Applied Computing and Geosciences 16 (2022) 100105

Contents lists available at ScienceDirect

Applied Computing and Geosciences

journal homepage: www.sciencedirect.com/journal/applied-computing-and-geosciences

A framework for constructing machine learning models with feature set optimisation for evapotranspiration partitioning

Adam Stapleton [a,*,1], Elke Eichelmann [b,2], Mark Roantree [c,3]

[a] School of Computing, Dublin City University, Dublin 9, Ireland
[b] School of Biology and Environmental Science, University College Dublin, Dublin 4, Ireland
[c] Insight Centre for Data Analytics, Dublin City University, Dublin 9, Ireland

ARTICLE INFO

Keywords:
Machine learning
Evapotranspiration
Optimisation
Atmospheric sciences
Geophysics
Environmental sciences

ABSTRACT

A deeper understanding of the drivers of evapotranspiration and the modelling of its constituent parts (evaporation and transpiration) may be of significant importance to the monitoring and management of water resources globally over the coming decades. In this work a framework was developed to identify the best performing machine learning algorithm from a candidate set, select optimal predictive features and rank features in terms of their importance to predictive accuracy. The experiments conducted in this work used 3 separate feature sets across 4 wetland sites as input into 8 candidate machine learning algorithms, providing 96 sets of experimental configurations. Given this high number of parameters, our results show strong evidence that there is no singularly optimal machine learning algorithm or feature set across all of the wetland sites studied despite their similarities. At each of the sites at least one model was identified that improved on the predictive performance of our baseline. A key finding discovered when examining feature importance is that methane flux, a feature whose relationship with evapotranspiration is not generally examined, may contribute to further biophysical process understanding. This work demonstrates the applicability of a machine learning framework for evapotranspiration partitioning that is independent of domain knowledge, producing improved models for partitioning and identifying new and useful predictive features.

1. Introduction

Evapotranspiration (ET) is the process by which water is exchanged between the biosphere and the atmosphere. Better understanding of ET processes and their drivers in various environments is important for the entire terrestrial hydrological cycle that governs the transport and recycling of the water that supports, for example, our fresh water supplies (Oki and Kanae, 2006; Zeng et al., 2018). Observations of the Earth's atmosphere and biosphere over the last number of decades have indicated an intensifying hydrological cycle (Brutsaert and Parlange, 1998; Pascolini-Campbell et al., 2021) and an increase in the number of people living in water stressed areas (Oki and Kanae, 2006). Modelling efforts over this period have shown disagreements, with evidence indicating a decline in global terrestrial ET caused by a reduction in available moisture supply (Jung et al., 2010) and more recently, indication of an increase in global terrestrial ET due to increasing land temperature (Pascolini-Campbell et al., 2021). ET is a process composed of two main parts: Evaporation (E), the physical process, and Transpiration (T), a biologically modulated process that occurs through the stomata of plants. A better understanding of the drivers of ET and the modelling of each of its constituent parts may be of significant importance to the monitoring and management of water resources globally over the coming decades. ET research contributes to many important components of global climate modelling including cloud formation (of relevance due to their role in the absorption and reflection of solar radiation and the transfer of energy between environments) and moisture availability (Gerken et al., 2018; Green et al., 2017; Pielke et al., 1998; Schlesinger and Jasechko, 2014; Trenberth et al., 2009). The partitioning of ET into its constituents is vital in reducing the associated uncertainty in climate land surface models and satellite remote sensing projects such as ECOSTRESS (Fisher et al., 2020) as current models are validated on combined ET data only (Stoy et al., 2019). The usage of machine learning (ML) in the domain of biosphere–atmosphere exchange has seen an increase in recent years with the availability of large, open source Eddy Covariance (EC) data sets such

[*] Corresponding author.
E-mail addresses: adam.stapleton9@mail.dcu.ie (A. Stapleton), elke.eichelmann@ucd.ie (E. Eichelmann), mark.roantree@dcu.ie (M. Roantree).
[1] http://orcid.org/0000-0003-1233-211X
[2] http://orcid.org/0000-0001-9516-7951
[3] http://orcid.org/0000-0002-1329-2570

https://doi.org/10.1016/j.acags.2022.100105
Received 8 March 2022; Received in revised form 2 September 2022; Accepted 1 November 2022
Available online 8 November 2022

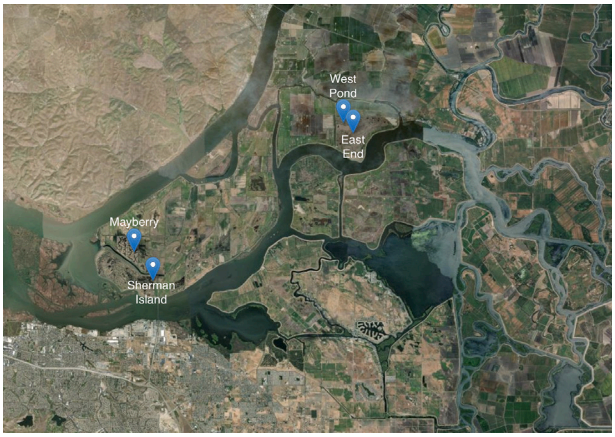

Fig. 1. Satellite view of wetland sites included in this study.

as FLUXNET (Baldocchi et al., 2001) and AmeriFlux (Novick et al., 2018) enabling more data intensive approaches. Applications of ML in the domain of biosphere–atmosphere exchange have mostly focused on gap-filling of EC data (Irvin et al., 2021) but some success has been achieved in the application of ML techniques to the partitioning of gas fluxes (Tramontana et al., 2020), prediction of fluxes (Tramontana et al., 2016), spatial interpolation (Lin et al., 2002), and upscaling of EC data (Bodesheim et al., 2018; Jung et al., 2009). As the EC method measures total water flux, the goal of partitioning in this work is to determine the individual contributions of E and T to the net flux.

Our previous work (Eichelmann et al., 2021b) introduced a novel, data-driven ET partitioning method and applied neural networks on micro-meteorological data collected from four wetland sites in California (Eichelmann et al., 2021b). Artificial neural networks (NN) were used to partition ET into E and T by training these networks to predict E during periods where T can assumed to be negligible. From this, T can be estimated by subtracting the predicted E from total ET. In this paper a broader range of ML algorithms are compared alongside the NN tested in the previous work, expanding on the complexity of the models via a novel feature selection process. The previous work utilised predictive features via domain expertise only and in this work additional features are selected via their correlation with the target and their effect on increasing predictive performance. This work seeks to address three research questions. Firstly, is there a ML algorithm that performs as well or better than those tested in our previous work on the task of predicting E? Secondly, can ML be utilised to identify an optimal set of predictive features that improves predictive performance? Thirdly, do the features identified contribute to our understanding of the processes mediating ET in the wetland sites in this study? Identical datasets to the previous work are used in this work (as described in Section 2.1) and the results from this previous work are used as a partial baseline for comparison in Section 4.

