Machine Learning for Inferring CO_2 Fluxes: The New Metaphysics of Neural Nets

P. Nguyen[1], M. Halem[1]

[1]University of Maryland, Baltimore County, MD USA.

Corresponding authors: Phuong Nguyen phuong3@umbc.edu and Milton Halem halem@umbc.edu

Key Points:

- Evaluated and benchmarked deep learning models for inferring CO_2 fluxes
- Conducted sensitivity tests on the influence of surface variables and model hyper-parameters on the accuracy of the training and predicting.
- Proposed to use Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM) architecture for improved CO_2 flux inferences.

Abstract

The advent of direct high-resolution global surface measurements of CO_2 from the recently launched NASA Orbiting Carbon Observatory (OCO-2) satellite offers an opportunity to improve the estimate of Net Ecosystem Exchange (NEE) over land. Long-term measurements of CO_2 flux obtained from eddy covariance instruments on Flux towers show large annual differences with NEE calculated by inverse methods using Land Surface Photosynthetic models. This suggests consideration of alternative approaches for calculating seasonal to annual global CO_2 flux over land. Recent advances in deep machine learning models, including recurrent neural nets, have been successfully applied to many inverse measurement problems in the Earth and space sciences. We present evaluations of two deep machine learning models for estimating CO_2 flux or NEE using station tower data acquired from the DOE Atmospheric Radiation Measurement (ARM), AmeriFlux and Fluxnet2015 station datasets. Our results indicate that deep learning models employing Recurrent Neural Networks (RNN) with the Long Short Term Memory (LSTM) provide significantly more accurate predictions of CO_2 flux (~22% -28% improvements) than Feed Forward Neural Nets (FFNN) in terms of Root Mean Square Errors, correlation coefficients and anomaly correlations with observations. It was found that using heat flux as input variables also produce more accurate CO_2 flux or NEE predictions. A non-intuitive machine learning metaphysical result was observed by the omission of CO_2 concentrations as an input variable. Neural net models, in most cases, produce comparable accuracies of CO_2 flux or NEE inferences, when trained with and without CO2 for the same station data.

Plain Language Summary

Vegetation annually removes 40% of global CO_2 emitted to the atmosphere by natural and anthropogenic processes through photosynthetic absorption of CO_2 (sinks) and CO_2 respiration (sources). Scientists have developed ecosystem surface models employing photosynthetic and respiratory processes to infer the annual absorption of CO_2 by plants in order to account for the influence of factors that affect the amount of CO_2 that remains in the atmosphere. However, these model predictions are largely inconsistent with station measurements. In this study, we apply deep machine learning models to train neural nets to predict the turbulent CO_2 flux or NEE based on station data from the Department of Energy (DOE) Atmospheric Radiation Measurement (ARM), AmeriFlux and Fluxnet2015 for different ecosystem sites. We used Recurrent Neural Networks (RNN) with the Long Short Term Memory (LSTM), which produced ~22% - 28% more accurate predictions than a more common Feed Forward Backward Propagation Neural Net.

1 Introduction

The successful launch of the NASA Orbiting Carbon Observatory-2 and its multi-year availability of high spatial resolution of CO_2 concentration and Sun-Induced Fluorescence (SIF) provide a new source of input for arriving at improved estimates of the global annual CO_2 uptake over land. The capacity of continents to take up CO_2 is largely affected by climate variability. Natural sinks over land seen in Figure 1 to have grown in response to the increasing emissions, but appears relatively level in the last decade, though year-to-year variability is large (Le Quéré et al., 2016). The large 2015 El Niño is suspected of producing the most significant anomalous low in the local estimate of continental CO_2 uptake over the past 60-year record. The annual global budget estimate of carbon sink over land has varying levels of uncertainty since it is calculated as a residual difference between the net of CO_2 emissions of fossil fuel and land cover change from the net growth rate of CO_2 in the atmosphere and the ocean sink (Le Quéré et al., 2016). Thus, improvements in the direct calculation of CO_2 flux over land has the potential of mitigating the largest uncertainty among the above individual components of the global CO_2 budget estimates (Li, W. et al. 2016). Machine learning algorithms trained on the network of CO2 flux stations hold the promise of leading to better continental estimates of annual CO_2 flux uptake.

The total global sink of CO_2 is estimated to remove about 55% of all CO_2 emissions for the period 2000-2008, while terrestrial CO_2 sinks account for removal of 29% of all anthropogenic emissions for the same period (Land 53%). Investigations of mid to high latitude atmospheric CO_2, shows growing amplitude in seasonal variations over the past several decades (Graven H.D. et. al.,2013; Piao et al.2017;Bousquet et al.2000;Bala G et al. 2013). Land surface ecosystem models poorly predict the net annual CO_2 flux exchange (NEE) (Keenan, T.F et al. 2012; Chen, B, 2008;Piao, S. et al. 2009). The Atmospheric Radiation Measurement (ARM) program of the ARM, AmeriFlux and Fluxnet2015 stations have been making long-term collections of CO_2flux measurements (in addition to CO_2 concentration and an array of other meteorological quantities) at several towers and mobile sites located around the globe at half-hour and hour frequencies. In addition, since the mid 90's, the DOE, NASA, NOAA and the US Forest service have sponsored over 75 flux tower CO_2 measurement stations distributed over the contiguous US, Canada, Alaska, Brazil and now throughout North and South American continents.

Yet, inverse ecosystem model inferences of continental scale annual CO_2 flux by assimilating the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) CO_2 satellite observations have been shown to be largely inconsistent (2x - 4x factor) with station observations (M. Reuter et. al., 2017), as well as inverse climate ecosystem models (Ott, L. E., et al. 2015). High-resolution satellite surface measurements of CO_2 concentration are now available for 4 years from the NASA OCO -2providing a global view of the seasonal cycles and spatial patterns of atmospheric CO_2 (Eldering, A et al. 2017). Recent advances in computing technologies have enabled significant improvements in the performance and accuracies of deep learning algorithms. We present in this paper, the evaluation of hyper parameters such as list of input variables, learning rate, number of layers and neurons etc.) of the Feed Forward and Recurrent Neural Net algorithms for determining the best learned parameters (optimized weights of neural network models) to employ for application with micro meteorological tower station data for inferring annual turbulent CO_2 flux from station towers at different latitudes and ecosystem types.

1.1 Prediction using neural networks

Can the deep learning models contribute to improved assessments of environmental and atmospheric variances in CO_2 flux or NEE estimation? An investigation of how the biosphere has reacted to changes in atmospheric CO_2 is essential to our understanding of potential climate-vegetation feedbacks. Further, given longer-term extreme climate events such as El Nino's, can regional changes of CO_2 affect positive regional ecosystem feedbacks? The monitoring and understanding of carbon sinks and sources can be inferred from the amount of carbon dioxide (CO2) fluxes into and out of the atmosphere. Thus, global or regional season to annual investigations into new approaches for calculating CO_2 flux should improve the prediction of annual NEE.

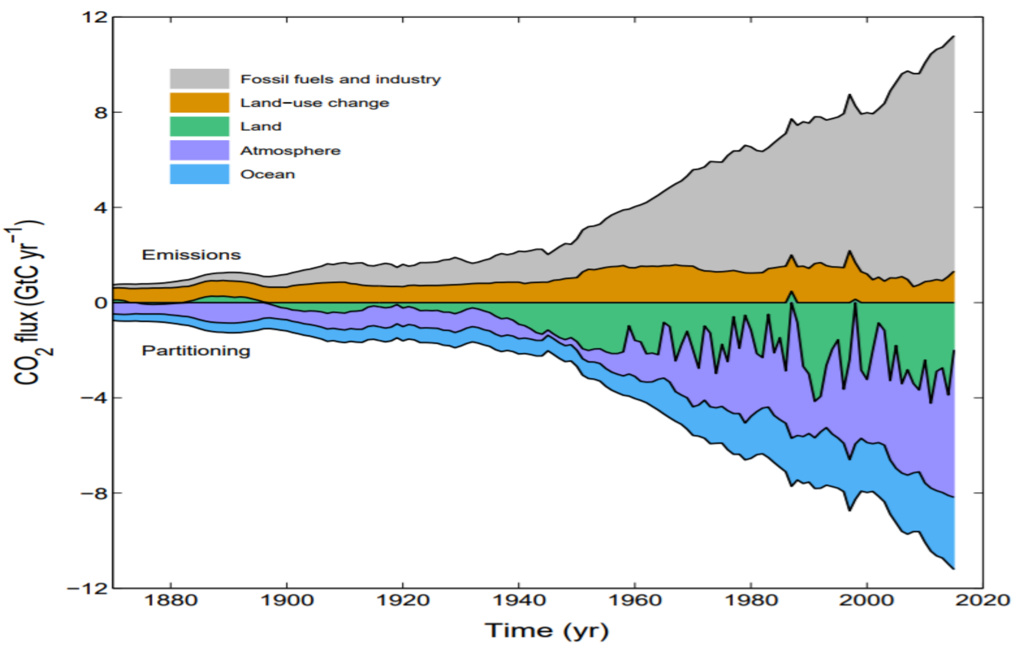


Figure 1 Global CO_2 flux sources and sinks. Source: Global carbon budget 2016(Le Quéré et al, 2016)

