Improving net ecosystem CO_2 flux prediction using memory-based interpretable machine learning

1st Siyan Liu
Computational Sciences and Engineering Division
Oak Ridge National Laboratory
Oak Ridge, TN USA
lius1@ornl.gov

2nd Dan Lu
Computational Sciences and Engineering Division
Oak Ridge National Laboratory
Oak Ridge, TN USA
lud1@ornl.gov

3rd Daniel Ricciuto
Environmental Sciences Division
Oak Ridge National Laboratory
Oak Ridge, TN USA
ricciutodm@ornl.gov

4th Anthony Walker
Environmental Sciences Division
Oak Ridge National Laboratory
Oak Ridge, TN USA
walkerap@ornl.gov

***Abstract*—Terrestrial ecosystems play a central role in the global carbon cycle and affect climate change. However, our predictive understanding of these systems is still limited due to their complexity and uncertainty about how key drivers and their legacy effects influence carbon fluxes. Here, we propose an interpretable Long Short-Term Memory (iLSTM) network for predicting net ecosystem CO_2 exchange (NEE) and interpreting the influence on the NEE prediction from environmental drivers and their memory effects. We consider five drivers and apply the method to three forest sites in the United States. Besides performing the prediction in each site, we also conduct transfer learning by using the iLSTM model trained in one site to predict at other sites. Results show that the iLSTM model produces good NEE predictions for all three sites and, more importantly, it provides reasonable interpretations on the input driver's importance as well as their temporal importance on the NEE prediction. Additionally, the iLSTM model demonstrates good across-site transferability in terms of both prediction accuracy and interpretability. The transferability can improve the NEE prediction in unobserved forest sites, and the interpretability advances our predictive understanding and guides process-based model development.**

***Index Terms*—Long Short-Term Memory networks, Interpretable machine learning, net ecosystem CO_2 exchange, driver importance, memory effects**

I. Introduction and motivation

Terrestrial ecosystems play a key role in the global carbon cycle and affect climate change. Accurate prediction of net CO_2 exchange of terrestrial ecosystems (NEE) is important to advance our understanding of terrestrial carbon dynamics and the net CO_2 balance of the ecosystem [1]. Traditionally, process-based terrestrial ecosystem models are mainly used to simulate the land carbon cycle. In recent years, data-driven machine learning (ML) models, such as gradient boosting and random forest methods [2], fully-connected artificial neural networks [3], [4], recurrent neural networks [4], and Bayesian Neural Networks [5] have increasingly been applied with successful results. Among these ML models, Long Short-Term Memory (LSTM) networks [6] have often demonstrated superior performance in carbon flux prediction against other methods [4], [7]–[9].

LSTM networks, trained over multi-variable time series consisting of exogenous and target variables, capture nonlinear correlations of historical values of environmental drivers and carbon fluxes to predict future carbon fluxes. Because of their data-driven nature that learning system patterns and dynamical behaviors from observations, LSTM models have been criticized for their lack of interpretability, despite their successful applications in terrestrial ecosystem modeling [10], [11]. An interpretable ML model should be able to explain, for a carbon flux variable, what are the most important environmental drivers, and at which time scales do the environmental drivers have a strong impact on the carbon flux estimation? Such an interpretation is crucial for improving carbon flux predictions and understanding of the terrestrial ecosystems. Essentially, an interpretable ML model opens the black box of the data-driven models and enables researchers to understand what the model learns. Additionally, it can help researchers check whether the accurate model prediction is resulted from that the model actually captures the physically reasonable input-output relationship, or simply because it memorizes the input-output correlations from the training data. Thus, after we have a good understanding of the learning process and verify that the ML model identified driver importance and temporal importance are consistent with our domain knowledge, we will be more confident in applying the model to the future projection and to unobserved sites for prediction.

However, developing an interpretable ML model is not easy, and it is more difficult to ensure the method maintains high prediction accuracy while demonstrating interpretability. Some studies developed physics-informed LSTM models to provide physical constraints and explanations, but their predictions may be less accurate than the purely data-driven models [12], [13]. Some studies performed sensitivity or permutation analysis on trained LSTM networks to explain the relative

importance of inputs to the outputs [14]–[17]. However, this post-hoc analysis method cannot explain what the model actually learned during the training process, and thus it is less effective to help us check whether the model truly learns the underlying input-output relationships. Moreover, recent work in [18] argued that the application of the post-hoc method had a large uncertainty in its inference process and the inferred explanation should be used cautiously as it may produce inconsistent and unreliable results.

In this work, we propose an interpretable LSTM (iLSTM) model for NEE prediction from five environmental driver observations. iLSTM explores the internal structures of the LSTM network and overcomes the opacity of its hidden states for improved interpretability. It jointly learns network parameters, variables, and temporal importance with respect to the target prediction. Thus, iLSTM not only produces an accurate prediction but also provides interpretable insights into the data, which advances our understanding about the influence of environmental drivers and their temporal influence on the target predictions.

Additionally, the transferability of an ML model is crucial in terrestrial ecosystem modeling. Although a large amount of data is collected in FLUXNET [19], and some sites even have a long record of data, the data diversity, coverage, and quantity are still limited compared to the high complexity and heterogeneity of terrestrial ecosystems. We expect that an ML model trained in one site could be well transferred to other unobserved sites for a good prediction. Meanwhile, we realize that ML models are data driven and they could suffer from large extrapolation errors when the target sites have largely different features than the site in training. Thus, in this effort, we also evaluate the iLSTM model's transferability, i.e., investigate whether iLSTM can reliably provide consistent results when trained in one site and being used to predict other sites.

To demonstrate iLSTM's prediction performance, interpretability and transferability, we apply it to three temperate deciduous forest sites in the United States (US) for NEE prediction from five environmental drivers. The three sites are Morgan Monroe State Forest (MMS) in south-central Indiana, US, Oho Oak Opening (Oho) within the Oak Openings Preserve Metropark of northwest Ohio, and University of Michigan Biological Station (UMB) within protected forest owned by the University of Michigan. We first apply iLSTM to predict NEE in each site individually, and then perform the across-site performance evaluation where we use the iLSTM model trained in one site to predict the NEE in other sites. For both the individual and across-site studies, we evaluate the iLSTM model's prediction performance and interpretability. And for the across-site study, we additionally assess the transferability of the iLSTM.