2. Background

2.1. Data

The data utilised in this work are obtained using the Eddy Covariance (EC) method (Aubinet et al., 2012) from measurement towers across four wetland sites in the Sacramento–San Joaquin river delta in Northern California: Twitchell Wetland West Pond (AmeriFlux ID: US-TW1) (Valach et al., 2021b)(WP), Twitchell East End Wetland (AmeriFlux ID: US-TW4) (Eichelmann et al., 2021a)(EE), Mayberry Wetland (AmeriFlux ID: US-MYB) (Matthes et al., 2021)(MB), and Sherman Island Restored Wetland (AmeriFlux ID: US-Sne) (Shortt et al., 2021)(SW). The locations of the sites are displayed graphically in Fig. 1. This method ascertains the flux of trace gases by measuring the covariance between fluctuations in vertical wind velocity and the mixing ratio of the gas in question. The data from all sites are available under an open-source license as part of the AmeriFlux network and can be accessed through the AmeriFlux data sharing platform (Laboratory, 2021; Novick et al., 2018). The sites have been described in detail elsewhere (Detto et al., 2010; Eichelmann et al., 2018; Hatala et al., 2012; Knox et al., 2015) and the reader is referred to these works for a more complete description. The four sites are all freshwater marsh wetlands that have been constructed by the Department of Water Resources to manage soil subsidence in the area. The observation period for each site differs in length with approximately 10 years of data for MB (October 2010 to October 2020), 8 for WP (July 2012 to September 2020), 7 for EE (November 2013 to September 2020) and 4 for SW (May 2016 to April 2020). All sites, with the exception of WP, underwent flooding within the measurement period. The longest standing of the four sites is WP having been established in 1998. The initial flooding period is of note as it provides a period in which vegetation has not yet been established and thus, it can be assumed that T is negligible during this period. The vegetation cover (within the EC footprint at the latest measurement in 2018 (Valach et al., 2021a)) varies between the sites with 97% cover at WP, 64% at MB, 96% at EE and 45% at SW. The lower vegetation cover at SW can be explained by the fact that it is the newest wetland to be established, constructed in 2016. The dominant vegetation species at all sites are tule (*Schoenoplectus acutus*) and cattail (*Typha* spp.) (O'Connell et al., 2015). Continuous fluxes of water vapour and other trace gases were measured using the EC method. In addition to the EC data, micrometeorological and environmental data were also obtained for each of the sites including the following variables with a known relationship with ET; air temperature (TA); water temperature (TW); soil temperature (TS); relative humidity (RH); atmospheric pressure (AP); net radiation ($RNET$); water table depth (WT); vapour pressure deficit (VPD); sensible heat exchange (H); friction velocity ($u*$); vegetation greenness index from camera data (GCC) and the target variable water flux (labelled total ET or wq). The data frequency is at 30 min intervals and where the data were recorded at higher frequencies, the mean was computed for that 30 min period. Pre-processing of the data to remove spikes, filter for instrument malfunctioning and gap-filling procedure for certain data has been described in detail in Eichelmann et al. (2018). For EC flux features with missing data, a NN procedure was used for imputation (Baldocchi et al., 2015; Knox et al., 2015) and is detailed in our previous work (Eichelmann et al., 2021b). Meteorological variables were imputed using data from nearby weather stations where data were available. For any remaining features with missing data, linear interpolation was used.

2.2. Machine learning algorithms

In this paper, a variety of supervised ML algorithms were utilised and the resulting models compared for performance on the prediction task described in Section 3.1. The algorithms tested can be broadly grouped into 3 categories: parametric regressors, non-parametric regressors and ensembles (Géron, 2019). The scikit-learn library (Pedregosa et al., 2011) was used for model building in addition to the XGBoost (Chen and Guestrin, 2016) and LightGBM (Ke et al., 2017) libraries.

Parametric models, such as linear and ridge regression models or NNs, produce a predictive function by assuming a model with a fixed number of parameters and improving the performance of the model by adjusting the weights of the parameters until a minimum loss is obtained. Due to their simplicity, linear parametric models are extremely fast in training and prediction but can suffer from underfitting if the true distribution of the data is more complex.

Non-parametric models, such as K-nearest neighbours (KNN), decision trees (DT) and Support Vector Machines (SVM) make little or

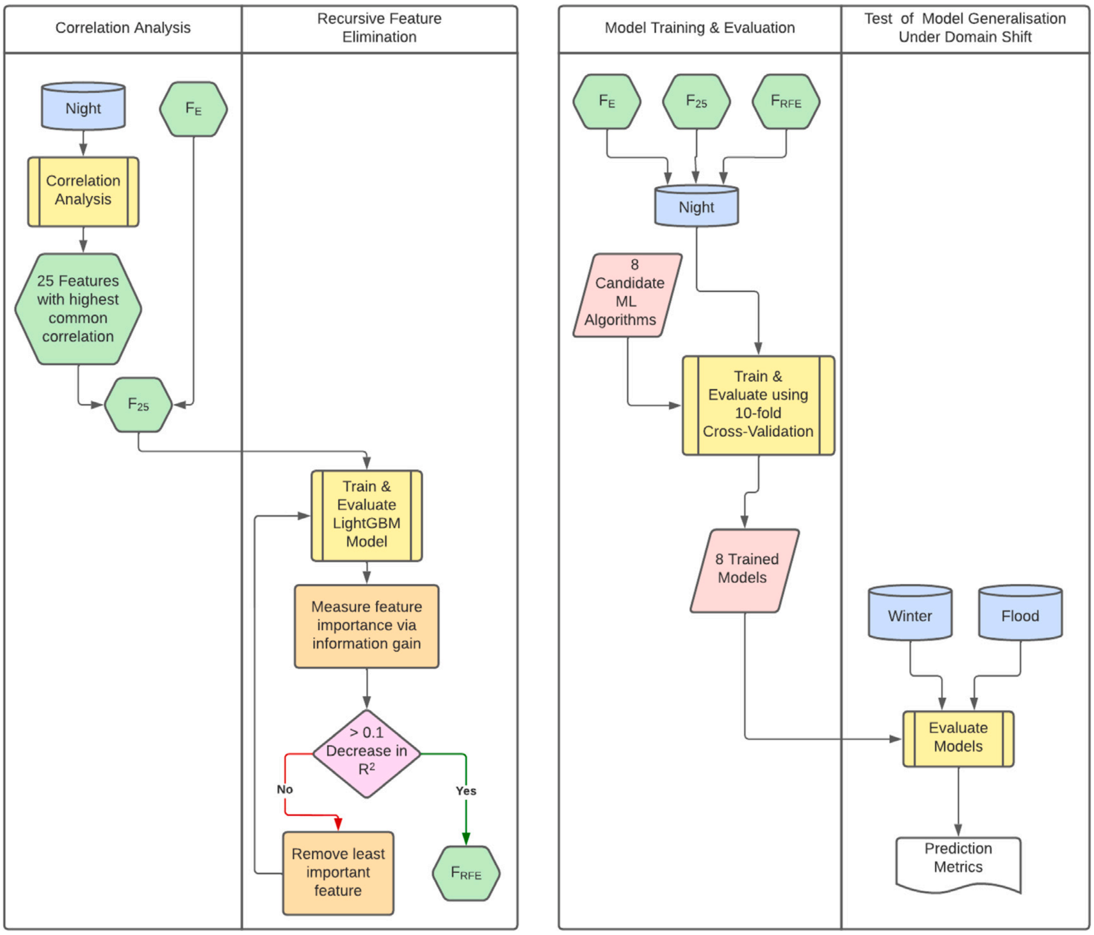
Correlation Analysis
Recursive Feature Elimination
Night
F_E
Correlation Analysis
25 Features with highest common correlation
F_{25}
Train & Evaluate LightGBM Model
Measure feature importance via information gain
> 0.1 Decrease in R^2
No
Yes
Remove least important feature
F_{RFE}
Model Training & Evaluation
Test of Model Generalisation Under Domain Shift
F_E
F_{25}
F_{RFE}
Night
8 Candidate ML Algorithms
Train & Evaluate using 10-fold Cross-Validation
8 Trained Models
Winter
Flood
Evaluate Models
Prediction Metrics

