Reviewer comments:

In the revised manuscript, I believe that most of my concerns from my previous rounds have been addressed. I have no further comments and do appreciate the authors' efforts to improve this manuscript. As a reminder, please make sure to check so that there are no similar sentences in the two papers.

Response & Actions:

Thank you for your suggestion. By using the 'compare' function of MS Word, we have checked (on the next attached page) and modified the found similar/duplicate sentences. The following changes were made to similar sentences to make the text different enough from the BG paper.

Similar sentences/ duplicate words (red)
Revised (blue)

However, systematic evaluation of such models is still limited. We therefore performed a meta-analysis of 32 such studies, derived 139 model records, and evaluated the impact of various features on model accuracy throughout the modeling flow.

However, it is unclear how various model features affect prediction accuracy. To fill this gap, we evaluated this issue based on records of 139 developed models collected from 32 such studies. **(line 20)**

Publication bias and weighting: Due to the relatively limited number of articles that could be included in the meta-analysis, this study did not focus much on publication bias. Meta-analytic studies in other fields typically measure the quality of journals and the public availability of research data (Borenstein et al., 2011; Field and Gillett, 2010) to determine the weighting of the literature in a comprehensive assessment. However, most of the articles did not publicly provide flux observations or share developed models. Meta-analysis studies in other fields typically measure the impact of included studies based on sample size and variance of experimental results (Adams et al., 1997; Don et al., 2011; Liu et al., 2018a). In this study, due to the lack of a convincing manner to determine weights among articles, we assigned the same weight to the results for all the literature.

Publication bias and weighting: In a meta-analysis in other fields, weights for different studies can be assigned based on the quality of the journal and the extent to which the research data are publicly available (Borenstein et al., 2011; Field and Gillett, 2010). However, most of the articles included in this study did not fully publish the flux data they used, the models they developed, and the predicted ET data. Given this limitation, we were unable to assign them small weights due to the relatively limited available sample size of this study. Besides, in meta-analyses in other fields, the sample size and the variance of the results of the experiments can also be used to adjust the weights of the effects among studies (Adams et al., 1997; Don et al., 2011; Liu et al., 2018a). However, for this study, due to the lack of a convincing way to determine the weights, we briefly assigned equal weight values to all the included studies. **(line 417)**

Attached Duplicate Check by MS Word: (Duplicate text is displayed in black)

~~Variability and Uncertainty in Flux-Site Scale Net Ecosystem Exchange Simulations Based on Machine Learning and Remote Sensing: A Systematic Evaluation~~

Evaluation of water flux predictive models developed using eddy covariance observations and machine learning: a meta-analysis

Haiyang Shi[1,2,4,5], Geping Luo[1,2,3,5], Olaf Hellwich[6], Mingjuan Xie[1,2,4,5], Chen Zhang[1,2], Yu Zhang[1,2], Yuangang Wang[1,2], Xiuliang Yuan[1], Xiaofei Ma[1], Wenqiang Zhang[1,2,4,5], Alishir Kurban[1,2,3,5], Philippe De Maeyer[1,2,4,5] and Tim Van de Voorde[4,5]

[1] State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, Xinjiang, 830011, China.

[2] University of Chinese Academy of Sciences, 19 (A) Yuquan Road, Beijing, 100049, China.

[3] Research Centre for Ecology and Environment of Central Asia, Chinese Academy of Sciences, Urumqi, China.

[4] Department of Geography, Ghent University, Ghent 9000, Belgium.

[5] Sino-Belgian Joint Laboratory of Geo-Information, Ghent, Belgium and Urumqi, China.

[6] Department of Computer Vision & Remote Sensing, Technische Universität Berlin, 10587 Berlin, Germany.

Correspondence to: **Geping Luo (luogp@ms.xjb.ac.cn) and Olaf Hellwich (olaf.hellwich@tu-berlin.de)**

Submitted to ~~*Biogeosciences*~~*Hydrology and Earth System Sciences*

带格式的: 行距: 1.5 倍行距

设置了格式: 字体: 加粗

设置了格式: 字体: 加粗, 字体颜色: 文字 1, 英语(美国)

设置了格式: 字体: 加粗, 字体颜色: 文字 1, 英语(美国)

设置了格式: 字体: 加粗, 字体颜色: 文字 1, 英语(美国)

设置了格式: 字体: 加粗, 字体颜色: 文字 1, 英语(美国)

带格式的: 段落间距段前: 0 磅

设置了格式: 字体: 加粗, 字体颜色: 文字 1, 英语(美国)

设置了格式: 字体: 加粗, 字体颜色: 文字 1, 英语(美国)

设置了格式: 字体颜色: 文字 1, 英语(美国)

设置了格式: 字体: 10 磅

Abstract. ~~Net ecosystem exchange (NEE) is an important indicator of carbon cycling in terrestrial ecosystems. Many previous studies have combined~~

With the rapid accumulation of water flux observations~~, meteorological, biophysical, and ancillary predictors using~~ from global eddy-covariance flux sites, many studies have used data-driven approaches to model water fluxes with various predictors and machine learning ~~to simulate the site-scale NEE.~~algorithms used. However, systematic evaluation of ~~the performance of~~ such models is still limited. ~~Therefore, we~~We therefore performed a meta-analysis of ~~these NEE simulations. A total of 40~~32 such studies ~~and 178~~, derived 139 model records ~~were included. The impacts~~, and evaluated the impact of various features on model accuracy throughout the modeling ~~process on the accuracy of the model were evaluated. Random Forests and Support Vector Machines performed better than other algorithms. Models with larger time scales have lower average R-squared, especially when the time scale exceeds the monthly scale. Half-hourly models~~flow. SVM (average R-squared = 0.~~73) were significantly more accurate than daily models~~82) and RF (average R-squared = 0.~~5~~). There are significant differences in the predictors used and their impacts on model accuracy for different plant functional types (PFTs). Studies at continental and global scales~~81) outperformed over evaluated algorithms with sufficient sample size in both cross-study and intra-study (with the same data) comparisons. The average accuracy of the model applied to arid regions is higher than in other climate types. The average accuracy of the model was slightly lower for forest sites (average R-squared = 0.~~37) with multiple PFTs, more sites, and a large span of years correspond to lower R-squared than studies at local~~76) than for croplands and grasslands (average R-squared = 0.~~69)~~8 and 0.79), but higher than for shrubland sites (average R-squared = 0.67). Using Rn/Rs, precipitation, Ta, and ~~regional scales (average R-squared = 0.7). Also, the site-scale NEE predictions need more focus on the internal heterogeneity of the NEE dataset~~FAPAR improved the model accuracy. The combined use of Ta and ~~the matching of the training set~~Rn/Rs is very effective especially in forests, while in grasslands the combination of Ws and Rn/Rs is also effective. Random cross-validation ~~set~~showed higher model accuracy than spatial cross-validation and temporal cross-validation, but spatial cross-validation is more important in spatial extrapolation. The findings of this study are promising to guide future research on such machine learning-based modeling.

设置了格式: 英语(美国)

设置了格式: 英语(美国)

1 Introduction

带格式的: 行距: 1.5 倍行距

Net ecosystem exchange (NEE) of CO_2 is an important indicator of carbon cycling in terrestrial ecosystems (Fu et al., 2019), and accurate estimation of NEE is important for the development of global carbon neutral policies. Although process-based models have been used for NEE simulations (Mitchell et al., 2009), their accuracy and spatial resolutions of the model outputs are limited probably due to the lack of understanding and quantification of complex processes. Many researchers have tried to use a data-driven approach as an alternative (Fu et al., 2014; Tian et al., 2017; Tramontana et al., 2016; Jung et al., 2011). On the one hand, it was made possible by the increase in the growth of global carbon flux observations and the large amount of flux observation data being accumulated. Since the 1990s, the use of the eddy covariance technique to monitor NEE has been rapidly promoted (Baldocchi, 2003). Several regional and global flux measurement networks have been established for the big data management of the flux sites, including CarboEuro-flux (Europe), AmeriFlux (North America), OzFlux (Australia), ChinaFlux (China), FLUXNET (global), etc. On the other hand, machine learning approaches are increasingly used to extract patterns and insights from the ever-increasing stream of geospatial data (Reichstein et al., 2019). The rapid development of various algorithms and high public availability of model tools in the field of machine learning have made these techniques easily available to more researchers in the field of geography and ecology (Reichstein et al., 2019). Since the above two major advances (i.e., increasing availability of flux data and machine learning techniques) in the last two decades, various machine learning algorithms have been used to simulate NEE at the flux station scale with various predictor variables (e.g., meteorological variables, biophysical variables) incorporated for spatial and temporal mapping of NEE or understanding the driving mechanisms of NEE.

To date, studies on using machine learning to predict NEE have a high diversity in terms of modeling approaches. To obtain a comprehensive understanding of machine learning-based NEE prediction, a synthesis evaluation of these machine learning models is necessary. Since the beginning of this century, when machine learning approaches were still rarely used in geography and ecology research, neural networks were already used to perform simulations and mapping of NEE in European forests (Papale and Valentini, 2003). Subsequently, considerable efforts have been made by researchers to improve such predictive models. Many studies have demonstrated the effectiveness of their proposed improvements

~~(i.e., using predictors with a higher spatial resolution (Reitz et al., 2021) and using data from the local flux site network (Cho et al., 2021)) by comparing with previous studies. However, the improvements achieved in these studies may be limited to smaller areas and specific conditions and may not be generalizable (Cleverly et al., 2020; Reed et al., 2021; Cho et al., 2021). We are more interested in guidelines with universal applicability that improve the model accuracy, such as the selection of appropriate predictors and algorithms under different conditions. Therefore, we should synthesize the results of models applied to different conditions and regions to obtain general insights.~~

~~Many factors may affect the performance of these NEE prediction models, such as the predictor variables, the spatial and temporal span of the observed flux data~~Evapotranspiration (ET) is one of the most important components of the water cycle in terrestrial ecosystems. It also represents the key variable in linking ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources (Fisher et al., 2017). The quantification of ET for regional, continents, or the globe can improve our understanding of the water, heat, and carbon interactions, which is important for global change research (Xu et al., 2018). Information on ET has been used in many fields, including, but not limited to, droughts and heatwaves (Miralles et al., 2014), regional water balance closures (Chen et al., 2014; Sahoo et al., 2011), agricultural management (Allen et al., 2011), water resources management (Anderson et al., 2012), biodiversity patterns (Gaston, 2000). In addition, accurate large-scale and long-time series ET prediction at high spatial and temporal resolution has been of great interest (Fisher et al., 2017).

Currently, there are three main approaches for simulation and spatial and temporal prediction of ET: (i) physical models based on remote sensing such as surface energy balance models (Minacapilli et al., 2009; Wagle et al., 2017), Penman-Monteith equation (Mu et al., 2011; Zhang et al., 2010), Priestley-Taylor equation (Miralles et al., 2011); (ii) process-based land surface models, biogeochemical models and hydrological models (Barman et al., 2014; Pan et al., 2015; Sándor et al., 2016; Chen et al., 2019); and (iii) the observation-based machine learning modeling approach with in situ eddy covariance (EC) observations of water flux (Jung et al., 2011; Li et al., 2018; Van Wijk and Bouten, 1999; Xie et al., 2021; Xu et al., 2018; Yang et al., 2006; Zhang et al., 2021). For remote sensing-based physical models and process-based land surface models, some physical

processes have not been well characterized due to the lack of understanding of the detailed mechanisms influencing ET under different environmental conditions. For example, the inaccurate representation and estimation of stomatal conductance (Li et al., 2019) and the linearization (McColl, 2020) of the Clausius-Clapeyron relation in the Penman-Monteith equation may introduce both empirical and conceptual errors into estimates of ET. Limited by complicated assumptions and model parametrizations, these process-based models face challenges in the accuracy of their ET estimations over heterogeneous landscapes (Pan et al., 2020; Zhang et al., 2021). Therefore, many researchers have used data-driven approaches for the simulation and prediction of ET with the accumulation of a large volume of measured observational data of water fluxes in the past decades. Various machine learning models have been developed to simulate water fluxes at the flux site scale. Besides, various predictor variables (e.g., meteorological factors, vegetation conditions, and moisture supply conditions) have been incorporated into such models for upscaling (Fang et al., 2020; Jung et al., 2009) of water flux to a larger scale or understanding the driving mechanisms with the variable importance analysis performed in such models.

However, to date, the systematic assessment of the uncertainty in the processes of water flux prediction models constructed using the machine learning approach is limited. Although considerable effort has been invested in improving the accuracy of such prediction models, our understanding of the expected accuracy of such models under different conditions is still limited. It is still not easy for us to give the general guidelines for selecting appropriate predictor variables and models. Questions such as 'Which predictor variables are the best in water flux simulations?' and 'How to improve the prediction accuracy of water flux effectively?' etc. still confuse the researchers in the field. Therefore, we should synthesize the findings from published studies to determine which predictor variables, machine learning models, and other features can significantly improve the prediction accuracy of water flux. Also, we are interested in understanding under which specific conditions they are more effective.

