



**HAL**  
open science

## Memory effects of climate and vegetation affecting net ecosystem CO<sub>2</sub> fluxes in global forests

Simon Besnard, Nuno Carvalhais, M. Altaf Arain, Andrew Black, Benjamin Brede, Nina Buchmann, Jiquan Chen, Jan G. P. W Clevers, Loïc P. Dutrieux, Fabian Gans, et al.

### ► To cite this version:

Simon Besnard, Nuno Carvalhais, M. Altaf Arain, Andrew Black, Benjamin Brede, et al.. Memory effects of climate and vegetation affecting net ecosystem CO<sub>2</sub> fluxes in global forests. PLoS ONE, 2019, 14 (2), 10.1371/journal.pone.0211510 . hal-02626146

**HAL Id: hal-02626146**

**<https://hal.inrae.fr/hal-02626146>**

Submitted on 26 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

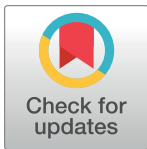


Distributed under a Creative Commons CC0 - Public Domain Dedication 4.0 International License

RESEARCH ARTICLE

# Memory effects of climate and vegetation affecting net ecosystem CO<sub>2</sub> fluxes in global forests

Simon Besnard<sup>1,2\*</sup>, Nuno Carvalhais<sup>1,3</sup>, M. Altaf Arain<sup>4</sup>, Andrew Black<sup>5</sup>, Benjamin Brede<sup>2</sup>, Nina Buchmann<sup>6</sup>, Jiquan Chen<sup>7</sup>, Jan G. P. W. Clevers<sup>2</sup>, Loïc P. Dutrieux<sup>8</sup>, Fabian Gans<sup>1</sup>, Martin Herold<sup>2</sup>, Martin Jung<sup>1</sup>, Yoshiko Kosugi<sup>9</sup>, Alexander Knohl<sup>10</sup>, Beverly E. Law<sup>11</sup>, Eugénie Paul-Limoges<sup>6</sup>, Annalea Lohila<sup>12</sup>, Lutz Merbold<sup>13</sup>, Olivier Roupsard<sup>14,15</sup>, Riccardo Valentini<sup>16</sup>, Sebastian Wolf<sup>17</sup>, Xudong Zhang<sup>18</sup>, Markus Reichstein<sup>1</sup>



**1** Department for Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Jena, Germany, **2** Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Netherlands, **3** CENSE, Departamento de Ciências e Engenharia do Ambiente, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, Caparica, Portugal, **4** School of Geography and Earth Sciences and McMaster Center For Climate Change, McMaster University, Hamilton, Ontario, Canada, **5** Faculty of Land and Food Systems, University of British Columbia, Vancouver, Canada, **6** ETH Zurich, Department of Environmental Systems Sciences, Zurich, Switzerland, **7** CGCEO/Geography, Michigan State University, East Lansing, MI, United States of America, **8** National Commission for the Knowledge and Use of Biodiversity (CONABIO), Mexico City, México, **9** Laboratory of Forest Hydrology, Graduate School of Agriculture, Kyoto University, Kyoto, Japan, **10** Faculty of Forest Sciences, University of Goettingen, Göttingen, Germany, **11** College of Forestry, Oregon State University, Corvallis, OR, United States of America, **12** Finnish Meteorological Institute, Helsinki, Finland, **13** Mazingira Centre, International Livestock Research Institute (ILRI), Nairobi, Kenya, **14** CIRAD, UMR Eco&Sols, LMI IESOL, Dakar, Senegal, **15** Eco&Sols, University Montpellier, CIRAD, INRA, IRD, Montpellier SupAgro, Montpellier, France, **16** University of Tuscia, Department for Innovation on Biological, Agro-food and Forest Systems (DIBAF), Viterbo, Italy, **17** ETH Zurich, Department of Environmental Systems Science, Physics of Environmental Systems, Zurich, Switzerland, **18** Research Institute of Forestry, Chinese Academy of Forestry, Haidian District, Beijing, P.R.China

**OPEN ACCESS**

**Citation:** Besnard S, Carvalhais N, Arain MA, Black A, Brede B, Buchmann N, et al. (2019) Memory effects of climate and vegetation affecting net ecosystem CO<sub>2</sub> fluxes in global forests. PLoS ONE 14(2): e0211510. <https://doi.org/10.1371/journal.pone.0211510>

**Received:** October 18, 2018

**Accepted:** January 15, 2019

**Published:** February 6, 2019

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

**Data Availability Statement:** Data cannot be shared publicly because of Fluxnet2015 (<https://fluxnet.fluxdata.org/data/data-policy/>) and Lathuille (<https://fluxnet.fluxdata.org/data/la-thuille-dataset/>) data policies. However, one can access these data in the same manner as the authors via the Fluxnet2015 dataset (<https://fluxnet.fluxdata.org/data/fluxnet2015-dataset/>) and the LaThuille dataset (<https://fluxnet.fluxdata.org/data/la-thuille-dataset/>). The authors did not have any special access privileges that others would not have.

**Funding:** The author(s) received no specific funding for this work.

\* [sbesnard@bgc-jena.mpg.de](mailto:sbesnard@bgc-jena.mpg.de)

## Abstract

Forests play a crucial role in the global carbon (C) cycle by storing and sequestering a substantial amount of C in the terrestrial biosphere. Due to temporal dynamics in climate and vegetation activity, there are significant regional variations in carbon dioxide (CO<sub>2</sub>) fluxes between the biosphere and atmosphere in forests that are affecting the global C cycle. Current forest CO<sub>2</sub> flux dynamics are controlled by instantaneous climate, soil, and vegetation conditions, which carry legacy effects from disturbances and extreme climate events. Our level of understanding from the legacies of these processes on net CO<sub>2</sub> fluxes is still limited due to their complexities and their long-term effects. Here, we combined remote sensing, climate, and eddy-covariance flux data to study net ecosystem CO<sub>2</sub> exchange (NEE) at 185 forest sites globally. Instead of commonly used non-dynamic statistical methods, we employed a type of recurrent neural network (RNN), called Long Short-Term Memory network (LSTM) that captures information from the vegetation and climate's temporal dynamics. The resulting data-driven model integrates interannual and seasonal variations of climate and vegetation by using Landsat and climate data at each site. The presented

**Competing interests:** The authors have declared that no competing interests exist.

LSTM algorithm was able to effectively describe the overall seasonal variability (Nash-Sutcliffe efficiency, NSE = 0.66) and across-site (NSE = 0.42) variations in NEE, while it had less success in predicting specific seasonal and interannual anomalies (NSE = 0.07). This analysis demonstrated that an LSTM approach with embedded climate and vegetation memory effects outperformed a non-dynamic statistical model (i.e. Random Forest) for estimating NEE. Additionally, it is shown that the vegetation mean seasonal cycle embeds most of the information content to realistically explain the spatial and seasonal variations in NEE. These findings show the relevance of capturing memory effects from both climate and vegetation in quantifying spatio-temporal variations in forest NEE.

## Introduction

Forests cover about 30% of the terrestrial surface of our planet, accounting for 75% of gross primary production (GPP), and store 45% of all terrestrial carbon (C) [1–3]. This fundamental role highlights the importance of understanding forest C dynamics, which are generally driven by climatic conditions and vegetation dynamics as well as natural and anthropogenic disturbances [4–6]. Changes in climate and disturbance regime can influence the development, structure, and functioning of forest ecosystems [7–12], therefore causing anomalies in the net carbon dioxide (CO<sub>2</sub>) exchange of terrestrial ecosystems (NEE). As a result, quantifying the effects of climatic variations and forest disturbances on biosphere-atmosphere CO<sub>2</sub> fluxes across-scales has considerable importance for understanding the net CO<sub>2</sub> balance of forest ecosystems [13–16].

Disturbances, such as fire, disease, insect outbreaks, drought, windthrow, or harvesting, can shift forest ecosystems into early stages of ecological succession [17, 18]. These events can potentially trigger an accelerated release of stored C back to the atmosphere by reducing the amount of photosynthetic tissue and also by increasing the pool of respiring detritus material for subsequent gradual release [14, 19–21]. During recovery, forests accumulate biomass and potentially sequester C from the atmosphere at rates that could alter current trends of atmospheric C cycling [10]. Post-disturbance successional trajectories are often complex, depending on pre-disturbance forest structure and function, disturbance type, frequency, and intensity [22, 23] as well as the climate of the region [24, 25] and land management. Some disturbances, such as insect outbreaks, can slow down recovery process during regeneration or transform forests from closed to open canopies [26], while other low to moderate severity disturbances increase structural complexity leading to less of an impact on mid-succession net primary productivity than is often assumed [27]. Therefore, climate and disturbance regimes contribute to interannual and seasonal variations in forest net CO<sub>2</sub> fluxes. Changes in climate may also exacerbate the frequency and intensity of extreme meteorological events (e.g. droughts, [6] or associated fire events [28, 29]), thereby increasing both mortality rates and the vulnerability of forest ecosystems [30], which would necessarily impact the dynamics of ecosystem C cycle.

Current response patterns observed in forest CO<sub>2</sub> fluxes depend on the contemporaneous environment conditions as well as on the so-called memory effects of disturbances, climatic variation, and their interactions [30, 31]. In fact, disturbances or climate extreme events exert both instantaneous and lagged impacts on biosphere-atmosphere C fluxes [6, 32]. Memory effects can be defined as the influence that past events have on the present or future responses of an ecosystem to environmental conditions. Extensive research has been done to understand

climate and disturbance memory effects on CO<sub>2</sub> fluxes (i.e. NEE, gross primary productivity, and ecosystem respiration) [33–38]. However, given the complexity of NEE responses to disturbances and climate extremes and highly non-linear processes, the legacies of these events on CO<sub>2</sub> fluxes remain unclear [32, 39], and thus they are rarely implemented in current C cycle models. As such, statistical models capable of dynamically incorporating temporal information related to episodic disturbances and climatic fluctuations are required to enhance our understanding and predictive capabilities of the global C budget [8, 40]. Recently, deep learning (DL) techniques, such as Recurrent Neural Networks (RNNs), have shown the potential to capture long-term temporal dependencies and variable-length observations [41–43]. Yet, DL is early in its application for CO<sub>2</sub> flux predictions [44]; questions related to the potential of extracting temporal information for estimating CO<sub>2</sub> fluxes across-scales have yet to be investigated.

In this study, we explore the potential of a dynamic statistical approach—a type of RNN, called Long Short-Term Memory model (LSTM)—to characterize the memory effects of disturbance and climate variations on NEE across temporal and spatial scales at 185 forest and woodland FLUXNET sites globally utilizing remote sensing, climate, and eddy-covariance (EC) flux datasets. In particular, this study focuses on: (1) comparing the statistical power of an LSTM approach to a Random Forest algorithm in predicting ecosystem level NEE, and (2) assessing the importance of capturing the memory effects of vegetation and climate to predict forest NEE using data-driven LSTMs. The analysis focuses on the variations in NEE spatially and temporally for seasonal, monthly, and interannual anomalies, for which a factorial experiment was designed as explained below. We propose that the application of dynamic statistical approaches results in estimating net CO<sub>2</sub> fluxes across-scales more realistically by including the responses of NEE to antecedent climate and disturbance conditions.

