Understanding and modeling climate impacts on ecosystem dynamics with FLUXNET data and artificial intelligence

NY Zhu\textsuperscript{1}, XL Yu\textsuperscript{2}, SR Zhang\textsuperscript{3}, ZS Liu\textsuperscript{4}, YW Tong\textsuperscript{5}

1 Graduate study of art and science, Columbia University, New York City, USA
2 School of Science, Beijing University of Posts and Telecommunications, Beijing, China
3 School of Biological Science and Medical Engineering, Southeast University, Nanjing, Jiangsu, China
4 College of Agriculture, Shandong Agricultural University, Taian, Shandong, China
5 School of economics and management, Nanchang University, Nanchang, China
E-mail: nanyan.zhu@columbia.edu

Abstract. Since preindustrial era, the radiative energy balance of the earth system has been largely perturbed by anthropogenic activities such as CO\textsubscript{2} emissions from fossil fuel burning. As a net effect, global temperature increasingly warms up and will further increase in the future if CO\textsubscript{2} concentration in the atmosphere keeps going up. Plants sequester a large amount of atmospheric CO\textsubscript{2} via photosynthesis, thus greatly mediate the global warming. In this study, we aim to model the temporal dynamics of photosynthesis for various different vegetation types and further understand controlling factors of photosynthesis machinery. Our results showed that the photosynthesis and its interactions with climate drivers, such as temperature, precipitation, radiation, and vapor pressure deficit, has an internal system memory about 14 days. Thus, the predictive model could be best trained with historical data of the past two weeks and could best predict future temporal evolution of photosynthesis in the following two weeks. Our leave-one-out experiment also showed that temperature and solar radiation dramatically control grassland and forest photosynthesis activity.

1. Introduction
Climate change and global warming have been threatening the development and stability of both natural ecosystems and human society [1]. The dynamics of greenhouse gases, such as CO\textsubscript{2}, play pivotal roles in controlling the radiative forcing and the energy balance of the whole earth system [2]. Therefore, understanding the full cycle (sources and sinks) of atmospheric CO\textsubscript{2} is critically important in order to better control the growth of CO\textsubscript{2} concentrations in the future. Nowadays, photosynthesis pathway is the largest land surface CO\textsubscript{2} sink that sequester atmospheric CO\textsubscript{2} into vegetation biomass and store them as living biomass. The magnitude of photosynthetic C sequestration pathway is around 120-130 Pg C per year (1Pg = 10\textsuperscript{15} g), globally [3]. Ecosystem respiration, including autotrophic and heterotrophic, consumes most of the photosynthetic carbon sink, cycles it back into atmosphere, and leading to 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\textsubscript{2} concentration and mitigate the anthropogenic CO\textsubscript{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\textsubscript{2} concentrations, radiative balance of earth system, and global warming, it is critically
important to gain mechanistic understand of photosynthesis machinery and its relationship to climate factors such as temperature, precipitation, and radiation. Observational network has been established globally to measure and understand various different components of land surface carbon cycle including photosynthesis [5]. At leaf scale, photosynthesis reaction is carried out by RUBISCO enzyme, which combines CO2 and water molecules to generate carbohydrate products. It occurs at two stages: 1) first light-dependent reactions to capture the energy of light store in ATP and NADPH; 2) second, light-independent reactions capture and reduce carbon dioxide. This biological reaction relies on substrate concentration (CO2 and water), activity of temperature-sensitive RUBISCO enzyme, and driven by solar energy. Thus, theoretically, one is able to build an effective photosynthesis model that takes all those important factors into account and predict 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, photosynthesis rate could also be modeled with data-driven machine learning models that are trained by a large amount of observed photosynthesis data. The latter approach is gaining popularity because the growth of observational network of photosynthesis such as FLUXNET and the advancing of effective machine learning techniques. In this study, the objective is to model the temporal dynamics of plant photosynthesis with two different artificial neural network architectures. With appropriate design model experiments, we also aim to explore and understand the controlling of different climate drivers on photosynthesis.

2. Methodology

2.1. FLUXNET Data

We use Eddy Covariance data collected from four different sites from FLUXNET observational network [5]. It measures plant Gross Primary Productivity (GPP, hereafter we refer photosynthesis to be GPP) and micrometeorological measurements of temperature, precipitation, radiation, vapor pressure deficit. At present, hundreds of sites are operating on a long-term and continuous basis. In this study, we include both boreal and temperate forested sites and grass sites (Table 1). Figure 1 shows the probability density distribution of GPP and climate drivers at the four sites (Red: DE-Tha, Green: NL-Loo, Blue: AT-Neu, Orange: US-Var).

Table 1. Site information of four FLUXNET towers

| Name   | Latitude (N) | Longitude (E) | Elevation (m) | Land cover type          | MAP (mm$^1$ yr$^{-1}$) | MAT (°C) | GPP (gC day$^{-1}$) |
|--------|--------------|---------------|--------------|--------------------------|-------------------------|----------|---------------------|
| DE-Tha | 50.96        | 13.56         | 380          | Boreal Evergreen Needleleaf tree | 334±814                 | 6.8±8.2  | 6±6                 |
| NL-Loo | 52.16        | 5.74          | 25           | Temperate Evergreen Needleleaf tree | 420±988                 | 8.8±7.9  | 5±4                 |
| AT-Neu | 47.11        | 11.31         | 970          | Grasslands               | 419±827                 | 10.1±6.4 | 4±3                 |
| US-Var | 38.41        | -120.95       | 129          | Grasslands               | 282±980                 | 15.8±6.8 | 2±3                 |
2.2. Artificial Intelligence modeling frameworks

We first build an fully connected artificial neural network that has three hidden layers and ten nodes each layer to model the temporal dynamics of GPP using temperature, precipitation, radiation, and vapor pressure deficit. Eqn. 1 denotes the general mathematical design of each hidden layer neuron.

\[ y = f(\sum_{i=1}^{n} x_i \cdot w_i) \]  

(1)

where \( f \) is rectifier activation function, \( W_i \) are neural network parameters for each input feature. The input features \( (x_i) \) represent those climate drivers; and the target output \( (y_i) \) is GPP.

