance from each predictor is averaged in a forest and predictors are ranked according to this measure. We used the Sklearn algorithm in python 3.7 to calculate the importance score for each predictor after training and the score is scaled to 1 to calculate the influence of each predictor on ET. Therefore, the sum of the importance of all predictors is equal to one, and the higher the value associated with a predictor, the more important that predictor. The importance of model predictors was calculated for both prediction and forecast model versions of RF.

The LSTM algorithm does not have a built-in variable importance selection criterion. So predictors' importance was measured in terms of change in NSE by removing certain predictors and by comparing the change in NSE with the NSE obtained from the original LSTM_16 model.

2.6. Forecast model

After evaluation of ET prediction, we also proposed a multistep forecast model that can forecast ET three days ahead of time using RF and LSTM models as described above (Fig. 2). At each daily time step, there are three ET forecasts: 1) day 1 ET (tomorrow ET), day 2 ET (ET day after tomorrow), day 3 ET (ET three days from today). Forecasts were made by integrating the uncertainty of forecast meteorology through ensemble simulation. Hence, along with 16 model predictions that were used for the prediction model (Table 3), input meteorological predictors from re-forecasts from the National Oceanic and Atmospheric Administration's (NOAA's) National Centers for Environmental Prediction (NCEP) 11-member Global Ensemble Forecast System Reforecast version 2 (GEFSRv2) were propagated into each model to make forecasts (Hamill et al., 2013). The uncertainty in meteorological forecasts of GEFS was quantified by generating ten ensembles of multiple ET forecasts, each perturbed from the original observations (or control). RF_16 and LSTM_16 versions were used for forecasting ET.

RF_16 prediction model and initial forecasting model runs provided us with identification of important variables as described above. We also measured the Pearson correlation between predictors and ET to evaluate forecast reliability. In addition, for LSTM_16 we did some initial model runs with different combinations of predictors and only used those predictors that helped to improve the accuracy of the model (using ubRMSE, MAE, AIC criteria). Hence based on initial model runs, information from the prediction model, and literature review (Fang et al., 2018), only those meteorological predictors were selected that were the main drivers of future ET, i.e., maximum and incoming solar radiation, minimum temperature, and precipitation. So for the day 1 ET forecast, forecast meteorology for the next day was included in the model. For day 2 ET, forecast meteorology of days 1 and 2 were included. For day 3 ET, forecast meteorology of days 1, 2, and 3 were included.

2.7. Model evaluations

There are 19 sites with 14 rain-fed sites, and five irrigated sites, with a total of seventeen site-years (growing season April-October) of observations, or 26,331 daily observations of ET. Thirteen of the 19 sites were used for training where for one of the sites 80% of data was used in training and the remaining 20% of data from the same site was used in testing. These thirteen sites were used for training while seven sites were held-out and used exclusively for testing. In total, 70% of observed ET data (18,481 daily ET observations), from the 13 different agricultural sites for corn, soybeans, and potatoes, was used for calibration, and data from the remaining seven agricultural sites were used for evaluation for the time period 2003–2019 (7,850 daily ET observations). To test the accuracy of the calibrated models, a subset of data was used to determine the optimal number of trees in RF and hidden neurons and layers in LSTM and an optimum or satisfactory point for the calibration without overfitting the models for one set of data.

For statistical analysis, coefficient of determination R^2, Pearson correlation coefficient, Nash–Sutcliffe (NSE), Willmott's skill score or index of model performance (Willmott, 1981), mean absolute error (MAE), unbiased root mean square error (ubRMSE), RMSE-observations standard deviation ratio (RSR) (Moriasi et al., 2015), percentage bias (Pbias) were used to assess the predictive ability of the proposed RF and LSTM models. In addition, Akaike's Information Criteria (AIC) metric was also used to see the effect of penalization of additional drivers to the model (Akaike,1970). AIC adds penalty by including additional predictors in the model that leads to higher error. Hence a more parsimonious model will have lower AIC.

$$AIC = -2\ln(L) + 2k$$

where L is the likelihood and k is the number of parameters. Likelihood is calculated as the log of mean square error.

3. Results

3.1. RF versus LSTM prediction model evaluation

Fig. 3 illustrated the performance of the two ET prediction algorithms for the test data, which demonstrated the ability of the calibrated models to generalize to unseen ET observations (test data) from eddy covariance flux towers across multiple crop types. The evaluation statistics shown in Fig. 3 indicated that there is a good agreement between the predicted and observed ET values across corn, soybeans, and potatoes. For RF 16 model, R^2 and NSE values for the corn vary from 0.53 to 0.70 (Willmott's score 0.85–0.9) in the testing period and for LSTM_16 the R^2 range was 0.56–0.66 and Willmott's skill score (0.80–0.89). Further, LSTM_16 had less bias for the site with a smaller number of observations (potatoes in loamy sand) compared to RF_16 (Fig. 3).

The more complex model required a greater number of neurons for the LSTM hidden layer. The number of neurons for different versions of best-fit LSTM models varies from 25 to 100 (Table 4). For the LSTM_16 model, using more than two layers and more than 100 neurons did not improve the model performance on testing data. The run time for the LSTM_16 model and RF_16 model was ten and two minutes, respectively on an Intel CORE i7 9750H CPU, windows 10 X64 based processor.

Both model outputs products closely follow the seasonal growth of crops (Fig. 4). During the shoulder months (i.e. September to next May), ET is lower, and as percentage canopy cover increased in June-August, so did ET. In addition, both observations and models are consistent in showing that during dry years (2006, 2010, and 2012), ET is higher than compared to wet years (2014–2018) across crop types. For example, in the drought of 2012, the ET at US-Ro1 and US-Ro3 was above 6 mm day^{-1} while it was less than 6 mm day^{-1} in the wet summer of 2015.

The consistency of modeled ET against the ground truth differs based

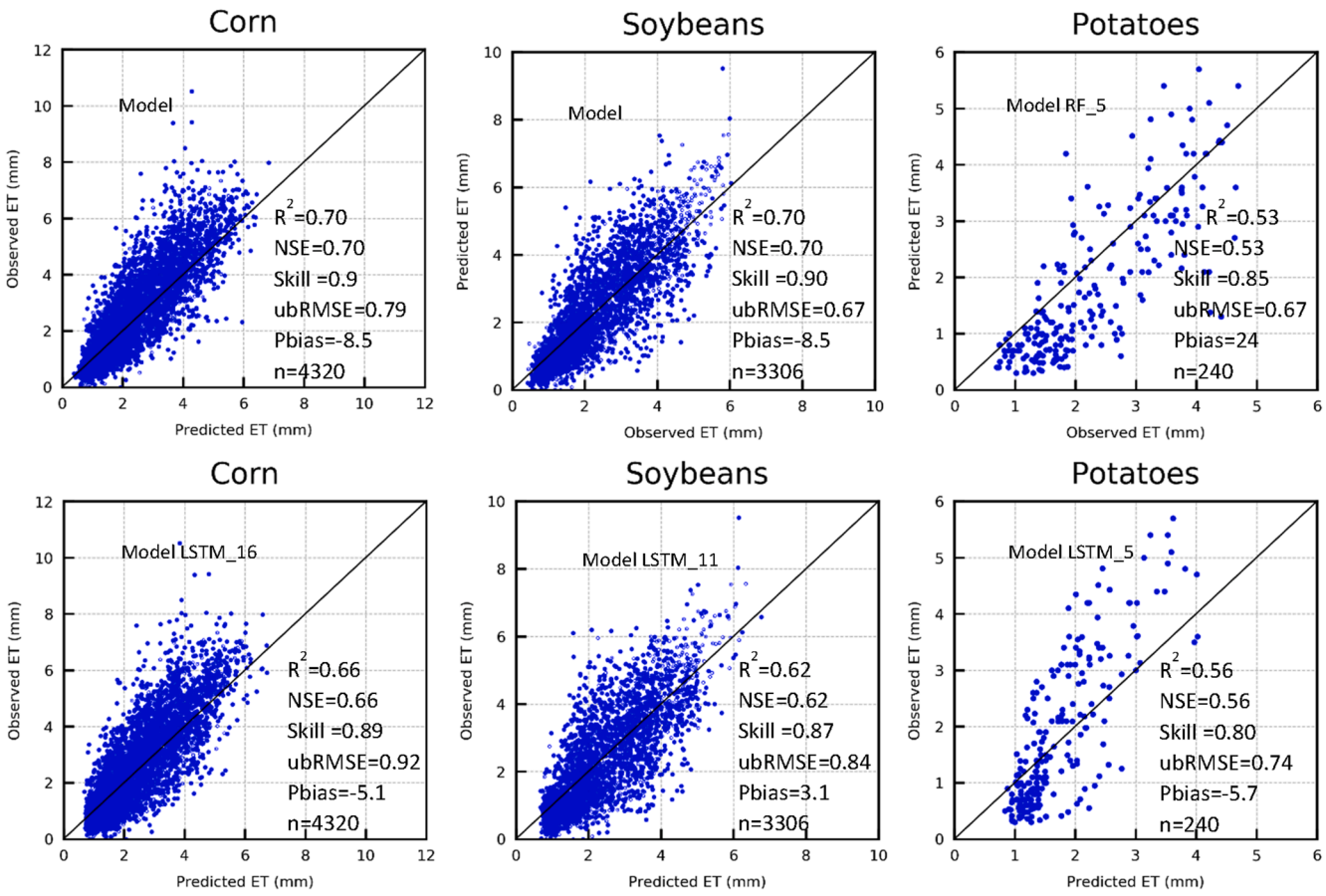

Fig. 3. Evaluation results of the proposed daily Random Forest (RF_16) and LSTM (LSTM_16) prediction models for various soil texture for flux tower locations in Midwest. 16 means model with 16 variables. n = sample size.

on the regional characteristic and amount of data available for calibration. For example, in sites US-CS1 and US-CS3, RF model predictors could well track the dynamics of the water loss caused by an increase in canopy cover. However, LSTM_16 had lower Pbias (−5.7%) but higher error (Figs. 3 and 4) than RF_16 (−24.1% Pbias). In general, RF_16 had a higher bias, but lower error compared to LSTM_16 during months when irrigation and ET are higher (June, July). We also computed the Empirical Cumulative Distribution Function (ECDF) for the evaluation period under different soil conditions, soil moisture (variable precipitation under wet, dry years), and crop types. The ECDFs of RF_16 and LSTM_16 models match closely with the observations. Compared to extreme events, the middle section of ECDFs curves is better represented by models.

3.2. Significant predictors

The significance of each predictor variable with respect to its effect on the RF and LSTM models is displayed by predictor variables' importance. Four predictor variables that explained most of the variance in the data include Enhanced vegetation Index (EVI), solar zenith angle, incoming SW radiation, and CumGDD. These four predictors combined explain 62% of the model variance. Fig. 5 also showed the Pearson correlation coefficient between predictors and ET, which is positively correlated with VP and EVI. Since most of the Midwest regions are not moisture limited and have a humid climate with warm summers, we expect to see a high correlation between ET and maximum daily temperature (during the growing season) compared to the correlation between ET and precipitation (i.e., our soil moisture proxy in the form of moving average precipitation). In irrigated fields, NSE was reduced from 0.7 to 0.52 by removing SW and SolarZenith predictors (Fig. 3S in Supplementary materials) in LSTM. In addition, a change in NSE from 0.6 to 0.47 was observed by removing SW and SolarZenith from rainfed or non-irrigated fields. The positive correlation between maximum daily temperature can be seen in the ranking of the CumGDD predictor among the four most important predictors for RF (Fig. 5). In contrast, despite the low direct correlation of soil moisture proxy (seven days average precipitation), it is among the five most important predictors for the RF model (Fig. 5). Crop coefficient also improved model performance by explaining the dynamics of canopies (cover fraction, LAI, greenness). Our analysis for RF model showed that VP and crop coefficients were the most important predictors for irrigated crops, while short wave radiation and enhanced vegetation index were key predictors for non-irrigated crops (Fig. 6).

3.3. Model performances

Different versions of the RF and LSTM models (complex versus simple models) were also evaluated on a daily timestep in comparison with the daily predictions from the mechanistic model – NLDAS-Noah (Table 5). Overall, the RF_16 model resulted in an R^2 of 0.7 with a Pbias of −4.7% while the RF_11 model had an R^2 of 0.7 with Pbias of −5.3% (Table 5). The NLDAS-Noah model had a 0.57 R^2 with the lowest Pbias of 0.3 (Fig. 7). The lowest Pbias for NLDAS-Noah was most likely a result of the averaged ET prediction across a larger geographical area. That leads to a wider spread from the mean estimate on the scatter plot with a ubRMSE of 1.1 mm/day and a lower R^2 of 0.57 for the NLDAS-Noah model (Fig. 7).

