

JGR Biogeosciences

RESEARCH ARTICLE

10.1029/2021JG006363

Special Section:

Advances in scaling and modeling of land-atmosphere interactions

Key Points:

- With as little as 2 years of data, ecosystem fluxes can be modeled with artificial neural network models
- Annual sums of net ecosystem exchange (NEE), ecosystem respiration (R_{eco}), and gross primary production (GPP) can be modeled with an uncertainty of 20%, 22%, and 8%
- At different land-covers, uncertainties of modeled annual sums are $NEE = 83%$, $R_{eco} = 12%$ and $GPP = 10%$

Correspondence to:

D. E. Reed,
david.edwin.reed@gmail.com

Citation:

Reed, D. E., Poe, J., Abraha, M., Dahlin, K. M., & Chen, J. (2021). Modeled surface-atmosphere fluxes from paired sites in the upper Great Lakes region using neural networks. *Journal of Geophysical Research: Biogeosciences*, 126, e2021JG006363. <https://doi.org/10.1029/2021JG006363>

Received 27 MAR 2021

Accepted 27 JUL 2021

Author Contributions:

Conceptualization: David E. Reed, Jeralyn Poe, Kyla M. Dahlin

Data curation: David E. Reed, Michael Abraha

Formal analysis: David E. Reed

Funding acquisition: Kyla M. Dahlin, Jiquan Chen

Methodology: David E. Reed, Jeralyn Poe, Michael Abraha, Kyla M. Dahlin




Project Administration: David E. Reed, Kyla M. Dahlin, Jiquan Chen

Resources: Michael Abraha

Supervision: Kyla M. Dahlin, Jiquan Chen

Visualization: Jeralyn Poe

Modeled Surface-Atmosphere Fluxes From Paired Sites in the Upper Great Lakes Region Using Neural Networks

David E. Reed^{1,2} , Jeralyn Poe^{1,3}, Michael Abraha¹, Kyla M. Dahlin¹ , and Jiquan Chen¹ 

¹Department of Geography, Environment, and Spatial Sciences, Michigan State University, East Lansing, MI, USA, ²Environmental Science, University of Science and Arts of Oklahoma, Chickasha, OK, USA, ³School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA

Abstract The eddy covariance (EC) method is one of the most widely used approaches to quantify surface-atmosphere fluxes. However, scaling up from a single EC tower to the landscape remains an open challenge. To address this, we used 63 site years of data to examine simulated annual and growing season sums of carbon fluxes from three paired land-cover type sites of corn, restored-prairie, and switchgrass ecosystems. This was also done across the landscape by modeling fluxes using different land-cover type input data. An artificial neural network (ANN) approach was used to model net ecosystem exchange (NEE), ecosystem respiration (R_{eco}), and gross primary production (GPP) at one paired site using environmental observations from the second site only. With a mean spatial separation of 11 km between paired sites, we were able to model annual sums of NEE, R_{eco} , and GPP with uncertainties of 20%, 22%, and 8%, respectively, relative to observation sums. When considering the growing season only, model uncertainties were 17%, 22%, and 9%, respectively for the three flux terms. We also show that ANN models can estimate sums of R_{eco} and GPP fluxes without needing the constraint of similar land-cover-type, with annual uncertainties of 12% and 10%. These results provide new insights to scaling up observations from one EC site beyond the footprint of the EC tower to multiple land-cover types across the landscape.

Plain Language Summary Measuring terrestrial carbon cycling is an important step in solving the global missing carbon sink budget, but direct measurements of carbon cycling are time intensive and costly. Eddy covariance methods are a commonly used technique to measure carbon cycling at scales of $\sim 1 \text{ km}^2$ and efforts to scale these measurements to the entire landscape would allow a better comparison of carbon measurements of global models. Here we use a long-term record of carbon cycling from six paired sites plus a seventh reference site that are all spatially close to each other, with an average of $\sim 11 \text{ km}$ separation. The goal is to use field observations from one of the paired sites, along with observations from the second site that can be measured easily, in order to estimate carbon cycling at the second site using artificial intelligence methods. We find that carbon cycling can be estimated with the same amount of uncertainty as the observations themselves. This is also true when using non-paired sites, for example, information from a corn site can model carbon fluxes at a grassland, and shows we could be able to extrapolate our understanding of carbon cycling much farther from a single measurement location.

1. Introduction

Spatial scaling up in environmental science is a long standing challenge (Levin, 1992; Running et al., 1999) with important implications to understanding ecological responses to changing future conditions (Dietze et al., 2018; Reichstein et al., 2019). Scaling up in ecology has been based on process-based mechanistic models, either built from field data or compared with field data to find areas to improve (Denny & Benedetti-Cecchi, 2012). A roadmap to tackling the scaling problem in the carbon cycle was outlined 20 years ago by Running et al. (1999), where local surface-atmosphere flux data would be scaled to the region via parameter based mechanistic modeling with satellite and ancillary data providing quantifying landscape heterogeneity. This allows site level flux data to be used to estimate surface-atmosphere fluxes at global scale in order to evaluate climate models (Jung et al., 2011, 2020). However, an evaluation of 26 models shows that none of the daily mechanistic model outputs matched estimated gross primary production (GPP) within observed uncertainty, with models being able to follow seasonal patterns but not correctly matching flux magnitudes (Schaefer et al., 2012). It is important to note that model uncertainties, as well as observational uncertainties, can increase when scaling to the landscape, sometimes in nonlinear fashion (Keenan et al., 2011).

Writing – original draft: David E. Reed

Writing – review & editing: David E. Reed, Jeralyn Poe, Michael Abraha, Kyla M. Dahlin, Jiquan Chen

Eddy covariance methods have multiple types of uncertainties (Richardson et al., 2012), including uncertainties due to—but not limited to—random measurement error (Hollinger & Richardson, 2005), gap-filling (Richardson & Hollinger, 2007), sensor choice (Frank et al., 2013), and site-specific landscape heterogeneity within the tower footprint (Chu et al., 2021; Stoy, Mauder, et al., 2013). Partitioning net ecosystem exchange (NEE) of carbon into GPP and ecosystem respiration (R_{eco}) incorporates additional uncertainties as well (Lasslop et al., 2010; Reichstein et al., 2005). The majority of these uncertainties can be quantified as average energy closure across all sites (Wilson et al., 2002) as well as temporal energy closure (Reed, Frank, et al., 2018), but not all uncertainties directly contribute to energy closure issues. While site to site variations exist, these uncertainties average to be $\sim 20\%$ when summing carbon fluxes to the annual timescale (Loescher et al., 2006). Observations from a single flux footprint have this $\sim 20\%$ uncertainty inherent to them, and the goal of scaling up from the footprint to the landscape level should be to model fluxes within this level of uncertainty.

The presence of both spectral similarity and spatial coherence in fluxes can be used to quantify spatial patterns and uncertainties in the data without needing to parametrize mechanistic models (Bishop, 2006). Using two clusters of surface-atmosphere flux sites from corn, switchgrass and restored-prairie fields, our previous work quantified the spatial extent that flux information from one eddy covariance site is coherent up to a distance ~ 35 km (Poe et al., 2020). Further analysis showed that fluxes have a high amount of spectral similarity, with the majority of information being at 24 and 12 h frequencies; and also at monthly and seasonal timescales. Katul et al. (2001) and Stoy et al. (2005) similarly showed that carbon fluxes as well as environmental drivers of fluxes such as air temperature, precipitation, and radiation vary on similar timescales. Spectral coherence on daily, weekly, monthly, and seasonal timescales were also observed in surface-atmosphere energy fluxes (Reed, Frank, et al., 2018), soil fluxes (Vargas et al., 2010), and aquatic fluxes (Reed, Dugan, et al., 2018), suggesting that parameter-based mechanistic modeling at discrete, coarsened time steps could miss some information from observational time-series. Applying wavelet coherence analysis on 20 ecosystem models, Stoy, Dietze, et al. (2013) found that the mechanisms driving daily and annual flux variability tend to be correct but the magnitude of these fluxes is often erroneous, particularly at seasonal and interannual timescales. However, there is an open question on what timescales and over what spectral frequencies is model performance optimal.

Recently, artificial neural networks (ANN) methods have been increasingly used to simulate annual sums, gap-fill, forecast, and spatially upscale surface-atmosphere fluxes with promising results. Using only 15 days of data for building and selecting the best models, Melesse and Hanley (2005) used environmental data to successfully simulate carbon fluxes at hourly time-scale and concluded ANN modeling was a reliable and efficient way to estimate regional carbon fluxes. Neural network methods have been shown to perform slightly better than other methods (interpolation, probabilistic filling, look-up tables, non-linear regression, and process-based models) in gap-filling of flux data (Moffat et al., 2007). Similar methods were also used to quantify uncertainties and forecast fluxes using future scenarios (Keenan et al., 2012). Papale et al. (2015) also used ANN methods to extrapolate fluxes to the continental scale, albeit with a large amount of uncertainty due to the number of sites being inadequate. An ongoing global scale project, FLUXCOM, used large scale datasets at 0.5° spatial resolution for scaling (Jung et al., 2020). Noted by Jung et al. (2020), one particular challenge is the available sample size and training length of datasets used to construct ANN models. The understanding is that more training data will improve models results, but the specific amount of training data is often not explicitly tested.

Here, we seek to model surface-atmosphere fluxes using neural networks in order to estimate fluxes without uncertainties from parameter-based models. We use paired flux sites to systemically train and test ANN models, directly testing performance over multiple lengths of input time-series. We hypothesize that fluxes can be directly modeled with the same amount of uncertainty as flux observations, at distances of ~ 10 km from the observation location based on the results of Poe et al. (2020). We use data from three paired long-term eddy covariance sites from a long-term land history study to demonstrate ANN models for simulating carbon fluxes over a 10 km spatial range. Using eight years of land-surface flux data from six flux sites in three pairs, we seek to answer the following research questions: (a) How does the length of time-series input data impact the accuracy of ANN modeled ecosystem fluxes? (b) How much improvement is observed in modeling time-series at daily time steps relative to 30-min time steps? (c) How much improvement is