## Materials and methods

### Data materials

**Eddy-covariance data and quality check.** The current dataset consists of 185 forest and woodland sites (S1 Table) composed of five plant functional types (PFTs): deciduous broadleaf forest (DBF;  $n = 42$ ), deciduous needleleaf forest (DNF;  $n = 1$ ), evergreen broadleaf forest (EBF;  $n = 27$ ), evergreen needleleaf forest (ENF;  $n = 81$ ), mixed forest (MF;  $n = 14$ ), woody savanna (WSA;  $n = 10$ ), and savanna (SAV,  $n = 10$ ); and four climate class: arid ( $n = 11$ ), boreal ( $n = 67$ ), temperate ( $n = 86$ ), and tropical ( $n = 21$ ). We aggregated DBF and DNF into a deciduous forest class, EBF and ENF into an evergreen forest class, and SAV and WSA into a savanna class. The sites were part of both version 2 of the LaThuile FLUXNET and the FLUXNET2015 datasets (<https://fluxnet.fluxdata.org>) of the FLUXNET network [45, 46]. For each site, continuously measured or gap-filled net CO<sub>2</sub> fluxes (i.e. NEE) and microclimatic variables (i.e. air temperature ( $T_{air}$ ), precipitation (P), global radiation (R<sub>g</sub>), and vapor pressure deficit (VPD)) were obtained at half-hourly time intervals from the FLUXNET datasets. The data processing included: storage-correction despiking, u-filtering [47], flux partitioning [48]. Half-hourly NEE were aggregated into monthly averages (i.e. seasonal cycle). Only monthly observations with more than 80% of the original or good quality gap-filled data were considered in this analysis [47]. A total of  $\approx 14,000$  observed or gapfilled monthly NEE flux data was used, from which  $\approx 1,500$  monthly observations were collected in forests younger than 30 years.

**Remote sensing data.** For each FLUXNET site, the entire multi-temporal collection 1 from the Landsat 4, 5, 7 and 8 archives (<https://www.usgs.gov/>) spanning the past 30 years at 30 meters resolution was collected. Landsat data have been pre-processed using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) [49] and the Landsat Surface

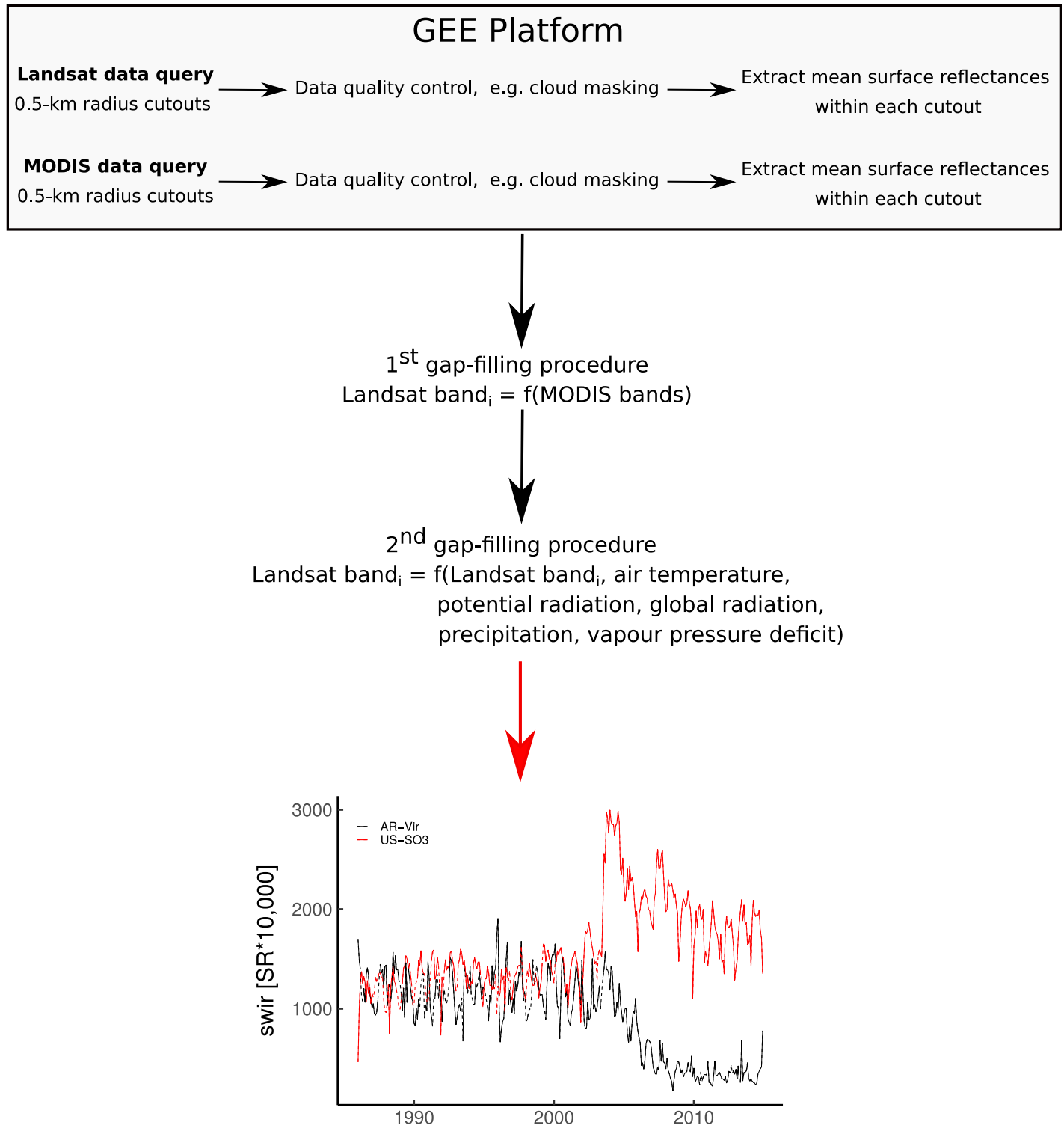
Reflectance Code (LaSRC) (<https://landsat.usgs.gov/landsat-surface-reflectance-data-products>) for atmospheric correction. Poor quality retrievals due to the clouds, cloud shadows, snow, and ice were masked out [50, 51]. The data extraction and the pre-processing chains (i.e. cloud, cloud shadow masking, and downloading) were implemented in the Google Earth Engine (GEE) platform [52] (<https://earthengine.google.com/>). The seven spectral bands of the Landsat product were used; i.e. blue, green, red, near infrared (NIR), shortwave infrared (SWIR) 1, SWIR 2, and thermal infrared (TIR) (<https://landsat.usgs.gov/what-are-band-designations-landsat-satellites>). To better represent the EC footprint area, a circular buffer of 500 m radius centered on each FLUXNET tower was defined for which a mean value within the different Landsat cutouts was extracted. Note that the proposed LSTM approach can only be implemented with regular time series, but most of the Landsat time series were irregular due to cloud cover or data quality issues. A first gap-filling procedure was conducted by predicting monthly reflectance values for each Landsat band with a Random Forest (RF) model [53, 54] using the monthly Moderate Resolution Imaging Spectroradiometer (MODIS, MCD43A4 version 6) bands as predictive variables (S1 Fig). The gap-filling procedure was completed for the remaining gaps in the entire Landsat time series (i.e. from the 1980s to now) by predicting each Landsat band with an RF model using climate variable (i.e.  $T_{air}$ , Precip, Rg, VPD, rpot), PFT, month of the year, and latitude as predictive variables (Fig 1 and S2 Fig). For the two aforementioned gap-filling procedures, the best set of the predictors for predicting each Landsat band independently was obtained with a feature selection analysis (i.e. the Boruta algorithm [55]).

**Climate data.** Long-term time series of  $T_{air}$ , P, Rg, and VPD were down-scaled for the period of 1979–2015 from the ERA-Interim datasets [56]. For each site, the three nearest grid cells in the ERA-Interim datasets were extracted and several statistical models were trained (i.e. relational logistic regression, kernel ridge regression, Gaussian processes regression, and neural networks) for each target variable (i.e.  $T_{air}$ , P, Rg, and VPD) using the time series of the three nearest gridcells as predictive variables. For each target variable and at each site, the best statistical model was consequently selected based on the highest Nash-Sutcliffe modeling efficiency (NSE) (S3 Fig). These down-scaled climate time series were used to gap-fill climate observations measured at the site level in order to have climatic data covering the entire remote sensing data period.

## Recurrent neural network model for dynamic modeling

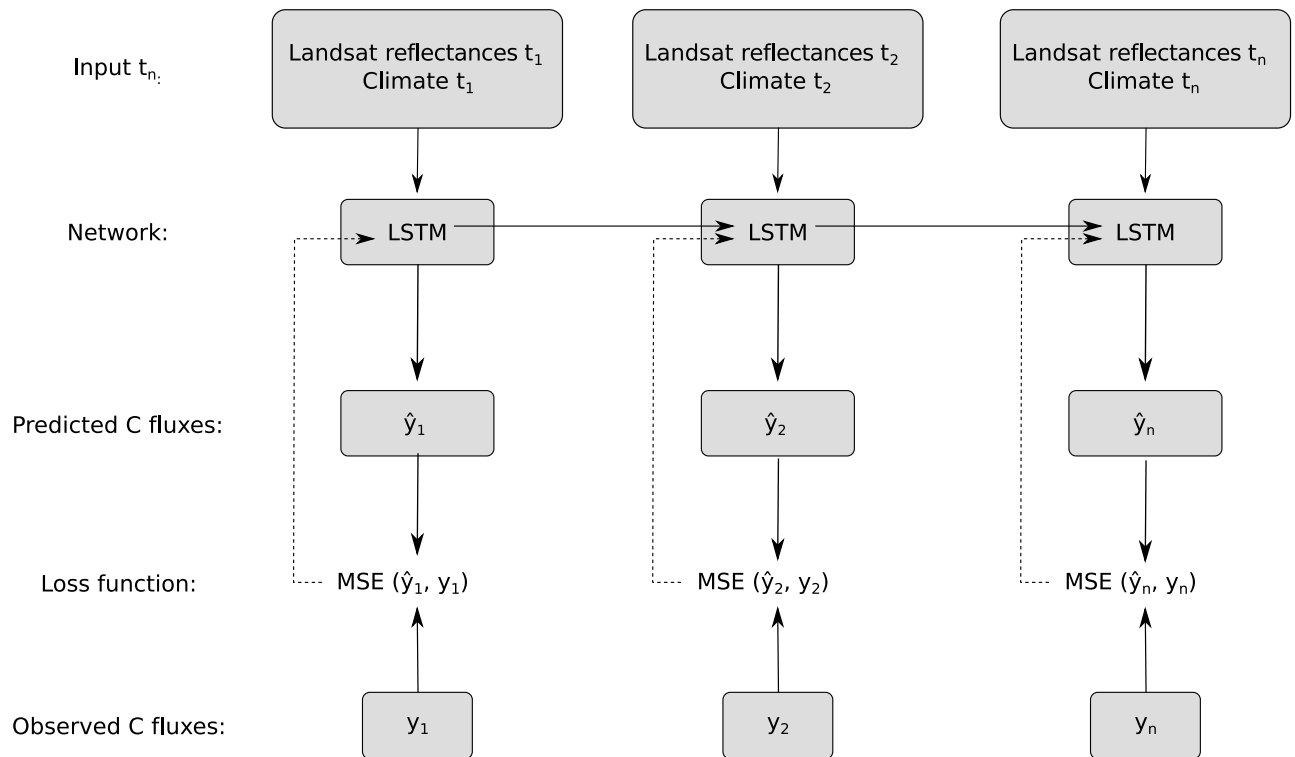
RNNs were employed to learn vegetation and climate history based on sequential observations (<https://github.com/bgi-jena/RNNFluxes.jl.git>) [44]. RNNs are effective tools for capturing temporal information from sequential/time series data. We used a special kind of RNNs, capable of learning long-term dependencies, called long short-term memory networks (LSTMs) [57]. LSTMs utilize relevant information from all previous observations and are suitable to model long-term temporal dependencies and memory effects.