A variety of features control the accuracy of such models, including the predictor variables used, the inherent heterogeneity within the dataset, the plant functional type (PFT) of the flux sites, the method of model construction and validation ~~method~~, and the ~~machine learning~~ algorithm ~~used~~.

带格式的: 正文, 行距: 1.5 倍行距
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅

~~as described below:~~ chosen:

~~a) Predictors: Various biophysical variables (Zeng et al., 2020; Cui et al., 2021; Huemmrich et al., 2019) and other meteorological and environmental factors have been used in the simulation of NEE. The most commonly used predictor variables include precipitation (Prec), air temperature (Ta), wind speed (Ws), net/sun radiation (Rn/Rs), soil temperature (Ts), soil texture, soil moisture (SM) (Zhou et al., 2020), vapor pressure deficit (VPD) (Moffat et al., 2010; Park et al., 2018), the fraction of absorbed photosynthetically active radiation (FAPAR) (Park et al., 2018; Tian et al., 2017), vegetation index (e.g., NDVI, EVI), LAI, and evapotranspiration (ET) (Berryman et al., 2018). The predictor variables used vary with the natural conditions and vegetation functional types of the study area. In contrast, in models that include multiple PFTs, some variables that play a significant role in the prediction of each of the multiple PFTs may have higher importance. For example, growing degree days (GDD) may be a more effective variable for NEE of tundra in the northern hemisphere high latitudes (Virkkala et al., 2021), while measured groundwater levels may be important for wetlands (Zhang et al., 2021). Some of these predictor variables are measured at flux stations (e.g., meteorological factors such as precipitation and temperature), while others are extracted from reanalyzed meteorological datasets and satellite remote sensing image data (e.g., vegetation indices). The spatial and temporal resolution of predictors can lead to differences in their relevance to NEE observations. Most measured in situ meteorological factors have a good spatio-temporal match to the observed NEE (site scale, half-hourly scale). However, the proportion of NEE explained by remotely sensed biophysical covariates may depend on their spatial and time scales. For example, the MODIS based 8-daily NDVI data may better capture temporal variation in the relationship between NEE and vegetation growth than the Landsat-based 16-daily NDVI data. In contrast, the interpretation of NEE by variables such as soil texture and soil organic content (SOC), which do not have temporal dynamic information, may be limited to the interpretation of spatial variability, although they are considered to be important drivers of NEE. Therefore, the importance of variables obtained from NEE simulations based on a data-driven approach may differ from that in process-based models as well as in the actual driving mechanisms. This may be related to the spatial and temporal resolution of the predictors used and the quality of the data. It is necessary to consider the spatio-temporal resolution of the data for the actual biophysical variables used in the different studies in the systematic evaluation of data-driven NEE simulations.~~

b) The spatio-temporal heterogeneity of data sets, and validation method: The spatio-temporal heterogeneity of the dataset may affect model accuracy. Typically, training data with larger regions, multiple sites, multiple PFTs, and longer spans of years may have a higher degree of imbalance (Kaur et al., 2019; Van Hulse et al., 2007; Virkkala et al., 2021; Zeng et al., 2020). Modeling with unbalanced data (where the difference between the distribution of the training and validation sets is significant even if selected at random) may result in lower model accuracy. To date, the most commonly used methods for validating such models include spatial (Virkkala et al., 2021), temporal (Reed et al., 2021), and random (Cui et al., 2021) cross-validation. The imbalance of data between the training and validation sets may affect the accuracy of the models when using these validation methods. Spatial validation is used to assess the ability of the model to adapt to different regions or flux sites of different PFTs, and a common method is 'leave one site out' cross-validation (Virkkala et al., 2021; Zeng et al., 2020). If the data from the site left out is not covered (or partially covered) by the distribution of the training dataset, the model's prediction performance at that site may be poor due to the absence of a similar type in the training set. Temporal validation typically uses some years of data as training and the remaining years as validation to assess the model's fitness for interannual variability. For a year that is left out (e.g. a special extreme drought year which does not occur in the training set), the accuracy of the model may be limited if there are no similar years (extreme drought years) in the training dataset. K-fold cross-validation is commonly used in random cross-validation to assess the fitness of the model to the spatio-temporal variability. In this case, different values of K may also have a significant impact on the model accuracy. For example, for an unbalanced dataset, the average model accuracy obtained from a 10-fold ($K = 10$) validation approach is likely to be higher than that of a 3-fold ($K = 3$) validation approach (Marcot and Hanea, 2021).

c) Machine learning algorithms used: Simulating NEE using different machine learning algorithms may influence the model accuracy, which may be induced by the characteristics of these algorithms themselves and the specific data distribution of the NEE training set. For example, Neural Networks can be used effectively to deal with nonlinearities, while as an ensemble learning method, Random Forests can avoid overfitting due to the introduction of randomness. Therefore, a comprehensive evaluation of this is necessary.

~~In this study, to evaluate the impacts of predictors use, algorithms, spatial/time scale, and validation methods on model accuracy, we performed a meta-analysis of papers with prediction models that combine NEE observations from flux towers, various predictors, and machine learning for the data-driven NEE simulations. In addition, we also analyzed the causality of multiple features in NEE simulations and the joint effects of multiple features on model accuracy using the Bayesian Network (BN) (a multivariate statistical analysis approach (Pearl, 1985)). The findings of this study can provide some general guidance for future NEE simulations.~~

a) Predictor variables used: Compared to process-based models, the data used may have a more significant impact on the final model performance in data-driven models. Various biophysical covariates and other environmental factors have been used for the simulation and prediction of water fluxes. The most commonly used factors include mainly precipitation (Prec), air temperature (Ta), wind speed (Ws), net/sun radiation (Rn/Rs), soil temperature (Ts), soil texture, vapor-pressure deficit (VPD), the fraction of absorbed photosynthetically active radiation (FAPAR), vegetation index (e.g., Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI)), Leaf area index (LAI), and carbon fluxes (e.g., Gross Primary Productivity (GPP)). These used predictor variables and their complex interactions drive the fluctuations and variability of water fluxes. They affect the accuracy of water flux simulations in two ways: their actual impact on water fluxes at the process-based level and their spatio-temporal resolution and inherent accuracy. The relationship between water fluxes and these variables at the process-based driving mechanism level is very different under different PFTs, different climate types, and different hydrometeorological conditions. For example, in irrigated croplands in arid regions, water fluxes may be highly correlated with irrigation practices, and thus soil moisture may be a very important predictor variable, and its importance may be significantly higher than in other PFTs. And in models that incorporate data from multiple PFTs, some variables that play important roles in multiple PFTs may have higher importance. In terms of data spatial and temporal resolution, the data for these predictor variables may have different scales. In terms of spatial resolution, meteorological observations such as precipitation and air temperature are at the flux site scale, while factors extracted from satellite remote sensing and reanalysis climate datasets cover a much larger spatial scale (i.e. the grid-scale). This leads to considerable differences in the degree of spatial match between different variables and the site scale EC observations (approximately 100 m

x 100 m). It is therefore difficult for some variables to be fairly compared in the subsequent importance analysis of driving factors. In terms of temporal resolution, the importance of predictor variables with different temporal resolutions may be variable for models with different time scales (e.g., half-hourly, daily, and monthly models). For example, the daily or 8-day NDVI data based on MODIS satellite imagery may better capture the temporal dynamics of water fluxes concerning vegetation growth than the 16-daily NDVI data derived from Landsat images. Besides, data on non-temporal dynamic variables such as soil texture cannot explain temporal variability in water fluxes in the data-driven simulations, although soil texture may be important in the interpretation of the actual driving mechanisms of ET (which may need to be quantified in detail in ET simulations by process-based models). In addition, some inherent accuracy issues (e.g., remote sensing-based NDVI may not be effective at high values) of the predictors may propagate into the consequent machine learning models, thus affecting the modeling and our understanding of its importance. Therefore, it is necessary to consider the spatial and temporal resolution of the data and their inherent accuracy for the predictors used in different studies in the systematic evaluation of data-driven water flux simulations.

b) The heterogeneity of the dataset and model validation: the volume and inherent spatiotemporal heterogeneity of the training dataset (with more variability and extremes incorporated) may affect model accuracy. Typically, training data with larger regions, multiple sites, multiple PFTs, and longer year spans may have a higher degree of imbalance (Kaur et al., 2019; Van Hulse et al., 2007; Virkkala et al., 2021; Zeng et al., 2020). And in machine learning, in general, modeling with unbalanced data (with significant differences in the distribution between the training and validation sets) may result in lower model accuracy. Currently, the most common ways of model validation include spatial, temporal, and random cross-validation. Spatial validation is mainly to evaluate the ability of the model to be applied in different regions or flux sites with different PFT types, and one of the common methods is 'leave one site out' (Fang et al., 2020; Papale et al., 2015; Zhang et al., 2021). If the data of the site left out for validation differs significantly from the distribution of the training data set, the expected accuracy of the model applied at that site may be low because the trained model may not capture the specific and local relationships between the water flux and the various predictor variables at that site. For temporal validation, to assess the ability of the models to adapt to the interannual variability, typically some years of data are used for training and

the remaining years for model validation (Lu and Zhuang, 2010). If a year with extreme climate is used for validation, the accuracy may be low because the training dataset may not contain such extreme climate conditions. In the case of PFTs that are significantly affected by human activities, such as cropland, the possible different crops grown and different land use practices (e.g., irrigation) across years can also lead to low accuracy in temporal validation.

c) Various machine learning algorithms: Some machine learning algorithms may have specific advantages when applied to model the relationships between water fluxes and covariates. For example, neural networks may have an advantage in nonlinear fitting, while random forests can avoid serious overfitting problems. However, which algorithm is better overall in different situations (i.e. applied to different data sets)? Which algorithm is generally more accurate than the others when using the same data set? A comprehensive evaluation is important.

Therefore, to systematically and comprehensively assess the impact of various features in such modeling, we perform a meta-analysis of published water flux simulation studies that combine the flux site water flux observations, various predictors, and machine learning. The accuracy of model records collected from the literature was linked with various model features to assess the impacts of predictor data types, algorithms, and other features on model accuracy. The findings of this study may be promising to improve our understanding of the impact of various features of the models to guide future research on such machine learning-based modeling.

2 Methodology

2.1 ~~Criteria~~Protocol for ~~including~~selecting the sample of articles

~~In the Scopus database, a literature query was~~We applied ~~to titles, abstracts~~a general query (on December 1st, 2021) on title, abstract, and keywords to include articles with the "OR" operator applied among expressions (Table 1) ~~according to~~in the Scopus database. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al., 2009) ~~(Fig. 1):~~are followed when filtering the papers. We first excluded articles that obviously did not fit the topic of this study based on the abstract, and then performed the article screening with the full-text

带格式的: 行距: 1.5 倍行距
设置了格式: 字体: 10 磅, 加粗
设置了格式: 字体: 10 磅, 加粗
设置了格式: 字体: 10 磅, 加粗
带格式的: 标题 2
设置了格式: 字体: 10 磅
带格式的: 正文, 行距: 1.5 倍行距
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅

reading.

The inclusion of articles follows the following criteria:

a) Articles were filtered for those ~~that modeled NEE. Articles that modeled other carbon~~ with water fluxes ~~such as methane~~(or latent heat) simulated.

~~a) The water~~ flux ~~were not included.~~

b) ~~Articles that~~or latent heat observations used in the prediction models should be from the eddy-covariance flux measurements.

c) Articles focusing only ~~univariate~~on gap-filling (Hui et al., 2004) techniques (i.e., the objective was not simulation and extrapolation of water fluxes using machine learning) were excluded.

~~b)~~d) Only articles that used multivariate regression ~~rather~~(with the number of covariates greater than ~~multiple regression~~or equal to 3) were ~~screened out.~~ included.

~~c) Articles reported the determination coefficient (R-squared) of the validation step (Shi et al., 2021; Tramontana et al., 2016; Zeng et al., 2020) as the measure of model performance. Although RMSE is also often used for model accuracy assessment, its dependence on the magnitude of water flux values makes it difficult to use for fair comparisons between studies.~~

~~d) Articles were published in journals with language limited to English.~~

~~e) Articles were filtered for those that were published in the specific journals (Table S1) for research quality control because the data, model implements, and peer review in these journals are often more reliable.~~

e) The determination coefficient (R-squared) of the validation step should be reported as the metric of model performance (Shi et al., 2021; Tramontana et al., 2016; Zeng et al., 2020) in the articles.

f) The articles should be published in English-language journals.

Although RMSE is also often used for model accuracy assessment, its dependence on the magnitude of water flux values makes it difficult to use for fair comparisons between studies. For example, due to the difference in the range of ET values, models developed from flux stations in dry grasslands will typically have lower RMSE than models developed by flux stations based on forests in humid regions. Therefore, RMSE may not be a good metric for cross-study comparisons

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

带格式的: 列表段落, 行距: 1.5 倍行距, 编号 + 级别: 1 + 编号样式: a, b, c, … + 起始编号: 1 + 对齐方式: 左侧 + 对齐位置: 0 厘米 + 缩进位置: 0.74 厘米

设置了格式: 字体: 10 磅

in this meta-analysis.