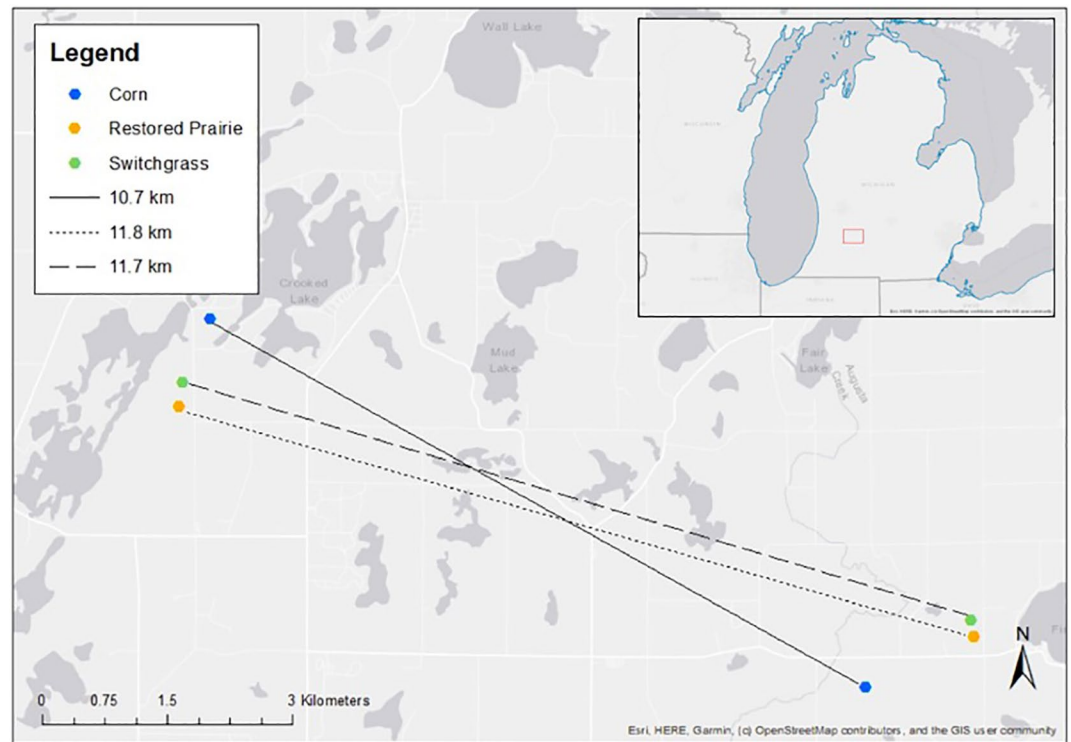


Figure 1. Map of study eddy covariance site locations, color coded by land-cover type, showing distance between paired land-cover type study sites.

observed in modeling growing seasons time scales relative to annual time scales? (d) Is ANN model performance variable across multiple land-surface types using input data from paired sites? and (e) How much is ANN model performance reduced when using non-paired land-surface data as input?

2. Methods

2.1. Site Description and Eddy Covariance Data

Data were collected from seven field sites located in Michigan, USA, at the Great Lakes Bioenergy Research Center of the Kellogg Biological Field Station's Long Term Ecological Research (KBS LTER) sites (Abraham et al., 2015, 2016; Reed et al., 2020). The climate is humid and continental, with a 30-year mean annual air temperature of 9.9°C, and monthly averages from −4.2°C in January to 22.8°C in July (KBS LTER data). Mean annual precipitation is 1,027 mm, with soils at the research site being well-drained sandy loam, classified as a Typic Hapludalf. The growing season was defined from April 1 to September 30 in this study.

The site cluster was established in 2008, with six of the sites planted as no-till soybean (*Glycine max*) in 2009. In 2010, two fields were planted to no-till continuous corn (US-KL1 and US-KM1), two to switchgrass (*Panicum virgatum*, US-KL2 and US-KM3), and two to mixed grass/forb restored-prairie (US-KL3 and US-KM2) (Figure 1). The seventh site was left as a grassland reference site and has remained unmanaged since 2009 (US-KM4). There are two groups of sites (each containing corn, switchgrass, and prairie fields) located ~11 km from the other group. Data used for analysis is from 2010 through 2018, with site data available from AmeriFlux.

Surface-atmosphere flux data were collected using eddy covariance methods. Each site was instrumented with a 3-D sonic anemometer (CSAT3, Campbell Scientific Inc.) and infrared gas analyzer (LI-7500, LI-COR Biosciences). The LI-7500 sensors were calibrated every 4–6 months and eddy covariance observations were sampled and logged at 10 Hz using a Campbell CR5000 datalogger. Ancillary measurements of net radiation (R_{net}) and air temperature (T_{air}), as well as the calculated value of atmospheric vapor pressure deficit (VPD)

were made at each site. Surface-atmosphere fluxes were processed following standard AmeriFlux guidelines, with 30-min average NEE of carbon fluxes, as well as latent (LE) and sensible heat (H) fluxes, which were computed using EdiRe (Clement, 1999). u^* filtering of low turbulence periods was done following Papale et al. (2006), removing ~25% of data with an average u^* threshold of 0.11 m s^{-1} . Fluxes were gap-filled using REddyProc (Wutzler et al., 2018) with an average of 25.3% missing data between sites. NEE was partitioned into GPP and R_{eco} following daytime partitioning methods (Lasslop et al., 2010). Thirty minute gap-filled fluxes were aggregated to daily timescales if >60% of each day had observations. Both 30-min and daily data were used in our analysis.

2.2. Artificial Neural Network Model

Model experimental design incorporated the paired flux site design of Abraha et al. (2015), with the goal of simulating carbon fluxes from matched sites (Figures 1 and 2). Papale and Valentini (2003) provide a complete background and description of neural network fundamentals as applied to eddy covariance flux datasets, with the large amount of 30-min data from multiple sites allowing creation of neural networks using a relatively limited number of input driving variables. Applications of ANN and related deep learning methods have been a new and growing tool in the earth sciences in the last decade (Reichstein et al., 2019).

With multiple neural network design options, construction of the model is largely left up to the investigator. Here, we used an input-output neural model set up as a nonlinear autoregressive model with exogenous inputs (Xie et al., 2009) to retain the time-series nature of the eddy covariance data (Reed, Frank, et al., 2018). Each model had a time-delay neural network, a specific type of feedforward model architecture, that used 3 input delays (to simulate a given flux at one time-step, using information from that time-step plus the previous three time-steps) and 20 hidden layers (output flux variable is simulated using 20 weighted connections from inputs, including input in the three previous time-steps). For example, a model simulating daily NEE would use data from that day and the previous three days as input to simulate NEE using 20 nonlinear autoregressive relationships between input data and NEE.

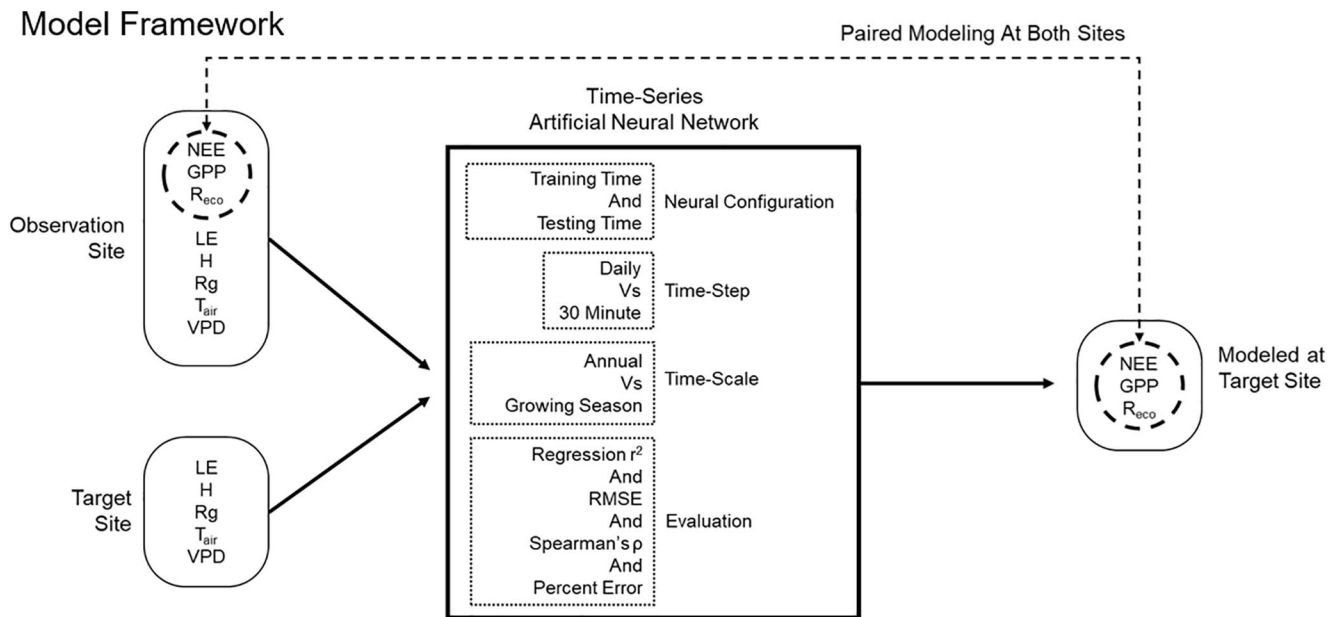
Based on this paired-site design, a model was trained and tested for each site, using input data from the target site and the paired site (Figure 2). At each site, data is used from that site and the paired land-cover-type site, using NEE, R_{eco} , GPP, LE, H, R_{net} , T_{air} , and VPD from the observation site, as well as LE, H, R_{net} , T_{air} , and VPD from the paired site. At the paired site, referred to as the target site, NEE, R_{eco} , and GPP is modeled, which then the model output of simulated fluxes at each target site was compared with measured fluxes from that target site, from the same time-period. No other ancillary or internal variables were used for modeling. This design allowed blind replicates between paired land-cover types, with one physical eddy covariance site being both an observation and target site at the same time.

Models were trained 300 times, using Levenberg-Marquardt backpropagation as a training function (Maier & Dandy, 1998). During training, mean squared error was used to define model performance and select the best option out of the 300 models. Models were trained at both the 30-min and daily timescales.

When quantifying the impact of variable training and testing time on model performance, between 1 and 4 whole-years of data was used for training, followed by between 1 and 4 whole-years of data for testing. Training always started with 2010, then chronologically successive whole years. Training, testing and validation datasets were whole-years, starting on January 1st of a given year. In this way, ANN models were always trained and then tested with observations at the chronological start of study, and validation data started between 2012 (for one year of training, one year of testing) and 2017 (four years of training, four years of testing). Each site had a total of 32 models created based on the experimental design, with sites having factors of variable training time and factors of variable testing time, for both timescales.