Monthly climate data (i.e.  $T_{air}$ , P, Rg, and VPD) and Landsat raw bands from the period of 1982 to 2015 were used to train the LSTM models (Fig 2). A single layered LSTM was used to learn information based on the input of the current and of all previous observations. At each training iteration, a loss function (Mean Squared Error) was calculated by comparing monthly predicted and observed NEE. The loss function was used to derive the gradients for backpropagation over the entire sequence [58]. The gradients were further used by an *Adam* optimizer [59] for adjusting the weights of the connections in the model so as to minimize the loss function. During the learning procedure, 20% of the training set (i.e. evaluation set) served for optimizing the weights of the networks. The learning procedure was stopped



**Fig 1. Flowchart of the Landsat data extraction and post-processing.** SWIR = Shortwave Infrared. SR = Surface Reflectance. Monthly temporal gap-filled Landsat time series from 1982 to 2015 of the shortwave Infrared band are shown for AR-Vir and US-SO3 sites where, respectively, afforestation-reforestation and fire followed by a regrowth were reported in 2003. The solid and the dashed lines depict the real observations and the gap-filled data, respectively.

<https://doi.org/10.1371/journal.pone.0211510.g001>



**Fig 2. Flowchart of the proposed LSTM approach.** Figure adapted from [61]. Each individual timestep is a monthly observation for the period 1982 to 2015. Landsat surface reflectances correspond to the seven spectral bands of the Landsat product; i.e. blue, green, red, near infrared, shortwave infrared 1, shortwave infrared 2, and thermal infrared. Climate corresponds to air temperature, precipitation, global radiation, and vapor pressure deficit. At each time step, an LSTM layer containing a set of cells or hidden neurons (10, 20, or 30) processes information based on the input of the current and of all previous observations. Predictions of net ecosystem exchange were performed at each monthly timestep by using information from both current and previous observations. The loss function was only calculated when net ecosystem exchange observations were available; i.e. measurement periods of LaThuile and FLUXNET2015 datasets.

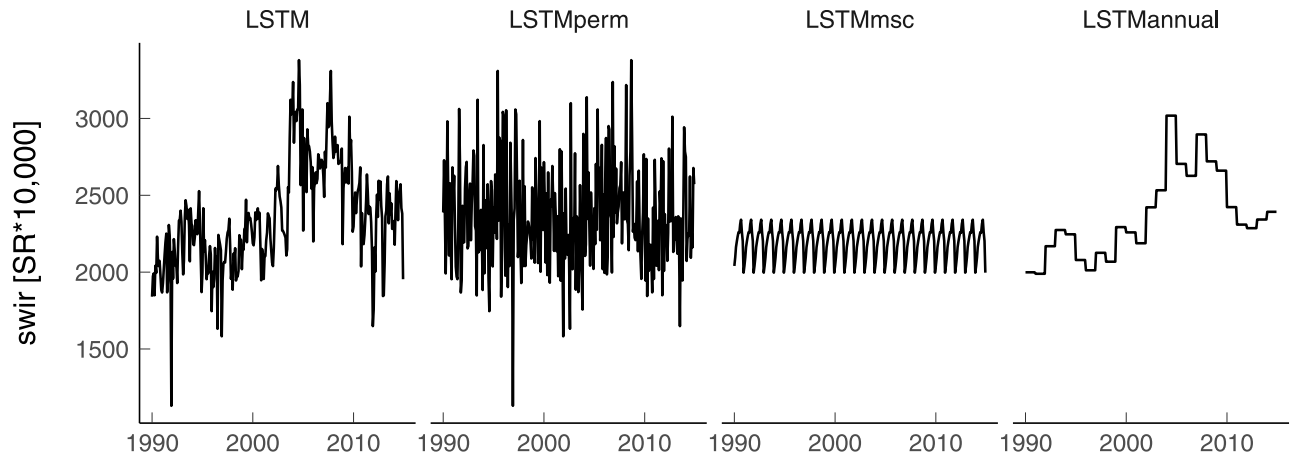
<https://doi.org/10.1371/journal.pone.0211510.g002>

when the calculated loss function on the evaluation set did not decrease after 500 iterations (i.e. early stopping). Additionally, there was a grid search of the LSTM’s hyperparameters; i.e. learning rate (0.1 or 0.01), number of hidden neurons (10, 20, or 30), and dropout (0 or 0.5) [60] to select the optimal set of hyperparameters. Due to the random initialization of LSTMs, 50 runs for each model set-up were performed to assess the uncertainty of the model outputs.

**Table 1. Design of the factorial experiment.** X means that the variant was used to study the respective topic of each row. *LSTM* = LSTM model using the full depth of the Landsat time series and climate data; *LSTM<sub>perm</sub>* = LSTM model but the temporal patterns of both the predictive and the target variables were randomly permuted while instantaneous relationships between predictive and target variables were kept; *LSTM<sub>msc</sub>* = LSTM model but the Landsat time series for each band were replaced by their mean seasonal cycle, while using the actual values of air temperature ( $T_{air}$ ), precipitation (P), global radiation (Rg), and vapor pressure deficit (VPD); *LSTM<sub>annual</sub>* = LSTM model but the Landsat time series for each band were replaced by their annual mean, while using the actual values of  $T_{air}$ , P, Rg, and VPD, RF = Random Forest model using the actual values of the Landsat time series and climate data.

	LSTM	LSTM <sub>perm</sub>	LSTM <sub>msc</sub>	LSTM <sub>annual</sub>	RF
Temporal feature extraction/Memory effects	X	X			X
Vegetation interannual seasonal variation	X		X		
Vegetation interannual variability	X			X	
Comparison to non-dynamic method	X				X

<https://doi.org/10.1371/journal.pone.0211510.t001>



**Fig 3. Illustration of the different Landsat time series temporal architectures of the different LSTM model set-ups for the SWIR band only for the period 1990–2015.** SWIR = Short Wave Infrared. SR = Surface Reflectance. The US-SO3 site where fire followed by regrowth was reported in 2003 is shown.

<https://doi.org/10.1371/journal.pone.0211510.g003>

### Experimental design

**Model set-ups.** In order to understand vegetation and climate memory effects on NEE, a trained LSTM with monthly climatic data (i.e.  $T_{air}$ , P, Rg, and VPD) and monthly Landsat data (i.e. blue, green, red, NIR, SWIR 1, SWIR 2, and TIR bands) was benchmarked against a series of different model set-ups (Table 1 and Fig 3). A comparison of the following five distinct experimental set-ups was performed: (1) LSTM model using the full depth of the Landsat time series and climate data (hereafter *LSTM*); (2) *LSTM* model where the orders of the predictor-target pairs were randomly permuted so that the instantaneous link between dependent and independent variables were kept but the realistic temporal sequences were destroyed (hereafter *LSTM<sub>perm</sub>*); (3) *LSTM* model where the Landsat time series for each band were replaced by their mean seasonal cycle (i.e. mean of each month over the Landsat time series period) while using the actual values of  $T_{air}$ , P, Rg, and VPD (hereafter *LSTM<sub>msc</sub>*); (4) *LSTM* model where the Landsat time series for each band were replaced by their annual mean (i.e. mean of the monthly observations within each year over the Landsat time series period) while using the actual values of  $T_{air}$ , P, Rg, and VPD (hereafter *LSTM<sub>annual</sub>*), and (5) a Random Forest (RF) model [53, 54] using the actual values of the Landsat time series and climate data (hereafter *RF*).

Comparing the *LSTM* with the *LSTM<sub>msc</sub>* model set-ups served to assess the importance of including and extracting information on interannual seasonal variation of vegetation to calculate NEE for each forest site across the globe. Contrasting the *LSTM<sub>annual</sub>* with the *LSTM* reflects lost in model fitness by not including the information contained in the monthly mean seasonal cycle of vegetation. The differences between the results from the *LSTM* and the *LSTM<sub>perm</sub>* as well as between the *LSTM* and the *RF* aimed to test the effects of extracting realistic temporal dependencies in the observations for predicting net CO<sub>2</sub> fluxes. The *RF* set-up also provided a comparison to commonly used data-driven statistical modeling approaches for NEE estimates [62, 63] (Table 1). The predictive variables used in the different model set-ups are listed in Table 2.

**Model training and evaluation.** The performance of each model set-up was evaluated by directly comparing model estimates with observed values of NEE for each site. These statistical models were evaluated using a 10-fold cross-validation strategy in which entire sites were assigned to each fold [63]. The training of each model set-up was done using data from  $n_{fold}-1$ ,



**Table 2. List of predictors used in the different model set-ups.**  $T_{air}$  = air temperature, P = precipitation, Rg = global radiation, and VPD = vapor pressure deficit. *LSTM* = LSTM model using the full depth of the Landsat time series and climate data; *LSTM<sub>perm</sub>* = *LSTM* model but the temporal patterns of both the predictive and the target variables were randomly permuted while instantaneous relationships between predictive and target variables were kept; *LSTM<sub>msc</sub>* = *LSTM* model but the Landsat time series for each band were replaced by their mean seasonal cycle, while using the actual values of air temperature ( $T_{air}$ ), precipitation (P), global radiation (Rg), and vapor pressure deficit (VPD); *LSTM<sub>annual</sub>* = *LSTM* model but the Landsat time series for each band were replaced by their annual mean, while using the actual values of  $T_{air}$ , P, Rg, and VPD, RF = Random Forest model using the actual values of the Landsat time series and climate data.

Model set-up	Predictors
<i>LSTM</i>	7 Landsat bands + $T_{air}$ + P + Rg + VPD
<i>LSTM<sub>perm</sub></i>	permuted [7 Landsat bands + $T_{air}$ + P + Rg + VPD]
<i>LSTM<sub>msc</sub></i>	7 MSC Landsat bands + $T_{air}$ + P + Rg + VPD
<i>LSTM<sub>annual</sub></i>	7 annual mean Landsat bands + $T_{air}$ + P + Rg + VPD
RF	actual values of 7 Landsat bands + $T_{air}$ + P + Rg + e VPD

<https://doi.org/10.1371/journal.pone.0211510.t002>

while predictions were made for the remaining fold, ensuring that the validation data were independent of the training data. The statistics used to assess the capability of the statistical models to estimate NEE were the coefficient of determination ( $R^2$ ), the NSE, the root mean squared error (RMSE), and the mean absolute error (MAE) [64]. The predictive capacity of the different algorithms was assessed for the seasonal cycle, the seasonal anomalies, the interannual anomalies, and the across-site variability. The seasonal anomalies were computed as the difference between the monthly NEE estimates of a considered month and those of the same month averaged over the observation period for each site. The interannual anomalies were computed as the difference between the annual NEE estimates of a considered year and the annual averaged over the entire observation period for each site. Both seasonal and interannual anomalies were calculated only for the sites with at least three years of complete observations after the data quality check. Across scales, the statistical models were trained using monthly time-series and the estimates were further aggregated to the corresponding scales, i.e. seasonal cycle, seasonal and interannual anomalies, and across-site.

Results were analyzed on the global dataset as well as according to PFT, bioclimatic and, forest age classes. PFT and climate classifications were found in the ancillary data files provided by the La Thuile or the FLUXNET2015 datasets <https://fluxnet.fluxdata.org>. Forest age data were derived from a published dataset [65]. Forest age estimates were aggregated in six classes: 0-10 years (n = 7 sites), 10-20 years (n = 8 sites), 20-50 years (n = 14 sites), 50-100 (n = 27 sites), 100-150 (n = 14 sites), and 150- $\geq$ 300 years (n = 15 sites).