设置了格式: 字体: 10 磅

2.2 Features of the prediction ~~models~~processes evaluated

设置了格式: 字体: 10 磅, 加粗

设置了格式: 字体: 10 磅, 加粗

带格式的: 标题 2

~~Typically, the flow of the NEE prediction modeling framework (Fig. 2) based on flux observations and machine learning is as follows: first, half-hourly scale NEE flux observations are aggregated into various time scale NEE data, and gap-filling techniques (Moffat et al., 2007) are often used in this step to obtain complete NEE series when data are missing. Various predictors including meteorological variables, remote sensing-based biophysical variables, etc. are extracted to match site-scale NEE series to generate a training dataset containing the target variable NEE and various covariates. Subsequently, various algorithms are used for the NEE prediction model construction and validated in different ways (e.g., leave-one-site-out validation (Zeng et al., 2020)). Finally, in some studies, prediction models were applied to gridded covariate data to map the regional or global-scale NEE spatial and temporal variations (Zeng et al., 2020; Papale and Valentini, 2003; Jung et al., 2020). The information of R-squared (at the validation phase) and the associated model features reported in the article are considered as one data record for the formal meta-analysis (i.e., each R-squared record corresponding to a prediction model). From the included papers, R-squared records and various features (Table 2) involved in the NEE modeling framework (Fig. 2) were extracted (including the used algorithms, modeling/validation methods, remote sensing data, meteorological data, biophysical data, and ancillary data). In some studies, multiple algorithms were applied to the same dataset, or models with different features were developed (Virkkala et al., 2021; Zhang et al., 2021; Cleverly et al., 2020; Tramontana et al., 2016). In these cases, multiple data records will be documented.~~

~~In the practical information extracting step, we categorized such features in a comparable manner. First, we categorized the various algorithms used in these papers, although the same algorithm may also have a variant form or an optimized parameter scheme. They are categorized into the following families of algorithms:~~ The various features (Table 2) involved in the water flux modeling framework (Fig. 1) include the PFTs of the sites, the predictors used, the

machine learning algorithms, the validation methods, and other features. Each model for which R-squared is reported is treated as a data record. If multiple algorithms were applied to the same dataset, then multiple records were extracted. Models using different data or features are also recorded as multiple records.

~~Random Forests (RF), Multiple Linear Regressions (MLR), Artificial Neural Networks (ANN), Support Vector Machines (SVM), Partial Least Squares Regression (PLSR), Generalized additive model (GAM), Boosted Regression Tree (BRT), Bayesian Additive Regression Trees (BART), Cubist, model tree ensembles (MTE). Second, we classified the spatial scales of these studies. Models with study areas (spatial extent covered by flux stations) smaller than 100x100 km were classified as 'local' scale models, those with study area sizes exceeding continental scale were classified as 'global' scale, and those with study area sizes in between were classified as 'regional' scale. Third, for various predictors, we only recorded whether the predictors were used or not without distinguishing the detailed data sources and categories (e.g., grid meteorological data from various reanalysis datasets and in-situ meteorological observations from flux stations), measurement methods (e.g., soil moisture measured/estimated by remote sensing or in situ sensors), etc. Fourth, we documented PFTs for the prediction models from the description of study areas or sites in these papers. They are classified into the following types: forest, grassland, cropland, wetland, savannah, tundra, and multi-PFTs (models containing a mixture of multiple PFTs). Models not belonging to the above PFTs were not given a PFT field and were not included in the subsequent analysis of the PFT differences. Other features (Table 2) are extracted directly from the corresponding descriptions in the papers in an explicit manner.~~

~~Subsequently, the model accuracies corresponding to different levels of various features are compared in a cross-study fashion. In the evaluation of algorithms and time scales, we also implement comparisons within individual studies. For example, in the evaluation of the effects of the algorithms, we compare the accuracy of models using the same training data and keeping other features as constants in individual studies. In this intra-study comparison step, only algorithms with relatively large sample sizes in the cross-study comparisons were selected. In this study, algorithms with less than 10 available model records are not considered to have a sufficient sample size and we do not give further conclusive opinions on the accuracy of these algorithms due to their small samples (e.g., PLSR and BART with high R-~~

squared but very few records as evidence). MLR, RF, SVM, and ANN were found to have large sample sizes (Fig. 5a), and thus their accuracies can be comparable. Based on this, in the intra-study comparison step, we only compare the accuracy differences between MLR, RF, SVM, and ANN in the context of using the same data and the same other model features (Fig. 5b).

Figure 2. Features of the machine learning-based NEE prediction process. The flux tower photo is from https://www.licor.com/env/support/Eddy-Covariance/videos/ec-method-02.html (last accessed: 23rd March 2022). The map in the lower part is from Harris et al., 2021. Prec, Ta, Rn, Ws, RH, and VPD represent precipitation, air temperature, net surface radiation, wind speed, relative humidity, and vapour-pressure deficit respectively. FAPAR is the fraction of absorbed photosynthetically active radiation. LST is the land surface temperature. LAI is the leaf area index.

2.3 Bayesian Network for analyzing joint effects

Based on the Bayesian network (BN), the joint impacts of multiple model features on the R-squared are analyzed. A BN can be represented by nodes (X_1,.., X_n) and the joint distribution (Pearl, 1985):

$$P(X) = P(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} P\left(X_i \middle| pa(X_i)\right) \quad (1)$$

where $pa(X_i)$ is the probability of the parent node X_i. Expectation-maximization (EM) approach (Moon, 1996) is used to incorporate the collected model records and compile the BN.

Sensitivity analysis is used for the evaluation of node influence based on mutual information (MI) which is calculated as the entropy reduction of the child node resulting from changes at the parent node (Shi et al., 2020):

$$MI = H(Q) - H(Q|F) = \sum_q \sum_f P(q, f) \log_2 \left(\frac{P(q,f)}{P(q)P(f)}\right) \quad (2)$$

where H represents the entropy, Q represents the target node, F represents the set of other nodes and q and f represent the status of Q and F.

带格式的: 正文, 行距: 1.5 倍行距

3 Results

3.1 Articles included in the meta-analysis

We included 40 articles (Table S2) and extracted 178 model records for the formal meta-analysis (Fig. 1). Most studies were implemented in Europe, North America, Oceania, and China (Fig. 3). The number of such papers is increasing recently (Fig. 4) and it shows the machine learning approach for NEE prediction has been of interest to more researchers. The main journals in which these articles have been published (Fig. 4) include Remote Sensing of Environment, Global Change Biology, Agricultural and Forest Meteorology, Biogeosciences, and Journal of Geophysical Research: Biogeosciences, etc.

Figure 3. Location of studies (a) included with the number of flux sites included and (b) their PFTs in the meta-analysis (total of 40 studies and 178 model records). Global (mainly based on FluxNet (Tramontana et al., 2016)) and continental-scale studies are not shown on the map due to the difficulty of identifying specific locations.

Figure 4. The number of studies published across journals and the total number of publications per year.

A total of 32 articles (Table S1) containing a total of 139 model records were included. The geographical scope of these articles was mainly Europe, North America, and China (Fig. 2).

3.2 The formal Meta-analysis

We assessed the impact of the features (e.g., algorithms, study area, PFTs, amount of data, validation methods, predictor variables, etc.) used in the different models based on differences in R-squared.

3.2.1 Algorithms

Among the more frequently used algorithms, ANN SVM and SVM performed betterRF outperformed (Fig. 5a) on average3a) across studies (lightly better than RF). On the other hand, since cross-study comparisons of algorithm accuracy include differences in data used in

设置了格式: 字体: 10 磅, 加粗
带格式的: 标题 2
设置了格式: 字体: 10 磅, 加粗
带格式的: 标题 2
设置了格式: 字体: 10 磅
带格式的: 正文, 行距: 1.5 倍行距
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅

~~model construction, we performed a pairwise comparison (Fig. 5b) of these four~~algorithms with sufficient sample size in Fig. 3a such as ANN). These three machine learning algorithms (i.e., ANN, SVM, RF~~, and MLR). In these~~) were significantly more accurate than the traditional MLR. Other algorithms such as MTE, ELM, Cubist, etc. also correspond to high accuracy, but with limited evidence sample size (Fig. 3a). In the internal comparison (different algorithms applied to the same data set) in single studies, ~~multiple models are developed for consistent training data with the interference of training data differences removed. It shows~~we also find that ~~RF and~~ SVM ~~perform best in the inter-study comparison (Fig. 5b). Whereas ANN performed~~and RF were slightly ~~worse~~more accurate than ~~RF~~ANN (Fig. 3b), and all these three (i.e., ANN, SVM, ~~all three of them were stronger~~RF) are considerably more accurate than MLR. Overall, ~~the performance of RF and~~ SVM ~~may be good~~ and ~~similar~~RF have shown higher accuracy in ~~the NEE~~water flux simulations in both inter and intra-study comparisons with sufficient sample size as evidence.

设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅

3.2.2 ~~Time scales~~Climate types and PFTs

We found higher average model accuracy in arid climate zones (Fig. 4a), such as the Cold semi-arid (steppe) climate (BSk) and Cold desert climate (BWk). Most of these studies were located in northwest China and the western USA. It may be caused by the simpler relationship between water fluxes and biophysical covariates in arid regions. In arid zones, due to the high potential ET, the variability in the actual ET may be largely explained by water availability (moisture supply) and vegetation change with the effect of variability in thermal conditions reduced. As for the various PFTs, the average model accuracy was slightly lower for forest types than for cropland and grassland types (Fig. 4b). The lowest average accuracy was found for shrub sites, which may be related to the difficulty of the remote sensing-based vegetation index (e.g., NDVI) to quantify the physiological and ecological conditions of shrubs (Zeng et al., 2022), and the heterogeneity of the spatial distribution of shrubs within the EC observation area may also cause difficulties in capturing their relationships with biophysical variables. We also found high model accuracy for the wetland type, although records as evidence to support this finding may be limited. Compared to other PFTs, the more steady and adequate water availability in the wetland type may make the variations of water fluxes less explained by other biophysical covariates.

3.3.3 Predictors and their combinations

On one hand, for the effects of individual predictors, the use of Rn/Rs, Prec, Ta, and FAPAR improved the accuracy of the model (Fig. S1). This pattern partially changed in the different PFTs. In the forest sites, the accuracy of the models with Rn/Rs and Ta used was higher than that of the models with Rn/Rs and Ta not used. For the grassland sites, the use of Ws, FAPAR, Prec, and Rn/Rs improved the model accuracy. For the cropland sites, Ta and FAPAR were more important for improving the model accuracy.

On the other hand, the evaluation of the effect of individual predictors on model accuracy is not necessarily reliable because some predictor variables are used together (e.g., the high model accuracy corresponding to a particular variable may be because it is often used together with another variable that plays the dominant role in improving accuracy). Therefore, we tested for independence between the use of variables and assessed the effect of the combination of variables on model accuracy. We calculated the correlation matrix (Fig. S2) between the use of various predictors (not used is set as 0 and used is set as 1). We found there was a dependence between the use of some predictors, the use of NDVI/EVI, LAI, and SM was significantly negatively correlated with the use of Rn/Rs and Ta (Fig. S2). ~~The impact of time scale on R-squared is considerable (Fig. 6), with models with larger time scales having lower average R-squared, especially when the time scale exceeds the monthly scale. The most frequently used scales were the daily, 8-day, and monthly scales. In studies where multiple time scales were used with other characteristics being the same, we found that models with half-hourly scales were significantly more accurate than models with daily scales (Fig. 6). However, the difference in accuracy between the day-scale and week-scale models is small. The accuracy of models with a monthly scale is the lowest.~~

~~Figure 6. Differences in model accuracy (R-squared) at different time scales across studies with the linear regression between R-squared and time scales (a), and comparison of the model accuracy (R-squared) of selected pairs of time scales within individual studies (b). All model records were included in panel (a), while studies that used multiple time scales (with other model characteristics unchanged) were included in panel (b). Time scales: 0.02 days (half-hourly), 0.04 days (hourly), 30 days (monthly), and 90 days (quarterly).~~

~~**3.2.3 Various predictors**~~It indicated that many of the models that used Rn/Rs and Ta did not use

NDVI/EVI, LAI, and SM, and the models that used NDVI/EVI, LAI, and SM also happened to not use Rn/Rs and Ta. Given this dependence, we evaluated the effect of the combination of variables on the model accuracy (Fig. 5). In Fig. 5, the three variable combinations on the left side are mainly meteorological variables while the three variable combinations on the right side are mainly vegetation-related variables based on remote sensing (e.g., NDVI, EVI, LAI, LSWI). We found that, overall, the accuracy of the models using only meteorological variable combinations was higher than that of the models using only remote sensing-based vegetation-related variables. It demonstrated the importance of using meteorological variables in machine learning-based ET prediction (probably especially for models with small time scales such as hourly scale, and daily scale). For example, in the forest type, the combination of Ta and Rn/Rs is very effective compared to using only remote sensing-based vegetation index variable combinations. The combination of Ta and Rn/Rs is also effective in the grassland and cropland types. The combination of Ws and Rn/Rs played an important role in the grassland type for improving model accuracy. Despite this, it does not negate the positive role of remote sensing-based vegetation-related variables in ET prediction. This effectiveness can be dependent on the time scale of the model as well as the PFTs. In models with large time scales (monthly scale, seasonal scale) and PFTs in which ET is sensitive to vegetation dynamics, remote sensing-based vegetation-related variables may also be of high importance.

