Understanding and Modeling Climate Impacts on Photosynthetic Dynamics with FLUXNET Data and Neural Networks

Nanyan Zhu 1, Chen Liu 2, Andrew F. Laine 3 and Jia Guo 4,*

1 Biological Sciences, Columbia University, New York, NY 10027, USA; nz2305@columbia.edu
2 Electrical Engineering, Columbia University, New York, NY 10027, USA; cl3760@columbia.edu
3 Biomedical Engineering, and Radiology, Columbia University, New York, NY 10027, USA; al418@columbia.edu
4 Psychiatry, and Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, USA
* Correspondence: jg3400@columbia.edu
The code is publicly available on https://github.com/RosalieZhu/DL-CO2.
This paper is adapted from a conference version presented on the 2019 9th International Conference on Future Environment and Energy.

Received: 1 February 2020; Accepted: 6 March 2020; Published: 12 March 2020

Abstract: Global warming, which largely results from excessive carbon emission, has become an increasingly heated international issue due to its ever-deteriorating trend and the profound consequences. Plants sequester a large amount of atmospheric CO\(_2\) via photosynthesis, thus greatly mediating global warming. In this study, we aim to model the temporal dynamics of photosynthesis for two different vegetation types to further understand the controlling factors of photosynthesis machinery. We experimented with a feedforward neural network that does not utilize past histories, as well as two networks that integrate past and present information, long short-term memory and transformer. Our results showed that one single climate driver, shortwave radiation, carries the most information with respect to prediction of upcoming photosynthetic activities. We also demonstrated that photosynthesis and its interactions with climate drivers, such as temperature, precipitation, radiation, and vapor pressure deficit, has an internal system memory of about two weeks. Thus, the predictive model could be best trained with historical data over the past two weeks and could best predict temporal evolution of photosynthesis two weeks into the future.

Keywords: neural network; deep learning; climate; photosynthesis; ecology; FLUXNET

1. Introduction

Climate change, specifically global warming, has long been considered a pressing issue to the international society as it disrupts the stability of the ecosystem and threatens the prosperity of mankind [1]. Anthropogenic activities that yield excessive carbon emission, such as fossil fuel burning, industrialization and animal husbandry, have been adversely affecting the ecosystem by raising the global temperature. The dynamics of greenhouse gases, such as CO\(_2\), play pivotal roles in controlling the radiative forcing and the energy balance of the whole earth system [2]. Therefore, understanding the full cycle, especially the sources and sinks, of atmospheric CO\(_2\) is critical to better control the CO\(_2\) concentrations in the future to a reasonable extent.

Photosynthesis pathway is the largest land surface CO\(_2\) sink that sequester atmospheric CO\(_2\) into vegetation biomass and stores them as living biomass. The magnitude of photosynthetic carbon sequestration pathway is around 120–130 Pg C per year (1 Pg = 10\(15\) grams) [3]. Ecosystem respiration,
including autotrophic and heterotrophic, consumes most of the photosynthetic carbon sink, cycles it back into the atmosphere, and results in a much smaller net carbon sink compared with gross carbon input from photosynthesis. The balance between photosynthesis carbon input and ecosystem reparation carbon output determines the fate of atmospheric CO$_2$ concentration and mitigates the anthropogenic CO$_2$ emissions, such as from fossil fuel and biomass burning [4].

Given the significant role of photosynthesis carbon uptake in determining the global carbon cycle, atmospheric CO$_2$ concentrations, radiative balance of earth system, and global warming, it is crucial to gain mechanistic understanding of photosynthesis machinery and its relationship to climate factors such as temperature, precipitation, and radiation. Observational networks, such as FLUXNET [5], have been established globally to measure and understand various different components of land surface carbon cycle including photosynthesis. At leaf scale, photosynthesis reaction is carried out by Ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO enzyme), which combines CO$_2$ and water molecules to generate carbohydrate products. It occurs at two stages: (1) first, light-dependent reactions capture the energy of light and store in adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide phosphate hydrogen (NADPH); (2) second, light-independent reactions capture and reduce carbon dioxide. This biological reaction relies on substrate concentration (CO$_2$ and water), depends on the activity of temperature-sensitive RUBISCO enzyme, and is driven by solar energy. Thus, theoretically, one is able to build an effective photosynthesis model that takes all those important factors into account and predicts the magnitude of land surface photosynthesis given relevant climate drivers.