2.3. Statistical Performance Evaluation

Model performance across the experimental design was evaluated using correlation coefficient of determination (r^2), root mean square error (RMSE) and Spearman's ρ (also known as Spearman's rank correlation coefficient; Spearman, 1904) by comparing modeled against observed NEE, R_{eco} , and GPP of the target site. In addition, percent error was calculated using summations of fluxes as:



Experimental Data Structure

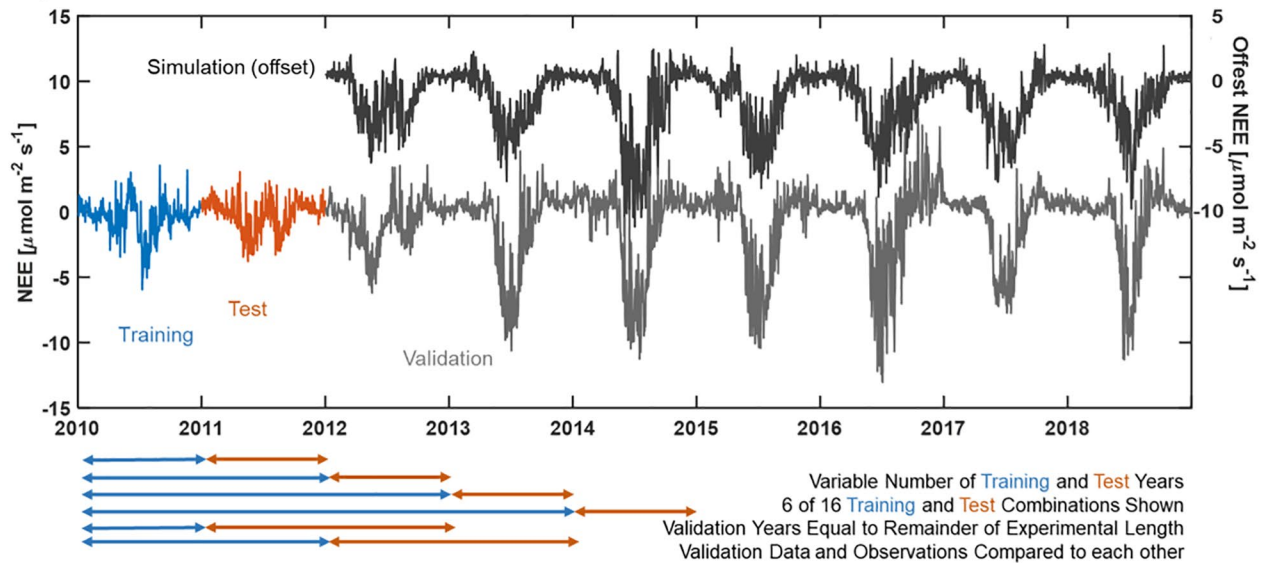


Figure 2. Conceptual framework of experimental design, showing model input data used from the observation and target sites that is, used to model carbon fluxes at the target site. Aspects of the model design are highlighted including model training and testing time, time-step and time-scale along with evaluation criteria. Also shown is an example of experimental data structure, showing measured net ecosystem exchange (NEE) fluxes from one target site, with one year of training data, one year of testing data, and seven years of fluxes use for validation of simulated NEE data. Simulation NEE data at the target site is show on an offset axis (right axis).

$$\text{Percent Error} = \frac{\sum \text{Modeled} - \sum \text{Obs}}{\sum \text{Obs}} * 100 \quad (1)$$

An $r^2 = 1$ (range: 0 to 1), $\text{RMSE} = 0$ (range: 0 to ∞), Spearman's $\rho = 1$ (range: -1 to 1), and percent error = 0% (range: 0 to 100) indicate perfect relationship between observed and modeled values. r^2 and percent error results are presented at daily and 30-min time-scales, as well as aggregated to the annual and growing season time-scales. Results are also presented by land cover type. Uncertainties in model performance metrics was calculated with models grouped by number of training and testing years (Figure 3) with all model output pooled between land-cover-type, and time-step (Figure 4) and time-scale (Figures 4 and 7).

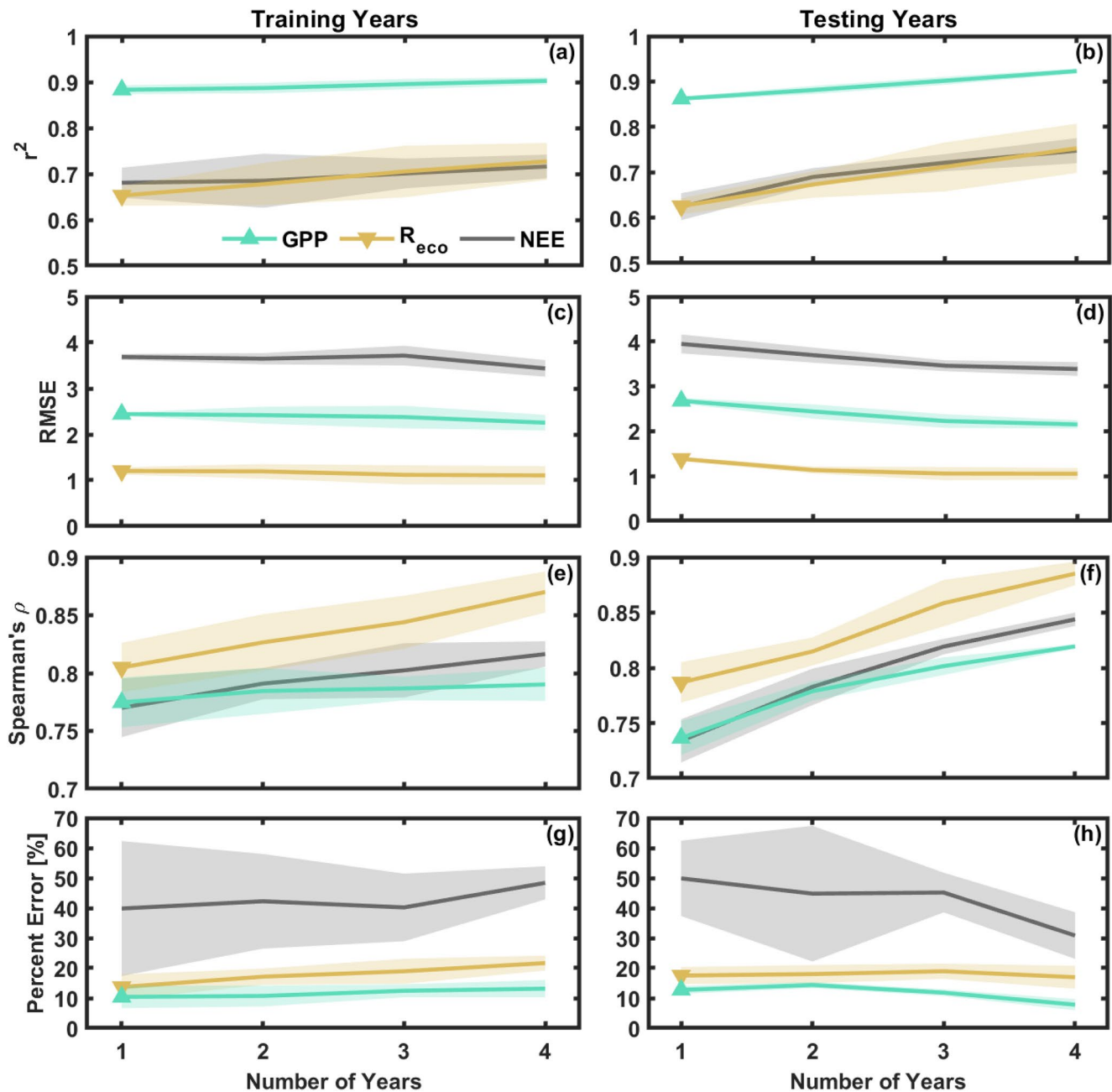


Figure 3. Average evaluation criteria of artificial neural network (ANN) training and testing results using 30-min flux data at annual time-scales showing modeled results of gross primary production (green, upward triangle), R_{eco} (yellow, downward triangle), and net ecosystem exchange (gray). Each point represents results from all possible ANN models from six sites with either 1–4 years of training years (panels a, c, e, and g) and 1–4 years of testing years (panels b, d, f, and h). Shaded areas represent standard deviations of evaluation criteria ($n = 24$, 6 site-years and 4 replicates).

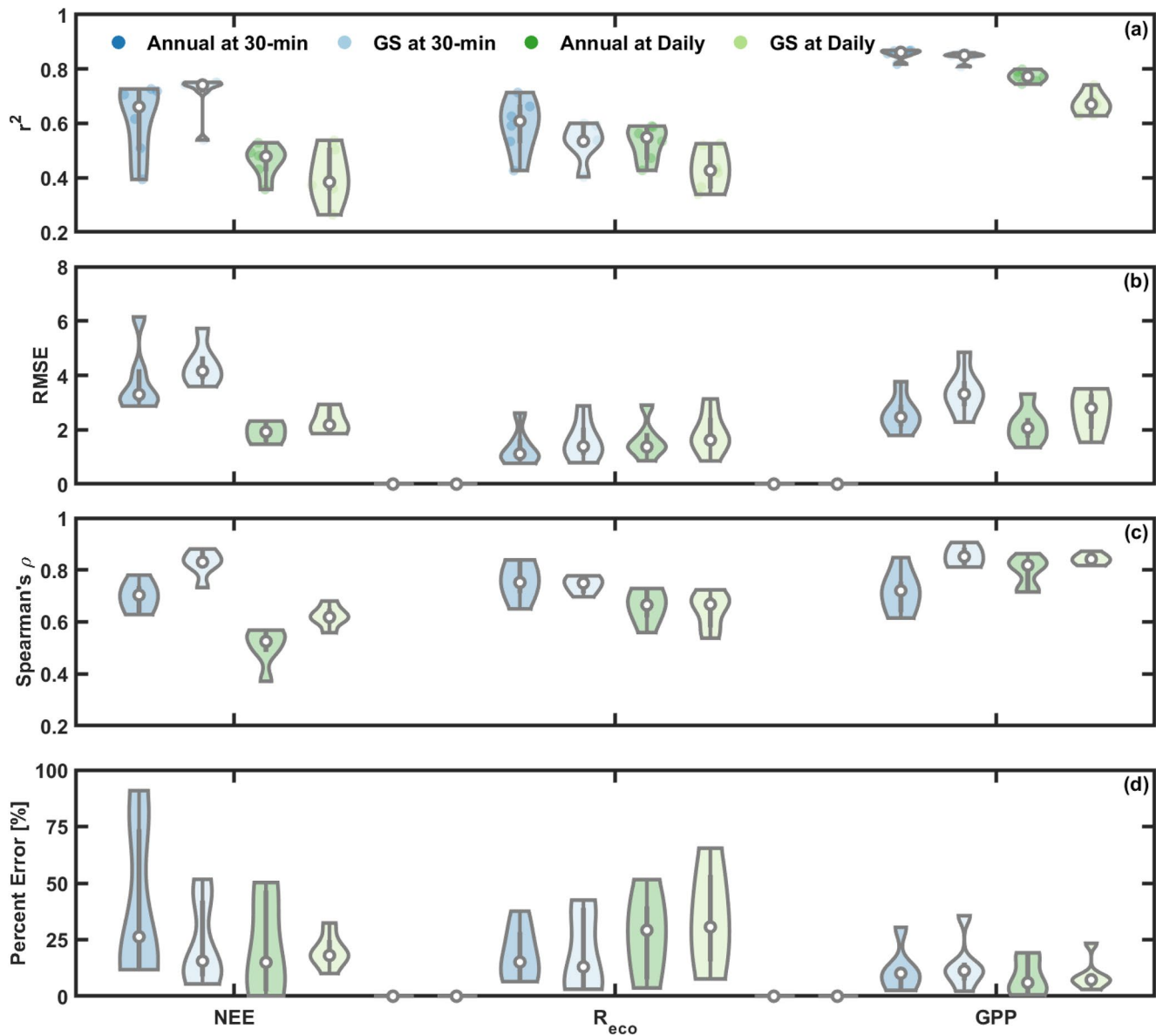


Figure 4. Violin plots showing performance results from multiple experimental design aspects (summation periods of growing season [GS] or annual as well as 30-min and daily time-step) of artificial neural network (ANN) model results. Shown are median (circle), 25th and 75th percentile (thick vertical line), 5th and 95th percentile (thin vertical line), data distribution (violin shape) of ANN model evaluation criteria of (a) r^2 , (b) root mean square error, (c) Spearman's ρ , (d) and percent error, at variable evaluation time-steps (30-minute data in blue, daily data in green) and time-scales (annual in darker colors, growing season in lighter colors). $N = 6$ for each grouping. Results shown are from models with 1 year of training data and 1 year testing data.

