

# Deep Learning Models for Predicting CO<sub>2</sub> Flux Employing Multivariate Time Series

Phuong Nguyen

Department Of Computer Science and Electrical Engineering  
University Of Maryland Baltimore County  
phuong3@umbc.edu

Milton Halem

Department Of Computer Science and Electrical Engineering  
University Of Maryland Baltimore County  
halem@umbc.edu

## ABSTRACT

Understanding temporal changes in regional sources and sinks of CO<sub>2</sub> (i.e. determined by CO<sub>2</sub> flux) is a challenging measurement that is needed for assessments of climate change. Long-term measurements of CO<sub>2</sub> flux are available for more than a decade from about 100 tower stations distributed globally containing instruments for the measurements of CO<sub>2</sub> flux along with an array of instruments for measuring meteorological variables. We explore the use of two deep machine learning models, a Feed Forward Neural Net (FFNN) and a Recurrent Neural net (RNN), to determine whether their accuracies are sufficient for inferring CO<sub>2</sub> flux over land. The RNN we employ is the long short-term memory (LSTM) and the FFNN is a standard feed forward backward propagation neural net approach. We evaluate both model predictions of CO<sub>2</sub> flux using long data records from distributed tower data sites from Fluxnet stations archived at the Oakridge Research National Lab. Our results indicate that the deep learning model employing LSTM provides significantly more accurate (~22% improvement) predictions of CO<sub>2</sub> flux than FFNN and can provide regional estimates. We suggest that these models be used as tools for data gap filling and temporal pattern analysis of CO<sub>2</sub> Flux and its correlated variables.

## CCS CONCEPTS

Computing methodologies---Modeling and simulation---Simulation types and techniques

## KEYWORDS

LSTM, Recurrent neural network, deep learning, feed forward neural network, CO<sub>2</sub> flux, multivariate, spatiotemporal, time series forecasting.

## 1 Introduction

Plants are responsible for absorbing ~30% of atmospheric CO<sub>2</sub> through the photosynthesis processes. The main factors affecting

photosynthesis rate are sunlight, CO<sub>2</sub> concentrations, temperature, moisture and surface winds [1]. The rate of change of CO<sub>2</sub> can be used to quantify photosynthesis activity or plant growth. In order to understand such biosphere-atmosphere interactions, spatial and temporal gradients of CO<sub>2</sub> flux need to be highly accurate to quantify the effects of such regional changes such as deforestation, wild fires. Measurements of CO<sub>2</sub> flux data are obtained from Eddy Covariance instruments installed on Flux tower stations and are known to show uncertainties due to turbulent meteorological fluctuations [2], [3]. Are the accuracies of deep learning models sufficient to contribute to improved assessments of regional atmospheric variances of CO<sub>2</sub> sources and sinks? An investigation of how the biosphere reacts to changes in atmospheric CO<sub>2</sub> is essential to our understanding of potential climate-vegetation feedbacks [4]-[8].

Artificial Neural Networks (ANNs) are often better candidates for modeling non-linear processes based on data driven input than other statistical methods. Feed Forward Backward Propagation Neural Nets (FFNN) have been used for the prediction of CO<sub>2</sub> fluxes using atmospheric CO<sub>2</sub> and other variables [9]. They showed that a FFNN is a promising technique which can be used to predict CO<sub>2</sub> flux in place of the eddy covariance method at a single point [10]. However, the study was limited to a very short period of data observations as well as the number of flux towers. In addition, there have been prior efforts to map fluxes across North America with machine learning regression tree models by [2] globally. In [11] and [12], authors proposed a methodology involving an FFNN to provide spatial (1 km×1 km) and temporal (weekly) estimates of carbon fluxes of European forests at continental scale.