Residuals were obtained for each model time step (daily) by subtracting the observed ET from the predicted ET. A negative residual value showed that the model underestimates ET while a positive residual

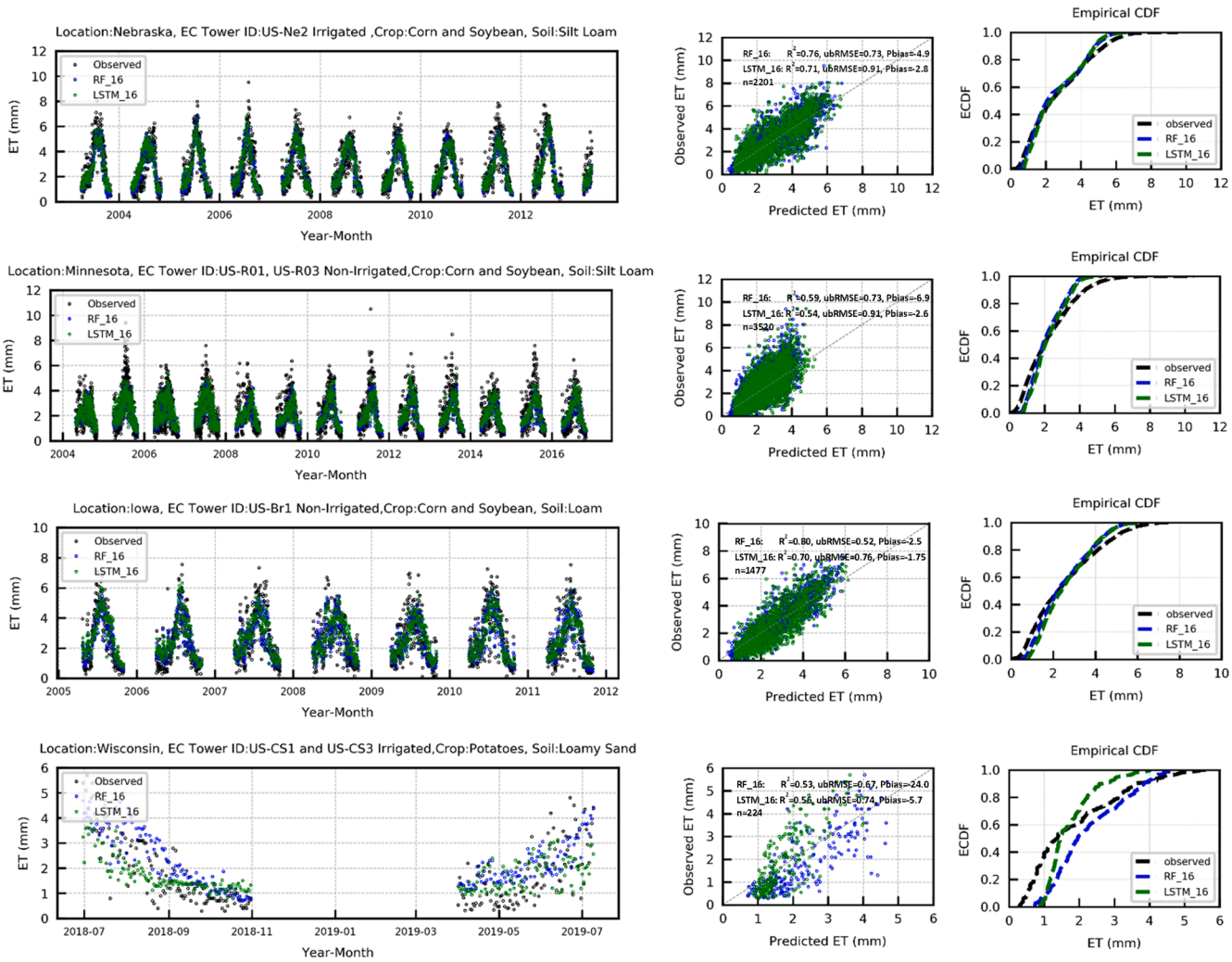

Fig. 4. Time series, scatter plot, and ECDF of modeled and observed ET; Observed = ET from six representative study sites; Data for US-CS1 and USCS3 were presented together in one graph. The black dots and dotted line show observed ET and the blue and green dotted line and points purple are indicating, respectively RF_16 and LSTM_16 model. RF_16 = random forest model with 16 input variables; LSTM_16 = LSTM model with 16 input variables; ECDF = Empirical Distribution Function; R^2 = Coefficient of determination; RMSE = Root Mean Square Error; Pbias = Percentage bias. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

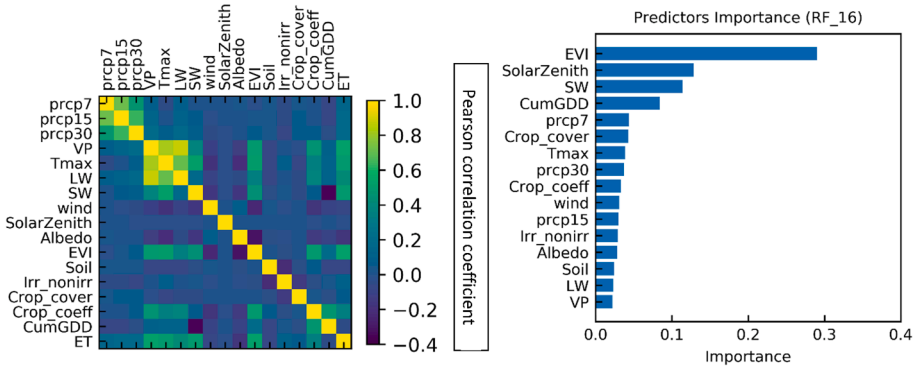

Fig. 5. Pearson Correlation coefficient between prediction model predictors and response (daily ET); Predictors importance for RF_16 model. Values were scaled to 1 to calculate the influence of each predictor on the response.

means that ET is overestimated. The distribution of residuals is the largest for the testing period. Based on residuals, the RF_11 produced the most accurate results in April and June (with 0.02 and −0.01 mm residuals, respectively) while the RF_16 was the most accurate model in September (Fig. 8). In July and August, the NLDAS-Noah model prediction was more accurate compared to other models. This could be because mechanistic models such as NLDAS_Noah has constrained ET by using soil moisture at different depths. If soil moisture storage is significantly variable due to large ET during the mid-growing seasons (July-August), the mechanistic model may outperform empirical

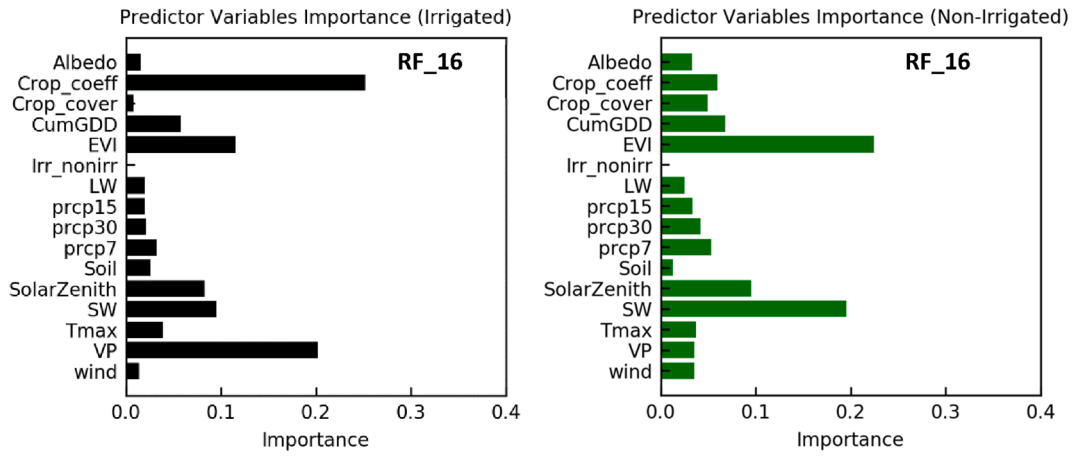

Fig. 6. Predictors importance for irrigated and non-irrigated crops based on RF_16 model. Predictor importance is scaled to one which means that sum of the contribution of all predictors is equal to one. Predictors with longer horizontal bars are more important in terms of explaining model variance.

Table 5
Model performance evaluation statistics for different versions of prediction models on daily timestamp. Number at end of each model name shows the number of predictors used to build model. e.g RF_16 is RF model with 16 predictors and LSTM_5 is LTM model with five predictors.

Model Evaluation Statistics for prediction model	R^2	NSE	Willmott skill score	Pearson Corr.	MAE	ubRMSE (mm/day)	RSR (mm/day)	Pbias (%)	AIC
RF_16	0.70	0.70	0.90	0.84	0.64	0.75	0.55	−4.7	0.0
LSTM_16	0.65	0.65	0.88	0.81	0.72	0.89	0.59	−1.9	0.34
NLDAS_Noah	0.57	0.57	0.86	0.76	0.79	1.1	0.65	0.3	Benchmark
RF_11	0.70	0.70	0.89	0.85	0.66	0.76	0.55	−5.3	0.04
LSTM_11	0.63	0.63	0.86	0.82	0.73	0.91	0.60	−6.0	0.42
RF_5	0.63	0.63	0.85	0.81	0.73	0.94	0.61	−5.4	0.46
LSTM_5	0.53	0.53	0.80	0.75	0.82	1.20	0.69	−9.3	0.94

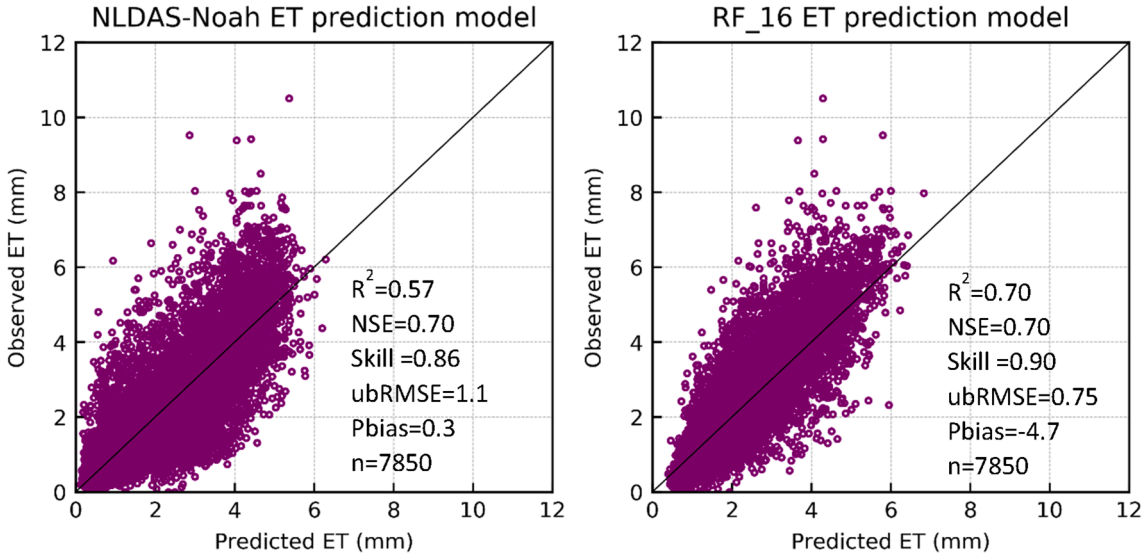

Fig. 7. Scatter plot of evaluation results of RF and NLDAS-Noah prediction models for sample size n = 7850 (30% of whole data). NLDAS-Noah is considered as a mechanistic benchmark model that is compared with the overall best model RF_16.

models. In shoulder months, since ET is lower, the coupling/interactions between soil moisture and ET is also lower. Overall models residuals were lower for the shoulder months of April, May, September and October and were in the range from 0.003 to 0.1 mm (overestimate of ET) while in peak warm months of June, July, and August, residuals range from −0.2 to −0.6 mm (underestimate of ET).

For the overall evaluation data set, RF_16 outperformed other models with the lowest AIC and R^2 of 0.7. The performance of the RF_11 was similar. RF_5 and LSTM_5 were the simplest version RF and LSTM, respectively, and produced the highest daily ubRMSE of 0.94–1.20 mm. As the model complexity reduced, ubRMSE and AIC error increased for both LSTM and RF and overall RF consistently outperformed LSTM. Supplementary materials include models results from training data (13 sites with 18,481 daily ET observations) and testing data (seven sites with 7850 daily ET observations) in irrigated and rain-fed fields and their comparison with benchmark model (Figs. S4–S7 in Supplementary

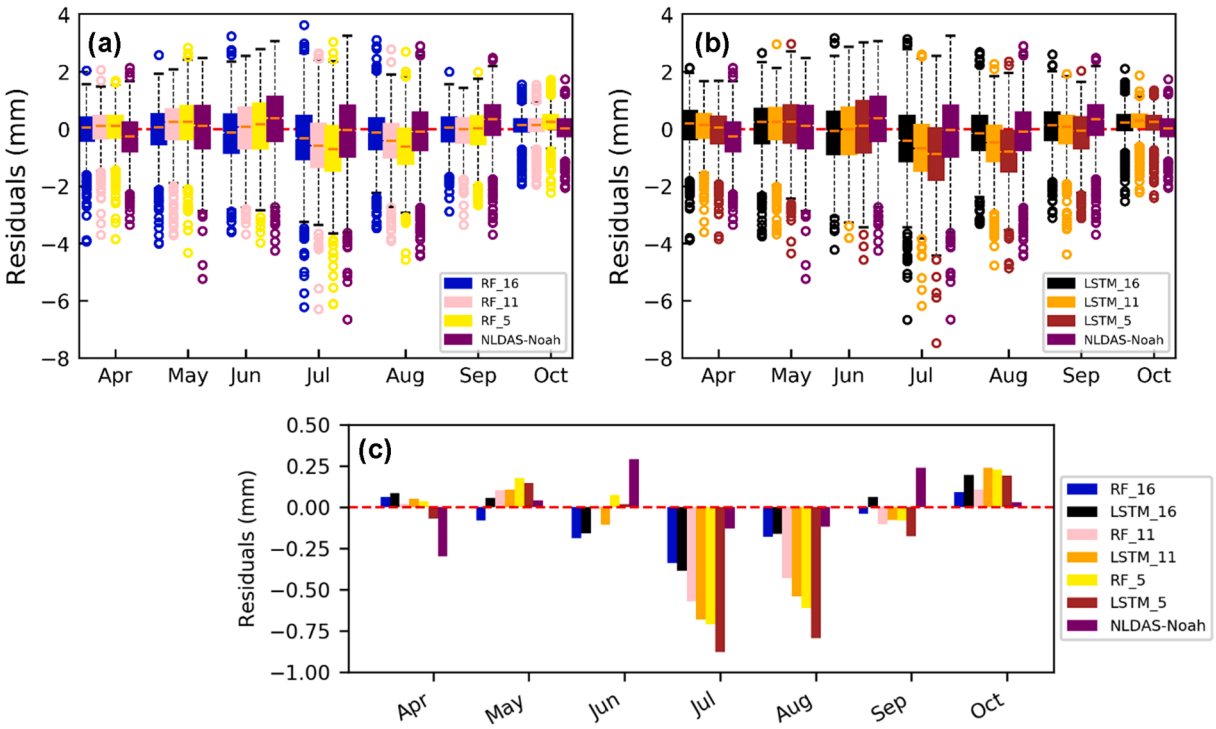

Fig. 8. Box plots for predicted ET residuals (simulated minus observed) for evaluation results of RF (a) and LSTM (b) prediction models for sample size n = 7850 (30% of whole data). RF and LSTM models are also compared with NLDAS-Noah. Median and the 25th and 75th percentiles are represented by boxes. The whiskers represent one and a half times the interquartile range (or ~ ±2.7σ). Circles show outliers. Figure (c) also shows the mean residual comparison between different versions of RF and LSTM models and NLDAS-Noah.