Historically, mechanistic models have been used to study the dynamics of land surface photosynthesis and its relevance to the fate of global warming. These models are based on either Monteith law of light use efficiency [6] or Farquhar photosynthesis modeling framework [7]. Alternatively, the photosynthesis rate could also be estimated by data-driven machine learning models that are trained on a large amount of observed photosynthesis data.

The latter approach is gaining popularity because of the growth of Eddy covariance observational networks that collect photosynthetic data on a daily basis over the years, along with the advancements of effective machine learning techniques. Data-driven models have been applied to a wide-range of areas recently, and have demonstrated their robustness in multiple tasks. Methods such as feedforward neural network (FFNN) [8], random forest [9], model trees ensemble [10–12] and support vector regression [13,14] have been utilized to estimate land surface–atmosphere fluxes from site level to regional or global scales [3,11,12,14–19]. Adaptive neuro-fuzzy inference system and general regression neural network have also been used to estimate daily carbon fluxes in forest ecosystems [20]. However, very little research has yet leveraged algorithms that specialize in the integration of temporal dependencies, such as hidden Markov model (HMM), long-short term memory (LSTM) [22] or transformer [23]. Thus, in this study, we employed LSTMs and transformers to predict photosynthetic activities from the past history of climate drivers, and compared their performances against an FFNN counterpart. Our intuition is that the temporal-aware neural network models shall be able to capture the multi-day dynamics of photosynthesis, thus yielding better predictions.

Gross primary product (GPP), a measure of the amount of energy that plants can trap from the sun, is used in this study as a correlate to the rate of photosynthesis. The objective of this study is to model the temporal dynamics of plant photosynthesis in terms of GPP with three different neural network architectures. With appropriately designed model experiments, we also aim to explore and understand the impact of different climate drivers on photosynthesis, as well as the natural ecosystem memory of the photosynthetic activities.

2. Results

We first modeled the dynamics of GPP across four different sites using Feedforward Neural Networks (FFNNs), one for each site and trained independently. The model architecture, formulas, and other details will be covered in the Materials and Methods section. In principle, each prediction of GPP on any day is made by the model based on the six climate drivers including GPP from the
Previous day. Afterwards, a leave-one-out experiment was conducted for climate drivers, where one driver is excluded from the input at a time, in the hope to draw inferences on which factors are the most influential on predicting GPP. The results are shown in the top panels in Figure 1. The experimental settings for subsequent models were the same unless otherwise specified.

**Figure 1.** The gross primary product (GPP) prediction performance of (1) the feedforward neural network (FFNN), where only the current climate drivers are utilized; (2) the long short-term memory (LSTM), where current and past climate drivers are taken into account; and (3) the transformer, where current and past climate drivers are comprehensively encoded. A leave-one-out experiment is also conducted for each model, where one of the climate drivers are excluded from the input. Within each neural network candidate, a one-way analysis of variance (ANOVA) followed by a post-hoc Tukey’s honestly significant difference (HSD) test [21] is performed to identify the climate drivers without which the model performance would deteriorate. Asterisks indicate the ones whose square errors significantly deviate from those of the version with all climate drivers provided. One, two and three asterisks respectively corresponds to HSD p-value of less than 0.05, 0.01, and 0.001. GPP: gross primary product; FSDS: shortwave radiation; Pre: precipitation; Temp: temperature; FLDS: longwave radiation; VPD: vapor pressure deficit.

Secondly, we employed Long Short-Term Memory (LSTM) Networks [22], a member of the Recurrent Neural Networks (RNNs) family which specializes in incorporating temporal information in time-series data, to replace the FFNN counterpart in the same analyses. Instead of merely utilizing the climate drivers in the day before, we let the LSTM candidate take into consideration the data within the past thirty days. Compared to FFNN, our LSTM model achieved a consistently lower mean square error and higher model-data correlation (the middle panels in Figure 1).