Fig. 2. The process flow for the entire framework is split into two for the sake of legibility. On the left hand side the processes for obtaining the two additional feature sets, F_{25} and F_{RFE} are described. On the right hand side the processes for training and evaluating the models are described. Each orange rectangle represents a standalone process, each yellow rectangle represents a process with multiple components (the details of which are included in the text). Each green hexagon represents a feature set, each blue cylinder represents a data set, each red parallelogram represents a set of ML models and a pink diamond represents a decision. Some processes are carried out across all 4 sites, such as the Correlation Analysis. Other processes are carried out on each site individually, such as the Recursive Feature Elimination.

no assumptions about the predictive function in advance and seek to learn both the functional form and the function's parameter values from the data. Non-parametric models produce a more flexible predictive function, thereby allowing them to better model more complex distributions. This increase in complexity can lead to overfitting and an increase in both training time and the volume of data required to fit more complex models.

Ensemble methods such as Gradient Boosting, XGBoost and LightGBM, combine the predictions of many simple models, referred to as weak learners or base learners, to produce a predictive model. Base learners can be trained in parallel and combined using methods such as bagging or stacking, or sequentially using methods such as boosting (Géron, 2019). For all ensemble methods tested, the base learners were DT. Ensemble methods are generally less prone to overfitting while still retaining sufficient complexity to arrive at a reasonable approximation of the underlying distribution.

3. Framework methodology

This section begins with a description of the approach used for ET Flux Partitioning. The proposed framework is then described as the following series of methodological steps: data preparation, feature selection, construction of baseline models, final model training, evaluation and comparison. These steps are represented as a flow in Fig. 2.

Each set of experiments were repeated across the four wetland sites for each set of algorithms.

3.1. Evapotranspiration partitioning

A novel, data-driven method to ET partitioning (drawing from previous work on carbon dioxide flux partitioning (Tramontana et al., 2020)) was presented previously in Eichelmann et al. (2021b), with a brief outline here. Given the difficulty in establishing ground-truth data for the component contributions of E and T to overall ET, a number of assumptions are used to establish periods during which T can be assumed to be negligible and therefore, taken to be 0 in calculations. During the night, plant stomata are assumed to be closed and therefore not transpiring (confirmed with leaf level measurements at these sites). Utilising this assumption, the night-time data (Night) are used to train models to predict E, which can then be subtracted from total ET to give predicted the values for T. Explicitly the relationship between E and T can be expressed using Eq. (1).

$$\begin{aligned} ET &= T + E \\ T_{Night} &\simeq 0 \\ ET_{Night} &= E \end{aligned} \tag{1}$$

As there are no measured ground truth data for the individual components E and T, assumptions about T during other periods of the year are used to determine two further test sets to evaluate the methodology, namely the day-time data from the initial flooding period (Flood) and day-time data from the winter senescent months (Winter). The principal purpose of evaluating with these test sets is to examine the performance of the models under a domain shift. Namely, this is the ability of the models to generalise to situations where the underlying distribution of the predictive features is different to that observed in training, which is performed on Night data only. Further information on this domain shift can be found in the supplementary material of our previous work (Eichelmann et al., 2021b). During the initial flooding period of each of the wetland sites, vegetation had not yet been established and therefore, T was not occurring. During the winter months the vegetation are observed to be senescent and here again, T is negligible. An additional set of core assumptions are used in determining the timing of the onset and duration of these periods. The zenith angle of the sun being greater than 90° is used to determine the night-time periods. Visual determination of the level of vegetation from camera observation of the sites is used to determine the onset of vegetation after the initial flooding period, also referred to as “greenup”. Lastly, the months of December, January and February are taken as the senescent periods. Limitations of these assumptions are discussed in Section 4.5.

3.2. Data preparation

In order to reduce the dimensionality (number of features) of the data to be computationally tractable for model building, the following approach to feature selection was undertaken. First reduce the candidate number of features using correlation and completeness analyses. Secondly, explore possible feature sets of different sizes and in different combinations using Recursive Feature Elimination. This process is described fully in Section 3.4.

As in Eichelmann et al. (2021b), domain knowledge was used to inform the selection of features that have a known relationship with water flux. These were *VPD, GCC, u*, TA, RNET, WT, H* and ecosystem respiration estimated from an exponential relationship between night-time carbon flux and temperature as performed in Reichstein et al. (2005) ($ER_{Reichstein}$). In addition, three time features were added: *year*, *month* and *the day of the year (DOY)*. This forms the first feature set for testing, denoted by the identifier F_E. All features with a completeness less than 80% (i.e. missing greater than 20% of the data) for the measurement period were discarded. Soil and water temperature measurements taken at various depths showed low levels of completeness and the depths at which measurements were taken was not consistent across sites. In order to obtain a useable feature for soil and water temperature that is comparable between sites, the measurements were consolidated by computing the mean of the sensor values across all depths to create two new features, *TS (mean)* and *TW (mean)*. For the remaining features, linear interpolation was used to replace the remaining missing data and a correlation analysis was undertaken to extract the most likely useful features. Of the 50 most highly correlated features at each site, the 25 features that were common across all sites in that subset of 50 were selected and added to the F_E feature set. The resultant feature set was labelled F_{25}. This approach was taken as the hypothesis is that features that do not have a correlation with the target feature that is common across multiple sites will be less likely to have an underlying physical causal relationship (i.e. the correlation is more likely to be spurious) and thus can be removed. It is noted that selecting the most highly correlated features that are common across the 4 sites is equivalent to removing the features that have no common correlation and the features that have a lower average correlation across the sites.

3.3. Model comparison

In order to test and compare a suitably diverse set of algorithms for model building, initial testing examined 38 algorithms from the scikit-learn library (Pedregosa et al., 2011) alongside two additional ensemble algorithms; LightGBM (Ke et al., 2017) and XGBoost (Chen and Guestrin, 2016). The models are compared in order to ascertain which algorithm will be most suited to the model building task, including but not limited to an improvement in model performance in terms of fitting well to the training data, generalising well to unseen data and computational cost. Any models whose predictions had a negative coefficient of determination (R^2) with the target at any site were immediately discarded. From the remaining models, a subset of the best performing models (or simplest model in the case of equal model performance) were selected across 3 different categories of models; parametric, non-parametric and ensemble. The inclusion of different categories of ML algorithms is undertaken to prevent a loss in diversity from the initial set of algorithms tested. The models selected for final testing and comparison were linear regression, ridge regression, KNN, DT, Gradient Boosting Decision Trees, LightGBM and XGBoost. Default hyper-parameters were used for all algorithms.

To reduce the effect of sampling bias in training and testing the models, 10-fold cross-validation was applied. The Night data are split into 10 randomly sampled subsets (folds) for cross-validation. The models are trained and evaluated 10 times wherein at each iteration, one of the folds is removed and the models are trained on the remaining 9 folds. At each iteration, the models are evaluated on the held-out fold of the Night data as well as the entire Winter and Flood data. The process is then repeated with the previous fold replaced for training and the next fold removed for testing. At the end of the procedure, the mean value for each of the metrics is obtained.