The major contributions of this work are:

- We propose an interpretable LSTM model for NEE prediction based on environmental drivers. The model calculates variable importance and variable-wise temporal importance to the prediction by exploring the network internal structures.
- We apply the proposed iLSTM model in three forest sites in the US. The model accurately predicts the NEE based on the environmental observations, and additionally identifies the important environmental drivers to NEE prediction and their temporal relevance to the prediction.
- We perform the across-site predictions and demonstrate iLSTM's transferability in terms of both prediction accuracy and interpretability. This promises the application of iLSTM for the prediction of NEE in unobserved sites.

II. Methodology

LSTM [20] is a special type of recurrent neural network to learn long-term dependence in time-series prediction, which makes it particularly suitable for daily NEE simulation where lag times between environmental drivers and NEE can be up to months. LSTM learns to map the inputs over time to an output, thus it knows what, when, and how past driving data are relevant for predictions, enabling dynamic learning of temporal dependence. For example, in the case that LSTM reads previous t days of environmental observations as inputs to predict NEE on the current day, each LSTM cell reads several environmental input sequences one time step at a time and the output from the previous time step is fed into the next cell as another input along with the input at the current time step to affect the prediction, and so on. The outputs from the chain of LSTM cells are saved in the hidden states h_t, which dynamically add, forget, and store information from the environmental input sequences. Lastly, the LSTM network maps the information in h_t to predict the current NEE. The standard LSTM networks blindly blend the information of all input drivers into the hidden states used for prediction. It is therefore difficult to distinguish the relative contribution of individual drivers to NEE, and the mixed multi-variable data in the hidden states neglect the different dynamics of the individual driver sequence [21]. Hence, the standard LSTM does not provide interpretations of the variable importance and the variable-wise temporal importance to the prediction.

The interpretable LSTM (iLSTM) proposed in [22] addresses these challenges. It explores the internal structure of the LSTM network for interpretability and improving prediction accuracy. The network structure and data flow can be referred to the Figure 1 in [22]. First, iLSTM enables hidden states to encode individual input variables, thus the relative importance or contribution from individual drivers to NEE prediction can be calculated. Specifically, in contrast to the standard LSTM, where a hidden state vector is used to summarize the input information, iLSTM uses a 2D hidden state *matrix* with a size of D by N, where the D rows represent the number of input variables, and the N columns stand for the number of network cells. In this way, each individual input variable has its own hidden state vector to capture its unique dynamics and its unique contribution to the NEE prediction. In addition, iLSTM uses a mixture attention mechanism to summarize the variable-wise hidden states and

jointly learns the network parameters for prediction and the importance weights for interpretation. Specifically, iLSTM first implements temporal attention on the hidden states to learn the time-dependent contributions of each input variable, and then applies variable attention to summarize the variable-wise hidden states. Lastly, iLSTM integrates the variable and temporal attention weights into a probabilistic mixture model for the training process [23] to obtain the weights for variable importance (β) and temporal importance (α). The temporal importance weights α, the sum of which is one for an individual input, describe the relative importance of the given input at a particular time to the output prediction. The variable importance weights β, the sum of which is one among all the inputs, represent the relative importance of a given input variable to the output prediction regardless of the time. In comparison to the standard LSTM network which produces the predicted NEE only, iLSTM produces three sets of outputs, the predicted NEE, the variable importance and temporal importance. When training iLSTM using the observations at site A, it learns the variable and temporal importance at this site A. And when we use the iLSTM model trained at site A to predict the NEE at site B, it will produce the variable and temporal importance at site B. By comparing the calculated variable and temporal importance at these two sites, we can investigate the transferability of the iLSTM model in terms of interpretability and check the similarity of the two sites in terms of important drivers and their memory effects. The iLSTM method explores multiple dynamics within the input sequences, and learns the variable and temporal importance of the inputs to the target output prediction. Therefore, it optimizes the time-series learning for accurate prediction and in the meantime achieves model interpretability.

III. Data

In this section, we present the datasets used in this study from three different observation sites: 1) MMS, Morgan Monroe State Forest, 2) Oho, Ohio Oak Openings site, and 3) UMB, University of Michigan Biological Station. These three sites are situated in the east-central region of the United States and have generally similar climatic, environmental, and ecological features. They are a deliberate choice for diagnosing iLSTM's prediction performance, interpretability, and transferability. As a new application of iLSTM for NEE prediction and the across-site validation, we first select some sites from AmeriFlux [24] with a relatively long record of observations and having similar terrestrial ecosystem properties for analyses. After a good understanding of the method, we will apply it to a larger number of AmeriFlux sites with different features in the future study.

The MMS site [25], [26] is located in south-central Indiana with an average elevation of 275 m above mean sea level, which is categorized as a deciduous broadleaf forest. The land is dominated by woody vegetation with more than 60% coverage and the height of most of the vegetation exceeds 2 m. This site consists of broadleaf tree communities with an annual cycle of leaf-on and leaf-of periods. It has a mean annual temperature of 10.85°C, and a mean annual precipitation of 1032 mm. The climate is categorized as Cfa (humid subtropical, mild with no dry season, hot summer) according to the Köeppen climate classification.

The time-series data used in this study include the output sequence of NEE ($gCm^{-2}d^{-1}$) and five input sequences of environmental drivers. These inputs are air temperature at tower height (TBOT, °C), incoming shortwave radiation (FSDS, W/m^2), wind speed (WS, m/s), rainfall (RAIN, mm), and vapor pressure deficit (VPD, hPa). Note that all environmental drivers are daily means data, thus very short-term influences from the variables are not considered in this work. We consider 15 years of daily data from 2000 to 2014, as shown in Fig. 1, where ten years of data are selected as a training set to calibrate the iLSTM model and the remaining five years of data are used as an unseen test set to evaluate the model performance. Note that, here we use the iLSTM model to learn the relationship between the environmental drivers and NEE, thus we randomly select the training period and test period to make a fair assessment that the trained iLSTM model can be generally applied for NEE prediction in any years. Also, we intended to keep an approximately 2:1 ratio for training and testing years in our experiments