The last neural network variant we analyzed was the transformers [23], which was recently introduced as a state-of-the-art natural language comprehension model whose utility covers the areas...
of machine translation [24], question-answering [25], and various other tasks. Similar to the LSTMs, thirty days’ worth of data was provided as the input. In our task, there were several cases that the transformers (the bottom panels in Figure 1) outperformed FFNN, such as it maintained a higher Pearson correlation when shortwave radiation data was absent, but overall it did not demonstrate superior performance over FFNN, let alone the even better-performing LSTM.

Given that the LSTM was the best candidate among the three, we then inspected multiple variants of the LSTM model to explore the natural temporal memory of the photosynthetic process. The photosynthetic activities as measured by GPP depend on the past history of environmental drivers, but that temporal dependency is not likely to extend infinitely to the past. For example, whether it rained a million years ago would minimally affect the GPP tomorrow. An LSTM variant that utilizes information within a reasonable temporal span is likely to achieve optimal prediction accuracy. On the other hand, incorporating information over a period much shorter or longer than the natural memory duration is likely to yield sub-optimal results. The natural length of memory may be deduced by the performance of different LSTM variants, and it may shed light on the internal linkage of photosynthesis process across time.

Finally, we experimented on the forecasting ability of the LSTM models such that all variants were given the historical data on climate drivers over the experimentally determined two-week period, but were asked to predict the GPP on varying days ahead in the future. We would like to find out the turning point at which the model was no longer able to yield faithful predictions, which may again imply the temporal memory of photosynthetic dynamics. The results are displayed in Figure 2.

3. Discussion

In all three network candidates, the prediction accuracy were fairly consistent across multiple sites when all climate drivers were provided as the input (Figure 1). The second grassland site seemed to be the easiest to predict, judging from the fact that all three candidates achieved the best performance on that site, as indicated by low mean square error and high Pearson correlation. The reason behind this observation is yet to be discussed.
Furthermore, our leave-one-out experiment helped us identify the most influential climate drivers on GPP prediction. The accurate estimation of GPP relies heavily on knowing the shortwave radiation in the region. This observation is reasonable, as shortwave radiation is radiation energy with wavelengths around the range of visible light—the primary energy source for photosynthesis. Apart from shortwave, knowledge of any other climate driver does not significantly influence the GPP prediction accuracy consistently across all three neural network candidates, and interestingly enough this is also the case for GPP itself: presence or absence of past GPP data does not significantly affect the prediction of future GPP levels.

After identification of the most influential climate driver, the phenomenon that the second grassland is the “most predictable” site becomes more explainable. If we look ahead into the shortwave radiation distribution across the four sites shown in Figure 3 in the upcoming Materials and Methods section, we could draw a plausible conclusion that the abundance of shortwave radiation might be the reason behind the high predictability. The shortwave radiation distribution in the second grassland implies a lot of non-zero, information-carrying observations of this climate driver, which would reasonably improve the prediction accuracy of GPP.

In both the initial study and the leave-one-out experiments, the performance of the LSTM models was consistently better than the FFNN counterparts, which we would attribute to the inclusion of information over a longer duration of past history. The less appreciable performance of transformers compared to LSTMs was likely due to the fact that transformers were originally designed for natural language translation rather than for time-series interpretation. Besides being less task-specific, the transformer we implemented had more parameters than the other two candidates, meaning that it would require more data to optimize. That also might have yielded the less competitive performance given the same amount of training data.

In the study of past-history dependency and forecasting ability of the LSTM models (Figure 2), our experiment implied that photosynthesis has an internal memory length of two weeks. In the past-history dependency study, the LSTM performance peaked when the historical data fed into it covered the past two weeks; providing a more distant history of the climate drivers no longer improved the performance. In the forecasting study, our LSTM model was able to predict GPP two weeks into the future at most. Attempts to predict beyond two weeks demonstrated a significant decrease in prediction accuracy as indicated by higher mean squared error and lower correlation between the prediction and the respective ground truth. Such deterioration was especially obvious in grassland sites (orange and green bars). Our results on the optimal historical system memory were quite consistent with the longest temporal predictability, which were both about two weeks.