Studies show that CO_2 fluxes are correlated with the energy fluxes and environmental variables such as temperature, moisture, heat and others non-linear processes (HeHonglin, 2006). Artificial neural networks (ANNs) are often better candidates for modeling the non-linear process based on data driven input than other statistical methods. ANNs, Feed Forward Backward Propagation Neural Net are being used for the simulation of CO_2 fluxes using atmospheric CO_2 and other variables (Melesse, A.M, 2005; Papale, D.; HeHonglin, 2006; Tramontana, G., 2016).There have been prior efforts to map fluxes across North America with machine learning regression tree models by Xiao et al and globally by Jung et al. For example, Papale and Valentini, 2003 proposed a new methodology involving Feed Forward Neural Network to provide spatial (1 km×1 km) and temporal (weekly) estimates of carbon fluxes of European forests at continental scale. First a separate neural network was used to predict the daily Net Ecosystem Exchange (NEE) at each of 16 different European sites. The dataset used in this study consisted of measurements of CO_2 flux taken at those 16 sites using the eddy covariance technique. The inputs of the network trained for each site were air temperature, air relative humidity, photo synthetically active radiation (PAR) and two series of four fuzzy sets (month of the year and time of day). The original data from each site consisted of about 13,573 samples (one measurement every half an hour) but the dataset was composed only of those examples with complete input and output data. Results were very good with Root Mean Square Errors (RMSE) of 1.149 and Mean Absolute Errors (MAE) of 0.859. The second part of the work involved using another ANN for an assessment of European NEE. The network specifications were twelve input variables, one hidden layer of 5 nodes and one output node of mean NEE expressed in grams Carbon per day (g C day−1). Results of this part were an annual estimate of 0.47 Gigatons of Carbon per year (Gt C yr−1), which is within the range of Europe's 0.2 - 0.7 Gt C yr−1 estimate in (Schulze, E.D., et al., 2001).

Melesse, A.M. and Hanley, R.S, 2005 showed that using a Feed Forward Neural Network (FFNN) to predict CO_2 flux over large areas is a promising technique which can be used instead of the eddy covariance method at a single point (Baldocchi, D., 2014). However, the study was limited to a very short period of data observations as well as the number of flux towers.

1.2 Objectives

Recent advances in deep learning (DL) models have been successfully applied to many applications including speed and image recognition (Hinton, G. E et al 2006;Bengio, Y., 2006;Ranzato, M.A, 2006; Krizhevsky, I. Sutskever et al, 2012;R. K. Srivastava et al 2015; Buduma, N., et al. 2015), which motivated us to investigate applying DL models for prediction of hourly CO_2 fluxes. In this study, we are going to build DL models including Feed Forward Neural Network (FFNN) and Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM) models to learn the physical complexity, space and time dependency of the measurements of CO_2 concentration, humidity, pressure, temperature, wind speed etc. for predicting the CO_2 flux from long term measurements at Atmospheric Radiation Measurement (ARM), AmeriFlux, and Fluxnet2015 station data for different ecosystem sites.

The motivation for using LSTM is to utilize the temporal dependencies or patterns within the long-term measurements of the data. Data was separated by year into Training, Validation and Test data sets. These deep learning models are shown to be efficient, dependable, and very powerful and can be employed in predicting CO_2 flux or NEE. We will also show that RMSE is lower at high latitude sites because there is more variation in the input data. Deep learning models can be used with historical data to further infer the present state of the atmosphere in the future.

2 Methods

Deep Learning (DL) neural networks coupled with new advances in computer technologies attempt to mimic the human brain activity of neurons in the neocortex. DL models can be trained to recognize complex patterns in digital representations of sounds, images, and other data (Krizhevsky, I. Sutskever, 2012). Because of advances in computer technologies, one can now model many more layers of virtual neurons than ever before (Y. LeCun, 2015;R. K. Srivastava, 2015; J.C. Duchi, E. Hazan et al 2011). With deep layer architectures exploiting computational acceleration, new training techniques are producing remarkable advances in speech and image recognition; and in medicine by identifying molecules that are leading to new drugs. Recurrent neural networks (RNN) with the Long Short Term Memory (LSTM) model is one of the deep learning model, which are able to successfully learn data with long range temporal dependencies such as time series data, or machine translation in language modeling (S. Hochreiter et al 1997;Sutskever, I., et al 2014). Extending DL models for applications such as Earth and space science as well as other science disciplines requires more conceptual breakthroughs and further advances in processing power. In this paper, we apply deep FFNN and RNN, LSTM to train and predict CO_2 flux or NEE using ARM/AmeriFlux/Fluxnet2015 station tower data.

2.1 Feed Forward Neural Network

The FFNN model simulates and learns the relationship and interaction between input variables and CO_2 flux (Bebis, G. and Georgiopoulos, M., 1994). The model does not need to know the complicated relationship and dependencies between the CO_2 flux and the input variables in advance. We choose to use FFNN and RNN to model the CO_2 flux since the relationship between CO_2 flux and its input variables are known to be nonlinear.

A FFNN includes artificial neurons, which are arranged into layers (figure 2). The input layer consists of neurons, which takes on the input values. In this application the input values are variables such as temperature, sensible heat flux, latent heat flux and friction velocity, etc. The last layer is called output layer having neurons, which holds the output values that are the prediction target CO_2 flux. There can be many layers in between input layer and output layer as needed by the DL model. Each node is only connected to a node in a prior adjacent layer. Nodes at the same layer are not connected with each other.

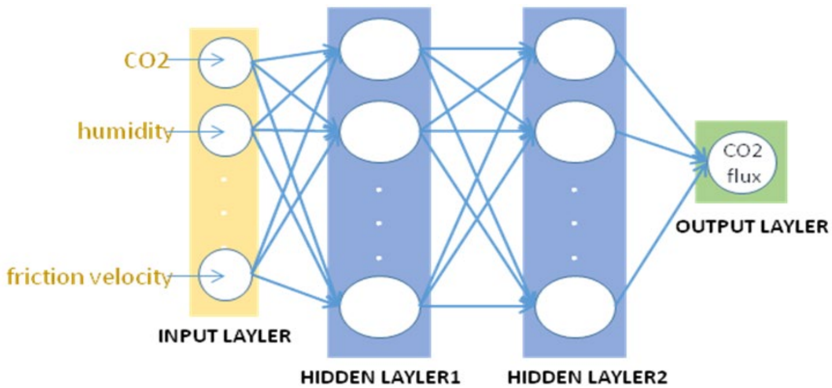


Figure 2 Feed Forward Neural Network model

Figure 2 presents the topological architecture of the multilayer FFNN. These layers are called the hidden layers. Let $w^{k}_{ii,}$ be the weight of the connection between the i^{th} neuron in the k^{th} layer with the j^{th} neuron in the $k + 1^{th}$ layer, x_i is the value given to the i^{th} input node and a bias value b_j is considered as an additional input node that acts to shift the point of activation. The output value of the i^{th} neuron is determined by the below equation.

$$y_j^k = f(\theta_j, x_i) = f(\sum w_{ij}^k x_i + b_j) \quad (1)$$

If we use θ_j as a vector, then the output of the node is modeled as a function $f(\theta_j, x_i)$ where x and y are the input and the output. These weights show the important nature of the connections in the network and the ability of solving the problem with a neural network depending on finding the optimal weights and the parameter value. The $f(\theta_j, x_i)$ function applied to the neuron is known as the activation function. The *sigmoid* function, $\sigma(x) = \frac{1}{1+e^{-x}}$ which is differentiable, monotonic, non-linear and is often used as the activation function. Non-linearity is fit to model non-linear patterns in our training dataset. The *elu* activation function, an Exponential Linear Unit,

$$\sigma(x) = \begin{cases} \alpha.\,(e^{x} - 1), & x < 0 \\ x, & x \geq 0 \end{cases}$$

converge cost to zero faster and produce more accurate results. The training process will find the θ_j vector, which produces the output $f(\theta_j, x_i)$, namely, the prediction value that is close to the desired output value y. In other words, given a set of inputs and their desired outputs, we try to find θ_j, which minimizes the difference between the desired outputs and the neural network outputs, for all of data points in the training dataset. The RMSE is used to measure the differences. Gradient-based methods such as Adaptive Gradient Algorithm (Adagrad) (J.C. Duchi 2011) are used to find the optimization weights and bias.

$$E(\theta, XY) = \sqrt{\sum_{x_i, y_{i \in XY}} (y_i - f(\theta, x_i))^2}$$

XY is the training set and x_i and y_i are the input and the desired output training values and $f(\theta, x_i)$ the neuron network output. A back propagation gradient descent training algorithm was performed using RMSE as error function to adjust the weight of neurons. We also use R^2, the coefficient of determination between true observation and model's prediction to measure the performance of the models.

The modeling dataset is divided into three groups: (i) a training dataset used to determine the weights during neural network training; (ii) a test dataset used during network training to calculate the errors to prevent overtraining; and (iii) a validation dataset used to assess the network's performance with 'new' data, which removes the possibility of the network over-fitting on training and test datasets. This is the case when the neural network models very well on training dataset and predicts poorly on unseen test dataset.