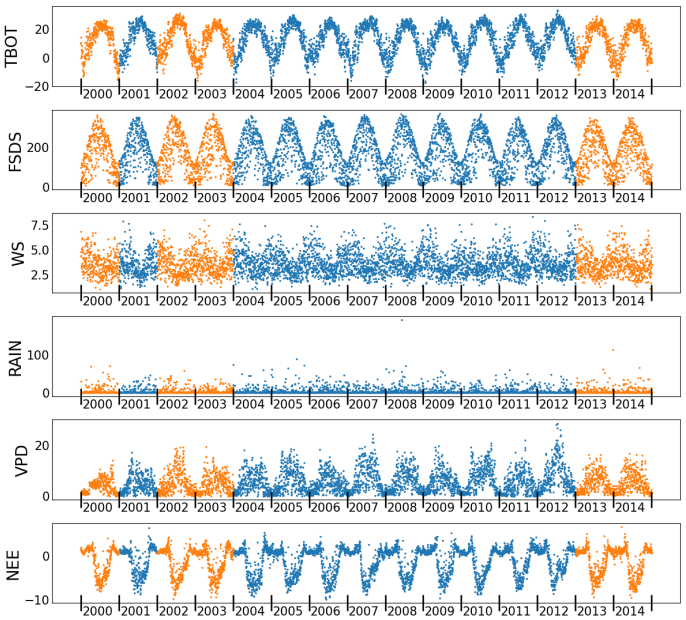

Fig. 1. MMS site. Daily observations of the five environmental drivers (TBOT, FSDS, WS, RAIN, and VPD) and the NEE between 2000-2014, where ten years of data (blue) are used for training the iLSTM model and the remaining five years of data (orange) are used for the model performance evaluation.

The Oho site [27], [28] locates in the Oak Openings Preserve Metropark of northwest Ohio with an average elevation of 230 m. The site consists of four major vegetation types including Oak Savanna, Oak Woodlands, Floodplain Forests, and Wet Prairies. The surroundings of the observation tower are dominated by black, white and red oaks with an abundance of red maples, which is an indication of high soil moisture retention and a limited fire disturbance. This site is

also categorized as a deciduous broadleaf forest where the vegetation has similar coverage, height, and an annual leaf-on and leaf-off cycle to the MMS site. The Oho site has a mean annual temperature of 10.1 °C, and a mean annual precipitation of 849 mm. The climate is categorized as Dfa (humid continental, humid with severe winter, no dry season, hot summer) according to the Köeppen climate classification. In site Oho, we consider the same five environmental drivers to predict NEE as in MMS, but Oho has a relatively short record of observations. The nine years of daily data from 2005 to 2013 in Oho are shown in Fig. 2, where we use six years of data as training and the remaining three years as test data.

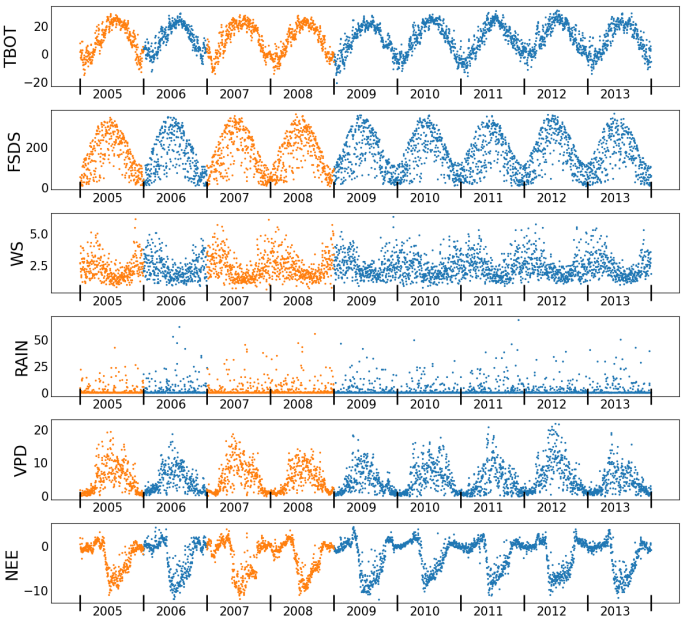

Fig. 2. Oho site. Daily observations of the five environmental drivers (TBOT, FSDS, WS, RAIN, and VPD) and the NEE between 2005-2013, where six years of data (blue) are used for training the iLSTM model and the remaining three years of data (orange) are used for the model performance evaluation.

The UMB site [29], [30] is located in the Northern Lower Peninsula of Michigan, in Cheboygan County. The site is located in a protected forest owned by the University of Michigan with an elevation of 234 m. It consists of middle-aged northern hardwoods, aspen, a conifer understory, and old growth hemlock. The site is also categorized as a deciduous broadleaf forest where the vegetation has similar coverage, height, and an annual leaf-on and leaf-off cycle to the MMS and Oho sites. The UMB site has a mean annual temperature of 5.83 °C, and a mean annual precipitation of 803 mm. The climate is categorized as Dfb (warm summer continental, significant precipitation in all seasons) according to the Köeppen climate classification. The daily data of the five environmental drivers and the NEE between 2001 to 2014 in UMB are presented in Fig. 3, where we use ten years of data as training and the other four years for testing.

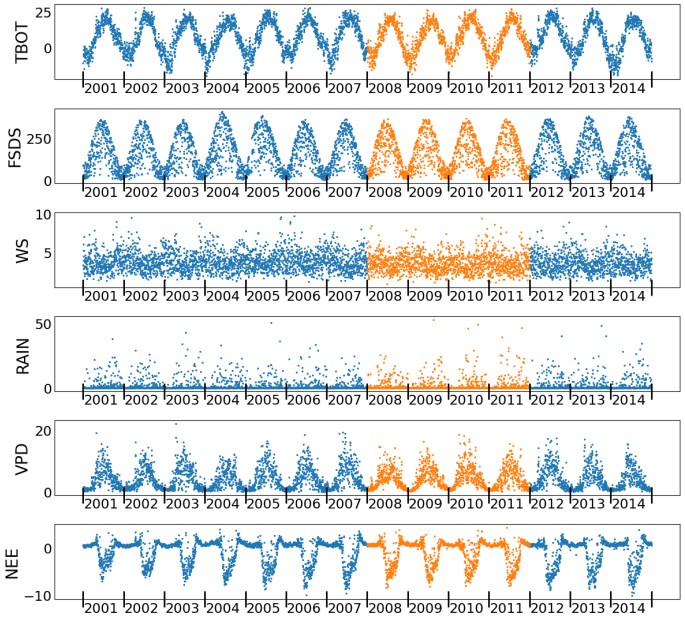

Fig. 3. UMB site. Daily observations of the five environmental drivers (TBOT, FSDS, WS, RAIN, and VPD) and the NEE between 2001-2014, where ten years of data (blue) are used for training the iLSTM model and the remaining four years of data (orange) are used for the model performance evaluation.

