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

Corresponding author: Gidon Eshel, Department of Physics, Bard College, Annandale-on-Hudson, New York, USA. (geshel@gmail.com)

¹Physics Department, Bard College, Annandale-on-Hudson, New York, USA.

²Department of Earth and Planetary Sciences and School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA.
Abstract.

The terrestrial biosphere strongly modulates atmospheric CO$_2$ 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 temperatures. 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–ANN errors we thus overcome ANNs’ notoriety for concealing the mechanisms underlying their predictive skills, demonstrating their ability to identify outstanding ecosystem science knowledge gaps and particularly fruitful corresponding model.
development directions, thus improving site specific and up-scaling flux mod-
eling.
1. Introduction

Roughly fourfold larger than the atmospheric pool [Schlesinger and Bernhardt, 2013], the global terrestrial biosphere carbon pool strongly influences atmospheric CO$_2$ concentration [Lovenduski and 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 or losses [Kirschbaum et al., 2001; Wehr et al., 2016]—has offset a quarter to a third [Hilton et al., 2013; Keenan et al., 2016; Fernandez-Martinez et al., 2017] of global anthropogenic emissions. NEE is thus a central element of the anthropogenically perturbed C cycle [Hilton et al., 2013; Tian et al., 2016] that requires robust observations for improved understanding, and skillful modeling for reliable projections of future atmospheric CO$_2$ levels [Schimel et al., 2015; Bloom et al., 2016] and their climate consequences, arguably the central geophysical forecast of our time [Harper and Snowden, 2017].

Developing these 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 PAR too often peaks, yielding a strong positive temperature
PAR covariance (we show this quantitatively and discuss it in later sections). With both mutually approximately canceling large fluxes rising in concert with temperature, their individual effects are masked in the timeseries of their smaller, noisier sum. Nonetheless, sufficiently large data sets, as the one we use here, sample enough exceptions to the above covariability—e.g., hot, cloudy days—to unmask individual effects.

The second challenge stems from the fact that because CO\textsubscript{2} is mostly well mixed in the global atmosphere, the relevant variations in atmospheric CO\textsubscript{2} are driven by \( \int \int A' \text{NEE}(x,y,t) \, dA \), NEE integrated over a regional to larger scale area \( A' \). Yet NEE varies widely in space \([Oren et al., 2006; Xia et al., 2015; Bloom et al., 2016]\), and our main NEE observational tool—the spatially heterogeneous ground based array of eddy covariance CO\textsubscript{2} flux measurements, most recently known as FLUXNET \([Baldocchi et al., 2001; Buchmann and Schulze, 2003; Jagermeyr et al., 2013; Pastorello et al., 2017; Chu et al., 2017]\)—undersamples this variability in both space and time \([Law et al., 2001]\). With its integrand imperfectly known locally and undersampled spatially, robustly estimating such integrals observationally is therefore still an outstanding challenge.

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 \(\text{Papale and Valentini} [2003]\) and methodologically expanding \(\text{Moffat et al.} [2010]; \text{Keenan et al.} [2012]; \text{Albert et al.} [2017]\), here we set out to develop such a flexible...
alternative based on an Artificial Neural Network [ANN, Anderson, 1972; Hassoun, 1995] time series modeling framework [Khashei and 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 and Crook, 1991; Malmgren and Nordlund, 1997; van der Baan and Jutten, 2000; Jain and Kumar, 2007] to a single (scalar) NEE time series measured by the eddy covariance $\text{CO}_2$ flux tower at Harvard Forest [HF, Wofsy et al., 1993; Goulden et al., 1996a, b; Urbanski et al., 2007; Munger and Wofsy, 2017], 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 photosynthetically active radiation (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 prior generating tool for 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 CO$_2$ observations [e.g., Matross et al., 2006; Dayalu et al., 2017]. 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 (note 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 model development efforts that 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 Harvard
Forest, hereafter 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 respiration is $\alpha \max(T, T_{\text{low}}) + \beta$,
where $T$ is the above canopy air temperature and $\alpha$, $\beta$ and $T_{\text{low}}$ are locally tunable
constants, with $T_{\text{low}} \approx 1-5^\circ \text{C}$ 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 de-
dendencies on environmental conditions characteristic of the $\approx 150^\circ$ meridional span of
the photosynthetically active terrestrial biosphere. Mahadevan et al. [2008], e.g., 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
2. Methods