4. Materials and Methods

4.1. FLUXNET Dataset

We used Eddy covariance data collected from four different sites in the FLUXNET observational network [5]. Eddy covariance [26] is a measurement technique that quantifies ecosystem gas exchange and areal emission rates. By simultaneously measuring the velocity of the swirling wind with anemometers and the gas concentrations with infrared gas analyzers, the technique can eventually generate estimations of the flux of gases into or out of the ecosystem. By now, hundreds of sites are operating on a long-term and continuous basis. In this study, we included two evergreen needleleaf forest sites and two grassland sites (Table 1). Figure 3 shows the probability density distributions of the climate drivers at the four selected sites.
Table 1. Summary of the geographic location and basic information of the four FLUXNET sites included in this study. Time-varying data is represented in mean ± standard deviation. AP: annual precipitation. GPP: gross primary production, a measure that correlates to the rate of photosynthesis in a region.

| Name   | Latitude (N) | Longitude (E) | Elevation (m) | Land Cover Type      | AP (mm yr⁻¹) | Temp (°C) | GPP (gC m⁻² day⁻¹) |
|--------|--------------|---------------|---------------|----------------------|--------------|-----------|-------------------|
| NL-Loo | 52.16        | 5.74          | 25            | Evergreen Needleleaf | 419 ± 829    | 10.1 ± 6.4 | 4.3 ± 3.1         |
| DE-Tha | 50.96        | 13.56         | 385           | Evergreen Needleleaf | 420 ± 988    | 8.8 ± 7.9  | 5.1 ± 4.1         |
| AT-Neu | 47.11        | 11.31         | 970           | Grassland           | 334 ± 814    | 6.8 ± 8.2  | 5.9 ± 5.9         |
| US-Var | 38.41        | -120.95       | 129           | Grassland           | 282 ± 980    | 15.8 ± 6.8 | 1.8 ± 2.8         |

Figure 3. Probability densities of the observed climate drivers at the four FLUXNET sites included in this study. Red and orange are grassland sites while blue and green are evergreen needleleaf forest sites.

4.2. Neural Network Architectures and Formulas

4.2.1. FFNN

The first neural network candidate we explored, FFNN, composes of five layers of cells as shown in Figure 4b. Each cell is connected to all cells in the previous and the next layer but intact from other cells. Information from the input feature vectors, which in our case are the six climate drivers, can only pass from a shallower layer to a deeper layer but not the other way around, hence the “feedforward” part in the name of the model. Figure 4a indicates the composition of the input feature vectors and prediction ground truths for not only the FFNN but also the two other candidates.

An arbitrary subset of the FFNN, with \( n \) pre-synaptic cells attaching to the same post-synaptic cell, is depicted in Figure 5a for a detailed demonstration of the network. Conceptually, the signal passed to the post-synaptic cell is a weighted sum of the outputs from the pre-synaptic cells, and that signal goes through an activation function before it passes on to the cells in the next layer.

The mathematical description of this scenario is denoted in Equation (1).

\[
y = f \left( \sum_{i=1}^{n} x_i \cdot w_i \right)
\]
where $x_i$ are the input features; $w_i$ are neural network parameters, or in technical terms, weights, for each cell; $f$ is the activation function, which is our implementation is a rectified linear unit (ReLU) [27]; and $y_i$ is the output. For a cell in the first layer, the input features are the climate drivers, whereas for other cells deeper down the network, the input features are abstractions extracted from previous cells. For the single cell in the last layer, its output is the predicted GPP.

**4.2.2. LSTM**

The second candidate, LSTM, models GPP based on not only climate drivers on the previous day but also climate drivers from the more distant past. As illustrated in Figure 4c, the six climate drivers
over the past \( k \) days are sequentially fed into the LSTM in the correct chronological order. The order at which they enter the LSTM matters to the model, as the LSTM would adapt its states depending on the past and present information. On a conceptual level, LSTM shares some similar design philosophies with a hidden Markov model who computes the conditional probability \( P(x_t|x_{t-k},...,x_{t-1},\text{state}) \). The linear layer immediately after the LSTM cell in the figure stands for a feedforward layer that very much resembles the last layer in the FFNN counterpart: all the extracted features from the LSTM cell are condensed into a single FFNN post-synaptic cell whose output is the predicted GPP at day \( t \).