2.2 Recurrent Neural Network

A Long Short Term Memory (LSTM) network is a Recurrent Neural Network (RNN) where connections between units in a layer form a directed graph along a sequence (see figure 3). This RNN architect exhibits dynamic temporal behavior for a time sequence. Unlike FFNN, where input variables at particular times are trained independently and there are no connections within or between two adjoining layers of a network, RNN enables the insertion of state operations or gates between layers. Thus, when learning the model for a particular time using FFNN, there is no consideration of the historical data from prior times. The RNNs can use their internal state (memory in LSTM unit) to process sequences of inputs. For example, in this application, 12 input measurements of CO_2 concentration, humidity, pressure, temperature, wind speed, etc have been observed by every half hour forming a time sequence. Time dependency can be learned by LSTM. It can selectively learn when to remember or forget things by controlling information flow through block conditions in each LSTM unit state called gates. There are three types of gates in each LSTM unit. The Forget Gate decides which information to discard from the block. The Input Gate conditionally makes decision on which values from the input will be used to update memory state. The output Gate will conditionally decide which value will be output based on the input and the memory of the block. Gates of the LSTM units have weights that are learned during the training procedure. LSTM units are often implemented in deep multiple layer architects. This model capability is one of the advances made in deep machine learning models, which can be applied to solve difficult sequence problems in machine learning. It has been successfully implemented (Sutskever, I., et al 2014). The detail of LSTM units and the architecture of this model are referenced in (Hochreiter, J., and Schmidhuber, J.,1997;Sutskever, I., et al 2014).

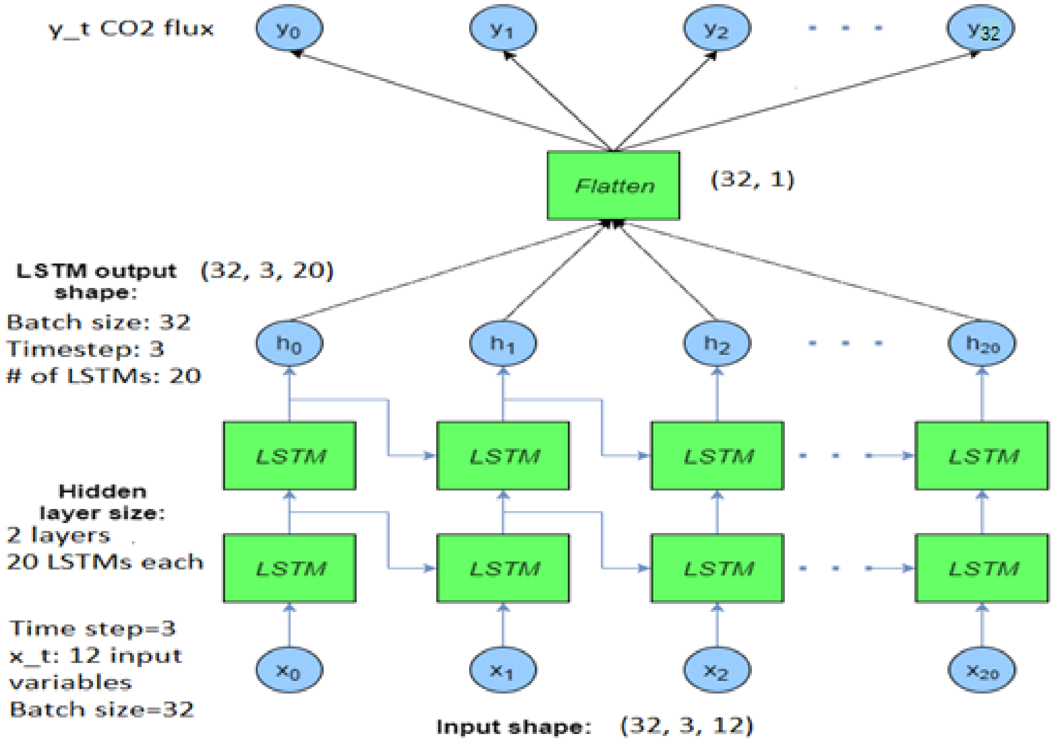


Figure 3 Recurrent neural network with LSTM model

Figure 3 shows the following structure of the RNN with LSTM model we will employ. Each observed time step consists of an observation of 12 input variables, and the number of time steps in each LSTM unit can be configured. A time step=3, means that the current observation is learned using two previous observations of the 12 input variables. The two layers of 20 LSTM units each are thus configured. A batch of 32 such input observations can be specified and trained at each LSTM. The Flatten unit is used to produce single output, which is the prediction of CO_2 flux. The DropOut technique is used for training to improve the performance of the neural network. On training, it drops x% of units in each layer (Baldi, Pierre and Sadowski, Peter, 2014). All de-activate units would not participate in the training propagation. This technique will block the error values from in each layer to the output level. The dropout technique only applies on the training phase. Advanced gradient algorithm such as SGD and AdaGrad (J.C. Duchi, E. Hazan, and Y. Singer, 2011) will be used to train the models.

3 Data description

AmeriFlux, Fluxnet2015 and Atmospheric Radiation Measurement (ARM) tower data at different sites have been used in this study. The site name, siteID, latitude, longitude, type, and date time ranges are showed in figure 4 and table 1. The half an hour and hourly data have been used for experiments.

Table 1 The ARM, AmeriFlux, Fluxnet2015 site list

Lat	Lon	Date Range	SiteID	Site Name	Type
43.732	10.291	2013-2014	FLX-IT-Sr2	San Rossore 2	Evergreen Needleleaf Forests
42.38	12.0222	2011-2014	FLX-IT-Ca3	Castel d Asso 3	Deciduous Broadleaf Forests
55.4859	11.6446	1996-2014	FLX-DK-Sor	Soroe	Deciduous Broadleaf Forests
61.8474	24.2948	1996-2014	FLX-FI-Hyy	Hyytiala	Evergreen Needleleaf Forests
48.2167	-82.1556	2003-2014	FLX-CA-Gro	Ontario	Groundhog River, Boreal Mixedwood Forest
45.8444	7.5781	2008-2014	FLX-IT-Tor	Torgnon	Grasslands
50.7867	13.7213	2008-2014	FLX-DE-Obe	Oberbarenburg	Evergreen Needleleaf Forests
46.0147	11.0458	2003-2013	FLX-IT-Mbo	Monte Bondone	Grasslands
43.7413	3.5957	2000-2014	FLX-FR-Pue	Puechabon	Evergreen Broadleaf Forests
-14.063	131.3181	2007-2013	FLX-AU-DaP	Daly River Savanna	Grasslands
51.0792	10.453	2000-2012	FLX-DE-Hai	Hainich	Deciduous Broadleaf Forests
-3.018	-54.9714	2000-2004	FLX-BR-Sa3	Santarem-Km83	Logged Forest
40.0062	-88.2904	1996-2008	Ameri-US-Bo1	Bondville	Croplands Annual rotation between corn (C4) and soybeans (C3)
39.3232	-86.4131	1999-april2018	Ameri-US-MMS	Morgan Monroe	Deciduous Broadleaf Forests
45.5598	-84.7138	2007-oct2017	Ameri-US-Umb	Univ. of Mich.	Deciduous Broadleaf Forests
39.8379	-74.3791	2005-2014	FLX-US-Ced	Cedar Bridge	Closed Shrublands
44.4523	-121.5574	2002-2014	FLX-US-Me2	Metolius	Evergreen Needleleaf Forests - mature ponderosa pine
-15.438	23.2528	2000-2009	FLX-ZM-Mon	CarboAfrica Mongu	Deciduous Broadleaf Forests
70.8291	147.4943	2003-2014	FLX-RU-Cok	Chokurdakh	Open Shrublands
40.0062	-88.2904	1999	Ameri-US-Pon	Ponca City	Croplands
35.1426	-111.7273	2005-2010	Ameri-US-Fmf	Flagstaff	Evergreen Needleleaf Forests Managed Forest
36.6058	-97.4888	2012-2014	ARM-US-SGP	ARM Southern Great Plains	Croplands
71.3225	-156.6092	2011-2017	ARM-US-Brw	Barrow	Permanent Wetlands
-3.2897	-60.6321	2014-2015	ARM-US-MAO	Manacapuru	Green Ocean Amazon
44.7143	-93.0898	2003-2017	Ameri-US-Ro1	Rosemount	Croplands Annual rotation between corn (C4) and soybeans (C3)

Two sets of input variables have been used. The first set uses heat fluxes including net radiance at the top of the atmosphere (W m-2), latent heat (W m-2), sensible heat (W m-2), soil heat flux (W m-2), air and soil temperatures (deg C). The second set of input variable does not use the heat fluxes since they are not typically available globally. In this case, we have incoming shortwave radiation (W m-2), outgoing shortwave radiation (W m-2), Carbon Dioxide (CO_2) mole fraction in wet air (μmolCO2 mol-1), Air temperature (deg C), and Precipitation (kPa) have been used as input variables. The output of both FFNN and RNN models predict CO2 flux (Carbon Dioxide (CO2) turbulent flux (μmolCO2 mol-1) for dataset at ARM/AmeriFlux or NEE (umolCO2 m-2 s-1) for dataset at Fluxnet2015. Figure 5 plots some of the variables of AmeriFlux tower data at MMS site from Jan 1, 1999 to April 12, 2018 including Temperature (deg C), CO_2 Flux (μmolCO2 m-2 s-1), Shortwave Radiation incoming (W m-2), Shortwave Radiation Outgoing (W m-2), Net Radiation (W m-2), CO2 (μmolCO2 mol-1), Precipitation (mm), Latent Heat (W m-2), Sensible Heat (W m-2), Soil Heat Flux (W m-2), Soil Temperature (deg C). As seen in figure 5, there are more variances in radiation, temperature, sensible heat flux and latent heat flux corresponding to the CO_2 flux.

CO_2 flux with reasonable accuracies. The preprocessing of the data includes two parts: filtering row data and normalizing all input variable data. Row data with filled values, which have flags 'y' and out of normal CO_2 flux ranges were filtered. Since the input variables have different value ranges (see fig. 5), which vary significantly, all input data are normalized to scale between 0 and 1 to improve the time to reach convergence.

