



RESEARCH ARTICLE

10.1029/2018JG004791

Key Points:

- A novel neural network C uptake model improves predictive skills and is a key step toward complete spatiotemporal land C uptake modeling
- Comparing the neural network and the VPRM errors reveals promising future model development and ecosystem science research directions
- Better representing respiration and temperature dependence are particularly promising future model development directions

Correspondence to:

G. Eshel,
geshel@gmail.com

Citation:

Eshel, G., Dayalu, A., Wofsy, S. C. C., Munger, J. W., & Tziperman, E. (2019). Listening to the forest: An artificial neural network-based model of carbon uptake at Harvard Forest. *Journal of Geophysical Research: Biogeosciences*, 124, 461–478. <https://doi.org/10.1029/2018JG004791>

Received 5 SEP 2018

Accepted 31 DEC 2018

Accepted article online 15 JAN 2019

Published online 2 MAR 2019

Listening to the Forest: An Artificial Neural Network-Based Model of Carbon Uptake at Harvard Forest

Gidon Eshel¹ , Archana Dayalu² , Steven C. Wofsy² , J. William Munger² , and Eli Tziperman²

¹Physics Department, Bard College, Annandale-on-Hudson, NY, USA, ²Department of Earth and Planetary Sciences and School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

Abstract The terrestrial biosphere strongly modulates atmospheric CO₂ mixing ratios, whose inexorable rise propels anthropogenic climate change. Modeling and mechanistically understanding C uptake by the terrestrial biosphere are thus of broad societal concerns. Yet despite considerable progress, scaling up point observations to landscape and larger scales continues to frustrate analyses of the anthropogenically perturbed global C cycle. While that up-scaling is our overarching motivation, here we focus on one of its elements, modeling C uptake at a given site. We devise a novel artificial neural network (ANN)-based model of C uptake at Harvard Forest that combines locally observed and remotely sensed variables. Most of our model predictors are those used by an established ecosystem C uptake model, the Vegetation Photosynthesis and Respiration Model (VPRM), easing comparisons. To those, we add observed cumulative antecedent precipitation and soil temperature. We find that model errors are much larger in winter, indicating that better understanding and modeling of respiration will likely discernibly improve model performance. Comparing the ANN and VPRM results reveals errors attributed to unrealistic treatment of temperature in the VPRM formulation, indicating that better representation of temperature dependencies is also likely to enhance model skill. By judiciously comparing VPRM and ANN errors we thus overcome ANNs' notoriety for concealing the mechanisms underlying their predictive skills. We demonstrate their ability to identify outstanding ecosystem science knowledge gaps and particularly fruitful corresponding model development directions, improving site specific and up-scaling flux modeling and understanding of the climate impacts of the northern forest.

Plain Language Summary Anthropogenic climate change due to atmospheric carbon dioxide (CO₂) buildup reflecting imbalances between emissions and uptake is a key challenge. Unfortunately, understanding photosynthetic C uptake by land biomes, our focus, is incomplete. The central source of insights into this uptake is a network of exchange measuring towers. However, this sparse network undersamples the ecosystems they strive to represent. Better understanding land C uptake thus depends on vegetation models that can upscale local C uptake to regional and global scales. To this end, we develop an artificial neural network (ANN) model of C uptake in a northern mixed forest measurement site in Massachusetts which reduces errors by the equivalent of the emissions of 5.5 million Americans. This ANN can be readily applied to any observed terrestrial C uptake record at a particular biome, or to all simultaneously. Coupling it to a geographically explicit model for filling spacetime gaps thus achieves the quest for a spatiotemporally complete land C uptake modeling. Our results can also guide future model development by identifying error sources. For example, we identify the representation of respiration and temperature dependence as promising avenues for future research, demonstrating the unique role of ANNs in improving land C uptake modeling.

1. Introduction

Roughly fourfold larger than the atmospheric carbon (C) pool (Schlesinger & Bernhardt, 2013), the global terrestrial biosphere pool strongly influences atmospheric CO₂ concentration (Lovenduski & Bonan, 2017). In recent decades, its net uptake of atmospheric carbon—net ecosystem exchange or NEE, whose long-term integral is the total (live + dead, above+below ground) biomass production in the limit of no dissolved C leaching (Kirschbaum et al., 2001; Wehr et al., 2016)—has offset a quarter to a third of global anthropogenic

©2019. The Authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

emissions (Fernandez-Martinez et al., 2017; Hilton et al., 2013; Keenan et al., 2016). NEE is thus a central element of the anthropogenically perturbed C cycle (Hilton et al., 2013; Tian et al., 2016) that requires robust observations and skillful modeling for improved understanding and projections of future atmospheric CO₂ levels and their climate consequences (Bloom et al., 2016; Schimel et al., 2015), arguably the central geophysical forecast of our time (Harper & Snowden, 2017).

Developing such understanding and forecasting capabilities poses several challenges. First, NEE is the generally small imbalance between Gross Primary Production (GPP, photosynthetic plant uptake) and respiration R , each at least tenfold larger than their residual. NEE observations thus contain considerable noise (Oren et al., 2006), and may not necessarily improve understanding of GPP and R individually or of their responses to environmental stimuli. A key example we analyze later involves temperature dependence. Except during extreme temperatures of either sign, upward C flux increases with rising temperature due to faster respiration. Yet high temperatures tend to roughly coincide with local noon, when photosynthetically active radiation (PAR) too often peaks, yielding a strong positive temperature-PAR covariance (we show this quantitatively and discuss it in later sections). With both of these large, approximately mutually canceling fluxes rising with temperature, their individual effects are masked in the time series of their smaller, noisier sum. Nonetheless, sufficiently large data sets, as the one we use here, sample enough exceptions to the above covariability—for example, hot, cloudy days—to unmask individual effects.

The second challenge stems from the fact that because CO₂ is mostly well mixed in the global atmosphere, the relevant variations in atmospheric CO₂ are driven by $\iint_{A'} \text{NEE}(x, y, t) dA$, NEE integrated over a regional to larger scale area A' (where x , y , and t denote longitude, latitude, and time, and A is general area, of which A' is a specific case of interest). Yet NEE varies widely in space (Bloom et al., 2016; Oren et al., 2006; Xia et al., 2015), and our main NEE observational tool—the spatially heterogeneous ground-based array of eddy covariance CO₂ flux measurements, most recently known as FLUXNET (Baldocchi et al., 2001; Buchmann & Schulze, 2003; Chu et al., 2017; Jagermeyr et al., 2013; Pastorello et al., 2017)—undersamples this variability in both space and time (Law et al., 2001). With their integrand imperfectly known locally and undersampled spatially, observational estimates of such integrals are therefore still uncertain.

These challenges highlight the need for a flexible and general NEE modeling framework that uses surface flux observations to capture C flux dependence on local environmental variables, and utilizes such spatially resolved information as satellite remote sensing to characterize patterns of spatiotemporal variability. Key to such a system are numerical tractability (in contrast with computationally demanding mechanistic ecosystem models) and predictive skills that match or exceed those of existing widely used models. Following and updating Papale and Valentini (2003) and methodologically expanding Moffat et al. (2010), Keenan et al. (2012), and Albert et al. (2017), here we set out to develop a flexible Artificial Neural Network (ANN, Anderson, 1972; Hassoun, 1995)-based time series modeling framework (Khashei & Bijari, 2010). While our overall objective is to use the developed methodology to improve estimates of the requisite regional to continental scale $\text{NEE}(x, y, t)$ integrand and forecasts of its future trajectory, in this paper we lay the methodological foundation by applying ANNs (Cromp & Crook, 1991; Jain & Kumar, 2007; Malmgren & Nordlund, 1997; van der Baan & Jutten, 2000) to a single (scalar) NEE time series measured by the eddy covariance CO₂ flux tower at Harvard Forest (HF, Goulden et al., 1996a, 1996b; Munger & Wofsy, 2017; Urbanski et al., 2007; Wofsy et al., 1993), in central Massachusetts. Beyond the specifics of this site, the devised ANN allows us to pose and answer the following questions. Can a well-configured ANN improve predictive skills beyond what is typical of current simple models? Can the lack of mechanistic underpinning characteristic of ANNs be overcome, illuminating prediction error sources and guiding future model development efforts?

Because predictive skill is key to the work presented here, we devote considerable effort to quantitative skill evaluation against a comparative benchmark. As this benchmark, here we use the Vegetation Photosynthesis and Respiration Model (VPRM, Mahadevan et al., 2008), which employs empirically derived functional dependencies to calculate NEE (Luus et al., 2017) from locally observed air temperature T and PAR, and remotely sensed solar reflectivity at four narrow bands: red (620–670 nm), blue (459–479 nm), near infrared (NIR, 841–876 nm), and shortwave infrared (SWIR, 1628–1652 nm). The choice is based on the considerable

similarity, in use and spirit, between the VPRM and ANNs. Both model types can be used to analyze a single location scalar NEE time series and forecast later values there (although the VPRM was not devised for this purpose); both can combine the same locally measured and remotely sensed environmental data as predictors, and both models are empirical, with their parameters subject to optimization by error minimization. These similarities make the VPRM a natural choice for the head-to-head skill comparison essential to the current paper. We favor the VPRM over its simpler Model-0 predecessor (Urbanski et al., 2007)—which also enjoys the above appealing attributes—because of the VPRM's richer, process-based analytic formulation, in which several multiplicative terms strive to treat each of the main rate limiting productivity constraints in analytic isolation.

Despite the above attractive attributes, the VPRM is not without limitations for our purposes. Most importantly, reproducing or forecasting individual station NEE observations is not the key impetus for the development of the VPRM. Rather, it was conceived as a tool for generating prior state estimates required by an inverse analysis-based data assimilation framework that describes the spatiotemporal C flux patterns by utilizing all available data while still simple enough to optimize against atmospheric CO₂ observations (e.g., Dayalu et al., 2017; Matross et al., 2006). As such, the VPRM emphasizes accurate representation of mean values over coarse spatiotemporal scales, not perfectly representing hourly exchanges at a single location.

The above attributes and caveats jointly shape the role the VPRM plays in this paper. First, it is a yardstick that provides a context for the predictive skills of the new ANN-based methodology, not the reigning, hereby dethroned standard bearer. Second, the VPRM serves as an analytic tool with which to probe the system by analyzing NEE sensitivities to various environmental variables and comparing them to those of the novel ANN (a brief description of ANNs in general and the specific one we use here, as well as of the predictor input variables is given below in section 2.2). Occasionally, we also identify situations in which the VPRM falls short in its representation of various physical processes, in the hope that these will guide future efforts to develop and modify the original VPRM into a tool for site specific forecasting and mechanistic analysis of observed NEE or GPP records.

Despite single site forecasting not being the main impetus behind its development, the VPRM reproduces extremely well-observed NEE in such calibration sites as the HF and an irrigated corn field near Mead, Nebraska (Mahadevan et al., 2008). Consistent with its deliberate simplicity, the VPRM also performs distinctly less well in other sites. For example, the representation of the upward C flux due to respiration is

$$R = \alpha \max(T, T_{\text{low}}) + \beta, \quad (1)$$

where T is the above canopy air temperature and α , β , and T_{low} are locally tunable constants, with $T_{\text{low}} \approx 1\text{--}5^\circ\text{C}$ a temperature floor that accounts for the observation that soils remain warm enough to permit aqueous biochemical reactions even when $T_{\text{air}} \ll 0^\circ\text{C}$. While such representation is a reasonable choice for normally encountered temperatures, it probably needs further refinements to adequately handle more extreme temperatures of either sign.