3.3.4 Other model features

We also evaluated the impact of some other features on accuracy. The differences in accuracy of models with different spatial scales, year spans, number of sites, and volume of data (Fig. 6) appear to be insignificant. This seems to be related to the fact that in large-scale water flux simulations, the sites of similar PFTs are selected such as for modeling multiple forest sites across Europe (Van Wijk and Bouten, 1999) which focus on 'forest' and multiple grassland sites across arid northern China (Xie et al., 2021; Zhang et al., 2021) which focus on 'grassland', rather than mixing different PFT types to train models as the way in machine learning modeling of carbon fluxes (Zeng et al., 2020). In terms of the time scales of the models, the 4-day, 8-day, and monthly scales appear to correspond to higher accuracy compared to the half-hourly and daily scales. The higher the ratio of the volume of data in the training and validation sets, the higher the model accuracy. Compared

to the models using Landsat data, the models using MODIS data showed slightly higher accuracy probably due to the advantage of MODIS data in capturing the temporal dynamics of biophysical covariates. There were significant differences in the accuracy of the models using different cross-validation methods, with the models using random cross-validation showing higher accuracy than those using temporal cross-validation. This suggests that interannual variability may have a high impact on the models in water flux simulations. The driving mechanism of ET may vary significantly across years, and the inclusion of some extreme climatic conditions in the training set may be important for model accuracy and robustness.

3.3.5 Linear correlation of quantitative features and R-squared

We also analyzed the linear correlation (Fig. 7) between multiple quantitative features and the R-squared. We found that the magnitude of the linear correlation coefficients between the use of predictor combinations and the R-squared was higher than other features. The use of the predictor combination 'Ta and Rn/Rs' significantly improved the model accuracy. 'Temporal scale', 'time span', 'training/validation ratio', and 'number of sites' showed weak positive correlations with R-squared (not significant, p-value > 0.1). The positive correlation between 'temporal scale' and R-squared is higher among these features, although not significant. It should also be paid more attention to in future studies. The feature 'training/validation ratio' and 'time span' are also positively correlated (although not significantly) with the R-squared, suggesting the importance of the volume of data in the training set in a data-driven machine learning model. Larger 'training/validation ratio' and 'time span' may correspond to greater proportional coverage of the scenarios/conditions in the training set over the validation set, and thus correspond to higher accuracy.

带格式的: 正文, 行距: 1.5 倍行距

~~Among the commonly used predictors for NEE, there are significant differences in the predictors used and their impacts on model accuracy for different PFTs (Fig. 7). Ancillary data (e.g. soil texture, soil organic content, topography) that do not have temporal variability are used less frequently because they can only explain spatial heterogeneity. In contrast, the biophysical variables LAI, FAPAR, and ET were used significantly less frequently than NDVI/EVI, especially in the cropland and wetland types. The meteorological variables Ta, Rn/Rs, and VPD were used most frequently. For forest sites, Rn/Rs and Ws~~

appear to be the variables that improve model accuracy. For grassland sites, we found that NDVI/EVI appears to be the most effective, despite the small sample size. For sites in croplands and wetlands, we did not find predictor variables that had a significant impact on model accuracy.

For different PFTs, the top three variables in the ranking of model importance differed (Fig. S1). SM, Rn/Rs, Ta, Ts, and VPD all showed high importance across PFTs. This suggests that the variability of measured site-scale moisture and temperature conditions is important for the simulation of NEE for all PFTs. In contrast, in the importance ranking, other variables such as precipitation and NDVI/EVI may not lead because of the lag in their effect on NEE (Hao et al., 2010; Cranko Page et al., 2022). And some other variables may improve model accuracy for specific PFTs such as groundwater table depth (GWT) for wetland sites and growing degree days (GDD) for tundra sites.

Figure 7. The impact of the various predictors incorporated in models of different PFTs (1-forest, 2-grassland, 3-cropland, 4-wetland, 6-tundra) on R-squared. Dark blue boxes indicate that the predictor was used in the model, while dark red boxes indicate that the predictor was not used. Predictors: soil organic content (Soil_OC), precipitation (Prec), soil moisture/land surface water index (SM_LSWI), net radiation/solar radiation (Rn_Rs), enhanced vegetation index (EVI), air temperature (Ta), vapor pressure deficit (VPD), the fraction of absorbed photosynthetically active radiation/photosynthetically active radiation (FAPAR_PAR), relative humidity (RH), evapotranspiration (ET), leaf area index (LAI).

3.2.4 Other features

In addition, we evaluated other features of the model construction that may contribute to differences in model accuracy (Fig. 8). Studies at continental and global scales with a large number of sites and a large span of years correspond to lower R-squared than studies at local and regional scales, suggesting that studies with a large number of sites across large regions are likely to have high variability in the relationship between NEE and covariates and that studies at small scales are more likely to have higher model accuracy. Spatial validation (usually 'leave-one-site-out') corresponds to lower model accuracy compared to random and temporal validation. This again confirms the dominant role of heterogeneity in the relationship between NEE and covariates across sites in explaining model accuracy. This seems to

be indirectly supported by the fact that a high ratio of training to validation sets corresponds to a low R-squared, as this high ratio tends to be accompanied by the use of the 'leave-one-site-out' validation approach. The accuracy of the models with a growing season period was slightly higher than that of the models with an annual period. For the satellite remote sensing data used, the models based on MODIS data with biophysical variables extracted were slightly less accurate than those based on Landsat data. For the daily scale models, Landsat data performed a little better than MODIS (Fig. S2). This suggests that the higher temporal resolution of MODIS compared to Landsat may not play a dominant role in improving model accuracy. This may also be partially attributed to studies using MODIS based explanatory data that tend to include too large surrounding areas around the site (e.g., 2x2 km), which can lead to a scale mismatch between the flux footprint and the explanatory variables.

Figure 8. The impacts of other features (i.e. spatial scale, study period, number of sites, year span, site year, cross-validation method, training/validation, and satellite imagery) on the model performance.

3.3. The joint causal impacts of multi-features based on the BN

We selected the features that had a more significant impact on model accuracy in the above assessment and further incorporated them into the BN-based multivariate assessment to understand the joint impact of multiple features on R-squared. The features incorporated included the spatial scale, the number of sites, the time scale, the span of years, the cross-validation method, and whether some specific predictors were used. We discretized the distribution of individual nodes and compiled the BN (Fig. 9a) using records from different PFTs as input. Sensitivity analysis of the R-squared node (Fig. 10) showed that R-squared was most sensitive to 'year span', cross-validation method, Rn/Rs, and time scale under multi-feature control. In the forest and cropland types, R-squared is more sensitive to Rn/Rs, while in the wetland type it is more sensitive to SM/LSWI and Ta. The sensitivity of R-squared to 'year span' was much higher in the cropland type compared to the other PFTs, which may suggest that the interannual variability in the NEE simulations of the cropland type is higher due to potential interannual variability of the planting structure and irrigation practices. For the cropland type, differences in the phenology, harvesting, and irrigation (water volume and frequency) in different years can lead to significant inter-annual differences in NEE simulations. Subsequently, using the constructed BN (with the empirical information in previous studies incorporated), for new studies we can instructively infer the probability

~~distribution of the possible R-squared (Fig. 9b) with some model features predetermined. In previous studies, spatio-temporal mapping of NEE based on statistical models has often lacked accuracy assessment since there are no grid-scale NEE observations, and this BN may have the potential to be used to validate the accuracy (R-squared) of the NEE time series output of the grid scale (i.e. inferring possible R-squared from model features, where the output of the grid-scale is considered to be of the form 'leave one site out').~~

~~Figure 9. The joint effects of multiple features on the R-squared based on the BN with all records input (a) and the inference on the probability distribution of R-squared based on the BN with the status of some nodes determined (b). The values before and after the "±" indicate the mean and standard deviation of the distribution, respectively. The gray boxes indicate that the status of the nodes has been determined. In panel (b), specific values of parent nodes such as 'spatial scale' are determined (shown in the red box), leading to an increase in the expected R-squared compared to the average scenario of panel (a) (as inferred from the posterior conditional probabilities with the status of the node 'spatial scale' are determined as 'local').~~

~~Figure 10. The sensitivity analysis of the R-squared node to other nodes based on the mutual information (MI) across PFTs. 'Cross-validation' is the cross-validation method including spatial, temporal, and random cross-validation.~~

4 Discussions

带格式的: 行距: 1.5 倍行距

With the accumulation of in situ EC observations around the world, the study of ET simulations based on data-driven approaches has received more attention from researchers in the last decade. Many studies have combined EC observations, various predictors, and machine learning algorithms to improve the prediction accuracy of water fluxes. To date, the results of these studies have not been comprehensively evaluated to provide clear guidance for feature selection in water flux prediction models. To better understand the approach and guide future research, we performed a meta-analysis of such studies. Machine learning-based water flux simulations and

predictions still suffer from high uncertainty. By investigating the expected improvements that can be achieved by incorporating different features, we can avoid practices that may reduce model accuracy in future research.

4.1 Opportunities and challenges in the water flux simulation

In the above meta-analysis of the models, we found that water flux simulations based on EC observations can achieve high accuracy but also have high uncertainty through the modeling workflow. The R-squared of many water flux simulation models exceeds 0.8, possibly higher than some remote sensing-based and process-based models, and possibly higher than carbon flux simulations such as the net ecosystem exchange (NEE) in a similar modeling framework (Shi et al., 2022). This may be because many data on important variables affecting carbon flux such as soil and biomass pools, disturbances, ecosystem age, management activities, and land use history are not yet effectively and continuously measured (Jung et al., 2011) with the global spatially and temporally explicit information. While ET simulations rely on observations of moisture and energy conditions and vegetation conditions, much of the current available meteorological and remote sensing data have been effective to represent and capture the spatial and temporal dynamics of these predictors well.

4.1.1 Comprehensive insights on model features

Biophysical and meteorological variables are considered both important in ET simulations. This study found that models using a combination of meteorological variables had higher accuracy than models using only remotely sensed vegetation dynamic information. However, due to the high proportion of models with small temporal scales (e.g., half-hourly scale, hourly scale, and daily scale) in this study, this advantage of the combination of meteorological variables may be more suitable for small temporal scales. A possible explanation is that vegetation-related variables such as NDVI and LAI at the daily scale, 8-day scale, and 16-day scale have limited explanatory ability for hourly or daily-scale variability in ET, especially under cloudy conditions (e.g., tropical rainforest regions), the temporal continuity of the vegetation index data may be greatly limited (Zeng et al., 2022). This should be given more attention and some vegetation indices derived from hourly temporal resolution satellite remote sensing data such as GOES (Zeng et al., 2022) can be

used for ET simulations to investigate the possible adding-values of vegetation indices at smaller time scales. In contrast, at a small temporal scale, the use of combinations of meteorological variables can capture moisture and energy conditions that control the rapid fluctuations of ET and thus has a dominant role in hourly or daily-scale ET prediction. This also corroborates the high accuracy of some physic-based ET estimation models (Rigden and Salvucci, 2015) that use only meteorological variables and not vegetation-related variables such NDVI (only an estimate of vegetation height derived from land cover maps is used to represent vegetation conditions (Rigden and Salvucci, 2015)).

There are differences in model accuracy among different PFTs. For example, in forest sites, limitations in data accuracy of factors were possible because some remote sensing-based predictors such as NDVI, FAPAR, and LAI have limited accuracy when applied to forest types (Liu et al., 2018b; Zeng et al., 2022). In addition, factors such as crown density, which may significantly affect the proportion of soil evaporation, transpiration, and evaporation of canopy interception, were not considered in these models, which may also lead to low model accuracy. This suggests that in water flux simulation, the driving mechanisms of water fluxes in different PFTs do affect the accuracy of machine learning models, and we need to consider more the actual and specific influencing factors in specific PFTs. More variables that can quantify the ratio of evaporation and transpiration should be considered for inclusion, which also appears to improve the mechanistic interpretability of such machine learning models. A previous study (Zhao et al., 2019) combined the physics-based approach (e.g., Penman-Monteith equation) and machine learning to build hybrid models to improve interpretability. We should make full use of empirical knowledge and experiences from process-based models to improve the accuracy and interpretability of the machine learning approach.

Among the ~~Many studies have evaluated the incorporation of various predictors and model features using machine learning for improving the site-scale NEE predictions (Tramontana et al., 2016; Zeng et al., 2020; Jung et al., 2011). A comprehensive evaluation of these studies to provide definitive guidance on the selection of features in NEE prediction modeling is limited. This study fills the research gap with a meta-analysis of the literature through statistics on the accuracy and performance of models. Machine~~

设置了格式: 字体: 非加粗

带格式的: 正文, 行距: 1.5 倍行距

learning-based NEE simulations and predictions still suffer from high uncertainty. By better understanding the expected improvements that can be achieved through the inclusion of different features, we can identify priorities for the consideration of different features in modeling efforts and avoid operations decreasing model accuracy.