Materials). In addition, evaluation metrics for RF_16 (overall best model) are calculated for daily ET is each year of testing data in Table S2 in Supplementary Materials.

For non-irrigated crops, the predictors that improved RF_16 and RF_11 performance were similar and additional predictors such as soil texture, crop cover, crop coefficients, and cumulative GDD did not significantly improve the model performance of RF_11 and RF_16. However, this was not the case for the irrigated crops. Here, ET prediction was improved by including additional information related to physical properties of sites (soil types, crop coefficient, cumulative GDD) and relative AIC error reduced from 0.16 to zero (Fig. 9), making RF_16 the best model for irrigated crops. AIC score and R^2 were also computed for sites with different crops and soil texture. For all crop types, the simplest versions of models such as RF_5 and LSTM_5 (Fig. 10, Table 6) increased ubRMSE and AIC errors. Soybean and corn on fine-grained soils such as silty loams did not show an increase in R^2 or decrease in ubRMSE and AIC in models by including additional 5 parameters in RF 16 and LSTM 16 model. However, corn and soybeans on coarser soil such as loam showed improved performance with additional information about crop planting and harvest dates, cumulative GDD, and crop coefficients.

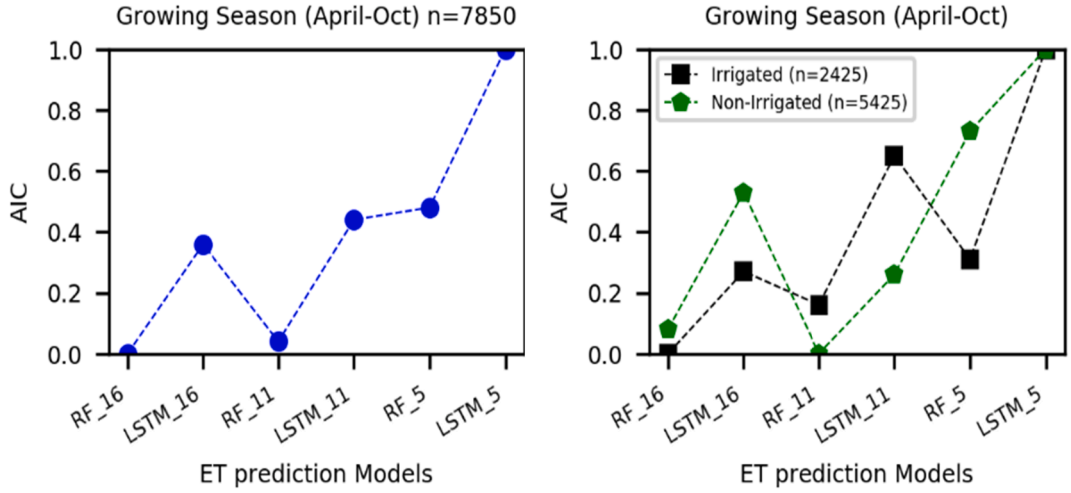

Fig. 9. AIC scores for different versions of prediction models on evaluation data. n represents the sample size. AIC score is normalized between 0 and 1 for comparison. First AIC was calculated for the whole data set (n = 7850) for different versions of prediction models. Then data are divided into irrigated and non-irrigated crops and AIC is calculated separately for irrigated and non-irrigated crops because of different sample sizes.

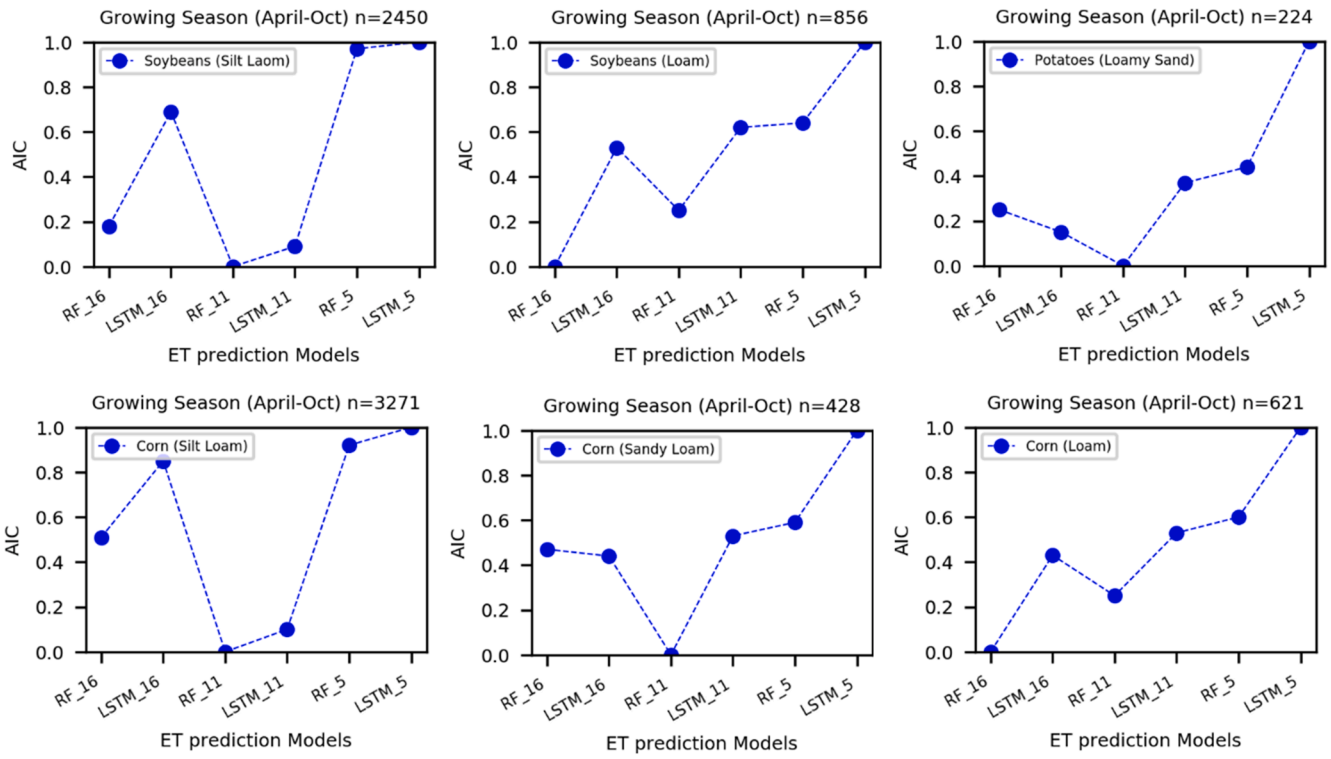

Fig. 10. AIC scores for different versions of prediction models on evaluation data based on soil types and crop types. n represents the sample size. AIC score is normalized between 0 and 1 for comparison to account for different parameter numbers among the models. Lower AIC means the model is more parsimonious than a model with higher AIC.

3.4. Forecast model results

The evaluation was performed for the retrospective period of 2003–2019 (Fig. 11). For both RF and LSTM, the overall ET estimate was comparable for day 1, day 2, and day 3 ET forecast. It is observed that as lead time increases, uncertainty and error in forecast increases but for proposed RF and LSTM models there was only a slight increase in MAE from 0.74 to 0.75 mm and from 0.75 to 0.80 mm (Table 7). The MAE for June, July, and August was higher, in concordance with higher variance on GEFS meteorology ensemble forecast spread (Fig. 11). This bias was more evident in LSTM models where ensembles estimates showed a wider spread from the mean estimate compared to RF. Overall, the RF forecast model produced results with high confidence (small ensemble standard deviation) compared to LSTM. RF was also more precise and less biased than the LSTM, for example, for day 3 ET forecasting MAE = 0.75 vs. 0.8 and Pbias = −4.1% vs. −5.1% (Table 7) . However, overall, the difference between the results of the two forecasting models is not significant (p value for two-tailed *t*-test is 0.2). Based on variable importance for RF forecast models (Fig. 11), VP and SolarZenith explained about 32% and 12% variance in the model. Other important model predictors include Crop_coeff, CumGDD, EVI, and SW_Day3. For LSTM removing day 2 and day 3 SW radiation reduced model NSE from 0.56 to 0.49 (Fig. 11) for day 3 ET, indicating significance of meteorological forecast predictors.

While the forecasts appear reliable, there are differences in soil type, climate conditions, and irrigation. RF and LSTM were consistent in prediction on sandy or loamy soils, but underpredicted ET on silty loam (Fig. 12). The performance of the daily ET forecast model decreases during extreme conditions. Fig. 13 showed that RF outperforms LSTM for ET forecast for day 3 for irrigated crops (RF NSE = 0.70 and Willmott's skill score = 0.91 vs LSTM NSE = 0.67 and Willmott's skill score = 0.90, p value 0.0001) and non-irrigated crops (RF NSE = 0.53 and Willmot' skill score of 0.81 versus LSTM NSE = 0.50 and Willmott's skill score of 0.80 for non-irrigated areas p value 0.07). The difference between RF and LSTM model performance was significant for irrigated sites.

Models performance was also tested for extreme events such as as floods and drought years. Fig. 14 showed that for 2012, a dry year with a flash drought, the difference between the model for day 3 ET forecast estimate is larger during days (July, August) with high temperatures and ET. Similarly, for the year 2017, a wet year, the model for day 3 ET forecast overestimated lower values (~1 mm) of ET. These analyses indicate that there are an under-estimation and over-estimation of the forecasted maximum and minimum values, respectively.

4. Discussion

4.1. Model evaluations

Overall, we found that empirical ML models can accurately and precisely predict ET across a range of crop and soil types in the upper Midwest USA, with R^2 and NSE equal to 0.70 and ubRMSE from 0.75 and 0.89 mm day^{-1} for RF_16 and LSTM_16 respectively. In general, different versions of RF models had higher R^2 and NSE and lower Pbias than the LSTM, except for irrigated potatoes in sandy loam. We suspect that this result is because we had data for only two growing seasons for irrigated potatoes, thus our results support that while RF can be more accurate, LSTM may be more useful when available data for model calibration is smaller. In addition, the prominent soil type for sites with irrigated potatoes (US-CS1 and US-CS3) is loamy sand, which stimulates rapid water movement through coarse grains after precipitation and irrigation. RF_16 could capture this pattern properly during months with high ET and irrigation during months when ET is higher but not during months with moderate or low ET, while LSTM_16 had a larger variance than the bias during such extreme events. This indicates that when irrigation and ET are higher (June and July), RF_16 had a higher bias, but lower error compared to LSTM_16. The high bias for RF_16 for that site is likely because of RF's greater sensitivity to the size of the training

Table 6
Daily ET prediction Models (RF_16, LSTM_16, RF_11, LSTM_11, RF_5, LSTM_5) performance for different soil types and crops based on R^2, ubRMSE and AIC. n = sample size.