2.1. VPRM

The VPRM [Mahadevan et al., 2008] is a widely employed model of biome generic canopy–boundary layer CO$_2$ 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, respiration ($R$) carbon flux, and their net balance, NEE [Kirschbaum et al., 2001]. The VPRM equations are,

$$\text{NEE}_↑ = - \text{GPP}_↓ + \text{R}_↑$$

$$= - \lambda T_{\text{scale}} P_{\text{scale}} W_{\text{scale}} \text{EVI PAR}_f + \text{R}_↑$$

$$= - \lambda T_{\text{scale}} \frac{(1 + LSWI)^2}{2(1 + LSWI_{\text{max}})} \text{EVI} \frac{\text{PAR}}{1 + \text{PAR/PAR}_0} + \alpha \max(T, T_{\text{low}}) + \beta$$

In Eq. 1, vertical arrows indicate the direction defined as positive, with actual fluxes in general assuming either negative or positive value (such that NEE > 0 indicates CO$_2$ loss to the boundary layer via upward net flux, and NEE < 0 corresponds to downward flux into or net C uptake by the terrestrial biosphere). The PAR$_f$ term is implicitly defined by the 2nd and 3rd rows, and

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

where, following Mahadevan et al. [2008], we set $T_{\text{min}} = 0^\circ$C and $T_{\text{max}} = 40^\circ$C. Eq. 2 differs in appearance from the original (their Eq. 6), because it incorporates (through the numerator’s first term) additional constraints presented in that paper. The parameters
in Eqs. 1 and 2, \( \{\lambda, T_{opt}, \text{PAR}_0, \alpha, \beta, T_{\text{low}}\} \) are optimized for a specific site of observations [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 1628-1652 nm from the nearest pixel to HF in the L3 8-day, 500 m MODIS data set [Vermote, 2015], and LSWI\(_{\text{max}}\) 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 \( \mu \)m), 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 eight 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 Artificial Neural Networks

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–December 2015 (the period over which both the tower data and MODIS observations are available) and sampling reasonably uniformly the diurnal and seasonal cycles [Urbanski et al., 2007; Munger and Wofsy, 2017].

We apply to the HF NEE record numerous ANNs with variable number of predictors (scalar inputs, out of the full 8 predictor set), covering the full set of ANNs comprising each one of the 255 possible combination of $N_p = [1, 8]$ predictors (where 255 is the sum of the binomial coefficients “$n$ choose $k$” with $n = 8$ and $k = N_p = [1, 8]$). When the considered ANN has fewer predictors than the VPRM’s six, it is denoted tANN, for “truncated”. When it depends on all six VPRM predictors, it is denoted fANN, for “full”. Finally, when the considered model contains novel predictors not included in the VPRM formalism, the model is denoted aANN, for “augmented”. These models are further described below. An extensive systematic search has revealed that the highest systematically reproducible cross validated forecasting skill is achieved by an ANN(6,5,4) architecture, explained below, although the performance is robust under small network architecture changes.

While an introduction to ANNs is outside of the current scope, Consider the following brief overview specifically pertinent to the current problem. An ANN(6,5,4) has three hidden layers with six, five and four nodes (or neurons) respectively. At each time point, weighted sums 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 non-linear 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 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 neural network toolbox im-
plementation of the Levenberg-Marquardt back propagation algorithm [Marquardt, 1963].
The training data are a randomly chosen half of the available time points for the chosen
predictors, amounting to about 22,450 hourly observations for each predictor. The model
is then cross validated for its NEE predictive skill over the remaining half.

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. 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 the Results section.

For aANN, we explore two additional daily resolution predictors that cannot be straightforwardly used in VPRM. The first is soil temperature \cite{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 \cite{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 $T_{\text{soil}}$ predictor is the mean value of the six Prospect Hill control plots.