Compared to previous comparisons of machine learning-based NEE prediction models, this study is more comprehensive. Previous studies (Abbasian et al., 2022) have also found advantages of RF over other algorithms in NEE prediction. This study consolidated this finding using a larger amount of evidence. Previous studies (Tramontana et al., 2016) have also compared the impact of different practices in NEE prediction models based on the R-squared, such as comparing the difference in accuracy between the two predictor combinations (i.e., using only remotely sensed data and using remotely sensed data and meteorological data together). In contrast, since this study incorporated more detailed factors influencing model accuracy, the understanding of such issues was deepened. However, there are still many uncertainties and challenges in NEE prediction not clarified in this study.

4.1 Challenges in the site-scale NEE simulation and implications for other carbon flux simulations

4.1.1 Variations in time scales

In the above analysis, we found that the effect of the time scale of the model is considerable. This suggests that we should be careful in determining the time scale of the model to consider whether the predictor variables used will work at this time scale. Previous studies have reported the dependence of the NEE variability and mechanism on the time scales. On the one hand, the importance of variables affecting NEE varies at different time scales. For example, in tropical and subtropical forests in southern China (Yan et al., 2013), seasonal NEE variability is predominantly controlled by soil temperature and moisture, while interannual NEE variability is controlled by the annual precipitation variation. A study (Jung et al., 2017) showed that for annual-scale NEE variability, water availability and temperature were the dominant drivers at the local and global scales, respectively. This indicates the need to recognize the temporal and spatial driving mechanisms of NEE in advance in the development of NEE prediction models. On the other hand, dependence may exist between NEE anomalies at various time scales. For example, previous studies (Luyssaert et al., 2007) showed that short-term temperature anomalies may interpret both the daily and seasonal NEE anomalies. This implies that the models at different time scales

may not be independent. In the previous studies, the relationship between prediction models at different scales has not been well investigated, and it may be valuable to compare the relations between data and models at different scales in depth. Larger time scales correspond to lower model accuracy, possibly related to the fact that some small time scale relations between NEE and covariates (especially meteorological variables) are smoothed. In particular, for models with time scales smaller than one day (e.g. half-hourly models), the 8-daily and 16-daily biophysical variable data obtained from satellite remote sensing are difficult to explain the temporal variation in the sub-daily NEE. Therefore, for models at small time scales (i.e. half-hourly, hourly, daily scale models), in situ meteorological variables may be more important. The inclusion of some ancillary variables (e.g. soil texture, topographic variables) with no temporal dynamic information may be ineffective unless many sites are included in the model and the spatial variability of the ancillary variables for these sites is sufficiently large (Virkkala et al., 2021).

In terms of completeness and purity of training data, hourly and daily models can be better compared to monthly and yearly models. Hourly and daily models can usually preclude those low-quality data and gaps in the flux observations. However, for monthly and yearly scale models, gap-filling (Ruppert et al., 2006; Moffat et al., 2007; Zhu et al., 2022) is necessary because there are few complete and continuous fluxes observations without data gaps on the monthly to yearly scales. Since various gap-filling techniques rely on environmental factors (Moffat et al., 2007) such as meteorological observations, this may introduce uncertainty in the predictive models (i.e., a small fraction of the observed information of NEE is estimated from a combination of independent variables). How it would affect the accuracy of prediction models at various time scales remains uncertain, although various gap-filling techniques have been widely used in the pre-processing of training data.

In addition, the impacts of lagged effects (Hao et al., 2010; Cranko Page et al., 2022) of covariates are not considered in most models, which may underestimate the degree of explanation of NEE for some predictor variables (e.g. precipitation). Most of the machine learning-based models use only the average Ta and do not take into account the maximum temperature, minimum temperature, daily difference in temperature, etc., as in the process-based ecological models (Mitchell et al., 2009). This suggests that the inclusion of different temporal characteristics of individual variables in machine learning-based NEE prediction models may be insufficient.

4.1.2 Scale mismatch of explanatory predictors and flux footprints

An excessively large extraction area of remote sensing data (e.g., 2x2 km) may be inappropriate. In the non-homogeneous underlying conditions, the agreement of the area of flux footprints with the scale of the predictors should be considered in the extraction of the predictor variables in various PFTs (Chu et al., 2021).

The effects of this mismatch between explanatory variables and flux footprints may be diverse for different PFTs. For example, for cropland types, the NEE is monitored at a range of several hundred meters around the flux towers, but remote sensing variables such as FAPAR, NDVI, LAI, etc. can be extracted at coarse scales (e.g., 2x2 km), some effects outside the extent of the flux footprint (Chu et al., 2021; Walther et al., 2021) are incorporated (e.g., planting structures with high spatial heterogeneity, agricultural practices such as irrigation). And for more homogeneous types such as grasslands, coarse-scale meteorological data may still cause spatial mismatches, even though the differences in land cover types within the 2x2 km and 200x200 m extent around the flux stations in grasslands may not be considerable. For example, precipitation with high spatial heterogeneity can dominate the spatial variability of soil moisture and thus affect the spatial variability of grassland NEE (Wu et al., 2011; Jongen et al., 2011). However, using 0.25°x0.25° reanalysis precipitation data (Zeng et al., 2020) may make it difficult for predictive models to capture this spatial heterogeneity around the flux station.

Since few of the studies included in this meta-analysis considered the effect of variation in flux footprint, this feature was difficult to consider in this study. However, its influence should still be further investigated in future studies. With flux footprints calculated (Kljun et al., 2015) and the factors around the flux site (Walther et al., 2021) that affect the flux footprint incorporated, .it is promising to clarify this issue.

4.1.3 Possible unbalance of training and validation sets

In addition to the time scale of the models, the most significant differences in model accuracy and performance were found in the heterogeneity within the NEE dataset and the match of the training set and validation set. Often NEE simulations can achieve high accuracy in local studies, where the main factor negatively affecting model accuracy may be the interannual variability in the relationship between

~~NEE and covariates. However, the complexity may increase when the dataset contains a large study area, many sites, PFTs, and year spans. Under this condition, the accuracy of the model in the 'leave one site out' validation may be more dependent on the correlation and match between the training and validation sets (Jung et al., 2020). When the model is applied to an outlier site (of which the NEE, covariates, and their relationship are very different compared with the remaining sites), it appears to be difficult to achieve a high prediction accuracy (Jung et al., 2020). If we further upscale the prediction model to large spatial and time scales, the uncertainties involved may be difficult to assess (Zeng et al., 2020). We can only infer the possible model accuracy based on the similarity of the distribution of predictors in the predicted grid to that of the existing sites in the model. In the upscaling process, reanalysis data with coarse spatial resolution are often used as an alternative for site-scale meteorological predictors. However, most studies did not assess in detail the possible errors associated with spatial mismatches in this operation.~~

~~In summary, the site-scale NEE predictions may require more focus on the internal heterogeneity of the NEE dataset and the matching of the training set and validation set, and also require a better understanding of the influence of different scales of the same variable (e.g. site-scale precipitation and grid-scale precipitation in the reanalysis meteorological data) across modeling and upscaling steps. For the prediction of other carbon fluxes such as methane fluxes (in the same framework as the NEE predictions), the results of this study may also be partially applicable, although there may be significant differences in the use of specific predictors (Peltola et al., 2019).~~

~~4.2 Uncertainties~~

~~The uncertainties in this analysis may include:~~

validation methods, random cross-validation has higher accuracy than spatial cross-validation and temporal cross-validation. However, spatial cross-validation and temporal cross-validation may be able to better help us recognize the robustness of the model when extrapolated (i.e., applied to new stations and new years). The lower accuracy in the temporal cross-validation approach implies that we need to focus on interannual hydrological and meteorological variability in the water flux simulations. In cropland sites, we may also need to pay more attention to the effects of interannual variability in anthropogenic cropping patterns. If some extreme weather years are not

included, the robustness of the model when extrapolated to other years may be challenged, especially in the context of the various extreme weather events of recent years. This can also inform the siting of future flux stations. Regions where climate extremes may occur and biogeographic types not covered by existing flux observation networks should be given more attention to achieve global-scale, accurate and robust machine learning-based spatio-temporal prediction of water fluxes. Furthermore, although the R-squared and the training/validation ratio show a positive correlation (Fig. 7) (i.e., a higher training/validation ratio may correspond to a higher R-squared), we should still be cautious in reducing this ratio in our modeling. For a really small validation set, it would be very challenging to determine which model is better given the potential uncertainty caused by the considerable randomness.

4.1.2 Differences from NEE predictions in the similar model framework

In general, predictors related to meteorological, vegetation, and soil conditions were common to both ET and NEE simulations in a similar framework (Shi et al., 2022). However, in NEE predictions, explanatory variables such as soil organic content, photosynthetic photon flux density, and growing degree days (Shi et al., 2022) are not necessary for ET predictions. The selection of these variables requires our prior knowledge of the dominant drivers of ET and NEE anomalies of particular ecosystems and their differences.

The accuracy of NEE predictions (Shi et al., 2022) can be more limited by global variability across biomes and locations (Nemani et al., 2003) given the lack of locally measured data on soil and biomass pools, disturbances, ecosystem age, management activities and land use history (Jung et al., 2011). It can result in a higher heterogeneity of the training data in large-scale modeling with multiple flux sites (Shi et al., 2022) and the weak ability to capture the NEE anomalies. In contrast, in ET predictions, meteorological variables and vegetation conditions appear to be already sufficient to capture a considerably large fraction of the ET variations in most conditions.

In future ET prediction studies, given that few current ET products have time scales smaller than daily scale (Jung et al., 2019; Pan et al., 2020), improvements in the accuracy of daily and hourly models may be necessary to fill this gap. Besides, the partitioning of ET components (i.e.,

transpiration, interception evaporation, and soil evaporation) can be more focused to better decouple the contributions of vegetation and soil to ET with machine learning (Eichelmann et al., 2022). It can be further matched with the partitioning of NEE (i.e., to GPP and ecosystem respiration) to increase our knowledge of the global water cycle and ecosystem functioning and obtain further refined global carbon-water fluxes coupling relations (Eichelmann et al., 2022). Also, the above two promising improvements can be beneficial for research on topics related to the global terrestrial water cycle (Fisher et al., 2017).

4.2 Uncertainties and limitations of this meta-analysis

4.2.1 The limited number of available literature and model records

Despite many articles and model records collected through our efforts to perform this meta-analysis, there still appears to be a long way to go to finally and completely understand the various mechanisms involved in water flux simulation with machine learning. Some of the insights provided by this study can be not robust (due to the limited sample size available when the goal is to assess the effects of multiple features), but this does not negate the fact that this study does obtain some meaningful findings. Therefore, researchers should treat the results of this study with caution, as they were obtained only statistically. Overall, it is still positive to conduct a meta-analysis of such studies, considering their rapid growth in number and lack of guiding directions.

4.2.2 Publication bias and weighting:

a) Publication bias is not refined dueand weighting: Due to the limitations of the relatively limited number of articles that cancould be included in the meta-analysis, this study did not focus much on publication bias. Meta-analyses often analytic studies in other fields typically measure the quality of journals and the datapublic availability of research data (Borenstein et al., 2011; Field and Gillett, 2010) to determine the weighting of the literature in a comprehensive assessment. However, a high proportionmost of the articles in this study did not makepublicly provide flux observations publicly available or share the NEE prediction models developed. Furthermore, meta models. Meta-analysis studies in other fields typically measure the impact of papers by evidence/data volume;included studies based on sample size and the variance of the evaluated

设置了格式: 字体: 10 磅, 加粗
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
带格式的: 正文, 行距: 1.5 倍行距, 无项目符号或编号
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅
设置了格式: 字体: 10 磅

effectsexperimental results (Adams et al., 1997; Don et al., 2011; Liu et al., 2018). However, in this study, because no convincing method is found to quantify the weights of results from included articles, some features (e.g. the number of flux sites, the span of years) were directly assessed rather than used to determine the weights of the articles(Adams et al., 1997; Don et al., 2011; Liu et al., 2018a). In this study, due to the lack of a convincing manner to determine weights among articles, we assigned the same weight to the results for all the literature.

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

b) Limitations of the criteria for inclusion in the literature: in the model accuracy-based evaluation, we selected only literature that developed multiple regression models. Potentially valuable information from univariate regression models was not included. In addition, only papers in high-quality English journals were included in this study to control for possible errors due to publication bias. However, many studies that fit this theme may have been published in other languages or other journals.

c) Independence between features: There is dependence between the evaluated features (e.g. the dependency between the spatial extent and the number of sites). It may negatively affect the assessment of the impact of individual features on the accuracy of the model, although the BN-based analysis of joint effects can reduce the impact of this dependence between variables by specifying causal relationships between features. The interference of unknown dependencies between features may still not be eliminated when we focus on the effects of an individual feature on the model performance. We should pay more attention to the effect of features on model accuracy individually in future studies, and it may be valuable to keep other features as constants while changing the level of only one feature and assessing the difference. It may help us to understand the real sensitivity of model accuracy to different features in specific conditions. The sample size collected in this study (178 records in total) is not very large. This also suggests that more future efforts should be devoted to the comprehensive evaluation and summarization of NEE simulations.