Uncertainty is shown as standard deviations or percentiles. MATLAB (The MathWorks, Inc., R2019a) was used for eddy covariance time-series processing and analysis and code is available at github.com/ClimateScienceResearchGroup/ANN.

3. Results

3.1. ANN Model Training and Testing Performance

To quantify the varying amounts of training and testing time in the neural network, annual time-scale summations of 30-min fluxes (Figure 3, Table 1) and daily fluxes (Table 1) show relatively similar performance metrics. With varying amounts of training data, ANN models show similar performance according to r^2 ,

Table 1

Average Evaluation Criteria of Artificial Neural Network Training and Testing Results Using 30-Min Flux Data and Daily Flux Data at Annual Time-Scales Showing Modeled Evaluation Criteria of r^2 , Root Mean Square Error (RMSE), Spearman's ρ , and Percent Error for Net Ecosystem Exchange (NEE), R_{eco} , and Gross Primary Production (GPP) Fluxes

30-Min data		Number of training years				Numbers of testing years			
		1	2	3	4	1	2	3	4
NEE	r^2	0.68 (0.03)	0.69 (0.06)	0.70 (0.03)	0.72 (0.03)	0.62 (0.03)	0.69 (0.02)	0.72 (0.02)	0.75 (0.03)
	RMSE	3.69 (0.07)	3.65 (0.12)	3.72 (0.21)	3.44 (0.18)	3.95 (0.21)	3.70 (0.17)	3.46 (0.12)	3.39 (0.15)
	Spearman's ρ	0.77 (0.03)	0.79 (0.01)	0.80 (0.02)	0.82 (0.01)	0.73 (0.02)	0.78 (0.02)	0.82 (0.01)	0.84 (0.01)
	Percent Error [%]	39.8 (22.5)	42.2 (15.8)	40.2 (11.2)	48.4 (5.5)	49.9 (12.5)	44.8 (22.6)	45.2 (6.6)	30.8 (7.8)
R_{eco}	r^2	0.65 (0.02)	0.68 (0.05)	0.70 (0.06)	0.73 (0.04)	0.63 (0.02)	0.67 (0.03)	0.71 (0.05)	0.75 (0.05)
	RMSE	1.20 (0.08)	1.20 (0.16)	1.12 (0.21)	1.11 (0.20)	1.38 (0.05)	1.13 (0.08)	1.06 (0.15)	1.05 (0.13)
	Spearman's ρ	0.80 (0.02)	0.83 (0.02)	0.84 (0.02)	0.87 (0.02)	0.79 (0.02)	0.81 (0.01)	0.86 (0.02)	0.89 (0.01)
	Percent Error [%]	13.7 (4.2)	17.1 (2.7)	18.9 (4.2)	21.7 (2.5)	17.5 (2.8)	18.0 (3.0)	18.9 (2.6)	16.9 (3.8)
GPP	r^2	0.88 (0.01)	0.89 (0.01)	0.90 (0.01)	0.90 (0.01)	0.86 (0.00)	0.88 (0.01)	0.90 (0.01)	0.92 (0.01)
	RMSE	2.44 (0.03)	2.42 (0.19)	2.38 (0.25)	2.25 (0.17)	2.68 (0.06)	2.44 (0.16)	2.23 (0.15)	2.15 (0.09)
	Spearman's ρ	0.77 (0.02)	0.78 (0.02)	0.79 (0.01)	0.79 (0.01)	0.74 (0.02)	0.78 (0.01)	0.80 (0.01)	0.82 (0.00)
	Percent Error [%]	10.4 (3.7)	10.7 (3.5)	12.4 (2.2)	13.2 (2.9)	12.7 (1.3)	14.4 (1.1)	11.8 (1.1)	7.8 (1.9)
Daily data		Number of training years				Numbers of testing years			
		1	2	3	4	1	2	3	4
NEE	r^2	0.56 (0.04)	0.56 (0.05)	0.57 (0.04)	0.57 (0.04)	0.45 (0.01)	0.56 (0.02)	0.62 (0.01)	0.63 (0.03)
	RMSE	1.77 (0.09)	1.60 (0.09)	1.58 (0.04)	1.60 (0.07)	1.71 (0.16)	1.66 (0.03)	1.57 (0.05)	1.60 (0.08)
	Spearman's ρ	0.58 (0.02)	0.58 (0.02)	0.61 (0.03)	0.59 (0.02)	0.51 (0.01)	0.58 (0.01)	0.63 (0.01)	0.63 (0.01)
	Percent Error [%]	30.0 (8.0)	35.0 (6.2)	49.3 (10.6)	47.6 (15.9)	41.0 (12.1)	33.3 (8.2)	42.6 (12.6)	45.1 (11.6)
R_{eco}	r^2	0.57 (0.02)	0.58 (0.04)	0.61 (0.07)	0.65 (0.05)	0.52 (0.01)	0.56 (0.04)	0.64 (0.05)	0.70 (0.05)
	RMSE	1.42 (0.07)	1.30 (0.13)	1.31 (0.22)	1.23 (0.11)	1.56 (0.06)	1.34 (0.13)	1.17 (0.14)	1.18 (0.13)
	Spearman's ρ	0.71 (0.01)	0.73 (0.02)	0.75 (0.03)	0.78 (0.04)	0.67 (0.01)	0.71 (0.02)	0.78 (0.02)	0.81 (0.01)
	Percent Error [%]	21.2 (3.9)	20.1 (6.4)	19.8 (8.3)	19.5 (4.5)	28.3 (5.2)	19.2 (3.4)	19.8 (4.2)	13.3 (3.3)
GPP	r^2	0.83 (0.01)	0.84 (0.01)	0.85 (0.01)	0.85 (0.02)	0.79 (0.01)	0.83 (0.01)	0.86 (0.01)	0.88 (0.00)
	RMSE	1.94 (0.12)	1.77 (0.10)	1.74 (0.10)	1.72 (0.13)	1.97 (0.11)	1.84 (0.06)	1.67 (0.04)	1.69 (0.09)
	Spearman's ρ	0.81 (0.01)	0.80 (0.03)	0.80 (0.03)	0.81 (0.01)	0.79 (0.01)	0.80 (0.01)	0.83 (0.02)	0.81 (0.03)
	Percent Error [%]	8.4 (2.0)	7.8 (5.7)	11.0 (3.6)	12.9 (1.4)	15.0 (2.9)	8.2 (3.3)	11.2 (3.3)	5.7 (2.7)

Note. Values in parenthesis represent standard deviations of evaluation criteria ($n = 24$, 6 site-years and 4 replicates).

RMSE, and percent error, but small improvements in Spearman's ρ with increasing amounts of data. With increasing amounts of testing data, model performance improved for all metrics, with the exception of percent error on GPP and R_{eco} , where large improvements were not possible due to already high-performance metrics using one year of testing data (Figures 3a and 3b). Overall, performance of GPP was higher than that of R_{eco} , with NEE performance being the lowest of all—except for Spearman's ρ where NEE and GPP were lowest (Figures 3e and 3f). NEE performance, particularly based on RMSE and percent error metrics, is penalized due to NEE values typically fluctuating around zero. This issue is present in all following NEE percent error results.

Using one year of training and testing data, annual summations of NEE, R_{eco} , and GPP can be modeled with an average accuracy of 40%, 18%, and 11%, based on 30-min time step percent error data (Figures 3g and 3h). Small improvements in model performance, primarily increases in Spearman's ρ , were noted when additional training and testing data beyond one year each was used. With that finding, and with GPP and R_{eco} percent error being well within the 20% error target, we found that it is possible to adequately model

carbon fluxes with two years of data. Hence, the remaining results are for one year of training along with one year of testing data.

3.2. ANN Model Performance at 30-min and Daily Timesteps

To compare model performance at daily time steps relative to 30-min time step, as well as annual time scales relative to growing season timescales, metrics for all possible combinations are shown in Figure 4. Model performance is generally improved when simulating annual sums of daily flux data, relative to annual sums of 30-min flux with decreases noted for r^2 and Spearman's ρ for all carbon fluxes relative to 30-min data ($p < 0.05$ for all cases), while RMSE and percent error metrics showed increases in performance for the majority of carbon fluxes. NEE showed large increases in performance for RMSE and percent error while decreased in performance using the r^2 and Spearman's ρ metrics. Percent error of annual sums of daily data were 14.9%, 29.2%, and 5.9%, for NEE, R_{eco} , and GPP fluxes.

3.3. ANN Model Performance at Annual and Growing Season Timescales

Using either 30-min or daily flux data, the difference in model performance was small if flux data were summed over annual (Figure 4, dark colors) or growing season timescales (Figure 4, light colors). NEE showed a large improvement in Spearman's ρ , from 0.53 at annual scales to 0.70 over the growing season ($p < 0.05$) at daily time steps, and from 0.62 (annual) to 0.83 (growing season) ($p < 0.05$) at 30-min time steps. The changes in other metrics were relatively smaller. Summed over the growing season, percent error of NEE, R_{eco} , and GPP daily fluxes were 18.0%, 30.6%, and 7.3%, respectively. Here, as noted above, one year of training and testing data is shown.

3.4. ANN Model Performance Across Land-Cover Types

When quantifying ANN modeled carbon fluxes from unique land-cover types (Figure 5, Table 2), percent error is shown over both annual and growing season summation periods as well as at 30-min and daily timescales. The three land-cover types (switchgrass, prairie, and corn) had two site replicates each. Over both timescales and for both summation periods, R_{eco} and GPP from switchgrass and prairie ecosystems, as well as GPP from corn, were within the 20% uncertainty target, while NEE from switchgrass and prairie and R_{eco} from corn were higher than the 20% uncertainty target. In general, model results were better at growing season than annual summation periods, and at daily than at 30-min timescale for both NEE and GPP, with the opposite being true for R_{eco} .

Simulated carbon fluxes are dynamic in time and therefore model performance appeared dependent on which year is used for testing. Using one year of training data and one year of testing data (2010 and 2011 respectively), percent error of model output for carbon fluxes is shown from 2012 through 2018 in Figure 6 and Table 3. For growing season sums, NEE was within the 20% uncertainty range, but two years were outside of the target uncertainty range at the annual scale. R_{eco} modeled results were within this uncertainty range, with the clear exception of 2014, where model uncertainty was high for both model timescales and both summation scales. At both annual and growing season summation periods, GPP was within the 20% uncertainty target for all years, with relatively small interannual variability in uncertainty.

