Exploring the Potential of Long Short-Term Memory Networks for Predicting Net CO2 Exchange Across Various Ecosystems With Multi-Source Data

Chengcheng Huang[1], Wei He[2], Jinxiu Liu[3], Ngoc Tu Nguyen[4], Hua Yang[5], Yiming Lv[6], Hui Chen[1], and Mengyao Zhao[7]

[1]School of Information Engineering, China University of Geosciences
[2]International Institute for Earth System Science, Nanjing University
[3]China University of Geosciences
[4]Hohai University
[5]Beijing Normal University
[6]China University of Geosciences, Beijing
[7]School of Geography and Tourism, Anhui Normal University

November 14, 2023

Abstract

Upscaling flux tower measurements based on machine learning (ML) algorithms is an essential approach for large-scale net ecosystem CO2 exchange (NEE) estimation, but existing ML upscaling methods face some challenges, particularly in capturing NEE interannual variations (IAVs) that may relate to lagged effects. With the capacity of characterizing temporal memory effects, the Long Short-Term Memory (LSTM) networks are expected to help solve this problem. Here we explored the potential of LSTM for predicting NEE across various ecosystems using flux tower data over 82 sites in North America. The LSTM model with differentiated plant function types (PFTs) demonstrates the capability to explain 79.19% (R2 = 0.79) of the monthly variations in NEE within the testing set, with RMSE and MAE values of 0.89 and 0.57 g C m-2 d-1 respectively (r = 0.89, p < 0.001). Moreover, the LSTM model performed robustly in predicting cross-site variability, with 67.19% of the sites that can be predicted by both LSTM models with and without distinguished PFTs showing improved predictive ability. Most importantly, the IAV of predicted NEE highly correlated with that in flux observations (r = 0.81, p < 0.001), clearly outperforming that by the random forest model (r = -0.21, p = 0.011). Among all nine PFTs, solar-induced chlorophyll fluorescence, downward shortwave radiation, and leaf area index are the most important variables for explaining NEE variations, collectively accounting for approximately 54.01% in total. This study highlights the great potential of LSTM for improving carbon flux upscaling with multi-source remote sensing data.

Hosted file

979209_0_art_file_11580370_s3yzrp.docx available at https://authorea.com/users/699314/articles/686738-exploring-the-potential-of-long-short-term-memory-networks-for-predicting-net-co2-exchange-across-various-ecosystems-with-multi-source-data

Hosted file

979209_0_supp_11580371_s3yd1v.docx available at https://authorea.com/users/699314/articles/686738-exploring-the-potential-of-long-short-term-memory-networks-for-predicting-net-co2-exchange-across-various-ecosystems-with-multi-source-data

Exploring the Potential of Long Short-Term Memory Networks for Predicting Net CO_2 Exchange Across Various Ecosystems With Multi-Source Data

Chengcheng Huang[1], Wei He[2,3*], Jinxiu Liu[1], Ngoc Tu Nguyen[4], Hua Yang[5,6], Yiming Lv[1], Hui Chen[1], and Mengyao Zhao[7]

[1]School of Information Engineering, China University of Geosciences, Beijing 100083, China.

[2]International Institute for Earth System Science, Nanjing University, Nanjing, Jiangsu 210023, China.

[3]Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Key Laboratory for Land Satellite Remote Sensing Applications of Ministry of Natural Resources, School of Geography and Ocean Science, Nanjing University, Nanjing, Jiangsu 210023, China.

[4]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, College of Hydrology and Water Resources, Hohai University, Nanjing, Jiangsu 210024, China.

[5]State Key Laboratory of Remote Sensing Science Jointly Sponsored by Beijing Normal University and Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100854, China.

[6]Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China.

[7]School of Geography and Tourism, Anhui Normal University, Wuhu, Anhui 241002, China.

Corresponding author: Wei He (weihe@nju.edu.cn)

Key Points:

- The LSTM model with differentiated PFTs demonstrates the capability to explain 79.19% of the monthly variations in NEE.
- The LSTM model exhibited clear advantages over the RF model in capturing the interannual variations of NEE.
- The relative importance of feature variables for predicting monthly NEE dynamics across different PFTs in North America was quantified.

Abstract

Upscaling flux tower measurements based on machine learning (ML) algorithms is an essential approach for large-scale net ecosystem CO_2 exchange (NEE) estimation, but existing ML upscaling methods face some challenges, particularly in capturing NEE interannual variations (IAVs) that may relate to lagged effects. With the capacity of characterizing temporal memory effects, the Long Short-Term Memory (LSTM) networks are expected to help solve this problem. Here we explored the potential of LSTM for predicting NEE across various ecosystems using flux tower data over 82 sites in North America. The LSTM model with differentiated plant function types (PFTs) demonstrates the capability to explain 79.19% (R^2 = 0.79) of the monthly variations in NEE within the testing set, with RMSE and MAE values of 0.89 and 0.57 g C m^{-2} d^{-1} respectively (r = 0.89, p < 0.001). Moreover, the LSTM model performed robustly in predicting cross-site variability, with 67.19% of the sites that can be predicted by both LSTM models with and without distinguished PFTs showing improved predictive ability. Most importantly, the IAV of predicted NEE highly correlated with that in flux observations (r = 0.81, p < 0.001), clearly outperforming that by the random forest model (r = -0.21, p = 0.011). Among all nine PFTs, solar-induced chlorophyll fluorescence, downward shortwave radiation, and leaf area index are the most important variables for explaining NEE variations, collectively accounting for approximately 54.01% in total. This study highlights the great potential of LSTM for improving carbon flux upscaling with multi-source remote sensing data.

Plain Language Summary

Net ecosystem exchange (NEE) of CO_2 is a crucial process that regulates carbon exchange between terrestrial ecosystems and the atmosphere. Currently, the growing availability of NEE measurement data, multi-source remote sensing data and meteorological data, has made machine learning algorithms a popular approach for estimating large-scale NEE. Various types of NEE datasets have been derived with different methods; however, the ability in representing the memory effects of climate and environmental factors remains a significant source of uncertainty contributed to NEE estimates. To address this issue, we constructed site-level LSTM training models by plant function types in North America for improving the monthly-scale simulation of NEE and its interannual variations. The established LSTM model enables the prediction of the temporal variability of NEE and effectively captures the memory effects over time, showing a great potential for improving carbon flux upscaling.

1 Introduction

The net exchange of CO_2 between terrestrial ecosystems and the atmosphere (NEE) is an essential component of the global carbon cycle (Bonan, 2008; Shevliakova et al., 2013). Accurately estimating NEE is an essential step towards enhancing our understanding of the feedback between the terrestrial carbon cycle and climate change and better predicting future climate status. Accurately quantifying terrestrial NEE is also a prerequisite for implementing net-zero policies. However, estimating large-scale NEE faces great challenges due to the complex relationships among the physical, chemical, and biological processes.

Currently, there are three main ways for large-scale NEE estimation, including top-down atmospheric CO_2 inversions, terrestrial biosphere models (TBMs) and eddy flux upscaling, the latter two also calls bottom-up approaches. The top-down approach infers biosphere CO_2 fluxes from atmospheric CO_2 observations onboard different observation platforms, such as tall towers, aircraft, ships, and satellites (Ciais et al., 2014), which utilizes atmospheric CO_2 data and a transport model to deduce the spatiotemporal distribution of carbon fluxes. Atmospheric inversions are particularly beneficial for constraining large-scale carbon fluxes (He et al., 2023a; He et al., 2023b), but providing limited spatial information on smaller scales, as uncertainties increase with spatial scale decreases. The process-based TBMs consider the physical processes of energy, carbon, and water cycle regulation. Nevertheless, the complexity of the model structure and the inherent assumptions of specific parameters contribute to substantial discrepancies in NEE simulations among various ecosystem models (Huntzinger et al., 2012). Recently, the bottom-up approach for extrapolating eddy covariance (EC) data, i.e., flux upscaling, shows advantages in accurately quantifying large-scale carbon fluxes. Traditionally, EC technology has been used for continuous measurements at flux sites to develop and evaluate NEE models at the site level. Subsequently, by using spatial variability predominantly driven by Earth observation data, the net exchange of CO_2 and energy between terrestrial ecosystems and the atmosphere can be estimated through spatial extrapolation. Empirical models use statistics to identify certain patterns between meteorological and satellite remote sensing observations, enabling them to capture even highly nonlinear relationships among explanatory variables and carbon fluxes. With the growing availability of global flux observation data and multi-source remote sensing data, there is an increasing interest in encouraging machine learning (ML) technology to become another promising method for NEE prediction. Data-driven ML methods are simple and effective in evaluating NEE, as they are entirely adaptable to the data and do not rely on assumptions about terrestrial ecosystem patterns (Peylin et al., 2013). Various ML algorithms have made advancements in estimating ecosystem carbon fluxes and exchange, including Artificial Neural Networks (Papale & Valentini, 2003), Model Tree Ensemble (Liang et al., 2020), and Random Forest (RF; Guo et al., 2023).

However, despite significant progress have made in the field of empirically upscaling NEE from in-situ EC measurements, various sources of uncertainty remain (Jung et al., 2020). Firstly, many regions around the world only provide point measurements from sparse flux site networks (Tramontana et al., 2016), which contributes to a significant uncertainty regarding NEE upscaling at the regional scale. Moreover, the accuracy of ML methods for estimating carbon fluxes depends heavily on the variables used as driving factors and the limited information available regarding all major ecosystem features that influence carbon fluxes (Huang et al., 2021). An essential aspect of data limitation is the accessibility of pertinent explanatory variables, which correspond to in-situ information at the site level and corresponding global networks. Additionally, predictive factors can also hinder the evaluation of NEE variability, emphasizing the need for more important feature variables to enhance our understanding of NEE. In addition, the upscaling method could also impact flux upscaling, since their abilities in characterizing the relationship between carbon flux and feature variables vary notably (Tramontana et al., 2016; Jung et al., 2020).

When employing ML methods for spatial estimation of different carbon and energy fluxes, NEE is recognized as the most challenging flux to predict (Bodesheim et al., 2018; Jung et al., 2011; Tramontana et al., 2016). Particularly, the interannual variation (IAV) of NEE has not been accurately estimated (Jung et al., 2020), predominantly due to the inability to represent temporal dynamics of climate and vegetation activities. Extreme climate events and human disturbances exhibit memory effects in the response of NEE. These effects refer to the influence of past climate and environmental conditions on current and future ecosystem responses (Ogle et al., 2015). This can lead to nonnegligible interannual changes in the terrestrial carbon budget. The FLUXNET and AmeriFlux networks are composed of EC flux towers, which offer long-term, high-temporal resolution measurements of the site-scale NEE. Remote sensing, being a potentially powerful technology, offers ecosystem observations with consistent spatial and temporal coverage. Recent rapid development of deep learning (DL) technology has shed new light on Earth system modeling (Irrgang et al., 2021). In particular, its capacity for mining historical time-series information from multi-source ecosystem observations offers a great potential for improving terrestrial carbon flux estimation (Besnard et al., 2019; Liu et al., 2023), which incorporates environmental memory into flux modeling while difficult to implement in state-of-the-art process models. The Long Short-Term Memory model (LSTM) is a dynamic statistical method that has demonstrated excellent performance on sequence data, such as crop field classification (Rußwurm & Körner, 2018). With its distinctive design, the LSTM model can effectively address long-term considerations and incorporate memory effects of climate and vegetation, thus aiding in the representation of interannual fluctuations in carbon fluxes (Besnard et al., 2019). To support this concept, we developed and applied an LSTM model to predict site-level NEE in North America. This model utilizes meteorological and flux data sets from FLUXNET and AmeriFlux networks, along with multi-source remote sensing data. We use continuous monthly NEE data, which represent direct samples of NEE from sites encompassing diverse biological communities and climate types in North America. The predictive performance of the LSTM model was assessed in combination with NEE data obtained from EC flux towers, regarding spatial variability and interannual changes in monthly NEE at both site and ecosystem levels. The advantage of the LSTM model to capture climate and vegetation memory effects in quantifying spatiotemporal variations in NEE was analyzed.