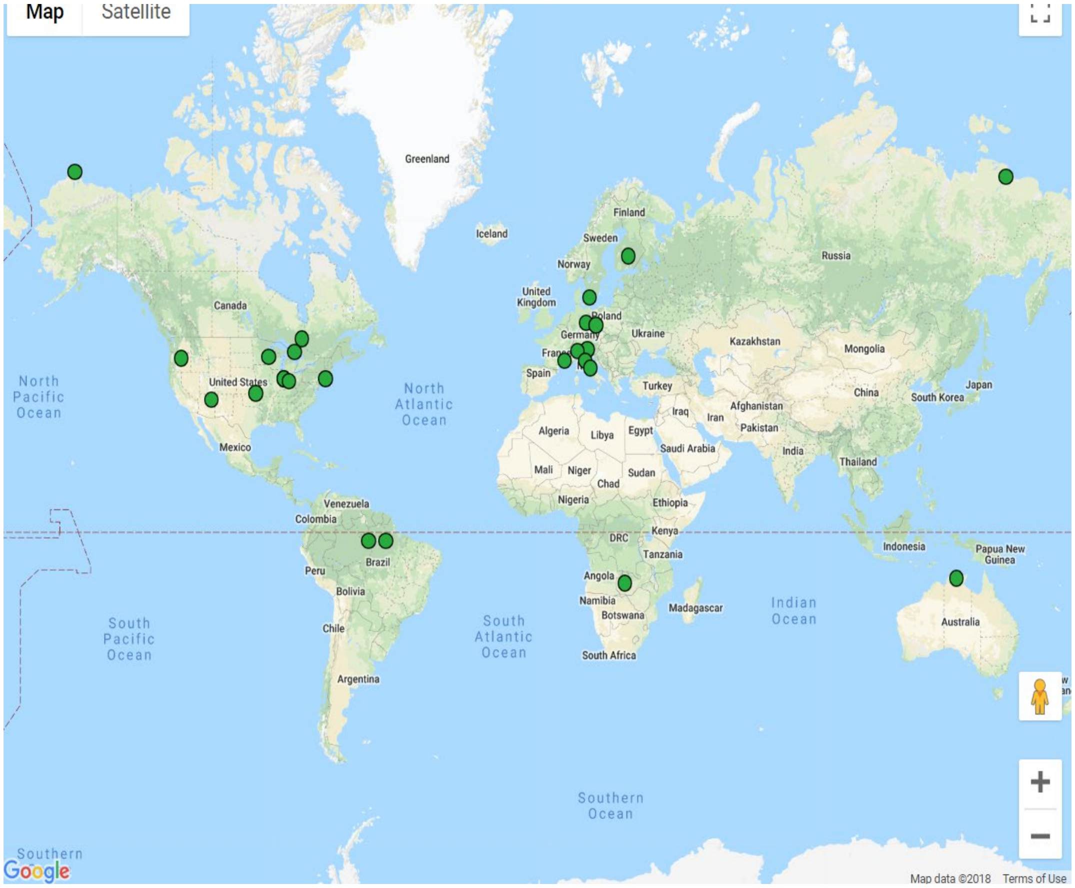


Figure 4 the map of ARM, AmeriFlux, Fluxnet all tower location used in this study

The input dataset needs to be preprocessed and cleaned in order for the neural nets to infer the The normalized equation is as follows:

$$X_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Where $X_{normalized}$ is the normalized value of input variable, X_{max} is the maximum value of the input variable and X_{min} is the minimum value of the input variable.The only output variable of the learning model is CO_2 flux and it is de-normalized into original data ranges.

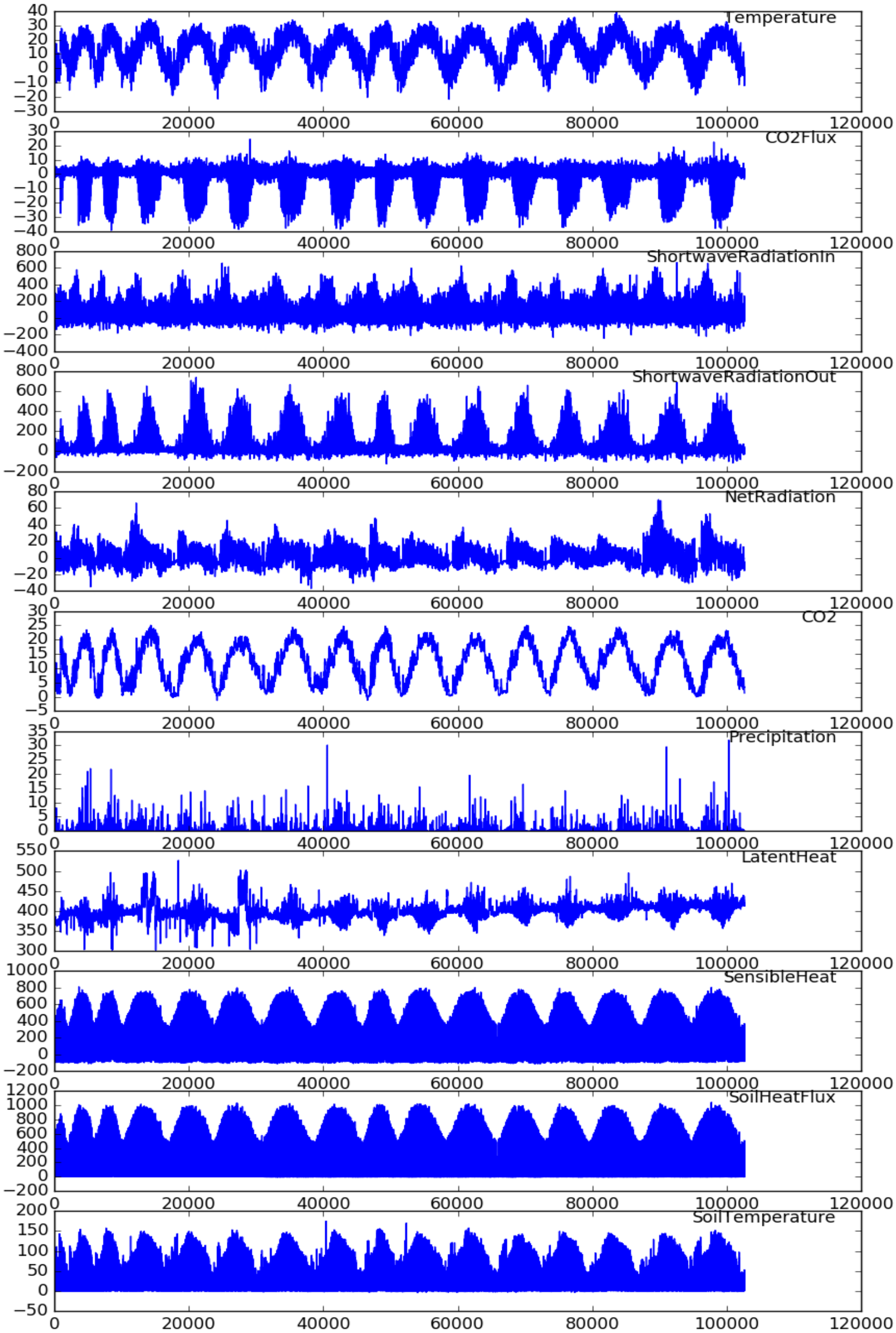


Figure 5 AmeriFlux hourly data at AmeriFlux at Morgan-Monroe, Indiana (MMS) Temperature, CO2 Flux, Shortwave RadiationIn, Shortwave RadiationOut, NetRadiation, CO2, Precipitation, Latent Heat, Sensible Heat, Soil Heat Flux, Soil Temperature from Jan 1, 1999 to April 12, 2018**.**

4 Experimental Results

In section 4.1, we present the experiments for the comparison of FFNN and RNN/LSTM models for predicting CO_2 flux from both the ARM and Ameriflux station data and predictions of NEE from the Fluxnet2015 stations. Neural net training and test results depend on a variety of selected hyperparameters. We use an algorithm, WEKA, that applies principal component analysis for the ranking of the input variables based on the station data. Slight differences in the rankings can occur for different AmeriFlux, ARM and Fluxnet2015 stations. Experimental results are presented utilizing a common set of input variables that are the highest or near highest ranked station variables obtained from WEKA for both FFNN and LSTM. The input variables selected for all the experiments in section 4.1 consisted of: net radiation, latent heat, sensible heat, soil heat flux, air and soil temperatures. It should be noted that CO_2 observations are not among the above input variables used in these experiments. When heat fluxes are not available at any station, the following set of input variables are used to predict CO_2 flux: shortwave radiation (incoming), shortwave radiation (outgoing), air temperature, and precipitation. These variables also appeared in the WEKA rankings. Again, CO_2 is not an input variable for both FFNN and RNN models for predicting CO_2 flux. Similar findings for calculating NEE without CO_2 as an input variable were obtained by Papali and Valentini, 2003 and Medlyn et. al., 2005. We show below that Recurrent Neural Networks (RNN) with the Long ShortTerm Memory (LSTM) improvements in the prediction of CO_2 flux by ~22% to 28% in terms of RMSE and R^2 scores, and 28% improvements in anomaly correlations. Section 4.2 shows the influence of selecting WEKA rankings of variables as input for both FFNN and LSTM. We also show that both neural nets are dependent on the influence of latitudinal variations of CO_2 flux and NEE station data. Both FFNN and LSTM produce their lowest CO_2 flux scores at high latitudes while stations near or at the equator produce the largest RMSE errors. We suspect this variability in performance is influenced by the variability of the input variables, i.e. larger seasonal variations temperature and net radiation occur at high latitudes and have less variability at low latitudes near or at the equator.