Additionally, there are still other potential factors not considered by this study such as the uncertainty of climate data (site-vs-reanalysis), footprint matching between site and satellite images, etc. Overall, although the quantitative results of this study should be used with caution, they still have positive implications for guiding future such studies.

4.2.3 Uncertainties in the information of the extracted features

At the information extraction level, the following issues may also introduce uncertainties:

a) Uncertainties caused by data quality control (e.g. gap-filling (Hui et al., 2004)) are difficult to assess effectively. Gap-filling is a commonly used technique to fill in low-quality data in flux observations. However, the impact of this practice on machine learning-based ET prediction models is unclear, due to the difficulty of directly assessing how this technique is performed in various studies by this meta-analysis. Typically, models with small time scales (e.g., hourly scale, daily scale) can exclude low-quality observations and use only high-quality data. However, for models with large time scales (e.g., monthly scales), gap-filling (e.g., based on meteorological data) may be unavoidable. This may lead to a decrease in training data purity and introduce uncertainty in the subsequent prediction model development.

b) Systematic uncertainties caused by the energy balance closure (EBC) issue in eddy-covariance flux measurements are also difficult to assess by this meta-analysis. EBC is a common problem (Eshonkulov et al., 2019) in eddy-covariance flux observations. For that reason, the latent heat flux measured potentially underestimates ET. Some prediction models corrected EBC (e.g., using Bowen ratio preserving (Mauder et al., 2013, 2018) and energy balance residuals (Charuchittipan et al., 2014; Mauder et al., 2018)) in the processing of training data, but some did not. How this will affect the accuracy of the prediction model is not clear due to multiple factors that need to be evaluated that influence EBC (Foken, 2008), including measurement errors of the energy balance components, incorrect sensor configurations, influences of heterogeneous canopy height, unconsidered energy storage terms in the soil-plant-atmosphere system, inadequate time averaging intervals, and long-wave eddies (Jacobs et al., 2008; Foken, 2008; Eshonkulov et al., 2019). To reduce this uncertainty, more attention to flux site characteristics (Eshonkulov et al., 2019) related to PFT, topography, flux footprint area, etc., to select the appropriate correction method is necessary for future studies.

c) As most studies used far more water flux observation records than the number of covariates in their regression models, we did not adjust the R-squared in this study to an adjusted R-squared.

d) The various specific ways in which the parameters of the model are optimized are not differentiated. They are broadly categorized into different families or kinds of algorithms, which may also introduce uncertainty into the assessment.

e) The assessment of some features is not detailed due to the limitations of the available model records. For example, the classification of PFT could be more detailed. 'Forest' could be further classified as broadleaf forest, coniferous forest, etc. while 'cropland' could be further classified as rainfed and irrigated cropland based on differences in their response mechanisms of water fluxes to environmental factors.

5 Conclusion

We performed a meta-analysis of the ~~site-scale NEE~~ water flux simulations combining in situ flux observations from flux stations/networks, meteorological, biophysical, and ancillary predictors, and machine learning. ~~The impacts of various features throughout the modeling process on the accuracy of the model were evaluated.~~ The main ~~findings of this study include:~~conclusions are as follows:

~~1. RF~~ SVM (average R-squared = 0.82) and ~~SVM performed better than other evaluated algorithms.~~

~~2. The impact of time scale on model performance is significant. Models with larger time scales have lower average R-squared, especially when the time scale exceeds the monthly scale. Models with half-hourly scales~~ RF (average R-squared = 0.~~73) were significantly more accurate than models with daily scales (average R-squared = 0.5).~~

~~3.~~1. ~~Among the commonly used predictors for NEE, there are significant differences~~81) outperformed over evaluated algorithms with sufficient sample size in ~~the predictors used~~both cross-study and ~~their impacts on model accuracy for different PFTs.~~ intra-study (with the same training dataset) comparisons.

2. ~~It~~The average accuracy of the model applied to arid regions is ~~necessary to focus on the potential imbalance between the training and validation sets in NEE simulations. Studies at continental and global scales~~higher than in other climate types.

3. The average accuracy of the model was slightly lower for forest sites (average R-squared = 0.~~37) with multiple PFTs, more~~76) than for cropland and grassland sites~~, and a large span of years correspond to lower~~ (average R-squared ~~than studies at local~~= 0.8 and 0.79), but higher than for shrub sites (average R-squared = 0.~~69)~~67).

带格式的: 行距: 1.5 倍行距

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

带格式的: 列表段落, 行距: 1.5 倍行距, 多级符号 + 级别: 1 + 编号样式: 1, 2, 3, … + 起始编号: 1 + 对齐方式: 左侧 + 对齐位置: 0 厘米 + 缩进位置: 0.63 厘米

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

4. Among various predictor variables, the use of Rn/Rs, Prec, Ta, and FAPAR improved the model accuracy. The combination of Ta and ~~regional scales (average R-squared = 0.7).~~ Rn/Rs is very effective especially in the forest type, while in the grassland type the combination of Ws and Rn/Rs is also effective.

带格式的: 列表段落, 行距: 1.5 倍行距, 多级符号 + 级别: 1 + 编号样式: 1, 2, 3, … + 起始编号: 1 + 对齐方式: 左侧 + 对齐位置: 0 厘米 + 缩进位置: 0.63 厘米

设置了格式: 字体: 10 磅

~~**Acknowledgments**~~

5. Among the different validation methods, random cross-validation shows higher model accuracy than spatial cross-validation and temporal cross-validation.

Acknowledgements

We thank the ~~editors~~editor and ~~three~~two anonymous ~~referees~~reviewers for their insightful comments ~~on this paper~~ which contributed substantially ~~improved.~~to the improvement of this manuscript.

Financial support

This research was supported by the National Natural Science Foundation of China (Grant No. U1803243), the Key projects of the Natural Science Foundation of Xinjiang Autonomous Region (Grant No. 2022D01D01), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA20060302), and High-End Foreign Experts Project.

Author ~~contributions~~ Contributions

~~H.S~~HS and ~~G.L initiated this research and~~ GL were responsible for the ~~integrity of the work as a whole. H.S performed~~conceptualization, methodology, formal analysis, ~~and calculations and drafted the manuscript. H.S, G.L, X.M, X.Y, Y.W, W.Z, M.X, C.Z, and Y.Z were responsible for the data collection and analysis. G.L, P.D.M, T.V.D.V, O.H, and A.K~~investigation, visualization, and writing. OH contributed ~~resources and financial support.~~to the investigation. XM, XY, YW, WZ, MX, CZ and YZ processed the data. AK, TVDV and PDM provided supervision.

Competing interests

The authors declare that they have no conflict of interest.

Code availability

The codes that were used for all analyses are available from the first author (shihaiyang16@mails.ucas.ac.cn) upon request.

带格式的: 行距: 1.5 倍行距

设置了格式: 字体: 宋体

带格式的: 行距: 1.5 倍行距

带格式的: 行距: 1.5 倍行距

带格式的: 行距: 1.5 倍行距

Data availability

The data used in this study can be accessed by contacting the first author (shihaiyang16@mails.ucas.ac.cn) ~~based on a reasonable~~upon request.

~~**Code availability**~~

~~The code used in this study can be accessed by contacting the first author (shihaiyang16@mails.ucas.ac.cn) based on a reasonable request.~~

带格式的: 行距: 1.5 倍行距

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

设置了格式: 字体: 10 磅

References

~~Abbasian, H., Solgi, E., Mohsen Hosseini, S., and Hossein Kia, S.: Modeling terrestrial net ecosystem exchange using machine learning techniques based on flux tower measurements, Ecological Modelling, 466, 109901, https://doi.org/10.1016/j.ecolmodel.2022.109901, 2022.~~

Adams, D. C., Gurevitch, J., and Rosenberg, M. S.: Resampling tests for meta‐analysis of ecological data, Ecology, 78, 1277–1283, 1997.

~~Baldocchi, D. D.: Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future, 9, 479–492, https://doi.org/10.1046/j.1365-2486.2003.00629.x, 2003.~~

~~Berryman, E. M., Vanderhoof, M. K., Bradford, J. B., Hawbaker, T. J., Henne, P. D., Burns, S. P., Frank, J. M., Birdsey, R. A., and Ryan, M. G.: Estimating soil respiration in a subalpine landscape using point, terrain, climate, and greenness data, Journal of Geophysical Research: Biogeosciences, 123, 3231–3249, 2018.~~

Allen, R. G., Pereira, L. S., Howell, T. A., and Jensen, M. E.: Evapotranspiration information reporting: I. Factors governing measurement accuracy, Agricultural Water Management, 98, 899–920, https://doi.org/10.1016/j.agwat.2010.12.015, 2011.

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

Barman, R., Jain, A. K., and Liang, M.: Climate-driven uncertainties in modeling terrestrial energy and water fluxes: a site-level to global-scale analysis, Global Change Biology, 20, 1885–1900, https://doi.org/10.1111/gcb.12473, 2014.

Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R.: Introduction to meta-analysis, John Wiley & Sons, 2011.

~~Cho, S., Kang, M., Ichii, K., Kim, J., Lim, J. H., Chun, J. H., Park, C. W., Kim, H. S., Choi, S. W., and Lee, S. H.: Evaluation of forest carbon uptake in South Korea using the national flux tower network, remote sensing, and data-driven technology, Agricultural and Forest Meteorology, 311, 108653, 2021.~~

~~Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., Torn, M. S., Metzger, S., Kumar, J., Arain, M. A., Arkebauer, T. J., Baldocchi, D., Bernacchi, C., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Bracho, R., Brown, S., Brunsell, N. A., Chen, J., Chen, X., Clark, K., Desai, A. R., Duman, T., Durden, D., Fares, S., Forbrich, I., Gamon, J. A., Gough, C. M., Griffis, T., Helbig, M., Hollinger, D., Humphreys, E., Ikawa, H., Iwata, H., Ju, Y., Knowles, J. F., Knox, S. H., Kobayashi, H., Kolb, T., Law, B., Lee, X., Litvak, M., Liu, H., Munger, J. W., Noormets, A., Novick, K., Oberbauer, S. F., Oechel, W., Oikawa, P., Papuga, S. A., Pendall, E., Prajapati, P., Prueger, J., Quinton, W. L., Richardson, A. D., Russell, E. S., Scott, R. L., Starr, G., Staebler, R., Stoy, P. C., Stuart-Haëntjens, E.,~~

设置了格式: 字体: 10 磅

Sonnentag, O., Sullivan, R. C., Suyker, A., Ueyama, M., Vargas, R., Wood, J. D., and Zona, D.: Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites, Agricultural and Forest Meteorology, 301–302, 108350, https://doi.org/10.1016/j.agrformet.2021.108350, 2021.

Cleverly, J., Vote, C., Isaac, P., Ewenz, C., Harahap, M., Beringer, J., Campbell, D. I., Daly, E., Eamus, D., He, L., Hunt, J., Grace, P., Hutley, L. B., Laubach, J., McCaskill, M., Rowlings, D., Rutledge Jonker, S., Schipper, L. A., Schroder, I., Teodosio, B., Yu, Q., Ward, P. R., Walker, J. P., Webb, J. A., and Grover, S. P. P.: Carbon, water and energy fluxes in agricultural systems of Australia and New Zealand, 287, https://doi.org/10.1016/j.agrformet.2020.107934, 2020.

Cranko Page, J., De Kauwe, M. G., Abramowitz, G., Cleverly, J., Hinko-Najera, N., Hovenden, M. J., Liu, Y., Pitman, A. J., and Ogle, K.: Examining the role of environmental memory in the predictability of carbon and water fluxes across Australian ecosystems, Biogeosciences, 19, 1913–1932, 2022.

Cui, X., Goff, T., Cui, S., Menefee, D., Wu, Q., Rajan, N., Nair, S., Phillips, N., and Walker, F.: Predicting carbon and water vapor fluxes using machine learning and novel feature ranking algorithms, Science of The Total Environment, 775, 145130, 2021.

Charuchittipan, D., Babel, W., Mauder, M., Leps, J.-P., and Foken, T.: Extension of the Averaging Time in Eddy-Covariance Measurements and Its Effect on the Energy Balance Closure, Boundary-Layer Meteorol, 152, 303–327, https://doi.org/10.1007/s10546-014-9922-6, 2014.

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

Chen, Y., Wang, S., Ren, Z., Huang, J., Wang, X., Liu, S., Deng, H., and Lin, W.: Increased evapotranspiration from land cover changes intensified water crisis in an arid river basin in northwest China, Journal of Hydrology, 574, 383–397, https://doi.org/10.1016/j.jhydrol.2019.04.045, 2019.

Don, A., Schumacher, J., and Freibauer, A.: Impact of tropical land-use change on soil organic carbon stocks – a meta-analysis, Global Change Biology, 17, 1658–1670, https://doi.org/10.1111/j.1365-2486.2010.02336.x, 2011.