The VPRM is also limited by its fixed functional form, which is likely too restrictive to robustly represent the widely variable limiting factors (Wieder et al., 2015) and dependencies on environmental conditions characteristic of the $\approx 150^\circ$ meridional span of the photosynthetically active terrestrial biosphere. For example, Mahadevan et al. (2008) indicate that the presence of significant inorganic carbon pools—such as those present in shrubland ecosystems—are outside the scope of a model like the VPRM. From a model development perspective, such wide structural variability requires different sets of truncated governing equations, each with unique and limited spatiotemporal applicability, and is unlikely to be successfully condensed into VPRM's single analytic formula.

2. Methods

2.1. VPRM

The VPRM (Mahadevan et al., 2008) is a widely employed model of biome generic canopy-boundary layer CO₂ exchanges, introduced earlier. It maps a small set of readily available remotely sensed and locally instrumentally obtained environmental measurements (Mahadevan et al., 2008) onto estimates of GPP and respiration carbon fluxes, and of their net balance, NEE (Kirschbaum et al., 2001). The VPRM equation is

$$\begin{aligned} \text{NEE}_{\uparrow} &= -\text{GPP}_{\downarrow} + R_{\uparrow} \\ &= -\lambda T_{\text{scale}} P_{\text{scale}} W_{\text{scale}} \text{EVI} \text{PAR}_f + R_{\uparrow} \\ &= -\lambda T_{\text{scale}} \frac{(1 + \text{LSWI})^2}{2(1 + \text{LSWI}_{\text{max}})} \text{EVI} \frac{\text{PAR}}{1 + \text{PAR}/\text{PAR}_0} + \alpha \max(T, T_{\text{low}}) + \beta. \end{aligned} \quad (2)$$

In equation (2), vertical arrows indicate the direction defined as positive, with actual fluxes in general assuming either sign (such that $\text{NEE} > 0$ indicates CO_2 loss to the boundary layer via upward net flux, and $\text{NEE} < 0$ corresponds to downward flux into or net C uptake by the terrestrial biosphere. The P_{scale} and W_{scale} terms account for leaf presence and state on photosynthesis and water availability, respectively, with formulations given in equations (7) and (8) of Mahadevan et al. (2008). For convenience, we introduce here $\text{PAR}_f^{\text{def}} = \text{PAR}/(1 + \text{PAR}/\text{PAR}_0)$ (which Mahadevan et al., 2008, do not explicitly define) and

$$T_{\text{scale}}^{\text{def}} = \frac{[\max(T, T_{\text{min}}) - T_{\text{min}}](T - T_{\text{max}})}{(T - T_{\text{min}})(T - T_{\text{max}}) - (T - T_{\text{opt}})^2}, \quad (3)$$

where, following Mahadevan et al. (2008), we set $T_{\text{min}} = 0^\circ\text{C}$ and $T_{\text{max}} = 40^\circ\text{C}$. Equation (3) differs in appearance from the original (their equation (6)), because it incorporates (through the numerator's first term) additional constraints that are presented in that paper in text following the equation. The parameters in equations (2) and (3), $\{\lambda, T_{\text{opt}}, \text{PAR}_0, \alpha, \beta, T_{\text{low}}\}$, are optimized for a specific observation site (Mahadevan et al., 2008) as described for the current paper below.

The VPRM inputs are as follows. Above canopy temperature T is instrumentally measured at HF. The LSWI (land surface water index, Xiao et al., 2004) is a function of rNIR and rSWIR, solar reflectivities at 841–876 and 1,628–1,652 nm, from the nearest pixel to HF in the L3 8-day, 500 m MODIS data set (Vermote, 2015), and LSWI_{max} is the growing season maximum LSWI. From the same source, we also use the enhanced vegetation index, EVI (Verhegghen et al., 2014), a function of rNIR, rBLUE, and rRED, where the latter two are solar reflectivities at 459–497 and 620–670 nm, respectively. Finally, PAR is photosynthetically active radiation (0.4–0.7 μm), radiometrically measured at the HF tower and reported hourly. We handle the mismatch between the hourly temporal resolution of the flux tower data and the 8-day resolution of the remotely sensed reflectivities by assuming smooth linear temporal phenological evolution between any two successive remotely sensed data points. While we initially experimented with using spatial aggregates of the 4 and 6 nearest pixels, we found the differences between results obtained with these aggregated time series and those based on the single nearest pixel to be inconsequential. The results reported here are thus based on the least conjectural input, time series of solar reflectivities from the single pixel whose center is nearest to HF.

2.2. Predictors and ANNs

To demonstrate the methodology, evaluate its performance, and compare it to that of the VPRM, we develop a forecasting scheme for the HF scalar NEE time series. In this time series, turbulence-scale measurements are temporally condensed into an hourly time series, available nearly continuously (excluding invalid observed data and measurements taken during calm, low u^* —or frictional velocity—conditions) over March 2000 to December 2015 (the period over which both the tower data and MODIS observations are available) and sampling reasonably uniformly the diurnal and seasonal cycles (Munger & Wofsy, 2017; Urbanski et al., 2007).

We apply to the HF NEE record 255 ANNs with number of predictors N_p ranging over one to eight out of the set of eight possible predictors described shortly. This $N_p \leq 8$ yields the above 255, which is the sum of the eight binomial coefficients “8-choose- N_p ” with $N_p = [1, 8]$. We thus explore all eight 1-predictor ANNs, the single possible 8-predictor ANN, and all possible combinations of ANNs with $N_p = [2, 7]$ predictors in between. Considered ANNs with fewer predictors than the VPRM's six are denoted tANN, for “truncated.” When they depend on all six VPRM predictors, we denote them fANN, for “full.” Finally, when considered ANNs contain novel predictors not included in the VPRM formalism, they are denoted aANN, for “augmented.” While an introduction to ANNs is outside of the current scope, the following brief overview is specifically pertinent to the current problem.

Because our focus is forecasting skills (i.e., the extrapolation problem, not the representation problem of an observed signal; Cybenko, 1989), there is no recipe or a theorem governing the optimality of an ANN-based forecasting model. Any combination of number of hidden layers and node numbers in each is thus poten-

tially interesting but must undergo careful, unbiased cross validation. We thus performed an extensive systematic search in which we rigorously cross validated all ANNs with up to four hidden layers and up to eight nodes apiece. Culminating this process is our choice, ANN(6,5,4), described below. It exhibits the highest systematically reproducible cross-validated forecasting skill, although this skill changes only minimally under small network architecture changes.

An ANN(6,5,4) has three hidden layers with six, five, and four nodes (or neurons), respectively. At each time point, weighted sums (possibly with an added constant called “bias”) of the scalar predictors constitute the input into each of the six nodes of the first hidden layer. The mapping of input into output at each node is calculated by a nonlinear tansig (hyperbolic tangent) activation function (Vogl et al., 1988). Weighted sums of the outputs of these six nodes then feed into each of the five nodes of the intermediate hidden layer, and similarly for the final hidden layer. Using a linear activation function for the output layer, the final output is a weighted sum of the outputs of the four nodes of the final hidden layer. The weights and biases for all nodes are calculated in a training (optimization) step, by requiring the network output to be as close as possible (in a least squares sense) to the observed NEE over all training data, using the Matlab (versions 2017b and 2018b) deep learning toolbox (version 12.0) implementation of the Levenberg-Marquardt back propagation algorithm (Marquardt, 1963). Starting from randomly assigned initial weights and biases, we train the ANN on a randomly chosen half of the available time points of the chosen predictors, about 22,450-hourly observations, and cross validate the ANN’s NEE predictive skill over the remaining $\approx 22,450$ observations. The choice of basing skill or model goodness quantification only on cross-validated results, as well as the exclusive reliance on hundreds of redundant models in the Monte Carlo (MC) formalism described below, reflects our deliberate, judicious efforts to avoid spurious skill due to overfitting. We revisit this point in section 3.

The set of predictor time series we consider starts with the raw inputs into the VPRM. For example, while one of the inputs into the VPRM is LSWI, we use rNIR and rSWIR (of which LSWI is a function) directly rather than the derived index itself, and allow the ANN to determine the functional dependence of NEE on these raw inputs. Our initial set of potential predictor inputs thus include flux tower PAR and T , and remotely sensed rBLUE, rNIR, rRED, and rSWIR observations with no flagged quality issues. The predictors are arranged in a matrix form, with each column representing all predictors at a given time, and each row holding the full time series of a given predictor. It is at least hypothetically possible that some VPRM inputs, while plausibly mechanistically related to NEE and thus expected to offer at least some NEE predictive skill, in fact do not. This possibility is made all the more actionable by the modularity of the ANN formulation, in which adding or removing a model predictor is as simple as adding or removing a predictor matrix row, a level of flexibility the VPRM does not offer because of its fixed functional form. Exploiting this flexibility, we thus test the above six potential predictors for individual utility for NEE prediction, with outcomes reported in section 3.

For aANN, we explore two additional daily resolution predictors that cannot be straightforwardly used in VPRM. The first is soil temperature (Melillo et al., 2011, 2017), which is far more directly pertinent to soil microbiology than is above canopy air temperature, potentially improving the treatment of respiration. Below ground temperature data are available (Melillo et al., 2017) for the nearby Prospect Hill site at HF, which is within the footprint of the flux tower and has been measured consistently for the period of interest. Our T_{soil} predictor is the mean value of the six Prospect Hill control plots.

The second novel predictor we test is cumulative antecedent precipitation, p_{ca} , defined for a specified accumulation time τ by an integral over the precipitation $p(t)$,

$$p_{\text{ca}}(t; \tau) \stackrel{\text{def}}{=} \int_{t-\tau}^t p(t') dt'. \quad (4)$$

We first construct a representative precipitation time series for the HF area by augmenting daily measurements taken at the Prospect Hill meteorological station, 1.6 km from the flux tower (Boose, 2018)—a direct measure of water delivered to the forest in the immediate vicinity of the flux tower—with the nearby NCDC (National Center for Environmental Information, NOAA, 2018) site at Orange Municipal Airport, MA, where valid data sometimes exist when the HF data are unavailable. This combination is guided and justified by the airport record being essentially a redundant realization of the Prospect Hill record, with a 0.91 temporal correlation between the two records over 6,154 overlapping daily precipitation values. Filling data

Table 1
The Vegetation Photosynthesis and Respiration Model Parameters

Parameter	Mahadevan et al. (2008) Estimate	Our estimates				Parameter units
		Monte Carlo, 100 means of $N = 3 \times 10^3$			over all 44,879 observations	
		Standard	Ratio,	Parameter		
		mean	deviation			
λ	0.127	0.221	0.003	1.2	0.221	mol CO ₂ (mol PAR) ⁻¹
T_{opt}	20	23.3	0.14	0.6	23.3	°C
PAR ₀	570	536.9	8.0	1.5	536.8	μmol·m ⁻² ·s ⁻¹
α	0.271	0.267	0.007	2.6	0.267	μmol CO ₂ ·m ⁻² ·(s K) ⁻¹
β	0.25	0.54	0.09	16.8	0.54	μmol CO ₂ ·m ⁻² ·s ⁻¹
T_{low}	5	4.63	0.27	5.7	4.63	°C

Note. The values given by Mahadevan et al. (2008) are reproduced in column 2 (from the left), and the ones estimated here are reported in columns 3–6. Mean and variability of our estimates (columns 3 and 4, with their ratios given in column 5) are calculated over 100 nonlinear least squares fits of the Vegetation Photosynthesis and Respiration Model equation (equation 11 of Mahadevan et al., 2008) to 100 subsets of the full data, each comprising 3,000 randomly chosen hourly data points. Column 6 shows the values that result from a single nonlinear least squares fit of the Vegetation Photosynthesis and Respiration Model equation to the full data set. Following Mahadevan et al. (2008), we set $T_{\text{min}} = 0^\circ\text{C}$ and $T_{\text{max}} = 40^\circ\text{C}$ throughout.

missing from the Prospect Hill record with data from the airport record results in the $p(t)$ time series we use, spanning 1 January 1991 to 10 February 2018 with over 70% of the daily data present.