The reliability of spatial-resolved NEE estimation over large regions is constrained by the predictive capability of ML- or DL- based upscaling models at the site level. Thus, it is a prerequisite to address significant challenges in accurately modeling site-level NEE before conducting large-scale flux estimation. The objectives of this study are to (a) investigate the differences of established LSTM models with and without distinguishing PFTs in describing monthly NEE variations at the plant functional type (PFTs) level, (b) analyze the variability of

LSTM model performances across sites, (c) evaluate the ability of the PFT-based LSTM models in capturing the IAV of NEE and compare with the modeling results using widely used RF models, and (d) quantify the relative importance of feature variables for predicting monthly NEE dynamics across different PFTs in North America.

2 Materials and Methods

2.1. Dataset and Preprocessing

The FLUXNET is a worldwide ecosystem observational network composing observation sites distributed around the globe. These sites are situated in diverse ecosystems, such as forests, grasslands, cropland, etc. AmeriFlux is a network especially dedicated to monitoring terrestrial ecosystems in the Americas. The observation stations affiliated with both networks utilize high-precision instruments and equipment to record meteorological and ecosystem data (Baldocchi, 2020; Novick et al., 2018). Researchers use data from the flux networks to analyze and comprehend factors related to climate change and energy and material exchange processes in terrestrial ecosystems, particularly NEE and GPP (Guo et al., 2023; Xu et al., 2019). These measurements are reliable, allowing for robust analysis of daily, monthly, and interannual variations in the North American region.

When training the site-level ML algorithm for each site, we constructed a feature dataset to indicate vegetation growth status. We retrieved monthly NEE and environmental variables from the FLUXNET2015 (Pastorello et al., 2020) and AmeriFlux data sets (Novick et al., 2018), including wind speed (WS), vapor pressure deficit (VPD), air temperature (TA), soil water content (SWC), downward shortwave radiation (DSR), and precipitation (P). To gain a deeper understanding of plant responses to extreme events like droughts and floods, we specifically chose SWC to analyze the impact of soil moisture conditions on NEE, despite this may result in a loss of some site candidates because of SWC unavailability.

The selected remote sensing variables included the normalized difference vegetation index (NDVI), leaf area index (LAI), solar-induced chlorophyll fluorescence (SIF), and the fraction of absorbed photosynthetically active radiation (FAPAR). This study utilized the Global Land Surface Satellite (GLASS) LAI and FAPAR products (Liang et al., 2021). LAI represents half of the total green leaf area per unit of horizontal land surface, and it is a fundamental land climate variable defined by the Global Climate Observing System (GCOS) (Fang et al., 2013). FAPAR is a crucial biophysical variable that directly reflects the photosynthetic activity of plants (Gower et al., 1999). NDVI, which is a normalized ratio of the near-infrared (NIR) and red bands, is valuable data for detecting vegetation status (Yin et al., 2022). We use the PKU Global Inventory Monitoring and Modeling Studies (GIMMS) NDVI product (Li et al., 2023). During the process of plant photosynthesis, leaves absorb photosynthetically active radiation (PAR) and release the unused portion of the absorbed energy in the form of fluorescence, which is referred to as SIF (Verrelst et al., 2016). SIF has a direct and close relationship with photosynthesis and is reported to highly correlate with NEE (Shiga et al., 2018). However, previous ML predictions of

NEE seldom incorporated SIF as a feature variable, which may relate to its low spatial resolution and spatially discontinuation in original satellite retrievals. Here we employed a high-resolution (0.05°) contiguous reanalysis SIF dataset (GOSIF) (Li & Xiao, 2019), primarily derived from OCO-2 SIF data, to characterize the response of NEE to climate and environment. To match the site level NEE data, we utilized the monthly remote sensing observation data from GLASS LAI and FAPAR products by averaging 8-day data, with a spatial resolution of 0.05° × 0.05°. For remote sensing data, the pixels covering the site were used to monitor vegetation growth at the site level. The values at the coordinates of each site are extracted for model training and validation. Following Ukkola et al. (2021), we employed the cubic spline function to fill in the blank of the monthly time series obtained, and any negative feature data was set to zero.

Through this approach, we created a comprehensive dataset at monthly scale. The dataset comprises one label data (NEE) and 10 feature variables (WS, VPD, TA, SWC, DSR, P, NDVI, LAI, SIF, and FAPAR) that are closely associated with NEE. To match the length of comprehensive memory effects in each ecosystem and ensure adequate volume of data for LSTM analysis, we only considered sites with at least one and a half years of NEE records. We ultimately selected 7471 monthly data records from 82 sites distributed in 9 biological communities in North America, covering the period from 2001 to 2020 (Figure S1 in Supporting Information S1). These records encompass measurements of carbon fluxes and meteorological data. The sites that were selected cover a diverse range of climatic conditions and ecosystems. Following the vegetation classification scheme of the International Geosphere-Biosphere Program (IGBP), those sites include 9 vegetation types: evergreen needleleaf forest (ENF; n = 22), grassland (GRA; n = 17), deciduous broadleaf forest (DBF; n = 11), open shrubland (OSH; n = 10), cropland (CRO; n = 9), permanent wetland (WET; n = 7), closed shrubland (CSH; n = 2), mixed forest (MF; n = 2) and woody savanna (WSA; n = 2). The type of cropland/natural vegetation mosaics (CVM; n = 1) was not used for model establishment due to limited site observation data. Our analysis involved 35 sites from FLUXNET and 47 sites from AmeriFlux. This analysis is based on NEE data and focuses on conducting monthly scale simulations and interannual variation predictions across different PFTs and sites.

2.2. The LSTM-based NEE model

2.2.1. LSTM Algorithm

Recurrent Neural Networks (RNNs) can learn to recursively use internal memory states to process sequential data (Thireou & Reczko, 2007). It has emerged as a valuable tool for studying vegetation and climate history through time series observations (Reichstein et al., 2018). By internally transmitting data, RNNs effectively encode the information seen at past time-steps, enabling them to capture temporal dependencies and patterns. As an enhanced variation of RNN, the Long Short-Term Memory Networks (LSTMs) adeptly model long-term dependencies by regulating the information flow (Hochreiter & Schmidhuber, 1997). The connections between units in the LSTM layer create a directed graph along the sequence, illustrating the dynamic temporal behavior of time series in this RNN architecture. LSTM can selectively store and extract information relevant to the problem at each time-step, thereby enabling a better adaption to the memory effect of environmental variables on the carbon cycle of terrestrial ecosystems.

2.2.2. Design of NEE Prediction Model

Different PFTs exhibit distinct characteristics and ecological processes, leading to diverse NEE responses to ecosystem carbon cycle and climate changes. To enhance the utilization of existing data resources and increase model flexibility, our method directly establishes the LSTM model for different PFTs to estimate NEE. NEE observations obtained from FLUXNET2015 and AmeriFlux networks were used as the label data for time series prediction. The site-level inputs are decomposed into 9 separate PFT groups (ENF, GRA, DBF, OSH, CRO, WET, CSH, MF, and WSA). Then, we create individual LSTM model for each PFT, optimizing the model parameters specifically for the PFT site being applied. Furthermore, we included all training data in an LSTM model and did not consider PFTs during the model establishment process. To evaluate whether distinguishing PFTs leads to improved model performance in NEE estimation, we compared the accuracy of each PFT DL model with that of the model without differentiating PFTs. The model that distinguishes PFTs are referred to as PFT_LSTM models, whereas the latter are referred to as nonPFT_LSTM models.

The developed LSTM deep learning (DL) model framework is illustrated in Figure 1. It employs an LSTM layer for the processing and modeling of time series data. Following the LSTM layer, a dropout layer is incorporated to randomly disregard a portion of neuron outputs during training, reducing the interdependence between neurons (Baldi & Sadowski, 2014). Early stopping is implemented to enhance the generalization ability of the networks. The final fully connected layer is responsible for mapping the output of the dropout layer to the target variable NEE. We calculate the Mean Squared Error between the predicted results and the label data (monthly NEE) as the loss function (Rumelhart et al., 1986). To obtain the optimal model, we employ the Adam optimizer to minimize this loss function (Kingma & Ba, 2017). To achieve the best model performance, a grid search was employed to determine major model parameters (Bergstra & Bengio, 2012): learning rate (0.01, 0.001), number of hidden neurons (32, 64, 128, 256), weight decay coefficient (0.01, 0.001, 0.0001), dropout rate (0.1, 0.2, 0.3), and batch size (8, 16, 32).

Throughout this process, we tested various parameter combinations. Thus, we selected the parameter set that showed the least deviation between the observed monthly NEE data provided by each PFT and the corresponding data predicted by the model. To ensure the comprehensive utilization of time information, the initial 70% of both the feature data set and label data set from each site served as training data to optimize the weights of networks. The remaining 30% was employed as test data to assess the model performance. We conducted time validation for each PFT. The training and testing data sets were used for the development and evaluation of prediction models, respectively. We utilize EC tower-measured data (WS, VPD, TA, DSR, SWC, and P) along with remote sensing data (NDVI, LAI, FAPAR, and SIF) to train the LSTM model and estimate site-level NEE in North America at monthly intervals.