Prediction Models	Crop Types	Soil Types	Sample Size (n)	R^2	NSE	Willmott's skill score	Pearson Corr.	MAE (mm/day)	ubRMSE (mm/day)	RSR	Pbias (%)	AIC
RF_16	Soybeans	Silt Loam	2450	0.63	0.63	0.87	0.79	0.65	0.76	0.61	−0.61	0.18
LSTM_16	Soybeans	Silt Loam	2450	0.56	0.56	0.85	0.75	0.73	0.89	0.66	3.5	0.69
RF_11	Soybeans	Silt Loam	2450	0.65	0.65	0.88	0.81	0.65	0.71	0.59	1.36	0.00
LSTM_11	Soybeans	Silt Loam	2450	0.61	0.61	0.85	0.79	0.79	0.80	0.62	0.97	0.09
RF_5	Soybeans	Silt Loam	2450	0.56	0.56	0.85	0.75	0.74	0.89	0.66	2.1	0.97
LSTM_5	Soybeans	Silt Loam	2450	0.48	0.48	0.80	0.69	0.79	1.1	0.72	−2.3	1.00
RF_16	Soybeans	Loam	856	0.84	0.84	0.95	0.92	0.51	0.43	0.4	−0.6	0.00
LSTM_16	Soybeans	Loam	856	0.75	0.75	0.91	0.87	0.65	0.68	0.5	2.0	0.53
RF_11	Soybeans	Loam	856	0.80	0.80	0.93	0.92	0.56	0.53	0.45	−4.6	0.25
LSTM_11	Soybeans	Loam	856	0.72	0.72	0.90	0.88	0.67	0.73	0.53	−5.6	0.62
RF_5	Soybeans	Loam	856	0.71	0.71	0.89	0.88	0.67	0.76	0.53	−4.9	0.64
LSTM_5	Soybeans	Loam	856	0.61	0.61	0.85	0.80	0.80	1.0	0.63	−7.8	1.00
RF_16	Potatoes	Loamy Sand	224	0.53	0.53	0.85	0.80	0.72	0.67	0.69	24	0.25
LSTM_16	Potatoes	Loamy Sand	224	0.56	0.56	0.80	0.80	0.69	0.74	0.66	−5.7	0.15
RF_11	Potatoes	Loamy Sand	224	0.58	0.58	0.84	0.79	0.70	0.68	0.65	13.1	0.00
LSTM_11	Potatoes	Loamy Sand	224	0.50	0.50	0.80	0.69	0.74	0.89	0.73	1.83	0.37
RF_5	Potatoes	Loamy Sand	224	0.42	0.42	0.7	0.65	0.77	0.99	0.76	1.45	0.44
LSTM_5	Potatoes	Loamy Sand	224	0.42	0.42	0.75	0.65	0.77	0.99	0.76	1.45	1.00
RF_16	Corn	Silt Loam	3271	0.70	0.70	0.90	0.85	0.68	0.86	0.56	−9.6	0.51
LSTM_16	Corn	Silt Loam	3271	0.66	0.66	0.88	0.82	0.78	1.0	0.59	−6.6	0.85
RF_11	Corn	Silt Loam	3271	0.69	0.69	0.88	0.96	0.70	0.88	0.56	−10.7	0.00
LSTM_11	Corn	Silt Loam	3271	0.62	0.62	0.85	0.83	0.78	1.1	0.62	−12	0.10
RF_5	Corn	Silt Loam	3271	0.63	0.63	0.86	0.84	0.76	1.0	0.61	−10.4	0.92
LSTM_5	Corn	Silt Loam	3271	0.53	0.53	0.8	0.79	0.87	1.3	0.68	−15	1.00
RF_16	Corn	Sandy Loam	428	0.66	0.66	0.88	0.82	0.51	0.45	0.58	−5.1	0.47
LSTM_16	Corn	Sandy Loam	428	0.66	0.66	0.88	0.82	0.53	0.45	0.58	−2.8	0.44
RF_11	Corn	Sandy Loam	428	0.7	0.7	0.9	0.84	0.49	0.41	0.55	2.9	0.00
LSTM_11	Corn	Sandy Loam	428	0.64	0.64	0.88	0.81	0.53	0.47	0.60	5.3	0.53
RF_5	Corn	Sandy Loam	428	0.63	0.63	0.87	0.80	0.54	0.50	0.61	1.6	0.59
LSTM_5	Corn	Sandy Loam	428	0.58	0.58	0.85	0.76	0.59	0.57	0.65	−0.42	1.00
RF_16	Corn	Loam	621	0.76	0.76	0.92	0.88	0.61	0.65	0.49	−5.1	0.00
LSTM_16	Corn	Loam	621	0.68	0.68	0.89	0.83	0.75	0.87	0.56	1.4	0.43
RF_11	Corn	Loam	621	0.71	0.71	0.89	0.88	0.66	0.75	0.54	−9.8	0.25
LSTM_11	Corn	Loam	621	0.65	0.65	0.86	0.84	0.75	0.92	0.59	−7.3	0.53
RF_5	Corn	Loam	621	0.63	0.63	0.85	0.85	0.75	0.97	0.61	−10.3	0.60
LSTM_5	Corn	Loam	621	0.52	0.52	0.79	0.77	0.87	1.3	0.69	−11	1.00

sample. The high bias of RF_16 can also be seen in Iowa (corn and soybeans rotation, loamy soil), Minnesota (corn, soybean rotation silty loam), and Michigan (Corn, Sandy loam). Thus, while RF models outperform LSTM for crop ET, they require more training data. Time series analysis of observed and predicted ET values (Fig. 4) shows that data-driven models are well trained for predicting the daily data. Hence errors in reproducing the daily anomalies are smaller when compared to the errors in the seasonal cycle because of their relative amplitude.

For the evaluation period of different versions of RF and LSTM, residuals (simulated-observed) are roughly normally distributed during the growing season. However, negative residuals in range of −0.25 to −0.75 for different versions of models during peak ET months (July, Aug) showed an underestimation of ET. This difference may also be due to errors in the input data from different sources, or complexities that the model cannot explain e.g., more irrigation during the dry year or not capturing fluxes through the root zone of fine-grained and coarse-grained soils. Our current models do not have irrigation data as a predictor so including it in future research can be useful. Our work is consistent with earlier studies on using ML to estimate water cycle elements. For example, Kratzert et al. (2019) used LSTM in an ungauged basin (with an aridity index from 0.22 to 5.20) to estimate stream flow using static predictors (e.g soil, geology, water content, max LAI) and non-static parameters (e.g precipitation, temperature, solar radiation) and found that ML models can be useful to predict information by extracting complex relationships between diverse data under heterogeneous condition.

4.1.1. Model complexity

The complexity of an RF tree grows with an increase in the number of trees in the forest as well as the number of training samples. Hence a simple RF tree with small training samples could not account for variability in the potatoes ET. In RF we also limit the number of variables to split on in each split that can lead to higher bias in each tree especially when the sample size is small.

Among the ML models, we found that the best overall model to be RF_16. We also observed that more than 300 decision trees for the RF model only improved the accuracy of training data but did not show significant improvement in model accuracy on testing data, and instead

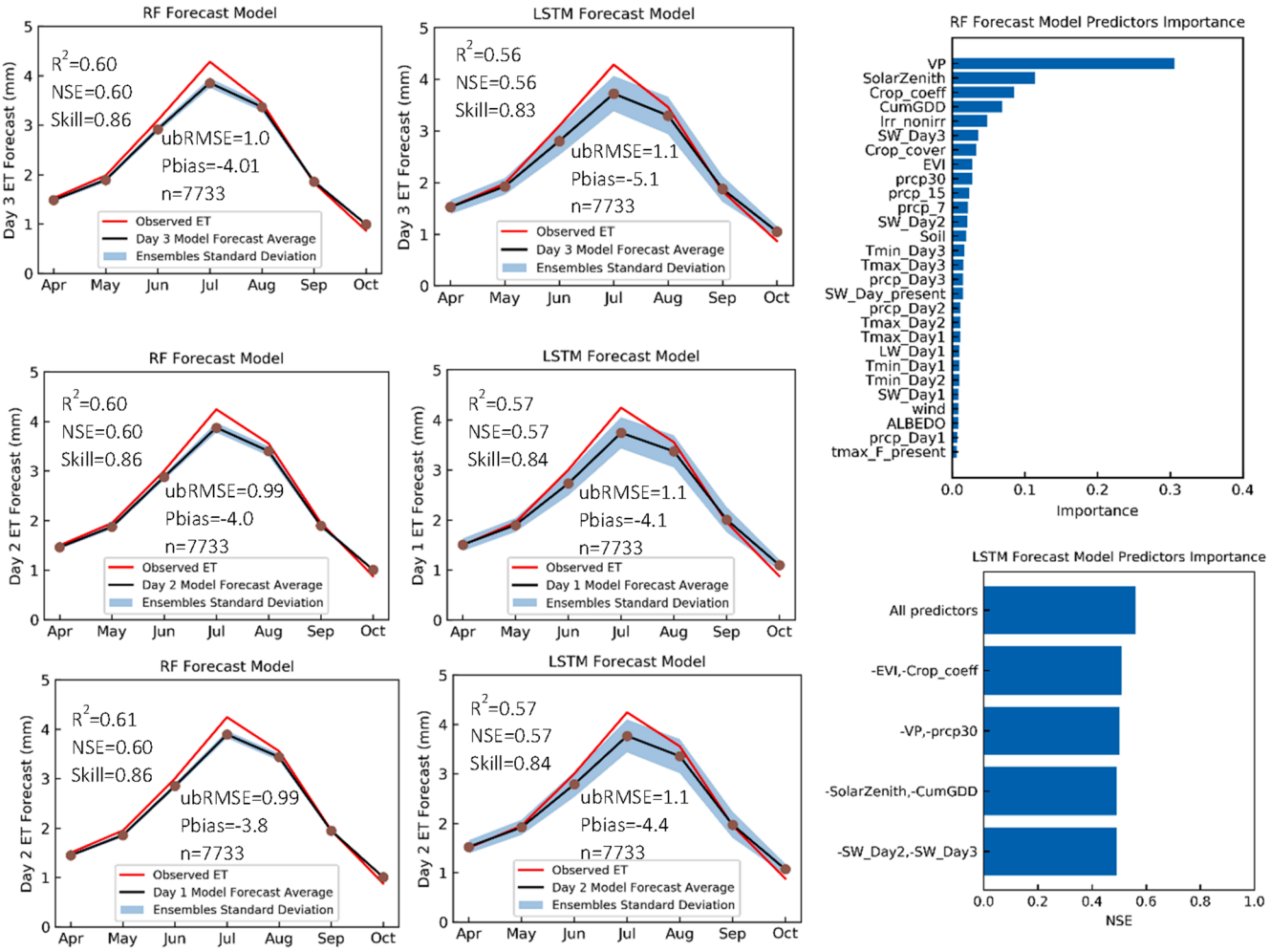

Fig. 11. RF and LSTM comparison for day1,2,3 ET forecast on evaluation data. ET values are averaged by month for visualization of ensemble spread for each month. Evaluation statistics are calculated on daily ET forecast estimate. Graphs on the right column show RF models and the graph in the middle column shows LSTM models. Variable importance is shown for day 3 ET forecast in the graph on the right column for RF and LSTM. The effect of only important variable removal on NSE was shown for the LSTM model, while the importance of all predictors is shown for the RF forecast model.

Table 7
Comparison of RF and LSTM model for day1, day2 and day 3 ET forecast. Models are evaluated on 7733 daily ET observations (2003–2019) from seven sites in Midwest. Model evaluation statistics are calculated on daily timestep.

Model Evaluation Statistics for prediction model	R^2	NSE	Willmott's skill score	Pearson Corr.	MAE	ubRMSE (mm/day)	RSR	Pbias (%)
RF_day 1	0.61	0.61	0.86	0.78	0.75	0.99	0.63	−3.8
RF_day 2	0.60	0.60	0.86	0.78	0.75	0.99	0.63	−4.0
RF_day 3	0.60	0.60	0.86	0.78	0.75	1.01	0.63	−4.1
LSTM_day 1	0.57	0.57	0.84	0.76	0.79	1.1	0.65	−4.4
LSTM_day 2	0.57	0.57	0.84	0.76	0.79	1.1	0.65	−4.1
LSTM_day 3	0.56	0.56	0.84	0.76	0.80	1.1	0.66	−5.1

only made the proposed approach computationally more intensive. The performance of the RF_16 model is comparable to the process-based NLDAS-Noah model and needs relatively fewer parameters and drivers to estimate ET. The inner structure of RF allows the model to explain the non-linear relationships among ET and important predictors such as EVI, solar zenith angle and incoming SW radiation. RF_16 models outperformed other smaller parameter number models for most of the locations except at corn and soybeans with silt loam soil texture and potatoes in sandy soils. At those locations, we found that a simple version of RF (RF_11) performed better at those locations as well as for non-irrigated crops. For these sites, the complex models (with 16 predictors) were overfitting on training data. In other words, for these crop and soil combinations, an overall simpler model was able to learn the appropriate non-linear relationship and memory (in the case of LSTM) between predictors and memory. Hence, we can expect the performance of LSTM and RF to decline when models are trained on drivers beyond the leading predictors of a hydrologic system. Tennant et al. (2020) observed this decline in performance in the LSTM discharge prediction model in snow-dominated catchment when trained on additional predictors.

Another reason for the divergence in model performance among sites may be related to the observation that irrigated crops have high variability in ET e.g. based on summary statistics in Table 2, irrigated crops in US-Ne2 have maximum daily ET at a higher end (e.g., ~9 mm day^{-1}) with sample variance more than 3 mm compared to non-irrigated crops. Although, we did not observe this high variability in irrigated potatoes

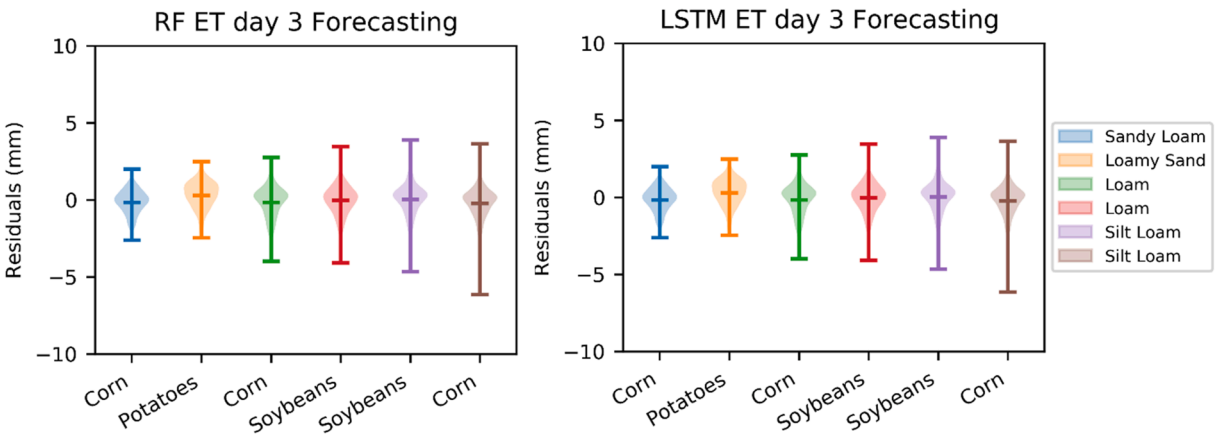

Fig. 12. Day 3 ET forecast for corn, soybeans, and potatoes under different soil textures for evaluation RF and LSTM. Different colors show the combination of crop types with soil textures.