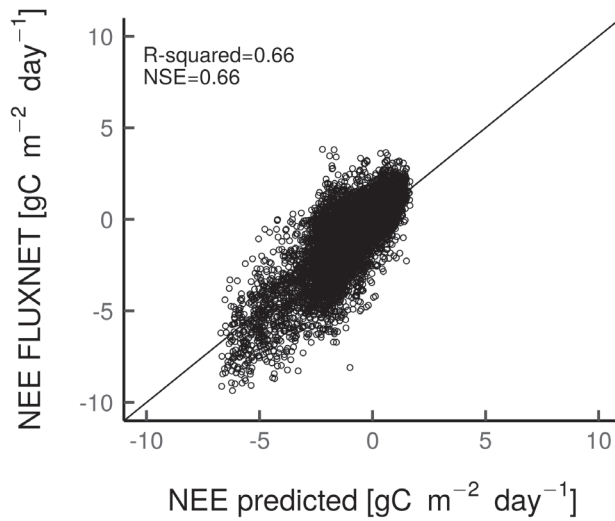
## Results and discussion

### Performance of the Recurrent Neural Networks

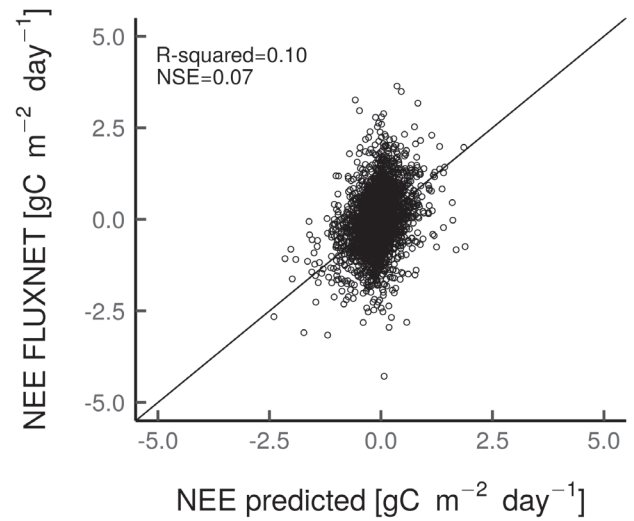
The proposed approach was generally able to capture the seasonal cycle well for *LSTM* set-ups (NSE = 0.66), but had moderate to poor predictive capacity to explain across-site variability (NSE = 0.42), monthly anomalies (NSE = 0.07) or interannual anomalies (NSE = 0.07) (Fig 4). However, the proposed approach achieved comparable predictive capacity than the most recent NEE estimates based on FLUXNET data across scales [62, 63].

Such dynamic statistical modeling approach (i.e. LSTM) was expected to achieve a better performance for predicting anomalies compared to [62] and [63] analysis. In fact, LSTMs are theoretically able to automatically learn informative features [66, 67], and as such could capture interannual and seasonal fluctuations in the remote sensing and climate data related to

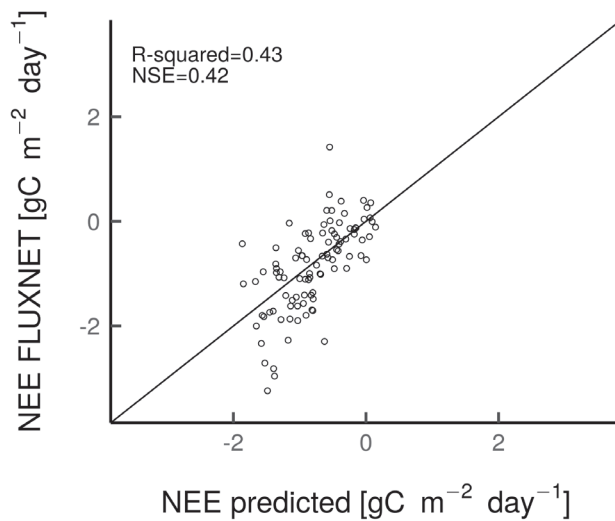
a) seasonal cycle



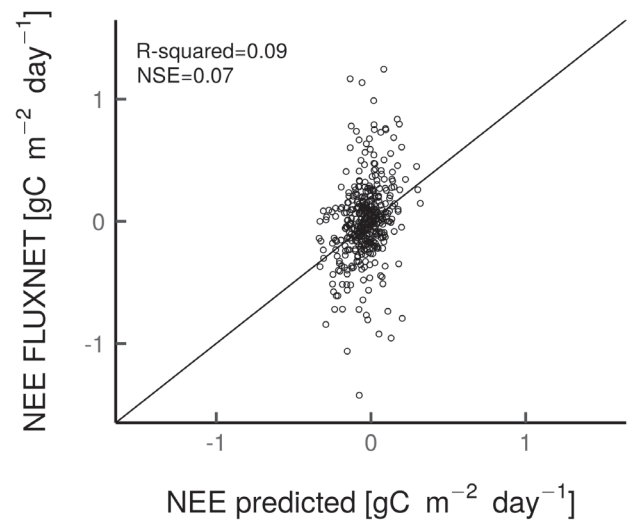
b) seasonal anomalies



c) across-sites



d) interannual anomalies



**Fig 4.** Scatterplots of observed data by eddy-covariance and the LSTM modeled fluxes for the seasonal cycle (Fig 4a), seasonal anomalies (Fig 4b), across-site variability (Fig 4c), and interannual anomalies (Fig 4d). The modeled estimates are derived from the mean ensemble of the 50 model runs.

<https://doi.org/10.1371/journal.pone.0211510.g004>

specific ecosystem impact events (e.g. anthropogenic disturbances, seasonal droughts). However, this appeared not to be the case. We assume that this could be due to:

1. the fact that anomaly signals were relatively small (Fig 4b and 4d) compared to the low signal-to-noise ratio in the remote sensing data because of atmospheric contamination;
2. the non-availability of complete Landsat time series, and necessary gapfilling step;
3. the fact that the training of the statistical models was performed at monthly scale and not at daily scale due to the temporal resolution of the Landsat data. Signatures of extreme events are likely more apparent at daily time scale, therefore one could argue that the temporal

scale used in the presented study is not appropriate to capture well the anomalies in the signals;

4. the limitations associated with the remote sensing signals in providing all necessary information regarding vegetation structure and growth trajectory, while being insensitive to C decomposition dynamics;
5. the fact that few disturbances events were observed during the observational period compared to undisturbed sites or to sites where disturbances occurred a long time ago in which no spectral recovery signals were captured during the training procedure (i.e. only 10% of the observations in recently disturbed forests); and
6. the lack of information related to the spatial context (e.g. landscape patchiness, fractal dimension) that could translate the development stage.

All these factors could suggest that few anomaly signals are captured during the training process of the proposed approach. Furthermore, we cannot ignore the fact that there is a lack of relevant information on the predictors used in this study to predict NEE variability across-scales. For instance, extracting temporal variation in the Landsat data could have been a good proxy for the age effects on NEE among sites, but this was not the case here (as discussed further), likely due to missing information in the input variables for the heterotrophic respiration component of NEE. The mismatch between the observed and predicted NEE at interannual scales could also be related to the fact that the training procedure does not learn site-specific characteristics due to the implemented cross-validation set-up (i.e. entire sites out cross-validation) [40], therefore limiting the capacity of a statistical model to predict NEE interannual anomalies accurately. Furthermore, the cost function was performed on the monthly observations during the training/evaluation procedure, which can potentially limit the capacity of the presented approach for calculating NEE signals realistically at annual scales. In addition, there are few very young sites (< 20 years old) or sites where disturbances occurred during the Landsat record in the training set, which can limit the ability of the proposed approach to have good predictive capacity in young recently disturbed sites. Another source of uncertainty is related to the mismatch between flux measurements and the Landsat time series cutouts around each flux tower. To overcome the latter issue, integration of footprint analysis could help to better describe the origin of the fluxes within the Landsat time series cutouts.

Differences in predictive capacity were apparent for different PFTs and climate levels (Table 3, S2 and S3 Tables). The NSE for different PFTs and climate regions at the seasonal scale ranged from 0.42 (i.e. evergreen forests) to 0.82 (i.e. deciduous forests), and from -0.0006 for the tropical forests to 0.68 for both temperate and boreal forests. The fact that the LSTM showed poor agreement with observations in the tropics can be explained by the very small signal in the input data due to the lack of seasonal variation in terms of reflective, thermal, and moisture properties [32]. In addition, the Landsat data tend to be very sparse in the tropics due to frequent cloud coverage, leading to a high fraction of gap-filled data, thus a potentially poor representation of the seasonal vegetation variation. The properties of the Landsat data might also not be suitable to characterize seasonality in the tropics, therefore other remote sensing products related to leaf development and demography (e.g. Multi-Angle Implementation of Atmospheric Correction data product) [68] could be explored. However, such products were not tested given their relatively short time series, although this shortcoming may be overcome in the future. EC flux data also have their own limitations in the tropics, not only due to sparse spatial coverage but also due to large gaps in data related to frequent rain events and severe issues with the night-time fluxes due to low wind speed and tall canopies. The LSTM was able to well predict NEE across-sites in evergreen forests (NSE = 0.43), while it showed poor

**Table 3. Nash-Sutcliffe modeling efficiency of the LSTM setup per vegetation type and climate region from the ensemble mean  $\pm$ sd estimate of the 50 runs.** Statistics for the anomalies were not calculated in the arid and tropical climate (i.e. NA) because there was no site with at least 3 years of complete data after data quality control. Savanna vegetation type includes both savanna and woody savanna sites.

	Seasonal cycle	Seasonal anomalies	Across-sites	Interannual anomalies
Deciduous forest	0.82 $\pm$ 0.01	0.16 $\pm$ 0.03	0.26 $\pm$ 0.07	0.17 $\pm$ 0.04
Evergreen forest	0.42 $\pm$ 0.02	0.03 $\pm$ 0.02	0.43 $\pm$ 0.06	0.008 $\pm$ 0.04
Mixed forest	0.63 $\pm$ 0.03	0.01 $\pm$ 0.02	0.40 $\pm$ 0.19	0.08 $\pm$ 0.04
Savanna	0.55 $\pm$ 0.02	0.03 $\pm$ 0.02	0.29 $\pm$ 0.48	0.37 $\pm$ 0.07
Arid	0.47 $\pm$ 0.04	NA	0.15 $\pm$ 0.75	NA
Boreal	0.68 $\pm$ 0.01	0.14 $\pm$ 0.03	0.33 $\pm$ 0.07	-0.04 $\pm$ 0.05
Temperate	0.68 $\pm$ 0.01	0.04 $\pm$ 0.02	0.29 $\pm$ 0.07	0.09 $\pm$ 0.03
Tropical	-0.0006 $\pm$ 0.12	NA	-0.15 $\pm$ 0.27	NA

<https://doi.org/10.1371/journal.pone.0211510.t003>

agreement with observations in tropical regions (NSE = -0.15). One could assume that the low number of tropical sites currently available in this dataset (n = 16 and n = 7) at the seasonal scale and across-sites, respectively) might limit an LSTM to predict spatial NEE variabilities in such an ecosystem and also lead to a systematically higher uncertainty across-scales compared to other PFTs (Table 3).

### Comparison of the different model set-ups

A comparison of the model performance was done between the different LSTM networks along with the non-dynamic statistical RF model (Table 4, S4, S5 and S6 Tables). In general, the performance metrics across the model set-ups differed. All model set-ups were capable of well predicting the seasonal cycle, with the LSTM achieving particularly better model fitness and lower errors (NSE = 0.66, RMSE = 1.12, and MAE = 0.81). Similarly, LSTM depicted better agreement between observations and predictions across-sites (NSE = 0.42, RMSE = 0.63, and MAE = 0.48). However, none of the presented model set-ups were able to successfully predict the anomaly signals, with LSTM having rather similar performance and level of errors than the other model set-ups for both seasonal (NSE = 0.07, RMSE = 0.61, and MAE = 0.31) and inter-annual anomalies (NSE = 0.07, RMSE = 0.31, and MAE = 0.22). Still, this results supported the importance of accounting for interannual and seasonal fluctuations of climate and vegetation to estimate net CO<sub>2</sub> fluxes, in particular at the seasonal scale and across-sites. This was evidenced by LSTM<sub>msc</sub>, LSTM<sub>perm</sub>, LSTM<sub>annual</sub> and the RF model set-ups, which depicted lower predictive capacities and higher errors than the original LSTM. However, these comparisons were done for the entire FLUXNET dataset, but the effects of memory were substantially different across-sites (S4 Fig).