Deep Learning (DL) neural networks, an evolution of ANNs, couple with new advances in computer technologies attempt to better 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 [13]. Because of advances in computer technologies, one can now model many more layers of virtual neurons than ever before [13]-[15]. With deep layer architectures exploiting computational acceleration, these 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 models, which are able to

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MileTS '19, August 5th, 2019, Anchorage, Alaska, USA

© 2019 Copyright held by the owner/author(s). 978-1-4503-0000-0/18/06...\$15.00

successfully learn time series data with long range temporal dependencies or machine translation for language modeling [16], [17]. Extending DL models for applications such as Earth and space science as well as other science disciplines will require both conceptual breakthroughs and further advances in processing power.

In this study, we train the DL models for both the FFNN and the RNN with an LSTM model to learn the weights for predicting the CO<sub>2</sub> flux from the long term measurements at tower stations of CO<sub>2</sub> concentration, humidity, pressure, temperature, wind speed etc. representing data from different ecosystem sites. The motivation for using LSTM is to utilize the temporal dependencies or patterns within the long-term measurements of the multivariate variables to improve the regional accuracies of the station data.

The organization of this paper is as following. Section 2 provides the description of input datasets and our methodologies. The experimental results and analysis are presented in section 3. Section 4 presents conclusions and recommendations.

## 2 Data description and Methodology

We use the AmeriFlux, and Fluxnet2015 tower data located at different sites (21 ecosystems) in this study. Figure 1 shows a map of the selected stations distributed at different latitudes, and having different types of environmental ecosystems (for example forest, plains, fire sites). The half an hour and hourly data have been used. Figure 2 shows the variables for the AmeriFlux tower data at Morgan-Monroe, Indiana (MMS) site from Jan 1, 1999 to Dec 31, 2013 labeled Temperature (deg C), CO<sub>2</sub> Flux ( $\mu\text{molCO}_2 \text{ m}^{-2} \text{ s}^{-1}$ ), Net Radiation ( $\text{Wm}^{-2}$ ), Shortwave Radiation ( $\text{Wm}^{-2}$ ), Longwave Radiation ( $\text{Wm}^{-2}$ ), CO<sub>2</sub> ( $\mu\text{molCO}_2 \text{ mol}^{-1}$ ), Precipitation (mm). MMS's hourly data from Jan 1, 1999 to Dec 31, 2013 has about 152,000 observations (the detail of datasets and full list of variables can be found and downloaded in links provided in the acknowledgement section).

### 2.1 Data preprocessing

The preprocessing of the data consists of two parts: filtering row data and normalizing all input variable data. Row data with filled values (-9999) or having flags 'y' or out of normal CO<sub>2</sub> flux ranges are filtered out. Since the input variables have different value ranges, which vary significantly, all input data are normalized to scale between 0 and 1 to improve the time to reach convergence. The only output variable of the learning model is the quantity CO<sub>2</sub> flux and it is de-normalized into its original data range.