The building block, or an LSTM cell, is demonstrated in Figure 5b. Conceptually, an LSTM cell keeps track of two internal states (i.e., cell state and hidden state) which respectively correspond to long-term and short-term memory) and consists of three gates (i.e., forget gate that determines what fraction of the previous cell state shall be maintained, input gate that decides how much new information from the current input shall be integrated to the cell state and hidden state, and output gate that updates the hidden state).

Mathematically speaking, given an input sequence \( x = (x_1, ..., x_T) \), a standard LSTM computes the hidden vector sequence \( h = (h_1, ..., h_T) \), which can be used as the output of the model, by iterating the following equations from \( t = 1 \) to \( T \):

\[
\begin{align*}
    f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\
    i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\
    o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\
    c_t &= f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c) \\
    h_t &= o_t \odot \tanh(c_t),
\end{align*}
\]

where \( \sigma \) is the sigmoid function, and \( f, i, o, c \) and \( h \) are respectively the input gate, forget gate, output gate, cell state and hidden state vectors. \( b_f, b_i, b_o \) and \( b_c \) are the biases (i.e., additive scalars) to their respective gates. Matrices \( W \) and \( U \) contain the model weights at the input and recurrent connections. \( x \) is the model input at each time step. The symbol “\( \odot \)” represents the point-wise multiplication operation.

4.2.3. Transformer

The third candidate, the transformer, also incorporates information from multiple time points in past history. Even though transformers and LSTMs may share similar high-level structures (Figure 4c,d), there exists a major difference between them. Transformers completely remove the concept of recurrence, or in other words, the way that LSTMs comprehend data as an ordered sequence that incrementally builds up information. They instead rely on two methods called positional encoding and self-attention to handle long series of data. The former means that although there is no longer a concept of “who comes before whom and after whom”, the position of each data point is nevertheless encoded so that transformers do have spatial awareness. The latter method helps transformers to figure out “which data points” (or specifically, data points from “which positions”) are more important and “pay more attention” to them, so that they can manage long sequence of data.

The fundamental building block of a transformer is shown in Figure 5c. All components should already be self-explanatory, except for the layer normalization blocks. The purpose of such blocks is to re-scale the output values from the previous layer in order to balance the data, prevent the model from being overwhelmed by extreme outliers, and prevent such outliers from propagating and deteriorating progressively.

4.3. Neural Network Implementation Details

Data was divided into train, validation and test sets at the ratio of 8:1:1, where the training set was used to train the model by updating model parameters according to the loss; the validation set
was left for sanity checking each time the entire training set has been exhausted; and the test set was held out for the quantitative evaluation that yielded the results in Figures 1 and 2.

For all three candidates, model parameters were initialized as zeros and were iteratively optimized by back-propagation of loss (i.e., amount of difference between the model prediction and the GPP ground truth as determined by a loss function) via gradient descent. The loss function implemented for all candidates was mean square error loss (MSE loss). Adaptive learning rate (i.e., adjust model parameters at a larger magnitude at early stages of training, and gradually decrease that magnitude when the model reaches the fine-tuning stage) and early stopping (i.e., stop training when the model performance on the validation set no longer improves) were both implemented to prevent over-fitting (i.e., the phenomenon that the model performs great on the training set, most likely by “memorizing” the answers, but fails to generalize on the test set) and improve performance on the test set. Random seed was controlled to secure reproducibility: for example, the pseudo-randomized order at which each training sample was visited would be kept the same each time we ran the experiments.

Other more technical and less inspirational hyper-parameters, such as learning rate, batch size, etc., will not be mentioned in this paper, but can be found in the GitHub repository https://github.com/RosalieZhu/DL-CO2.

5. Conclusions

Greenhouse gas emissions such CO$_2$ could dramatically warm up climate system via positive radiative forcing effect. Fortunately, terrestrial ecosystems are able to mitigate the anthropogenic CO$_2$ emissions via photosynthesis. In this study, we aim to model the dynamics of plant photosynthesis activity using advanced machine learning frameworks with temporal awareness. Our results showed that incorporation of past information is important for improving model predictability. Our memory-based neural network model was able to successfully capture the temporal dynamics of plant photosynthesis at all four sites of interest, and it identified shortwave radiation as the most informative climate driver for faithful prediction of photosynthetic rates. Our modeling experiment also demonstrated a two-week internal system memory of the photosynthesis machinery.