IV. Results and Discussion

In this section, we evaluate iLSTM's performance from three aspects: prediction accuracy, interpretability, and transferability. We assess the prediction accuracy using the coefficient of determination (R^2) and by comparing the observed and simulated NEE in the test period. The R^2 score is defined as,

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

where Y_i^{obs} and $\overline{Y}_i^{obs}$ stands for the ith observation value and mean observations, Y_i^{pred} is the model predictions. We investigate iLSTM's interpretability by analyzing the temporal and variable importance of the five environmental drivers to NEE. In all the numerical experiments, we use the previous 75 days of environmental data to predict NEE on the current day based on hyperparameter optimization. Thus, we specifically evaluate the temporal influence on this look-back window of 75 days. Lastly, we evaluate iLSTM's across-site transferability by using the trained model in one site to predict other sites and assessing the prediction performance from both accuracy and interpretability. We first analyze the performance in the individual sites and then move to the across-site investigation.

Fig. 4 shows the observed and iLSTM-predicted NEE in each year of the test period at the MMS site. The figure describes the annual NEE dynamics of a typical deciduous forest system. NEE has large negative values in summer indicating carbon uptake by vegetation, while its values are close to zero in winter and early spring. The iLSTM network is able to successfully simulate this dynamics by producing the

predicted NEE close to the observations based only on the five environmental drivers. The R^2 value for the entire test period is 0.81, suggesting a high prediction accuracy. Moreover, a closer examination of the individual year also shows a high consistency between the predicted and observed NEE, where the best R^2 is 0.84 in 2002 and the worst R^2 has 0.77 in 2003. This suggests that iLSTM learns the seasonal pattern of NEE very well, which requires learning a model of leaf phenology from lagged temperature effects. Phenology drives the seasonal cycle of leaf area and is thus strong control on carbon cycling in deciduous forest systems. The iLSTM prediction at a few time periods at finer temporal resolution demonstrates underperformance (e.g., May of 2013 test year), which might be attributed to the limited input features, the model itself, model training conditions, and hyper-parameters. We are conducting further investigation to address the issue. Meanwhile, we are developing a new residual-based multi-step approach LSTM model to tackle similar problems and obtain promising results, which will be included in our future work.

iLSTM not only produces good prediction accuracy, but also provides interpretation on the variable importance and temporal importance to the NEE prediction at the MMS site. Fig. 5 shows the variable importance of the five environmental drivers. We observe that WS and VPD are the least and second least important input variables to the NEE prediction. The drivers TBOT, FSDS, and RAIN have a relative importance of 0.257, 0.232, and 0.202 to the NEE prediction, respectively, where TBOT and FSDS have the highest contribution. This identified variable importance is consistent with our domain knowledge. MMS is a deciduous forest where temperature and radiation have a significant influence of the carbon cycle. From Fig. 1, we can see that TBOT and FSDS have more similar patterns to NEE, which suggests that they would have more impact on NEE compared to other environmental features. And iLSTM accurately identifies these two variables as the most important ones.

Additionally, Fig. 6 demonstrates the temporal importance of each of the five input variables where the results span the 75-day, look-back window. The figure indicates that different variables have different patterns in the temporal importance of the last 75 days which are used to predict the NEE at the current day. For instance, the temperature variable, TBOT, has a relatively long-term correlation to NEE, and the short-term data of the radiation variable (FSDS) contributes relatively more to the NEE prediction. This temporal importance analysis is also in line with our process understanding. The study site MMS is a deciduous forest, where the temperature can have both short-term and long-term memory effects on NEE [31]. The radiation variable (FSDS) affects latent and sensible heat fluxes and also has an impact on carbon fluxes by controlling photosynthetic uptake at shorter time scales.

This interpretability of iLSTM in variable and temporal importance can guide terrestrial ecosystem model development. Process-based terrestrial ecosystem models implement environmental variable related processes and specify the physics that considers the memory effect of the environmental drivers to the carbon fluxes. From the prediction analysis of iLSTM, we can gain insights about which environmental features are more important and which variable has a stronger memory effect on NEE. The iLSTM network may eventually also be applied to outputs from process-based terrestrial ecosystem models to better understand whether model equations and assumptions lead to consideration of memory effects that are consistent with the observations. Such information is crucial to guide process development and physics implementation in terrestrial ecosystem models and to advance our predictive understanding of the system.

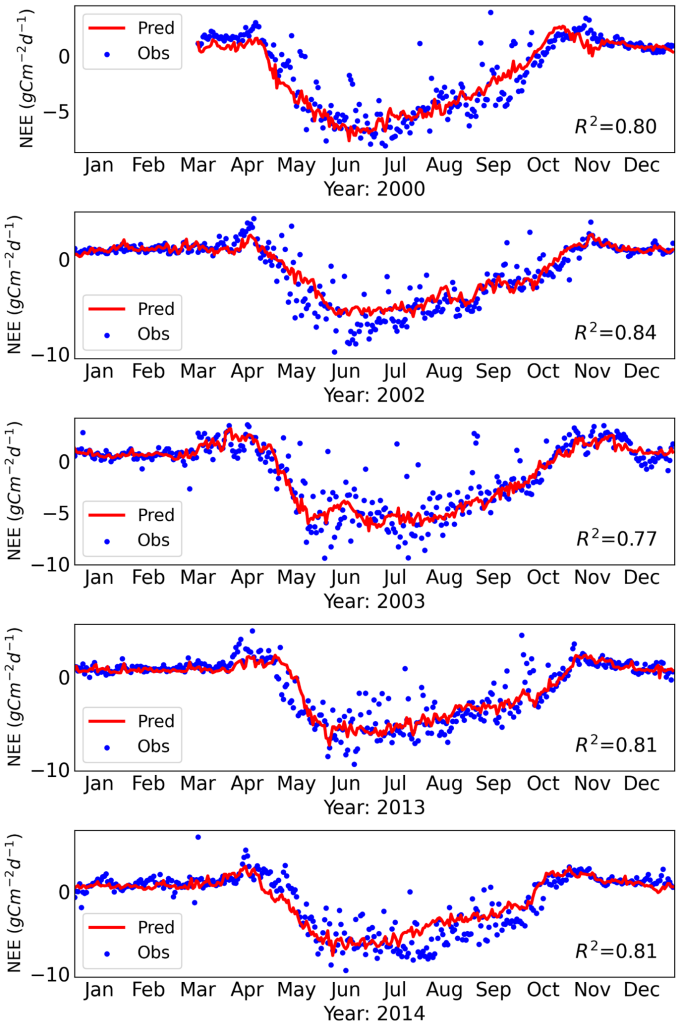

Fig. 4. The iLSTM prediction performance for individual test years at MMS site. Observed (blue dots) and iLSTM predicted (red lines) NEE in the test period. Note that we use the previous 75 days of input data to predict NEE, thus the first 75 days in 2000 do not have NEE predictions.