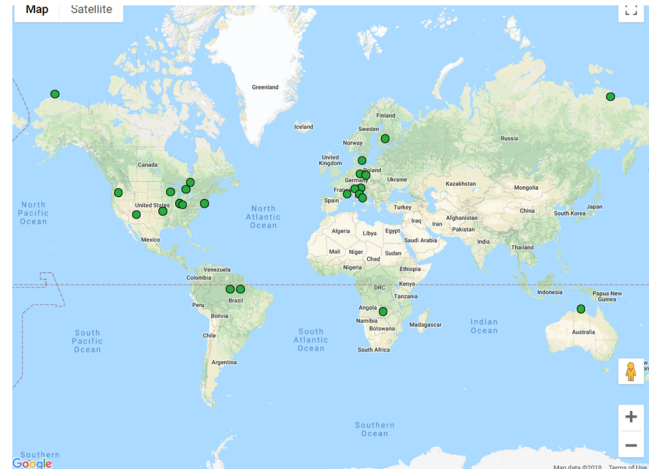


Figure 1 Map of selected tower sites for experiments

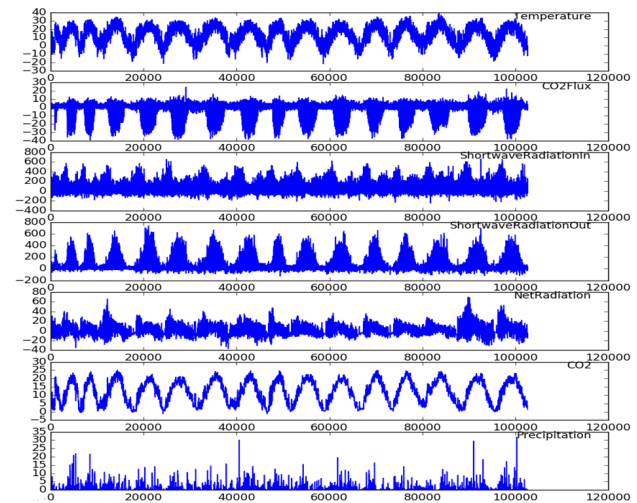


Figure 2 AmeriFlux hourly data at AmeriFlux at Morgan-Monroe, Indiana (MMS) from Jan 1, 1999 to Dec 31, 2013 (x axis shows index of hourly data by time, y axis are in above units of variables)

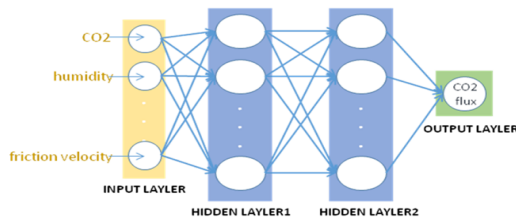
### 2.2 Selection of multivariate input variables

We used WEKA [18] (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. The output of the PCA was a set of components formed by means of linear combinations of the correlated attributes. The results from WEKA's PCA and Ranker attribute selector showed the following set of variables to have the highest rank. The first set uses heat fluxes including net radiance at the top of the atmosphere ( $\text{Wm}^{-2}$ ), latent heat ( $\text{Wm}^{-2}$ ), sensible heat ( $\text{Wm}^{-2}$ ), soil heat flux ( $\text{Wm}^{-2}$ ), 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 and they are derived from other variables such as radiation. In this case, we have incoming shortwave radiation ( $\text{Wm}^{-2}$ ), outgoing shortwave radiation ( $\text{Wm}^{-2}$ ), Carbon Dioxide (CO<sub>2</sub>) mole fraction in wet air ( $\mu\text{molCO}_2 \text{ mol}^{-1}$ ), Air temperature (deg C), and

Precipitation (kPa) have been used as input variables. The output of both FFNN and RNN models predict CO<sub>2</sub> flux (Carbon Dioxide (CO<sub>2</sub>) turbulent flux ( $\mu\text{molCO}_2 \text{ mol}^{-1}$ ) for dataset at AmeriFlux or CO<sub>2</sub> flux ( $\text{umolCO}_2 \text{ m}^{-2} \text{ s}^{-1}$ ) for the dataset at Fluxnet2015.

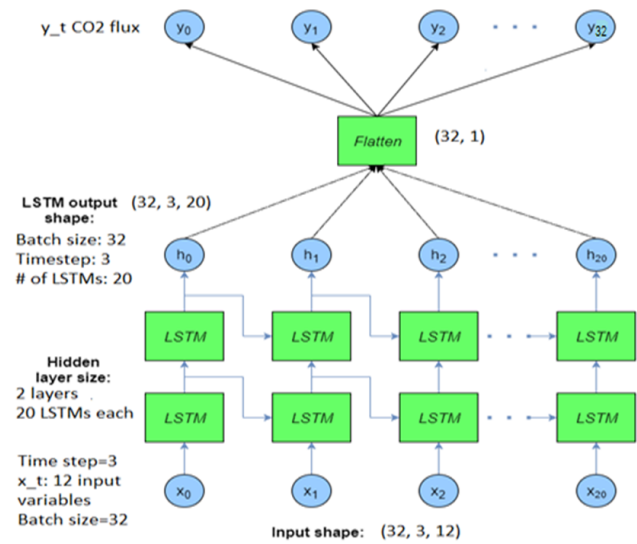
### 2.3 Building deep learning models for estimation of CO<sub>2</sub> Flux

Both FFNN and LSTM models use the same set of multivariate variables as listed in the previous session and estimate CO<sub>2</sub> flux as output. Figure 3 shows the layers of FFNN model and its input and output variables. 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 4). 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 neurons in the same layer, RNN enables the insertion of state operations or gates between layers, and neurons within the same layer have connections. Thus, when learning the model at a particular time using FFNN, there is no consideration of the historical data from prior times.