**Table 4. Nash-Sutcliffe modeling efficiency of the proposed approach against the other model set-ups from the ensemble mean  $\pm$ sd estimate of the 50 model runs.** LSTM = LSTM model using the full depth of the Landsat time series and climate data; LSTM<sub>perm</sub> = LSTM model but the temporal patterns of both the predictive and the target variables were randomly permuted while instantaneous relationships between predictive and target variables were kept; LSTM<sub>msc</sub> = LSTM model but the Landsat time series for each band were replaced by their mean seasonal cycle, while using the actual values of air temperature (T<sub>air</sub>), precipitation (P), global radiation (Rg), and vapor pressure deficit (VPD); LSTM<sub>annual</sub> = LSTM model but the Landsat time series for each band were replaced by their annual mean, while using the actual values of T<sub>air</sub>, P, Rg, and VPD, RF = Random Forest model using the actual values of the Landsat time series and climate data.

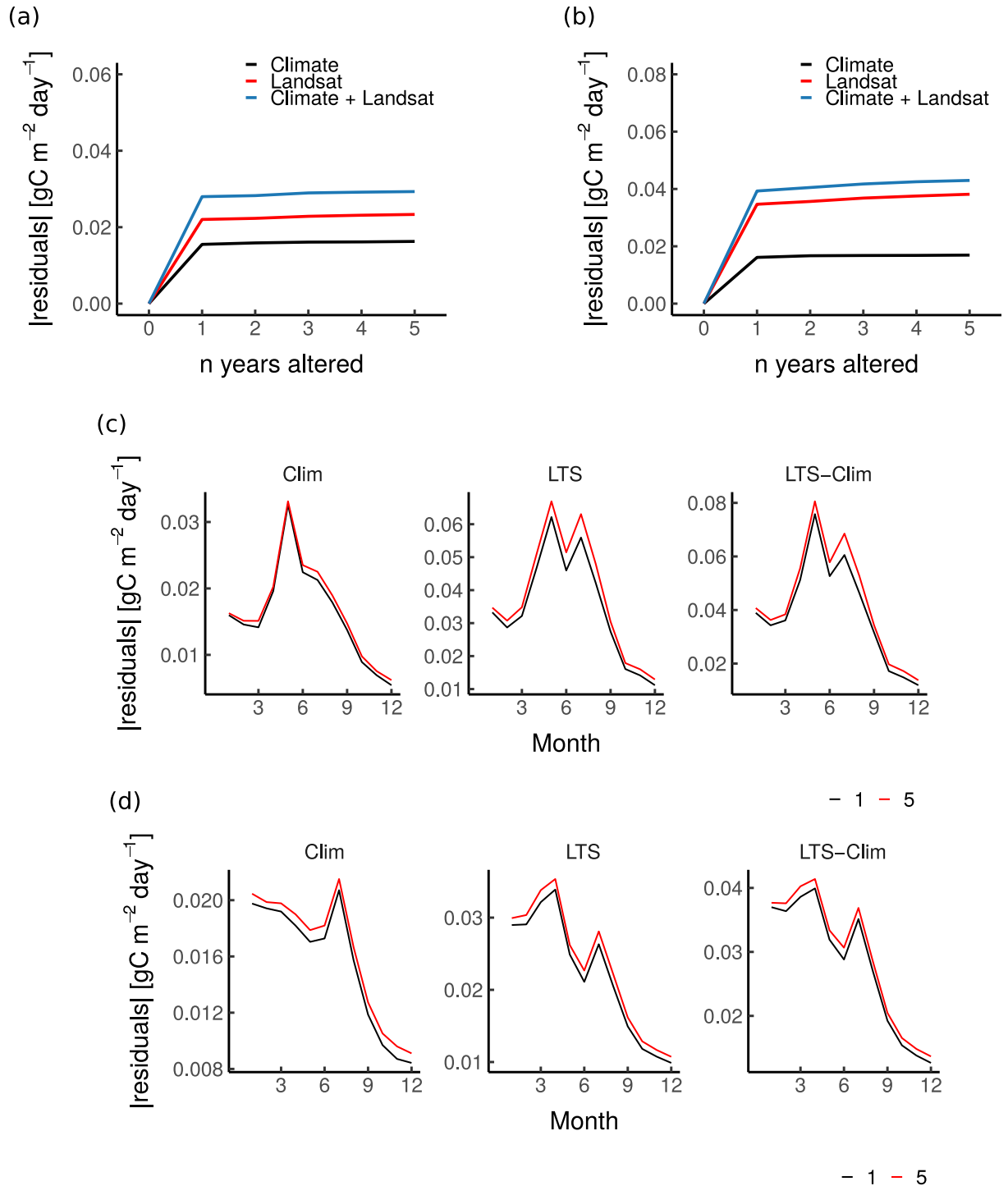
	Seasonal cycle	Seasonal anomalies	Across-sites	Interannual anomalies
LSTM	0.66 $\pm$ 0.01	0.07 $\pm$ 0.01	0.42 $\pm$ 0.05	0.07 $\pm$ 0.03
LSTM <sub>msc</sub>	0.64 $\pm$ 0.009	0.05 $\pm$ 0.008	0.39 $\pm$ 0.04	0.02 $\pm$ 0.01
LSTM <sub>annual</sub>	0.59 $\pm$ 0.02	0.06 $\pm$ 0.03	0.36 $\pm$ 0.05	0.04 $\pm$ 0.05
LSTM <sub>perm</sub>	0.61 $\pm$ 0.01	0.008 $\pm$ 0.02	0.38 $\pm$ 0.05	0.08 $\pm$ 0.02
RF	0.58 $\pm$ 0.00004	-0.30 $\pm$ 0.0006	0.38 $\pm$ 0.0002	-0.04 $\pm$ 0.0007

<https://doi.org/10.1371/journal.pone.0211510.t004>

The fact that the *LSTM* network exploits the history of the predictor variables could explain its overall better results in predicting CO<sub>2</sub> fluxes compared to other model set-ups, despite the differences being marginal. The CO<sub>2</sub> fluxes are not only linearly related to the instantaneous reflectance and meteorological conditions but also associated with the climate and vegetation dynamics several months to years prior [6], which may affect non-observed ecosystem states with direct consequences to C fluxes. To investigate this, an additional simulation experiment was conducted to understand how many years, before predicting a specific year, the proposed approach (i.e. *LSTM*) uses to achieve a better model performance (Fig 5). The *LSTM* model trained before was used, but during the prediction, the actual values for predictors in year  $i-n$  (where  $n$  is a number of years ranging from one to five) were replaced by their MSC when predicting year  $i$ . Hence, the interannual variations and seasonal deviations of year  $i-n$  were not included in the predictions of the *LSTMs* when calculating NEE for year  $i$ . For both deciduous and evergreen forests, there was a consistent increase in the mean absolute residuals from 0 to 1 years of altered forcings, while there were no substantial changes when the number of years since alteration was  $\geq 1$  year (Fig 5a and 5b). It is also interesting to see that altering only climate predictors has less of an effect on the deviations from the NEE estimates, compared to the other two scenarios. For deciduous forests, capturing information from the current and the previous years results in the highest differences in NEE estimates mainly during the growing season (April to September) (Fig 5c). On the other hand, altering the Landsat and climate time series of the previous years seemed to mainly have the highest effects on the predicted NEE from January to August-September. Overall, the magnitudes in the errors are substantially higher for deciduous forests. Note that these findings do not mean that only previous-year climate and vegetation memory effects are important for improving NEE estimates but indicate that their significance in the proposed approach diminishes to further improve its predictive capacity.

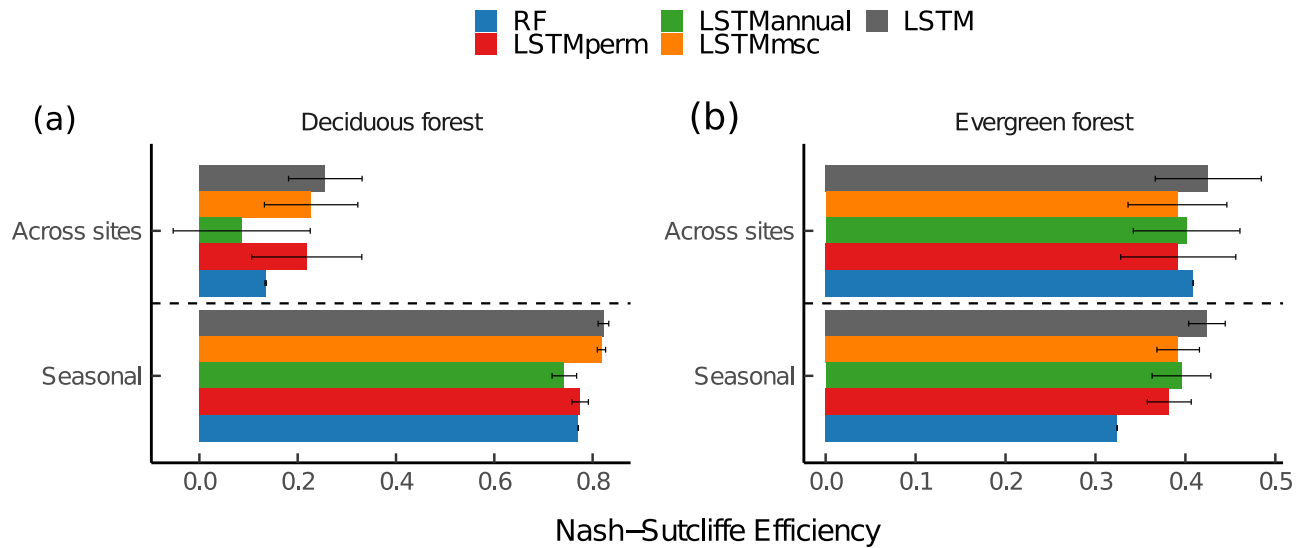
This study confirms that changes in historical climate and vegetation dynamics play a moderate role in shaping the temporal variability of ecosystem CO<sub>2</sub> fluxes, particularly at the seasonal scale (i.e. around 8% difference in model efficiency between *LSTM* and *LSTM<sub>perm</sub>*) and across-sites (i.e. 10% difference in model efficiency between *LSTM* and *LSTM<sub>perm</sub>*) (Table 4 and Fig 5). However, these findings differ markedly between forest types (Figs 5 and 6). For instance, NEE estimates calculated by *LSTM* and *LSTM<sub>msc</sub>* for deciduous forests are rather similar at the seasonal scale, suggesting that the interannual variation information carried by the remote sensing data does not help to achieve better performance capacity in predicting NEE at seasonal scale in such an ecosystem, while the interannual variation in climate is still considered. On the other hand, the highest modeling efficiency was achieved for evergreen forests using the *LSTM* model, suggesting that both interannual and seasonal fluctuations in vegetation are important drivers of NEE variabilities at the seasonal scale.

One outstanding result of this analysis is the importance of memory effect at the seasonal scale (Table 4 and Fig 6). Such finding can be better explored using the NEE mean seasonal variation residuals for deciduous and evergreen forests (Fig 7). For deciduous and evergreen forests, it is important to extract realistic temporal vegetation and climate information when predicting NEE as the *LSTM<sub>perm</sub>* model depicts the highest overall error in the residual seasonal patterns compared to the other model set-ups. For deciduous forests, both the onset and the peak of the growing season were better captured by *LSTM* and *LSTM<sub>msc</sub>* models (Fig 7a). This could suggest that the climatic conditions of the previous years (e.g. water limitations [33, 35], increased precipitation [34], or minimum air temperature during spring of the previous year [36]) not only control NEE seasonal patterns in deciduous forests but also mean seasonal vegetation fluctuations. It is therefore probable that seasonal leaf physiology due to leaf aging also drives the residual seasonal patterns [69]. The *LSTM<sub>annual</sub>* model set-up revealed that



**Fig 5. Effects on predicting monthly NEE by altering  $n$  year in the predictors for deciduous and evergreen forests.** Average of the absolute residuals calculated between predicted monthly NEE with 0 year altered in the predictors against predicted monthly NEE with year  $i-n$  altered in the predictors for deciduous and evergreen forests (Fig 5a and b, respectively). The absolute residuals for the mean seasonal cycle were also reported (Fig 5c and d for deciduous and evergreen forests, respectively). “1 year” means that only the last year was altered, “2 years” means that the last two years were altered, and so on. Months for the sites located in the Southern hemisphere have been adjusted to match the seasonal cycle of the sites in the Northern hemisphere.