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

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

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 \cite{Boose, 2018}—a direct measure of water delivered to the forest in the immediate vicinity of the flux tower—with the nearby NCDC \cite{National Center for Environmental Information and Administration, 2018} site at Orange Municipal Airport, MA, where valid data sometimes exist when the HF data is 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 6154 overlapping daily precipitation values. Filling data missing from the Prospect Hill record with data from the airport record results in a $p(t)$ time series spanning Jan. 1 1991 to Feb. 10 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 $\tau$) the correlation magnitude steadily declines as it rises back toward zero. Based on this behavior, our second additional 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—$T_{\text{min}} = 0^\circ C$ and $T_{\text{max}} = 40^\circ C$—and 6 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 [Eq. 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 $3 \times 10^3$ randomly chosen hourly data points. This yields the parameter statistics reported in columns 3–5 (from the left) of Table 1. 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 $\beta$ is an exception. Yet this too is reasonable given that $\beta$ 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.

The temperature grid we use spans $[-7^\circ \text{C}, 31^\circ \text{C}]$ in $2^\circ \text{C}$ increments. We treat each $T$ grid value as the center of $3^\circ \text{C}$ wide $T$ bins (e.g., the two successive bins centered at $1^\circ$ and $3^\circ \text{C}$ span $[-0.5^\circ, 2.5^\circ \text{C}]$ and $[1.5^\circ, 4.5^\circ \text{C}]$ 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 Monte Carlo (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 $\phi_2$ 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 ($T, \text{PAR}$)
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 nonlinear 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/$T$ bin 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 $\phi_2$ 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 potential ANN input variables as NEE predictors, we use each individually to train a single-predictor tANN (an ANN(6,5,4) with one scalar input). The performance criterion we employ is the mean ± standard deviation of cross-validated forecast $R^2$ evaluated over 100 samples of $10^3$ data points each, randomly drawn from the validation half. Note that this randomization in general eliminates the spuriously elevated explanatory power due to the strong subsynoptic temporal autocorrelations most hourly resolution meteorological time series exhibit [Moffat et al., 2010]. 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 Fig. 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 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 (SIF) measurements [Commane 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? Fig. 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 Fig. 1b, which presents the two most skillful models for each given value of the number of predictions \( N_p \). These results are shown for five values of the number of predictions, \( N_p = 3–7 \), and we note that all ten top models shown outperform the VPRM’s skill (shown by the gray band). In fact, of the 254 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 Fig. 2 amount to nontrivial errors in estimating C sequestration. A useful reference is the mean NEE (annual net CO\(_2\)-C uptake) at Harvard Forest, estimated to be roughly 1.5–3 metric ton (mt) C ha\(^{-1}\) y\(^{-1}\) 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\(^{-1}\) y\(^{-1}\). 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\(^{-1}\) y\(^{-1}\) with a mean of \( \geq 100 \) kg C ha\(^{-1}\) y\(^{-1}\). 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 $\approx 9 \times 10^5$ mt C, which is equivalent to the annual emissions of approximately 5.5 million Americans.

Returning briefly to the $N_p = 1$ results introduced earlier, the leftmost (black) curve in Fig. 1a shows that the expected dominance of PAR and $T$ in determining NEE—a basic expectation from any NEE model—indeed holds for the ANN based models. 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 dominance of PAR is expected given its role in photosynthetic rate determination. The secondary utility of $T$ as an 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, are less expected. This result too can guide future efforts to enhance the VPRM. One possible explanation of this peculiarity is that the representation of ecosystem respiration in the VPRM,

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

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 $T_{soil}$, not $T$ (a possibility with limited practical applicability for regional scale modeling regardless of merit because $T_{soil}$ data coverage is poor). 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 $T_{\text{soil}}$ 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 Fig. 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 com-
binations 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, Fig. 1b warns of the rising risks of overfitting and information satu-
ration that typically accompany the exhaustive undiscriminating 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 7th predictor adds no significant skill.