**Figure 3 FFNN model architecture, input: multiple variables (CO<sub>2</sub>, humidity, temperature etc.), hidden layers, output: CO<sub>2</sub> flux.**

The RNNs can use their internal state (memory in LSTM unit) to process sequences of inputs. For example, in this application, input measurements of CO<sub>2</sub> concentration, humidity, pressure, temperature, wind speed, etc. have been observed at every half hour forming a time sequence. Time dependency can be learned by LSTM. It can selectively learn when to forget things or remember by controlling the information flow through block or pass 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 multiple layer architects. This model capability is one of the advances made in deep machine learning models, which can be applied to solve difficult time sequence problems in machine learning. It has been successfully implemented [17]. The detail of LSTM units and the architecture of this model are referenced in [16], [17].



**Figure 4 Recurrent neural network with LSTM model**

Figure 4 shows the following structure of the RNN with the LSTM model that we employ. Each observed time step consists of an observation of multiple 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 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 a single output, which is the prediction of CO<sub>2</sub> flux. The DropOut technique is used for training to improve the performance of the neural network. During training, it drops x% of units in each layer [19]. All de-activate units would not participate in the training propagation. This technique will block the error values in each layer from passing to the output level. The dropout technique only applies on the training phase. Advanced gradient algorithm such as SGD and AdaGrad [15], [20] in Tensor flow have been used to train the models.

## 3 Experiment results and discussion

This section presents the experiments for the comparison of FFNN and RNN/LSTM models for predicting CO<sub>2</sub> flux from Ameriflux station data and from the Fluxnet2015 stations. We show below that Recurrent Neural Networks (RNN) with the Long Short Term Memory (LSTM) improvements in the prediction of CO<sub>2</sub> flux by ~22% interms of Root Mean Square Error and R<sup>2</sup> (correlation coefficient) scores.

### 3.1 Comparison of FFNN and LSTM models

Table 1 shows the comparisons between FFNN and RNN models using RMSE and R<sup>2</sup> correlation metrics for a selected set of tower sites (shown in figure 1, map of tower stations). These experiments use net radiation, latent heat, sensible heat, and soil

heat flux, air and soil temperatures as input variables and predict CO<sub>2</sub> flux for AmeriFlux stations and the incoming shortwave radiation, outgoing shortwave radiation, (CO<sub>2</sub>) mole fraction in wet air, Air temperature and Pressure (kPa) at 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 validate the errors to prevent overtraining; and (iii) 10% of dataset is used as test dataset to assess the network's performance with 'new' data. We used 10 years of hourly or half an hour observations. Eight (8) years of data has been used for training, 1 year for evaluation and the last year of data has been used for testing (new prediction or forecasting).

The FFNN model used in this section was configured to run using either 10 layers or 20 layers, with 20 neurons each layer. RNN models used 2 layers each have 20 LSTM units. Both models were trained using the back propagation gradient descent algorithm with 'elu' activation. The mean improvement for the average of the stations employed 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 1 Comparisons between FFNN and LSTM (RNN)**