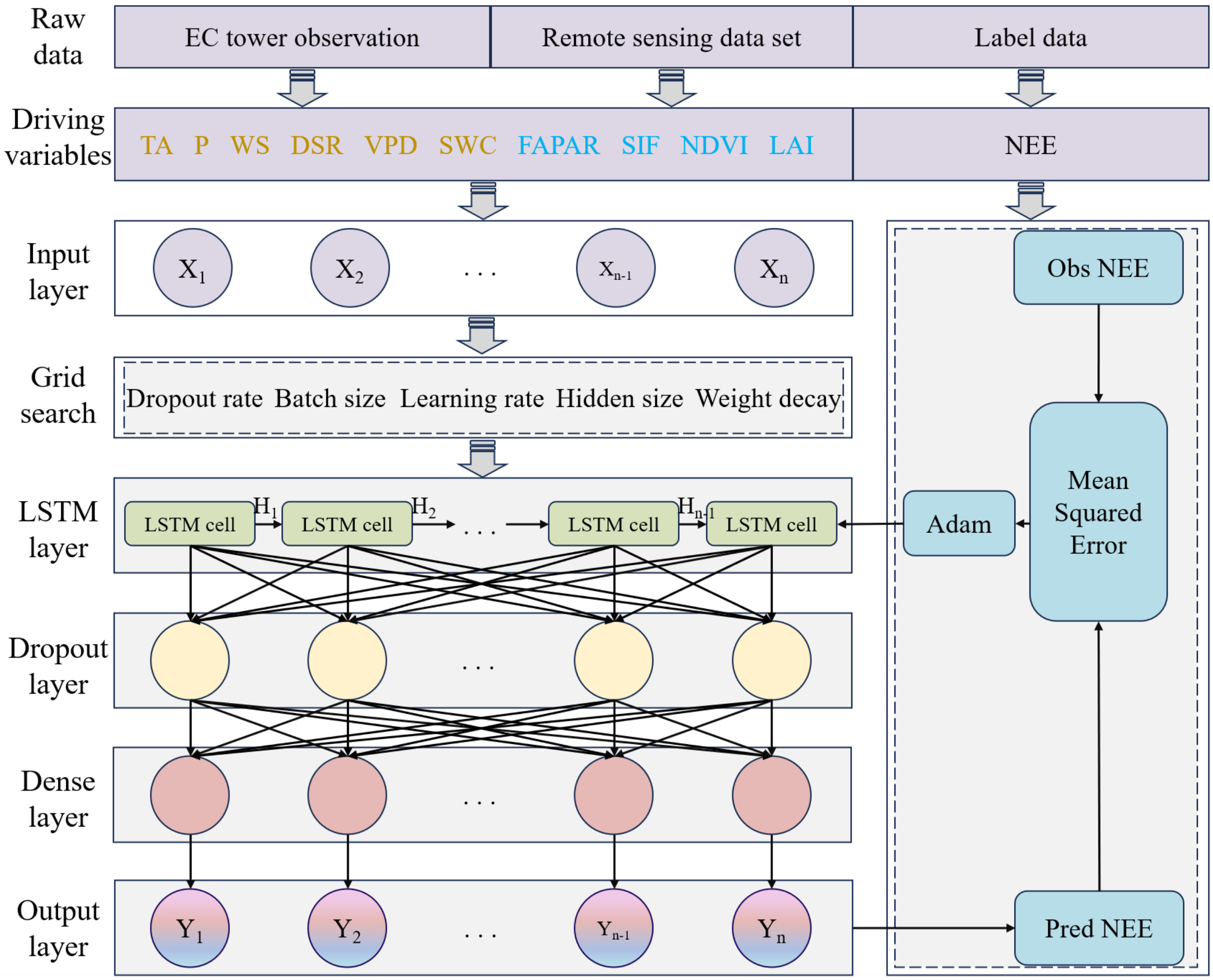

Figure 1. The architecture of the designed LSTM deep learning model.

2.3. Model evaluation and uncertainty assessment

In this study, we determined the optimal prediction by iterating the models used to fit the feature dataset. This approach allowed us to simplify, or at least quantify the empirical uncertainty caused by the random initialization of the LSTM model. We conducted 10 simulation training sessions using LSTM for each model to reflect the uncertainty in the model output (Besnard et al., 2019). Based on this training approach, 90 DL models with optimal parameters were eventually obtained. For each PFT, these models were employed to generate 10 sets of predictions for NEE. The uncertainty range of the model output was determined by the interquartile range of these 10 predicted sets, while the final estimate was derived from the median of the predicted set. We evaluated the accuracy of each model during the testing period using three indicators, coefficient of determination (R^2), Root Mean Square Error, and Mean Absolute Error (MAE). These indicators are defined as,

$$R^2 = 1 - \frac{\sum_{k=1}^{N}\left(Y_i^{obs} - Y_i^{pred}\right)^2}{\sum_{k=1}^{N}\left(Y_i^{obs} - \overline{Y}_i^{obs}\right)^2} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N}\left(Y_i^{obs} - Y_i^{pred}\right)^2}{N}} \tag{2}$$

$$MAE = \frac{\sum_{k=1}^{N}\left|Y_i^{obs} - Y_i^{pred}\right|}{N} \tag{3}$$

where Y_i^{obs} and $\overline{Y}_i^{obs}$ are the observation value and mean observations, and Y_i^{pred} is the model predictions.

2.4. Variable Importance Analysis

Due to the diverse nonlinear responses of ecosystems to climate conditions and environmental control, complex spatiotemporal variability in NEE exists within and across ecosystems. While traditional LSTM DL algorithms are capable of learning system modeling and capturing dynamic behavior from observations, they cannot provide explanations for the spatiotemporal variability of carbon fluxes (Perez-Suay et al., 2020). In the case of carbon fluxes, it is essential for an ML model to identify and clarify the most significant environmental driving factor. In revealing the interaction between vegetation biological characteristics and the environment, quantifying the contribution of these driving factors to monthly NEE changes poses a significant challenge. The presence of imbalanced sample data among PFT sites can hinder the effectiveness of statistical analysis, thereby limiting the reliability of traditional statistical methods in inferring the impact of variables on monthly NEE (Stoy et al., 2009).

To establish a quantitative framework for quantifying the importance of control factors on NEE changes at PFT sites, we used the boosted regression trees (BRT) model (Elith et al., 2008). The BRT model is a ML method that effectively connects environmental variables with monthly scale NEE data. It is able to capture physically complex and nonlinear relationships as well as interactions among variables (Kong et al., 2022; Li et al., 2020). This advantage makes it particularly suitable for quantifying the contribution of predicted variables to monthly NEE. The BRT model can identify key features related to the target variable by evaluating the importance of each feature within the model. We used the BRT model to discern the primary plant traits and environmental factors that drive NEE changes in each PFT.

3 Results

3.1. Prediction Performance at the PFT Level

The LSTM models were used to estimate monthly NEE for various PFTs over the flux sites in North America. We firstly investigated the impact of differentiating PFT to the prediction accuracy. Figure 2 shows the performance comparison between the LSTM models with distinguished PFT (PFT_LSTM) and the ones without differentiating PFTs (nonPFT_LSTM) for monthly NEE predictions. Among all PFTs, the NEE values predicted by the PFT_LSTM model highly correlated with the observations (r = 0.89, p < 0.001), slightly outperforming the nonPFT_LSTM models (r = 0.85, p < 0.001). In the PFT_LSTM models, their R^2 over all PFT sites increased from 0.72 to 0.79, RMSE decreased from 0.98 g C m^{-2} d^{-1} to 0.89 g C m^{-2} d^{-1}, and MAE decreased from 0.62 g C m^{-2} d^{-1} to 0.57 g C m^{-2} d^{-1}. PFT_LSTM models demonstrated higher accuracy compared to the nonPFT_LSTM models, with significantly higher R^2 and lower RMSE and MAE.

The performances of the LSTM models varied across PFTs (R^2 shown in Figure 2c and RMSE, MAE can be found in Table S2 in Supporting Information S1). Different PFT models employed various driver data and architectures, leading to slightly different performance and generalization abilities of the trained LSTM model for the NEE predictions. For PFT_LSTM models, the median R^2 ranged from 0.51 to 0.93 for each PFT test set, with RMSE ranging from 0.28 and 1.47 g C m^{-2} d^{-1}, and MAE ranging from 0.20 to 0.95 g C m^{-2} d^{-1}. Without distinguishing PFTs, the R^2 of each PFT test set ranged from 0.13 to 0.82. The RMSE ranged from 0.37 to 1.66 g C m^{-2} d^{-1}, and the MAE ranged from 0.27 to 1.09 g C m^{-2} d^{-1}. Among the nine PFTs, except for WSA, where the R^2 remained the same, the PFT_LSTM models outperformed the nonPFT_LSTM models in predicting NEE. The LSTM model performance was improved in terms of R^2, with an increase ranging from 0.05 to 0.38. The NEE estimation for the MF sites showed the most significant increase in R^2, with an improvement of 0.38. This was followed by CSH, OSH and WET sites, where the increase in median R^2 exceeding 0.10. Therefore, the differentiation of PFTs has improved the ability of the DL models to predict NEE.

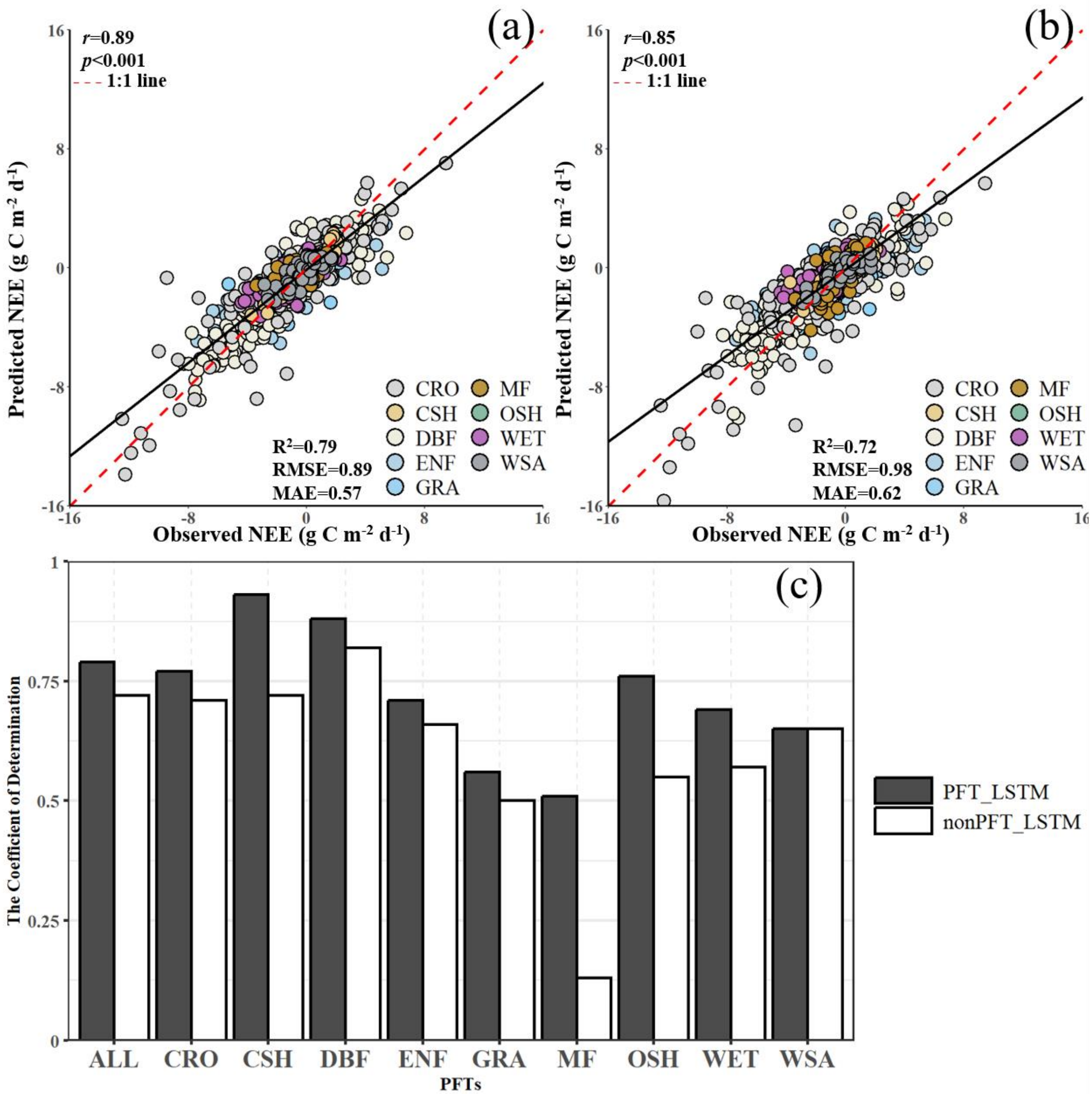

Figure 2. Comparative evaluation of the predicted monthly NEE by the LSTM models against the observed NEE. (a) The PFT_LSTM models for all PFTs; (b) The nonPFT_LSTM models for all PFTs; (c) Model performance comparison over nine different PFTs. The colors of points in (a) and (b) indicate the predominant PFT presented at respective sites (ENF: evergreen needleleaf forest, GRA: grassland, DBF: deciduous broadleaf forest, OSH: open shrubland, CRO: cropland, WET: permanent wetland, CSH: closed shrubland, MF: mixed forest, and WSA: woody savanna). Each data point corresponds to the modeled estimates derived from the median ensemble of the 10 model runs. The black line shows the best-fit line from the least-squares regression. The units of RMSE and MAE are g C m^{-2} d^{-1}.