We modeled carbon fluxes across the landscape using input data from the non-paired reference site (Figure 7). We found similar results as model performance from similar land-cover types. Model performance of r^2 and Spearman's ρ values were worse while RMSE were improved, with similar amount of variability in each of the performance metrics. NEE percent error was notably high (72.4% and 98.3% for annual and growing season time-scales), although R_{eco} (11.7% and 12.8% for annual and growing season) and GPP (6.7% and 11.9% for annual and growing season) were within the target 20% uncertainty range. For daily models, the performance of annual sum R_{eco} and GPP appeared better than that of growing season sums ($p < 0.05$ for r^2 and Spearman's ρ), while growing season sums of NEE were not significantly different than annual sums.

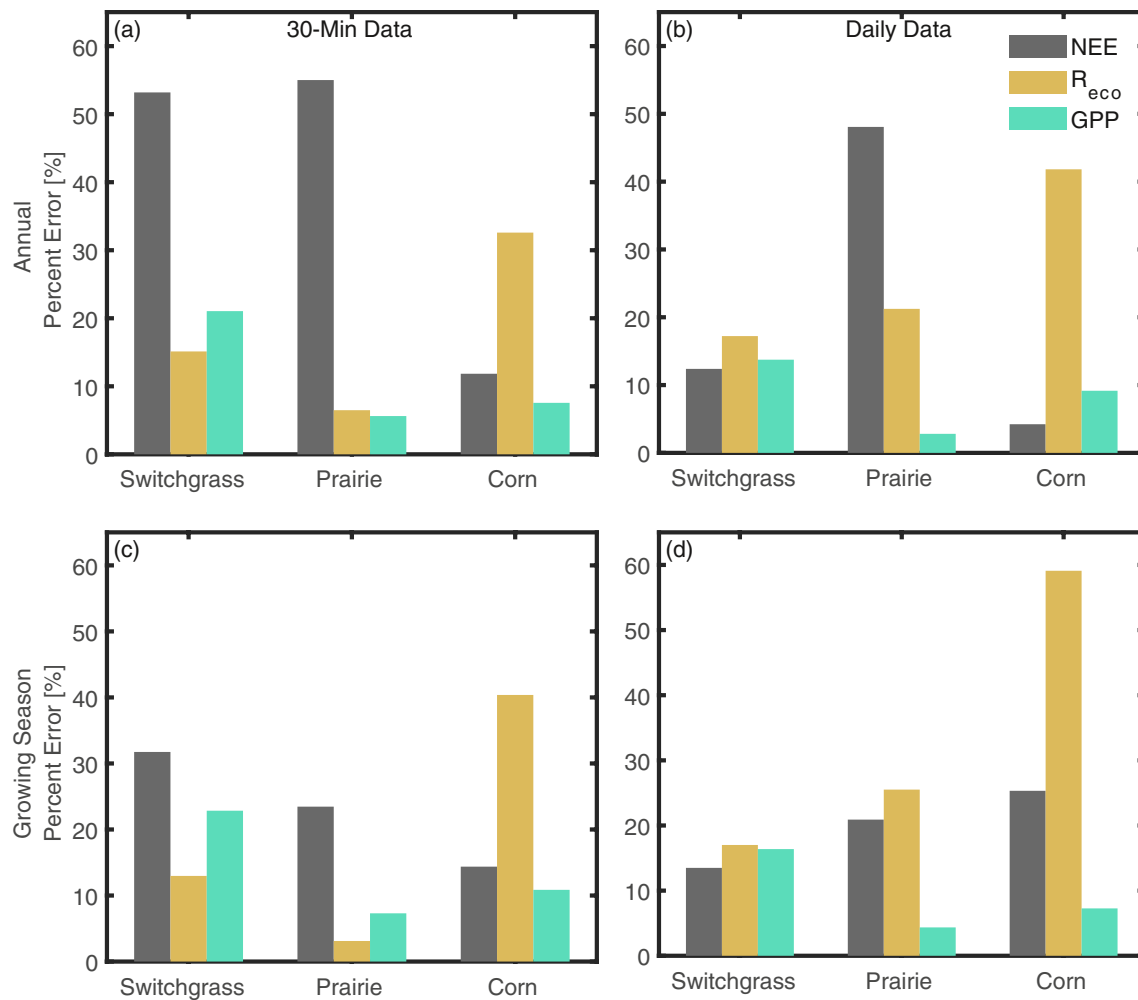


Figure 5. Model percent error of net ecosystem exchange (gray), R_{eco} (yellow), and gross primary production (green) by land-cover type (switchgrass, prairie, and corn), calculated from annual (panels a and b) and growing season summations (panels c and d), and using 30-min (panels a and c) and daily time-step data (panels b and d). Results shown are from models with 1 year of training data and 1 year of testing data.

4. Discussion

4.1. ANN Model Scaling

We show that annual and growing season flux summations can be modeled outside of the footprint well within the average 20% uncertainty of eddy covariance observations. For NEE, R_{eco} , and GPP we found uncertainties of 20%, 22%, and 8% annually and 17%, 22%, and 9% over the growing season, respectively. The same fluxes can also be modeled with uncertainties of 83%, 12%, and 10% using inputs from different land-cover types. Our findings also show that ANN modeling can be conducted with as little as two years of data, one-year data for training and one-year data for testing, although model years with unusual climatic patterns could be underconstrained in ANN models (Reichstein et al., 2019). Using landscape land-cover data, 850+ sites in the FLUXNET, 250+ sites in the AmeriFlux network, and ~200 sites in the Fluxnet2015 database (Pastorello et al., 2020) with at least two years of data, there is the potential for modeling and up-scaling fluxes beyond a site's footprint, given care is used in selection of training and testing data (Reichstein et al., 2019).

Previous works have attempted to upscale fluxes using ANN approaches with varying degrees of success. At single-site scales, Melesse and Hanley (2005) found that environmental data can be used to simulate carbon fluxes and concluded that ANN modeling was a reliable and efficient way to scale up fluxes. Papale and Valentini (2003) were one of the first to attempt continental scale land-surface flux modeling. They

Table 2

Artificial Neural Network Model Evaluation Criteria of r^2 , Root Mean Square Error (RMSE), Spearman's ρ , and Percent Error of Net Ecosystem Exchange (NEE), R_{eco} , and Gross Primary Production (GPP) by Land-Cover Type (Switch Grass, Prairie, and Corn), Calculated Over Annual and Growing Season Timescales, and Using 30-min and Daily Time-Step Data

		30-Min data			Daily data			
		Land-cover type			Land-cover type			
		Switchgrass	Praire	Corn	Switchgrass	Praire	Corn	
Annual timescale								
	NEE	r^2	0.62	0.71	0.51	0.42	0.49	0.48
		RMSE	3.30	2.97	5.15	2.26	1.54	1.92
		Spearman's ρ	0.64	0.76	0.7	0.53	0.56	0.43
	Percent Error [%]	53.2	55	11.8	12.4	48	4.2	
R_{eco}		r^2	0.6	0.67	0.51	0.58	0.56	0.45
		RMSE	0.92	1.04	2.11	1.10	1.37	2.38
		Spearman's ρ	0.79	0.74	0.73	0.70	0.69	0.59
		Percent Error [%]	15.1	6.5	32.6	17.2	21.2	42.2
GPP		r^2	0.84	0.86	0.85	0.78	0.78	0.75
		RMSE	2.19	2.26	3.30	1.86	1.79	2.82
		Spearman's ρ	0.68	0.79	0.67	0.82	0.84	0.72
		Percent Error [%]	21.0	5.6	7.6	13.7	2.8	9.2
		30-Min data			Daily data			
		Land-cover type			Land-cover type			
		Switchgrass	Praire	Corn	Switchgrass	Praire	Corn	
Growing season timescale								
	NEE	r^2	0.64	0.75	0.73	0.31	0.38	0.52
		RMSE	4.16	3.76	5.18	2.86	1.86	2.17
		Spearman's ρ	0.78	0.43	0.86	0.58	0.62	0.66
	Percent Error [%]	31.7	23.5	14.4	13.5	20.9	25.3	
R_{eco}		r^2	0.47	0.6	0.53	0.43	0.43	0.44
		RMSE	1.07	1.28	2.44	1.22	1.61	2.75
		Spearman's ρ	0.72	0.39	0.73	0.67	0.65	0.61
		Percent Error [%]	13.0	3.1	40.2	17.0	25.5	59.1
GPP		r^2	0.83	0.85	0.86	0.65	0.65	0.71
		RMSE	2.86	3.06	4.29	2.40	2.25	3.34
		Spearman's ρ	0.84	0.44	0.87	0.84	0.84	0.85
		Percent Error [%]	22.8	7.3	10.9	16.4	4.4	7.3

Note. Results shown are from models with 1 year of training data and 1 year of testing data.

found inaccuracies in annual NEE summations due to using a single network model for all sites, including known outliers and sites with divergent land management practices, information of which was not included in their model. In a follow up study, Papale et al. (2015) found that the number and distribution of sites across the scaling region to be critical for continental scaling. A similar number of input variables were needed, capturing the variability of environmental conditions that control land-surface fluxes. However, models were trained and tested at each site, and then scaled to the continent using gridded input variables. In our study, after quantifying the spatial coherence range of fluxes, we use a novel approach first to test the amount of data needed to build models, then to model fluxes beyond the extent of a site's footprint.