Our study demonstrated that LSTM, with its capability of capturing temporal dependencies, is a technique worth investigating when researches study time series data regarding photosynthetic activities, or even environmental data in general. In the future, similar methods can be directly applied to evaluate other important climate drivers besides GPP, such as ecosystem respiration and net ecosystem exchange. Specifically, a more comprehensive study on mapping photosynthesis internal memory can be conducted if more data from additional FLUXNET sites that cover all land types are incorporated, and ultimately an internal memory map can be derived based on the prediction of the proposed LSTM network.

Author Contributions: conceptualization, N.Z.; methodology, N.Z.; software, N.Z. and C.L.; validation, N.Z. and C.L.; formal analysis, N.Z. and C.L.; investigation, N.Z. and C.L.; resources, A.F.L. and J.G.; data curation, N.Z.; writing—original draft preparation, N.Z. and C.L.; writing—review and editing, N.Z., C.L., A.F.L. and J.G.; visualization, N.Z. and C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This study is conducted with the aid of the computational resources at the Mortimer B. Zuckerman Mind Brain Behavior Institute.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:
ANOVA analysis of variance
ATP adenosine triphosphate
CO₂ carbon dioxide
FFNN feedforward neural network
FLDS longwave radiation
FSDS shortwave/solar radiation
GPP gross primary product
HMM hidden Markov model
HSD honestly significant difference
LSTM long short-term memory
MSE mean square error
NADPH nicotinamide adenine dinucleotide phosphate hydrogen
Pre precipitation
RNN recurrent neural network
RuBisCO enzyme Ribulose-1,5-bisphosphate carboxylase oxygenase
VPD vapor pressure deficit

References
1. Stips, A.; Macias, D.; Coughlan, C.; Garcia-Gorriz, E.; San Liang, X. On the causal structure between CO₂ and global temperature. *Sci. Rep.* 2016, 6, 21691.
2. Falkowski, P.; Scholes, R.; Boyle, E.; Canadell, J.; Canfield, D.; Elser, J.; Gruber, N.; Hibbard, K.; Högberg, P.; Linder, S.; et al. The global carbon cycle: A test of our knowledge of earth as a system. *Science* 2000, 290, 291–296.
3. Beer, C.; Reichstein, M.; Tomelleri, E.; Ciais, P.; Jung, M.; Carvalhais, N.; Rödenbeck, C.; Arain, M.A.; Baldocchi, D.; Bonan, G.B.; et al. Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate. *Science* 2010, 329, 834–838.
4. Friedlingstein, P.; Houghton, R.; Marland, G.; Hackler, J.; Boden, T.A.; Conway, T.; Canadell, J.; Raupach, M.; Ciais, P.; Le Quéré, C. Update on CO₂ emissions. *Nat. Geosci.* 2010, 3, 811–812.
5. Baldocchi, D.; Falge, E.; Gu, L.; Olson, R.; Hollinger, D.; Running, S.; Anthoni, P.; Bernhofer, C.; Davis, K.; Evans, R.; et al. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Am. Meteorol. Soc.* 2001, 82, 2415–2434.
6. Medlyn, B.E. Physiological basis of the light use efficiency model. *Tree Physiol.* 1998, 18, 167–176.
7. Farquhar, G.D.; von Caemmerer, S.V.; Berry, J.A. A biochemical model of photosynthetic CO₂ assimilation in leaves of C3 species. *Planta* 1980, 149, 78–90.
8. Papale, D.; Valentini, R. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. *Glob. Chang. Biol.* 2003, 9, 525–535.
9. Tramontana, G.; Ichii, K.; Camps-Valls, G.; Tomelleri, E.; Papale, D. Uncertainty analysis of gross primary production upscaling using Random Forests, remote sensing and eddy covariance data. *Remote Sens. Environ.* 2015, 168, 360–373.
10. Jung, M.; Reichstein, M.; Bondeau, A. Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model. *Biogeosciences* 2009, 6, 2001–2013.
11. Xiao, J.; Zhuang, Q.; Baldocchi, D.D.; Law, B.E.; Richardson, A.D.; Chen, J.; Oren, R.; Starr, G.; Noormets, A.; Ma, S.; et al. Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data. *Agric. Forest Meteorol.* 2008, 148, 1827–1847.
12. Xiao, J.; Zhuang, Q.; Law, B.E.; Chen, J.; Baldocchi, D.D.; Cook, D.R.; Oren, R.; Richardson, A.D.; Wharton, S.; Ma, S.; et al. A continuous measure of gross primary production for the conterminous United States derived from MODIS and AmeriFlux data. *Remote Sens. Environ.* 2010, 114, 576–591.
13. Yang, F.; White, M.A.; Michaelis, A.R.; Ichii, K.; Hashimoto, H.; Votava, P.; Zhu, A.X.; Nemani, R.R. Prediction of continental-scale evapotranspiration by combining MODIS and AmeriFlux data through support vector machine. *IEEE Trans. Geosci. Remote Sens.* 2006, 44, 3452–3461.
14. Yang, F.; Ichii, K.; White, M.A.; Hashimoto, H.; Michaelis, A.R.; Votava, P.; Zhu, A.X.; Huete, A.; Running, S.W.; Nemani, R.R. Developing a continental-scale measure of gross primary production by
combining MODIS and AmeriFlux data through Support Vector Machine approach. Remote Sens. Environ. 2007, 110, 109–122.