| Site                   | RMSE     |      |                 |      | R2 Correlation |      |      |      |
|------------------------|----------|------|-----------------|------|----------------|------|------|------|
|                        | Training |      | Test Validation |      | Training       |      | Test |      |
|                        | 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 |
| FLX-IT-Ca3             | 2.03     | 2.14 | 0.91            | 0.87 | 1.22           | 1.08 | 0.83 | 0.81 |
| FLX-DK-Sor             | 4.53     | 3.01 | 4.51            | 3.23 | 5.06           | 3.03 | 0.68 | 0.86 |
| FLX-FI-Hyy             | 2.19     | 1.55 | 2.18            | 1.66 | 2.93           | 2.39 | 0.73 | 0.87 |
| FLX-CA-Gro             | 3.04     | 2.39 | 2.74            | 2.08 | 2.72           | 2.25 | 0.61 | 0.76 |
| FLX-IT-Tor             | 2.68     | 1.66 | 3.01            | 1.85 | 3.69           | 2.52 | 0.64 | 0.86 |
| FLX-DE-Obe             | 3.37     | 2.82 | 3.44            | 3.2  | 3.8            | 3.57 | 0.76 | 0.83 |
| FLX-IT-Mbo             | 4.06     | 2.87 | 4.7             | 3.38 | 4.04           | 2.88 | 0.52 | 0.76 |
| FLX-FR-Pue             | 2.21     | 1.69 | 2.06            | 1.69 | 1.84           | 1.88 | 0.73 | 0.84 |
| FLX-AU-DaP             | 4.14     | 2.98 | 4.32            | 3.18 | 4.13           | 3    | 0.58 | 0.78 |
| FLX-DE-Hai             | 4.67     | 2.8  | 5.07            | 3.36 | 4.83           | 3.04 | 0.64 | 0.87 |
| FLX-BR-Sa3             | 5.99     | 5.41 | 6.21            | 6.21 | 6.43           | 5.94 | 0.71 | 0.76 |
| Ameri-US-Bo1           | 5.33     | 4.62 | 5.84            | 4.81 | 6.15           | 5.58 | 0.39 | 0.54 |
| Ameri-US-MMS           | 4.28     | 2.93 | 4.31            | 2.94 | 4.62           | 3.07 | 0.63 | 0.83 |
| Ameri-US-Umb           | 4.23     | 2.93 | 4.87            | 3.56 | 4.47           | 3.12 | 0.55 | 0.78 |
| FLX-US-Ced             | 3.99     | 2.73 | 4.41            | 2.69 | 3.65           | 2.79 | 0.58 | 0.8  |
| FLX-US-Me2             | 3.12     | 2.76 | 3.72            | 3    | 4.14           | 3.56 | 0.69 | 0.76 |
| FLX-ZM-Mon             | 3.07     | 2.79 | 2.96            | 2.74 | 2.1            | 3.38 | 0.74 | 0.78 |
| FLX-RU-Cok             | 1.49     | 1.34 | 1.55            | 1.29 | 1.73           | 1.66 | 0.57 | 0.66 |
| Ameri-US-Ro1*          | 5.33     | 2.88 | 5.84            |      | 6.15           | 3.22 | 0.39 | 0.81 |
| Ameri-US-Pon*          | 4.01     | 2.46 |                 |      | 3.91           | 3.01 | 0.68 | 0.86 |
| Avg of RNN Improvement |          | 24%  |                 | 23%  |                | 18%  |      | 27%  |
|                        |          |      |                 |      |                |      | 20%  | 20%  |

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

### 3.2 Anomaly correlation analysis

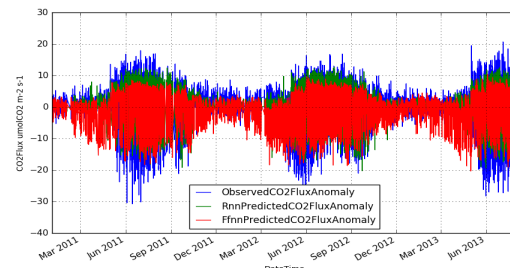
The Morgan Monroe State Forest station in Indiana, had the longest continuous record available to the end of 2013, and was selected for investigating the performance of Neural Nets using anomaly correlation metrics. Figure 5 shows a plot of the anomaly for the observed CO<sub>2</sub> flux for years 2011-to 2013, where the mean of the monthly data for the years 1999-2010 (training data) are removed.