Using multiple machine learning methods, Tramontana et al. (2016) reported highly similar results among 11 different methodologies. In this study we used time-series based training and testing of models, that is,

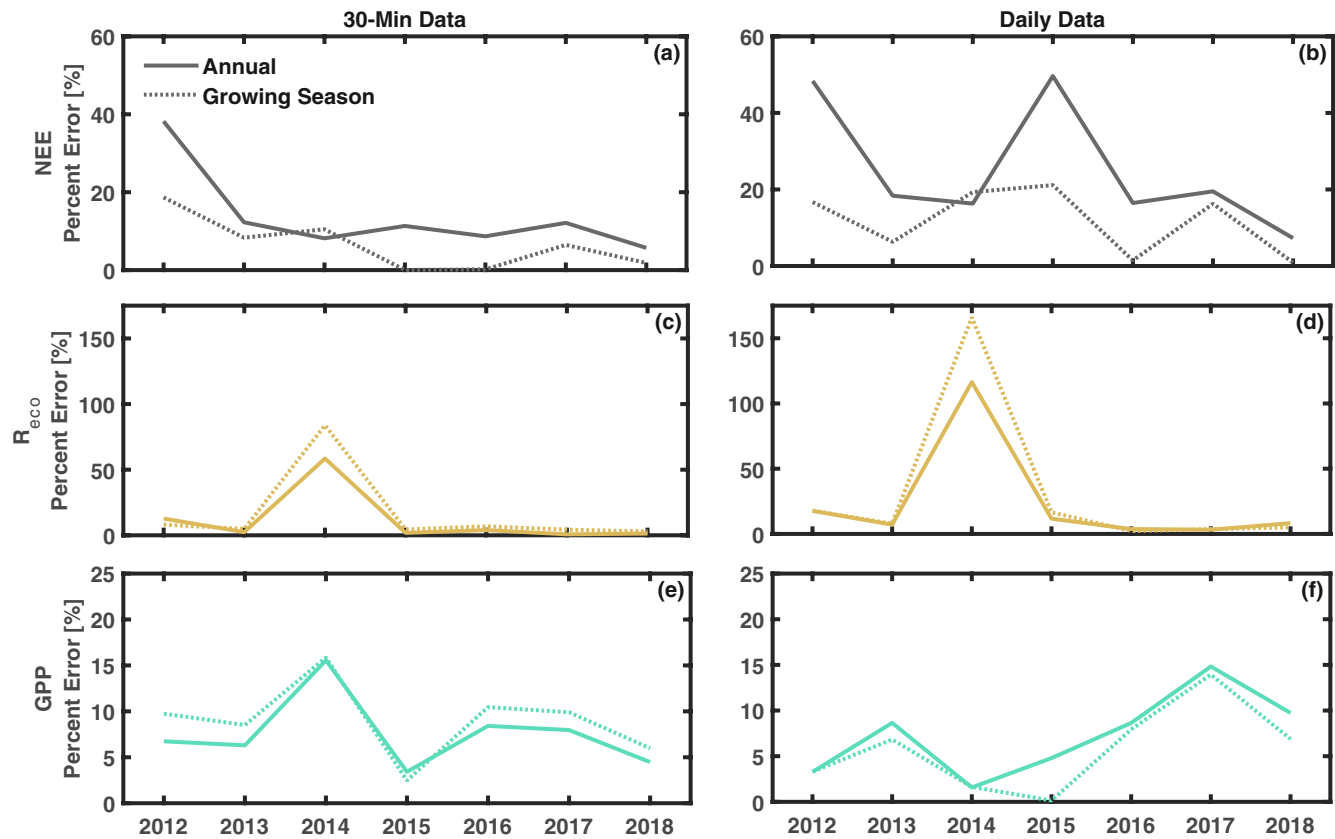


Figure 6. Time-series of model percent error for net ecosystem exchange (gray, panels (a and b)), R_{eco} (yellow, panels (c and d)), and gross primary production (green, panels (e and f)), using 30-minute (panels (a, c, and e)) and daily time-step data (panels b, d, and f), summed over annual (solid line) and growing season (dotted line) time-scales. Results shown are from models with 1 year of training data and 1 year of testing data.

the first chronological year was used for training and the following year for testing, the results might not have been impacted by that model design choice. While we show that the ANN methods need a relatively small amount of data to train and test models, limitations still exist given that interannual variation can be a large factor in terrestrial ecosystems. This suggests that additional drivers representing inter-annual variation of climate may be needed for future ANN modeling. Here, our modeled results of R_{eco} in 2014 were quite poor, while GPP and NEE were well within $\sim 20\%$ error acceptable range. In the Midwest, 2014 was an extreme cold and wet year. With one year of training, the weather extremes of 2014 were outside of the training data set (Arndt et al., 2015). Selection of a representative training database is an important consideration. Here we show one year of training can reproduce a typical year, while we take care to also note our models are still underconstrained for unusual conditions (Reichstein et al., 2019).

We chose to include H and LE fluxes as model inputs, even though this may limit the implications and use of this approach. A common source of H and LE data is eddy covariance flux measurements, but recent advances in thermal and shortwave remote sensing shows promise for water and energy fluxes to be modeled based on land surface temperature or ecosystem conductance through remote sensing models (Chen & Liu, 2020). Techniques using unmanned aerial systems has applied spatial thermal data to scale water and energy fluxes across the flux observation footprint (Wang et al., 2019). Additionally, Poe et al. (2020) shows that within spatial distances of 10's of km, H and LE should be spatially coherent. We expect that the availability and quality of spatial water and energy data will continue to rapidly improve, and therefore further exploring coupling of water and carbon fluxes through novel modeling techniques is a key area of future research (Chen & Liu, 2020).

Including additional ancillary flux measurements is one way to improve model performance. While many of the ancillary variables are cross-correlated, including additional predictors improved neural network

Table 3

Time-Series of Artificial Neural Network Model Evaluation Criteria of r^2 , Root Mean Square Error (RMSE), Spearman's ρ , and Percent Error of Net Ecosystem Exchange (NEE), R_{eco} , and Gross Primary Production (GPP), Calculated Over Annual and Growing Season Timescales, and Using 30-min and Daily Time-Step Data

Annual timescale		30-Min data							Daily data						
		2012	2013	2014	2015	2016	2017	2018	2012	2013	2014	2015	2016	2017	2018
NEE	r^2	0.56	0.77	0.59	0.61	0.71	0.78	0.65	0.30	0.51	0.52	0.55	0.44	0.63	0.40
	RMSE	2.83	2.98	3.47	4.41	3.31	2.96	3.89	1.62	1.72	1.94	2.00	1.83	1.56	1.94
	Spearman's ρ	0.72	0.73	0.58	0.69	0.72	0.78	0.76	0.42	0.54	0.48	0.58	0.50	0.61	0.46
	Percent Error [%]	38.2	12.3	8.2	11.3	8.7	12.1	5.7	48.3	18.3	16.3	49.6	16.5	19.5	7.3
R_{eco}	r^2	0.56	0.71	0.09	0.62	0.69	0.67	0.71	0.43	0.67	0.08	0.61	0.62	0.64	0.65
	RMSE	0.89	1.11	1.42	1.77	1.10	1.10	1.29	1.11	1.31	1.80	1.90	1.36	1.27	1.53
	Spearman's ρ	0.75	0.81	0.35	0.85	0.84	0.80	0.85	0.65	0.75	0.28	0.76	0.74	0.71	0.75
	Percent Error [%]	12.7	2.4	58.5	1.8	4.0	0.7	1.3	17.7	7.1	116.4	11.7	3.7	3.2	8.2
GPP	r^2	0.84	0.88	0.82	0.84	0.87	0.89	0.87	0.65	0.80	0.73	0.81	0.78	0.84	0.79
	RMSE	1.80	2.28	2.59	3.19	2.40	2.33	2.52	1.72	2.06	2.14	2.41	2.09	1.76	2.19
	Spearman's ρ	0.76	0.73	0.71	0.69	0.73	0.74	0.72	0.78	0.81	0.74	0.81	0.83	0.81	0.80
	Percent Error [%]	6.7	6.3	15.6	3.4	8.4	8.0	4.5	3.3	8.6	1.6	4.8	8.6	14.8	9.7

Growing season timescale		30-Min data							Daily data						
		2012	2013	2014	2015	2016	2017	2018	2012	2013	2014	2015	2016	2017	2018
NEE	r^2	0.63	0.79	0.61	0.71	0.73	0.80	0.78	0.25	0.42	0.42	0.48	0.35	0.56	0.42
	RMSE	3.43	3.82	4.42	5.00	4.23	3.83	4.07	1.96	2.09	2.44	2.37	2.17	1.90	2.14
	Spearman's ρ	0.81	0.87	0.73	0.82	0.83	0.86	0.88	0.44	0.66	0.59	0.70	0.58	0.73	0.60
	Percent Error [%]	18.7	8.3	10.5	0.0	0.2	6.5	1.8	16.7	6.3	19.3	21.1	1.5	16.3	0.9
R_{eco}	r^2	0.47	0.64	0.06	0.58	0.59	0.58	0.58	0.29	0.57	0.03	0.55	0.49	0.56	0.49
	RMSE	1.03	1.35	1.72	1.85	1.35	1.34	1.63	1.23	1.58	1.94	1.95	1.65	1.46	1.91
	Spearman's ρ	0.72	0.84	0.26	0.78	0.81	0.77	0.80	0.57	0.77	0.15	0.73	0.73	0.72	0.72
	Percent Error [%]	8.1	4.8	83.9	4.4	6.8	4.3	2.9	17.9	7.8	166.1	16.6	2.3	3.6	5.0
GPP	r^2	0.84	0.88	0.82	0.84	0.86	0.88	0.87	0.39	0.69	0.65	0.72	0.65	0.78	0.71
	RMSE	2.32	3.06	3.30	3.99	3.16	3.11	3.29	2.04	2.58	2.47	2.93	2.61	2.14	2.69
	Spearman's ρ	0.86	0.87	0.85	0.86	0.85	0.86	0.84	0.61	0.84	0.81	0.86	0.82	0.88	0.86
	Percent Error [%]	9.7	8.5	15.9	2.5	10.5	9.9	6.0	3.3	6.8	1.6	0.1	7.9	13.9	6.8

Note. Results shown are from models with 1 year of training data and 1 year of testing data.

models of water fluxes relative to process-based models (Talib et al., 2021). Additional training data would improve results, however there is a computational cost and data availability limitations. Here we focused on a limited training data set that could be supplemented with remote sensed products in the future (Chen & Liu, 2020). Additionally, many flux tower sites are not homogeneous (Chu et al., 2021; Giannico et al., 2018), suggesting that moving from the footprint to the landscape scale will require information of landscape composition around flux towers. Neural network models can be trained across site-to-site differences and variation when landscape information is incorporated into their design (Talib et al., 2021). Including remotely sensed data of site and landscape characteristics into the correct experimental design could go a long way to quantify model uncertainty while not excluding sites for not having ancillary measurements available.