Overall, the PFT_LSTM models demonstrated a satisfactory performance in predicting monthly NEE across various PFTs (Figure 3). The model performed best for CSH and DBF, with the median R^2 of 0.93 and 0.88, respectively. In contrast, the predictive ability for MF is relatively poor, with the median R^2 close to 0.50. This can be explained by multiple factors, including the limited number of sites (n = 2), limited observation data (only 253 site months of NEE), and limited variation of NEE between these investigated sites. These factors collectively constrain the performance of the LSTM models. The PFT_LSTM models performed relatively well in ENF, DBF, OSH, CRO, WET, CSH, and WSA, with R^2 greater than 0.65. In comparison, the nonPFT_LSTM models performed best at DBF sites, with the R^2 of 0.82, while they performed the worst at the MF sites, with the R^2 of only 0.13.

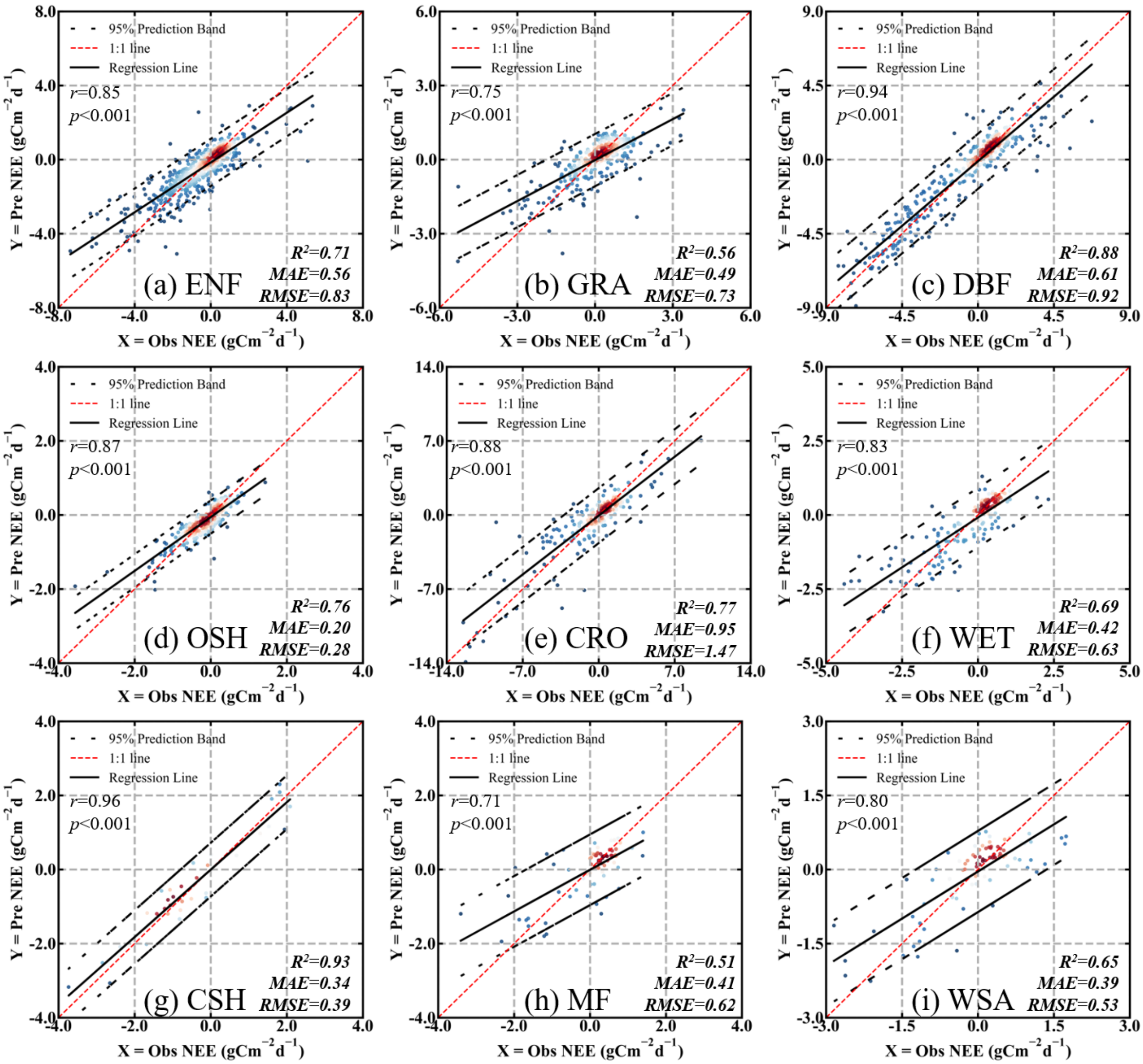

Figure 3. Scatter plots of the predicted NEE by the PFT_LSTM models against the observed NEE across various PFTs. The color in the scatter density thermogram indicates data density. The range covered by the black dashed line is the 95% prediction band of the models. The units of RMSE and MAE are g C m^{-2} d^{-1}.

3.2. Across-Site Variability Estimation

We further evaluated the prediction accuracy of the LSTM models at the site level. In the across-site study, we employed trained PFT_LSTM and nonPFT_LSTM models to generate the simulated data specially for each site during the testing phase. The validation analysis reveals that the PFT_LSTM models demonstrated a good performance in capturing cross-site variability, with R^2 exceeding 0.2 for 74 sites, which accounts for 90.24% of the total amount of tested sites. However, the performance of the PFT_LSTM models was unsatisfactory at 8 sites, e.g., R^2 = 0.06 at the US-KS2 site (Figure S2a in Supporting Information S1). In comparison, for the nonPFT_LSTM models, only 66 sites achieved an R^2 above 0.2, representing 80.49% of the total participated sites. Meanwhile, the nonPFT_LSTM models showed poor prediction at 16 sites, e.g., R^2 = 0.01 at the CA-NS6 site (Figure S2b in Supporting Information S1).

For cross-site validation, we viewed that the model failed to predict NEE changes at that site if the R^2 was lower than 0.2. These failed predictions can be attributed to limited input data, the choice of feature variables, and the limitation of the model design. Uncertain factors may obscure the relationship between the target variable and the predictors. Among the sites that can be predicted by LSTM models, when using PFT_LSTM models for prediction, the median R^2 across sites was greater than 0.65, including CSH, DBF, WET, and ENF sites. Especially, PFT_LSTM models can effectively explain monthly NEE variations at DBF sites, with a median R^2 of 0.92 for site-level predictions, RMSE = 0.66 g C m^{-2} d^{-1}, and MAE = 0.5 g C m^{-2} d^{-1}. In addition, for all PFTs except CRO and GRA, the fitting accuracy of PFT_LSTM models for cross-site monthly NEE was higher than that of nonPFT_LSTM models. We compared the model performance at the sites that can be predicted by both PFT_LSTM models and nonPFT_LSTM models (Figure 4). 43 out of these 64 sites exhibited an improvement in R^2 along with obvious reductions in RMSE and MAE, indicating that 67.19% of the sites have enhanced the performance of LSTM models.

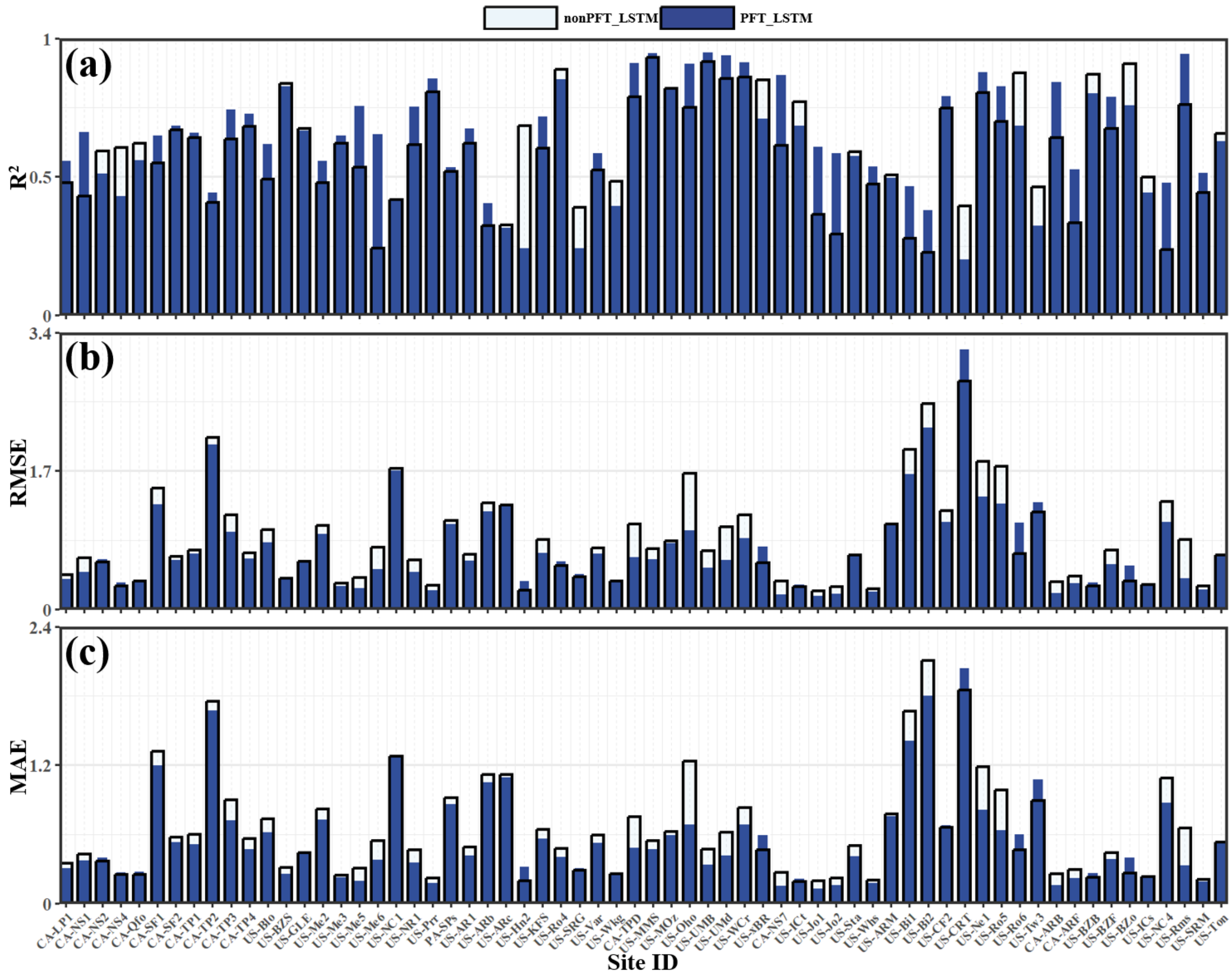

Figure 4. Model performance comparison between the PFT_LSTM models and the nonPFT_LSTM models across different sites. (a) R^2, (b) RMSE and (c) MAE. The unit for RMSE and MAE is g C m^{-2} d^{-1}.