We construct 121 p_{ca} candidate time series by exploring $\tau = [0, 120]$ days, and evaluate the temporal correlation of each with the time series of NEE at HF. These correlations start near -0.04 at $\tau = 0$ days and steadily and smoothly become more negative until reaching -0.16 at $\tau = 63$ days. Beyond that point (toward longer τ) the correlation magnitude steadily declines as it rises back toward zero. Based on this behavior, our second novel predictor is p_{ca} gotten by setting $\tau = 63$ days.

2.3. Setting the Prediction Standard: Quantifying the VPRM Skill

As formulated by Mahadevan et al. (2008), the VPRM has two specified parameters— $T_{\text{min}} = 0^\circ\text{C}$ and $T_{\text{max}} = 40^\circ\text{C}$ —and six additional parameters, $\{\lambda, T_{\text{opt}}, \text{PAR}_0, \alpha, \beta, T_{\text{low}}\}$, which we calculate by a nonlinear least squares fit of the VPRM equation (equation 11 of Mahadevan et al., 2008) to the HF NEE time series using the direct search Nelder-Mead simplex algorithm (Lagarias et al., 1998).

Table 1 shows the results of deriving these parameters in two ways. First, to reduce sensitivity to potential parameter determinacy issues, we derive 100 sets, each based on nonlinear least squares fit of the VPRM equation (Mahadevan et al., 2008) to a subset of the full NEE data comprising 3×10^3 randomly chosen hourly data points. Second, we repeat the parameter estimation using all available data points together. The first yields the parameter statistics reported in columns 3–5 (from the left) of Table 1, while the results of the second are shown in column 6. Column 3 shows that our parameter values clearly differ somewhat from those reported by Mahadevan et al. (2008), but well within what is expected given that their analysis was based on under 4 years of data, as compared to 16 years here. Column 5 offers a measure of parameter sensitivity to input data, and thus of robustness of the optimized parameter values. The measure is the sample standard deviation s of a given parameter calculated over the 100 randomized fits, expressed as percent of the respective mean estimate m (calculated over the same 100 randomized fits), namely, $100s/m$. Most parameters vary little and are clearly robust, but β is an exception. Yet this too is reasonable given that β is the only additive parameter in the VPRM, thus shouldering a disproportionate portion of the burden of reproducing the mean NEE, which varies considerably among data subsets due to the randomly varying proportions of night and winter data in the total data subset being considered.

2.4. Deriving the Model Specific $\partial\text{NEE}/\partial\text{PAR}$ Estimates of Section 5

In section 5, we derive estimates of $\partial\text{NEE}/\partial\text{PAR}$ (NEE dependence on light availability) for both the ANN model and the VPRM in deep winter (December–February), spring–early summer (mid-April–May), and later summer (June–August), as follows. Because these seasonal partial derivatives depend strongly on tem-

perature, we carry out their quantification on a temperature grid spanning $[-7, 31^{\circ}\text{C}]$ in 2°C increments. We treat each T grid value as the center of 3°C wide T bins (e.g., the two successive bins centered at 1 and 3°C span $[-0.5, 2.5^{\circ}\text{C}]$ and $[1.5, 4.5^{\circ}\text{C}]$, respectively) whose slight overlap ensures more robust statistics and added smoothness.

For each temperature bin at each of the above three considered seasons, we identify all observed hourly data that fall within the season and temperature range. For observed tower NEE data, we devise 100 MC $\partial\text{NEE}/\partial\text{PAR}|_{T',s}$ estimates at each grid value T' and season s . Each is based on randomly choosing 90% of the season/ T bin data (typically comprising several thousand points), using the subset to solve $\text{NEE}_i = \phi_1 + \phi_2\text{PAR}_i$ by least squares within the day and temperature ranges, and record the resultant 100 ϕ_2 values. Note that this is not really the sought partial derivative, but an estimate of it that fails to control for the impact of other variables that jointly determine the observed NEE. Most importantly, because of the considerable diurnal scale (T, PAR) covariance, systematic T differences among the model points may exist, and impact NEE, which the above model is unable to resolve.

For the two models (aANN and the VPRM), we are able to eliminate this ambiguity. For each of these non-linear NEE models, we use each of the 100 MC parameter sets described earlier in this section to simulate the same observed season/ T range specific NEE data subsets discussed above and use the median of those 100 simulated data sets as the modeled NEE for that model (aANN or VPRM). Importantly, the input into the models comprise the actual individual hourly PAR values, but the season/ T bin medians for all other input variables. With this, the input vectors on which all simulated NEE values of a given season/ T in combination depend differ only in PAR values, with all other input variables being the same for all data points in the combination. This results in two simulated NEE data subsets (one for each model) whose internal variability stems exclusively from PAR differences. We then use the same linear model and the same 90% based randomization as above to fit 100 random subsets of these simulated NEE values to corresponding observed PAR values and record the two resultant sets (one for aANN and one for VPRM) of 100 ϕ_2 values per model per T bin/season combination.

3. Results and Discussion

We answer the key question of this paper—whether a predictive NEE model can outperform the VPRM skills—in section 3.2 below. Yet the answer can only be as definitive and persuasive as the ANN-based NEE model from which it is derived. To build a model that incorporates the most skilled predictors while excluding spurious or minimally skillful ones, in the following section we test each of the potential ANN predictors for its relevance to NEE forecasting.

3.1. Individual Predictor Skills

To test the utility of each of the eight potential predictors in the ANN, we use each individually to train and cross validate a single-predictor tANN [an ANN(6,5,4) with one scalar input]. The performance criterion we employ is the mean $\pm$ standard deviation of cross-validated forecast R^2 evaluated over 100 samples of 10^3 data points each, randomly drawn from the validation half. This choice reflects our effort to avoid model overfitting by relying only on cross-validated results. The likelihood of overfitting is further reduced by the above time oblivious randomization protocol, which in general produces a temporally punctuated and discontinuous samples. This reduces contamination by spuriously high explanatory power due to strong subsynoptic temporal autocorrelations most hourly resolution meteorological time series exhibit (Moffat et al., 2010). Finally, parameter dependence on the application of specific numerical methods to specific samples is also reduced by randomization of initial weights and biases.

These tests yield individual NEE predictive skills ranging from 0.44 ± 0.03 for PAR to 0.06 ± 0.01 for rSWIR (leftmost black curve in Figure 1). All six skills are $p \ll 0.05$ significantly above zero, despite the modest NEE predictive skills of trailing reflectivities. Yet some or all of these less individually skillful predictors may add nontrivial predictive power to that of more powerful predictors by governing subspaces of NEE variability that are nearly orthogonal to the NEE variability subspaces the more powerful predictors govern. For example, as shown below, rSWIR is in fact a valuable secondary predictor.

Note that the low predictive contributions of some reflectivities may assist future efforts to improve the performance of the VPRM. For example, LSWI(rNIR, rSWIR) is a key element of P_{scale} and W_{scale} , which represent phenology and water availability, respectively, and which jointly multiplicatively determine VPRM

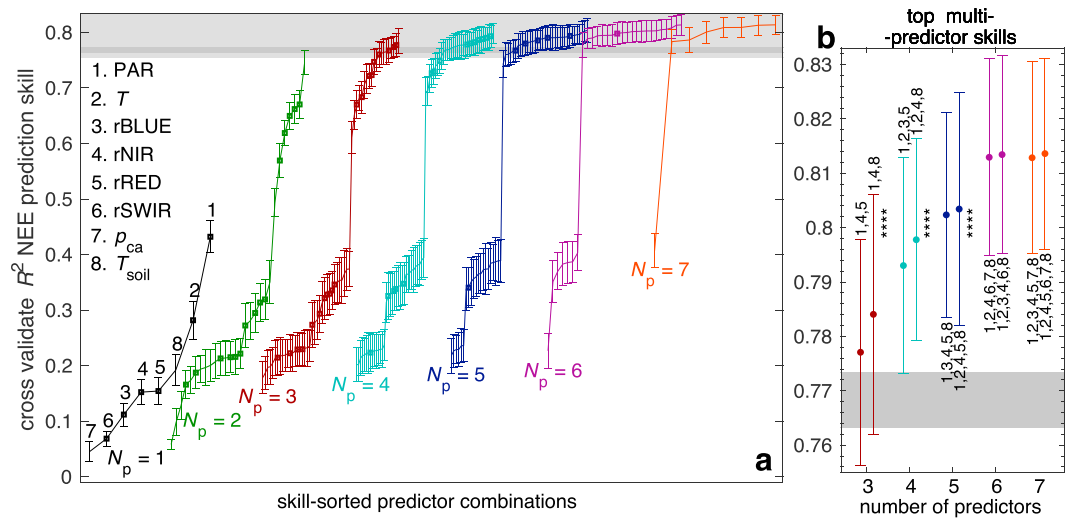


Figure 1. Cross-validated NEE predictive skill of various artificial neural network(6,5,4) models. We explore such models with one predictor ($N_p = 1$, leftmost black curve, with predictors identified numerically along the curve and by name on the upper left), and $N_p = 2-7$ (other color curves). For each N_p value, we consider all possible predictor combinations, and plot the sorted mean NEE predictive skill of each set $\pm$ standard deviations calculated for each model over a set of 100 Monte Carlo realizations, each 10^3 data points long, drawn at random from the validation set that was not used in the optimization. Models (predictor combinations) whose predictors are all members of VPRM's six predictors are highlighted with square symbols. In both panels, dark gray shading around $R^2 \approx 0.77$ shows the cross-validated R^2 skill of the VPRM nonlinear regression estimated by the same randomization. In (a), light gray shading shows the R^2 range of panel (b), which highlights the NEE predictive skill ranges of the two most skillful predictor combinations with three to seven predictors. Panel (b) asterisks indicate the significance of the R^2 difference from the left neighbor, with one to four asterisks denoting $p < 0.05$, 0.01, 0.005, and 0.001. NEE = net ecosystem exchange; PAR = photosynthetically active radiation; NIR = near infrared; SWIR = shortwave infrared.

GPP. Given the above limited contributions of rSWIR, future model development efforts may wish to explore alternatives to P_{scale} and W_{scale} that better resolve early spring C uptake burst following leaf budding and water stress, perhaps using rNIR or previously unutilized bands directly rather than the compound LSWI, or by further developing methods that incorporate remotely sensed solar-induced chlorophyll fluorescence measurements (Commene et al., 2017).