<https://doi.org/10.1371/journal.pone.0211510.g005>

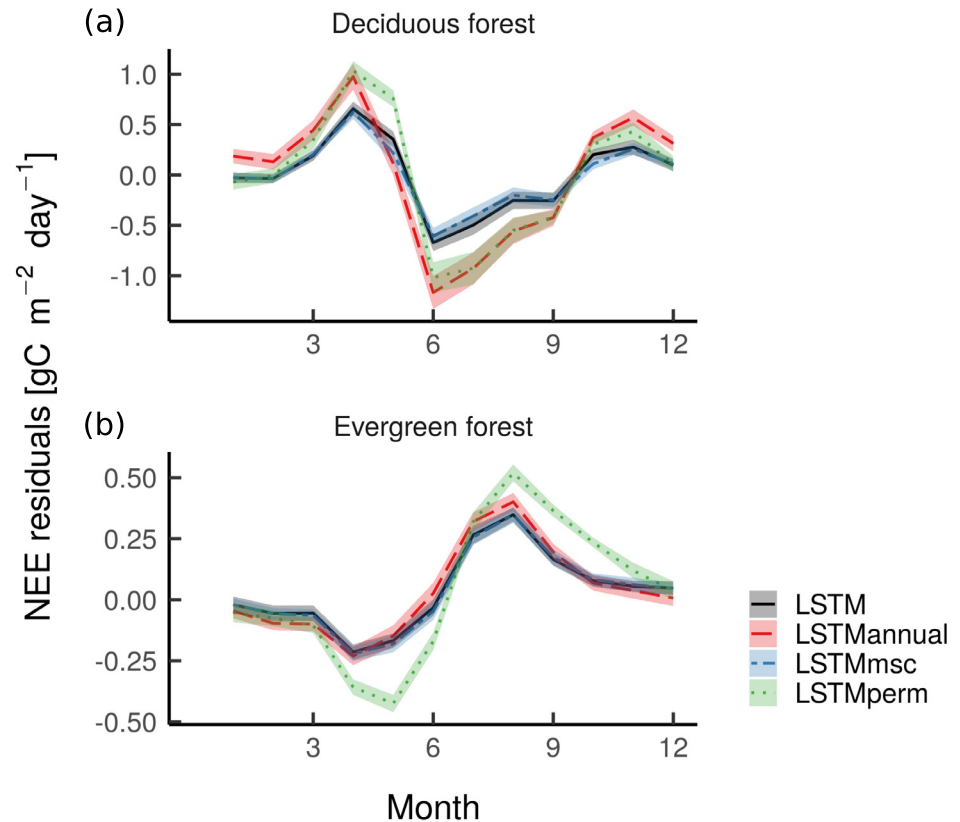


**Fig 6.** Nash-Sutcliffe modeling efficiency comparison between the proposed LSTM-based models and the other model set-ups for (a) deciduous and (b) evergreen forests. Nash-Sutcliffe modeling efficiency values have been calculated based on the mean ensemble  $\pm$ sd of the 50 model runs. *LSTM* = LSTM model using the full depth of the Landsat time series and climate data; *LSTM<sub>perm</sub>* = LSTM model but the temporal patterns of both the predictive and the target variables were randomly permuted while instantaneous relationships between predictive and target variables were kept; *LSTM<sub>msc</sub>* = LSTM model but the Landsat time series for each band were replaced by their mean seasonal cycle, while using the actual values of air temperature ( $T_{air}$ ), precipitation (P), global radiation (Rg), and vapor pressure deficit (VPD); *LSTM<sub>annual</sub>* = LSTM model but the Landsat time series for each band were replaced by their annual mean, while using the actual values of  $T_{air}$ , P, Rg, and VPD, RF = Random Forest model using the actual values of the Landsat time series and climate data.

<https://doi.org/10.1371/journal.pone.0211510.g006>

capturing interannual variations in vegetation activities does not help in representing NEE estimates at the seasonal scale. However, all the model set-ups showed rather similar errors when representing the senescence phase in deciduous forests, suggesting that the processes that control these dynamics are not expressed in any of the observational datasets used here. Interestingly, *LSTM*, *LSTM<sub>msc</sub>*, and *LSTM<sub>annual</sub>* model set-ups depicted relatively similar errors over the course of the growing season for evergreen forests (Fig 7b). This means that both the current climate conditions and the ones of the previous months or years control NEE seasonal cycle in such an ecosystem. These findings confirmed the existence of different ecosystem type-specific memory or lagged effects [33].

The *LSTM* model set-up outperformed the other models (i.e. *LSTM<sub>perm</sub>*, *LSTM<sub>msc</sub>*, *LSTM<sub>annual</sub>*, and *RF*) across sites, suggesting that it is able to better capture the complexity of the relation between past dynamics and current functions of the forests across-space. One hypothesis could be that net CO<sub>2</sub> fluxes in recently disturbed forests are better predicted with a method that captures disturbance regimes. However, this hypothesis could not be confirmed since: (1) the *LSTM* model set-up did not outperform the other model set-ups for young forests (i.e. 0-20 years old) and for recently disturbed forests (Fig 8); and (2) training an LSTM adding forest age as predictor or training it only for young forests (forest age < 40 years) did not correct for the bias in young forests (Fig 8). However, it is not possible to be conclusive on the ability of the LSTMs to predict young sites since: (1) there was only a small sample of young forests and recently disturbed sites in this dataset; and that methodologically (2) no *in-situ* proxies for productivity were used in the analysis (e.g. related gross primary productivity); and (3) the LSTMs were trained with monthly observation. Therefore, it is very likely that the better performance of the *LSTM* model set-up compared to the other model set-ups at the seasonal cycle could explain its overall better capacity in explaining NEE spatial variation (e.g. spring NEE accounts for most of the annual NEE in the temperate deciduous forests [36]).



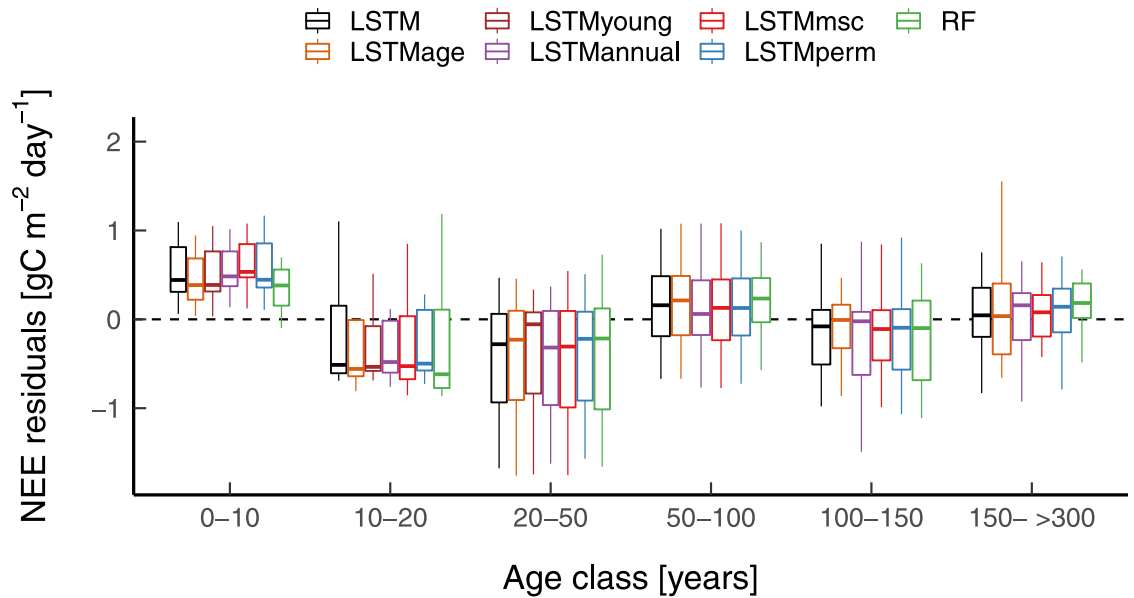
**Fig 7.** Mean seasonal variation of NEE residuals for  $LSTM$ ,  $LSTM_{perm}$ ,  $LSTM_{msc}$  and  $LSTM_{annual}$  models for (a) deciduous and (b) evergreen forests.  $NEE\ residuals = [NEE\ observed_{i,j} - \text{mean}(NEE\ observed_i)] - [NEE\ predicted_{i,j} - \text{mean}(NEE\ predicted_i)]$ , where  $i$  is a unique Fluxnet site and  $j$  is a monthly observation. Residual estimates have been calculated based on the mean ensemble  $\pm$ sd of the 50 model runs.  $LSTM$  = LSTM model using the full depth of the Landsat time series and climate data;  $LSTM_{perm}$  = LSTM model but the temporal patterns of both the predictive and the target variables were randomly permuted while instantaneous relationships between predictive and target variables were kept;  $LSTM_{msc}$  = LSTM model but the Landsat time series for each band were replaced by their mean seasonal cycle, while using the actual values of air temperature ( $T_{air}$ ), precipitation (P), global radiation (Rg), and vapor pressure deficit (VPD);  $LSTM_{annual}$  = LSTM model but the Landsat time series for each band were replaced by their annual mean, while using the actual values of  $T_{air}$ , P, Rg, and VPD, RF = Random Forest model using the actual values of the Landsat time series and climate data. Months for the sites located in the Southern hemisphere have been adjusted to match the seasonal cycle of the sites in the Northern hemisphere.

<https://doi.org/10.1371/journal.pone.0211510.g007>

## Conclusion

In this study, we present a novel framework for assessing the potential of the memory effects of climate and vegetation on forests' NEE using the Landsat satellite imagery and *in-situ* eddy covariance observations. The results presented here for the whole FLUXNET dataset reveal a variable memory effect on NEE across-scales, but that is mainly apparent at the seasonal scale and across-sites. We also find that the effects of memory vary between FLUXNET sites suggesting site-specific memory effects. Although instantaneous observations of the contemporaneous vegetation states may already carry information from the past, current analysis suggests that extracting antecedent observations of vegetation and climate are beneficial for estimating NEE more realistically (the difference between  $LSTM$  and  $LSTM_{perm}$ , as well as between  $LSTM$  and  $RF$ ). Such effects can emerge from the information contained in the course of the seasonal cycle or from the effects of interannual variation on NEE. However, the close agreement between  $LSTM$  and  $LSTM_{msc}$  suggests that either the effect is smeared out by the impact of





**Fig 8. Model residuals per age class for LSTM, LSTM<sub>perm</sub>, LSTM<sub>msc</sub>, LSTM<sub>annual</sub> and RF models based on site-average NEE.** LSTM = LSTM model using the full depth of the Landsat time series and climate data; LSTM<sub>perm</sub> = LSTM model but the temporal patterns of both the predictive and the target variables were randomly permuted while instantaneous relationships between predictive and target variables were kept; LSTM<sub>msc</sub> = LSTM model but the Landsat time series for each band were replaced by their mean seasonal cycle, while using the actual values of air temperature ( $T_{air}$ ), precipitation (P), global radiation (Rg), and vapor pressure deficit (VPD); LSTM<sub>annual</sub> = LSTM model but the Landsat time series for each band were replaced by their annual mean, while using the actual values of  $T_{air}$ , P, Rg, and VPD, RF = Random Forest model using the actual values of the Landsat time series and climate data.; LSTM<sub>age</sub> = LSTM + forest age as a predictive variable; LSTM<sub>young</sub> = LSTM only trained with forests younger than 40 years.