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

16. Jung, M.; Reichstein, M.; Margolis, H.A.; Cescatti, A.; Richardson, A.D.; Arain, M.A.; Arneth, A.; Bernhofer, C.; Bonal, D.; Chen, J.; et al. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. J. Geophys. Res.: Biogeosci. 2011, 116, doi:10.1029/2010JG001566.

17. Kondo, M.; Ichii, K.; Takagi, H.; Sasakawa, M. Comparison of the data-driven top-down and bottom-up global terrestrial CO₂ exchanges: GOSAT CO₂ inversion and empirical eddy flux upscaling. J. Geophys. Res.: Biogeosci. 2015, 120, 1226–1245.

18. Schwalm, C.R.; Williams, C.A.; Schaefer, K.; Arneth, A.; Bonal, D.; Buchmann, N.; Chen, J.; Law, B.E.; Lindroth, A.; Luyssaert, S.; et al. Assimilation exceeds respiration sensitivity to drought: A FLUXNET synthesis. Glob. Chang. Biol. 2010, 16, 657–670.

19. Schwalm, C.R.; Williams, C.A.; Schaefer, K.; Baldocchi, D.; Black, T.A.; Goldstein, A.H.; Law, B.E.; Oechel, W.C.; Scott, R.L. Reduction in carbon uptake during turn of the century drought in western North America. Nat. Geosci. 2012, 5, 551–556.

20. Dou, X.; Yang, Y.; Luo, J. Estimating Forest Carbon Fluxes Using Machine Learning Techniques Based on Eddy Covariance Measurements. Sustainability 2018, 10, 203.

21. Tukey J.W. Comparing individual means in the analysis of variance. In Biometrics; pp. 99–114.

22. Hochreiter, S.; Schmidhuber, J. Long short-term memory. Neural Comput. 1997, 9, 1735–1780.

23. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Advances in Neural Information Processing Systems: Cambridge, MA, USA, 2017; pp. 5998–6008.

24. Wang, Q.; Li, B.; Xiao, T.; Zhu, J.; Li, C.; Wong, D.F.; Chao, L.S. Learning deep transformer models for machine translation. arXiv 2019, arXiv:1906.01787.

25. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv 2018, arXiv:1810.04805.

26. Burba, G. Eddy Covariance Method for Scientific, Industrial, Agricultural and Regulatory Applications: A Field Book on Measuring Ecosystem Gas Exchange and Areal Emission Rates; LI-Cor Biosciences: Lincoln, NE, USA, 2013.

27. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10): Haifa, Israel, 21–24 June 2010; pp. 807–814.