Eichelmann, E., Mantoani, M. C., Chamberlain, S. D., Hemes, K. S., Oikawa, P. Y., Szutu, D., Valach, A., Verfaillie, J., and Baldocchi, D. D.: A novel approach to partitioning evapotranspiration into evaporation and transpiration in flooded ecosystems, Global Change Biology, 28, 990–1007, https://doi.org/10.1111/gcb.15974, 2022.

Eshonkulov, R., Poyda, A., Ingwersen, J., Wizemann, H.-D., Weber, T. K. D., Kremer, P., Högy, P., Pulatov, A., and Streck, T.: Evaluating multi-year, multi-site data on the energy balance closure of eddy-covariance flux measurements at cropland sites in southwestern Germany, Biogeosciences, 16, 521–540, https://doi.org/10.5194/bg-16-521-2019, 2019.

Fang, B., Lei, H., Zhang, Y., Quan, Q., and Yang, D.: Spatio-temporal patterns of evapotranspiration based on upscaling eddy covariance measurements in the dryland of the North China Plain, Agricultural and Forest Meteorology, 281, https://doi.org/10.1016/j.agrformet.2019.107844, 2020.

Field, A. P. and Gillett, R.: How to do a meta‐analysis, British Journal of Mathematical and Statistical Psychology, 63, 665–694, 2010.

~~Fu, D., Chen, B., Zhang, H., Wang, J., Black, T. A., Amiro, B. D., Bohrer, G., Bolstad, P., Coulter, R., and Rahman, A. F.: Estimating landscape net ecosystem exchange at high-spatial-temporal resolution based on Landsat data, an improved upscaling model framework, and eddy covariance flux measurements, Remote Sensing of Environment, 141, 90–104, 2014.~~

~~Fu, Z., Stoy, P. C., Poulter~~Fisher, J. B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M. F., Hook, S., Baldocchi, D., Townsend, P. A., Kilic, A., Tu, K., Miralles, D. D., Perret, J., Lagouarde, J.-P., Waliser, D., Purdy, A. J., French, A., Schimel, D., Famiglietti, J. S., Stephens, G., and Wood, E. F.: The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources, Water Resources Research, 53, 2618–2626, https://doi.org/10.1002/2016WR020175, 2017.

Foken, T.: The energy balance closure problem: An overview, Ecological Applications, 18, 1351–1367, 2008.

Gaston, K. J.: Global patterns in biodiversity, Nature, 405, 220–227, https://doi.org/10.1038/35012228, 2000.

Hui, D., Wan, S., Su, B., Katul, G., Monson, R., and Luo, Y.: Gap-filling missing data in eddy covariance measurements using multiple imputation (MI) for annual estimations, Agricultural and Forest Meteorology, 121, 93–111, https://doi.org/10.1016/S0168-1923(03)00158-8, 2004.

Jacobs, A. F. G., Heusinkveld, B~~., Gerken, T., Zhang, Z., Wakbulcho, G., and Niu, S.: Maximum carbon uptake rate dominates~~. G., and Holtslag, A. A. M.: Towards Closing the ~~interannual variability~~Surface Energy Budget of ~~global net ecosystem exchange, Global Change Biology, 25, 3381–3394, 2019.~~

~~Hao, Y., Wang, Y., Mei, X., and Cui, X.: The response of ecosystem CO2 exchange to small precipitation pulses over~~ a ~~temperate steppe, Plant Ecol, 209, 335–347~~Mid-latitude Grassland,

Boundary-Layer Meteorol, 126, 125–136, https://doi.org/10.1007/s11258-010-9766-1, 2010s10546-007-9209-2, 2008.

Harris, N. L., Gibbs, D. A., Baccini, A., Birdsey, R. A., de Bruin, S., Farina, M., Fatoyinbo, L., Hansen, M. C., Herold, M., Houghton, R. A., Potapov, P. V., Suarez, D. R., Roman-Cuesta, R. M., Saatchi, S. S., Slay, C. M., Turubanova, S. A., and Tyukavina, A.: Global maps of twenty-first century forest carbon fluxes, Nat. Clim. Chang., 11, 234–240, https://doi.org/10.1038/s41558-020-00976-6, 2021.

Huemmrich, K. F., Campbell, P., Landis, D., and Middleton, E.: Developing a common globally applicable method for optical remote sensing of ecosystem light use efficiency, Remote Sensing of Environment, 230, 111190, 2019.

Jongen, M., Pereira, J. S., Aires, L. M. I., and Pio, C. A.: The effects of drought and timing of precipitation on the inter-annual variation in ecosystem-atmosphere exchange in a Mediterranean grassland, Agricultural and Forest Meteorology, 151, 595–606, https://doi.org/10.1016/j.agrformet.2011.01.008, 2011.

Jung, M., Reichstein, M., and Bondeau, A.: Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model, Biogeosciences, 6, 2001–2013, https://doi.org/10.5194/bg-6-2001-2009, 2009.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., and Chen, J.: Global patterns of land - Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, Journal of Geophysical Research: Biogeosciences, 116, 2011.https://doi.org/10.1029/2010JG001566, 2011.

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., RödenbeckSchwalm, C., Tramontana, G., Viovy, N., Wang, Y.-P., Weber, U., Zaehle, S., and Zeng, N.: Compensatory water effects link yearly Reichstein, M.: The FLUXCOM ensemble of global land-CO2 sink changes to temperature, 541, 516–520-atmosphere energy fluxes, Sci Data, 6, 74, https://doi.org/10.1038/nature20780, 2017.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., S Goll, D., Haverd, V., Köhler, P., Ichii, K., K Jain, A., Liu, J., Lombardozzi, D., E M S Nabel, J., A Nelson, J., O'Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., and Reichstein, M.: Scaling carbon fluxes from eddy-

~~covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach, 17, 1343–1365, https://doi.org/10.5194/bg-17-1343-2020, 2020~~s41597-019-0076-8, 2019.

Kaur, H., Pannu, H. S., and Malhi, A. K.: A Systematic Review on Imbalanced Data Challenges in Machine Learning: Applications and Solutions, ACM Comput. Surv., 52, 79:1-79:36, https://doi.org/10.1145/3343440, 2019.

~~Kljun, N., Calanca, P., Rotach, M., and Schmid, H. P.: A simple two-dimensional parameterisation for Flux Footprint Prediction (FFP), Geoscientific Model Development, 8, 3695–3713, 2015.~~

Li, X., He, Y., Zeng, Z., Lian, X., Wang, X., Du, M., Jia, G., Li, Y., Ma, Y., Tang, Y., Wang, W., Wu, Z., Yan, J., Yao, Y., Ciais, P., Zhang, X., Zhang, Y., Zhang, Y., Zhou, G., and Piao, S.: Spatiotemporal pattern of terrestrial evapotranspiration in China during the past thirty years, Agricultural and Forest Meteorology, 259, 131–140, https://doi.org/10.1016/j.agrformet.2018.04.020, 2018.

Li, X., Kang, S., Niu, J., Huo, Z., and Liu, J.: Improving the representation of stomatal responses to CO2 within the Penman–Monteith model to better estimate evapotranspiration responses to climate change, Journal of Hydrology, 572, 692–705, https://doi.org/10.1016/j.jhydrol.2019.03.029, 2019.

Liu, Q., Zhang, Y., Liu, B., Amonette, J. E., Lin, Z., Liu, G., Ambus, P., and Xie, Z.: How does biochar influence soil N cycle? A meta-analysis, Plant and soil, 426, 211–225, ~~2018~~2018a.

~~Luyssaert, S., Janssens, I. A., Sulkava, M., Papale, D., Dolman, A. J., Reichstein, M., Hollmén, J., Martin, J. G., Suni, T., Vesala, T., Loustau, D., Law, B. E., and Moors, E. J.: Photosynthesis drives anomalies in net carbon-exchange of pine forests at different latitudes, 13, 2110–2127, https://doi.org/10.1111/j.1365-2486.2007.01432.x, 2007.~~

~~Marcot, B. G. and Hanea, A. M.: What is an optimal value~~Liu, Y., Xiao, J., Ju, W., Zhu, G., Wu, X., Fan, W., Li, D., and Zhou, Y.: Satellite-derived LAI products exhibit large discrepancies and can lead to substantial uncertainty in simulated carbon and water fluxes, Remote Sensing of Environment, 206, 174–188, https://doi.org/10.1016/j.rse.2017.12.024, 2018b.

Lu, X. and Zhuang, Q.: Evaluating evapotranspiration and water-use efficiency of ~~k~~terrestrial ecosystems in ~~k-fold cross-validation in discrete Bayesian network analysis?, Comput Stat, 36, 2009–2031, https://doi.org/10.1007/s00180-020-00999-9, 2021.~~

~~Mitchell, S., Beven, K., and Freer, J.: Multiple sources~~the conterminous United States using MODIS and AmeriFlux data, Remote Sensing of ~~predictive uncertainty in modeled estimates of net ecosystem CO2 exchange, Ecological Modelling, 220, 3259–3270~~Environment, https://doi.org/10.1016/j.~~ecolmodel.2009.08.021, 2009~~rse.2010.04.001, 2010.

Moffat, A. M., Papale, D., Reichstein, M., Hollinger, D. Y., Richardson, A. D., Barr, A. G., Beckstein Mauder, M., Cuntz, M., Drüe, C., Braswell, B. Graf, A., Rebmann, C., Schmid, H., Churkina, G., Desai, A. . P., Schmidt, M., and Steinbrecher, R., Falge, E., Gove, J. H., Heimann, M., Hui, D., Jarvis, A. J., Kattge, J., Noormets, A., and Stauch, V. J.: Comprehensive comparison of gap-filling techniques.: A strategy for quality and uncertainty assessment of long-term eddy-covariance net carbon fluxes, 147, 209–232measurements, Agricultural and Forest Meteorology, 169, 122–135, https://doi.org/10.1016/j.agrformet.2007.08.011, 2007 2012.09.006, 2013.

Moffat, A. M., Beckstein, C., Churkina, G., Mund, M., and Heimann, M.: Characterization of ecosystem responses to climatic controls using artificial neural networks, 16, 2737–2749, https://doi.org/10.1111/j.1365-2486.2010.02171.x, 2010.

Mauder, M., Genzel, S., Fu, J., Kiese, R., Soltani, M., Steinbrecher, R., Zeeman, M., Banerjee, T., De Roo, F., and Kunstmann, H.: Evaluation of energy balance closure adjustment methods by independent evapotranspiration estimates from lysimeters and hydrological simulations, Hydrological Processes, 32, 39–50, https://doi.org/10.1002/hyp.11397, 2018.

McColl, K. A.: Practical and Theoretical Benefits of an Alternative to the Penman-Monteith Evapotranspiration Equation, Water Resources Research, 56, e2020WR027106, https://doi.org/10.1029/2020WR027106, 2020.

Minacapilli, M., Agnese, C., Blanda, F., Cammalleri, C., Ciraolo, G., D'Urso, G., Iovino, M., Pumo, D., Provenzano, G., and Rallo, G.: Estimation of actual evapotranspiration of Mediterranean perennial crops by means of remote-sensing based surface energy balance models, Hydrology and Earth System Sciences, 13, 1061–1074, https://doi.org/10.5194/hess-13-1061-2009, 2009.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. a. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrology and Earth System Sciences, 15, 453–469, https://doi.org/10.5194/hess-15-453-2011, 2011.

Miralles, D. G., Teuling, A. J., van Heerwaarden, C. C., and Vilà-Guerau de Arellano, J.: Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation, Nature Geosci, 7, 345–349, https://doi.org/10.1038/ngeo2141, 2014.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Prisma Group: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, PLoS medicine, 6, e1000097, 2009.

Moon, T. K.: The expectation-maximization algorithm, 13, 47–60, 1996.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sensing of Environment, 115, 1781–1800, https://doi.org/10.1016/j.rse.2011.02.019, 2011.

Nemani, R. R., Keeling, C. D., Hashimoto, H., Jolly, W. M., Piper, S. C., Tucker, C. J., Myneni, R. B., ~~Papale, D. and Valentini, R.: A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization, 9, 525–535, https://doi.org/10.1046/j.1365-2486.2003.00609.x, 2003.~~

~~Park, S.~~ and Running, S. W.: Climate-Driven Increases in Global Terrestrial Net Primary Production from 1982 to 1999, Science, 300, 1560–1563, https://doi.org/10.1126/science.1082750, 2003.

Pan, S., Tian, H., Dangal, S. R. S., Yang, Q., Yang, J., Lu, C., Tao, B., Ren, W., and Ouyang, Z.: Responses of global terrestrial evapotranspiration to climate change and increasing atmospheric CO2 in the 21st century, Earth's Future, 3, 15–35, https://doi.org/10.1002/2014EF000263, 2015.

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V. K., Haverd, V., Jain, A. K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J. E. M. S., Ottlé, C., Poulter, B., Zaehle, S., and Running, S. W.: Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling, Hydrology and Earth System Sciences, 24, 1485–1509, https://doi.org/10.5194/hess-24-1485-2020, 2020.