3.2. Performance Comparison Between the VPRM and ANN-Based Models

So can alternative NEE models with predictive skills superior to those of VPRM be devised? Figure 1 answers this question affirmatively for the HF site, and possibly more generally for the mixed forest ecosystems it represents. It shows that NEE predictive skill of ANN models with various predictor combinations, including combinations with fewer predictors than the VPRM's 6, exceed the VPRM skill (whose range is shown in dark gray shading). This point is made most clearly by Figure 1b, which presents the two most skillful models for each given value of the number of predictors N_p . These results are shown for five values of the number of predictors, $N_p = 3-7$, and we note that all 10 top models shown outperform the VPRM's skill (shown by the gray band). In fact, of all possible predictor combinations, a full third of the ANN-based alternatives to the VPRM outperform the VPRM. This demonstrates the outstanding robustness of the high prediction skill of ANN models.

The normalized NEE prediction errors presented in Figure 2 amount to nontrivial errors in estimating C sequestration. A useful reference is the mean NEE (annual net CO_2 -C uptake) at HF, estimated to be roughly 1.5–3 metric ton (mt) C/ha/year by Goulden et al. (1996b) and broadened by the more recent and comprehensive analysis of Urbanski et al. (2007) to 1.0–4.7 with an average of 2.5 mt C/ha/year. Relative to this mean NEE, the shown VPRM error is 62% of this annual uptake, while the six shown ANN errors span 57–59%. In absolute terms, the differences in dimensional errors fall inside 75–115 kg C/ha/year with an approximate mean of 100 kg C/ha/year. If narrowly applied to the areal extent of only the northeastern mixed forest (Olson et al., 2001; roughly 9 million ha), this reduced error amounts to lowering the estimation error of this biome's annual C sequestration by approximately 9×10^5 mt C, which is equivalent to

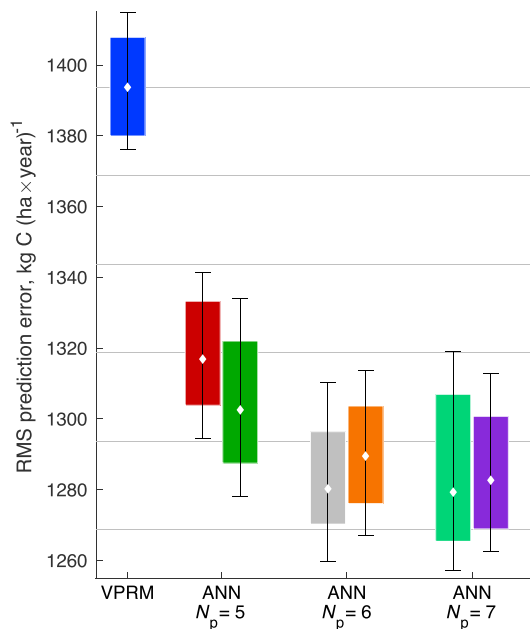


Figure 2. Spread of net ecosystem exchange prediction error by the VPRM (leftmost, blue) and the six top performing ANNs shown in Figure 1b. Colored boxes and whiskers show the central 50% and 70% of the Monte Carlo distributions, with white diamonds presenting medians. As a yardstick, the gray horizontal lines show the median VPRM error minus {0, 25, 50, 75, 100, 125} kg C (ha × year)^{−1}. VPRM = Vegetation Photosynthesis and Respiration Model; ANN = artificial neural network; RMS = root-mean-square.

the annual emissions of approximately 5.5 million Americans (EPA, 2018). Note that because the VPRM is designed as a prior generating tool for inverse analyses, not a predictive model of scalar NEE time series per se, its somewhat larger error may be unique and not generalizable to other process-based models.

Returning briefly to the $N_p = 1$ results introduced earlier, the leftmost (black) curve in Figure 1a shows that PAR and T , which are disproportionately important to determining photosynthetic rates and are thus expected to powerfully control any reasonable NEE model, are indeed the first and second most skillful predictors. PAR also enters as a predictor in each of the $N_p > 1$ VPRM-outperforming ANNs. Yet it is followed not by T , but by rNIR and T_{soil} , which are featured in two thirds of all outperforming models, with rSWIR and p_{ca} the least ubiquitous predictors among ANN models outperforming VPRM. The secondary utility of T as an hourly NEE predictor, and its inferiority to T_{soil} despite their respective individual NEE predictive skills (cross-validated R^2) of 0.28 and 0.19, is less expected and potentially useful for future efforts to enhance the VPRM. One possible explanation of this peculiarity is that the representation of ecosystem respiration in the VPRM (equation (1)) is too simplistic. It is also possible that the above general functional representation is adequate, but would have better predicted NEE if its input temperature were the more directly pertinent to respiration rate T_{soil} , not T . While this possibility is currently of limited practical applicability for regional scale modeling regardless of merit because T_{soil} data coverage is at the moment poor, this may well be gradually rectified by expanding availability and coverage of remotely sensed observations (e.g., Holgate et al., 2016; Jin et al., 2018). A third possibility is that with PAR and T partly marching to the same seasonal and diurnal drums, the information T holds is largely contained within (redundant with) the PAR data. While physically plausible,

these speculations must first be carefully tested, which we hope future efforts to improve the VPRM will take on. Regardless of which of the above scenarios holds under future scrutiny, the ubiquity of T_{soil} in VPRM-outperforming ANN-based models highlights again the general modularity advantage of the ANN modeling framework. With predictors effortlessly added or removed based on selecting and integrating into the ANN model those best representative of a particular ecosystem, model improvements are easier to achieve.

Of the 84 ANN-based VPRM-outperforming models, 14 (17%, highlighted by square symbols in Figure 1a) employ only VPRM predictors. For $N_p = 3$ –6, there are 2, 7, 4, and 1 such models, amounting to 10%, 47%, 67%, and 100% of the respective number of possible combinations out of the VPRM's 6. The improvements over the VPRM performance the ANN formalism offers are thus not a simple artifact of the additional predictors. Rather, they partly reflect the inherent advantage of the ANN modeling framework over the VPRM. At the same time, Figure 1b warns of the rising risks of overfitting and information saturation that typically accompany the exhaustive indiscriminating search for skill among numerous combinations. It shows that while NEE predictive skills rise in concert with N_p over $N_p = 3$ –6 (the differences among like- N_p skill pairs are all minimally significant or insignificant), adding a seventh predictor adds no significant skill.

3.3. Seasonal Dependence and a Mechanistic Interpretation

The results Figure 1 reports have two key limitations. The first is the often cited failure of ANNs to explain mechanistically their predictive skills (Wilby et al., 2002). Second, Figure 1 does not resolve seasonal dependence of the inherently time dependent forest C uptake (Funk & Brown, 2006; Running et al., 2004). Addressing both limitations, we exploit the significant NEE dependence on PAR and T to obtain explicitly seasonally dependent results that offer some indirect mechanistic insights. Figure 3 presents NEE as a function of these two observed predictors for (from the top row down) five time ranges during the year: (1) 27 November to 1 March; (2) 2 March to 10 May; (3) 11 May to 23 August; (4) 24 August to 17 October; and (5) 18 October to 26 November. These time brackets differ somewhat from the eight Urbanski et al. (2007) use because they balance the quest for roughly uniform seasonal data coverage with the need to adequately

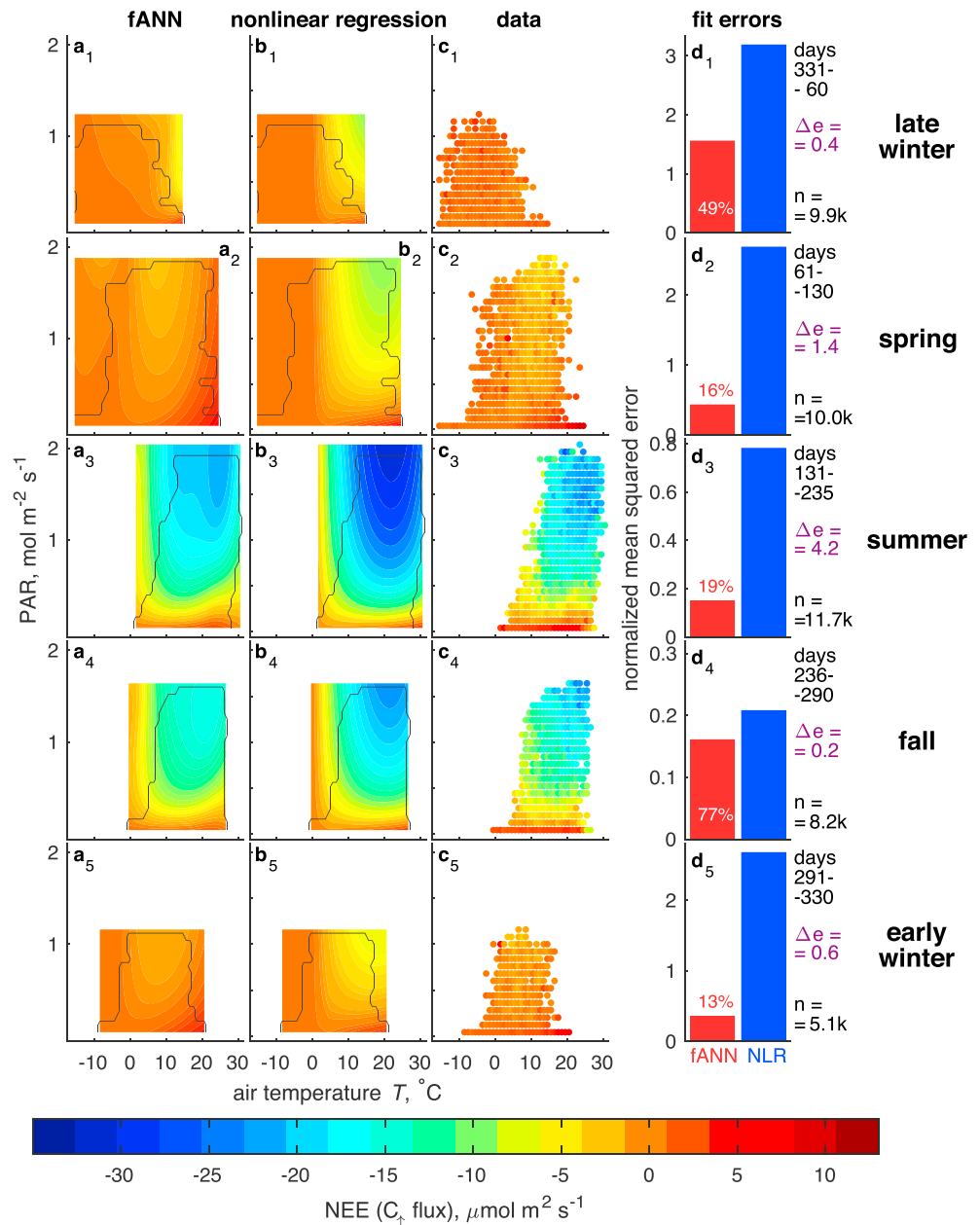


Figure 3. Comparison of NEE dependence on air temperature and PAR over different parts of the seasonal cycle (panels (a)–(c) of each row, all sharing the shown NEE color scale). Panels (a) show this dependence for the ANN(6,5,4) using VPRM's six predictors. NEE calculated by VPRM is shown in panels (b), followed in panels (c) by the same for observed data in (T , PAR) bins populated with observations. The parts of the plane spanned by these populated bins are shown in panels (a)–(c) by a black contour. The parameters of both models are optimized once (globally over all available data spanning the full year), and their inputs are (T , PAR) pairs on the shown uniform grids, combined with seasonal mean rBLUE, rNIR, rRED, and rSWIR values. Panels (d) present mean square errors over all populated bins within the day ranges normalized by the corresponding NEE observations, with the fANN and the VPRM in red and blue. The smaller errors as percentage of the larger ones are indicated (e.g., for the top, deep winter row, the fANN global error is 49% of the VPRM's). PAR = photosynthetically active radiation; ANN = artificial neural network; NEE = net ecosystem exchange; VPRM = Vegetation Photosynthesis and Respiration Model.

resolve the phenological seasonal cycle for the data set analyzed here. Results are shown in panels (a)–(c) of each row for our fANN model using VPRM's six predictors, for the VPRM itself, and as a scatter plot of observed NEE values.