Fig. 7 illustrates the prediction performance at the Oho site and Fig. 10 shows the results at the UMB site in their test periods. The figures once again demonstrate the good prediction accuracy of iLSTM where the overall R^2 score is 0.83 for Oho and is 0.73 for UMB. For the Oho, the highest R^2 is up to 0.89 in 2008, and for the UMB, the highest R^2 is 0.83 in 2008. Fig. 8 and Fig. 9 depict the variable importance

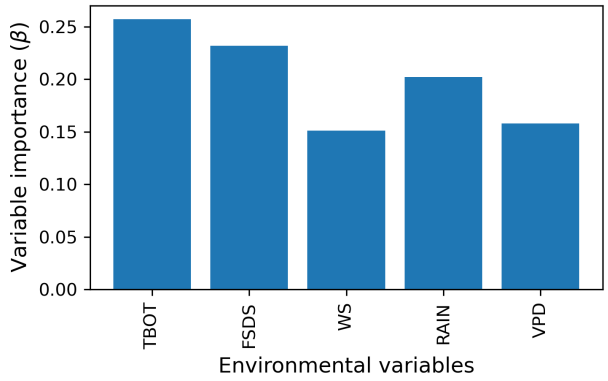

Fig. 5. MMS site. Variable importance β measures the relative importance of the five environmental drivers to NEE prediction; a large value indicates a high importance.

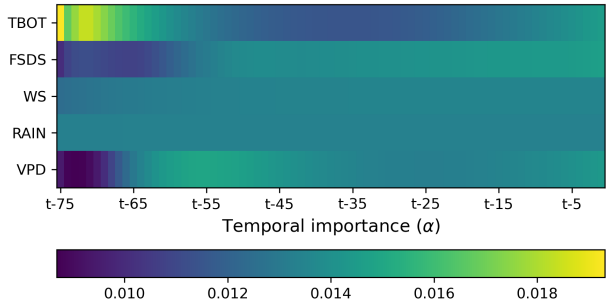

Fig. 6. MMS site. Temporal importance α measures the relative contribution of the time steps of the five environmental variables to the NEE prediction. Each row represents one variable; the light color on the left indicates a long-term contribution of the variable to NEE prediction and the light color on the fight indicates a relatively short-term impact.

and their temporal importance to NEE, respectively, at the Oho site. We observe similar patterns as the MMS site. The variables TBOT and FSDS are the two most important drivers, and TBOT shows a relatively long-term correlation to NEE prediction and FSDS demonstrates relatively short-term impacts. On the other hand, the UMB site presents a slightly different pattern compared to MMS and Oho on the variable importance shown in Fig. 11 and the temporal importance shown in Fig. 12. In UMB, variable RAIN overtakes FSDS and becomes the second important driver, indicating this site may be more sensitive to moisture controls on NEE. Generally in a deciduous forest, temperature, radiation, and water are usually important drivers for NEE. Both MMS and Oho sites have no dry seasons all year round and have hot summer, so the temperature (TBOT) and radiation (FSDS) are more important than water (RAIN). In comparison to MMS and Oho, the UMB site in the high latitude has a relatively lower annual precipitation and a cooler summer, so RAIN becomes more important than FSDS. In all three deciduous forests, the temperature is always the most important driver to NEE and TBOT consistently shows the long-term effect on NEE prediction.

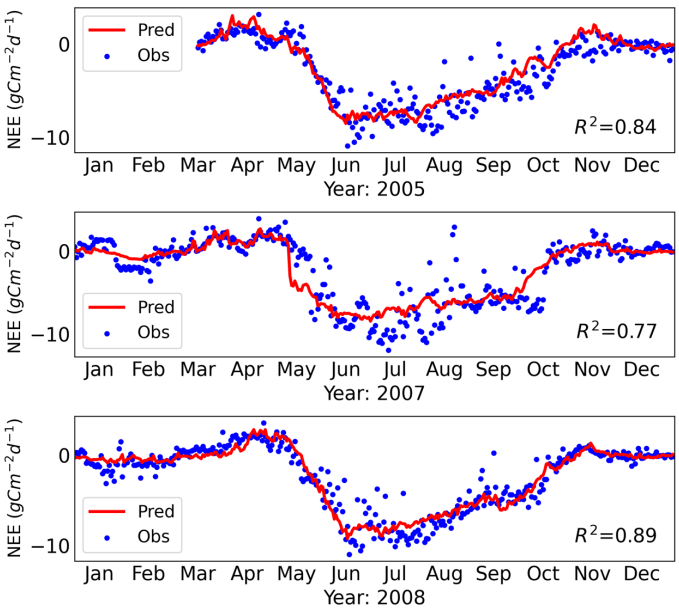

Fig. 7. The iLSTM prediction performance for individual test years at Oho site. Observed (blue dots) and iLSTM predicted (red lines) NEE in the test period. Note that we use the previous 75 days of input data to predict NEE, thus the first 75 days in 2005 do not have NEE predictions.

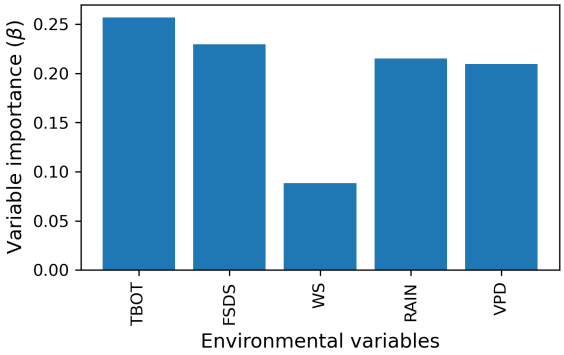

Fig. 8. Oho site. Variable importance β measures the relative importance of the five environmental drivers to NEE prediction; a large value indicates a high importance.