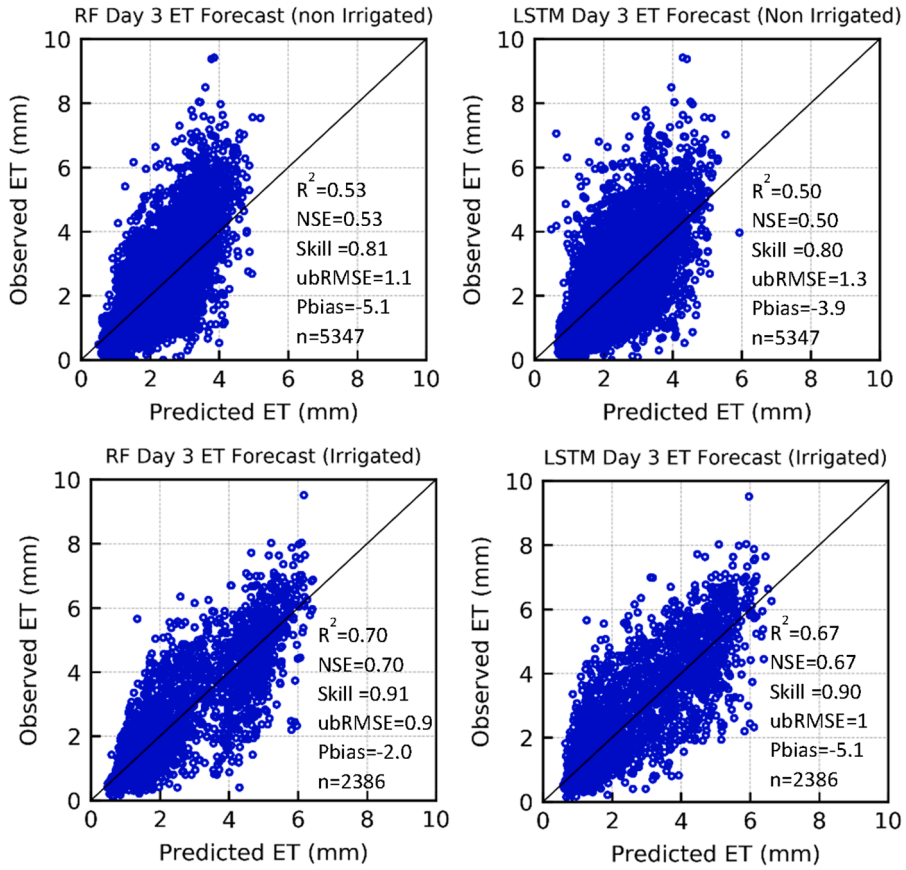

Fig. 13. Comparison of R^2, MAE and Pbias for day 3 ET forecast results. Both RF and LSTM models were tested in irrigated and non-irrigation crops. n represents the sample size of evaluation data.

in US-CS1 and US-CS3 and irrigated corn in MI sites (US-JCK, Jackson 1), because available data were only from wet years of 2018 and 2019. In addition, when water is sufficient or close to sufficient, the importance of additional predictors such as storage capacity (soil texture) and crop phenology (crop coefficients) become stronger and have a critical role in predicting ET. However, this effect is masked when irrigation is not available, or soil water storage is relatively low in non-irrigated crops (Seneviratne et al., 2010).

When predictors were reduced to only 5, both RF and LSTM performance contained large errors, limiting their utility. This outcome showed the importance of wind speed, solar zenith angle, maximum temperature, albedo, and 30 days average precipitation (as soil moisture proxy) that were excluded in the RF_5 and LSTM_5. Oliveira et al. (2018) also noted that surface energy fluxes that drive ET depend on rainfall and soil moisture, and albedo's influence on net radiation estimates. Thus, we argue that our 11-parameter model is the baseline minimum inputs required to predict ET across a range of crop, soil, and irrigation types. This also suggests that a number of predictors lower than 11 could not explain the variance in ET and it is possible for some sites to build more robust models with 11 predictors instead of 16. However, it's worth noting that the improvement of performance in LSTM and RF is not just from more parameters, but also the more complex models

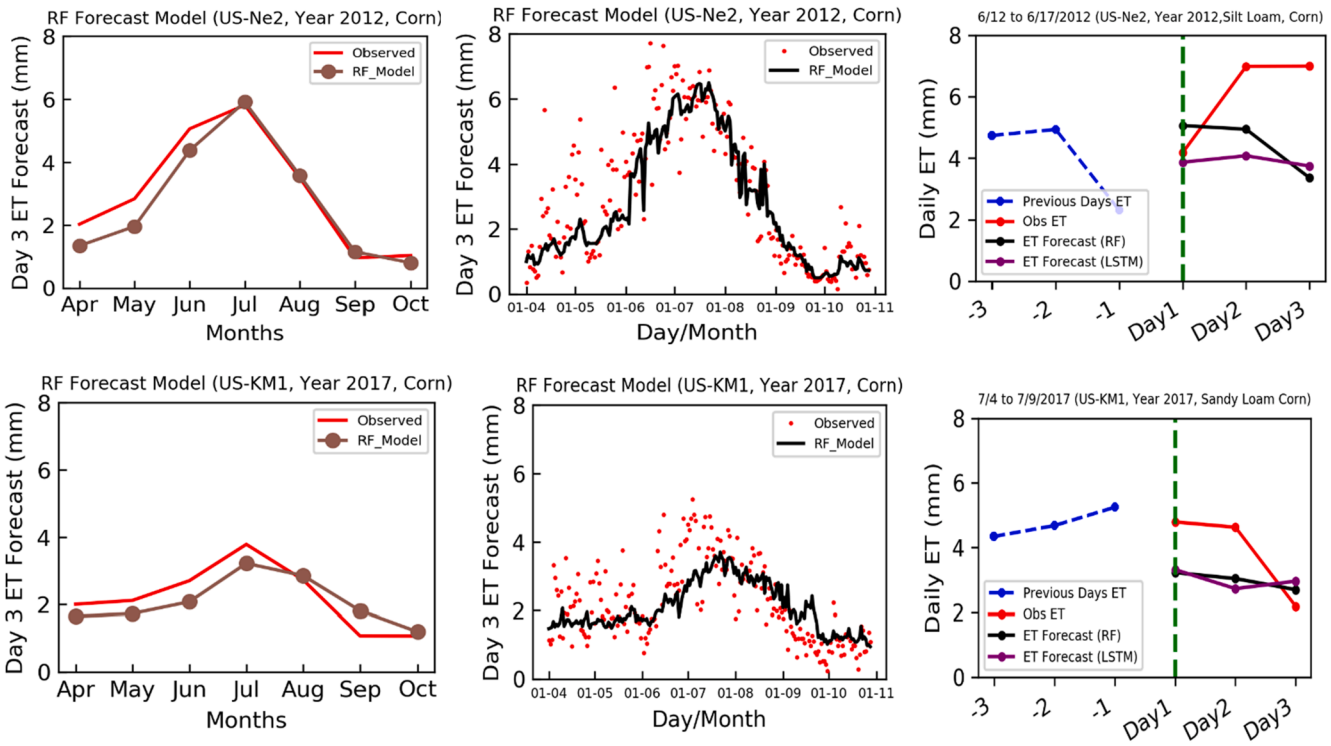

Fig. 14. RF and LSTM model Day 3 ET forecast performance on evaluation data for a dry year (2012) with flash drought and wet year (2017). Graphs on the top and bottom right show the performance of RF and LSTM models for forecasting day1, day 2, and day 3 ET in 2012 (US-Ne2) and 2017 (US-KM1). Graphs in the middle show comparison based on daily data for growing season and graphs on left shows average monthly ET.

include a greater number of hidden neurons, in the case of LSTM, or decision trees subsets and nodes in the case of RF. The additional elements provide an additional benefit over easily implementable regression-based models that cannot account for the non-linear interactions among the predictors (e.g. temperature) and ET. For example, Chen et al. (2020a) found that temperature and humidity-based ML models (RF and LSTM) outperformed temperature and humidity based empirical models in areas with limited meteorological data.

Compared to other techniques for ET estimation, the advantage of the proposed ML modeling approach is that these models monitor ET by using fewer parameters and do not rely on the accurate parameterization of mechanistic models or collections of labor-intensive field-scale data (e.g. field-scale leaf area index). However, care must be taken in appropriate model selection because the models are location-dependent and require sufficient calibration and testing data. For example, for soybeans and corn in silt loam, corn with sandy loam soil texture, and potatoes with loamy and sandy texture a comparable level of ET prediction performance can be achieved without using additional parameters about crop coefficients, crop cover, or CumGDD. Hence, ET can be predicted by the readily available biophysical predictors for such locations, in contrast to ET prediction for corn and soybeans with loam soil texture, where model performance is improved by including those biophysical parameters. The importance of cumGDD in daily ET prediction is encouraging as it is readily derived from low-frequency temperature observations and more readily available across more sites than soil moisture.

We found good performance using the same crop coefficients for irrigated and non-irrigated crops. Depending upon the objective and availability of data for a study, different models can be built for a specific crop type and soil texture at a daily time step.

4.2. Significant predictors

The predictors importance of the RF model (Fig. 5) highlights driving predictors and combats with the black-box nature of some ML models. Our study showed that EVI, solar zenith angle, incoming SW radiation, and cumulative growing degree days are important predictors to predict daily ET for the growing season (April-October) in the Midwest. Similarly, studies based on empirical models (Priestley and Taylor, 1972; Jensen et al., 1990) and data-driven ML framework (Chen et al., 2020a) evaluated that most of the variation in reference ET can be explained by solar radiation. This result is consistent with our study where incoming solar radiation explaind about 10%-20% variance for irrigated and non-irrigated crops. However, in our study, an additional variation of about 20% was explained by other variables such as EVI and crop coefficient. Zhao et al. (2018) also found that crop coefficients not only correlate with canopy development but also controls seasonal ET partitioning and surface soil moisture. This shows the importance of variable crop coefficients and EVI in predicting ET.

Noting that LSTM_5 and RF_5 residuals were high especially during peak ET months models suggests that wind speed, albedo, and EVI are leading factors that promote enhanced ET. For example, the potential of plants to extract water from soil depth varies during different stages of crop growth, so we can surmise through the lower residuals that it was captured by 11 and 16 predictors model versions. EVI has been used for agricultural drought monitoring (Song and Ma, 2011) and the results of this study also suggest the potential of EVI and ET as good indicators of short-term and long-term drought.

Our work is consistent with earlier studies to estimate ET. For example, Cobaner (2011) used fuzzy inference system-based grid partition to estimate reference ET in the moderate Mediterranean climate of California and found that solar radiation, air temperature, and relative humidity as important drivers for ET prediction. Our model was built to estimate daily actual ET for agricultural lands and found that EVI (non-irrigated crops), crop coefficients, and VP (irrigated crops) to be better predictors than solar radiation in Midwest humid-temperate climate.

Feng et al. (2017) used temperature-based RF and generalized

regression neural networks (GRNN) to estimate reference ET and found that RF outperformed GRNN. They also noted that without using solar radiation temperature-based RF and GRNN underestimated reference ET. Our model also found that incoming SW radiation was a more important driver than Tmax for actual ET. Walls et al. (2020) used RNN model based on ReLU and sigmoid activation function to estimate actual ET and found that without net radiation, model performance goes down. In our study, we found incoming SW radiation explained higher variance for ET compared to LW radiation, and thus net radiation could be omitted. In terms of RF and LSTM comparison for other hydrological variables, such as snowfall retrievals from microwave humidity sounders, Adhikari et al. (2020) found that RF is more robust than LSTM.

Chen et al. (2020b), which developed an LSTM based actual ET prediction model irrigated maize/corn, found that leaf area index, relative humidity, and solar radiation as important features that drive corn dynamics in a continental monsoon climate. Those predictors are in agreement with physical processes that can affect corn ET. We also found that VP and crop coefficients were more important predictors for irrigated crops compared to non-irrigated, while incoming SW radiation explained more variation in non-irrigated compared to irrigated crops. Irrigation influences surface temperature, convection, cloud formation (Lohar and Pal 1995), and humidity (Jianping et al. 2002). In irrigated crops, additional water vapor (Boucher et al., 2004) in the atmosphere due to evaporation of irrigated water can explain why vapor pressure is an important driver for irrigated crops, while less surface cooling in non-irrigated land can make incoming SW radiation important driver for those sites.

Since our study had shown that EVI is the most important variable for rain-fed crops, the uncertainty of EVI and associated parameters used in other models (e.g. for deriving leaf area index, LAI) will greatly affect ET estimation/mapping across the globe and improvement in estimating LAI can improve hydrologic and land surface models for ET mapping. Thus, methods to reduce uncertainty in EVI can improve remote sensing estimate of ET (Sharma et al., 2016).

We also found that soil texture is important in improving ET estimation in irrigated fields, which suggests the use of soil texture maps for ET estimation in ET mechanistic models in addition to soil moisture as a limiting factor. Dong et al. (2020) showed that soil moisture and ET coupling strength bias is caused by oversimplification of soil texture effects on soil evaporation stress. A data-driven based hydrodynamic prediction model can benefit from data sets of appropriate temporal and spatial coverage, readily available meteorological, biophysical variables, and advanced RNN such as LSTM (Kratzert et al., 2019) as well as robust simple ensemble tree-based RF algorithms.