Panels (d) of each row in Figure 3 show the fit errors for ANN and VPRM, demonstrating that the novel ANN-based NEE model outperforms the VPRM at all “seasons.” Rows 1, 2, and 5 (covering mid-October–mid-May) show that—like in the results of Urbanski et al. (2007)—both models perform least skillfully in late fall, winter, and early spring. The (T, PAR) panels in these rows clearly reveal that these errors stem from the tendency of both models, but more so of the VPRM, to overestimate C uptake (i.e., to produce unrealistically strongly negative NEE estimates) when $T > 2\text{--}3^\circ\text{C}$ and PAR fluxes are ample. This model generated small but firmly negative NEEs in the upper-right quadrants contrast with the corresponding observed data (panels $c_{1,2,5}$) showing a featureless $\text{NEE} \gtrsim 0$ field with no systematic left–right NEE changes. In the real forest, $R > -\text{GPP}$ thus holds in this timespan despite the adequate PAR, because the deciduous part of the local forest lacks photosynthetically active leaves with which to exploit this PAR abundance (consistent with Urbanski et al., 2007, their Figure 15). Winter errors are thus largest under warmer daytime conditions, are weakly PAR dependent, and are mostly T dependent. We return to these winter error patterns momentarily.

During the active growing season (summer and early fall; rows 3–4 of Figure 3), when NEE varies mostly in response to variability in PAR dependent GPP, both models perform better, but the ANN improvement is considerably larger. Here too, errors mostly reflect overly negative NEE estimates in high- T , high-PAR bins due to either R underestimation or GPP overestimation. The VPRM errors thus combine underestimating R and overestimating GPP, and stem from the treatment of NEE's dependence on T , PAR, and their covariability with each other and with other input variables.

The above results answer our main question here, whether ANN-based models can outperform the VPRM in NEE forecasting. Indeed they can, and their predictive skills can, probably further improve (e.g., our reported skills are well below the $R^2 \approx 0.93$ Moffat et al., 2010, report for an ANN with 14 predictors).

4. Preliminary Analysis of the VPRM Errors

Since the VPRM's skill can in all likelihood similarly improve, one of our objectives in this paper is to engage the VPRM community in a dialog focused on improving NEE predictions. To that end, and to gain further insights into potential model developments that may enhance the VPRM's skill, in Figure 4 we analyze the VPRM's T dependence over a one-dimensional temperature grid, T_g , that samples uniformly the central 95% of the observed seasonal temperature ranges. Respiration R is explicitly T dependent and is straightforwardly calculated in $\mu\text{mol CO}_2\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ using equation (1) (with α , T_{low} , and β calculated as described in section 2). Also explicitly T dependent in VPRM is T_{scale} (equation (3)). Its impacts on NEE predictions are trickier to analyze because T_{scale} determines predicted NEE only after multiplication by the remaining terms in the compound expression for GPP (i.e., by $P_{\text{scale}} W_{\text{scale}} \text{EVI PAR}_f$; equation (2)). We overcome this by transforming T_{scale} into two VPRM-based NEE estimates, denoted $\text{VPRM}_{1,2}$. In VPRM_1 , we multiply the unique $-\lambda T_{\text{scale}}(T_g)$ function characterizing a given seasonal/PAR regime combination (a specific panel in Figure 4) by the overall seasonal median of $P_{\text{scale}} W_{\text{scale}} \text{EVI PAR}_f$ (where “overall” means “evaluated irrespective of temperature” or “not binned according to T_g ”). Conversely, in VPRM_2 , we bin all GPP terms by temperature, with a unique and in general distinct $P_{\text{scale}} W_{\text{scale}} \text{EVI PAR}_f$ median for each value of the T_g grid.

These transforms convert the dimensionless, arbitrary magnitude T_{scale} into the same physical units as R ($\mu\text{mol CO}_2\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) and endow its magnitude with physical meaning that permits mechanistic interpretation. They differ in whether temperature covariabilities are permitted to impact the VPRM NEE estimates. In VPRM_1 , only the two explicitly T dependent terms affect NEE dependence on T_g , with $\text{GPP} = T_{\text{scale}} \times$ the seasonal all T median of all GPP terms other than T_{scale} , a single scalar that thus leaves the analytic T_{scale} shape unaltered. In VPRM_2 , conversely, the T covariabilities of all predictors modify predicted NEE. We analyze $\text{VPRM}_{1,2}$ in Figure 4, which condenses Figure 3's five seasons into two, defined by bare (mid-October to mid-May) and leafed (mid-May to mid-October) canopy state (top and bottom panels in Figure 4, respectively), comprising 25.0 and 19.8 thousand hourly observations, respectively.

During midday, when PAR is abundant (Figures 4a and 4c), the VPRM general functional form is reasonably but imperfectly consistent with observed NEE T dependence. In winter (Figure 4a) this consistency

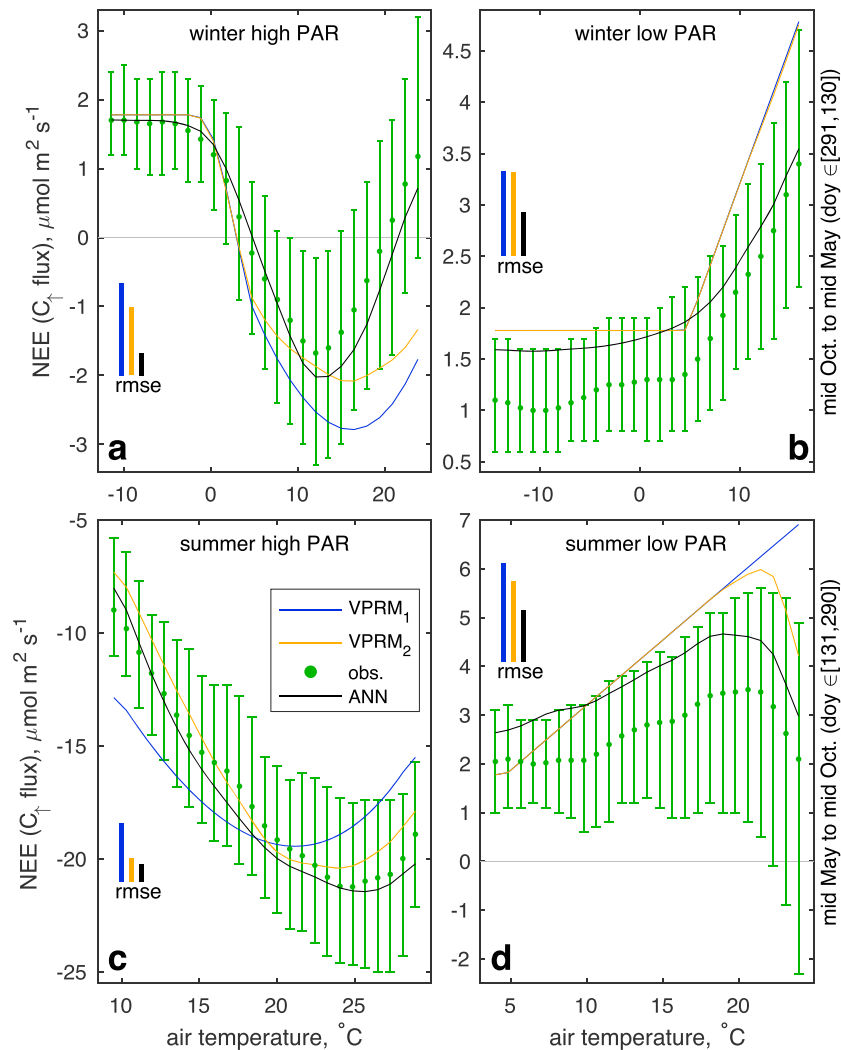


Figure 4. Dependence of NEE estimates on temperature and other environmental variables for winter (a–b) and summer (c–d), distinguishing low (a, c) and high (b, d) PAR conditions. Curves represent the medians over 100 Monte Carlo realizations and available hourly observations. Observed seasonal NEE Medians and thirtieth to seventieth percentiles are shown in green. Black presents median NEE estimates by the fANN. VPRM_{1,2}-based predicted NEE (explained in section 4) are shown in blue and yellow. Root-mean-square errors are given by the corresponding color bars (with arbitrary vertical location but heights conforming with the respective vertical axes). Note the distinct and nonuniform temperature and NEE ranges of individual panels. NEE = net ecosystem exchange; PAR = photosynthetically active radiation; VPRM = Vegetation Photosynthesis and Respiration Model; ANN = artificial neural network.

steadily erodes, however, when $T > T_{low}$ by a functional mismatch between observations and T_{scale} , yielding a sixfold higher error than the fANN's (compare the blue and black vertical bars). When T covariates also guide NEE predictions (Figure 4a, yellow), the error decreases but is still fourfold larger than the fANN's. Exhibiting the inherent limitation of its current analytic formulation, in high PAR summertimes (Figure 4c) T_{scale} alone (blue) systematically overestimates GPP for $T \leq 20^\circ\text{C}$, begins its decline with rising T prematurely, and strongly exaggerates this decline above 23°C . During high PAR conditions in both summer and winter (Figures 4a and 4c), considering T covariates (yellow) improves the estimated NEE somewhat (see the lower yellow root-mean-square error bars). This is expected, because actual C exchanges—and thus the NEE observations—are impacted by all environmental variables, not only temperature, and even though Figure 4 explicitly addresses only temperature, considering both temperature and its covariates offers a back door for other environmental variables to modify the NEE prediction beyond the effect of temperature alone. Yet the NEE prediction error is still considerably higher than the error of the fANN-based estimate (see the left-to-right declining errors the vertical color bars in Figures 4a and 4c show).