3.3. Seasonal Dependence and a Mechanistic Interpretation

The results Fig. 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,
Fig. 1 does not resolve seasonal dependence of the inherently time dependent forest C
uptake [Running et al., 2004; Funk and Brown, 2006]. 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. Fig. 3 presents NEE as
a function of these two observed predictors for (from the top row down) five time ranges
during the year: (1) Nov. 27–Mar. 1; (2) Mar. 2–May 10; (3) May 11–Aug. 23; (4)
Aug. 24–Oct. 17; and (5) Oct. 18–Nov. 26. 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 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 Fig. 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-Oct.–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, when
GPP is small and respiration dominates NEE. The $(T,\text{PAR})$ 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 $T > 2^\circ–3^\circ\text{C}$ and PAR fluxes are ample. These 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 NEE $\gtrapprox 0$ 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
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 Fig. 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 Fig. 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 $\mu$mol CO$_2$ m$^{-2}$ s$^{-1}$ using Eq. 4 (with $\alpha$, $T_{low}$ and $\beta$ calculated as described in the Methods section). Also explicitly $T$ dependent in VPRM
is $T_{\text{scale}}$ (Eq. 2). Its impacts on NEE predictions are trickier to analyze because $T_{\text{scale}}$ determines predicted NEE only after multiplication by the remaining terms in the compound expression for GPP (i.e., by $P_{\text{scale}} W_{\text{scale}} \text{EVI PAR}_f$; Eq. 1). We overcome this by transforming $T_{\text{scale}}$ into two VPRM-based NEE estimates, denoted 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 Fig. 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 $T_{\text{scale}}$ into the same physical units as $R \, (\mu\text{mol CO}_2\text{ m}^{-2}\text{ s}^{-1})$ 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 $T_{\text{scale}}$, a single scalar that thus leaves the analytic $T_{\text{scale}}$ shape unaltered. In VPRM$_2$, conversely, the $T$ covariabilities of all predictors modify predicted NEE. We analyze VPRM$_{1,2}$ in Fig. 4, which condenses Fig. 3’s five seasons into two, defined by bare (mid October–mid May) and leafed (mid May - mid Oct) canopy state (top and bottom panels in Fig. 4 respectively), comprising 25.0 and 19.8 thousand hourly observations respectively.

During midday, when PAR is abundant (Fig. 4a,c), the VPRM general functional form is reasonably but imperfectly consistent with observed NEE $T$ dependence. In winter (panel a) this consistency steadily erodes, however, when $T > T_{\text{low}}$ by a functional
mismatch between observations and $T_{\text{scale}}$, yielding a 6-fold higher error than the fANN’s (compare the blue and black vertical bars). When $T$ covariates also guide NEE predictions (4a, yellow), the error decreases but is still 4-fold larger than the fANN’s. In high PAR summer times (4c) $T_{\text{scale}}$ 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 $23^\circ\text{C}$. While considering $T$ covariates (yellow) improves the estimated NEE somewhat, 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 Fig. 4c show).

Fig. 4b,d address respiration dominated, GPP $\approx 0$ times. While in winter the fANN (Fig. 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 (Fig. 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 (Fig. 4d), the VPRM’s overly positive uniform $\partial \hat{\text{NEE}}/\partial T$ 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 Fig. 4a,c to spurious variability of $T_{\text{opt}}$, the only tunable parameter in the $T_{\text{scale}}$ 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 $< 0.2 K$ ($< 1\%$, and clearly far too small to materially improve the agreement between Fig. 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, 2nd row).

Fig. 4 makes clear that both explicitly $T$ dependent terms in the VPRM—$R(T)$ and $T_{scale}(T)$—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 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 a different productivity limiting process, will overcome the simplification motivated imperfections of each individual representation. Yet Fig. 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 Fig. 5. To ease interpretation, we address summer only, and isolate the effect of PAR by distinguishing four narrow PAR ranges (the
four colors, ±50 µmol m\(^{-2}\) s\(^{-1}\) of the stated central PAR values). Fig. 5a presents observed NEE\(\left( T \right)\) for each of these PAR ranges. It makes clear that the heat stress induced GPP reduction at \(T \geq 23^\circ C\) the shape of \(T_{\text{scale}}\) dictates (panel c) only occurs during very high PARs characteristic of local solar noon (the occurrence times of the data yielding Fig. 5a’s cyan curve are centered on 11:30 AM). If the sought compensation were to occur, it can only arise from increasing \(W_{\text{scale}}P_{\text{scale}}EVI\) over \(T \geq 23^\circ C\) preventing predicted VPRM GPP from declining in concert with declining \(T_{\text{scale}}\). Yet panel b shows that by 23°C, the rapid increase of this product has slowed considerably, falling well short of what is needed to undo the \(T_{\text{scale}}\) controlled declining GPP. For example, from its maximum at \(T = 23^\circ C\), \(T_{\text{scale}}\) declines to 0.94 and 0.91 at 28 and 29°C while \(W_{\text{scale}}P_{\text{scale}}EVI\) rises from 0.3246 to 0.335 and 0.34, a 3–4% increase that falls well short of what is necessary to undo the 6–9% decrease in \(T_{\text{scale}}\). The erroneous GPP decline dictated by \(T_{\text{scale}}\) (Fig. 5c) thus remains unopposed, inconsistently with observation at most PAR values save the highest. It is possible that this product’s failure to undo \(T_{\text{scale}}\) 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 (Fig. 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 \( \partial \text{NEE}/\partial \text{PAR} \) during three key seasons as a function of air temperature, using the methods described in section 2.4. These results are presented in Fig. 6, with solid curves and spread bars presenting medians and the full range of the 100 realizations MC populations.