4.3. Forecast models evaluations

We found that RF and LSTM framework can be used for forecasting for three days in advance using gridded forecast meteorology. Based on our hindcast analysis, the RF forecast model provided higher accuracy overall than LSTM, consistent with prediction model evaluation. LSTM forecast model was more sensitive to GEFS meteorology ensembles, where a higher spread from mean forecast ET was observed compared to the RF forecast model and RF can handle multivariate dimensionality (Belgiu and Drăguţ, 2016) better than RNN.

ML-based actual ET forecasts are a novel contribution of our research here and demonstrate significant performance across multiple irrigated and non-irrigated crops and soil texture. Short-term ET forecasts have value for irrigation planning considering under-irrigation and over-irrigation can be detrimental for crops and local water supply quantity and quality. We find that vapor pressure, solar zenith angle, and third-day forecasted incoming SW radiation are important predictors for accurate and precise ET forecasts. Ferreira and da Cunha (2020) used similar meteorological predictors (maximum air, solar radiation) for multistep forecasting of reference ET and found that deep learning models such as LSTM performed better than classic machine learning models. This is because LSTM process input in its sequential order and overcomes the problem of learning lagged dependencies. In addition, connections between neurons, that allow data to move in forward and backward direction within the modeling framework of LSTM and helps to learn temporal dependencies. Perera et al. (2014) used numeric prediction output for reference ET forecast in Australia and found that forecasting based on air and dew point temperatures leads to better performance for all lead times compared to incoming SW radiation and attributed the poor performance of incoming SW radiation to error forecast weather meteorology. Our study found incoming SW radiation (forecast) a more important predictor compared to day air temperature for actual ET forecast at all lead time. Higher ET during dry seasons showed that water was not limited due to irrigation.

Yin et al. (2020) applied bi-directional LSTM (Bi-LSTM) to forecast short term reference ET (one day lead time) in areas with limited meteorological data by using inputs of maximum, minimum temperature, sunshine duration and observed that sunshine duration has a higher correlation with reference ET than solar radiation. Hence including the sunshine duration in the forecast model can improve model accuracy. This study was also able to the ability of Bi-LSTM to represents the temporal variability of reference ET over the year.

For our study, in terms of accuracy, the forecast showed a greater skill for irrigated crops compared to non-irrigated crops. We also found higher accuracy for coarse-grained soils (sandy loam, loam). Results suggest that developed forecasting models are promising for simulating ET in the growing season, but the methods need to be improved for fine-textured and non-irrigated conditions. The performance of ET forecasting can be improved by selecting appropriate meteorological parameters as the input features of the model. At the same time, ET had strong regional characteristics such as different accuracy for different soil types. Future work will involve testing how such forecasts could be directly implemented for irrigation management and what changes can be made to reduce model bias.

4.4. Limitations and future directions

ML models such as RF and LSTM models show better generalization than linear models and can perform well in space and time compared to one-layer ANNs or autoregressive models (Fang et al., 2017). While ML models are useful for ET modeling, they have limits. For example, the models here are locally calibrated. While the calibration was pooled across multiple crop and soil types, it is possible that some combinations of crops and soils were not well trained and could lead to inaccurate prediction of ET at those locations. Significant training data is a limitation to the ML models. Long-term climatic data can help data-driven models to extract the climatic cycle influence on ET. Hence models developed on those domains with long-term flux tower locations would be more reliable to predict ET and less sensitive to uncertainty than those regions with shorter-term and fewer ET data. In cases with limited training data, mechanistic models do have an indisputable advantage of estimating hydrological variables for any set of inputs as long as the limitations and assumptions of the model are valid.

In terms of parameters, one limitation of our proposed model is the lack of root zone water dynamics. For example, when soils have enough water stored in them during the wet year, actual ET under non-irrigated conditions is assumed to be equal to the potential crop ET. However, during dry conditions, limited soil water storage is often observed, which can reduce actual ET, and plant ET is more a function of soil moisture. We also observed that soil moisture proxy predictors (in form of prcp 7 and prcp 30) were of particular importance for non-irrigated crops. This could be because spells of heat waves during dry years (e.g., 2012, 2010) can lead to a more rapid decline in soil moisture in non-irrigated sites compared to irrigated sites. ML models also tend to perform poorly on extrapolation to conditions not observed in the data or during extreme or rare events. We saw these results in Fig. 14 for extreme events in a dry year (2012) and a wet year (2017). The tendency

of all ML models to "regress to the mean" limits their usefulness in flash drought or flooding type conditions that may become more prevalent with ongoing anthropogenic climate change. In addition, Gupta et al. (2009) also found that this result is more expected when using MSE as a calibration objective function.

The future application of LSTM and RF models will be catalyzed with the availability of more data under more conditions. There is also promising research in improving the representation of processes within ML, using reinforcement learning or physical constraint type approaches (Zhao et al., 2019a, 2019b). For example, it is possible to add physical properties to account for subsurface dynamics by including an additional input layer of tree nodes. Even though the proposed model does not have a representation of water balance, it is possible to link neurons and trees to atmospheric and hydrological patterns, such as heat fluxes, so that water is conserved and allowing for less realistic ET estimates to be rejected. However, this might come at a cost of requiring more input predictors that must be derived from data products that may or may not be available. It is also possible to physically constraint ML models (O'Gorman and Dwyer, 2018; Zaherpour et al., 2019; Zhao et al., 2019b), which can help to conserve energy budget while accounting for physical transport processes of water vapor, leading to a better generalization of physical processes during extremes. Camporeale (2019) also underscores the need to do more research into probabilistic-based uncertainty estimates and the development of gray box models by combining mechanistic and ML approaches.

It will also be useful to collect more data from other climate regimes, crops, and soil types that can help us understand if the conclusions found here and related papers can be generalized to other regions and other crops. This can be used to study the scale- and location- dependence of the drivers on ET and help improve ET forecasting in regional scales.

5. Conclusion

We proposed a new framework based on a machine learning data-driven network to estimate and forecast ET and its uncertainty for corn, soybeans, and potatoes under different soil texture types in agricultural areas of the Midwest, USA. The model was built by using biophysical and meteorological information acquired from ground observations and satellite sensor data. The data sets used in the proposed model have been widely utilized in many studies for ET prediction and related to ancillary data used in hydrological models such as SWAT and HSPF. The proposed model was calibrated using 13 field-based eddy covariance ET time series distributed across the region for the period of 2003–2019 for irrigated and rainfed agricultural areas in the Midwest. The model was evaluated in seven independent locations for the time period of 2003–2019.

The evaluation results based on observed ET measurements collected from seven different sites confirmed that the predicted models can be used for daily ET estimates with ubRMSE from 0.67 to 0.92 mm, Willmott's skill score from 0.80 to 0.90 and simulate the spatial heterogeneity of agricultural parameters and dynamics of water use by crops. The prediction model estimates were reliable and on par with mechanistic model estimates from NLDAS. The results of this study also revealed that the inclusion of EVI, solar zenith angle, incoming SW radiation, and CumGDD were the most important input predictors. Vapor pressure was of greater importance for forecasting future ET. The proposed model can also be applied to both rainfed and irrigated crop types. Overall, our work supports the use of ML, especially random forest approaches for prediction and short-term forecasting of ET in both rainfed and irrigated crops, which had a range of valuable uses for irrigation management and water cycling evaluation. Expanding this work outward to tropical or semi-arid regions may require further evaluation of additional predictors, but overall, the results here find that a general field-scale regional ET model is realizable across a range of soil characteristics and climatic patterns. ET prediction and forecasting by using this modeling framework can help policy makers to allocate water sustainability for irrigation and assist growers to spot water stress areas in farms.

CRediT authorship contribution statement

Ammara Talib: Conceptualization, Methodology, Writing - original draft. **Ankur R. Desai:** Visualization, Investigation, Writing - review & editing. **Jingyi Huang:** Visualization, Investigation, Writing - review & editing. **Tim J. Griffis:** Visualization, Investigation, Writing - review & editing. **David E. Reed:** Visualization, Investigation, Writing - review & editing. **Jiquan Chen:** Visualization, Investigation, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgements

ARD and AT acknowledge support from the Wisconsin Vegetable and Potato Growers Association award to UW-Madison, the Wisconsin Department of Natural Resources, and the UW Center for Climatic Research Climate, People, and Environment Program, and thank J Thom, T Houlihan, J Pavelski for site support at US-CS1 and US-CS3. DR and JC are supported by the NASA Carbon Cycle & Ecosystems program (NNX17AE16G).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2021.126579.

References

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

Adhikari, A., Ehsani, M.R., Song, Y., Behrangi, A., 2020. Comparative Assessment of snowfall retrieval from microwave humidity sounders using machine learning methods. Earth Space Sci. 7 (11) https://doi.org/10.1029/2020EA001357.

Akaike, H., 1970. Statistical predictor identification. Ann. Inst. Stat. Math. 22 (1), 203–217. https://doi.org/10.1007/BF02506337.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration: guidelines for computing crop requirements. Irrig. Drain. Pap. No. 56, FAO. https://doi.org/10.1016/j.eja.2010.12.001.

Anandhi, A., 2016. Growing degree days – Ecosystem indicator for changing diurnal temperatures and their impact on corn growth stages in Kansas. Ecol. Indic. 61, 149–158. https://doi.org/10.1016/j.ecolind.2015.08.023.

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

Anderson, M.C., Yang, Y., Xue, J., Knipper, K.R., Yang, Y., Gao, F., Hain, C.R., Kustas, W. P., Cawse-Nicholson, K., Hulley, G., Fisher, J.B., Alfieri, J.G., Meyers, T.P., Prueger, J., Baldocchi, D.D., Rey-Sanchez, C., 2021. Interoperability of ECOSTRESS and Landsat for mapping evapotranspiration time series at sub-field scales. Remote Sens. Environ. 252, 112189. https://doi.org/10.1016/j.rse.2020.112189.

Auret, L., Aldrich, C., 2012. Interpretation of nonlinear relationships between process variables by use of random forests. Miner. Eng. 35, 27–42. https://doi.org/10.1016/j.mineng.2012.05.008.

Baker,J., Griffis, T., 2017. AmeriFlux US-Ro5 Rosemount I18_South, Dataset. https://doi.org/10.17190/AMF/1419508.

Baker,J., Griffis,T., 2003-2017a. AmeriFlux US-Ro1 Rosemount- G21, Dataset. https://doi.org/10.17190/AMF/1246092.

Baker,J., Griffis ,T., 2003-2010. AmeriFlux US-Ro3 Rosemount- G19, Dataset. https://doi.org/10.17190/AMF/1246093.

Baker,J., Griffis,T., 2003-2017b. AmeriFlux US-Ro2 Rosemount- G21, Dataset. https://doi.org/10.17190/AMF/1418683.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, U.K.T., Pilegaard, K., Schmid, H.P., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001.

FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. Bull. Am. Meteorol. Soc. https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2.

Barr, C., Tibby, J., Gell, P., Tyler, J., Zawadzki, A., Jacobsen, G.E., 2014. Climate variability in south-eastern Australia over the last 1500 years inferred from the high-resolution diatom records of two crater lakes. Quat. Sci. Rev. 95, 115–131. https://doi.org/10.1016/j.quascirev.2014.05.001.

Barr, A.G., van der Kamp, G., Black, T.A., McCaughey, J.H., Nesic, Z., 2012. Energy balance closure at the BERMS flux towers in relation to the water balance of the White Gull Creek watershed 1999–2009. Agric. For. Meteorol. 153, 3–13. https://doi.org/10.1016/j.agrformet.2011.05.017.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Netw. 5 (2), 157–166.

Bodesheim, P., Jung, M., Gans, F., Mahecha, M.D., Reichstein, M., 2018. Upscaled diurnal cycles of land-Atmosphere fluxes: A new global half-hourly data product. Earth Syst. Sci. Data. https://doi.org/10.5194/essd-10-1327-2018.

Boucher, O., Myhre, G., Myhre, A., 2004. Direct human influence of irrigation on atmospheric water vapour and climate. Clim. Dyn. 22 (6-7), 597–603. https://doi.org/10.1007/s00382-004-0402-4.

Breiman L., 2001. Machine Learning, 45(1), 5–32. Stat. Dep. Univ. California, Berkeley, CA 94720. https://doi.org/10.1023/A:1010933404324.

Breiman, L., 2002. Manual on setting up, using, and understanding random forests v3. 1. Tech. Report, http//oz.berkeley.edu/users/breiman, Stat. Dep. Univ. Calif. Berkeley, …. https://doi.org/10.2776/85168.

Camporeale, E., 2019. The challenge of machine learning in space weather: nowcasting and forecasting. Sp. Weather. 17 (8), 1166–1207. https://doi.org/10.1029/2018SW002061.

Carriere, P., Mohaghegh, S., Gaskari, R., 1996. Performance of a virtual runoff hydrograph system. J. Water Resour. Plan. Manage. 122 (6), 421–427. https://doi.org/10.1061/(ASCE)0733-9496(1996)122:6(421).

Chen,J., Chu, H., 2011-2013. AmeriFlux US-CRT Curtice Walter-Berger cropland, Dataset. https://doi.org/10.17190/AMF/1246156.

Chen, F., Dudhia, J., 2001. Coupling and advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. Mon. Weather Rev. https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V., Duan, Q.Y., Ek, M., Betts, A., 1996. Modeling of land surface evaporation by four schemes and comparison with FIFE observations. J. Geophys. Res. Atmos. 101 (D3), 7251–7268. https://doi.org/10.1029/95JD02165.