3.4. Recursive feature elimination

In order to identify a feature set that contains maximal information with the minimum number of features a recursive feature elimination (RFE) method is used. A lower number of features is desired to reduce model complexity and subsequently reduce the chance of overfitting and to combat the so-called “curse of dimensionality” (Han et al., 2011) whereby an increase in the number of features leads to a lower number of samples per unit volume of the feature space. In this method, a LightGBM model is trained on Night data with the features from the F_{25} feature set for each of the four sites. Each model is then used to obtain a metric for the relative importance of each feature at that site. The metric that is used to measure feature importance is the sum of the gains in model performance, as measured by reduction in Root Mean Squared Error (RMSE), of all branches of base learner DT using that feature. The least important feature is then removed from the feature set. A new model is trained on the resulting, smaller feature set and the process is repeated until no features remain. Cross-validation is applied at each iteration and the mean of the model performance metrics are recorded. The optimal features for each site are determined to be the feature set that preceded a 0.1% decrease in R^2 for the hold-out Night data as the number of features is iteratively decreased. The feature sets obtained for each site are then compared for commonalities and those features that were of low significance and in the optimal feature set for only one site are discarded and the remaining feature set is labelled F_{RFE}. It is hypothesised that this feature set approaches the minimum number of features needed to capture all information needed to model E and T from the available data.

3.5. Evaluation

In order to select the most appropriate set of metrics for evaluating model performance, the nature of the data must be taken into consideration. In contrast to a conventional supervised learning problem where the ground-truth data were obtained under known conditions, the ground-truth data for the experiments in this study are based on an assumption about approximate levels of T occurring under different conditions. During the night-time, winter and initial flooding periods, the assumptions governing negligible T are slightly different. Therefore, we expect that some common metrics for the evaluation of a regressor (e.g. RMSE and mean average error) may lead to difficulty in comparing model performance across test sets as the level of actual T occurring may vary and be non-negligible in some cases. This may lead to increases in measures of predictive error that are not attributable to poor predictive performance but rather to deviations in the data caused by a confounding variable that is not present in the training data (namely T arising in total measured ET where the model assumes that the total measured ET should be measuring E only).

Each of the Night, Flood and Winter datasets have different data distributions (Eichelmann et al., 2021b) and it is the performance of the models on data whose values lie outside the range of the training data (referred to as unseen data) that must be evaluated.

Therefore, a metric that determines how closely the variations in predictions of E follow the variations in total measured ET across all test sets is required. For this reason, the metrics chosen for evaluation are R^2, Adjusted R^2 ($\mathrm{R}^2{}_{\mathrm{Adj}}$) and slope of line of best fit between ground truth and predictions (m). $\mathrm{R}^2{}_{\mathrm{Adj}}$ enables comparison between feature sets as this metric adjusts for the number of features used in order to account for the often spurious increase in R^2 when additional features are added to a model.

$$R^2_{Adj} = 1 - \frac{(1-R^2)\times(p-1)}{p-q-1} \tag{2}$$

Eq. (2) describes Adjusted R^2 where R^2 is the R^2 of the model, p is the number of samples and q is the number of features.

Slope is chosen in order to validate one of the biophysical constraints of any partitioning model, namely that the slope of the line of best fit between the predicted E and the ground truth (total ET) never exceeds 1 for any of the data. A slope greater than 1 would indicate that E had exceeded net ET which would lead to negative T, violating the biophysical constraint that negative T cannot occur.

At each iteration of the cross-validation procedure, the metrics are obtained for the removed fold of the Night data in addition to the entire Winter and Flood data. The mean of the metrics for all 10 iterations is then reported as the metric for that model.

4. Results & discussion

In this section experimental results are presented where, for all four wetland sites, identical feature sets and experimental configurations were used. The results are reported on a per site basis as the goal is to compare how each of the models generalise to unseen data for the same site they were trained on. All results are the mean values of the metric across 10 cross-validation folds.

4.1. Model comparison results

Fig. 3 shows the $\mathrm{R}^2{}_{\mathrm{Adj}}$ values for all sites, algorithms and feature sets tested for the Night, Winter and Flood data.

Fig. 3 shows that an improvement in model performance was obtained on Night data as well as in generalising to Winter and Flood data over and above that of the baseline results (Eichelmann et al., 2021b), where the baseline results are those that utilised NN-based models and the F_{E} feature set, indicated by a grey circular icon.

In general, results show that addition of the extra 25 features from the correlation analysis gave some improvement in model performance across all model types when compared to the baseline feature set, F_E. In addition, it is seen that reduction in features from 36 to 19 in going from F_{25} to F_{RFE} either resulted in further incremental improvement for the best performing models, or did not drastically decrease model performance. All sites had more than one model which failed to generalise well to the Winter and Flood data. This observation is important as it indicates that a site-specific approach to model building may be more favourable. The sites modelled in this paper are all biologically similar: all wetlands with the same species composition, same climate and similar management. As described in Section 2.1 there are some known differences between the sites, such as the ratio of open water to vegetation cover and the utility of this framework may be best realised when building models that contain not only the general features relevant for modelling a particular ecosystem but also those features that are relevant for modelling that particular site. Gradient Boosting and LightGBM based models performed well at all sites except WP with LightGBM notably performing best on the Night data at all sites. At WP linear parametric models such as Ridge or Linear Regression performed best with most other models failing to generalise to Winter data at this site. The low computational time and resource requirements and high predictive performance on Night, Flood and Winter data would suggest that LightGBM or Ridge regression would be ideal candidate models in most cases. Many of the features are known to exhibit a high-level of non-linearity in their relationship with ET, particularly when considering the shift from night to day. For example, *RNET* is approximately constant and negative at night while positive during the day. This may go towards explaining why simpler models (such as linear parametric models or models that used less features) performed worse in some cases — the generated hyper-plane may not have sufficient complexity to model the underlying relationship between the predictive features and the target feature. It is also noted that most of the models failed to generalise well for the Winter data at WP, indicating that there may be particularities about this site that were not captured in the features or in the learned predictive function. This may be due to the fact that WP has differences in its composition to the other sites being the oldest of the 4 sites, as previously discussed in Section 2.1.

4.2. RFE results

Fig. 4 displays the results of the RFE process wherein features are iteratively removed from the F_{25} feature set until only 1 feature remains. The feature set selected for testing ($\mathrm{F}_{\mathrm{RFE}}$) is that which precedes a 1% reduction in $\mathrm{R}^2{}_{\mathrm{Adj}}$ on the Night data. F_{RFE}, the feature set generated by the RFE process, contains all features from F_E except *TA* and $ER_{Reichstein}$. It is also noted that the temperature information may already be captured sufficiently in the *TW* or *TS* variables. It is evident that the number of features needed for an optimally performing model varies from site to site, indicating the difficulties in determining a universally optimal feature set. For example, at SW a model with just 3 features generalises best on Winter data and generalises better than the feature set chosen by the RFE process for that site. In contrast, a model with 6 features for that site generalises best on Flood data with a reduction in performance in generalising to Winter data.