Secondly, we establish a memory based Recurrent Neural Network (RNN) (Figure 2) to model GPP based on not only climate drivers but also the historical memory of the GPP and climate drivers.

A implementation structure of RNN is denoted by Eqn. 2 - 7 (Figure 2). Given an input sequence \( x = (x_1, \ldots, x_T) \) a standard recurrent neural network computes the hidden vector sequence \( h = (h_1, \ldots, h_T) \) and output vector sequence \( y = (y_1, \ldots, y_T) \) by iterating the following equations from \( t = 1 \) to \( T \):

\[ h_t = H(W_{hi} x_t + W_{hh} h_{t-1} + b_h) \]  

(2)

\[ y_t = W_{ho} h_t + b_o \]  

(3)

\[ i_t = \sigma(W_{xi} x_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i) \]  

(4)
\[ c_t = f_t c_{t-1} + i_t \tanh(W_{xc} x_t + W_{hc} h_{t-1} + b_c) \]  
\[ o_t = \sigma(W_{xo} x_t + W_{ho} h_t - 1 + W_{co} c_t + b_o) \]  
\[ h_t = o_t \tanh(c_t) \]

where \( \sigma \) is the logistic sigmoid function, and \( i, f, o, \) and \( c \) are respectively the input gate, forget gate, output gate, and cell activation vectors. \( W \) and \( x \) are model parameters and input at each time step.

3. Results and discussion

We first used fully connected Artificial Neural Network (ANN) to model the dynamics of GPP across the four different sites (trained independently). We found that the model performance at a forest site was better than at a grassland site in general (Figure 3). Furthermore, our leave-one-out experiment showed that removing temperature from input features dramatically degraded model performance the Flux AT-Neu grass site, which suggested a significant temperature control at this site; while removing shortwave radiation from input features at another forest site (FLUX DE-Tha) led to a much worse model performance suggested a significant energy control. ANN model was generally successful at three sites but failed at FLUX US-Var grassland site (Figure 3 right panel, red bars mean negative correlation).

Secondly, we employed Recurrent Neural Network (RNN) to model the dynamics of GPP and conducted a leave-one-out experiment for climate drivers (similar to ANN modeling). Compared with ANN, our RNN model predicted a consistently low mean square error and high model-data correlation (Figure 3 and 4). At forested sites (green and blue bars), temperature, vapor pressure deficit and shortwave radiation played a more important role in controlling GPP; while at grassland sites (orange and red bars) temperature significantly affect the performance of RNN.
Figure 4. RNN model performance of predicted mean square error (left) and prediction-observation correlation (right) at four FLUXNET towers with different combination of climate drivers.

Given that RNN model performed much better than ANN model, we then conducted experiment with RNN model to explore the possible internal memory of the photosynthesis process itself. Our hypothesis is that time series is often self-explained to some extent, which corresponding to the dependency of future state on historical memory of the system. An optimal length of memory is helpful to understand the internal linkage of photosynthesis process across time. Our experiment implied that photosynthesis has an internal memory length of two weeks (Figure 5). Feeding historical memory longer than two weeks into RNN resulted a worse performance compared with 14 days memory length.

Figure 5. RNN model performance of predicted mean square error (left) and prediction-observation correlation (right) at four FLUXNET towers with different lengths of historical memories (from 7 days to 60 days) that RNN considers.

Finally, predictive length of photosynthesis process is critically important to understand the reliability of RNN model in the future time step. We further conducted a modeling experiment that predicted different lengths of future time step from 1 day to up to 30 days. Figure 6 revealed that our RNN modeling framework was able to predict future GPP in two weeks at most. Predictions that were longer than two weeks had large bias and small model-date correlation. Our results on the optimal
historical system memory was quite consistent with the model longest temporal predictability, which were both about two weeks.

4. Conclusion
Green house 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 with advanced machine learning frameworks. Our results showed that system memory is critically important for improving model predictability especially at grassland site. Our memory-based neural network model was able to successfully capture the temporal dynamics of plant photosynthesis at all four sites of interest. Our modeling experiment also demonstrated a clear internal system memory of the photosynthesis machinery.

5. Reference:
[1] Stips, A., et al. (2016). On the causal structure between CO$_2$ and global temperature. Scientific reports, 6, 21691.
[2] Falkowski, P. et al. (2000). The global carbon cycle: a test of our knowledge of earth as a system. science, 290(5490), 291-296.
[3] Beer, C et al., (2010). Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. Science, 1184984.
[4] Friedlingstein, P et al., (2010). Update on CO2 emissions, Nat. Geosci., 3, 811–812
[5] Baldocchi, D et al., (2001). FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bulletin of the American Meteorological Society, 82(11), 2415-2434.
[6] Medlyn, B. E. (1998). Physiological basis of the light use efficiency model. Tree physiology, 18(3), 167-176.
[7] Farquhar, G. V. et al., (1980). A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. Planta, 149(1), 78-90.