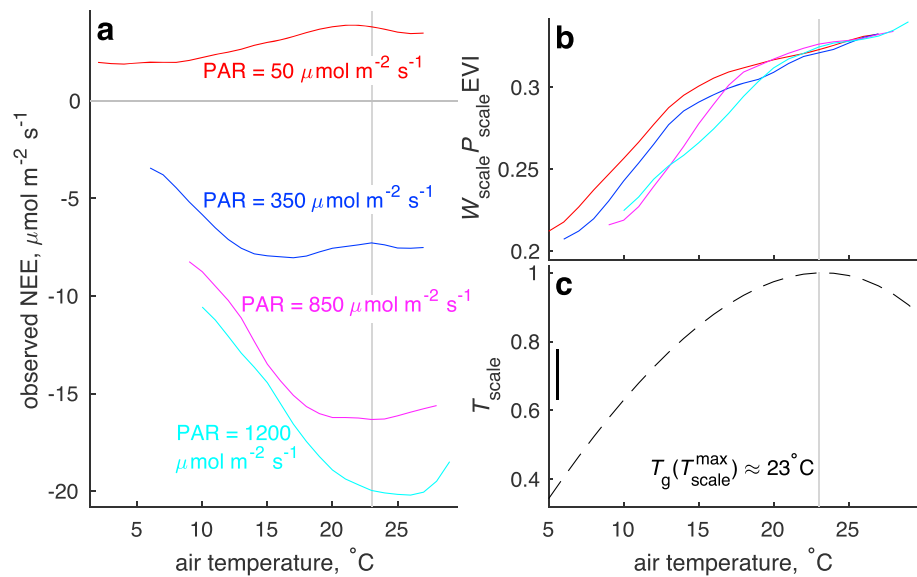


Figure 5. (a) Summer median NEE observations at four narrow PAR ranges (where each curve uses all hourly summer observations that fall within $\pm 50 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ of its stated central PAR value). (b) Median of summer $W_{\text{scale}}P_{\text{scale}}\text{EVI}$ temperature dependence. (c) $T_{\text{scale}}(T_g)$ evaluated on the same temperature grid T_g as panels (a) and (b). For easy comparison, the full vertical range of panel (b) is shown in panel (c) as a thick black vertical bar at $T_g = 6^\circ\text{C}$. The T_g value at which T_{scale} attains its maximum, approximately 23°C , is indicated in all panels by vertical gray lines. NEE = net ecosystem exchange; PAR = photosynthetically active radiation; EVI = enhanced vegetation index.

Figures 4b and 4d address times during which respiration dominates the NEE balance and $\text{GPP} \approx 0$. While in winter the fANN (Figure 4b, black curve) systematically overestimates R for all but the highest temperatures, VPRM-based estimates exhibit functional T dependence that is entirely at odds with observations. NEE observations and fANN-based NEE predictions (Figure 4b green and black) both rise quasi-exponentially, albeit with an offset and somewhat distinct e -foldings. Yet the two straight line segments the VPRM predicted NEE constitutes (yellow and blue) track this functional form quite poorly, yielding roughly double the error of the fANN prediction (colored vertical bars). Similarly, while the fANN captures—however imperfectly—the observed NEE rise over $T \leq 21^\circ\text{C}$ and its rapid decline over $T \geq 22^\circ\text{C}$ in summer (Figure 4d), the VPRM's overly positive uniform $\partial\text{NEE}/\partial T$ below $\approx 21^\circ\text{C}$ rises too high for the correctly reproduced subsequent decline to have an appreciable impact, again accruing roughly twice the error the fANN does. These errors suggest that neither the $T = \max(T, T_{\text{low}})$ switch nor the general uniform $R \propto \beta T$ form serve the VPRM well.

It is tempting to attribute the systematic errors of the blue curves in Figures 4a and 4c to spurious variability of T_{opt} , the only tunable parameter in the T_{scale} formulation, due to the parameter redundancy and resultant numerical instability Mahadevan et al. (2008) warn about. Yet this possibility is readily ruled out. First, when the optimization for the VPRM parameters is carried out seasonally rather than globally, the winter and summer T_{opt} values differ by $< 0.2\text{ K}$ ($< 1\%$, and clearly far too small to materially improve the agreement between Figure 4's blue curves and observations). Second, row 2 of Table 1 shows that the mean T_{opt} over all MC realizations is essentially the same as that based on the full data (columns 3 and 6 from the left, respectively), and not dramatically different from the Mahadevan et al. (2008) value (column 2). Third, T_{opt} proves very stable over the 100 MC realizations, with a standard deviation that is only $< 1\%$ of the mean value (Table 1 column 5, second row).

Figure 4 makes clear that both explicitly T dependent terms in the VPRM— $R(T)$ and $T_{\text{scale}}(T)$ —inadequately capture the temperature dependence of their intended targets, respiration and GPP. Thus, while logically compelling, the explicit T dependence of the VPRM does not represent that of the actual forest particularly well.

While Mahadevan et al. (2008) recognized this, they assumed (their section 4, and the three reduced models therein) that representing GPP in the VPRM as the product of several functions, each logically representing

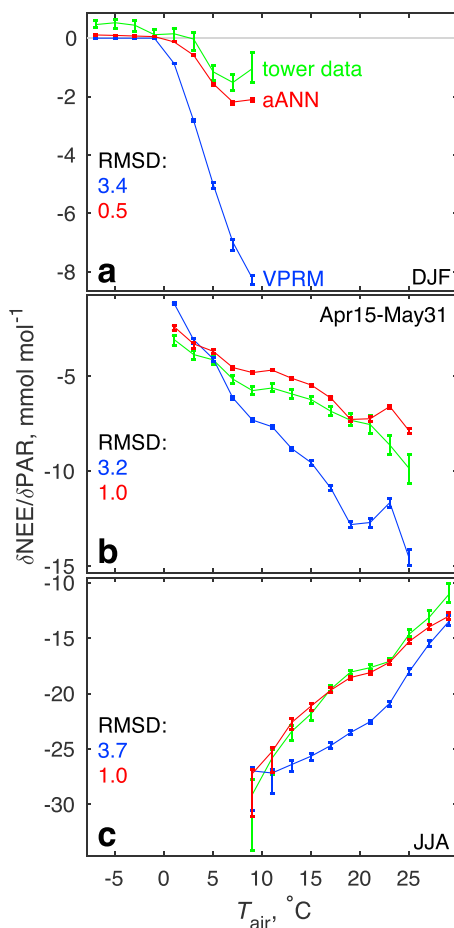


Figure 6. Sensitivity of modeled (red and blue) and observed (green) NEE to PAR changes at a range of air temperatures (horizontal axes) during winter (a), spring (b), and summer (c). Modeled NEE values use all available data points and 100 Monte Carlo based parameter sets, and may thus also reflect some covariabilities of the various solar reflectivities with T , PAR, or both. For each season, we consider the shown temperature grid ($T = -7, -5 \dots 27, 29^\circ\text{C}$). We then use all seasonal PAR data that fall within each of the considered 3°C wide T bins (from $[-8.5, -5.5^\circ\text{C}]$ to $[27.5, 30.5^\circ\text{C}]$) to derive a linear NEE model of the form $NEE_i = \phi_1 + \phi_2 PAR_i$, and present the mean and Monte Carlo-derived range of the ϕ_2 values thus obtained. The RMSD values are the RMSD between either one of the models and the tower data. NEE = net ecosystem exchange; PAR = photosynthetically active radiation; ANN = artificial neural network; RMSD = root-mean-square difference; DJF = December-January-February; JJA = June-July-August.

a different productivity limiting process, will overcome the simplification-motivated imperfections of each individual representation. Yet Figure 4 shows that this compensatory mechanism only works up to a point.

A simple demonstration of the mechanism by which this compensatory mechanism falls short of correcting the considerable mismatch between the analytic form of T_{scale} and NEE observations' T dependence is addressed in Figure 5. To ease interpretation, we address summer only and isolate the effect of PAR by distinguishing four narrow PAR ranges (the four colors, $\pm 50 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ of the stated central PAR values). Figure 5a presents observed NEE(T) for each of these PAR ranges. It makes clear that the heat stress induced GPP reduction at $T \geq 23^\circ\text{C}$ the shape of T_{scale} dictates (Figure 5c) only occurs during very high PARs characteristic of local solar noon (the occurrence times of the data yielding Figure 5a's cyan curve are centered on 11:30 a.m.). If the sought compensation were to occur, it can only arise from increasing $W_{scale} P_{scale} EVI$ over $T \geq 23^\circ\text{C}$ preventing predicted VPRM GPP from declining in concert with declining T_{scale} . Yet Figure 5b shows that by 23°C , the rapid increase of this product has slowed considerably, falling well short of what is needed to undo the T_{scale} controlled declining GPP. For example, from its maximum at $T = 23^\circ\text{C}$, T_{scale} declines to 0.94 and 0.91 at 28 and 29°C while $W_{scale} P_{scale} EVI$ rises from 0.325 to 0.335 and 0.340, a 3–4% increase that falls well short of what is necessary to undo the 6–9% decrease in T_{scale} . The erroneous GPP decline dictated by T_{scale} (Figure 5c), thus remains unopposed, at odds with observation at most PAR values save the highest. It is possible that this product's failure to undo T_{scale} -governed predicted GPP decreases stems from the relatively low sampling frequency of the MODIS data. Yet for the foreseeable future no higher frequency solar reflectivity data are in the offing, a limitation future VPRM-based models must negotiate.

5. NEE Dependence on Hourly Observed Variables

The preceding discussion highlights the potential challenge posed by blending remotely sensed variables observed at coarse temporal resolution with hourly flux tower observations. This challenge, and the fact that NEE is principally affected by PAR and temperature (Figure 1a), naturally leads to the question of the dependence of observed or modeled NEE on these two most impactful and finely observed variables. We examine this question by estimating $\delta NEE / \delta PAR$ during three key seasons as a function of air temperature, using the methods described in section 2.4. These results are presented in Figure 6, with curves and whiskers presenting medians and the full range of the 100 realizations MC populations.

In winter (Figure 6a) during below freezing temperatures, observed NEE rises with increasing PAR (the green $\delta NEE / \delta PAR$ curve is positive below 0°C). Since more positive NEE means enhanced respiration, which is unrelated directly to PAR, this result likely reflects the high (T, PAR) covariance, which means that here increasing PAR is a proxy for rising T , with both reaching their diurnal maxima at roughly the same time of day and the latter accelerating respiration. The aANN PAR dependence (red curve) seems to capture this, albeit with a diminutive amplitude. The VPRM (blue) cannot reproduce this observation because at this temperature range neither R nor GPP can change. For R , this is so because its rate rises with $\max(T, T_{low})$, which throughout this range is reduced to the invariant T_{low} . For GPP, it is so because it is made zero by the numerator $\max(T, T_{min}) - T_{min}$ term in T_{scale} (equation (3)), which here reduces to $T_{min} - T_{min}$ and thus vanishes throughout this T range.

Above 2–3 °C in winter, both the data and aANN (green and red in Figure 6a, respectively) show that the higher winter temperatures rise, the more robustly NEE becomes more negative (i.e., GPP rises) with added PAR. This negative $\delta\text{NEE}/\delta\text{PAR}$ above freezing indicates that evergreen conifer GPP is rising faster than whole ecosystem R in response to added PAR. For the data, which is affected in unknown proportions by both T and PAR, this can only hold if (T, PAR) covariance is weak in winter, or else T would be higher for higher PAR, increasing R faster. Indeed, in the northeast, some of the coldest days are characterized by anomalously strong surface radiative cooling due to subsidence-induced cloudless sky and intense midday solar radiation accompanying equatorward retreat of the jet and the subpolar air to its north. While the VPRM (blue in Figure 6a) also exhibits the same behavior, its response is strongly exaggerated. Over the full T range, the above yield a root-mean-square difference (RMSD) between the $\delta\text{NEE}/\delta\text{PAR}$ curves of the data and the aANN—0.5 mmol/mol—that is only 15% of the 3.4 separating the data and the VPRM curves.