While including H, LE, and ancillary flux measurements increased model performance, there is also the risk of overfitting the ANN models. The close proximity of the flux sites also could factor into potential model overfitting, as all the flux sites experience similar environmental conditions at both daily and seasonal timescales. Regardless of this potential limitation, geospatial coherence between fluxes has been reported

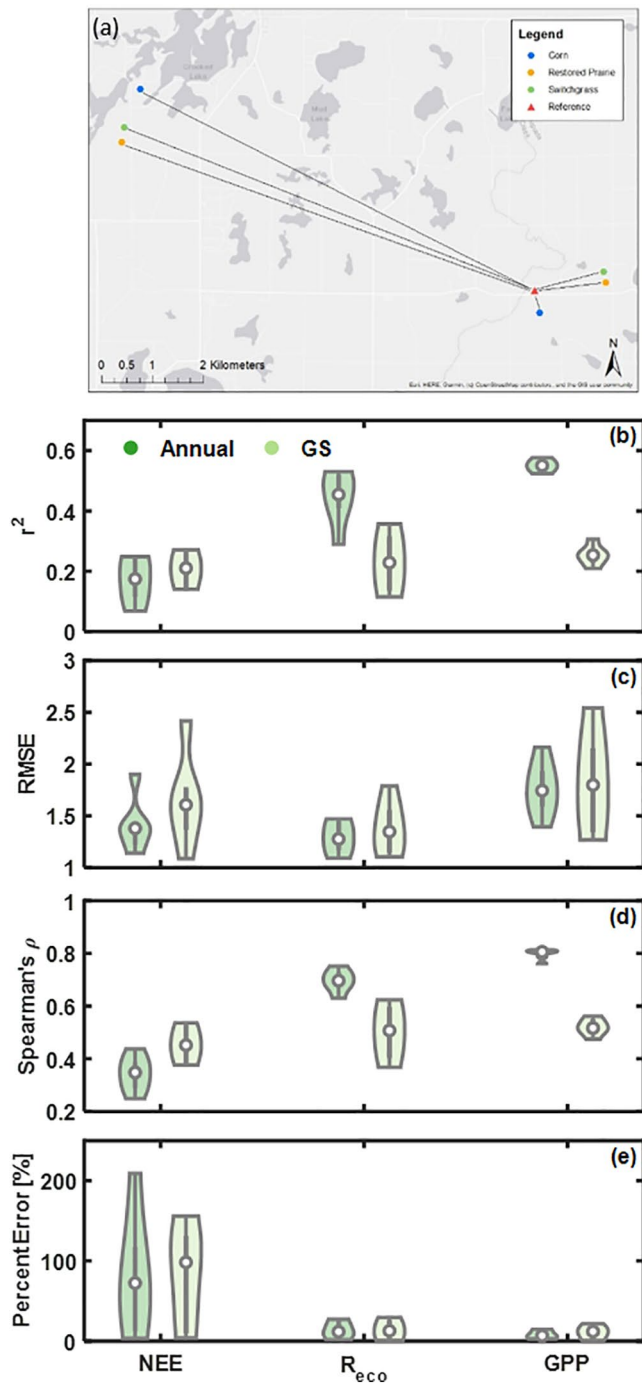


Figure 7. Map of six study sites used as observation sites, modeling the target reference site (panel (a)). Violin plots of artificial neural network model evaluation criteria at daily timescales of (b) r^2 , (c) root mean square error, (d) Spearman's ρ , and (e) percent error, at variable evaluation time-steps and time-scales (annual in dark green, growing season in light green). $N = 6$ for each grouping. Results shown are from models with 1 year of training data and 1 year or testing data.

(Poe et al., 2020) and in the future, should be explored more. Finally, modeling fluxes just beyond the edge of a site's footprint may need to be considered due to this coherent behavior.

Our work advances scaling up point observations across the landscape within similar land-cover types and shows promise for scaling across land-cover boundaries. With the goal of scaling up flux observations, similar work has been done across land-cover boundaries. Using 17 flux sites within a 5×5 km study area, the Multi-Scale Observation Experiment on Evapotranspiration was able to quantify the spatial representativeness and uncertainty of fluxes (Ran et al., 2016). They found that a single flux site, when scaled to a 25 km^2 grid scale, could overestimate NEE by more than $40 \pm 33\%$. Ran et al. (2016) concluded that the variability of NEE was primarily due to fragmentation of land-cover type within the 25 km^2 grid. Xiao et al. (2011) concluded that accurate land-cover maps are needed to scale fluxes to the regional scale. However, using model parameters derived from 17 flux sites, Xiao et al. (2011) found that detailed land-cover maps were needed for scaling as well as significant variability of parameters within a given plant functional type, suggesting that scaling up a single site using plant models could result in substantial variability in regional flux estimates. A major finding from these studies was scaling regional fluxes from a single observation was just as inaccurate as using an observation cluster to first build models and then modeling fluxes across the region. Model parameterization methods are expected to estimate fluxes more accurately, however, it is important to reiterate that model uncertainty is also included in model-based scaling which is a significant source of additive uncertainty (Keenan et al., 2011).

4.2. Scaling ANN Model Uncertainties

Attempts to scale up carbon fluxes to the landscape, regional, or continental scales had been based on either observational data or data-model fusion work with similar uncertainty of flux estimates. Combining 42 AmeriFlux sites with MODIS to make a predictive NEE model at 1×1 km resolution across North America, Xiao et al. (2008) showed that NEE could be modeled from flux data but with a percent error of $\sim 34\%$. Using ~ 25 models and ~ 40 sites of data, the North American Carbon Program synthesis project found that model performance of NEE (Schwalm et al., 2010) and GPP (Schaefer et al., 2012) fluxes were poor. The difference between NEE observations and simulations was ~ 10 times observational uncertainty, while none of the GPP models were able to simulate GPP within observed uncertainty ranges. While the uncertainties were high at the continental scale, they were largely due to model uncertainty. If such model uncertainty could be removed, the observational uncertainties of $\sim 20\%$ represent a significant improvement in scaling.

Realistically, the 20% uncertainty target used here is just a benchmark and while a more precise uncertainty target would be difficult to quantify precisely and accurately, further work is needed in this area. Uncertainty in observations has been shown to vary significantly temporally as well as from site-to-site (Loescher et al., 2006). This uncertainty target should also be different for partitioned fluxes, as GPP and R_{eco} estimates are derived from NEE (Reichstein et al., 2005). Additionally, uncertainties

for NEE uncertainties have a potential "divide by zero" issue, which is due to NEE values typically being close to zero. Uncertainty analysis has been an active area of investigation for the community with large

improvements made on reducing both random and systemic errors (Richardson et al., 2012). Other studies have reported annual carbon flux summation uncertainties ranging from 15% ($200 \pm 30 \text{ g C m}^{-2} \text{ yr}^{-1}$) (Goulden et al., 1996) to 60% ($300 \pm 180 \text{ g C m}^{-2} \text{ yr}^{-1}$) (Anthoni et al., 1999). Using energy balance closure regression slopes, a parallel way to benchmark flux uncertainty, annual energy closure ranged from 55% to 99% (Wilson et al., 2002) on a monthly basis, from energy closure of 55% in the summer and 94% in the winter (Reed, Frank, et al., 2018). One significant advancement in uncertainty analysis is standardization of flux processing, which allows for standardization of uncertainty assessment among sites (Pastorello et al., 2020). Being able to quantify uncertainty with a high degree of certainty is ultimately needed to scale point observations spatially across the landscape or region.

5. Conclusions

Ecosystem fluxes can be estimated from ANNs with uncertainties of annual and seasonal sums are on average less than 20%, approximately equal to the observational uncertainty commonly assumed for flux measurements. These metrics also hold for R_{eco} and GPP estimates with inputs from a different land-cover type. We show that a single site can estimate fluxes beyond its footprint edge across the landscape. This work could be extended by joining 1000's of site-years of eddy covariance data in the AmeriFlux database with remotely sensed or downscaled data products, which would create large scale training data to build ANN models from. Conceptually, flux observations that are on the scale of 1 km^2 could be scaled to regional or even continental scales.

Scaling of ecosystem data has been a long-standing challenge for several reasons. With large time-series datasets such as AmeriFlux and NEON, along with remote sensing products coming online in ever-finer resolutions, data availability is diminishing as a problem. Computational power is more accessible, which has allowed new methodological approaches, including neural network modeling. Current work in ecosystem forecasting has been showing great promise in scaling and modeling forward into the future; an observation today can be used to predict ecosystem response tomorrow. Here, we demonstrated, using a powerful novel approach for scaling time-series data that so-far has been limited spatially to just where the observation was made, that an observation here can be used to predict ecosystem response just over there.

Data Availability Statement

Flux data was collected with the Great Lakes Bioenergy Research Center and the Kellogg Biological Field Station's Long Term Ecological Research. All data is available from the Ameriflux Network for the following sites: US-KL1, doi.org/10.17190/AMF/1660344; US-KL2, doi.org/10.17190/AMF/1644212; US-KL3, doi.org/10.17190/AMF/1647438; US-KM1, doi.org/10.17190/AMF/1647439; US-KM2, doi.org/10.17190/AMF/1647440; US-KM3, doi.org/10.17190/AMF/1660345; US-KM4, doi.org/10.17190/AMF/1634882. Data used for this project are available in the AmeriFlux repository and Climate Science Research Group data repository (<https://sites.google.com/view/climate-science-lab/data-and-code>), while analysis code is availed at <https://github.com/ClimateScienceResearchGroup/ANN>.

Acknowledgments

This work is supported in part by the NASA Carbon Cycle & Ecosystems program (NNX17AE16G); the Great Lakes Bioenergy Research Center funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-SC0018409 and DE-FC02-07ER64494; and the Long-term Ecological Research Program (DEB 1637653) at the Kellogg Biological Station. We thank Yost R. for assistance with keeping slow ANN code running through multiple power outages.

References

- Abraha, M., Chen, J., Chu, H., Zenone, T., John, R., Su, Y. J., et al. (2015). Evapotranspiration of annual and perennial biofuel crops in a variable climate. *Gcb Bioenergy*, 7(6), 1344–1356. <https://doi.org/10.1111/gcbb.12239>
- Abraha, M., Gelfand, I., Hamilton, S. K., Shao, C. L., Su, Y. J., Robertson, G. P., & Chen, J. Q. (2016). Ecosystem water-use efficiency of annual corn and perennial grasslands: Contributions from land-use history and species composition. *Ecosystems*, 19(6), 1001–1012. <https://doi.org/10.1007/s10021-016-9981-2>
- Anthoni, P. M., Law, B. E., & Unsworth, M. H. (1999). Carbon and water vapor exchange of an open-canopied ponderosa pine ecosystem. *Agricultural and Forest Meteorology*, 95(3), 151–168. [https://doi.org/10.1016/S0168-1923\(99\)00029-5](https://doi.org/10.1016/S0168-1923(99)00029-5)
- Arndt, D., Blunden, J., Willett, K., Aaron-Morrison, A. P., Ackerman, S. A., Adamu, J., et al. (2015). *State of the climate in 2014*.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Chen, J. M., & Liu, J. (2020). Evolution of evapotranspiration models using thermal and shortwave remote sensing data. *Remote Sensing of Environment*, 237, 111594. <https://doi.org/10.1016/j.rse.2019.111594>
- Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., et al. (2021). Representativeness of eddy-covariance flux footprints for areas surrounding AmeriFlux sites. *Agricultural and Forest Meteorology*, 301–302, 108350.
- Clement, R. (1999). *EdiRe data software, v. 1.5. 0.32*. Edinburgh, UK: University of Edinburgh.