In winter (Fig. 6a) during below freezing temperatures, observed NEE rises with increasing PAR (the blue \( \delta \text{NEE}/\delta \text{PAR} \) curve is positive below 0\(^\circ\)C). Since more positive NEE means enhanced respiration, which is unrelated to PAR, this result likely reflects the high \((T, \text{PAR})\) 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 (green) 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 \(\max(T, T_{\text{low}})\), which throughout this range is reduced to the invariant \(T_{\text{low}}\). For GPP, it is so because it is made zero by the numerator \(\max(T, T_{\text{min}}) - T_{\text{min}}\) term in \(T_{\text{scale}}\) (Eq. 2), which here reduces to \(T_{\text{min}} - T_{\text{min}}\) throughout this \(T\) range.

Above 2–3\(^\circ\)C, both the data (blue) and the aANN (red) show that the warmer it is, 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 \((T, \text{PAR})\) 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 radia-

tion accompanying equatorward retreat of the polar high. While the VPRM also exhibits

the same behavior (green), its response is strongly exaggerated. Over the full $T$ range,

the above yield a root mean squared difference (RMSD) between the $\delta$NEE/$\delta$PAR curves

of the data and the aANN—0.5 mmol mol$^{-1}$—that is only 15% of the 3.4 separating the

data and the VPRM curves.

In spring (Fig. 6b), leaves of the areally more dominant deciduous part of the forest

have already emerged, and consequently NEE becomes more negative (GPP rises) with

added PAR at all temperatures. The data (blue) and aANN (red) curves are very close,

with their RMSD of 1 mmol mol$^{-1}$ being $\lesssim$17% of the -6 mmol mol$^{-1}$ mean of the data

curve. Yet their differences are clearly structured, with the aANN underestimating NEE

PAR sensitivity 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 blue and red curves, we can

take the blue 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.

The increased NEE negativity with increasing PAR itself increases with temperature

(i.e., $\delta^2$NEE/(δPAR δ$T$) < 0, as the left to right decreases of the blue and red curves

indicate). This likely reflects the unique challenges of early season photosynthesis, with

C uptake on some spring days (represented by the curves’ left, low $T$ ends) inhibited
by temperatures too low for the leaves to take full advantage of available PAR. It may also arise from the inability of the 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 data $\delta \text{NEE}/\delta \text{PAR}$ is quite limited, resulting in over 3-fold larger RMSD.

In the heart of summer (JJA, Fig. 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., $\delta^2 \text{NEE}/(\delta \text{PAR} \delta T) > 0$, as the left to right increases of panel c’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$^{-1}$ for a mean $\delta \text{NEE}/\delta \text{PAR}$ of -19 mmol mol$^{-1}$. The VPRM again falls well short of this performance, with an RMSD almost 4-fold larger than that of the aANN.

6. Conclusions

Motivated by the potentially large C uptake by midlatitude deciduous forests, and the prominence of the Vegetation Photosynthesis and Respiration Model [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 Artificial Neural Networks (ANN).
We devised, trained and tested multiple ANN-based models of C uptake, using the extensive Harvard Forest NEE record as a testbed. Many of those models outperform the VPRM itself. Our analysis reveals that the ANN models based on the same inputs as VPRM show higher skill arising from under-performance by both temperature dependent terms in the VPRM, representing soil respiration and photosynthetic uptake. Furthermore, the flexibility of ANN allowed us to judge which of the inputs used in the VPRM are most helpful, as well as test 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.