Cross-site validation analyses consistently demonstrated that the DBF and CSH sites exhibit the best predictive capability, which aligns with the performance evaluations at the PTF level. The PFT_LSTM models demonstrated the ability to predict monthly NEE spatiotemporal variability at more than 90% of EC tower sites in North America, with satisfactory performance (i.e., R^2 > 0.6) at over 50% of the sites. Therefore, distinguishing PFTs in predicting terrestrial ecosystem carbon fluxes is pretty crucial. Figure 5 illustrates the time series of the monthly NEE simulations by the PFT_LSTM models at typical sites within each PFT. The PFT_LSTM models can effectively capture the seasonal variations of terrestrial NEE during both the training period and the testing period.

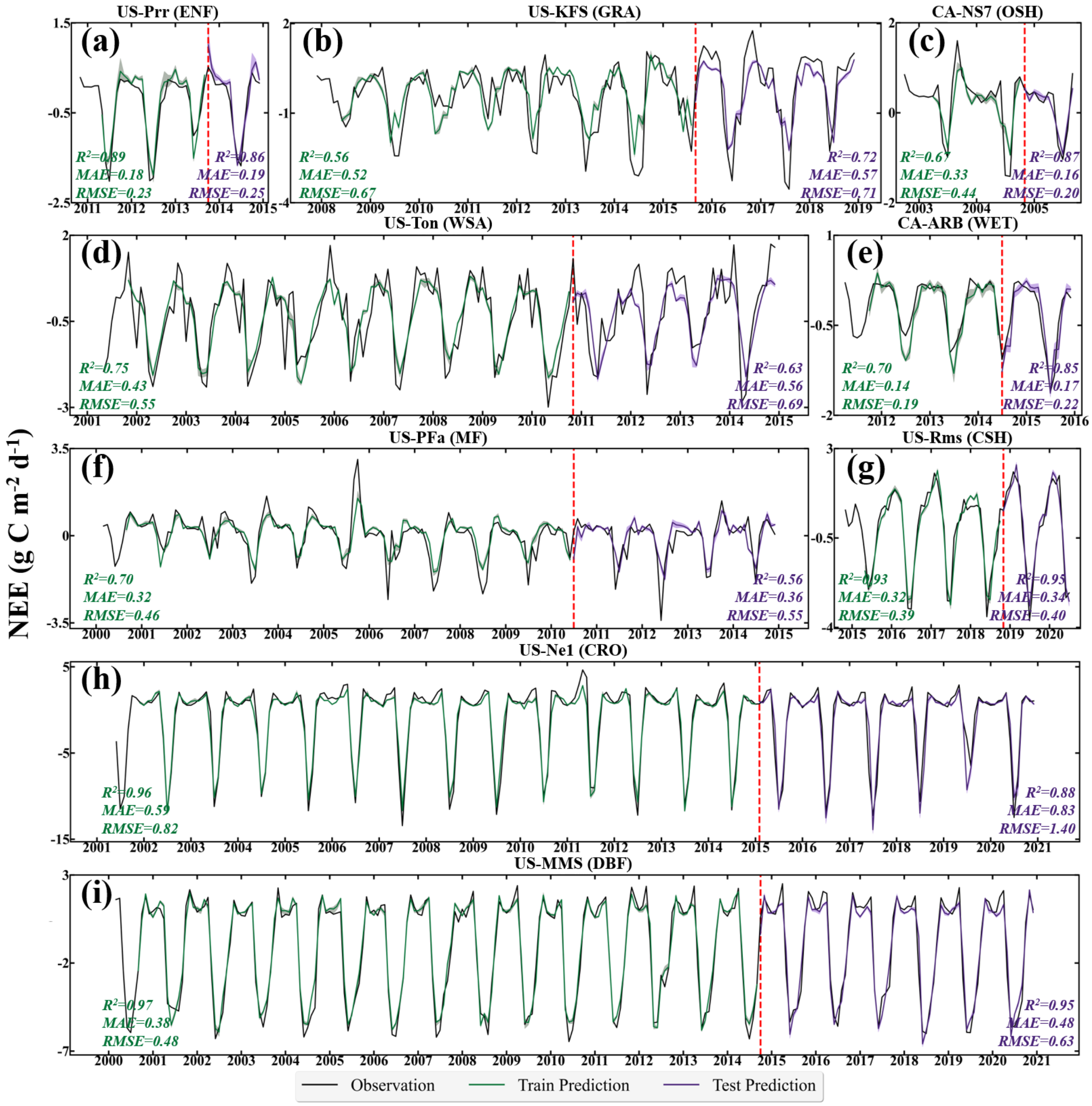

Figure 5. Fitting of the predicted NEE by the PFT_LSTM models against the observed NEE across various representative sites for each PFT. The shaded bands around the lines indicate the uncertainty ranges of the prediction ensemble members. The red dashed line indicates the start of the site testing period. Note that we use the previous 6 months of input data to predict NEE, thus no NEE predictions were made for the initial 6 months. The unit for RMSE and MAE is g C m^{-2} d^{-1}.

3.3. Advantages of LSTM over RF in Predicting NEE and its IAV

Random Forest (RF) is a widely recognized ML algorithm that performs well in handling complex datasets and features by constructing multiple decision trees for prediction (Belgiu & Drăguţ, 2016). It has been successfully employed to predict NEE variability at the site level (Huang et al., 2021), as well as in various endeavors aiming to upscale carbon fluxes to continental or global scales (Kondo et al., 2015; Reitz et al., 2021). RF and LSTM are two representatives ML or DL models with different theoretical and algorithmic implementations. For each PFT, we established an RF model and labeled them as PFT_RF Model. During the training phase of the RF models, we employed a tenfold cross-validation process to conduct a grid search to determine the optimal parameter set. Subsequently, a prediction model was built using the training data, and its performance was evaluated using test data. Similar to the LSTM model, each PFT model run 10 times, and the median estimate from these results was considered the best prediction.

The performance on predicting monthly NEE was compared between the RF model and the LSTM model (R^2 shown in Figure 6 and RMSE, MAE can be found in Table S2 in Supporting Information S1). It showed that the RF models provided monthly NEE estimations with R^2=0.59, RMSE=1.19 g C m^{-2} d^{-1}, and MAE=0.70 g C m^{-2} d^{-1}. PFT_RF models also exhibited acceptable performance in predicting NEE for 9 PFTs, with R^2 ranging from 0.1 to 0.8. Both the RF and LSTM models displayed consistent predictive abilities, with the best performance observed for the PFTs of CSH and DBF, with the R^2 of 0.80 and 0.73, respectively. Notably, the RF model performed poorly for the PFT of WET, with the R^2 value close to 0.1. The PFT_LSTM models demonstrated better predictive ability compared to the RF models across all PFTs.

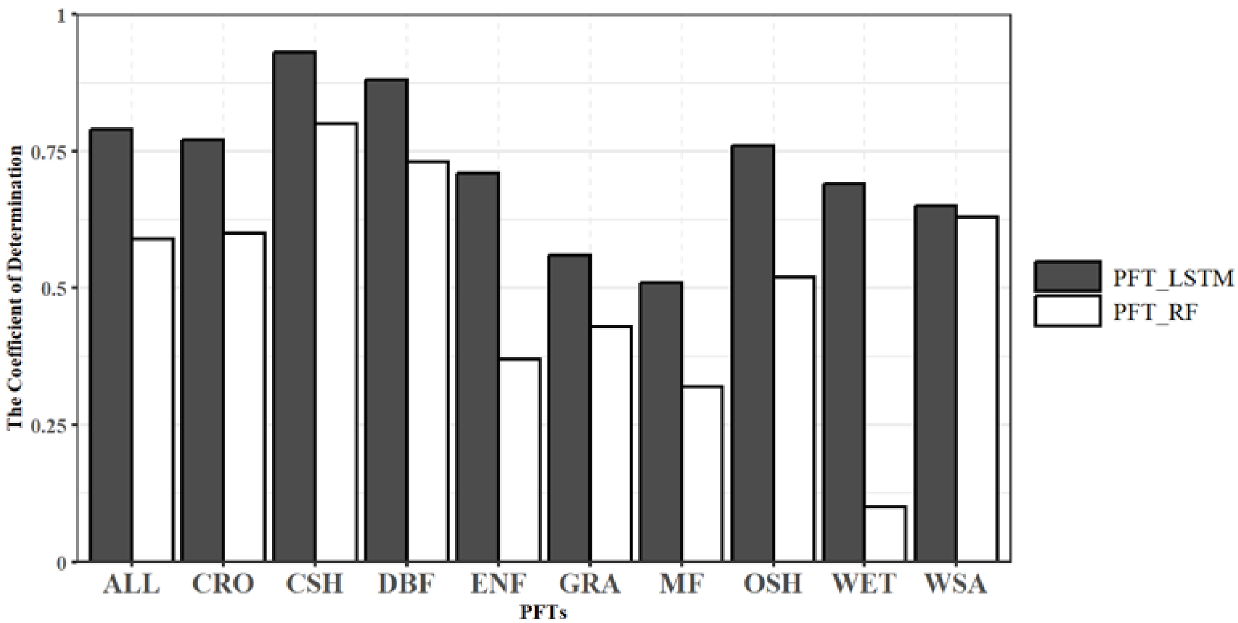

Figure 6. Model performance comparison between the LSTM model and the RF model over nine PFTs.

We further made comparative evaluations on the predictive capacities of NEE IAV by the PFT_LSTM models and PFT_RF models against flux tower observations over various PFTs. For each EC tower, we filtered the data to ensure only NEE data with sufficient 12 months per year is included, and calculated the annual NEE (g C m^{-2} yr^{-1}) from these monthly NEE data. This allowed us to obtain both annual NEE observations and model predictions. The evaluation was based on the analysis of annual anomalies. NEE IAV was calculated as the difference between yearly NEE during the testing period and the average value over the entire observation period. The results indicate that PFT_LSTM models reasonably captured the IAV of NEE in North America, showing a significant positive correlation between the observations and the model predictions in the IAV of NEE (r = 0.81, p < 0.001). By contrast, the PFT_RF models generally failed to predict the NEE IAV, with a low correlation coefficient r = -0.21 (p = 0.011). Figure 7 shows the combined time series of NEE IAV predicted by the LSTM and RF models by PFTs. The LSTM models performed relatively well in representing the IAV of NEE for ENF, WET, MF, and WSA, with the r exceeding 0.75, among which the WET performed the best (r = 0.970, p < 0.001). The CRO had strong IAV of NEE, ranging from -783.52 g C m^{-2} yr^{-1} to 271.93 g C m^{-2} yr^{-1}. Consequently, LSTM models predicted NEE IAV poorly for CRO, with r = 0.330 (p = 0.250). In terms of indicating NEE IAV, the RF models performed clearly less effectively compared to the LSTM models at the PFT level.