Chen, Z., Zhu, Z., Jiang, H., Sun, S., 2020a. Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. J. Hydrol. 591, 125286. https://doi.org/10.1016/j.jhydrol.2020.125286.

Chen, Z., Sun, S., Wang, Y., Wang, Q., Zhang, X., 2020b. Temporal convolution-network-based models for modeling maize evapotranspiration under mulched drip irrigation. Comput. Electron. Agric. 169, 105206. https://doi.org/10.1016/j.compag.2019.105206.

Chen,J., 2018c. AmeriFlux Dataset https://ameriflux.lbl.gov/sites.

Cleland, E., Chuine, I., Menzel, A., Mooney, H., Schwartz, M., 2007. Shifting plant phenology in response to global change. Trends Ecol. Evol. 22 (7), 357–365. https://doi.org/10.1016/j.tree.2007.04.003.

Cobaner, M., 2011. Evapotranspiration estimation by two different neuro-fuzzy inference systems. J. Hydrol. 398 (3-4), 292–302. https://doi.org/10.1016/j.jhydrol.2010.12.030.

Crosbie, R.S., Davies, P., Harrington, N., Lamontagne, S., 2015. Ground truthing groundwater-recharge estimates derived from remotely sensed evapotranspiration: a case in South Australia. Hydrogeol. J. https://doi.org/10.1007/s10040-014-1200-7.

Desai.A., 2018-2019. AmeriFlux US-CS1 Central Sands Irrigated Agricultural Field, Dataset. https://doi.org/10.17190/AMF/1617710.

Desai.A., 2019-2020. AmeriFlux US-CS3 Central Sands Irrigated Agricultural Field, Dataset. https://doi.org/10.17190/AMF/1617713.

Djaman, K., Smeal, D., Koudahe, K., Allen, S., 2020. Hay yield and water use efficiency of alfalfa under different irrigation and fungicide regimes in a semiarid climate. Water (Switzerland) 12 (6), 1721. https://doi.org/10.3390/w12061721.

Dong, J., Dirmeyer, P.A., Lei, F., Anderson, M.C., Holmes, T.R.H., Hain, C., Crow, W.T., 2020. Soil evaporation stress determines soil moisture-evapotranspiration coupling strength in land surface modeling. Geophys. Res. Lett. 47 (21) https://doi.org/10.1029/2020GL090391.

Donohue, R.J., McVicar, T.R., Roderick, M.L., 2010. Assessing the ability of potential evaporation formulations to capture the dynamics in evaporative demand within a changing climate. J. Hydrol. 386 (1-4), 186–197. https://doi.org/10.1016/j.jhydrol.2010.03.020.

Fang, Wei, Huang, Shengzhi, Huang, Qiang, Huang, Guohe, Meng, Erhao, Luan, Jinkai, 2018. Reference evapotranspiration forecasting based on local meteorological and global climate information screened by partial mutual information. J. Hydrol. 561, 764–779. https://doi.org/10.1016/j.jhydrol.2018.04.038.

Fang, Kuai, Shen, Chaopeng, Kifer, Daniel, Yang, Xiao, 2017. Prolongation of SMAP to spatiotemporally seamless coverage of continental U.S. using a deep learning neural network. Geophys. Res. Lett. 44 (21) https://doi.org/10.1002/2017GL075619.

FAO. 2015. Chapter 7 - ETc - Dual crop coefficient. Food and Agriculture Organization of the United Nations Retrieved from http://www.fao.org/docrep/x0490e/x0490e0c.

Feng, Yu, Cui, Ningbo, Gong, Daozhi, Zhang, Qingwen, Zhao, Lu, 2017. Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. Agric. Water Manage. 193, 163–173. https://doi.org/10.1016/j.agwat.2017.08.003.

Ferreira, L.B., da Cunha, F.F., 2020. Multi-step ahead forecasting of daily reference evapotranspiration using deep learning. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2020.105728.

Ferreira, Lucas Borges, da Cunha, Fernando França, de Oliveira, Rubens Alves, Fernandes Filho, Elpídio Inácio, 2019. Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM – A new approach. J. Hydrol. 572, 556–570. https://doi.org/10.1016/j.jhydrol.2019.03.028.

Filgueiras, Roberto, Almeida, Thomé Simpliciano, Mantovani, Everardo Chartuni, Dias, Santos Henrique Brant, Fernandes-Filho, Elpídio Inácio, da Cunha, Fernando França, Venancio, Luan Peroni, 2020. Soil water content and actual evapotranspiration predictions using regression algorithms and remote sensing data. Agric. Water Manage 241, 106346. https://doi.org/10.1016/j.agwat.2020.106346.

Fisher, J.B., Lee, B., Purdy, A.J., Halverson, G.H., Dohlen, M.B., Cawse-Nicholson, K., Wang, A., Anderson, R.G., Aragon, B., Arain, M.A., Baldocchi, D.D., Baker, J.M., Barral, H., Bernacchi, C.J., Bernhofer, C., Biraud, S.C., Bohrer, G., Brunsell, N., Cappelaere, B., Castro-Contreras, S., Chun, J., Conrad, B.J., Cremonese, E., Demarty, J., Desai, A.R., De Ligne, A., Foltýnová, L., Goulden, M.L., Griffis, T.J., Grünwald, T., Johnson, M.S., Kang, M., Kelbe, D., Kowalska, N., Lim, J.H., Maïnassara, I., McCabe, M.F., Missik, J.E.C., Mohanty, B.P., Moore, C.E., Morillas, L., Morrison, R., Munger, J.W., Posse, G., Richardson, A.D., Russell, E.S., Ryu, Y., Sanchez-Azofeifa, A., Schmidt, M., Schwartz, E., Sharp, I., Šigut, L., Tang, Y., Hulley, G., Anderson, M., Hain, C., French, A., Wood, E., Hook, S., 2020. ECOSTRESS: NASA's Next Generation Mission to Measure Evapotranspiration From the International Space Station. Water Resour. Res. https://doi.org/10.1029/2019WR026058.

Gupta, Hoshin V., Kling, Harald, Yilmaz, Koray K., Martinez, Guillermo F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377 (1-2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Hamill, Thomas M., Bates, Gary T., Whitaker, Jeffrey S., Murray, Donald R., Fiorino, Michael, Galarneau, Thomas J., Zhu, Yuejian, Lapenta, William, 2013. NOAA's second-generation global medium-range ensemble reforecast dataset. Bull. Am. Meteorol. Soc. 94 (10), 1553–1565. https://doi.org/10.1175/BAMS-D-12-00014.1.

Izadifar, Zohreh, Elshorbagy, Amin, 2010. Prediction of hourly actual evapotranspiration using neural networks, genetic programming, and statistical models. Hydrol. Process 24 (23), 3413–3425. https://doi.org/10.1002/hyp.v24:2310.1002/hyp.7771.

Jensen, M.E., Burmann, R.D., Allen, R.G., 1990. Evaporation and irrigation water requirements, ASCE manual and reports on engineering practice.

Jianping, Zhang, Zhong, Yang, Daojie, Wang, Xinbao, Zhang, 2002. Climate change and causes in the Yuanmou dry-hot valley of Yunnan. J. Arid Environ. 51 (1), 153–162. https://doi.org/10.1006/jare.2001.0851.

Jung, Martin, Reichstein, Markus, Margolis, Hank A., Cescatti, Alessandro, Richardson, Andrew D., Altaf Arain, M., Arneth, Almut, Bernhofer, Christian, Bonal, Damien, Chen, Jiquan, Gianelle, Damiano, Gobron, Nadine, Kiely, Gerald, Kutsch, Werner, Lasslop, Gitta, Law, Beverly E., Lindroth, Anders, Merbold, Lutz, Montagnani, Leonardo, Moors, Eddy J., Papale, Dario, Sottocornola, Matteo, Vaccari, Francesco, Williams, Christopher, 2011. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. J. Geophys. Res. Biogeosci. 116 https://doi.org/10.1029/2010JG001566.

Kao, I-Feng, Zhou, Yanlai, Chang, Li-Chiu, Chang, Fi-John, 2020. Exploring a long short-term memory based encoder-decoder framework for multi-step-ahead flood forecasting. J. Hydrol. 583, 124631. https://doi.org/10.1016/j.jhydrol.2020.124631.

Kim, Daeha, Lee, Woo-Seop, Kim, Seon Tae, Chun, Jong Ahn, 2019. Historical drought assessment over the contiguous united states using the generalized complementary principle of evapotranspiration. Water Resour. Res. 55 (7), 6244–6267. https://doi.org/10.1029/2019WR024991.

Kimball, B.A., Boote, K.J., Hatfield, J.L., Ahuja, L.R., Stockle, C., Archontoulis, S., Baron, C., Basso, B., Bertuzzi, P., Constantin, J., Deryng, D., Dumont, B., Durand, J.L., Ewert, F., Gaiser, T., Gayler, S., Hoffmann, M.P., Jiang, Q., Kim, S.H., Lizaso, J., Moulin, S., Nendel, C., Parker, P., Palosuo, T., Priesack, E., Qi, Z., Srivastava, A., Stella, T., Tao, F., Thorp, K.R., Timlin, D., Twine, T.E., Webber, H., Willaume, M., Williams, K., 2019. Simulation of maize evapotranspiration: An inter-comparison among 29 maize models. Agric. For. Meteorol. https://doi.org/10.1016/j.agrformet.2019.02.037.

Kisi, Ozgur, Alizamir, Meysam, 2018. Modelling reference evapotranspiration using a new wavelet conjunction heuristic method: Wavelet extreme learning machine vs wavelet neural networks. Agric. For. Meteorol 263, 41–48. https://doi.org/10.1016/j.agrformet.2018.08.007.

Koster, R.D., Suarez, M.J., 1996. Energy and Water Balance Calculations in the Mosaic LSM, NASA Technical Memorandum 104606, Technical Report Series on Global Modeling and Data Assimilation.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G., 2019. Benchmarking a Catchment-Aware Long Short-Term Memory Network (LSTM) for Large-Scale Hydrological Modeling. Hydrol. Earth Syst. Sci. Discuss. https://doi.org/10.5194/hess-2019-368.

Landeras, Gorka, Ortiz-Barredo, Amaia, López, José Javier, 2009. Forecasting weekly evapotranspiration with ARIMA and artificial neural network models. J. Irrig. Drain Eng. 135 (3), 323–334. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000008.

Li, Sien, Kang, Shaozhong, Zhang, Lu, Zhang, Jianhua, Du, Taisheng, Tong, Ling, Ding, Risheng, 2016. Evaluation of six potential evapotranspiration models for