<https://doi.org/10.1371/journal.pone.0211510.g008>

instantaneous climate on NEE or the interannual variation’s memory effect in NEE is implicitly captured by this approach. The results are contingent on the length of observations and few recently disturbed forests but do emphasize the possibility of dynamic statistical methods that include memory effects to better estimate the contribution of forest ecosystems in the global terrestrial C cycle, hence for further improving statistically-based prediction methods.

### Supporting information

**S1 Table. List of sites used in this study.** DBF = Deciduous broadleaf forest, DNF = deciduous needleleaf forest, EBF = evergreen broadleaf forest, ENF = evergreen needle-leaf forest, MF = mixed forest, WSA = woody savanna, and SAV = savanna. (PDF)

**S2 Table. RMSE of the LSTM setup per PFT and climate region from the ensemble mean mean  $\pm$ sd estimate of the 50 runs.** Statistics for the anomalies were not calculated in the arid and tropical climate (i.e. NA) because there was no site with at least 2 years of complete data after data quality control. (PDF)

**S3 Table. MAE of the LSTM setup per PFT and climate region from the ensemble mean mean  $\pm$ sd estimate of the 50 runs.** Statistics for the anomalies were not calculated in the arid and tropical climate (i.e. NA) because there was no site with at least 2 years of complete data after data quality control. (PDF)

**S4 Table. Coefficient of determination of the proposed approach against the other model set-ups from the ensemble mean mean  $\pm$ sd estimate of the 50 runs.** *LSTM* = LSTM model using the full depth of the Landsat time series and climate data; *LSTM<sub>perm</sub>* = LSTM model but the temporal patterns of both the predictive and the target variables were randomly permuted while instantaneous relationships between predictive and target variables were kept; *LSTM<sub>msc</sub>* = LSTM model but the Landsat time series for each band were replaced by their mean seasonal cycle, while using the actual values of air temperature ( $T_{air}$ ), precipitation (P), global radiation (Rg), and vapor pressure deficit (VPD); *LSTM<sub>annual</sub>* = LSTM model but the Landsat time series for each band were replaced by their annual mean, while using the actual values of  $T_{air}$ , P, Rg, and VPD, RF = Random Forest model using the actual values of the Landsat time series and climate data.

(PDF)

**S5 Table. RMSE of the proposed approach against the other model set-ups from the ensemble mean mean  $\pm$ sd estimate of the 50 runs.** *LSTM* = LSTM model using the full depth of the Landsat time series and climate data; *LSTM<sub>perm</sub>* = LSTM model but the temporal patterns of both the predictive and the target variables were randomly permuted while instantaneous relationships between predictive and target variables were kept; *LSTM<sub>msc</sub>* = LSTM model but the Landsat time series for each band were replaced by their mean seasonal cycle, while using the actual values of air temperature ( $T_{air}$ ), precipitation (P), global radiation (Rg), and vapor pressure deficit (VPD); *LSTM<sub>annual</sub>* = LSTM model but the Landsat time series for each band were replaced by their annual mean, while using the actual values of  $T_{air}$ , P, Rg, and VPD, RF = Random Forest model using the actual values of the Landsat time series and climate data.

(PDF)

**S6 Table. MAE of the proposed approach against the other model set-ups from the ensemble mean mean  $\pm$ sd estimate of the 50 runs.** *LSTM* = LSTM model using the full depth of the Landsat time series and climate data; *LSTM<sub>perm</sub>* = LSTM model but the temporal patterns of both the predictive and the target variables were randomly permuted while instantaneous relationships between predictive and target variables were kept; *LSTM<sub>msc</sub>* = LSTM model but the Landsat time series for each band were replaced by their mean seasonal cycle, while using the actual values of air temperature ( $T_{air}$ ), precipitation (P), global radiation (Rg), and vapor pressure deficit (VPD); *LSTM<sub>annual</sub>* = LSTM model but the Landsat time series for each band were replaced by their annual mean, while using the actual values of  $T_{air}$ , P, Rg, and VPD, RF = Random Forest model using the actual values of the Landsat time series and climate data.

(PDF)

**S1 Fig. Performance of the gap-filling procedure of each Landsat band using a Random Forest model and the MODIS bands as predictive variables.** The model was trained on 70% of the data and evaluated on 30% of the left out data. nir = near-infrared, swir1 = shortwave infrared 1, swir2 = shortwave infrared 2, and tir = thermal infrared.

(PDF)

**S2 Fig. Performance of the gap-filling procedure of each Landsat band using a Random Forest model and climate variables (i.e.  $T_{air}$ , Precip, Rg, VPD, rpot), PFT, month of the year, and latitude as predictive variables.** The model was trained on 70% of the data and evaluated on 30% of the left out data. nir = near-infrared, swir1 = shortwave infrared 1, swir2 = shortwave infrared 2, and tir = thermal infrared.

(PDF)

**S3 Fig. Performance of the gap-filling procedure for the different climate variables.**

Assessment of the gap-filling procedure was done for  $T_{air}$ , Precip,  $R_g$ , and VPD. For  $T_{air}$ ,  $R_g$ , and VPD, the Nash-Sutcliffe efficiency (NSE) is reported, while the root mean squared error (RMSE) is reported for Precip. (PDF)

**S4 Fig. Scatterplots of the coefficient of determination of the proposed approach against the other model set-ups at site level.** The coefficient of determination was computed using monthly observed and predicted NEE estimates for each site. Each point represents one site and only the sites with at least one complete year of good quality data ( $n_{site} = 81$ ) are shown. (PDF)

## Acknowledgments

This work used eddy covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux (U.S. Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program (DE-FG02-04ER63917 and DE-FG02-04ER63911)), AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCan), GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux, TCOS-Siberia, USCCC. We acknowledge the financial support to the eddy covariance data harmonization provided by CarboEuropeIP, FAO-GTOS-TCO, iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia, Université Laval and Environment Canada and US Department of Energy and the database development and technical support from Berkeley Water Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California—Berkeley, University of Virginia. We would like to thank Connor Crank for proofreading the manuscript.

## Author Contributions

**Conceptualization:** Simon Besnard, Nuno Carvalhais, Martin Jung, Markus Reichstein.

**Data curation:** Simon Besnard, M. Altaf Arain, Andrew Black, Nina Buchmann, Jiquan Chen, Loïc P. Dutrieux, Yoshiko Kosugi, Alexander Knohl, Beverly E. Law, Eugénie Paul-Limoges, Annalea Lohila, Lutz Merbold, Olivier Roupsard, Riccardo Valentini, Sebastian Wolf, Xudong Zhang.

**Formal analysis:** Simon Besnard, Nuno Carvalhais.

**Funding acquisition:** Martin Herold.

**Methodology:** Simon Besnard, Nuno Carvalhais, Fabian Gans, Markus Reichstein.

**Software:** Fabian Gans.

**Supervision:** Nuno Carvalhais, Martin Herold, Markus Reichstein.

**Writing – original draft:** Simon Besnard.

**Writing – review & editing:** Simon Besnard, Nuno Carvalhais, M. Altaf Arain, Andrew Black, Benjamin Brede, Nina Buchmann, Jiquan Chen, Jan G. P. W. Clevers, Loïc P. Dutrieux, Martin Jung, Alexander Knohl, Beverly E. Law, Eugénie Paul-Limoges, Annalea Lohila, Lutz Merbold, Olivier Roupsard, Sebastian Wolf, Markus Reichstein.

## References

1. Pan Y, Birdsey RA, Fang J, Houghton R, Kauppi PE, Kurz WA, et al. A large and persistent carbon sink in the world's forests. *Science*. 2011; 333(6045):988–993. <https://doi.org/10.1126/science.1201609> PMID: 21764754
2. Beer C, Reichstein M, Tomelleri E, Ciais P, Jung M, Carvalhais N, et al. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science*. 2010; 329(5993):834–838. <https://doi.org/10.1126/science.1184984> PMID: 20603496
3. Gower ST. Patterns and mechanisms of the forest carbon cycle 1. *Annual Review of Environment and Resources*. 2003; 28(1):169–204. <https://doi.org/10.1146/annurev.energy.28.050302.105515>
4. Le Quéré C, Andrew RM, Canadell JG, Sitch S, Korsbakken JI, Peters GP, et al. Global carbon budget 2016. *Earth System Science Data (Online)*. 2016; 8(2).
5. Zhu Z, Piao S, Myneni RB, Huang M, Zeng Z, Canadell JG, et al. Greening of the Earth and its drivers. *Nature Climate Change*. 2016; 6(8):791–795. <https://doi.org/10.1038/nclimate3004>
6. Reichstein M, Bahn M, Ciais P, Frank D, Mahecha MD, Seneviratne SI, et al. Climate extremes and the carbon cycle. *Nature*. 2013; 500(7462):287. <https://doi.org/10.1038/nature12350> PMID: 23955228
7. Williams CA, Collatz GJ, Masek J, Goward SN. Carbon consequences of forest disturbance and recovery across the conterminous United States. *Global Biogeochemical Cycles*. 2012; 26(1). <https://doi.org/10.1029/2010GB003947>
8. Liu S, Bond-Lamberty B, Hicke JA, Vargas R, Zhao S, Chen J, et al. Simulating the impacts of disturbances on forest carbon cycling in North America: Processes, data, models, and challenges. *Journal of Geophysical Research: Biogeosciences*. 2011; 116(G4).
9. Woodbury PB, Smith JE, Heath LS. Carbon sequestration in the US forest sector from 1990 to 2010. *Forest Ecology and Management*. 2007; 241(1):14–27. <https://doi.org/10.1016/j.foreco.2006.12.008>
10. Schimel D. Carbon cycle conundrums. *Proceedings of the National Academy of Sciences*. 2007; 104(47):18353–18354. <https://doi.org/10.1073/pnas.0709331104>
11. Birdsey R, Pregitzer K, Lucier A. Forest carbon management in the United States. *Journal of Environmental Quality*. 2006; 35(4):1461–1469. <https://doi.org/10.2134/jeq2005.0162> PMID: 16825466
12. Johnson DW, Curtis PS. Effects of forest management on soil C and N storage: meta analysis. *Forest Ecology and Management*. 2001; 140(2):227–238. [https://doi.org/10.1016/S0378-1127\(00\)00282-6](https://doi.org/10.1016/S0378-1127(00)00282-6)
13. Zscheischler J, Mahecha MD, Avitabile V, Calle L, Carvalhais N, Ciais P, et al. An empirical spatiotemporal description of the global surface-atmosphere carbon fluxes: opportunities and data limitations. *Biogeosciences*. 2017; 14(15):3685–3703.
14. Amiro B, Barr A, Barr J, Black TA, Bracho R, Brown M, et al. Ecosystem carbon dioxide fluxes after disturbance in forests of North America. *Journal of Geophysical Research: Biogeosciences*. 2010; 115(G4).
15. Carvalhais N, Reichstein M, Ciais P, Collatz GJ, Mahecha MD, Montagnani L, et al. Identification of vegetation and soil carbon pools out of equilibrium in a process model via eddy covariance and biometric constraints. *Global Change Biology*. 2010; 16(10):2813–2829. <https://doi.org/10.1111/j.1365-2486.2010.02173.x>
16. Thornton P, Law B, Gholz HL, Clark KL, Falge E, Ellsworth D, et al. Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests. *Agricultural and Forest Meteorology*. 2002; 113(1):185–222. [https://doi.org/10.1016/S0168-1923\(02\)00108-9](https://doi.org/10.1016/S0168-1923(02)00108-9)
17. Franklin JF, Spies TA, Van Pelt R, Carey AB, Thornburgh DA, Berg DR, et al. Disturbances and structural development of natural forest ecosystems with silvicultural implications, using Douglas-fir forests as an example. *Forest Ecology and Management*. 2002; 155(1-3):399–423. [https://doi.org/10.1016/S0378-1127\(01\)00575-8](https://doi.org/10.1016/S0378-1127(01)00575-8)
18. Odum EP. The strategy of ecosystem development. *Science*. 1969; 164(3877):262–270. <https://doi.org/10.1126/science.164.3877.262> PMID: 5776636
19. Ciais P, Dolman A, Bombelli A, Duren R, Peregon A, Rayner P, et al. Current systematic carbon-cycle observations and the need for implementing a policy-relevant carbon observing system. *Biogeosciences*. 2014; 11:3547–3602. <https://doi.org/10.5194/bg-11-3547-2014>
20. Moore DJ, Trahan NA, Wilkes P, Quaife T, Stephens BB, Elder K, et al. Persistent reduced ecosystem respiration after insect disturbance in high elevation forests. *Ecology Letters*. 2013; 16(6):731–737. <https://doi.org/10.1111/ele.12097> PMID: 23496289
21. Bowman DM, Balch JK, Artaxo P, Bond WJ, Carlson JM, Cochrane MA, et al. Fire in the Earth system. *Science*. 2009; 324(5926):481–484. <https://doi.org/10.1126/science.1163886> PMID: 19390038