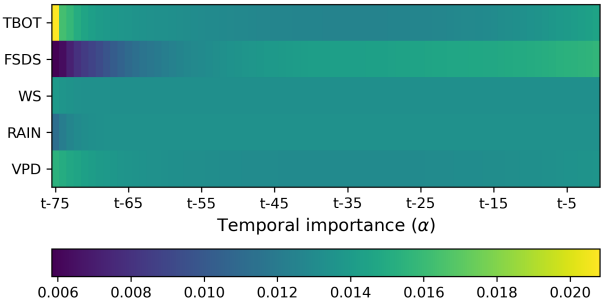

Fig. 9. Oho site. Temporal importance α measures the relative contribution of the time steps of the five environmental variables to the NEE prediction. Each row represents one variable; the light color on the left indicates a long-term contribution of the variable to NEE prediction and the light color on the right indicates a relatively short-term impact.

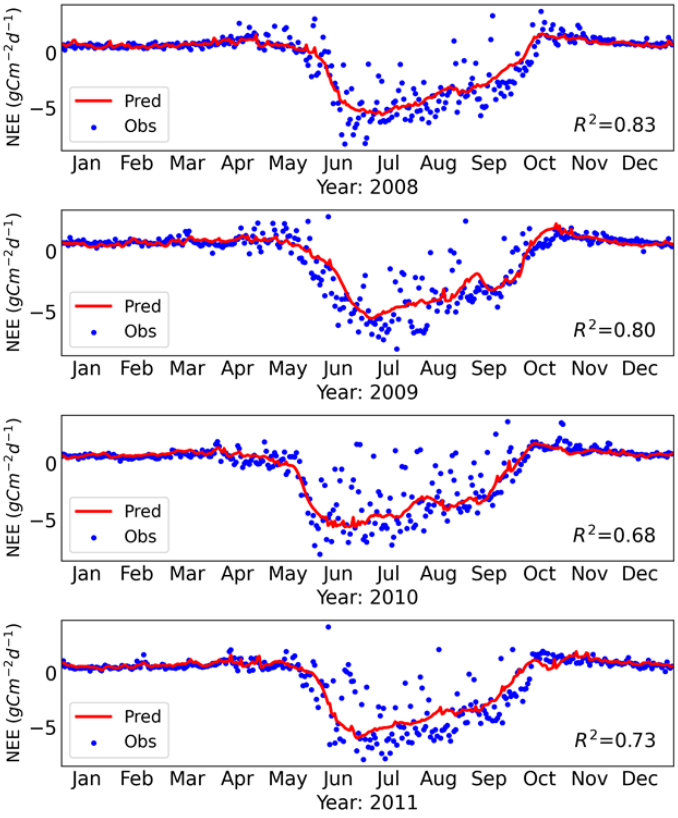

Fig. 10. The iLSTM prediction performance for individual test years at UMB site. Observed (blue dots) and iLSTM predicted (red lines) NEE in the test period.

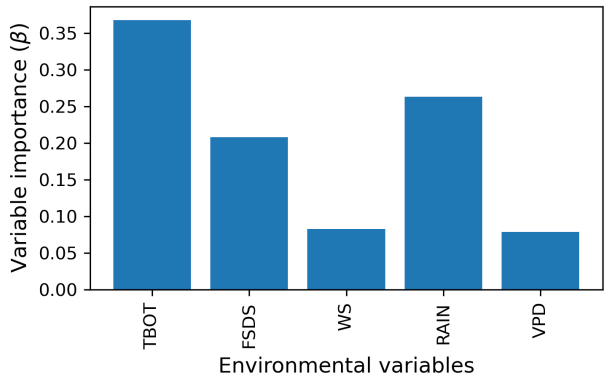

Fig. 11. UMB site. Variable importance β measures the relative importance of the five environmental drivers to NEE prediction; a large value indicates a high importance.

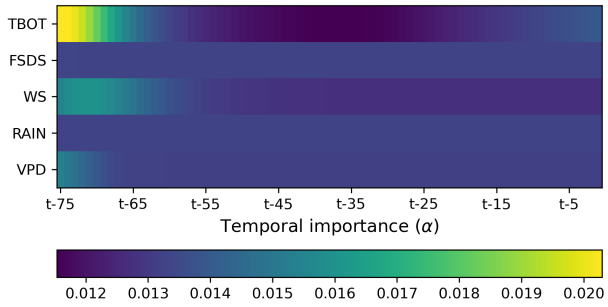

Fig. 12. UMB site. Temporal importance α measures the relative contribution of the time steps of the five environmental variables to the NEE prediction. Each row represents one variable; the light color on the left indicates a long-term contribution of the variable to NEE prediction and the light color on the fight indicates a relatively short-term impact.

In the following section, we discuss iLSTM's transferability. We design nine numerical experiments, i.e., we train the iLSTM model using one site of data and then apply the trained model to predict the three sites. For example, we use the training data in MMS to train iLSTM, and then use the trained model to predict NEE in the test period of the MMS, Oho, and UMB sites; this includes three numerical experiments, i.e., MMS-MMS, MMS-Oho, and MMS-UMB. As we have three sites, the combination constitutes nine experiments. Table I summarizes the prediction performance in the test periods of these nine experiments. It is not surprising that the model has the best prediction accuracy (the R^2 scores in the diagonal of the table) when it is trained by the data from the same site. Although the prediction accuracy decreases when the test data differ from the training set, the performance is still acceptable. For example, when iLSTM is trained by MMS data, it can predict the NEE in the Oho site with a R^2 of 0.68. Similarly, when iLSTM is trained by Oho data, we also get a R^2 of 0.68 in predicting the NEE at the MMS site. This high prediction accuracy in the across-site evaluation between MMS and Oho suggests the good transferability of the iLSTM model.

Meanwhile, we notice that MMS and UMB have relatively poor across-site prediction performance. The R^2 is 0.42 in predicting the NEE of UMB when iLSTM is trained by MMS, and the R^2 is 0.34 in predicting the NEE of MMS when the model is trained by UMB. Among the three sites, MMS and UMB are the most distant pair of sites geographically and have quite different climate and environmental conditions compared to the closeness between Oho and UMB, which are both located in the Great Lakes region. So, it seems that the transferability of iLSTM depends on the similarity of the data in the training and test sites; the closer the data are, the high the transferability. More investigation about this point will be performed in the future study using more sites.