- Denny, M., & Benedetti-Cecchi, L. (2012). Scaling up in ecology: Mechanistic approaches. *Annual Review of Ecology, Evolution, and Systematics*, 43(1), 1–22. <https://doi.org/10.1146/annurev-ecolsys-102710-145103>
- Dietze, M. C., Fox, A., Beck-Johnson, L. M., Betancourt, J. L., Hooten, M. B., Jarnevich, C. S., et al. (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences of the United States of America*, 115(7), 1424–1432. <https://doi.org/10.1073/pnas.1710231115>
- Frank, J. M., Massman, W. J., & Ewers, B. E. (2013). Underestimates of sensible heat flux due to vertical velocity measurement errors in non-orthogonal sonic anemometers. *Agricultural and Forest Meteorology*, 171, 72–81. <https://doi.org/10.1016/j.agrformet.2012.11.005>
- Giannico, V., Chen, J., Shao, C., Ouyang, Z., John, R., & Laforzezza, R. (2018). Contributions of landscape heterogeneity within the footprint of eddy-covariance towers to flux measurements. *Agricultural and Forest Meteorology*, 260, 144–153. <https://doi.org/10.1016/j.agrformet.2018.06.004>
- Goulden, M. L., Munger, J. W., Fan, S. M., Daube, B. C., & Wofsy, S. C. (1996). 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>
- Hollinger, D. Y., & Richardson, A. D. (2005). Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiology*, 25(7), 873–885. <https://doi.org/10.1093/treephys/25.7.873>
- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., et al. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research*, 116(G3), G00J07. <https://doi.org/10.1029/2010jg001566>
- Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., et al. (2020). Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach. *Biogeosciences*, 17(5), 1343–1365. <https://doi.org/10.5194/bg-17-1343-2020>
- Katul, G., Lai, C.-T., Schäfer, K., Vidakovic, B., Albertson, J., Ellsworth, D., & Oren, R. (2001). Multiscale analysis of vegetation surface fluxes: From seconds to years. *Advances in Water Resources*, 24(9), 1119–1132. [https://doi.org/10.1016/s0309-1708\(01\)00029-x](https://doi.org/10.1016/s0309-1708(01)00029-x)
- Keenan, T. F., Carbone, M. S., Reichstein, M., & Richardson, A. D. (2011). The model–data fusion pitfall: Assuming certainty in an uncertain world. *Oecologia*, 167(3), 587–597. <https://doi.org/10.1007/s00442-011-2106-x>
- Keenan, T. F., Davidson, E., Moffat, A. M., Munger, 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(8), 2555–2569. <https://doi.org/10.1111/j.1365-2486.2012.02684.x>
- Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneth, A., Barr, A., et al. (2010). Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: Critical issues and global evaluation. *Global Change Biology*, 16(1), 187–208. <https://doi.org/10.1111/j.1365-2486.2009.02041.x>
- Levin, S. A. (1992). The problem of pattern and scale in ecology: The Robert H. MacArthur Award Lecture. *Ecology*, 73(6), 1943–1967. <https://doi.org/10.2307/1941447>
- Loescher, H. W., Law, B. E., Mahrt, L., Hollinger, D. Y., Campbell, J., & Wofsy, S. C. (2006). Uncertainties in, and interpretation of, carbon flux estimates using the eddy covariance technique. *Journal of Geophysical Research*, 111(D21), D21S90. <https://doi.org/10.1029/2005jd006932>
- Maier, H. R., & Dandy, G. C. (1998). Understanding the behaviour and optimising the performance of back-propagation neural networks: An empirical study. *Environmental Modelling & Software*, 13(2), 179–191. [https://doi.org/10.1016/s1364-8152\(98\)00019-x](https://doi.org/10.1016/s1364-8152(98)00019-x)
- Melesse, A. M., & Hanley, R. S. (2005). Artificial neural network application for multi-ecosystem carbon flux simulation. *Ecological Modelling*, 189(3), 305–314. <https://doi.org/10.1016/j.ecolmodel.2005.03.014>
- Moffat, A. M., Papale, D., Reichstein, M., Hollinger, D. Y., Richardson, A. D., Barr, A. G., et al. (2007). Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agricultural and Forest Meteorology*, 147(3), 209–232. <https://doi.org/10.1016/j.agrformet.2007.08.011>
- Papale, D., Black, T. A., Carvalhais, N., Cescatti, A., Chen, J., Jung, M., et al. (2015). Effect of spatial sampling from European flux towers for estimating carbon and water fluxes with artificial neural networks. *Journal of Geophysical Research: Biogeosciences*, 120(10), 1941–1957. <https://doi.org/10.1002/2015jg002997>
- Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., et al. (2006). Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: Algorithms and uncertainty estimation. *Biogeosciences*, 3(4), 571–583. <https://doi.org/10.5194/bg-3-571-2006>
- Papale, D., & Valentini, A. (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., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y. W., et al. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data*, 7(1), 225.
- Poe, J., Reed, D. E., Abraha, M., Chen, J., Dahlin, K. M., & Desai, A. R. (2020). Geospatial coherence of surface-atmosphere fluxes in the upper Great Lakes region. *Agricultural and Forest Meteorology*, 295, 108188. <https://doi.org/10.1016/j.agrformet.2020.108188>
- Ran, Y., Li, X., Sun, R., Kljun, N., Zhang, L., Wang, X., & Zhu, G. (2016). Spatial representativeness and uncertainty of eddy covariance carbon flux measurements for upscaling net ecosystem productivity to the grid scale. *Agricultural and Forest Meteorology*, 230–231, 114–127. <https://doi.org/10.1016/j.agrformet.2016.05.008>
- Reed, D. E., Chen, J., Abraha, M., Robertson, G. P., & Dahlin, K. M. (2020). The shifting role of mRUE for regulating ecosystem production. *Ecosystems*, 23(2), 359–369. <https://doi.org/10.1007/s10021-019-00407-4>
- Reed, D. E., Dugan, H. A., Flannery, A. L., & Desai, A. R. (2018). Carbon sink and source dynamics of a eutrophic deep lake using multiple flux observations over multiple years. *Limnology and Oceanography Letters*, 3(3), 285–292. <https://doi.org/10.1002/lol2.10075>
- Reed, D. E., Frank, J. M., Ewers, B. E., & Desai, A. R. (2018). Time dependency of eddy covariance site energy balance. *Agricultural and Forest Meteorology*, 249, 467–478. <https://doi.org/10.1016/j.agrformet.2017.08.008>
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., et al. (2005). On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm. *Global Change Biology*, 11(9), 1424–1439. <https://doi.org/10.1111/j.1365-2486.2005.001002.x>
- Richardson, A. D., Aubinet, M., Barr, A. G., Hollinger, D. Y., Ibrom, A., Lasslop, G., & Reichstein, M. (2012). Uncertainty quantification. In M. Aubinet, T. Vesala, & D. Papale (Eds.), *Eddy covariance: A practical guide to measurement and data analysis* (pp. 173–209). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-2351-1_7

- Richardson, A. D., & Hollinger, D. Y. (2007). A method to estimate the additional uncertainty in gap-filled NEE resulting from long gaps in the CO₂ flux record. *Agricultural and Forest Meteorology*, *147*(3), 199–208. <https://doi.org/10.1016/j.agrformet.2007.06.004>
- Running, S. W., Baldocchi, D. D., Turner, D. P., Gower, S. T., Bakwin, P. S., & Hibbard, K. A. (1999). A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. *Remote Sensing of Environment*, *70*(1), 108–127. [https://doi.org/10.1016/s0034-4257\(99\)00061-9](https://doi.org/10.1016/s0034-4257(99)00061-9)
- Schaefer, K., Schwalm, C. R., Williams, C., Arain, M. A., Barr, A., Chen, J. M., et al. (2012). A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis. *Journal of Geophysical Research*, *117*(G3), G03010. <https://doi.org/10.1029/2012jg001960>
- Schwalm, C. R., Williams, C. A., Schaefer, K., Anderson, R., Arain, M. A., Baker, I., et al. (2010). A model-data intercomparison of CO₂ exchange across North America: Results from the North American Carbon Program site synthesis. *Journal of Geophysical Research*, *115*(G3), G00H05. <https://doi.org/10.1029/2009JG001229>
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101. <https://doi.org/10.2307/1412159>
- Stoy, P. C., Dietze, M. C., Richardson, A. D., Vargas, R., Barr, A. G., Anderson, R., et al. (2013). Evaluating the agreement between measurements and models of net ecosystem exchange at different times and timescales using wavelet coherence: An example using data from the North American Carbon Program site-level interim synthesis. *Biogeosciences*, *10*(11), 6893–6909. <https://doi.org/10.5194/bg-10-6893-2013>
- Stoy, P. C., Katul, G. G., Siqueira, M. B., Juang, J.-Y., McCarthy, H. R., Kim, H.-S., et al. (2005). Variability in net ecosystem exchange from hourly to inter-annual time scales at adjacent pine and hardwood forests: A wavelet analysis. *Tree Physiology*, *25*(7), 887–902. <https://doi.org/10.1093/treephys/25.7.887>
- Stoy, P. C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., et al. (2013). A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity. *Agricultural and Forest Meteorology*, *171*, 137–152. <https://doi.org/10.1016/j.agrformet.2012.11.004>
- Talib, A., Desai, A. R., Huang, J., Griffis, T. J., Reed, D. E., & Chen, J. (2021). Evaluation of prediction and forecasting models for evapotranspiration of agricultural lands in the Midwest U.S. *Journal of Hydrology*, *600*, 126579. <https://doi.org/10.1016/j.jhydrol.2021.126579>
- Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., et al. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, *13*(14), 4291–4313. <https://doi.org/10.5194/bg-13-4291-2016>
- Vargas, R., Detto, M., Baldocchi, D. D., & Allen, M. F. (2010). Multiscale analysis of temporal variability of soil CO₂ production as influenced by weather and vegetation. *Global Change Biology*, *16*(5), 1589–1605. <https://doi.org/10.1111/j.1365-2486.2009.02111.x>
- Wang, S., Garcia, M., Bauer-Gottwein, P., Jakobsen, J., Zarco-Tejada, P. J., Bandini, F., et al. (2019). High spatial resolution monitoring land surface energy, water and CO₂ fluxes from an unmanned aerial system. *Remote Sensing of Environment*, *229*, 14–31. <https://doi.org/10.1016/j.rse.2019.03.040>
- Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, P., et al. (2002). Energy balance closure at FLUXNET sites. *Agricultural and Forest Meteorology*, *113*(1–4), 223–243. [https://doi.org/10.1016/s0168-1923\(02\)00109-0](https://doi.org/10.1016/s0168-1923(02)00109-0)
- Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., et al. (2018). Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences*, *15*(16), 5015–5030. <https://doi.org/10.5194/bg-15-5015-2018>
- Xiao, J., Davis, K. J., Urban, N. M., Keller, K., & Saliendra, N. Z. (2011). Upscaling carbon fluxes from towers to the regional scale: Influence of parameter variability and land cover representation on regional flux estimates. *Journal of Geophysical Research*, *116*(G3), G00J06. <https://doi.org/10.1029/2010jg001568>
- Xiao, J., Zhuang, Q., Baldocchi, D. D., Law, B. E., Richardson, A. D., Chen, J., et al. (2008). Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data. *Agricultural and Forest Meteorology*, *148*(11), 1827–1847. <https://doi.org/10.1016/j.agrformet.2008.06.015>
- Xie, H., Tang, H., & Liao, Y.-H. (2009). Time series prediction based on NARX neural networks: An advanced approach, paper presented at 2009 International conference on machine learning and cybernetics, IEEE.