Superimposed are the predicted CO<sub>2</sub> 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 from this image. Thus, in table 2, 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 for the 21 station sites (table 1).

**Table 2 Anomaly correlations for FFNN and LSTM prediction vs. observation**

| Site         | Anomaly correlation |      |                 |      |      |      |
|--------------|---------------------|------|-----------------|------|------|------|
|              | Train               |      | Test Validation |      | Test |      |
|              | 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 5 Anomaly correlation comparisons between FFNN and RNN, LSTM**

### 4 Conclusions

We have presented two deep learning models, FFNN and RNN (LSTM) for prediction of CO<sub>2</sub> flux at annual scales. We have used micrometeorological data from AmeriFlux, and Fluxnet2015 sites with different vegetation and regional climates distributed globally. We presented CO<sub>2</sub> flux machine learning experiments from more than 21 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 about 22% improved predictions compared with FFNN models. We suggest that these models can be used as a tool for data gap filling or analysis of temporal patterns and its correlated variables for predicting CO<sub>2</sub> flux.

### ACKNOWLEDGMENTS

This study was funded by the NASA grant number NNH16ZDA001N-AIST16-0091. We also acknowledge the support of NSF Center for Accelerated Real Time Analytics – UMBC. The data underlying the analyses was downloaded from the DOE ARM station at <https://adc.arm.gov>, AmeriFlux data at <http://ameriflux.lbl.gov/> and <http://fluxnet.fluxdata.org/>.

## REFERENCES

- [1] HeHonglin, YU Guirui, ZHANG Leiming, SUN Xiaomin & SU Wen (2006). Simulating CO<sub>2</sub> flux of three different ecosystems in ChinaFLUX based on artificial neural networks, *Science in China Series D: Earth Sciences*
- [2] 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.
- [3] Medlyn, B.E., Robinson, A.P., Clement, R. and McMurtrie, R.E., 2005. On the validation of models of forest CO<sub>2</sub> exchange using eddy covariance data: some perils and pitfalls. *Tree Physiology*, 25(7): 839-857.
- [4] 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
- [5] H. D. Graven et al., Enhanced Seasonal Exchange of CO<sub>2</sub> by Northern Ecosystems Since 1960. (2013) *Science* 341, 1085; DOI 10.1126/science.1239207
- [6] 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<sub>2</sub>. *Global Change Biology*.
- [7] Bousquet, P., Peylin, P., Ciais, P., Le Quééré, 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.
- [8] Bala, G., (2013). Digesting 400 ppm for global mean CO<sub>2</sub> concentration. *Current science*, 104(11), pp.1471-1472.
- [9] Melesse, A.M. and Hanley, R.S., 2005. Artificial neural network application for multi-ecosystem carbon flux simulation. *Ecological Modelling*, 189(3), pp.305-314.
- [10] 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
- [11] 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.
- [12] 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.
- [13] Krizhevsky, I., Sutskever, and G. Hinton (2012) Imagenet classification with deep convolutional neural networks. In NIPS.
- [14] LeCun, Y., Y. Bengio, and G. Hinton. (2015) Deep learning. *Nature*, 521:436–444.
- [15] Duchi, J.C., E. Hazan, and Y. Singer (2011) Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*
- [16] Hochreiter S., and J. Schmidhuber. (1997) Long short-term memory. *Neural computation*, 9(8):1735–1780.
- [17] Sutskever, Ilya, Vinyals, Oriol, and Le, Quoc VV. (2014) Sequence to sequence learning with neural networks. In NIPS, pp. 3104– 3112
- [18] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*, Morgan Kaufmann, Fourth Edition, 2016.
- [19] Baldi, Pierre and Sadowski, Peter. (2014) The dropout learning algorithm. *Artificial intelligence*, 210:78–122.
- [20] Buduma, N. (2015) *Fundamentals of Deep Learning Designing Next-Generation Machine Intelligence Algorithms*, O'Reilly Media