4.1 Comparison of FFNN and RNN models.

Table 2 reports the comparisons between FFNN and RNN models using RMSE and R^2 correlation metrics. These experiments use net radiation, latent heat, sensible heat, and soil heat flux, air and soil temperatures as input variables and predict CO_2 flux for ARM and AmeriFlux stations and predict NEE for the Fluxnet2015 tower stations. The dataset is divided into three groups: (i) 80% of dataset has been used to determine the weights during neural network training; (ii) 10% of dataset has been used during network training to calculate the errors to prevent overtraining; and (iii) 10% of dataset is used as validation dataset to assess the network's performance with 'new' data. FFNN model used in this section was configured to run using 10 layers to 20 layers, 20 neurons each layer. RNN models used 2 layers each have 20 LSTM units. Both models were trained using back propagation gradient descent algorithm with 'elu' activation. The mean improvement for the 20 stations is 22% and the correlation improvement is 20%. Experiments also show that for the 14 stations with data having longer records than a decade, the mean improvement for the LSTM with respect to FFNN is 28% and the R2 correlation score is 27%. This is indicative of the influence of incorporating the time variations into the machine learning algorithm of RNNs.

Table 2 Comparisons between FFNN and RNN, LSTM models

	RMSE						R2 Correlation					
	Training		Test Validation		Test		Training		Test Validation		Test	
Site	FFNN	RNN	FFNN	RNN	FFNN	RNN	FFNN	RNN	FFNN	RNN	FFNN	RNN
FLX-IT-Sr2	3.74	3.39	4.62	3.57	3.23	3.24	0.8	0.84	0.79	0.88	0.67	0.67
FLX-IT-Ca3	2.03	2.14	0.91	0.87	1.22	1.08	0.83	0.81	0.64	0.67	0.81	0.85
FLX-DK-Sor	4.53	3.01	4.51	3.23	5.06	3.03	0.68	0.86	0.71	0.85	0.64	0.87
FLX-FI-Hyy	2.19	1.55	2.18	1.66	2.93	2.39	0.73	0.87	0.7	0.83	0.71	0.81
FLX-CA-Gro	3.04	2.39	2.74	2.08	2.72	2.25	0.61	0.76	0.61	0.78	0.55	0.69
FLX-IT-Tor	2.68	1.66	3.01	1.85	3.69	2.52	0.64	0.86	0.55	0.83	0.68	0.85
FLX-DE-Obe	3.37	2.82	3.44	3.2	3.8	3.57	0.76	0.83	0.75	0.78	0.79	0.82
FLX-IT-Mbo	4.06	2.87	4.7	3.38	4.04	2.88	0.52	0.76	0.53	0.76	0.52	0.75
FLX-FR-Pue	2.21	1.69	2.06	1.69	1.84	1.88	0.73	0.84	0.77	0.85	0.5	0.48
FLX-AU-DaP	4.14	2.98	4.32	3.18	4.13	3	0.58	0.78	0.64	0.81	0.66	0.82
FLX-DE-Hai	4.67	2.8	5.07	3.36	4.83	3.04	0.64	0.87	0.55	0.8	0.36	0.75
FLX-BR-Sa3	5.99	5.41	6.21	6.21	6.43	5.94	0.71	0.76	0.7	0.7	0.73	0.77
Ameri-US-Bo1	5.33	4.62	5.84	4.81	6.15	5.58	0.39	0.54	0.43	0.61	0.38	0.46
Ameri-US-MMS	4.28	2.93	4.31	2.94	4.62	3.07	0.63	0.83	0.57	0.8	0.58	0.82
Ameri-US-Umb	4.23	2.93	4.87	3.56	4.47	3.12	0.55	0.78	0.6	0.79	0.55	0.78
FLX-US-Ced	3.99	2.73	4.41	2.69	3.65	2.79	0.58	0.8	0.24	0.72	0.67	0.8
FLX-US-Me2	3.12	2.76	3.72	3	4.14	3.56	0.69	0.76	0.55	0.71	0.67	0.76
FLX-ZM-Mon	3.07	2.79	2.96	2.74	2.1	3.38	0.74	0.78	0.83	0.85	0.21	-1.04
FLX-RU-Cok	1.49	1.34	1.55	1.29	1.73	1.66	0.57	0.66	0.48	0.61	0.48	0.52
Ameri-US-Ro1*	5.33	2.88	5.84		6.15	3.22	0.39	0.81	0.43		0.38	0.81
Ameri-US-Pon*	4.01	2.46			3.91	3.01	0.68	0.86			0.21	0.5

* indicates did not use 80%,10%,10% data split

The Morgan Monroe State Forest station in Indiana, having the longest continuous record available to 2018, was selected for investigating the performance of Neural Nets using anomaly correlation metrics. Figure 7 shows a plot of the anomaly for the observed CO_2 flux for years 2011 to 2013, where the mean of the monthly data for the years 1999 to 2010 are removed. Superimposed are the predicted CO_2 flux anomalies for FFNN and LSTM anomalies with the same observed monthly means removed. The image shows that the LSTM fits the observed anomaly better for positive anomalies, but one cannot infer the values of the LSTM for negative anomalies form this image. Thus, in table 3, we present the anomaly correlation metrics for the above MMS station as well as two additional stations, one in Finland and the other in France. The anomaly correlation of three stations (Ameri- US- MMS, FLX-FI-Hyy, FLX-FR-Pue) shows that LSTM produces about ~29% improvements in the anomaly correlation metric compared with FFNN. This result is consistent with the mean performance improvements obtained in section 4.1 for the 20 station sites.

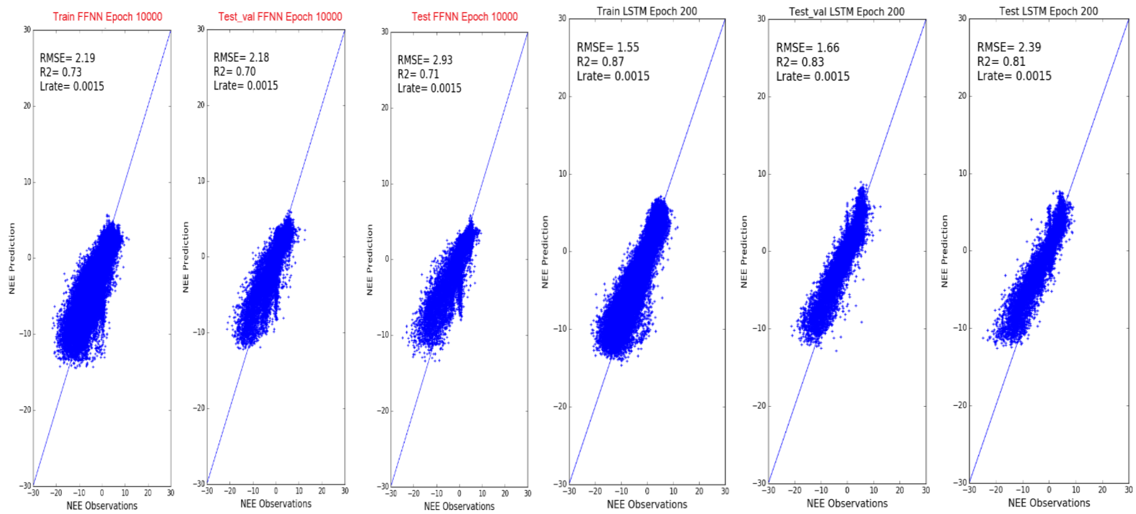


Figure 6Comparison of FFNN and RNN using Fluxnet2015, FLX-FI-Hyy station

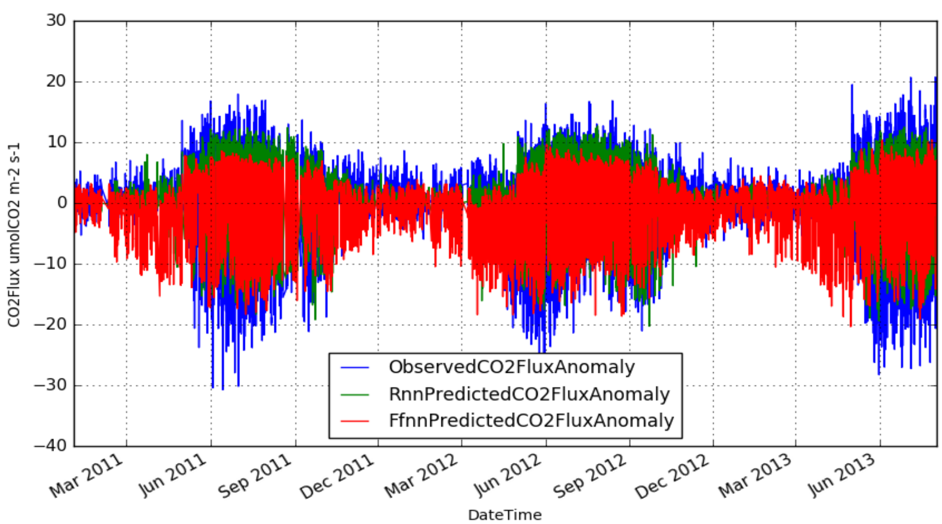


Figure 7 Anomaly correlation comparisons between FFNN and RNN, LSTM

Figure 6 shows that both FFNN and RNN models produce good predictions at a high latitude station with a long data history.