Acknowledgments. ET 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. GE 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 Energys Office of Science under Contract No. DE-AC02-05CH11231, and additionally as a part of the Harvard Forest LTER site supported by the National Science Foundation (DEB-1237491).
References


Table 1. The parameters of the VPRM given by Mahadevan et al. [2008] and estimated here, where mean and variability are calculated over 100 nonlinear least squares fits of the VPRM equation (Eq. 11 of Mahadevan et al. [2008]) to 100 subsets of the full data, each comprising \(3 \cdot 10^3\) randomly chosen hourly data points. Following Mahadevan et al. [2008], we set \(T_{\text{min}} = 0^\circ\text{C}\) and \(T_{\text{max}} = 40^\circ\text{C}\) throughout.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Mean</th>
<th>Deviation</th>
<th>%</th>
<th>Observations</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda)</td>
<td>0.127</td>
<td>0.221</td>
<td>0.003</td>
<td>1.2</td>
<td>0.221</td>
<td>mol CO(_2) (mol PAR(^{-1}))</td>
</tr>
<tr>
<td>(T_{\text{opt}})</td>
<td>20</td>
<td>23.3</td>
<td>0.14</td>
<td>0.6</td>
<td>23.3</td>
<td>°C</td>
</tr>
<tr>
<td>(\text{PAR}_0)</td>
<td>570</td>
<td>536.9</td>
<td>8.0</td>
<td>1.5</td>
<td>536.8</td>
<td>(\mu\text{mol m}^{-2}\ \text{s}^{-1})</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>0.271</td>
<td>0.267</td>
<td>0.007</td>
<td>2.6</td>
<td>0.267</td>
<td>(\mu\text{mol CO}_2\ \text{m}^{-2}\ \text{(s K)}^{-1})</td>
</tr>
<tr>
<td>(\beta)</td>
<td>0.25</td>
<td>0.54</td>
<td>0.09</td>
<td>16.8</td>
<td>0.54</td>
<td>(\mu\text{mol CO}_2\ \text{m}^{-2}\ \text{s}^{-1})</td>
</tr>
<tr>
<td>(T_{\text{low}})</td>
<td>5</td>
<td>4.63</td>
<td>0.27</td>
<td>5.7</td>
<td>4.63</td>
<td>°C</td>
</tr>
</tbody>
</table>
Figure 1. Cross-validated NEE predictive skill of various ANN(6,5,4) models. We explore such models with 1 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 skill of each set, sorted by NEE predictive skills. The range whiskers show means ± 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 6 predictors are highlighted with square symbols. The darker 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. The lighter gray shading indicates the $R^2$ range of panel b, which highlights the NEE predictive skill ranges of two most skillful predictor combinations with $N_p = [3–7]$. Panel b asterisks indicate the significance of the $R^2$ difference between neighboring skills, with 1–4 asterisks denoting $p < 0.05, 0.01, 0.005,$ and 0.001.
Figure 2. Spread of NEE prediction error by the VPRM (leftmost, blue) and the six top performing ANNs shown in Fig. 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\times y)^{-1}.
Figure 3. Comparison of NEE dependence on air temperature and PAR over different parts of the seasonal cycle (panels a–c on 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. 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 squared errors over all populated bins within the day ranges normalized by the corresponding NEE observations, with the fANN (the VPRM) in red (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.)
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 30th–70th percentiles are shown in green. Black presents median NEE estimates by the fANN. VPRM based NEE using $s_{1,2}$ (explained in the text) are shown in blue and yellow. Root mean squared errors are given by the corresponding color bars (with arbitrary vertical location but heights conforming with the respective vertical axes).
Figure 5. Panel a: Summer median NEE observations at four narrow PAR ranges (where each curve uses all hourly summer observations that fall within ±50 µmol m⁻² s⁻¹ of its stated central PAR value). Panel b: median of summer $W_{\text{scale}}P_{\text{scale}}EVI$ temperature dependence. Panel 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 $T_g = 6^\circ$C.
Figure 6. Sensitivity of modeled (red and green) and observed (blue) 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 ($T = -7^\circ C, -5^\circ C \cdots 27^\circ C, 29^\circ C$). We then use all seasonal PAR data that fall within each of the considered 3$^\circ$C wide $T$ bins (from $[-8.5,-5.5]^\circ C$ to $[27.5,30.5]^\circ C$) to derive a linear NEE model of the form $\text{NEE}_i = \phi_1 + \phi_2 \text{PAR}_i$, and present the mean and MC derived range of the $\phi_2$ values thus obtained.