Fig. 13 and 14 depict the variable and temporal importance, respectively, in the nine numerical experiments for the across-site transferability examination in term of interpretability. In both figures, the first column presents the prediction results in MMS when the model is trained by the data from different

Table I. The prediction accuracy measured in R^2 scores for the across-site transferability investigation.

Train \ Test	MMS	Oho	UMB
MMS	0.81	0.68	0.42
Oho	0.68	0.83	0.59
UMB	0.34	0.65	0.76

sites. In a similar manner, the second column shows the prediction results in the Oho site and the third column is for the predictions in the UMB site. We can see that although iLSTM is trained by the data from different sites, it seems the patterns of variable importance are robust across sites, and the model can consistently identify that TBOT and FSDS are generally the two most important drivers of NEE in the MMS and Oho sites. TBOT, FSDS, and RAIN generally are identified as the influential factors in the UMB site. One exception is when the model is trained by UMB (the third row of Fig. 13), variable RAIN shows slightly higher importance than FSDS in predicting the NEE at the MMS and Oho sites. This indicates a relatively poor transferability of UMB to the other two sites, which is consistent with the above discussion on the transferability of prediction accuracy in Table I.

Additionally, Fig. 14 indicates that the iLSTM model can consistently identify that the temperature variable TBOT has a long-term contribution to the NEE prediction and the radiation has a relatively short-term influence, no matter the training data from which site. This observation promises the transferability of iLSTM in identifying the important environmental drivers and their temporal importance for NEE prediction. In the future, we will perform more studies for further investigation using a larger range of forest sites.

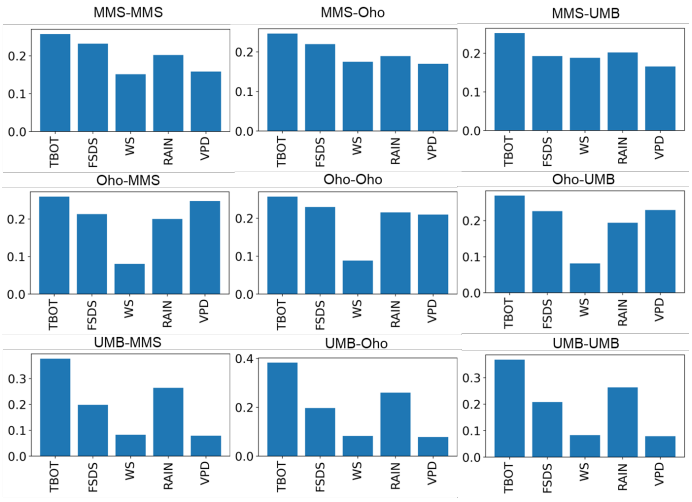

Fig. 13. Bar plots show the importance of the five environmental drivers to NEE prediction in the across-site transferability study, where MMS-Oho represents using the iLSTM model trained by the data at the MMS site to predict the NEE at the Oho site. The higher the bar is, the larger the importance.

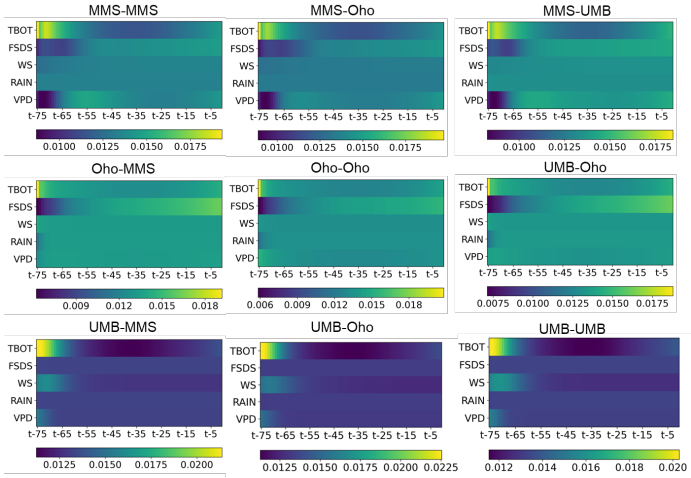

Fig. 14. Temporal importance of the five environmental variables to the NEE prediction at the across-site transferability study, where MMS-Oho represents using the iLSTM model trained by the data at MMS site to predict the NEE at the Oho site.

V. Conclusion and Future Work

In this study, we propose an interpretable LSTM (iLSTM) model for NEE prediction from five environmental drivers at three forest sites. The iLSTM model demonstrates high accuracy in predicting NEE by capturing the correlations and the individual dynamics of the environmental driver sequences. Additionally, iLSTM produces a reasonable interpretation about the input variables' importance as well as their temporal importance to the prediction. The interpretability of iLSTM not only agrees with our domain knowledge, but also offers scientific insights into the mechanisms of the ecological processes. Furthermore, iLSTM shows reasonable across-site transferability in terms of both prediction accuracy and interpretability. It is also worth noting that based on our numerical experiments, iLSTM demonstrated comparable point estimate prediction performance with the traditional LSTM model. We only include iLSTM results in this paper since we mainly focus on the interpretability capability of the ML model in this specific domain application. In the future, we will extend our investigation to more observation sites with diverse forest and climate types to explore and improve the prediction and interpretability of the iLSTM model. Meanwhile, we will use our gained mechanical understanding to guide the process-based terrestrial ecosystem model development. At last, we will include a more comprehensive study incorporating performance comparisons between multiple models and methods with data from various domains.

Acknowledgment

This research was supported by the Artificial Intelligence Initiative as part of the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the US DOE under contract DE-AC05-00OR22725. It is also sponsored by the TES-SFA Project funded by the US DOE, Office of Biological

and Environmental Research, and the Data-Driven Decision Control for Complex Systems (DnC2S) project funded by the US DOE, Office of Advanced Scientific Computing Research.

REFERENCES

[1] R. H. Waring and S. W. Running, "Chapter 3 - carbon cycle," in *Forest Ecosystems (Third Edition)*, third edition ed., R. H. Waring and S. W. Running, Eds. San Diego: Academic Press, 2007, pp. 59–98. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123706058500086

[2] J. Cai, K. Xu, Y. Zhu, F. Hu, and L. Li, "Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest," *Applied Energy*, vol. 262, p. 114566, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261920300787

[3] A. M. Melesse and R. S. Hanley, "Artificial neural network application for multi-ecosystem carbon flux simulation," *Ecological Modelling*, vol. 189, no. 3, pp. 305–314, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304380005002061

[4] P. Nguyen and M. Halem, "Deep learning models for predicting co2 flux employing multivariate time series," *Mile TS: Anchorage, AK, USA*, 2019.