4.3. Feature importance

Table 1 lists the features selected using the RFE process for *each* site. Feature (F) columns rank features by their importance while Importance (I) columns give the relative proportion of total gain in performance contributed by that feature, normalised to sum to 1. The features that were not included in $\mathrm{F}_{\mathrm{RFE}}$ are denoted by a strike-through. A full list of the features tested and their descriptions can be found in the supplementary material. As noted in Section 4.2, there is an overlap with the features previously selected using domain knowledge and many of the new features selected relate to processes that are known mediators of E and T. Our previous work highlighted the importance of *VPD* and $u*$ as they both relate to energy transport (Eichelmann et al., 2021b) which affects E as it is a form of latent energy. *VPD* is a measure of dryness of the air which increases transport of water across this gradient from high to low moisture and $u*$ is a measure of turbulence which also increases the transport of energy away from the surface. Most of the features identified by this framework can be grouped into their relationships with E as the energy available for evaporation (*TW*, *H*, *RNET*, *TS*, *tbar*), the moisture gradient driving E (*RH*, *VPD* and *WT* to a lesser degree), the turbulent processes transporting water vapour away from the surface ($u*$, *u (mean)*, *uw*, *ww* and *WD*) and the temporal patterns of ET (*year*, *DOY* and *time*). For description of variable labels please refer to Table 2 in the supplementary material.

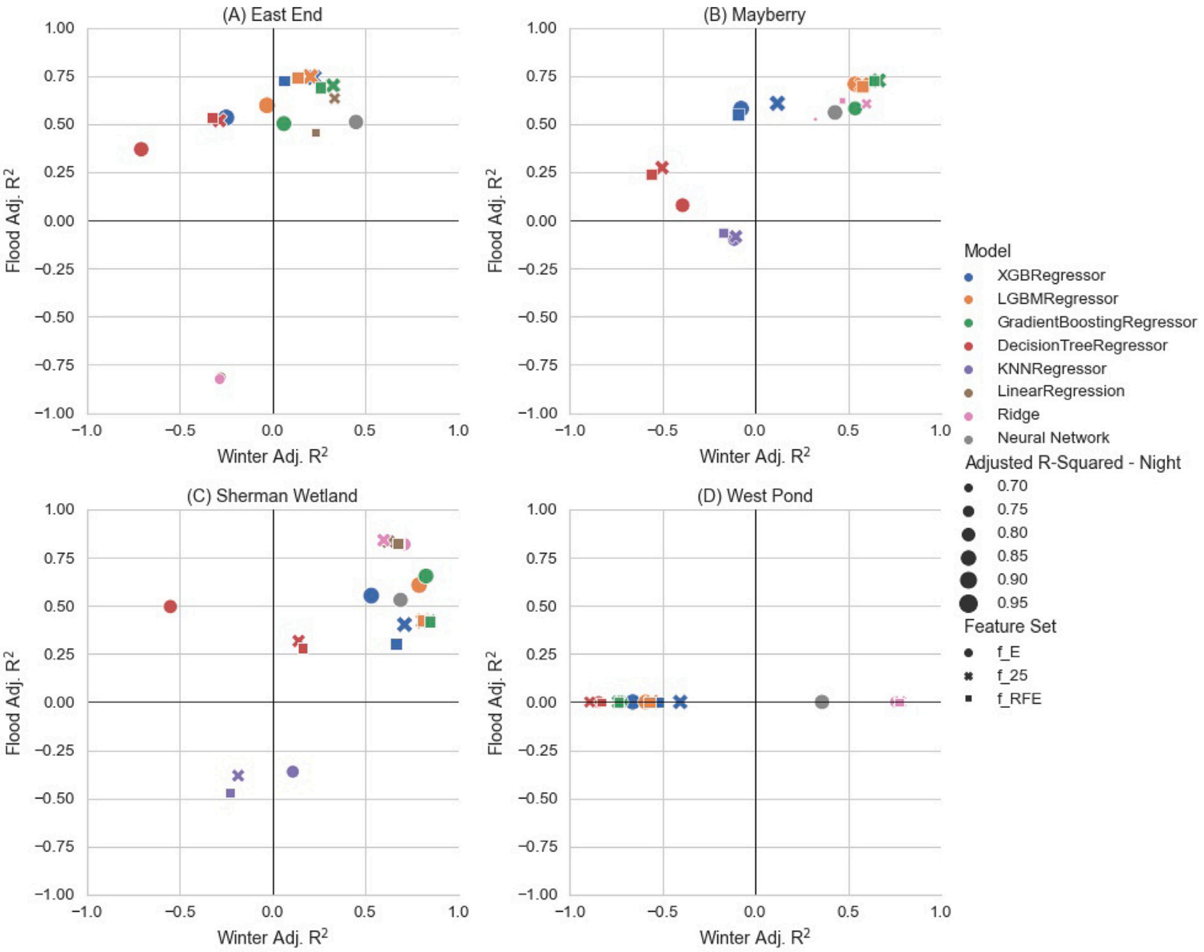

Fig. 3. Results of model comparison for the four sites being studied. The *x*-axis plots the Adjusted R^2 (R$^2_{\text{Adj}}$) values for predictions on data from winter month and the *y*-axis plots the R$^2_{\text{Adj}}$ values for predictions on data from the initial flooding period, testing the ability of the models to generalise to unseen data. The colour of the marker indicates the algorithm used in model building and the shape of the marker indicates the feature set being tested. The size of the marker indicates the R$^2_{\text{Adj}}$ values for predictions on the hold-out Night data, demonstrating how well the models perform on data that is identically distributed to the training data. Therefore, the best performing models are those with the largest markers that are closest to the upper right corner of the graph. The *x*- and *y*-axis lines along the origin are displayed to allow for ease of identification of those models that fail to generalise well (i.e. models with $R^2_{Adj} < 0$). As WP does not have data from the initial flooding period, the results are displayed along the *x*-axis only.

Table 1
Feature importance ranked in order of importance for each site where the features obtained by the RFE process are denoted by F followed by the site label and the relative importance of that feature at that site is given by I followed by the site label. Features that were omitted from the final feature set (F$_{\text{RFE}}$) are indicated by a strike-through, highly important features indicated in bold, features of interest in italics and the threshold for significant feature importance indicated by a horizontal line for each site.

Rank	F (EE)	I (EE)	F (SW)	I (SW)	F (MB)	I (MB)	F (WP)	I (WP)
1	**TW (mean)**	0.282	**u (mean)**	0.546	**u***	0.273	H	0.413
2	**u (mean)**	0.218	VPD	0.179	*wm*	0.146	**u***	0.208
3	*c (mean)*	0.123	u*	0.053	year	0.113	RNET	0.098
4	RH	0.087	H	0.044	RH	0.099	VPD	0.053
5	year	0.060	*wm*	0.032	VPD	0.094	RH	0.048
6	VPD	0.042	TW (mean)	0.028	u (mean)	0.082	DOY	0.024
7	WT	0.037	RH	0.025	TW (mean)	0.052	year	0.021
8	H	0.031	t (mean)	0.020	DOY	0.050	TA	0.018
9	u*	0.030	DOY	0.017	GCC	0.030	uw	0.018
10	uw	0.027	WT	0.015	H	0.027	GCC	0.016
11	DOY	0.015	uw	0.009	WD	0.014	~~uu~~	0.015
12	RNET	0.012	time	0.008			*wm*	0.013
13	TS (mean)	0.011	ww	0.007			WT	0.010
14	ww	0.004	RNET	0.004			WD	0.009
15	*wm*	0.004	GCC	0.004			time	0.008
16	~~ze~~	0.004	TS (mean)	0.004			TW (mean)	0.006
17	GCC	0.004	year	0.003			TS (mean)	0.005
18	~~stat$_q$~~	0.003					~~ts (mean)~~	0.004
19	t (mean)	0.002					~~sos~~	0.004
20	WD	0.002					~~er$_{linear}$~~	0.003
21							~~vv~~	0.003