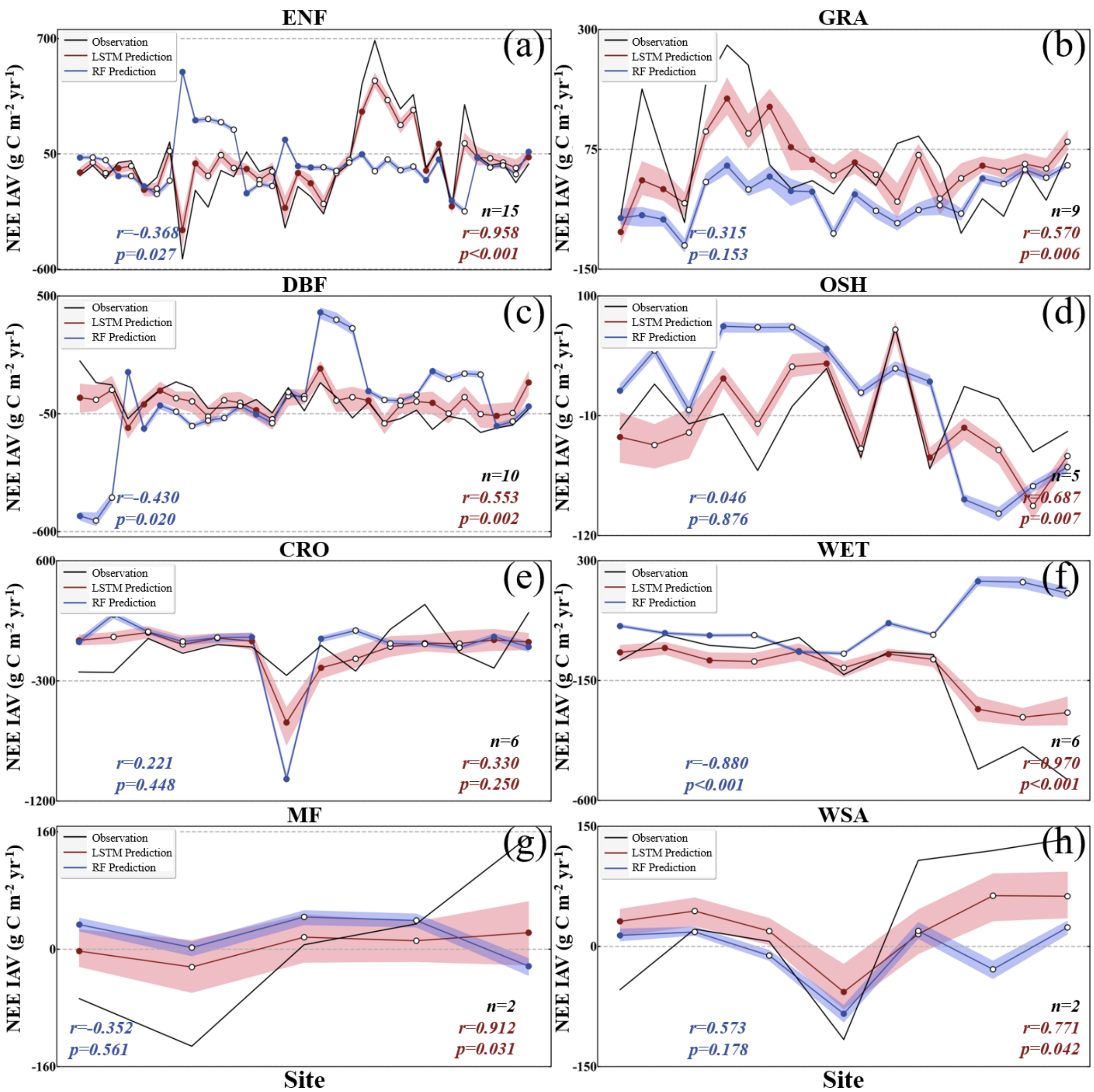

Figure 7. Evaluation of combined time series of NEE IAV predicted by the LSTM and RF models against flux tower observations at the PFT level. The shaded bands around the lines indicate the uncertainty ranges of the prediction ensemble members. Each scatter represents a site-year, while a solid scatter represents the start of the site testing period. *n* represents the number of sites available for estimating NEE IAVs across each PFT. Note that the IAV predictions for CSH are not included in the plots due to the limited availability of complete observational data; only two full years of observations were available for CSH during the testing period.

Furthermore, we investigated the performance of the LSTM and RF models to predict the IAV of NEE at the site level (Figure 8). These selected sites typically cover at least four complete years of flux observations during the testing period, ensuring sufficient data for conducting analysis. The LSTM models exhibited a much stronger correlation with the observed NEE IAVs at these sites than the RF models did, e.g., at US-GLE and US-Me2.

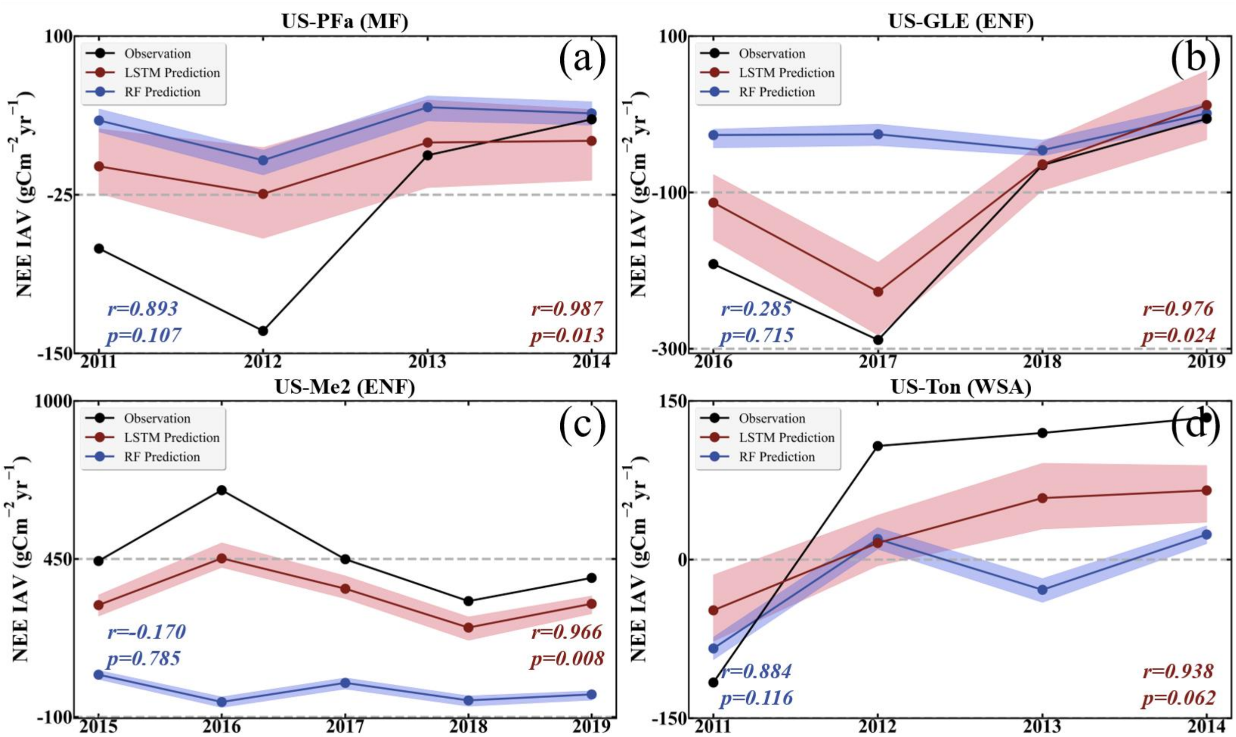

Figure 8. Evaluation of time series of NEE IAV predicted by the LSTM and RF models against flux tower observations at the site level. The shaded bands around the lines indicate the uncertainty ranges of the prediction ensemble members.

3.4. Relative Contributions of Environmental Controls to Monthly NEE Variations Across PFTs

We quantified the importance of the 10 predictive variables (WS, VPD, TA, SWC, DSR, P, NDVI, LAI, SIF, and FAPAR) on predicting monthly NEE using the BRT model (Figure 9). Since different PFTs exhibit distinct responses to NEE, individual BRT models were established for the 9 PFTs. Overall, among all PFTs, SIF was the most powerful predictor for monthly NEE variability, with an average contribution of 26.32%, followed by DSR and LAI. The combined contributions of SIF, DSR, and LAI to monthly NEE variability accounted for approximately 52.02%. Compared with SIF, DSR, and LAI, other variables showed much weaker controls over monthly NEE, with an average contribution of less than 10%. This analysis is in line with the findings of a previous study (Kong et al., 2022), which identified DSR and LAI as the primary environmental controls of daily NEE changes for most PFTs, while the contributions of TA, SWC, and other variables were relatively small. It is worth noting that Kong et al.'s study did not include the SIF variable.

The relative importance of predictive variables in driving NEE changes diverged among PFTs. For most PFTs, including GRA, DBF, OSH, WET, CRO, MF, and WSA, SIF is the most powerful predictor for monthly NEE variability. Particularly, for DBF, WET, and WSA, SIF contributed for more than 30% of monthly NEE variability. For WSA, SIF even contributed for more than 50% of monthly NEE variations. In contrast, SIF played a much weaker role for ENF and CSH, where DSR emerged as the most critical predictor for ENF, and LAI additionally made a remarkable contribution for CSH.

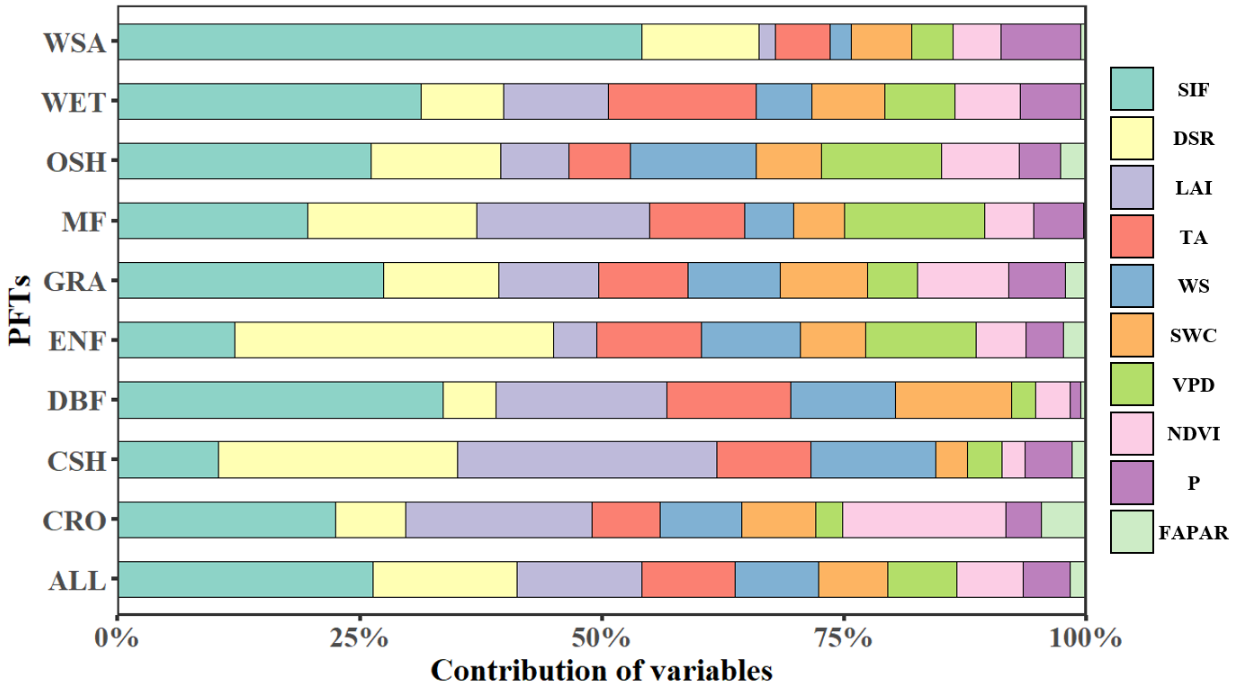

Figure 9. The relative contributions of downward shortwave radiation (DSR), solar-induced chlorophyll fluorescence (SIF), air temperature (TA), normalized difference vegetation index (NDVI), leaf area index (LAI), wind speed (WS), vapor pressure deficit (VPD), soil water content (SWC), precipitation (P) and fraction of absorbed photosynthetically active radiation (FAPAR) to monthly net ecosystem exchange of CO_2 from the BRT method across all North American vegetation types.