In spring (Figure 6b), leaves of the spatially more dominant deciduous part of the forest have already at least partially emerged, and consequently NEE becomes more negative (GPP rises) with added PAR at all temperatures (i.e., $\delta\text{NEE}/\delta\text{PAR} < 0$ throughout the considered temperature range). The data and aANN curves (green and red in Figure 6b) are very close, with their RMSD of 1 mmol/mol being $\lesssim 17\%$ of the -6 mmol/mol mean of the data curve. Yet their differences are clearly structured, with the aANN underestimating NEE light response at all considered temperatures. To a reasonable degree, this systematic difference can be taken as the amplitude of the contribution of (T, PAR) covariance to NEE variability. That is, because of the way we constructed the green and red curves, we can take the green curve to approximately represent NEE growing negativity with concurrently increasing T and PAR, and the red curve to represent NEE growing negativity with increasing PAR alone, without the effect of higher temperatures that characteristically accompany higher PAR values. With this interpretation, 83% of the effect is due to PAR alone, with the remaining 17% attributable to (T, PAR) covariance.

Throughout the winter and spring (Figures 6a and 6b), the increased NEE negativity with increasing PAR itself increases with temperature [i.e., $\delta(\delta\text{NEE}/\delta\text{PAR})/\delta T < 0$, as the left to right decreases of the curves in Figures 6a and 6b indicate]. In part, this likely reflects the unique challenges of early season photosynthesis, with C uptake on some cold spring days (represented by the curves' left, low T ends) inhibited by temperatures too low for the leaves to fully exploit increasing PAR (Hadley & Schedlbauer, 2002; Korner, 2015). Additionally, since green deciduous canopy is mostly absent or minimal during the time span of Figures 6a and 6b (Toomey et al., 2015), the green curves of these panels primarily represent hemlocks, which exist in the understory throughout the stand as scattered patches of mature trees. The influence of these conifers, which remain photosynthetically active as long as the daily low temperature remains above freezing (Hadley & Schedlbauer, 2002), is evident in the observed light response (green curves) shown in Figures 6a and 6b and is captured adequately or better by the aANN (red curves). Finally, $\delta(\delta\text{NEE}/\delta\text{PAR})/\delta T < 0$ may also arise from the inability of the 8-day MODIS aggregates to resolve the daily progression of leaf photosynthetic capacity characteristic of this transitional period. Either way, here again while the VPRM exhibits roughly the general functional form, its ability to reproduce the temperature dependence of the actual forest NEE light response (the green $\delta\text{NEE}/\delta\text{PAR}$ data curves in Figures 6a and 6b) is quite limited, resulting in over threefold larger RMSD.

In the heart of summer (June–July–August, Figure 6c), with ample PAR availability and temperatures that are at least high enough for full pace photosynthesis but often too high, the situation reverses [i.e., $\delta(\delta\text{NEE}/\delta\text{PAR})/\delta T > 0$, as the left to right increases of Figure 6c's curves indicate]. To be sure, $\delta\text{NEE}/\delta\text{PAR}$ is still negative throughout, which means that all else being equal, more PAR still means more vigorous C uptake by the forest. Yet as the temperature rises, the additional C uptake that an increment of added PAR yields gets smaller [which likely blends genuine T inhibition with (PAR, T) covariance, with higher T and PAR values more likely to occur simultaneously and with reduced impact of added PAR when PAR is already ample]. The aANN is able to reproduce this behavior quite well, accruing an RMSD of 1 mmol/mol for a mean $\delta\text{NEE}/\delta\text{PAR}$ of -19 mmol/mol. The VPRM again falls well short of this performance, with an RMSD almost fourfold larger than that of the aANN.

6. Conclusions

Motivated by the potentially large C uptake by midlatitude deciduous forests, and the prominence of the VPRM (Mahadevan et al., 2008) for scaling this uptake to biome-relevant scales, we set out to explore the possibility of improving uptake estimates using machine learning tools, specifically using ANNs.

We devised, trained, and tested multiple ANN-based models of C uptake, using the extensive HF NEE record as a testbed. Many of these ANN-based models outperform the VPRM itself. Most tellingly, the ANN model which incorporates the same six input variables the VPRM uses outperforms it due to inadequacies of both temperature dependent terms in the VPRM (representing soil respiration and photosynthetic uptake). Of the 28 possible 6-predictor ANN models, the cross-validated R^2 NEE predictive skills of 21, or 75%, exceed the VPRM's ≈ 0.77 , and 11 satisfy $R^2 > 0.8$. The flexibility of ANNs also permits identifying the most impactful of the explanatory variables used as VPRM inputs and testing two additional inputs based on soil temperature and precipitation, leading to further improvements. Since the VPRM offers valuable specific mechanistic insights its ANN alternative does not, its improvement is necessary, and the work reported here offers specific guidance for such needed improvements that stand to broaden the use of the versatile VPRM beyond just an inverse analysis prior generator. More broadly, the results presented here suggest a large and currently minimally tapped potential for ANN-based models in the analysis of biosphere-atmospheric interactions.

References

- Albert, L. P., Keenan, T. F., Burns, S. P., Huxman, T. E., & Monson, R. K. (2017). Climate controls over ecosystem metabolism: Insights from a fifteen-year inductive artificial neural network synthesis for a subalpine forest. *Oecologia*, 184(1), 25–41. <https://doi.org/10.1007/s00442-017-3853-0>
- Anderson, J. A. (1972). A simple neural network generating an interactive memory. *Mathematical Biosciences*, 14(3-4), 197–220.
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., 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.
- Bloom, A. A., Exbrayat, J.-F., van der Velde, I. R., Feng, L., & Williams, M. (2016). The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences of the U.S.A.*, 113(5), 1285–1290. <https://doi.org/10.1073/pnas.1515160113>
- Boose, E. (2018). Fisher meteorological station at Harvard Forest since 2001, Harvard Forest Data Archive: HF001-06, Accessed Marh 12 2018.
- Buchmann, N., & Schulze, E.-D. (2003). Net carbon dioxide and water fluxes of global terrestrial ecosystems, 1969–1998. Oak Ridge, TN: Oak Ridge national Laboratory. <https://doi.org/10.3334/ORNLDAAAC/662>
- Chu, H., Baldocchi, D. D., John, R., Wolf, S., & Reichstein, M. (2017). Fluxes all of the time? A primer on the temporal representativeness of FLUXNET. *Journal of Geophysical Research: Biogeosciences*, 122, 289–307. <https://doi.org/10.1002/2016JG003576>
- Commene, R., Lindaas, J., Benmergui, J., Luus, K. A., Chang, R. Y.-W., Daube, B. C., et al. (2017). Carbon dioxide sources from Alaska driven by increasing early winter respiration from Arctic tundra. *Proceedings of the National Academy of Sciences of the USA*, 114(21), 5361–5366. <https://doi.org/10.1073/pnas.1618567114>
- Crompt, R. F., & Crook, S. M. (1991). Automated extraction of metadata from remotely sensed satellite imagery, *Technical papers - ACSM-ASPRS Annual Convention* (Vol. 3, pp. 111–120). Bethesda, MD: American Congress on Surveying and Mapping: American Society for Photogrammetry and Remote Sensing.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4), 303–314. <https://doi.org/10.1007/BF02551274>
- Dayalu, A., Munger, J. W., Wofsy, S. C., Wang, Y., Nehrkorn, T., Zhao, Y., et al. (2017). VPRM-China: Using the vegetation, photosynthesis, and respiration model to partition contributions to CO₂ measurements in northern China during the 200–2009 growing seasons. *Biogeosciences Discussions*, 2017, 1–29. <https://doi.org/10.5194/bg-2017-504>
- EPA, U. S. E. P. A. (2018). Inventory of U.S. Greenhouse Gas Emissions and Sinks, accessed April 27 2018.
- Fernandez-Martinez, M., Vicca, S., Janssens, I. A., Ciais, P., Obersteiner, M., Bartrons, M., et al. (2017). Atmospheric deposition, CO₂, and change in the land carbon sink. *Scientific Reports*, 7(9632), 1–13. <https://doi.org/10.1038/s41598-017-08755-8>
- Funk, C. C., & Brown, M. E. (2006). Intra-seasonal NDVI change projections in semi-arid Africa. *Remote Sensing of Environment*, 101(2), 249–256. <https://doi.org/10.1016/j.rse.2005.12.014>
- Goulden, M. L., Munger, J. W., Fan, S., Daube, B. C., & Wofsy, S. C. (1996a). Measurements of carbon sequestration by long-term eddy covariance: Methods and a critical evaluation of accuracy. *Global Change Biology*, 2(3), 169–182. <https://doi.org/10.1111/j.1365-2486.1996.tb00070.x>
- Goulden, M. L., Munger, J. W., Fan, S., Daube, B. C., & Wofsy, S. C. (1996b). Exchange of carbon dioxide by a deciduous forest: Response to interannual climate variability. *Science*, 271(5255), 1576–1578. <https://doi.org/10.1126/science.271.5255.1576>
- Hadley, J. L., & Schedlbauer, J. L. (2002). Carbon exchange of an old-growth eastern hemlock (*Tsuga Canadensis*) forest in central New England. *Tree Physiology*, 22(15-16), 1079–1092.
- Harper, C., & Snowden, M. (2017). *Environment and society: Human perspectives on environmental issues* (6th ed., 466 pp.). New York: Routledge (Taylor & Francis).
- Hassoun, M. H. (1995). *Fundamentals of artificial neural networks*. Cambridge: MIT Press.
- Hilton, T. W., Davis, K. J., Keller, K., & Urban, N. M. (2013). Improving North American terrestrial CO₂ flux diagnosis using spatial structure in land surface model residuals. *Biogeosciences*, 10(7), 4607–4625. <https://doi.org/10.5194/bg-10-4607-2013>
- Holgate, C., Jeu, R. D., van Dijk, A., Liu, Y., Renzullo, L., Vinodkumar, I. Dharssi, et al. (2016). Comparison of remotely sensed and modelled soil moisture data sets across Australia. *Remote Sensing of Environment*, 186, 479–500. <https://doi.org/10.1016/j.rse.2016.09.015>

Acknowledgments

E.T. has been funded by National Aeronautics and Space Administration Habitable Worlds programme (grant FP062796-A) and thanks the Weizmann Institute for its hospitality during parts of this work. G.E. thanks Harvard University and its Radcliffe Institute for Advanced Studies for their generosity and hospitality during his Radcliffe Fellowship year. Operation of the Harvard Forest flux tower site is supported by the AmeriFlux Management Project with funding by the U.S. Department of Energy's Office of Science under contract DE-AC02-05CH11231, and additionally as a part of the Harvard Forest LTER site supported by the National Science Foundation (DEB-1237491). All relevant code and data are archived and wiki documented on the Center for Open Science web site at https://osf.io/kjqva/?view_only=238b7e65c7394d3fa40c8b32f7bb075, doi:10.17605/OSF.IO/KJQVA. The tower flux data are from Munger and Wofsy (2017), available at <https://portal.lternet.edu/nis/mapbrowse?packageid=knb-lter-hfr.4.28>, doi:10.6073/pasta/dd9351a3ab5316c844848c3505a8149d. The soil temperature data are from the hf005-04 data set of the Harvard Forest data repository, available from <http://harvardforest.fas.harvard.edu:8080/exist/apps/datasets/showData.html?id=hf005>, doi:10.6073/pasta/cea45e7fb060024359907b2da691b55a. The precipitation data are the hf001-06 set [entitled “daily (metric) since 2001”], also from the Harvard Forest data repository, available from <http://harvardforest.fas.harvard.edu:8080/exist/apps/datasets/showData.html?id=hf001>, doi:10.6073/pasta/04076dfd30b286c6c29301b6345a63f5. The surface reflectance data are from the Terra MOD09A1 MODIS Surface Reflectance 8-Day L3 Global 500 m SIN Grid V006 data set available at <https://modis.gsfc.nasa.gov/data/dataproduct/mod09.php>, doi:10.5067/MODIS/MOD09A1.006.