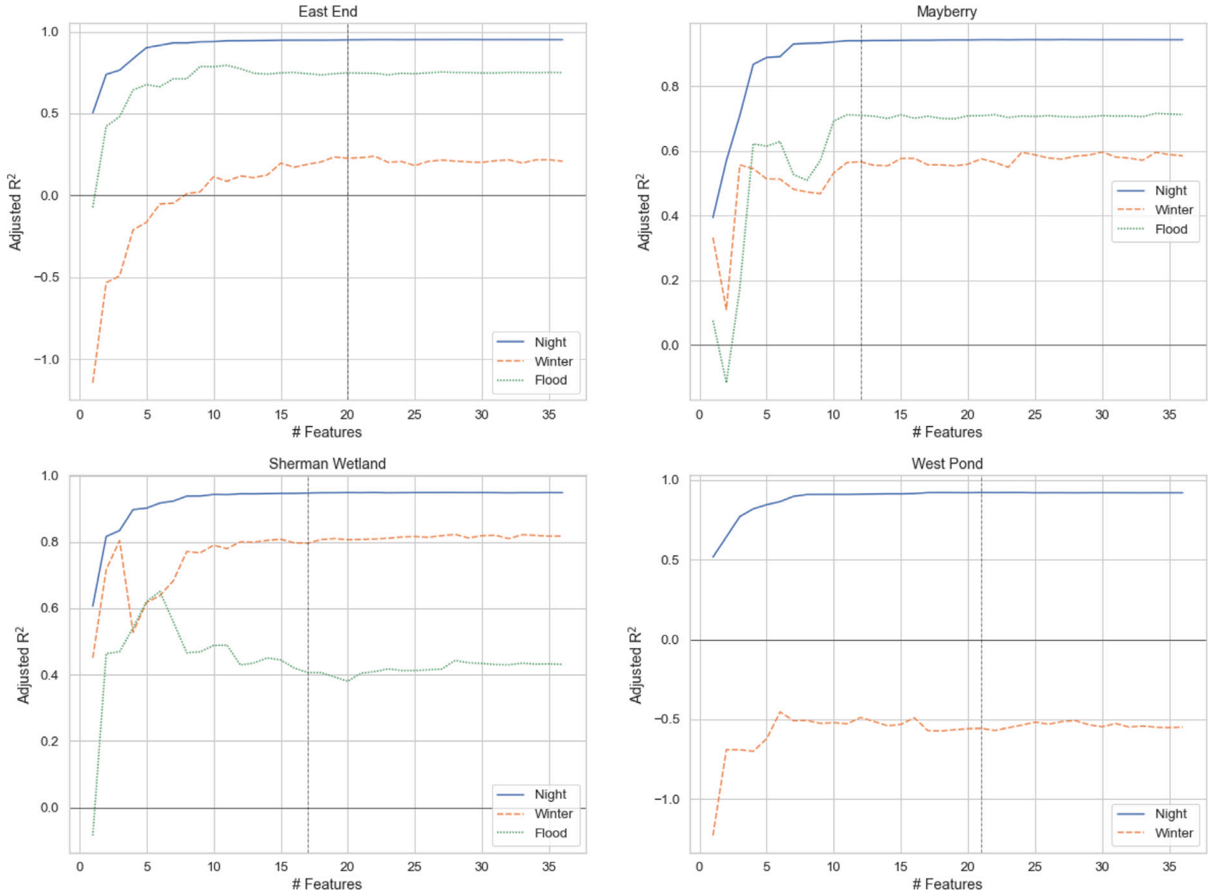

Fig. 4. Results of the RFE process for each of the 4 sites tested with number of features on the *x*-axis and R^2_{Adj} results on the *y*-axis. The iterations start on the right and move towards 0 as RFE iteratively decreases the number of features until only 1 feature remains for each of the sites and each of the test sets; Night, Winter and (where available) Flood. A vertical line on each graph indicates the number of features selected, where the optimal feature set is determined to be the last feature set preceding a 0.1% reduction in R^2_{Adj}.

If a threshold of 0.2 is set for highly important and 0.05 (±10%) for significantly important, an examination of Table 1 indicates 4 features (highlighted in bold) as being of high importance in accurately predicting ET: *u(mean)* at the EE and SW sites; *u** at the MB and WP sites; *H* at the WP site and *TW (mean)* at the EE site. If the features that are deemed to be highly or significantly important are examined it is observed that EE has 6, SW has 4, MB has 8, and WP has 5 features. This indicates that the majority of the predictive performance is attributable to these features. Two variables (highlighted in italics) which were unexpectedly ranked as important were carbon dioxide concentration (*c (mean)*) at EE and methane flux (*wm*) at MB. It is hypothesised that the relevance of *c (mean)* may be due to its connection to microbial activity via soil respiration wherein carbon dioxide and water are transported in the same way. The connection with *wm* is not as clear as there are multiple pathways through which methane can be released; diffusion, ebullition, and plant mediated transport. The fact that *wm* appears as an important predictive feature for *E* could indicate that there is mostly diffusive transport occurring which would follow the same physical processes as evaporation.

Identifying new features may reveal previously unknown connections between components of the system for further study with the potential to improve understanding of the underlying biophysical processes. This process is significantly enabled by this objective and data-driven framework.

While this work focused on using half hourly flux data, recent research on the use of high-frequency (10 or 20 Hz) EC data in the partitioning of methane fluxes (Iwata et al., 2018; Taoka et al., 2020) and in the partitioning of water vapour and carbon dioxide fluxes (Klosterhalfen et al., 2019; Scanlon and Sahu, 2008; Scanlon and Kustas, 2010; Scanlon et al., 2019; Skaggs et al., 2018; Zahn et al., 2022) provide an avenue for further research. The latter utilises the similarity between non-stomatal (respiration and E) and stomatal (photosynthesis and T) components, a methodology that could possibly be amenable to ML techniques or that could serve as a comparison for the outputs of our methodology.

4.4. Additional results

All other sites were tested for the slope of line of best fit between total ET and the predictions for the Night, Winter, Flood periods as well as for the daytime data outside of these periods. The single site for which a slope greater than 1 was observed was SW during the Winter period and only for the Decision Tree and Linear models. It should be noted however that the slopes for other models at SW were quite close to 1, indicating that the modelling of the Winter data for this site requires further investigation as these predictions violate the biophysical constraints outlined in Section 3.5. These findings are in line with those of our previous study (Eichelmann et al., 2021b). The details of these results can be found in the supplementary materials in Tables 4, 6, 8 and 9.

4.5. Limitations

Combining features across sites as part of the RFE feature selection process may have led to the inclusion of features that were site specific i.e. relevant to the predictions at one site but not adding useful information at another site. Therefore, this methodological pipeline may be more useful on a site specific basis to identify useful features for that site only and reduce them to the optimal number of features.

A large percentage of the data for the target has been imputed for all sites and additionally a small percentage of features were imputed with a variety of methods being used for imputation. Building models that use this data carry the errors and limitations of the imputation methods and may introduce noise to the data, particularly where linear interpolation was used. Gap-filling of the target data as well as H and wm was carried out using NN-based methods as discussed in our previous work (Eichelmann et al., 2021b), as well as linear interpolation of the remaining missing data used in these experiments, which may have effects on the error of our models. It is noted however, that there is no clear relationship between the linear interpolation carried out in this study and the performance of the models at any particular site. This is a topic that requires further investigation and the inclusion of more comprehensive methods for gap-filling that are outside the scope of the current work. More information on gaps in the data can be found in the supplementary materials (Figure 5). Assumptions around the onset of the different periods where T is considered to be negligible may also lead to the introduction of noise to the target feature where T could be low but non-negligible.