22. Meigs GW, Donato DC, Campbell JL, Martin JG, Law BE. Forest fire impacts on carbon uptake, storage, and emission: the role of burn severity in the Eastern Cascades, Oregon. *Ecosystems*. 2009; 12(8):1246–1267. <https://doi.org/10.1007/s10021-009-9285-x>
23. Gough CM, Vogel CS, Harrold KH, George K, Curtis PS. The legacy of harvest and fire on ecosystem carbon storage in a north temperate forest. *Global Change Biology*. 2007; 13(9):1935–1949. <https://doi.org/10.1111/j.1365-2486.2007.01406.x>
24. Chazdon RL, Broadbent EN, Rozendaal DM, Bongers F, Zambrano AMA, Aide TM, et al. Carbon sequestration potential of second-growth forest regeneration in the Latin American tropics. *Science Advances*. 2016; 2(5):e1501639. <https://doi.org/10.1126/sciadv.1501639> PMID: 27386528
25. Anderson-Teixeira KJ, Miller AD, Mohan JE, Hudiburg TW, Duval BD, DeLucia EH. Altered dynamics of forest recovery under a changing climate. *Global Change Biology*. 2013; 19(7):2001–2021. <https://doi.org/10.1111/gcb.12194> PMID: 23529980
26. Donato DC, Harvey BJ, Romme WH, Simard M, Turner MG. Bark beetle effects on fuel profiles across a range of stand structures in Douglas-fir forests of Greater Yellowstone. *Ecological Applications*. 2013; 23(1):3–20. <https://doi.org/10.1890/12-0772.1> PMID: 23495632
27. Gough CM, Curtis PS, Hardiman BS, Scheuermann CM, Bond-Lamberty B. Disturbance, complexity, and succession of net ecosystem production in North America's temperate deciduous forests. *Ecosphere*. 2016; 7(6). <https://doi.org/10.1002/ecs2.1375>
28. Seidl R, Schelhaas MJ, Lexer MJ. Unraveling the drivers of intensifying forest disturbance regimes in Europe. *Global Change Biology*. 2011; 17(9):2842–2852. <https://doi.org/10.1111/j.1365-2486.2011.02452.x>
29. Turner MG. Disturbance and landscape dynamics in a changing world. *Ecology*. 2010; 91(10):2833–2849. <https://doi.org/10.1890/10-0097.1> PMID: 21058545
30. Seidl R, Rammer W, Spies TA. Disturbance legacies increase the resilience of forest ecosystem structure, composition, and functioning. *Ecological Applications*. 2014; 24(8):2063–2077. <https://doi.org/10.1890/14-0255.1> PMID: 27053913
31. Monger C, Sala OE, Duniway MC, Goldfus H, Meir IA, Poch RM, et al. Legacy effects in linked ecological–soil–geomorphic systems of drylands. *Frontiers in Ecology and the Environment*. 2015; 13(1):13–19. <https://doi.org/10.1890/140269>
32. Frank D, Reichstein M, Bahn M, Thonicke K, Frank D, Mahecha MD, et al. Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts. *Global Change Biology*. 2015; 21(8):2861–2880. <https://doi.org/10.1111/gcb.12916> PMID: 25752680
33. Aubinet M, Hurdebise Q, Chopin H, Debacq A, De Ligne A, Heinesch B, et al. Inter-annual variability of Net Ecosystem Productivity for a temperate mixed forest: A predominance of carry-over effects? *Agricultural and Forest Meteorology*. 2018; 262:340–353. <https://doi.org/10.1016/j.agrformet.2018.07.024>
34. Shen W, Jenerette G, Hui D, Scott R. Precipitation legacy effects on dryland ecosystem carbon fluxes: direction, magnitude and biogeochemical carryovers. *Biogeosciences*. 2016; 13(2):425–439. <https://doi.org/10.5194/bg-13-425-2016>
35. Desai AR. Influence and predictive capacity of climate anomalies on daily to decadal extremes in canopy photosynthesis. *Photosynthesis Research*. 2014; 119(1):31–47. <https://doi.org/10.1007/s11120-013-9925-z> PMID: 24078353
36. Zielis S, Etzold S, Zweifel R, Eugster W, Haeni M, Buchmann N. NEP of a Swiss subalpine forest is significantly driven not only by current but also by previous year's weather. *Biogeosciences*. 2014; 11(6):1627–1635. <https://doi.org/10.5194/bg-11-1627-2014>
37. Zhang T, Xu M, Xi Y, Zhu J, Tian L, Zhang X, et al. Lagged climatic effects on carbon fluxes over three grassland ecosystems in China. *Journal of Plant Ecology*. 2014; 8(3):291–302. <https://doi.org/10.1093/jpe/rtu026>
38. van der Molen MK, Dolman AJ, Ciais P, Eglin T, Gobron N, Law BE, et al. Drought and ecosystem carbon cycling. *Agricultural and Forest Meteorology*. 2011; 151(7):765–773. <https://doi.org/10.1016/j.agrformet.2011.01.018>
39. Vicca S, Bahn M, Estiarte M, Van Loon E, Vargas R, Alberti G, et al. Can current moisture responses predict soil CO<sub>2</sub> efflux under altered precipitation regimes? A synthesis of manipulation experiments. *Biogeosciences*. 2014; 11(12):3307–3308. <https://doi.org/10.5194/bg-11-3307-2014>
40. Bodesheim P, Jung M, Gans F, Mahecha MD, Reichstein M. Upscaled diurnal cycles of land-atmosphere fluxes: a new global half-hourly data product. *Earth System Science Data*. 2018; 10:1327–1365. <https://doi.org/10.5194/essd-10-1327-2018>
41. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473. 2014.

42. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: *Advances in Neural Information Processing Systems*; 2014. p. 3104–3112.
43. Hinton G, Deng L, Yu D, Dahl GE, Mohamed Ar, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*. 2012; 29(6):82–97. <https://doi.org/10.1109/MSP.2012.2205597>
44. Reichstein M, Besnard S, Carvalhais N, Gans F, Jung M, Kraft B, et al. Modelling Landsurface Time-Series with Recurrent Neural Nets. *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE; 2018; 7640–7643.
45. Baldocchi D. 'Breathing' of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany*. 2008; 56(1):1–26. <https://doi.org/10.1071/BT07151>
46. Baldocchi D, Falge E, Gu L, Olson R, Hollinger D, Running S, et al. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*. 2001; 82(11):2415–2434. [https://doi.org/10.1175/1520-0477\(2001\)082%3C2415:FANTTS%3E2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082%3C2415:FANTTS%3E2.3.CO;2)
47. Papale D, Reichstein M, Aubinet M, Canfora E, Bernhofer C, Kutsch W, et al. Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences*. 2006; 3(4):571–583. <https://doi.org/10.5194/bg-3-571-2006>
48. Reichstein M, Falge E, Baldocchi D, Papale D, Aubinet M, Berbigier P, et al. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology*. 2005; 11(9):1424–1439. <https://doi.org/10.1111/j.1365-2486.2005.001002.x>
49. Schmidt G, Jenkerson C, Masek J, Vermote E, Gao F. Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description. *US Geological Survey*; 2013.
50. Zhu Z, Wang S, Woodcock CE. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment*. 2015; 159:269–277. <https://doi.org/10.1016/j.rse.2014.12.014>
51. Zhu Z, Woodcock CE. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*. 2012; 118:83–94. <https://doi.org/10.1016/j.rse.2011.10.028>
52. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. 2017; 202:18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
53. Kuhn M, Wing J, Weston S, Williams A, Keefer C, et al. caret: Classification and regression training. R package version 5.15–044; 2012.
54. Breiman L. Random forests. *Machine Learning*. 2001; 45(1):5–32.
55. Kursa MB, Rudnicki WR, et al. Feature selection with the Boruta package. *Journal of Statistical Software*. 2010; 36(11):1–13.
56. Dee DP, Uppala SM, Simmons A, Berrisford P, Poli P, Kobayashi S, et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*. 2011; 137(656):553–597. <https://doi.org/10.1002/qj.828>
57. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*. 1997; 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735> PMID: 9377276
58. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986; 323(6088):533. <https://doi.org/10.1038/323533a0>
59. Kinga D, Adam JB. A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*. vol. 5; 2015.
60. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*. 2014; 15(1):1929–1958.
61. Rußwurm M, Körner M. Temporal Vegetation Modelling using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-Spectral Satellite Images. *Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2017; 1496–1504.
62. Tramontana G, Jung M, Schwalm CR, Ichii K, Camps-Valls G, Ráduly B, et al. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*. 2016; 13(14):4291–4313. <https://doi.org/10.5194/bg-13-4291-2016>
63. Jung M, Reichstein M, Margolis HA, Cescatti A, Richardson AD, Arain MA, et al. 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: Biogeosciences*. 2011; 116(G3).
64. Omlin M, Reichert P. A comparison of techniques for the estimation of model prediction uncertainty. *Ecological Modelling*. 1999; 115(1):45–59. [https://doi.org/10.1016/S0304-3800\(98\)00174-4](https://doi.org/10.1016/S0304-3800(98)00174-4)

65. Besnard S, Carvalhais N, Arain A, Black A, de Bruin S, Buchmann N, et al. Quantifying the effect of forest age in annual net forest carbon balance. *Environmental Research Letters*. 2018; 13(12):124018. <https://doi.org/10.1088/1748-9326/aaeaeab>
66. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015; 521(7553):436. <https://doi.org/10.1038/nature14539> PMID: 26017442
67. Schmidhuber J. Deep learning in neural networks: An overview. *Neural Networks*. 2015; 61:85–117. <https://doi.org/10.1016/j.neunet.2014.09.003> PMID: 25462637
68. Wu J, Albert LP, Lopes AP, Restrepo-Coupe N, Hayek M, Wiedemann KT, et al. Leaf development and demography explain photosynthetic seasonality in Amazon evergreen forests. *Science*. 2016; 351(6276):972–976. <https://doi.org/10.1126/science.aad5068> PMID: 26917771
69. Rodríguez-Calcerrada J, Limousin JM, Martin-StPaul NK, Jaeger C, Rambal S. Gas exchange and leaf aging in an evergreen oak: causes and consequences for leaf carbon balance and canopy respiration. *Tree Physiology*. 2012; 32(4):464–477. <https://doi.org/10.1093/treephys/tps020> PMID: 22491489