- Jagermeyr, J., Gerten, D., Lucht, W., Hostert, P., Migliavacca, M., & Nemani, R. (2013). A high-resolution approach to estimating ecosystem respiration at continental scales using operational satellite data. *Global Change Biology*, 20(4), 1191–1210. <https://doi.org/10.1111/gcb.12443>
- Jain, A., & Kumar, A. M. (2007). Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing*, 7(2), 585–592. <https://doi.org/10.1016/j.asoc.2006.03.002>
- Jin, Y., Ge, Y., Wang, J., & Heuvelink, G. B. (2018). Deriving temporally continuous soil moisture estimations at fine resolution by downscaling remotely sensed product. *International Journal of Applied Earth Observation and Geoinformation*, 68, 8–19. <https://doi.org/10.1016/j.jag.2018.01.010>
- Keenan, T. F., Davidson, E., Moffat, A. M., Munger, J. W., & Richardson, A. D. (2012). Using model-data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling. *Global Change Biology*, 18, 2555–2569. <https://doi.org/10.1111/j.1365-2486.2012.02684.x>
- Keenan, T. F., Prentice, I. C., Canadell, J. G., Williams, C. A., Wang, H., Raupach, M., & Collatz, G. J. (2016). Recent pause in the growth rate of atmospheric CO₂ due to enhanced terrestrial carbon uptake. *Nature Communications*, 7(13428), 1–9. <https://doi.org/10.1038/ncomms13428>
- Khashei, M., & Bijari, M. (2010). An artificial neural network (p,d,q) model for timeseries forecasting. *Expert Systems with Applications*, 37(1), 479–489. <https://doi.org/10.1016/j.eswa.2009.05.044>
- Kirschbaum, M., Eamus, D., Gifford, R., Roxburgh, S., & Sands, P. (2001). Definitions of some ecological terms commonly used in carbon accounting. *Net ecosystem exchange workshop proceedings* pp. 2–5). Canberra: Cooperative Research Centre for Greenhouse Accounting.
- Korner, C. (2015). Paradigm shift in plant growth control. *Current Opinion in Plant Biology*, 25, 107–114. <https://doi.org/10.1016/j.pbi.2015.05.003>
- Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9(1), 112–147. <https://doi.org/10.1137/S1052623496303470>
- Law, B., Kelliher, F., Baldocchi, D., Anthoni, P., Irvine, J., Moore, D., & Tuyl, S. V. (2001). Spatial and temporal variation in respiration in a young ponderosa pine forest during a summer drought. *Agricultural and Forest Meteorology*, 110(1), 27–43. [https://doi.org/10.1016/S0168-1923\(01\)00279-9](https://doi.org/10.1016/S0168-1923(01)00279-9)
- Lovenduski, N. S., & Bonan, G. B. (2017). Reducing uncertainty in projections of terrestrial carbon uptake. *Environmental Research Letters*, 12(4), 044020. <https://doi.org/10.1088/1748-9326/aa66b8>
- Luus, K. A., Commene, R., Parazoo, N. C., Benmergui, J., Euskirchen, E. S., Frankenberg, C., et al. (2017). Tundra photosynthesis captured by satellite-observed solar-induced chlorophyll fluorescence. *Geophysical Research Letters*, 44, 1564–1573. <https://doi.org/10.1002/2016GL070842>
- Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X., Dunn, A. L., Lin, J. C., et al. (2008). A satellite-based biosphere parameterization for net ecosystem CO₂ exchange: Vegetation Photosynthesis and Respiration Model (VPRM). *Global Biogeochemical Cycles*, 22, GB2005. <https://doi.org/10.1029/2006GB002735>
- Malmgren, B. A., & Nordlund, U. (1997). Application of artificial neural networks to paleoceanographic data. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 136(1), 359–373. [https://doi.org/10.1016/S0031-0182\(97\)00031-X](https://doi.org/10.1016/S0031-0182(97)00031-X)
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431–441. <https://doi.org/10.1137/0111030>
- Matross, D. M., Andrews, A., Pathmathevan, M., Gerbig, C., Lin, J. C., Wofsy, S. C., et al. (2006). Estimating regional carbon exchange in New England and Quebec by combining atmospheric, ground-based and satellite data. *Tellus B: Chemical and Physical Meteorology*, 58(5), 344–358. <https://doi.org/10.1111/j.1600-0889.2006.00206.x>
- Melillo, J. M., Butler, S., Johnson, J., Mohan, J., Steudler, P., Lux, H., et al. (2011). Soil warming, carbon-nitrogen interactions, and forest carbon budgets. *Proceedings of the National Academy of Sciences*, 108(23), 9508–9512. <https://doi.org/10.1073/pnas.1018189108>
- Melillo, J., Steudler, P., & Mohan, J. (2017). Prospect hill soil warming experiment at Harvard Forest since 1991, Harvard Forest Data Archive: HF005-04, Accessed Marh 3 2018.
- Moffat, A. M., Beckstein, C., Churkina, G., Mund, M., & Heimann, M. (2010). Characterization of ecosystem responses to climatic controls using artificial neural networks. *Global Change Biology*, 2737–2749. <https://doi.org/10.1111/j.1365-2486.2010.02171.x>
- Munger, J. W., & Wofsy, S. C. (2017). Canopy-atmosphere exchange of carbon water and energy at Harvard Forest EMS tower since 1991. <https://doi.org/10.6073/pasta/dd9351a3ab5316c844848c3505a8149d>, environmental Data Initiative; dataset accessed 8/03/2018.
- NOAA (2018). Climate data online, daily summaries, daily Summaries, Accessed Marh 4 2018.
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., et al. (2001). Terrestrial ecoregions of the world: A new map of life on Earth a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, 51(11), 933–938. [https://doi.org/10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)
- Oren, R., Hsieh, C., Stoy, P., Albertson, J., McCarthy, H. R., Harrell, P., & Katulk, G. G. (2006). Estimating the uncertainty in annual net ecosystem carbon exchange: Spatial variation in turbulent fluxes and sampling errors in eddy-covariance measurements. *Global Change Biology*, 12(5), 883–896. <https://doi.org/10.1111/j.1365-2486.2006.01131.x>
- Papale, D., & Valentini, R. (2003). A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. *Global Change Biology*, 9(4), 525–535. <https://doi.org/10.1046/j.1365-2486.2003.00609.x>
- Pastorello, G. Z., Papale, D., Chu, H., Trotta, C., Agarwal, D. A., Canfora, E., et al. (2017). A new data set to keep a sharper eye on land-air exchanges. *Eos*, 98. <https://doi.org/10.1029/2017EO071597>
- Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A continuous satellite-derived measure of global terrestrial primary production. *BioScience*, 54(6), 547–560. [https://doi.org/10.1641/0006-3568\(2004\)054\[0547:ACSMOG\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2)
- Schimel, D., Stephens, B. B., & Fisher, J. B. (2015). Effect of increasing CO₂ on the terrestrial carbon cycle. *Proceedings of the National Academy of Sciences*, 112(2), 436–441. <https://doi.org/10.1073/pnas.1407302112>
- Schlesinger, W. H., & Bernhardt, E. S. (2013). *Biogeochemistry: An analysis of global change* (3rd ed., 688 pp.): Academic Press.
- Tian, H., Lu, C., Ciais, P., Michalak, A. M., Canadell, J. G., Saikawa, E., et al. (2016). The terrestrial biosphere as a net source of greenhouse gases to the atmosphere. *Nature*, 531(7593), 225–228. <https://doi.org/10.1038/nature16946>
- Toomey, M., Friedl, M. A., Frolking, S., Hufkens, K., Klosterman, S., Sonnentag, O., et al. (2015). Greenness indices from digital cameras predict the timing and seasonal dynamics of canopy-scale photosynthesis. *Ecological Applications*, 25(1), 99–115. <https://doi.org/10.1890/14-0005.1>
- Urbanski, S., Barford, C., Wofsy, S., Kucharik, C., Pyle, E., Budney, J., et al. (2007). Factors controlling CO₂ exchange on timescales from hourly to decadal at Harvard Forest. *Journal of Geophysical Research*, 112, 1–25. <https://doi.org/10.1029/2006JG000293>
- van der Baan, M., & Jutten, C. (2000). Neural networks in geophysical applications. *Geophysics*, 65(4), 1032–1047. <https://doi.org/10.1190/1.1444797>

- Verhegghen, A., Bontemps, S., & Defourny, P. (2014). A global NDVI and EVI reference data set for land-surface phenology using 13 years of daily SPOT-VEGETATION observations. *International Journal of Remote Sensing*, 35(7), 2440–2471. <https://doi.org/10.1080/01431161.2014.883105>
- Vermote, E. (2015). MOD09A1 MODIS/Terra surface reflectance 8-day 13 global 500 m SIN Grid V006. <https://doi.org/10.5067/MODIS/MOD09A1.006>, NASA EOSDIS LP DAAC, accessed Marh 14 2018.
- Vogl, T. P., Mangis, J., Rigler, A., Zink, W., & Alkon, D. (1988). Accelerating the convergence of the back-propagation method. *Biological Cybernetics*, 59(4-5), 257–263. <https://doi.org/10.1007/BF0033291>
- Wehr, R., Munger, J. W., McManus, J. B., Nelson, D. D., Zahniser, M. S., Davidson, E. A., et al. (2016). Seasonality of temperate forest photosynthesis and daytime respiration. *Nature*, 534(7609), 680–683. <https://doi.org/10.1038/nature17966>
- Wieder, W. R., Cleveland, C. C., Smith, W. K., & Todd-Brown, K. (2015). Future productivity and carbon storage limited by terrestrial nutrient availability. *Nature Geoscience*, 8, 441–444. <https://doi.org/10.1038/ngeo2413>
- Wilby, R. L., Abrahart, R. J., & Dawson, C. W. (2002). Detection of conceptual model rainfall-runoff processes inside an artificial neural network. *Hydrological Sciences Journal*, 48(2), 163–181. <https://doi.org/10.1623/hysj.48.2.163.44699>
- Wofsy, S. C., Goulden, M. L., Munger, J. W., Fan, S., Bakwin, P. S., Daube, B. C., et al. (1993). Net exchange of CO₂ in a mid-latitude forest. *Science*, 260(5112), 1314–1317. <https://doi.org/10.1126/science.260.5112.1314>
- Xia, J., Niu, S., Ciais, P., Janssens, I. A., Chen, J., Ammann, C., et al. (2015). Joint control of terrestrial gross primary productivity by plant phenology and physiology. *Proceedings of the National Academy of Sciences*, 112(9), 2788–2793. <https://doi.org/10.1073/pnas.1413090112>
- Xiao, X., Zhang, Q., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., et al. (2004). Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. *Remote Sensing of Environment*, 91(2), 256–270. <https://doi.org/10.1016/j.rse.2004.03.010>