Further model improvements could be obtained through the tuning and optimisation of the hyper-parameters of the models implemented. This forms a potential direction for further experimentation along with the testing and optimisation of different NN architectures exploiting the feature sets obtained in this research to allow for better comparison with previous modelling efforts (Eichelmann et al., 2021b). An investigation into the performance of the SW models which used the two smaller feature sets obtained from RFE may also yield further model improvements. Further research should focus on determining if these methods generalise to other sites, including other freshwater marsh sites and other sites in the FLUXNET network of different types. The challenge for generalising this methodology to sites of a different land, vegetation or climate class is that the underlying physical assumptions may be different, potentially rendering the method inapplicable or requiring modification of the core methodology. A definitive choice of feature set or algorithm across all sites was not possible from the results of our model comparison, indicating that while some features may be common and of relevance across similar ecosystems (i.e. modelling two different wetlands) some features may be specific to a particular site. It is the combination of these more general features with more specific features that may lead to more accurate data-driven modelling of more heterogeneous systems and the potential identification of previously unknown drivers or mediators of E and T for further study.

5. Conclusions

In this work a new framework by which climate scientists can test the efficacy of multiple ML algorithms and identify suitable predictive features from a high-dimensional candidate set has been presented. The result is a ranking of the candidate algorithms, a generally optimal feature set and an understanding as to how features contribute to model performance (predictive accuracy). For validation, micrometeorological datasets were used with this framework to produce a model with an optimal balance between complexity and model performance. The framework adopts an objective (i.e. without usage of domain knowledge) view of feature selection and demonstrated an improvement on the baseline (Eichelmann et al., 2021b) which used a subjective approach to feature selection.

Algorithm ranking identified that ensemble models (such as LightGBM or Gradient Boosting) or linear parametric models would likely perform well on this task at other sites, generalising well to unseen data. However this was not a universal result, with simpler linear parametric models performing best at WP, indicating that there are key differences between the sites that necessitate an approach that tailors the models to individual sites.

The RFE process identified new features from the data that improved model performance. The use of information gain as a metric to iteratively remove features also allows for a direct comparison as to which features were most important at each site, providing the basis for further work, either in transferring these learnings to new sites or refining the models for these sites.

The examination of feature importance highlighted an obscure biophysical link in the case of carbon dioxide concentration and methane flux which improves our understanding of the physical and biological processes involved.

In conclusion, this method provides new evidence of the contribution of ML to ET partitioning. The independence of the framework from explicit domain knowledge indicates that this approach may be domain agnostic, meaning that this method may have applications on other datasets, either for different EC flux sites or on entirely unrelated data.

Code availability

The code for this work is available at: https://github.com/AdamStapleton/ETPartitioning. Potential users of this code are advised that as this work is still under development it may not be suitable for reuse and requires further testing and refinement before implementation as a publicly available tool for researchers.

CRediT authorship contribution statement

Adam Stapleton: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Elke Eichelmann:** Conceptualization, Methodology, Validation, Data curation, Writing – review & editing, Supervision, Funding acquisition. **Mark Roantree:** Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

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

This work was funded by Science Foundation Ireland through the SFI Centre for Research Training in Machine Learning (18/CRT/6183) and by Science Foundation Ireland Grant Number SFI/12/RC/2289_P2, co-funded by the European Regional Development Fund. Funding for the AmeriFlux data portal was provided by the U.S. Department of Energy Office of Science. The authors would like to thank the site PI Dennis D. Baldocchi for enabling access to the extensive and high quality dataset without which this work would not have been possible. We would also like to acknowledge the technicians Joseph Verfaillie and Daphne Szutu and the many Berkeley Biometeorology Lab postdocs, Ph.D. students and field assistants for their long years of work maintaining the AmeriFlux field sites and collecting and processing the data used in this study.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.acags.2022.100105.

References

Aubinet, M., Vesala, T., Papale, D., 2012. Eddy Covariance: A Practical Guide to Measurement and Data Analysis. Springer Science & Business Media.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., et al., 2001. Fluxnet: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bull. Am. Meteorol. Soc. 82 (11), 2415–2434.

Baldocchi, D., Sturtevant, C., Contributors, F., 2015. Does day and night sampling reduce spurious correlation between canopy photosynthesis and ecosystem respiration?. Agricult. Forest Meteorol. 207, 117–126.

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 10 (3), 1327–1365.

Brutsaert, W., Parlange, M., 1998. Hydrologic cycle explains the evaporation paradox. Nature 396 (6706), 30.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. pp. 785–794.

Detto, M., Baldocchi, D., Katul, G.G., 2010. Scaling properties of biologically active scalar concentration fluctuations in the atmospheric surface layer over a managed peatland. Bound.-Lay. Meteorol. 136 (3), 407–430.

Eichelmann, E., Hemes, K.S., Knox, S.H., Oikawa, P.Y., Chamberlain, S.D., Sturtevant, C., Verfaillie, J., Baldocchi, D.D., 2018. The effect of land cover type and structure on evapotranspiration from agricultural and wetland sites in the sacramento–san joaquin river delta, california. Agricult. Forest Meteorol. 256, 179–195.

Eichelmann, E., Knox, S., Rey Sanchez, C., Valach, A., Sturtevant, C., Szutu, D., Verfaillie, J., Baldocchi, D., 2021a. AmeriFlux US-Tw4 twitchell east end wetland, ver. 11-5, AmeriFlux AMP, (dataset). http://dx.doi.org/10.17190/AMF/1246151.

Eichelmann, E., Mantoani, M.C., Chamberlain, S.D., Hemes, K.S., Oikawa, P.Y., Szutu, D., Valach, A., Verfaillie, J., Baldocchi, D.D., 2021b. A novel approach to partitioning evapotranspiration into evaporation and transpiration in flooded ecosystems. bioRxiv.

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., et al., 2020. Ecostress: Nasa's next generation mission to measure evapotranspiration from the international space station. Water Resour. Res. 56 (4), e2019WR026058.

Gerken, T., Bromley, G.T., Stoy, P.C., 2018. Surface moistening trends in the northern north american great plains increase the likelihood of convective initiation. J. Hydrometeorol. 19 (1), 227–244.

Géron, A., 2019. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc.

Green, J.K., Konings, A.G., Alemohammad, S.H., Berry, J., Entekhabi, D., Kolassa, J., Lee, J.-E., Gentine, P., 2017. Regionally strong feedbacks between the atmosphere and terrestrial biosphere. Nat. Geosci. 10 (6), 410–414.

Han, J., Pei, J., Kamber, M., 2011. Data Mining: Concepts and Techniques, Vol. 3. Elsevier.

Hatala, J.A., Detto, M., Sonnentag, O., Deverel, S.J., Verfaillie, J., Baldocchi, D.D., 2012. Greenhouse gas (co2, ch4, h2o) fluxes from drained and flooded agricultural peatlands in the sacramento-san joaquin delta. Agric. Ecosyst. Environ. 150, 1–18.

Irvin, J., Zhou, S., McNicol, G., Lu, F., Liu, V., Fluet-Chouinard, E., Ouyang, Z., Knox, S.H., Lucas-Moffat, A., Trotta, C., et al., 2021. Gap-filling eddy covariance methane fluxes: Comparison of machine learning model predictions and uncertainties at fluxnet-ch4 wetlands. Agricult. Forest Meteorol. 308, 108528.