Papale, D., Black, T. A., Carvalhais, N., Cescatti, A., Chen, J., Jung, M., Kiely, G., Lasslop, G., Mahecha, M. D., Margolis, H., Merbold, L., Montagnani, L., Moors, E., Olesen, Jø. E., Reichstein, M., Tramontana, G., Van Gorsel, E., Wohlfahrt, G., and Ráduly, B.: Effect of spatial sampling from European flux towers for estimating carbon and water fluxes with artificial neural networks, Journal of Geophysical Research: Biogeosciences, 120, 1941–1957, https://doi.org/10.1002/2015JG002997, 2015.

Paul-Limoges, E., Wolf, S., Schneider, F. D., Longo, M., Moorcroft, P., Gharun, M., and Damm~~B., Knohl, A., Lucas-Moffat, A. M., Migliavacca, M., Gerbig, C., Vesala, T., Peltola, O., Mammarella, I., Kolle, O., Lavrič, J. V., Prokushkin~~, A~~., and Heimann, M.: Strong radiative effect induced by clouds and smoke on~~.: Partitioning evapotranspiration with concurrent eddy covariance measurements in a mixed forest net ecosystem productivity in central Siberia~~, Agricultural and Forest Meteorology, ~~250–251, 376–387~~280, 107786, https://doi.org/10.1016/j.agrformet.~~2017.09.009, 2018~~2019.107786, 2020.

~~Pearl, J.: Bayesian netwerks: A model of self-activated memory for evidential reasoning, in: Proceedings of the 7th Conference of the Cognitive Science Society, University of California, Irvine, CA, USA, 15–17, 1985.~~

~~Peltola, O., Vesala, T., Gao, Y., Räty, O., Alekseychik, P., Aurela, M., Chojnicki, B., Desai, A. R., Dolman, A. J., Euskirchen, E. S., Friborg, T., Göckede, M., Helbig, M., Humphreys,~~

E., Jackson, R. B., Jocher, G., Joos, F., Klatt, J., Knox, S. H., Kowalska, N., Kutzbach, L., Lienert, S., Lohila, A., Mammarella, I., Nadeau, D. F., Nilsson, M. B., Oechel, W. C., Peichl, M., Pypker, T., Quinton, W., Rinne, J., Sachs, T., Samson, M., Schmid, H. P., Sonnentag, O., Wille, C., Zona, D., and Aalto, T.: Monthly gridded data product of northern wetland methane emissions based on upscaling eddy covariance observations, Earth System Science Data, 11, 1263–1289, https://doi.org/10.5194/essd-11-1263-2019, 2019.

Reed, D. E., Poe, J., Abraha, M., Dahlin, K. M., and Chen, J.: Modeled Surface-Atmosphere Fluxes From Paired Sites in the Upper Great Lakes Region Using Neural Networks, Journal of Geophysical Research: Biogeosciences, 126, https://doi.org/10.1029/2021JG006363, 2021.

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

Reitz, O., Graf, A., Schmidt, M., Ketzler, G., and Leuchner, M.: Upscaling Net Ecosystem Exchange Over Heterogeneous Landscapes With Machine Learning, 126, e2020JG005814, https://doi.org/10.1029/2020JG005814, 2021.

Ruppert, J., Mauder, M., Thomas, C., and Lüers, J.: Innovative gap-filling strategy for annual sums of CO2 net ecosystem exchange, 138, 5–18, https://doi.org/10.1016/j.agrformet.2006.03.003, 2006.

Shi, H., Luo, G., Zheng, H., Chen, C., Bai, J., Liu, T., Ochege, F. U., and De Maeyer, P.: Coupling the water-energy-food-ecology nexus into a Bayesian network for water resources analysis and management in the Syr Darya River basin, Journal of Hydrology, 581, 124387, https://doi.org/10.1016/j.jhydrol.2019.124387, 2020.

Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, Hydrology and Earth System Sciences, 11, 1633–1644, https://doi.org/10.5194/hess-11-1633-2007, 2007.

Rigden, A. J. and Salvucci, G. D.: Evapotranspiration based on equilibrated relative humidity (ETRHEQ): Evaluation over the continental U.S., Water Resources Research, 51, 2951–2973, https://doi.org/10.1002/2014WR016072, 2015.

Sahoo, A. K., Pan, M., Troy, T. J., Vinukollu, R. K., Sheffield, J., and Wood, E. F.: Reconciling the global terrestrial water budget using satellite remote sensing, Remote Sensing of Environment, 115, 1850–1865, https://doi.org/10.1016/j.rse.2011.03.009, 2011.

Sándor, R., Barcza, Z., Hidy, D., Lellei-Kovács, E., Ma, S., and Bellocchi, G.: Modelling of grassland fluxes in Europe: Evaluation of two biogeochemical models, Agriculture, Ecosystems & Environment, 215, 1–19, https://doi.org/10.1016/j.agee.2015.09.001, 2016.

Shi, H., Hellwich, O., Luo, G., Chen, C., He, H., Ochege, F. U., Van de Voorde, T., Kurban, A., and de Maeyer, P.: A global meta-analysis of soil salinity prediction integrating satellite

remote sensing, soil sampling, and machine learning, IEEE Transactions on Geoscience and Remote Sensing, 1–15, https://doi.org/10.1109/TGRS.2021.3109819, 2021.

~~Tian, X., Yan, M., van der Tol, C., Li, Z., Su, Z., Chen, E., Li, X., Li, L., Wang, X., Pan, X., Gao, L., and Han, Z.: Modeling forest above-ground biomass dynamics using multi-source data and incorporated models: A case study over the qilian mountains, Agricultural and Forest Meteorology, 246, 1–14, https://doi.org/10.1016/j.agrformet.2017.05.026, 2017.~~

Shi, H., Luo, G., Hellwich, O., Xie, M., Zhang, C., Zhang, Y., Wang, Y., Yuan, X., Ma, X., and Zhang, W.: Variability and Uncertainty in Flux-Site Scale Net Ecosystem Exchange Simulations Based on Machine Learning and Remote Sensing: A Systematic Evaluation, Biogeosciences Discussions, 1–25, 2022.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences, 13, 4291–4313, https://doi.org/10.5194/bg-13-4291-2016, 2016.

Van Hulse, J., Khoshgoftaar, T. M., and Napolitano, A.: Experimental perspectives on learning from imbalanced data, in: Proceedings of the 24th international conference on Machine learning, New York, NY, USA, 935–942, https://doi.org/10.1145/1273496.1273614, 2007.

Van Wijk, M. T. and Bouten, W.: Water and carbon fluxes above European coniferous forests modelled with artificial neural networks, Ecological Modelling, https://doi.org/10.1016/S0304-3800(99)00101-5, 1999.

Virkkala, A.-M., Aalto, J., Rogers, B. M., Tagesson, T., Treat, C. C., Natali, S. M., Watts, J. D., Potter, S., Lehtonen, A., Mauritz, M., Schuur, E. A. G., Kochendorfer, J., Zona, D., Oechel, W., Kobayashi, H., Humphreys, E., Goeckede, M., Iwata, H., Lafleur, P. M., Euskirchen, E. S., Bokhorst, S., Marushchak, M., Martikainen, P. J., Elberling, B., Voigt, C., Biasi, C., Sonnentag, O., Parmentier, F.-J. W., Ueyama, M., Celis, G., St.Louis, V. L., Emmerton, C. A., Peichl, M., Chi, J., Järveoja, J., Nilsson, M. B., Oberbauer, S. F., Torn, M. S., Park, S.-J., Dolman, H., Mammarella, I., Chae, N., Poyatos, R., López-Blanco, E., Christensen, T. R., Kwon, M. J., Sachs, T., Holl, D., and Luoto, M.: Statistical upscaling of ecosystem CO2 fluxes across the terrestrial tundra and boreal domain: Regional patterns and uncertainties, Global Change Biology, 27, 4040–4059, https://doi.org/10.1111/gcb.15659, 2021.

~~Walther, S., Besnard, S., Nelson, J. A., El-Madany, T. S., Migliavacca, M., Weber, U., Ermida, S. L., Brümmer, C., Schrader, F., Prokushkin, A. S., Panov, A. V., and Jung, M.: Technical note: A view from space on global flux towers by MODIS and Landsat: The FluxnetEO dataset, Biogeosciences Discussions, 1–40, https://doi.org/10.5194/bg-2021-314, 2021.~~

~~Wu, Z., Dijkstra, P., Koch, G. W., Peñuelas, J., and Hungate, B. A.: Responses of terrestrial ecosystems to temperature and precipitation change: a meta-analysis of experimental manipulation, 17, 927–942, https://doi.org/10.1111/j.1365-2486.2010.02302.x, 2011.~~

~~Yan, J., Zhang, Y., Yu, G., Zhou, G., Zhang, L., Li, K., Tan, Z., and Sha, L.: Seasonal and inter-annual variations in net ecosystem exchange of two old-growth forests in southern China, Agricultural and Forest Meteorology, 182–183, 257–265, https://doi.org/10.1016/j.agrformet.2013.03.002, 2013.~~

Wagle, P., Bhattarai, N., Gowda, P. H., and Kakani, V. G.: Performance of five surface energy balance models for estimating daily evapotranspiration in high biomass sorghum, ISPRS Journal of Photogrammetry and Remote Sensing, 128, 192–203, https://doi.org/10.1016/j.isprsjprs.2017.03.022, 2017.

Xie, M., Luo, G., Hellwich, O., Frankl, A., Zhang, W., Chen, C., Zhang, C., and De Maeyer, P.: Simulation of site-scale water fluxes in desert and natural oasis ecosystems of the arid region in Northwest China, Hydrological Processes, 35, e14444, https://doi.org/10.1002/hyp.14444, 2021.

Xu, T., Guo, Z., Liu, S., He, X., Meng, Y., Xu, Z., Xia, Y., Xiao, J., Zhang, Y., Ma, Y., and Song, L.: Evaluating Different Machine Learning Methods for Upscaling Evapotranspiration from Flux Towers to the Regional Scale, Journal of Geophysical Research: Atmospheres, 123, 8674–8690, https://doi.org/10.1029/2018JD028447, 2018.

Yang, F., White, M. A., Michaelis, A. R., Ichii, K., Hashimoto, H., Votava, P., Zhu, A.-X., and Nemani, R. R.: Prediction of Continental-Scale Evapotranspiration by Combining MODIS and AmeriFlux Data Through Support Vector Machine, IEEE Transactions on Geoscience and Remote Sensing, 44, 3452–3461, https://doi.org/10.1109/TGRS.2006.876297, 2006.

Zeng, J., Matsunaga, T., Tan, Z.-H., Saigusa, N., Shirai, T., Tang, Y., Peng, S., and Fukuda, Y.: Global terrestrial carbon fluxes of 1999–2019 estimated by upscaling eddy covariance data with a random forest, Scientific Data, 7, https://doi.org/10.1038/s41597-020-00653-5, 2020.

Zeng, Y., Hao, D., Huete, A., Dechant, B., Berry, J., Chen, J. M., Joiner, J., Frankenberg, C., Bond-Lamberty, B., Ryu, Y., Xiao, J., Asrar, G. R., and Chen, M.: Optical vegetation indices for monitoring terrestrial ecosystems globally, Nat Rev Earth Environ, 1–17, https://doi.org/10.1038/s43017-022-00298-5, 2022.

Zhang, C., ~~Brodylo, D., Sirianni, M. J., Li, T., Comas, X., Douglas, T. A., and Starr, G.: Mapping CO2 fluxes of cypress swamp~~Luo, G., Hellwich, O., Chen, C., Zhang, W., Xie, M., He, H., Shi, H., and ~~marshes in the Greater Everglades using eddy covariance measurements and Landsat~~Wang, Y.: A framework for estimating actual evapotranspiration at weather stations without flux observations by combining data~~, Remote Sensing of Environment, 262~~

from MODIS and flux towers through a machine learning approach, Journal of Hydrology, 603, 127047, https://doi.org/10.1016/j.~~rse~~jhydrol.2021.~~112523~~127047, 2021.

~~Zhou, Y., Li, X., Gao, Y., He, M., Wang, M., Wang, Y., Zhao,~~ Zhang, K., Kimball, J. S., Nemani, R. R., ~~L.,~~ ~~and Li, Y.: Carbon fluxes response of an artificial sand-binding vegetation system to rainfall variation during the growing season in the Tengger Desert, Journal of Environmental Management, 266, https://doi.org/10.1016/j.jenvman.2020.110556, 2020.~~

~~Zhu, S., Clement, R., McCalmont, J., Davies, C. A., and Hill, T.: Stable gap-filling for longer eddy covariance data gaps: A globally validated machine-learning approach for carbon dioxide, water, and energy fluxes, Agricultural and Forest Meteorology, 314, 108777, https://doi.org/10.1016/j.agrformet.2021.108777, 2022.~~

and Running, S. W.: A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006, Water Resources Research, 46, https://doi.org/10.1029/2009WR008800, 2010.

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

设置了格式: 字体: 10 磅

带格式的: 行距: 单倍行距