Table 3 Anomaly correlations for FFNN and RNN

	Anomaly correlation					
	Train		Test Validation		Test	
Site	FFNN	RNN	FFNN	RNN	FFNN	RNN
Ameri US-MMS	0.41	0.78	0.36	0.76	0.38	0.78
FLX-FI-Hyy	0.64	0.85	0.61	0.81	0.62	0.79
FLX-FR-Pue	0.69	0.84	0.73	0.84	0.45	0.47

Figure 8 shows the LSTM model predictions of CO_2flux at the Ameriflux-US-Ro1 station, where the crop is rotated yearly between a C3 and C4 plant. The RMS error for the test is 3.22 and the model produces a good R2 correlation with observation of 0.81. The purpose of this image is to show that the inferences of CO2 flux are comparable for both crops.

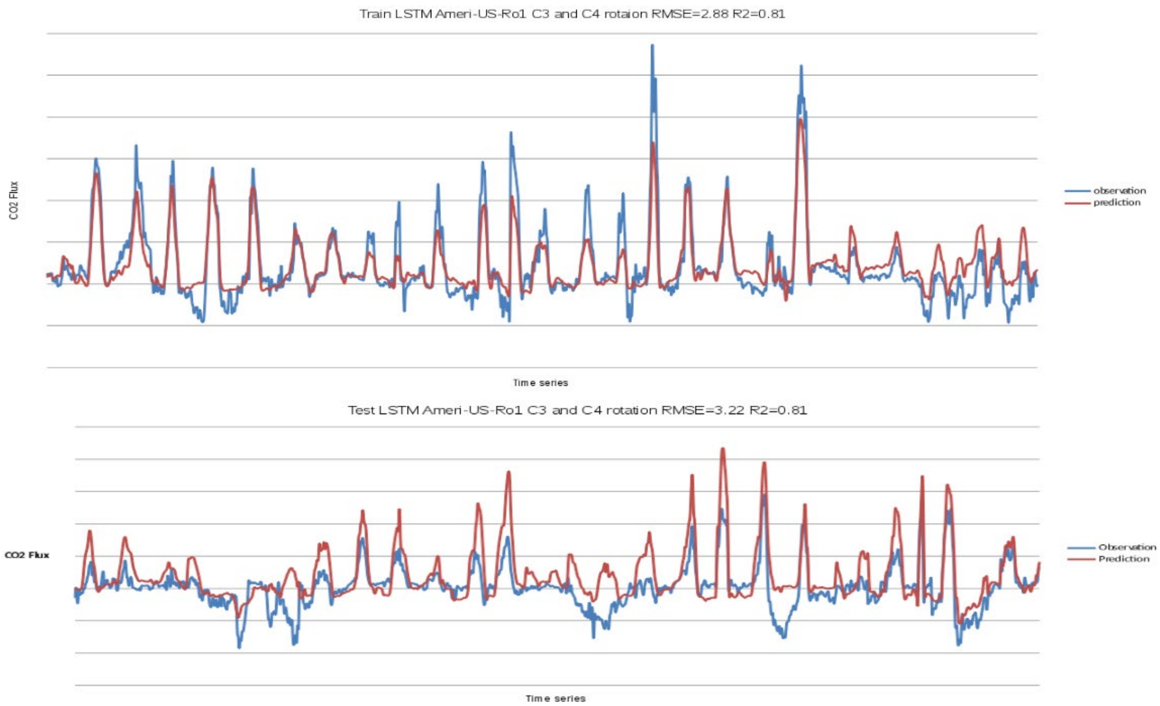


Figure 8 prediction of CO2 flux using Ameri-US-Ro1 C3 and C4 rotation station

4.2 The influence of input variables on model prediction

We used WEKA (a data mining tool) to perform a principal component analysis (PCA) in conjunction with a Ranker search which ranks attributes by their individual evaluations of the top eigen value on some of the selected sites. This software encapsulates a collection of Machine Learning techniques and provides a nice graphical user interface for quick exploration of those techniques on user's provided dataset. The output of the PCA was a set of components formed by means of linear combinations of the correlated attributes.

Table 4 PCA ranking of input variables

Ranked	ARM SGP	Ameri-US-Ced	Ameri-US-UMB
1	-0.376SW_IN-0.373NETRAD-0.334SW_OUT-0.325G-0.321H...	-0.339NETRAD-0.339PPFD_IN-0.339SW_IN-0.304LE-0.296VPD_PI...	-0.358SW_IN-0.357LW_OUT-0.357LW_IN-0.356SW_OUT-0.349NETRAD...
2	-0.604H2O-0.54TA-0.537NEE_PI+0.094P+0.094SWC_1...	0.48 TS_1+0.468TA+0.306G-0.284USTAR-0.256WS...	0.421LE+0.369WS+0.312H2O+0.312CO2+0.29 NEE_PI...
3	-0.422TS_1-0.385LW_IN+0.379SWC_2+0.372SWC_1+0.332PA...	-0.618TIMESTAMP_START-0.618TIMESTAMP_END-0.204CO2_1-0.204CO2_2+0.19 WS...	-0.533TS_1_1_1-0.514SWC_1_1_1+0.443SC+0.32 NEE_PI+0.275RECO_PI...
4	-0.56SWC_1-0.543SWC_2+0.289PA-0.264TS_1-0.236LW_OUT...	-0.507WS+0.476PA-0.434USTAR-0.251NEE_PI-0.233P...	0.748GPP_PI-0.535RECO_PI-0.181P-0.137TA-0.135TS_1_1_1...
5	-0.9P-0.208TA-0.204CO2_1-0.195RH+0.154SWC_2...	-0.471WD-0.345CO2_1-0.345CO2_2+0.339RH+0.309P...	0.495PA-0.42NEE_PI-0.359RECO_PI+0.268TA+0.248SC...
6	0.726CO2_1-0.372PA-0.344P+0.341RH+0.153LE...	0.45 CO2_2+0.45 CO2_1-0.328SC+0.302RH-0.263PA...	-0.478PA+0.422CO2+0.421H2O-0.347H-0.309SWC_1_1_1...
7	0.7 PA+0.29 LW_IN+0.269TS_1-0.234NEE_PI+0.22 H...	0.611WD-0.522SC-0.36PA-0.204P-0.191NEE_PI...	0.461P-0.404WS+0.372NEE_PI+0.37 GPP_PI-0.326LE...
8	0.633RH-0.461CO2_1+0.349LE+0.308LW_IN-0.226TS_1...	-0.836P-0.371WD+0.226USTAR-0.172SC+0.15 WS...	0.554PA-0.495P-0.32TA-0.296WS+0.261CO2...
9	-0.545LE+0.446H+0.337NEE_PI-0.319TA-0.26G...	-0.601SC+0.449PA+0.286NEE_PI-0.246LE-0.223RH...	0.803SC+0.468SWC_1_1_1-0.222PA+0.175TS_1_1_1-0.148P...
10	0.512TA-0.472NEE_PI-0.365LE-0.337LW_IN+0.301H...	-0.508PA-0.417WD-0.382RH-0.352TS_1+0.295VPD_PI...	-0.709TS_1_1_1+0.602SWC_1_1_1-0.202H+0.164PA-0.152SC...
11	0.598LW_IN-0.492TS_1-0.275RH-0.257H+0.24 TA...	0.457G-0.42TS_1-0.351LE+0.33 RH+0.285NEE_PI...	-0.527P+0.504TA+0.452H-0.261PA-0.258WS...

The results from WEKA's PCA and Ranker attribute selector in table 4 showed the following set of variables to have the highest rank. The following variables had higher ranking attributes and

were selected for experiments with both FFNN and LSTM in section 4.1 and 4.2: net radiation (NETRAD) (shortwave radiation incoming- SW_IN, outgoing- SW_OUT), temperature, latent heat (LE), sensible heat (H), soil heat (G), soil temperature(TS), pressure (PA),CO_2, H_2O, precipitation (P).
Since heat fluxes are not always available and they are a derived data product calculated from algorithms, we designed the experiments using the following two sets of input variables for evaluating the influence of variables for model prediction. The results are presented in table 5 and figure 9.

- Input variables with using heat fluxes (HF all): net radiation, latent heat, sensible heat, and soil heat flux, air and soil temperatures.
- Input variables without using heat fluxes (HF all except heat fluxes with CO_2): Shortwave radiation (incoming), Shortwave radiation (outgoing), Carbon Dioxide (CO_2) mole fraction in wet air, Air temperature, and Precipitation.
- Input variables without using heat fluxes (HF all except heat fluxes without CO_2): Shortwave radiation (incoming), Shortwave radiation (outgoing), Air temperature, and Precipitation.