Iwata, H., Hirata, R., Takahashi, Y., Miyabara, Y., Itoh, M., Iizuka, K., 2018. Partitioning eddy-covariance methane fluxes from a shallow lake into diffusive and ebullitive fluxes. Bound.-Lay. Meteorol. 169 (3), 413–428.

Jung, M., Reichstein, M., Bondeau, A., 2009. Towards global empirical upscaling of fluxnet eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. Biogeosciences 6 (10), 2001–2013.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S.I., Sheffield, J., Goulden, M.L., Bonan, G., Cescatti, A., Chen, J., De Jeu, R., et al., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. Nature 467 (7318), 951–954.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 30, 3146–3154.

Klosterhalfen, A., Graf, A., Brüggemann, N., Drüe, C., Esser, O., González-Dugo, M.P., Heinemann, G., Jacobs, C.M., Mauder, M., Moene, A.F., et al., 2019. Source partitioning of h2o and co2 fluxes based on high-frequency eddy covariance data: a comparison between study sites. Biogeosciences 16 (6), 1111–1132.

Knox, S.H., Sturtevant, C., Matthes, J.H., Koteen, L., Verfaillie, J., Baldocchi, D., 2015. Agricultural peatland restoration: effects of land-use change on greenhouse gas (co2 and ch4) fluxes in the sacramento-san joaquin delta. Global Change Biol. 21 (2), 750–765.

Laboratory, L.B.N., 2021. Ameriflux data sharing platform. https://ameriflux.lbl.gov/.

Lin, Z., Mo, X., Li, H., Li, H., 2002. Comparison of three spatial interpolation methods for climate variables in china. Acta Geograph. Sinica 57 (1), 47–56.

Matthes, J.H., Sturtevant, C., Oikawa, P., Chamberlain, S.D., Szutu, D., Ortiz, A.A., Verfaillie, J., Baldocchi, D., 2021. AmeriFlux US-myb mayberry wetland, ver. 11-5, AmeriFlux AMP, (dataset). http://dx.doi.org/10.17190/AMF/1246139.

Novick, K.A., Biederman, J., Desai, A., Litvak, M., Moore, D.J., Scott, R., Torn, M., 2018. The ameriflux network: A coalition of the willing. Agricult. Forest Meteorol. 249, 444–456.

O'Connell, J.L., Byrd, K.B., Kelly, M., 2015. A hybrid model for mapping relative differences in belowground biomass and root: shoot ratios using spectral reflectance, foliar n and plant biophysical data within coastal marsh. Remote Sens. 7 (12), 16480–16503.

Oki, T., Kanae, S., 2006. Global hydrological cycles and world water resources. Science 313 (5790), 1068–1072.

Pascolini-Campbell, M., Reager, J.T., Chandanpurkar, H.A., Rodell, M., 2021. A 10 per cent increase in global land evapotranspiration from 2003 to 2019. Nature 593 (7860), 543–547.

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, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pielke, R.A., Avissar, R., Raupach, M., Dolman, A.J., Zeng, X., Denning, A.S., 1998. Interactions between the atmosphere and terrestrial ecosystems: influence on weather and climate. Global Change Biol. 4 (5), 461–475.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., et al., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Global Change Biol. 11 (9), 1424–1439.

Scanlon, T.M., Kustas, W.P., 2010. Partitioning carbon dioxide and water vapor fluxes using correlation analysis. Agricult. Forest Meteorol. 150 (1), 89–99.

Scanlon, T.M., Sahu, P., 2008. On the correlation structure of water vapor and carbon dioxide in the atmospheric surface layer: A basis for flux partitioning. Water Resour. Res. 44 (10).

Scanlon, T.M., Schmidt, D.F., Skaggs, T.H., 2019. Correlation-based flux partitioning of water vapor and carbon dioxide fluxes: Method simplification and estimation of canopy water use efficiency. Agricult. Forest Meteorol. 279, 107732.

Schlesinger, W.H., Jasechko, S., 2014. Transpiration in the global water cycle. Agricult. Forest Meteorol. 189, 115–117.

Shortt, R., Hemes, K., Szutu, D., Verfaillie, J., Baldocchi, D., 2021. AmeriFlux US-sne sherman island restored wetland, ver. 7-5, AmeriFlux AMP, (dataset). http://dx.doi.org/10.17190/AMF/1418684.

Skaggs, T.H., Anderson, R.G., Alfieri, J., Scanlon, T., Kustas, W., 2018. Fluxpart: Open source software for partitioning carbon dioxide and water vapor fluxes. Agricult. Forest Meteorol. 253, 218–224.

Stoy, P., El-Madany, T., Fisher, J., Gentine, P., Gerken, T., Good, S., Liu, S., Miralles, D., Perez-Priego, O., Skaggs, T., et al., 2019. Reviews and syntheses: Turning the challenges of partitioning ecosystem evaporation and transpiration into opportunities. Biogeosci. Discuss.

Taoka, T., Iwata, H., Hirata, R., Takahashi, Y., Miyabara, Y., Itoh, M., 2020. Environmental controls of diffusive and ebullitive methane emissions at a subdaily time scale in the littoral zone of a midlatitude shallow lake. J. Geophys. Res.: Biogeosci. 125 (9), e2020JG005753.

Tramontana, G., Jung, M., Schwalm, C.R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M.A., Cescatti, A., Kiely, G., et al., 2016. Predicting carbon dioxide and energy fluxes across global fluxnet sites with regression algorithms. Biogeosciences 13 (14), 4291–4313.

Tramontana, G., Migliavacca, M., Jung, M., Reichstein, M., Keenan, T.F., Camps-Valls, G., Ogee, J., Verrelst, J., Papale, D., 2020. Partitioning net carbon dioxide fluxes into photosynthesis and respiration using neural networks. Global Change Biol. 26 (9), 5235–5253.

Trenberth, K.E., Fasullo, J.T., Kiehl, J., 2009. Earth's global energy budget. Bull. Am. Meteorol. Soc. 90 (3), 311–324.

Valach, A.C., Kasak, K., Hemes, K.S., Anthony, T.L., Dronova, I., Taddeo, S., Silver, W.L., Szutu, D., Verfaillie, J., Baldocchi, D.D., 2021a. Productive wetlands restored for carbon sequestration quickly become net co2 sinks with site-level factors driving uptake variability. PLoS One 16 (3), e0248398.

Valach, A., Szutu, D., Eichelmann, E., Knox, S., Verfaillie, J., Baldocchi, D., 2021b. AmeriFlux US-Tw1 twitchell wetland west pond, ver. 9-5, AmeriFlux AMP, (dataset). http://dx.doi.org/10.17190/AMF/1246147.

Zahn, E., Bou-Zeid, E., Good, S.P., Katul, G.G., Thomas, C.K., Ghannam, K., Smith, J.A., Chamecki, M., Dias, N.L., Fuentes, J.D., et al., 2022. Direct partitioning of eddy-covariance water and carbon dioxide fluxes into ground and plant components. Agricult. Forest Meteorol. 315, 108790.

Zeng, Z., Piao, S., Li, L.Z., Wang, T., Ciais, P., Lian, X., Yang, Y., Mao, J., Shi, X., Myneni, R.B., 2018. Impact of earth greening on the terrestrial water cycle. J. Clim. 31 (7), 2633–2650.