estimating crop potential and actual evapotranspiration in arid regions. J. Hydrol. 543, 450–461. https://doi.org/10.1016/j.jhydrol.2016.10.022.
Lian, Xu, Piao, Shilong, Huntingford, Chris, Li, Yue, Zeng, Zhenzhong, Wang, Xuhui, Ciais, Philippe, McVicar, Tim R., Peng, Shushi, Ottlé, Catherine, Yang, Hui, Yang, Yuting, Zhang, Yongqiang, Wang, Tao, 2018. Partitioning global land evapotranspiration using CMIP5 models constrained by observations. Nat. Clim. Change. 8 (7), 640–646. https://doi.org/10.1038/s41558-018-0207-9.
Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News.
Lohar, D., Pal, B., 1995. The effect of irrigation on premonsoon season precipitation over south West Bengal, India. J. Clim. https://doi.org/10.1175/1520-0442(1995)008<2567:TEOIOP>2.0.CO;2.
Long, Di, Longuevergne, Laurent, Scanlon, Bridget R., 2014. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. Water Resour. Res. 50 (2), 1131–1151. https://doi.org/10.1002/2013WR014581.
Meng, C.L., Li, Z.-L., Zhan, X., Shi, J.C., Liu, C.Y., 2009. Land surface temperature data assimilation and its impact on evapotranspiration estimates from the common land model. Water Resour. Res. 45 (2) https://doi.org/10.1029/2008WR006971.
Moratiel, R., Bravo, R., Saa, A., Tarquis, A.M., Almorox, J., 2020. Estimation of evapotranspiration by the Food and Agricultural Organization of the United Nations (FAO) Penman-Monteith temperature (PMT) and Hargreaves-Samani (HS) models under temporal and spatial criteria - a case study in Duero basin (Spain). Nat. Hazards Earth Syst. Sci. https://doi.org/10.5194/nhess-20-859-2020.
Moriasi, D.N., Gitau, M.W., Pai, N., Daggupati, P., 2015. Hydrologic and water quality models: Performance measures and evaluation criteria. Trans. ASABE. https://doi.org/10.13031/trans.58.10715.
Narasimhan, B., Srinivasan, R., 2005. Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring, in: Agricultural and Forest Meteorology. https://doi.org/10.1016/j.agrformet.2005.07.012.
O'Gorman, Paul A., Dwyer, John G., 2018. Using machine learning to parameterize moist convection: potential for modeling of climate, climate change, and extreme events. J. Adv. Model. Earth Syst. 10 (10), 2548–2563. https://doi.org/10.1029/2018MS001351.
Otkin, Jason A., Anderson, Martha C., Hain, Christopher, Svoboda, Mark, Johnson, David, Mueller, Richard, Tadesse, Tsegaye, Wardlow, Brian, Brown, Jesslyn, 2016. Assessing the evolution of soil moisture and vegetation conditions during the 2012 United States flash drought. Agric. For. Meteorol. 218-219, 230–242. https://doi.org/10.1016/j.agrformet.2015.12.065.
Pandey, P.K., Nyori, Topi, Pandey, Vanita, 2017. Estimation of reference evapotranspiration using data driven techniques under limited data conditions. Earth Syst. Environ. 3 (4), 1449–1461. https://doi.org/10.1007/s40808-017-0367-z.
Pauwels, Valentijn R.N., Verhoest, Niko E.C., De Lannoy, Gabriëlle J.M., Guissard, Vincent, Lucau, Cozmin, Defourny, Pierre, 2007. Optimization of a coupled hydrology-crop growth model through the assimilation of observed soil moisture and leaf area index values using an ensemble Kalman filter. Water Resour. Res. 43 (4) https://doi.org/10.1029/2006WR004942.
Payero, José O., Irmak, Suat, 2013. Daily energy fluxes, evapotranspiration and crop coefficient of soybean. Agric. Water Manage. 129, 31–43. https://doi.org/10.1016/j.agwat.2013.06.018.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res.
Perera, Kushan C., Western, Andrew W., Nawarathna, Bandara, George, Biju, 2014. Forecasting daily reference evapotranspiration for Australia using numerical weather prediction outputs. Agric. For. Meteorol. 194, 50–63. https://doi.org/10.1016/j.agrformet.2014.03.014.
PRIESTLEY, C.H.B., TAYLOR, R.J., 1972. On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. Mon. Weather Rev. https://doi.org/10.1175/1520-0493(1972)100<0081:otaosh>2.3.co;2.
Rangapuram, S.S., Seeger, M., Gasthaus, J., Stella, L., Wang, Y., Januschowski, T., 2018. Deep state space models for time series forecasting, in: Advances in Neural Information Processing Systems.
Richards, L.A., 1931. Capillary conduction of liquids through porous mediums. J. Appl. Phys. 1 (5), 318–333.
Rosenberry, Donald O., Winter, Thomas C., Buso, Donald C., Likens, Gene E., 2007. Comparison of 15 evaporation methods applied to a small mountain lake in the northeastern USA. J. Hydrol. 340 (3-4), 149–166. https://doi.org/10.1016/j.jhydrol.2007.03.018.
Sahoo, S., Russo, T.A., Elliott, J., Foster, I., 2017. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S. Water Resour. Res. 53 (5), 3878–3895. https://doi.org/10.1002/wrcr.v53.510.1002/2016WR019933.
Scott, R.L., Cable, W.L., Huxman, T.E., Nagler, P.L., Hernandez, M., Goodrich, D.C., 2008. Multiyear riparian evapotranspiration and groundwater use for a semiarid watershed. J. Arid Environ. 72 (7), 1232–1246. https://doi.org/10.1016/j.jaridenv.2008.01.001.
Seneviratne, Sonia I., Corti, Thierry, Davin, Edouard L., Hirschi, Martin, Jaeger, Eric B., Lehner, Irene, Orlowsky, Boris, Teuling, Adriaan J., 2010. Investigating soil moisture-climate interactions in a changing climate: a review. Earth Sci. Rev 99 (3-4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004.
Sharma, Vivek, Kilic, Ayse, Irmak, Suat, 2016. Impact of scale/resolution on evapotranspiration from Landsat and MODIS images. Water Resour. Res. 52 (3), 1800–1819. https://doi.org/10.1002/wrcr.v52.310.1002/2015WR017772.
Shiri, Jalal, 2018. Improving the performance of the mass transfer-based reference evapotranspiration estimation approaches through a coupled wavelet-random forest methodology. J. Hydrol. 561, 737–750. https://doi.org/10.1016/j.jhydrol.2018.04.042.
Shugart, H. H.,1998. Terrestrial ecosystems in changing environments. Cambridge University Press.https://doi.org/10.1016/s0304-3800(99)00031-9.
Song, Yi, Ma, Mingguo, 2011. A statistical analysis of the relationship between climatic factors and the normalized difference vegetation index in China. Int. J. Remote Sens. 32 (14), 3947–3965. https://doi.org/10.1080/01431161003801336.
Suyker, A., 2001a. AmeriFlux US-Ne2 Mead - irrigated maize-soybean rotation site, Dataset. https://doi.org/10.17190/AMF/1246085.
Suyker, A., 2001b. AmeriFlux US-Ne3 Mead - rainfed maize-soybean rotation site, Dataset. https://doi.org/10.17190/AMF/1246086.
Suyker, A., 2001c. AmeriFlux US-Ne1 Mead - rainfed maize-soybean rotation site, Dataset. https://doi.org/10.17190/AMF/1246084.
Tabari, Hossein, Kisi, Ozgur, Ezani, Azadeh, Hosseinzadeh Talaee, P., 2012. SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. J. Hydrol. 444-445, 78–89. https://doi.org/10.1016/j.jhydrol.2012.04.007.
Tavares, L.D., Saldanha, R.R., Vieira, D.A.G., 2015. Extreme learning machine with parallel layer perceptrons. Neurocomputing 166, 164–171. https://doi.org/10.1016/j.neucom.2015.04.018.
te Beest, D.E., Mes, S.W., Wilting, S.M., Brakenhoff, R.H., van de Wiel, M.A., 2017. Improved high-dimensional prediction with Random Forests by the use of co-data. BMC Bioinf. 18 (1) https://doi.org/10.1186/s12859-017-1993-1.
Tennant, Christopher, Larsen, Laurel, Bellugi, Dino, Moges, Edom, Zhang, Liang, Ma, Hongxu, 2020. The utility of information flow in formulating discharge forecast models: a case study from an arid snow-dominated catchment. Water Resour. Res. 56 (8) https://doi.org/10.1029/2019WR024908.
Thornton, P.E., Thornton, M.M., Mayer, B.W., Wilhelmi, N., Wei, Y., Devarakonda, R., Cook, R.B., 2014. Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2. Data set. [WWW Document]. Oak Ridge Natl. Lab. Distrib. Act. Arch. Center, Oak Ridge, Tennessee, USA. https://doi.org/https://doi.org/10.3334/ORNLDAAC/1219.
Trenberth, Kevin E., Smith, Lesley, Qian, Taotao, Dai, Aiguo, Fasullo, John, 2007. Estimates of the global water budget and its annual cycle using observational and model Data. J. Hydrometeorol. 8 (4), 758–769. https://doi.org/10.1175/JHM600.1.
Valipour, Mohammad, Gholami Sefidkouhi, Mohammad Ali, Raeini−Sarjaz, Mahmoud, 2017. Selecting the best model to estimate potential evapotranspiration with respect to climate change and magnitudes of extreme events. Agric. Water Manage. 180, 50–60. https://doi.org/10.1016/j.agwat.2016.08.025.
Velpuri, N.M., Senay, G.B., Singh, R.K., Bohms, S., Verdin, J.P., 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. Remote Sens. Environ. 139, 35–49. https://doi.org/10.1016/j.rse.2013.07.013.
E. Vermote, R.W., 2015. MOD09GQ MODIS/Terra Surface Reflectance Daily L2G Global 250m SIN Grid V006 [WWW Document]. Distrib. by NASA EOSDIS L. Process. DAAC. https://doi.org/doi.org/10.5067/MODIS/MOD09GQ.006.
Vinukollu, Raghuveer K., Sheffield, Justin, Wood, Eric F, Bosilovich, Michael G., Mocko, David, 2012. Multimodel analysis of energy and water fluxes: Intercomparisons between operational analyses, a land surface model, and remote sensing. J. Hydrometeorol. 13 (1), 3–26. https://doi.org/10.1175/2011JHM1372.1.
Walls, Spencer, Binns, Andrew D., Levison, Jana, MacRitchie, Scott, 2020. Prediction of actual evapotranspiration by artificial neural network models using data from a Bowen ratio energy balance station. Neural Comput. Appl. 32 (17), 14001–14018. https://doi.org/10.1007/s00521-020-04800-2.
Wang, Kaicun, Dickinson, Robert E., 2012. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. Rev. Geophys 50 (2). https://doi.org/10.1029/2011RG000373.
Wei, Zhongwang, Yoshimura, Kei, Wang, Lixin, Miralles, Diego G., Jasechko, Scott, Lee, Xuhui, 2017. Revisiting the contribution of transpiration to global terrestrial evapotranspiration. Geophys. Res. Lett. 44 (6), 2792–2801. https://doi.org/10.1002/2016GL072235.
Willmott, Cort J., 1981. On the validation of models. Phys. Geogr. 2 (2), 184–194. https://doi.org/10.1080/02723646.1981.10642213.
Wilson, K.B., Baldocchi, D.D., Hanson, P.J., 2001. Leaf age affects the seasonal pattern of photosynthetic capacity and net ecosystem exchange of carbon in a deciduous forest. Plant Cell Environ. 24 (6), 571–583. https://doi.org/10.1046/j.0016-8025.2001.00706.x.
Wutzler, Thomas, Lucas-Moffat, Antje, Migliavacca, Mirco, Knauer, Jürgen, Sickel, Kerstin, Šigut, Ladislav, Menzer, Olaf, Reichstein, Markus, 2018. Basic and extensible post-processing of eddy covariance flux data with REddyProc. Biogeosciences 15 (16), 5015–5030. https://doi.org/10.5194/bg-15-5015-2018.
Xia, Youlong, Mitchell, Kenneth, Ek, Michael, Sheffield, Justin, Cosgrove, Brian, Wood, Eric, Luo, Lifeng, Alonge, Charles, Wei, Helin, Meng, Jesse, Livneh, Ben, Lettenmaier, Dennis, Koren, Victor, Duan, Qingyun, Mo, Kingtse, Fan, Yun, Mocko, David, 2012. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. J. Geophys. Res. Atmos. 117 (D3) https://doi.org/10.1029/2011JD016048.
Xu, Tongren, Guo, Zhixia, Liu, Shaomin, He, Xinlei, Meng, Yangfanyu, Xu, Ziwei, Xia, Youlong, Xiao, Jingfeng, Zhang, Yuan, Ma, Yanfei, Song, Lisheng, 2018. Evaluating Different Machine Learning Methods for Upscaling Evapotranspiration from Flux Towers to the Regional Scale. J. Geophys. Res. Atmos. 123 (16), 8674–8690. https://doi.org/10.1029/2018JD028447.
Yang, F., White, M.A., Michaelis, A.R., Ichii, K., Hashimoto, H., Votava, P., Zhu, A-X., Nemani, R.R., 2006. Prediction of continental-scale evapotranspiration by

combining MODIS and AmeriFlux data through support vector machine. IEEE Trans. Geosci. Remote Sens. 44 (11), 3452–3461. https://doi.org/10.1109/TGRS.2006.876297.

Yao, Yunjun, Liang, Shunlin, Zhao, Shaohua, Zhang, Yuhu, Qin, Qiming, Cheng, Jie, Jia, Kun, Xie, Xianhong, Zhang, Nannan, Liu, Meng, 2013. Validation and application of the modified satellite-based Priestley-Taylor algorithm for mapping terrestrial evapotranspiration. Remote Sens. 6 (1), 880–904. https://doi.org/10.3390/rs6010880.

Yassin, Mohamed A., Alazba, A.A., Mattar, Mohamed A., 2016. Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid climate. Agric. Water Manage. 163, 110–124. https://doi.org/10.1016/j.agwat.2015.09.009.

Yin, Juan, Deng, Zhen, Ines, Amor V.M., Wu, Junbin, Rasu, Eeswaran, 2020. Forecast of short-term daily reference evapotranspiration under limited meteorological variables using a hybrid bi-directional long short-term memory model (Bi-LSTM). Agric. Water Manage. 242, 106386. https://doi.org/10.1016/j.agwat.2020.106386.

Zaherpour, Jamal, Mount, Nick, Gosling, Simon N., Dankers, Rutger, Eisner, Stephanie, Gerten, Dieter, Liu, Xingcai, Masaki, Yoshimitsu, Müller Schmied, Hannes, Tang, Qiuhong, Wada, Yoshihide, 2019. Exploring the value of machine learning for weighted multi-model combination of an ensemble of global hydrological models. Environ. Model. Softw. 114, 112–128. https://doi.org/10.1016/j.envsoft.2019.01.003.

Zhang, Jianfeng, Zhu, Yan, Zhang, Xiaoping, Ye, Ming, Yang, Jinzhong, 2018. Developing a Long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. J. Hydrol. 561, 918–929. https://doi.org/10.1016/j.jhydrol.2018.04.065.

Zhao, Wen Li, Gentine, Pierre, Reichstein, Markus, Zhang, Yao, Zhou, Sha, Wen, Yeqiang, Lin, Changjie, Li, Xi, Qiu, Guo Yu, 2019b. Physics-Constrained Machine Learning of Evapotranspiration. Geophys. Res. Lett. 46 (24), 14496–14507. https://doi.org/10.1029/2019GL085291.

Zhao, Peng, Kang, Shaozhong, Li, Sien, Ding, Risheng, Tong, Ling, Du, Taisheng, 2018. Seasonal variations in vineyard ET partitioning and dual crop coefficients correlate with canopy development and surface soil moisture. Agric. Water Manage. 197, 19–33. https://doi.org/10.1016/j.agwat.2017.11.004.

Zhao, Tongtiegang, Wang, Quan J., Schepen, Andrew, Griffiths, Morwenna, 2019a. Ensemble forecasting of monthly and seasonal reference crop evapotranspiration based on global climate model outputs. Agric. For. Meteorol. 264, 114–124. https://doi.org/10.1016/j.agrformet.2018.10.001.

Zou, Lei, Zhan, Chesheng, Xia, Jun, Wang, Tiejun, Gippel, Christopher J., 2017. Implementation of evapotranspiration data assimilation with catchment scale distributed hydrological model via an ensemble Kalman Filter. J. Hydrol. 549, 685–702. https://doi.org/10.1016/j.jhydrol.2017.04.036.