Table 5 The influence of input variables to the model prediction

		Train				Test			
Site	Input Vector	RMSE		R2		RMSE		R2	
		FFNN	RNN	FFNN	RNN	FFNN	RNN	FFNN	RNN
Ameri-US-Pon	HF all	4.01	2.46	0.68	0.86	3.91	3.01	0.21	0.5
	HF all except heat fluxes with CO2	4.81	3.34	0.53	0.77	4.32	3.28	0.62	0.39
	HF all except heat fluxes without CO2	4.8	3.38	0.55	0.77	4.1	2.88	0.1	0.53
Ameri-US-MMS 1999-2014	HF all	3.95	2.77	0.7	0.85	3.39	2.76	0.67	0.84
	HF all except heat fluxes with CO2	5.42	3.45	0.46	0.78	5.09	3.34	0.44	0.76
	HF all except heat fluxes without CO2	4.71	3.5	0.59	0.77	4.49	3.36	0.57	0.76
Ameri-US-Fmf	HF all except heat fluxes with CO2		1.82		0.7		1.91		0.66
	HF all except heat fluxes without CO2		1.85		0.69		1.84		0.68
Ameri-US-MMS 1999-april 2018	HF all except heat fluxes with CO2		3.42		0.78		3.5		0.76
	HF all except heat fluxes without CO2		3.49		0.77		3.53		0.75

Table 5 presents results of calculating CO_2 fluxes at the 4 stations in table 5 for all heat flux input variables with and without CO_2 concentration. LSTM shows less than 0.1% difference while FFNN shows a 10% variation in RMSE.

Figure 9a shows the comparison of FFNN and RNN using AmeriFlux data at Morgan-Monroe State Forest, Indiana (MMS) using input variables: net radiation, latent heat, sensible heat, and soil heat flux, air and soil temperatures. The data used was from Jan 1, 1999 to Dec 31, 2014. Half of the data was used for training results in RMSE= 3.95 with R^2 =0.70 and RMSE=2.77 with R^2=0.85 for FFNN and RNN respectively. The models tested using another half of MMS data show RMSE= 3.93 with R^2 =0.67 and RMSE=2.76 with R^2=0.84 for FFNN and RNN, respectively.

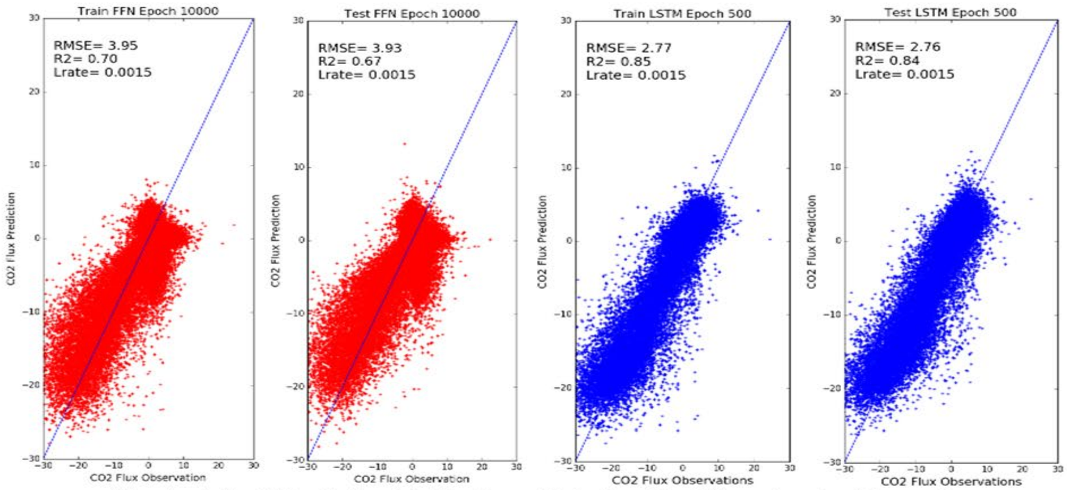
Figures 9b and 9c show the difference with and without CO_2 for non-heat flux data. The input variables in Figure 9b for experiments using AmeriFlux data at Morgan-Monroe, Indiana (MMS) are: Shortwave radiation (incoming), Shortwave radiation (outgoing), *CO_2 concentration*, Air temperature, and Precipitation. The hourly data from Jan 1, 1999 to Dec 31, 2014 (131,400 samples) have been used. Half of the data was used for training results in RMSE= 5.42 with R^2

=0.46 and RMSE=3.45 with R^2=0.78 for FFNN and RNN respectively. The models tested using another half of MMS data show RMSE= 5.09 with R^2 =0.44 and RMSE=3.34 with R^2=0.76 for FFNN and RNN, respectively.

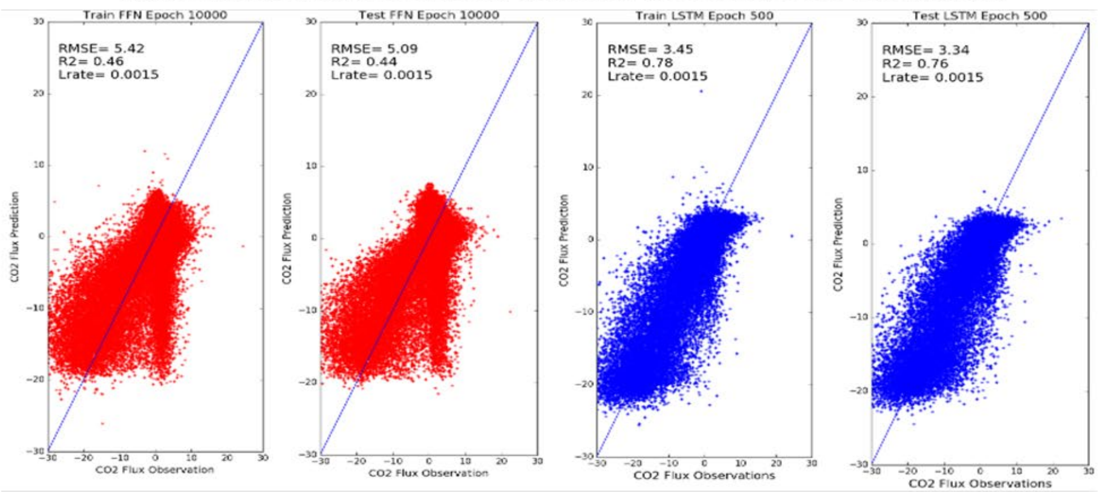
The input variables in figure 9c for experiments at AmeriFlux at Morgan-Monroe, Indiana (MMS) are: Shortwave radiation (incoming), Shortwave radiation (outgoing), Air temperature, and Precipitation. The data from Jan 1, 1999 to Dec 31, 2014 have been used. Half of the data was used for training results in RMSE= 4.71 with R^2 =0.59 and RMSE=3.50 with R^2=0.77 for FFNN and RNN respectively. The models tested using another half of MMS data show RMSE= 4.49 with R^2 =0.57 and RMSE=3.36 with R^2=0.76 for FFNN and RNN, respectively. The CO_2 concentration *was absent* in the input variables compared to the previous experiments in the figure 9b. The RMSE errors for MMS with and without CO_2 for RNN are 3.34 and 3.36, respectively, to the results for table 5.

FFNN and LSTM produce their lowest CO_2 flux scores of 2.30 and 1.75 resp., at high latitudes while stations near or at the equator produce the largest RMSE errors of 4.82 and 4.08. This variability in RMSE is a result of larger seasonal variations in temperature and net radiation at higher latitudes and less variability at low latitudes.

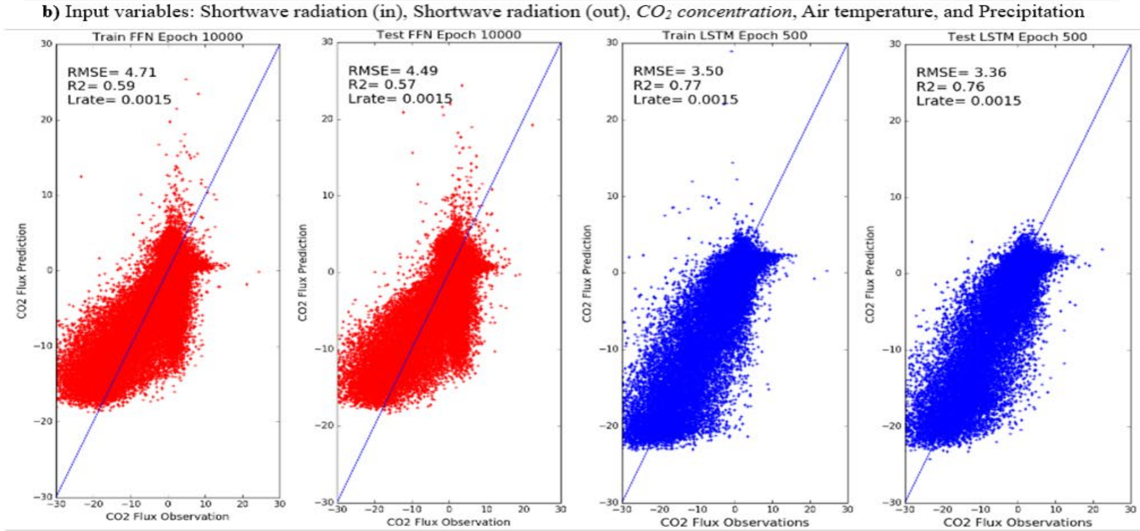
In conclusion, our experiments show that the RNN model produces on average about 28% improvement in RMSE with better R^2 for both CO_2 fluxes (for ARM/AmeriFlux stations) and NEE (Fluxnet2015 stations) prediction over FFNN model. It was also found that using heat fluxes as input variables produce improved CO_2 flux and NEE predictions. The most surprising result obtained from the machine learning studies was the omission of CO_2 concentrations as an input parameter produced comparable inferences of CO_2 fluxes. We refer to this property as the metaphysical nature of neural nets.



a) Input variables: Net radiation, latent heat, sensible heat, and heat flux, air and soil temperatures.



b) Input variables: Shortwave radiation (in), Shortwave radiation (out), *CO_2 concentration*, Air temperature, and Precipitation



c) Input variables: Shortwave radiation (in), Shortwave radiation (out), Air temperature, Precipitation **Excluding CO_2 concentration**

Figure 9 Comparisons of FFNN and RNN, LSTM using AmeriFlux tower data at Morgan-Monroe, Indiana. Data from 1999 to 2014 was used.