4 Discussion

4.1. Advantages of LSTM in Predicting NEE and Its IAV

The proposed LSTM models in this study exhibited a satisfactory performance in predicting the temporal dynamics and cross-site variability of monthly NEE, and clearly superior performance over the traditional ML models (e.g., RF investigated in this study) that doesn’t consider temporal memory effect.

Extreme climate events and disturbances can affect the development, structure, and function of terrestrial ecosystems(S. Liu et al., 2011; Williams et al., 2012). Due to the complex carbon cycling process between terrestrial ecosystems and the atmosphere, these impacts typically persist for a long period (Frank et al., 2015). This memory effect results in a delayed IAV in the growth rate of atmospheric CO_2 concentration, which hinders the accurate prediction of long-term changes in the terrestrial carbon budget under climate change and human influence. Capturing this impact on NEE was quite challenging in a long past period. The widely used ML-based NEE dataset (FLUXCOM NEE) is an upscaling of remote sensing data and meteorological reanalysis data to the flux towers data. However, FLUXCOM fails to accurately reproduce the long-term trends and IAV of NEE (Piao et al., 2020), which could largely associate with characterizing memory effect. The memory effect is important factor in controlling the IAV of NEE (Bloom et al., 2020), but has not been considered in FLUXCOM. Recent researches demonstrated the potential of considering memory effects in improving terrestrial carbon flux simulations (Besnard et al., 2019; Liu et al., 2023).

In terms of representing NEE IAV, the LSTM DL algorithm outperforms the widely used RF ML model when using the same input data and model configuration. The LSTM algorithm is able to dynamically incorporate temporal information into the estimation of CO_2 fluxes, allowing for the characterization of the memory effect caused by disturbances and climate change on NEE. The strong long-term dependency modeling abilities of LSTM make it suitable for characterizing memory effect relationships in sequence data, leading to more realistic estimations of NEE dynamics (Schmidhuber, 2015). By capturing the memory effects of climate and vegetation, we can enhance our understanding and predictive ability of regional C budgets. We anticipate that PFT_LSTM models will deliver enhanced performance in future carbon flux upscaling research.

4.2. Uncertainties and Prospects

While LSTM can generally fit monthly NEE from North American sites well, avoiding prediction bias is also challenging due to imbalanced input sampling caused by spatial and temporal differences in NEE data. In the model training process, the model is more frequently exposed to sites with spatial correlation and longer observation data, enabling better learning of the NEE variability of these specific sites (He et al., 2015). However, if certain spatiotemporal changes in the training samples are not adequately represented, the model may not accurately predict or adapt to those changes, resulting in significant bias and uncertainty. Although the LSTM model was specifically designed for PFT, there are still 8 sites scattered across three PFTs in North America that are not predictable by LSTM (GRA; n = 5, DBF; n = 2, and CSH; n = 1).

Furthermore, despite PFT_LSTM models performed better than nonPFT_LSTM models for the PFT level of CRO and GRA, their performances were slightly weaker than nonPFT_LSTM models in predicting across-site variability, with the median R^2 reduced by 0.01 for CRO and 0.09 for GRA, respectively. This is attributed to the challenge of identifying similar trends, patterns, or relationships between CRO and GRA sites, along with the substantial influence of human management on croplands (Marcolla et al., 2017). Thus, the integration of more observations could be crucial for ML algorithms to accurately capture monthly NEE changes at CRO and GRA sites. The Unbalanced sampling leads to a lack of representativeness in the data, which may be the primary factor contributing to uncertainty in the NEE simulation for these sites. Adding more observations is crucial for improving the ability to fit the NEE variability at these sites.

We notice that the time length of memory effect could influence the modeling power for different PFTs and sites. In this study, we used 6 months for all PFTs. In fact, for different PFTs, the memory length could be different (Aubinet et al., 2018; Zhang et al., 2022). In the study by Liu et al., (2023), they found the optimal memory effect lengths diverged across PFTs. From their study, 6 months are proper for most PFTs. In the future, in order to achieve a more reliable upscaling, designing different memory effect lengths for different PFTs could be helpful.

With LSTM models, the NEE fluxes in the context of global climate change are expected to be more accurately predicted. Using LSTM DL algorithms to upscale carbon estimation at the regional, continental, and even global scales would emerge as a popular approach for future NEE modeling. The investment and use of more flux towers will provide a substantial volume of high-quality continuous observation data for future studies. Enhancing the representativeness of flux tower data and the availability of predictive variables is an important undertaking in accurately estimating future carbon fluxes.

5 Conclusions

This study explored the potential of LSTM models in predicting monthly NEE over 82 sites in North America based on FLUXNET 2015 and the AmeriFlux datasets and multiple satellite land surface products. After distinguishing PFTs, the overall R^2 of monthly NEE increased by 9.72%, RMSE decreased by 0.09 g C m^{-2} d^{-1}, and MAE decreased by 0.05 g C m^{-2} d^{-1}. The model performance of each PFT has been improved, highlighting the importance of differentiating PFTs during model training. The use of time series data as model inputs allows the LSTM algorithm to effectively capture the memory effect of climate and environmental factors on the time scale. A significant positive correlation exists between the observation of NEE IAV and the model prediction results (r = 0.81, p < 0.001). While commonly used non-temporal dynamic statistical RF ML algorithms demonstrate acceptable performance in predicting monthly NEE, their ability to predict IAV is very poor. Among the selected predictive variables, SIF exhibits the strongest correlation with monthly NEE changes, contributing an average of 26.32%, followed by DSR (14.83%) and LAI (12.87%). Including these variables into ML or DL models is critical for predicting monthly NEE. Overall, the combination of LSTM and PFTs classification shows potential in predicting the temporal variability of NEE and correcting for NEE IAVs. This study provides a reference for modeling terrestrial carbon cycle, especially for upscaling in-situ carbon flux observations to larger scales.

Acknowledgments

This research is funded by the National Natural Science Foundation of China (Grant No. 42277453 and 41907378). We acknowledge FLUXNET and AmeriFlux for providing the flux tower data. We sincerely acknowledge Jingfeng Xiao from New Hampshire University for providing the GOSIF data. We sincerely thank Zaichun Zhu from Peking University for sharing the PKU GIMMIS NDVI dataset. We also sincerely acknowledge Shunlin Liang’s group for sharing GLASS LAI and FAPAR data.

Data Availability Statement

The data sets used in this paper are available from open resources. Eddy covariance data for the 35 FLUXNET sites utilized in this study are available from the FLUXNET2015 data set (Pastorello et al., 2020). Eddy covariance data for the remaining 47 sites are attained from the AmeriFlux website (Novick et al., 2018). The GLASS LAI and FAPAR products are available on the official website for the GLASS project (Liang et al., 2021). The PKU GIMMS NDVI data is publicly available in the Zenodo repository (Li et al., 2023b). GOSIF is available in the Global Ecology Group's data repository (Li & Xiao, 2023).

References

Aubinet, M., Hurdebise, Q., Chopin, H., Debacq, A., De Ligne, A., Heinesch, B., et al. (2018). Inter-annual variability of Net Ecosystem Productivity for a temperate mixed forest: A predominance of carry-over effects? *Agricultural and Forest Meteorology*, *262*, 340–353. https://doi.org/10.1016/j.agrformet.2018.07.024

Baldi, P., & Sadowski, P. (2014). The dropout learning algorithm. *Artificial Intelligence*, *210*, 78–122. https://doi.org/10.1016/j.artint.2014.02.004

Baldocchi, D. D. (2020). How eddy covariance flux measurements have contributed to our understanding of *Global Change Biology*. *Global Change Biology*, *26*(1), 242–260. https://doi.org/10.1111/gcb.14807

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of machine learning research*, 13(2). https://dl.acm.org/doi/pdf/10.5555/2188385.2188395

Besnard, S., Carvalhais, N., Arain, M. A., Black, A., Brede, B., Buchmann, N., et al. (2019). Memory effects of climate and vegetation affecting net ecosystem CO_2 fluxes in global forests. *PLOS ONE*, *14*(2), e0211510. https://doi.org/10.1371/journal.pone.0211510

Bloom, A. A., Bowman, K. W., Liu, J., Konings, A. G., Worden, J. R., Parazoo, N. C., et al. (2020). Lagged effects regulate the inter-annual variability of the tropical carbon balance. *Biogeosciences*, *17*(24), 6393–6422. https://doi.org/10.5194/bg-17-6393-2020

Bodesheim, P., Jung, M., Gans, F., Mahecha, M. D., & Reichstein, M. (2018). Upscaled diurnal cycles of land–atmosphere fluxes: a new global half-hourly data product.

Bonan, G. B. (2008). Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science*, *320*(5882), 1444–1449. https://doi.org/10.1126/science.1155121

Ciais, P., Dolman, A. J., Bombelli, A., Duren, R., Peregon, A., Rayner, P. J., et al. (2014). Current systematic carbon-cycle observations and the need for implementing a policy-relevant carbon observing system. *Biogeosciences*, *11*(13), 3547–3602. https://doi.org/10.5194/bg-11-3547-2014

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Fang, H., Jiang, C., Li, W., Wei, S., Baret, F., Chen, J. M., et al. (2013). Characterization and intercomparison of global moderate resolution leaf area index (LAI) products: Analysis of climatologies and theoretical uncertainties. *Journal of Geophysical Research: Biogeosciences*, *118*(2), 529–548. https://doi.org/10.1002/jgrg.20051

Frank, D., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M. D., et al. (2015). Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts. *Global Change Biology*, *21*(8), 2861–2880. https://doi.org/10.1111/gcb.12916

Gower, S. T., Kucharik, C. J., & Norman, J. M. (1999). Direct and Indirect Estimation of Leaf Area Index, fAPAR, and Net Primary Production of Terrestrial Ecosystems. *Remote Sensing of Environment*, *70*(1), 29–51. https://doi.org/10.1016/S0034-4257(99)00056-5

Guo, R., Chen, T., Chen, X., Yuan, W., Liu, S., He, B., et al. (2023). Estimating Global GPP From the Plant Functional Type Perspective Using a Machine Learning Approach. *Journal of Geophysical Research: Biogeosciences*, *128*(4), e2022JG007100. https://doi.org/10.1029/2022JG007100

He, H., Zhang, L., Gao, Y., Ren, X., Zhang, L., Yu, G., & Wang, S. (2015). Regional representativeness assessment and improvement of eddy flux observations in China. *Science of The Total Environment*, *502*, 688–698. https://doi.org/10.1016/j.scitotenv.2014.09.073

He, W., Jiang, F., Ju, W., Byrne, B., Xiao, J., Nguyen, N. T., et al. (2023a). Do State-Of-The-Art Atmospheric CO_2 Inverse Models Capture Drought Impacts on the European Land Carbon Uptake? *Journal of Advances in Modeling Earth Systems*, *15*(6), e2022MS003150. https://doi.org/10.1029/2022MS003150

He, W., Jiang, F., Ju, W., Chevallier, F., Baker, D. F., Wang, J., et al. (2023b). Improved Constraints on the Recent Terrestrial Carbon Sink Over China by Assimilating OCO-2 XCO_2 Retrievals. *Journal of Geophysical Research: Atmospheres*, *128*(14), e2022JD037773. https://doi.org/10.1029/2022JD037773

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, N., Wang, L., Zhang, Y., Gao, S., & Niu, Z. (2021). Estimating the Net Ecosystem Exchange at Global FLUXNET Sites Using a Random Forest Model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 9826–9836. https://doi.org/10.1109/JSTARS.2021.3114190