[5] D. Lu, S. Liu, and D. Ricciuto, "An efficient bayesian method for advancing the application of deep learning in earth science," in *2019 International Conference on Data Mining Workshops (ICDMW)*, 2019, pp. 270–278.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[7] S. Besnard, N. Carvalhais, M. A. Arain, A. Black, B. Brede, N. Buchmann, J. Chen, J. G. P. W. Clevers, L. P. Dutrieux, F. Gans, M. Herold, M. Jung, Y. Kosugi, A. Knohl, B. E. Law, E. Paul-Limoges, A. Lohila, L. Merbold, O. Roupsard, R. Valentini, S. Wolf, X. Zhang, and M. Reichstein, "Memory effects of climate and vegetation affecting net ecosystem co2 fluxes in global forests," *PLOS ONE*, vol. 14, no. 2, pp. 1–22, 02 2019. [Online]. Available: https://doi.org/10.1371/journal.pone.0211510

[8] M. Reichstein, S. Besnard, N. Carvalhais, F. Gans, M. Jung, B. Kraft, and M. Mahecha, "Modelling landsurface time-series with recurrent neural nets," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 7640–7643.

[9] B. Kraft, M. Jung, M. Körner, C. Requena Mesa, J. Cortés, and M. Reichstein, "Identifying dynamic memory effects on vegetation state using recurrent neural networks," *Frontiers in big Data*, vol. 2, p. 31, 2019.

[10] A. Pérez-Suay, J. E. Adsuara, M. Piles, L. Martínez-Ferrer, E. Díaz, A. Moreno-Martínez, and G. Camps-Valls, "Interpretability of recurrent neural networks in remote sensing," in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2020, pp. 3991–3994.

[11] G. Camps-Valls, M. Reichstein, X. Zhu, and D. Tuia, "Advancing deep learning for earth sciences: From hybrid modeling to interpretability," in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2020, pp. 3979–3982.

[12] G. Konapala, S.-C. Kao, S. L. Painter, and D. Lu, "Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US," *Environmental Research Letters*, vol. 15, no. 10, p. 104022, sep 2020. [Online]. Available: https://doi.org/10.1088/1748-9326/aba927

[13] W.-P. Tsai, D. Feng, M. Pan, H. Beck, K. Lawson, Y. Yang, J. Liu, and C. Shen, "From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling," *Nature communications*, vol. 12, no. 1, pp. 1–13, 2021.

[14] W. L. Zhao, P. Gentine, M. Reichstein, Y. Zhang, S. Zhou, Y. Wen, C. Lin, X. Li, and G. Y. Qiu, "Physics-constrained machine learning of evapotranspiration," *Geophysical Research Letters*, vol. 46, no. 24, pp. 14 496–14 507, 2019.

[15] M. Shahhosseini, G. Hu, I. Huber, and S. V. Archontoulis, "Coupling machine learning and crop modeling improves crop yield prediction in the us corn belt," *Scientific reports*, vol. 11, no. 1, pp. 1–15, 2021.

[16] M. Ancona, E. Ceolini, C. Öztireli, and M. Gross, "Towards better understanding of gradient-based attribution methods for deep neural networks," in *International Conference on Learning Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=Sy21R9JAW

[17] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," in *International conference on machine learning*. PMLR, 2017, pp. 3145–3153.

[18] Y. Yang and T. F. M. Chui, "Modeling and interpreting hydrological responses of sustainable urban drainage systems with explainable machine learning methods," *Hydrology and Earth System Sciences*, vol. 25, no. 11, pp. 5839–5858, 2021. [Online]. Available: https://hess.copernicus.org/articles/25/5839/2021/

[19] T. C. C. E. e. a. Pastorello, G., "The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data," *Scientific Data*, vol. 7, no. 1, p. 225, Jul. 2020. [Online]. Available: https://doi.org/10.1038/s41597-020-0534-3

[20] L. E. Jeffrey *et al.*, "Finding structure in time," *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.

[21] L. Zhang, C. Aggarwal, and G.-J. Qi, "Stock price prediction via discovering multi-frequency trading patterns," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2017, pp. 2141–2149.

[22] T. Guo, T. Lin, and N. Antulov-Fantulin, "Exploring interpretable lstm neural networks over multi-variable data," in *International conference on machine learning*. PMLR, 2019, pp. 2494–2504.

[23] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection," in *International conference on learning representations*, 2018.

[24] "AmeriFlux." [Online]. Available: https://ameriflux.lbl.gov/

[25] K. Novick and R. Phillips, "Ameriflux us-mms morgan monroe state forest, ver. 21-5, ameriflux amp, (dataset)," 2022. [Online]. Available: https://doi.org/10.17190/AMF/1246080

[26] D. Roman, K. Novick, E. Brzostek, D. Dragoni, F. Rahman, and R. Phillips, "The role of isohydric and anisohydric species in determining ecosystem-scale response to severe drought," *Oecologia*, vol. 179, no. 3, pp. 641–654, 2015.

[27] J. Chen, H. Chu, and A. Noormets, "Ameriflux base us-oho oak openings, ver. 7-5, ameriflux amp, (dataset)," 2021. [Online]. Available: https://doi.org/10.17190/AMF/1246089

[28] A. Noormets, S. G. McNulty, J. L. DeForest, G. Sun, Q. Li, and J. Chen, "Drought during canopy development has lasting effect on annual carbon balance in a deciduous temperate forest," *New Phytologist*, vol. 179, no. 3, pp. 818–828, 2008. [Online]. Available: https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8137.2008.02501.x

[29] C. Gough, G. Bohrer, and P. Curtis, "Ameriflux base us-umb univ. of mich. biological station, ver. 18-5, ameriflux amp, (dataset)," 2022. [Online]. Available: https://doi.org/10.17190/AMF/1246107

[30] C. M. Gough, B. S. Hardiman, L. E. Nave, G. Bohrer, K. D. Maurer, C. S. Vogel, K. J. Nadelhoffer, and P. S. Curtis, "Sustained carbon uptake and storage following moderate disturbance in a great lakes forest," *Ecological Applications*, vol. 23, no. 5, pp. 1202–1215, 2013.

[31] D. Lu, D. M. Ricciuto, and S. Liu, "An interpretable machine learning model for advancing terrestrial ecosystem predictions," Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), Tech. Rep., 2022.