5. Conclusions

We have presented two deep learning models, FFNN and RNN (LSTM) for prediction of CO_2 flux at annual scales. We have used micro meteorological data from ARM, AmeriFlux, and Fluxnet2015 sites with different vegetation and regional climates distributed globally. In addition, we have examined the influence of meteorological variables as input to these neural nets as well as the sensitivity to different C3 and C4 crop types. We also presented CO_2 flux or NEE machine learning experiments from more than 24 distributed stations using hourly data for multiple years for training, testing and validity tests. The experiments show for the global distribution of station data with more than a decade of observations that the LSTM model produces ~ 22% to 28% improved predictions compared with FFNN models.

We have learned from the machine learning approach in these station experiments that (i) supervised learning from long historical records are critical for improved inferences of NEE and CO2 flux, (ii) the temporal aspect of recurrent neural nets can yield significant improvements for those applications where the supervised data are time dependent such as occurring in the station data sets and (iii) there is still considerable room for improving both neural net algorithms by utilizing available tools that can experiment with the full set of hyper parameters for these algorithms. Some of the parameters are number of layers, epochs, number of time steps in LSTM, sensitivity to input variables and more careful analysis of station data like gap filling, use of fuzzy time steps to account for seasons, daytime hours, vegetation cover maps, nighttime stratification of unexpected low eddy fluxes.

A most surprising machine learning result for inferring CO_2 flux was the lack of sensitivity on CO_2 concentration itself as an input variable, when training with available CO_2 flux data. Whether this holds away from stations will require further experiments. We assume that training the data with CO_2 flux compensates for the lack of CO_2 concentration as an input variable. In conclusion, these techniques can be applied to tower stations to infer and evaluate the present state of net ecosystem exchange. In the future, we plan to use these models to predict CO_2 flux from remote sensing observations of OCO-2.

Acknowledgments

This study was funded by the NASA grant number NNH16ZDA001N-AIST16-0091. Special thanks are due to IBM for providing two Minsky nodes with each node having dual IBM Power8+CPUs and4 Nvidia P100 GPUs. We also acknowledge the support of the NSF supported Center for Hybrid Multicore Productivity Research at UMBC for providing access to this High Performance Computing resource and supporting staff. Authors acknowledge Dr. Pierre Gentine at the Dept of Earth and Environmental Engineering, Columbia University for providing comments. The data underlying the analyses was downloaded from the DOE ARM station athttps://adc.arm.gov, AmeriFlux data at http://ameriflux.lbl.gov/ and http://fluxnet.fluxdata.org/.

References

Le Quéré, C. et al. Global carbon budget (2016). *Earth Syst. Sci. Data* 8, 605–649 (2016)

Li, W. *et al.* (2016) Reducing uncertainties in decadal variability of the global carbon budget with multiple datasets. *Proc. Natl Acad. Sci. USA* 113, 13104–13108

H. D. Graven et al., Enhanced Seasonal Exchange of CO2 by Northern Ecosystems Since 1960. (2013) Science 341, 1085; DOI 10.1126/ science.1239207

Piao, S., Liu, Z., Wang, Y., Ciais, P., Yao, Y., Peng, S., Chevallier, F., Friedlingstein, P., Janssens, I.A., Peñuelas, J. and Sitch, S., (2017). On the causes of trends in the seasonal amplitude of atmospheric CO_2.*Global Change Biology*.

Bousquet, P., Peylin, P., Ciais, P., Le Quéré, C., Friedlingstein, P. and Tans, P.P., (2000). Regional changes in carbon dioxide fluxes of land and oceans since 1980.*Science*, *290*(5495), pp.1342-1346.

Bala, G., (2013). Digesting 400 ppm for global mean CO_2 concentration. *Current science*, *104*(11), pp.1471-1472.

Keenan, T.F., Baker, I., Barr, A., Ciais, P., Davis, K., Dietze, M., Dragoni, D., Gough, C.M., Grant, R., Hollinger, D. and Hufkens, K., (2012). Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO_2 exchange. *Global Change Biology*, 18(6), pp.1971-1987.

Chen, B., Chen, J.M., Mo, G., Black, T.A. and Worthy, D.E., (2008). Comparison of regional carbon flux estimates from CO2 concentration measurements and remote sensing based footprint integration. *Global Biogeochemical Cycles*, 22(2).Vancouver

Reuter, M.; Buchwitz, M.; Hilker, M.; Heymann, J.; Bovensmann, H.; Burrows, J.P.; Houweling, S.; Liu, Y.Y.; Nassar, R.; Chevallier, F.; et al (2017). How much CO2 is taken up by the European Terrestrial Biosphere? *Bull. Am. Meteorol. Soc.* 98, 665–671

Ott, L. E., et al. (2015), Assessing the magnitude of CO_2 flux uncertainty in atmospheric CO_2 records using products from NASA's Carbon Monitoring Flux Pilot Project, *J. Geophys. Res. Atmos.*, 120, 734–765, doi:10.1002/2014JD022411.

Eldering, A., Wennberg, P. O., Crisp, D., Schimel, D. S., Gunson, M. R., Chatterjee, A.,Weir, B.(2017). The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes. *Science*, 358, eaam5745. https://doi.org/10.1126/science.aam5745

Schneider, T., Lan, S., Stuart, A., Teixeira, J.(2017) Earth System Modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *arXiv preprint arXiv*:1709.00037

Piao, S., Fang, J., Ciais, P., Peylin, P., Huang, Y., Sitch, S. and Wang, T., (2009).The carbon balance of terrestrial ecosystems in China.*Nature*, 458(7241), p.1009.

HeHonglin, YU Guirui, ZHANG Leiming, SUN Xiaomin& SU Wen (2006). Simulating CO_2 flux of three different ecosystems in ChinaFLUX based on artificial neural networks, *Science in China Series D: Earth Sciences*

Melesse, A.M. and Hanley, R.S., 2005. Artificial neural network application for multi-ecosystem carbon flux simulation.*EcologicalModelling*, *189*(3), pp.305-314.

Papale, D. and Valentini, R., (2003). A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. *Global Change Biology*, 9(4), pp.525-535.

Tramontana, G. et al, (2016) Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms *Biogeosciences*, 13, 4291–4313

Papale, D. and Valentini, R(2003). A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization, *Glob.Change Biol*., 9, 525–535, doi:10.1046/j.1365-2486.2003.00609.x.

Baldocchi, D.2014 Measuring fluxes of trace gases and energy between ecosystems and the atmosphere – the state and future of the eddy covariance method, *Glob. Change Biol*., 20, 3600–3609 doi:10.1111/gcb.12649

Hinton, G. E., Osindero, S. and Teh, Y., (2006) A fast learning algorithm for deep belief nets. *Neural Computation* 18:1527-1554

Bengio, Y., Pascal Lamblin, Dan Popovici and Hugo Larochelle, (2006) Greedy Layer-Wise Training of Deep Networks ,*NIPS*, pp. 153-160

Ranzato, M.A., Christopher Poultney, Sumit Chopra and YannLeCun (2006) Efficient Learning of Sparse Representations with an Energy-Based Model, *NIPS*

Krizhevsky, I. Sutskever, and G. Hinton (2012) Imagenet classification with deep convolutional neural networks.*In NIPS*.

R. K. Srivastava, K. Greff, and J. Schmidhuber. (2015) Training very deep networks. 1507.06228

Bebis, G. and Georgiopoulos, M., (1994). Feed-forward neural networks. *IEEE Potentials*, *13*(4), pp.27-31.

Hochreiter S., and J. Schmidhuber. (1997) Long short-term memory. *Neural computation*, 9(8):1735–1780.

Baldi, Pierre and Sadowski, Peter. (2014) The dropout learning algorithm. *Artificial intelligence*, 210:78–122.

Sutskever, Ilya, Vinyals, Oriol, and Le, Quoc VV. (2014) Sequence to sequence learning with neural networks. *In NIPS*, pp. 3104– 3112

Buduma,N.(2015) Fundamentals of Deep Learning Designing Next-Generation Machine Intelligence Algorithms, *O'Reilly Media*

LeCun, Y., Y. Bengio, and G. Hinton. (2015) Deep learning.*Nature*, 521:436–444,

Duchi, J.C., E. Hazan, and Y. Singer (2011) Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*

U.S. Department of Energy, ARM Climate Research Facility Atmospheric Radiation Measurement Climate Research Facility., https://www.arm.gov/ ,.

H. D. Graven et al., Science 341, 1085 (2013); 10.1126/ *science*.1239207

Fung, I. (2013). *A Hyperventilating Biosphere.Science 341: 1075*

Jung, M. 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.*

Xiao, J. et al., 2011. Assessing net ecosystem carbon exchange of U.S. terrestrial ecosystems by integrating eddy covariance flux measurements and satellite observations. *Agricultural and Forest Meteorology, 151(1): 60-69.*

Medlyn, B.E., Robinson, A.P., Clement, R. and McMurtrie, R.E., 2005. On the validation of models of forest CO2 exchange using eddy covariance data: some perils and pitfalls. *Tree Physiology, 25(7): 839-857.*