Huntzinger, D. N., Post, W. M., Wei, Y., Michalak, A. M., West, T. O., Jacobson, A. R., et al. (2012). North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison. *Ecological Modelling*, *232*, 144–157. https://doi.org/10.1016/j.ecolmodel.2012.02.004

Irrgang, C., Boers, N., Sonnewald, M., Barnes, E. A., Kadow, C., Staneva, J., & Saynisch-Wagner, J. (2021). Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nature Machine Intelligence*, *3*(8), 667–674. https://doi.org/10.1038/s42256-021-00374-3

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., et al. (2011). 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*, *116*, G00J07. https://doi.org/10.1029/2010JG001566

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., et al. (2020). Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. *Biogeosciences*, *17*(5), 1343–1365. https://doi.org/10.5194/bg-17-1343-2020

Kingma, D. P., & Ba, J. (2017, January 29). Adam: A Method for Stochastic Optimization. arXiv. Retrieved from http://arxiv.org/abs/1412.6980

Kondo, M., Ichii, K., Takagi, H., & Sasakawa, M. (2015). Comparison of the data-driven top-down and bottom-up global terrestrial CO_2 exchanges: GOSAT CO_2 inversion and empirical eddy flux upscaling. *Journal of Geophysical Research: Biogeosciences*, *120*(7), 1226–1245. https://doi.org/10.1002/2014JG002866

Kong, Z., Wang, T., Han, Q., Dai, Y., Wang, L., & Chen, X. (2022). Evaluation of Environmental Controls on Terrestrial Net Ecosystem Exchange of CO_2: A Global Perspective From the FLUXNET Sites. *Journal of Geophysical Research: Atmospheres*, *127*(22). https://doi.org/10.1029/2022JD037217

Li, M., Cao, S., & Zhu, Z. (2023a). Spatiotemporally consistent global dataset of the GIMMS Normalized Difference Vegetation Index (PKU GIMMS NDVI) from 1982 to 2020. *Earth System Science Data*, *15(9): 4181–4203*. https://doi.org/10.5194/essd-2023-1

Li, M., Cao, S., Zhu, Z., Wang, Z., Myneni, R. B., & Piao, S. (2023b). Spatiotemporally consistent global dataset of the GIMMS Normalized Difference Vegetation Index (PKU GIMMS NDVI) from 1982 to 2022 (V1.2) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.8253971

Li, Xiangyi, Piao, S., Wang, K., Wang, X., Wang, T., Ciais, P., et al. (2020). Temporal trade-off between gymnosperm resistance and resilience increases forest sensitivity to extreme drought. *Nature Ecology & Evolution*, *4*(8), 1075–1083. https://doi.org/10.1038/s41559-020-1217-3

Li, Xing, & Xiao, J. (2019). A Global, 0.05-Degree Product of Solar-Induced Chlorophyll Fluorescence Derived from OCO-2, MODIS, and Reanalysis Data. *Remote Sensing*, *11*(5), 517. https://doi.org/10.3390/rs11050517

Li, Xing, & Xiao, J. (2023). GOSIF: A Global, 0.05-Degree Product of Solar-Induced Chlorophyll Fluorescence Derived from OCO-2, MODIS, and Reanalysis Data [Dataset]. Retrieved from http://data.globalecology.unh.edu/data/GOSIF_v2/

Liang, S., Zhao, X., Liu, S., Yuan, W., Cheng, X., Xiao, Z., et al. (2013). A long-term Global LAnd Surface Satellite (GLASS) data-set for environmental studies. *International Journal of Digital Earth*, *6*(sup1), 5–33. https://doi.org/10.1080/17538947.2013.805262

Liang, S., Cheng, J., Jia, K., Jiang, B., Liu, Q., Xiao, Z., et al. (2021). The Global Land Surface Satellite (GLASS) Product Suite: Leaf area index (LAI) and Fraction of absorbed photosynthetically active radiation (FAPAR) (version V60) [Dataset]. Retrieved from http://www.glass.umd.edu/Download.html

Liang, W., Zhang, W., Jin, Z., Yan, J., Lü, Y., Wang, S., et al. (2020). Estimation of Global Grassland Net Ecosystem Carbon Exchange Using a Model Tree Ensemble Approach. *Journal of Geophysical Research: Biogeosciences*, *125*(1). https://doi.org/10.1029/2019JG005034

Liu, S., Bond-Lamberty, B., Hicke, J. A., Vargas, R., Zhao, S., Chen, J., et al. (2011). Simulating the impacts of disturbances on forest carbon cycling in North America: Processes, data, models, and challenges. *Journal of Geophysical Research*, *116*, G00K08. https://doi.org/10.1029/2010JG001585

Liu, W., He, H., Wu, X., Ren, X., Zhang, L., Shi, L., et al. (2023). Importance of the memory effect for assessing interannual variation in net ecosystem exchange. *Agricultural and Forest Meteorology*, *341*, 109691. https://doi.org/10.1016/j.agrformet.2023.109691

Marcolla, B., Rödenbeck, C., & Cescatti, A. (2017). Patterns and controls of inter-annual variability in the terrestrial carbon budget. *Biogeosciences*, *14*(16), 3815–3829. https://doi.org/10.5194/bg-14-3815-2017

Novick, K. A., Biederman, J. A., Desai, A. R., Litvak, M. E., Moore, D. J. P., Scott, R. L., & Torn, M. S. (2018). The AmeriFlux network: A coalition of the willing. *Agricultural and Forest Meteorology*, *249*, 444–456. https://doi.org/10.1016/j.agrformet.2017.10.009

Novick, K. A., Biederman, J. A., Desai, A. R., Litvak, M. E., Moore, D. J. P., Scott, R. L., & Torn, M. S. (2018). The AmeriFlux network: A coalition of the willing. *Agricultural and Forest Meteorology*, *249*, 444–456. https://doi.org/10.1016/j.agrformet.2017.10.009

Ogle, K., Barber, J. J., Barron-Gafford, G. A., Bentley, L. P., Young, J. M., Huxman, T. E., et al. (2015). Quantifying ecological memory in plant and ecosystem processes. *Ecology Letters*, *18*(3), 221–235. https://doi.org/10.1111/ele.12399

Papale, D., & Valentini, R. (2003). A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization: A new assessment of European forests carbon. *Global Change Biology*, *9*(4), 525–535. https://doi.org/10.1046/j.1365-2486.2003.00609.x

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y. W., et al. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data*, *7*(1), 1–27. https://doi.org/10.1038/s41597-020-0534-3

Perez-Suay, A., Adsuara, J. E., Piles, M., Martinez-Ferrer, L., Diaz, E., Moreno-Martinez, A., & Camps-Valls, G. (2020). Interpretability of Recurrent Neural Networks in Remote Sensing. In *IGARSS 2020 - 2020*

IEEE International Geoscience and Remote Sensing Symposium (pp. 3991–3994). Waikoloa, HI, USA: IEEE. https://doi.org/10.1109/IGARSS39084.2020.9323898

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., et al. (2013). *Global atmospheric carbon budget: results from an ensemble of atmospheric CO₂ inversions* (preprint). Biogeochemistry: Greenhouse Gases. https://doi.org/10.5194/bgd-10-5301-2013

Piao, S., Wang, X., Wang, K., Li, X., Bastos, A., Canadell, J. G., et al. (2020). Interannual variation of terrestrial carbon cycle: Issues and perspectives. *Global Change Biology*, *26*(1), 300–318. https://doi.org/10.1111/gcb.14884

Reichstein, M., Besnard, S., Carvalhais, N., Gans, F., Jung, M., Kraft, B., & Mahecha, M. (2018). Modelling Landsurface Time-Series with Recurrent Neural Nets. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 7640–7643). Valencia: IEEE. https://doi.org/10.1109/IGARSS.2018.8518007

Reitz, O., Graf, A., Schmidt, M., Ketzler, G., & Leuchner, M. (2021). Upscaling Net Ecosystem Exchange Over Heterogeneous Landscapes With Machine Learning. *Journal of Geophysical Research: Biogeosciences*, *126*(2). https://doi.org/10.1029/2020JG005814

Rumelhart, D. E., Hintont, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors.

Rußwurm, M., & Körner, M. (2018). Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. *ISPRS International Journal of Geo-Information*, *7*(4), 129. https://doi.org/10.3390/ijgi7040129

Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Shevliakova, E., Stouffer, R. J., Malyshev, S., Krasting, J. P., Hurtt, G. C., & Pacala, S. W. (2013). Historical warming reduced due to enhanced land carbon uptake. *Proceedings of the National Academy of Sciences*, *110*(42), 16730–16735. https://doi.org/10.1073/pnas.1314047110

Shiga, Y. P., Tadić, J. M., Qiu, X., Yadav, V., Andrews, A. E., Berry, J. A., & Michalak, A. M. (2018). Atmospheric CO_2 Observations Reveal Strong Correlation Between Regional Net Biospheric Carbon Uptake and Solar-Induced Chlorophyll Fluorescence. *Geophysical Research Letters*, *45*(2), 1122–1132. https://doi.org/10.1002/2017GL076630

Stoy, P. C., Richardson, A. D., Baldocchi, D. D., Katul, G. G., Stanovick, J., Mahecha, M. D., et al. (2009). Biosphere-atmosphere exchange of CO_2 in relation to climate: a cross-biome analysis across multiple time scales.

Thireou, T., & Reczko, M. (2007). Bidirectional Long Short-Term Memory Networks for Predicting the Subcellular Localization of Eukaryotic Proteins. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *4*(3), 441–446. https://doi.org/10.1109/tcbb.2007.1015

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., et al. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites withregression algorithms. *Biogeosciences*, *13*(14), 4291–4313. https://doi.org/10.5194/bg-13-4291-2016

Ukkola, A. M., Abramowitz, G., & De Kauwe, M. G. (2021). *A flux tower dataset tailored for land model evaluation* (preprint). Data, Algorithms, and Models. https://doi.org/10.5194/essd-2021-181

Verrelst, J., Van Der Tol, C., Magnani, F., Sabater, N., Rivera, J. P., Mohammed, G., & Moreno, J. (2016). Evaluating the predictive power of sun-induced chlorophyll fluorescence to estimate net photosynthesis of vegetation canopies: A SCOPE modeling study. *Remote Sensing of Environment*, *176*, 139–151. https://doi.org/10.1016/j.rse.2016.01.018

Williams, C. A., Collatz, G. J., Masek, J., & Goward, S. N. (2012). Carbon consequences of forest disturbance and recovery across the conterminous United States. *Global Biogeochemical Cycles*, *26*(1), 2010GB003947. https://doi.org/10.1029/2010GB003947

Xu, X., Du, H., Fan, W., Hu, J., Mao, F., & Dong, H. (2019). Long-term trend in vegetation gross primary production, phenology and their relationships inferred from the FLUXNET data. *Journal of Environmental Management*, *246*, 605–616. https://doi.org/10.1016/j.jenvman.2019.06.023

Yin, G., Verger, A., Descals, A., Filella, I., & Peñuelas, J. (2022). A Broadband Green-Red Vegetation Index for Monitoring Gross Primary Production Phenology. *Journal of Remote Sensing*, *2022*, 2022/9764982. https://doi.org/10.34133/2022/9764982

Zhang, Z., Ju, W., Zhou, Y., & Li, X. (2022). Revisiting the cumulative effects of drought on global gross primary productivity based on new long-term series data (1982–2018). *Global Change Biology*, *28*(11), 3620–3635. https://doi.org/10.1111/gcb.16178