LSTM processing of experimental time series with varied quality.

Krzysztof Podlaski[1][0000−0002−2883−0773], Michał Durka[1], Tomasz Gwizdałła[1][0000−0002−3981−6037], Alicja Miniak-Górecka[1][0000−0002−1860−8853], Krzysztof Fortuniak[2][0000−0001−7043−8751], and Włodzimierz Pawlak[2][0000−0002−9785−4787]

[1] Faculty of Physics and Applied Informatics, University of Lodz, Lodz, Poland, {krzysztof.podlaski,tomasz.gwizdalla,alicja.miniak}@uni.lodz.pl
[2] Faculty of Geographical Sciences, University of Lodz, Lodz, Poland, {krzysztof.fortuniak,wlodzimierz.pawlak}@geo.uni.lodz.pl

Abstract. Automatic processing and verification of data obtained in experiments have an essential role in modern science. In the paper, we discuss the assessment of data obtained in meteorological measurements conducted in Biebrza National Park in Poland. The data is essential for understanding the complex environmental processes, such as global warming. The measurements of CO2 flux brings a vast amount of data but suffer from drawbacks like high uncertainty. Part of the data has a high-level of credibility while, others are not reliable. The method of automatic evaluation of data with varied quality is proposed. We use LSTM networks with a weighted square mean error loss function. This approach allows incorporating the information on data reliability in the training process.

Keywords: lstm · neural networks · time series · prediction · co2 flux

1 Introduction

Verification and prediction of real data are important and challenging. Many experiments conducted every day produce new raw data that has to be assessed and analyzed. The vast amount of new data acquired every day puts high demand on automatic methods of data processing. Moreover, some values obtained in the experiments have higher credibility than the others. Therefore, machine learning systems have to be sensitive to such issues. In the literature, prediction and forecasting are connected with the use of neural networks [4, 14, 20]. Here we discuss Long Short Term Memory (LSTM) networks as the ones that suit well for the task of analysis and prediction of time series [8, 18].

In the paper, we consider the data acquired in a continuous meteorological experiment. Some records in the dataset have higher quality (are more reliable) than others. The prediction of time series with varied quality is similar to the classification of imbalanced datasets [19]. The methods used in classification cannot be directly applied to time series prediction but can be used as suggestions.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_44

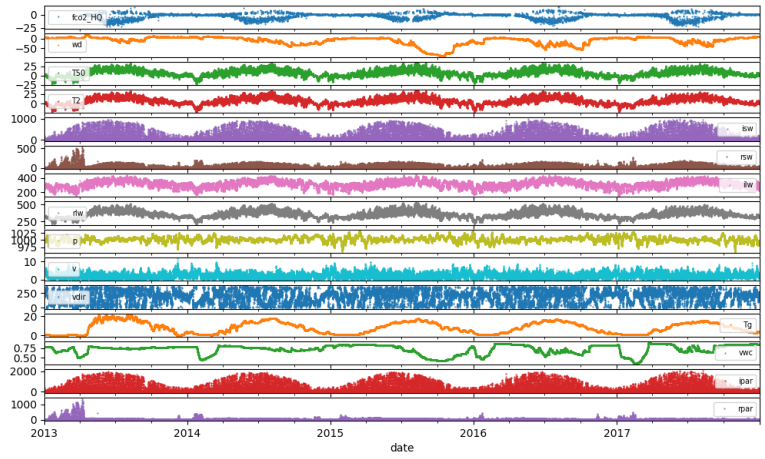
We propose a modification of the usual loss function in order to incorporate the issue of data credibility. The paper combines the computer and meteorological sciences. The importance of research conducted in the area of global warming is hard to neglect. The paper follows the idea of using computer science methods for a better future.

Understanding the complex environmental processes, such as global warming, belongs to the most challenging, both for cognitive reasons and social consequences. It is a truism to say that reliable data are of fundamental meaning for this purpose. The standard geophysical data provided by national and international services (such as WMO) are generally continuous and high-quality but do not include all parameters needed. Therefore, they must be supplemented by the results of experiments that often use very sophisticated measurement techniques. These experiments allow us to gain a vast amount of unique data but may suffer from various drawbacks, including a large number of unreliable or missing values. The missing values introduce high uncertainty when the data is analyzed from a long-term perspective and must be replaced by the most likely one. The short gaps in data sets might be filled up with simple interpolation methods, but the longer ones should be completed with adequately modeled values. This paper focuses on the time series of CO_2 flux collected in the wetlands of Biebrza National Park, northeastern Poland [9, 10]. The measured CO_2 flux represents the net exchange of this greenhouse gas between the surface and the atmosphere, which is vital to understand the role of such ecosystems in the global carbon cycle. The eddy-covariance method used, although considered to be the most adequate in measurements of surface-atmosphere exchange in the whole ecosystem scale [2, 3], results in a large number of non-randomly distributed missing data due to required quality control and sensitivity to unfavorable weather conditions. To evaluate the total uptake or emission of CO_2 by the ecosystem in an annual or multi-year perspective, these gaps must be filled up, taking into account the sensitivity of the CO_2 flux to changing hydrometeorological conditions [2]. Otherwise, the results could be biased towards the flux recorded in fine weather conditions. Various approaches to the gap-filling procedure have been suggested [2, 10, 15], [7, 16], but the problem has still not been fully standardized by the eddy-covariance community. Therefore, we propose using methods known from automatic prediction and verification of time series of CO_2 flux data, which can improve gap-filling methods and consequently reduce uncertainty in assessing the carbon balance in terrestrial ecosystems.

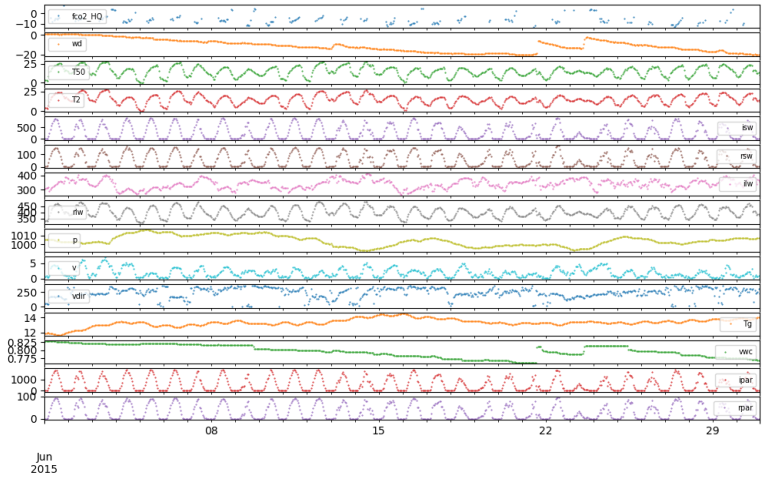
2 The dataset

In the paper, we use raw, real data obtained during measurements conducted at the Biebrza National Park's wetlands in northeastern Poland. The measurement site was located in the middle basin of Biebrza valley (53°35′30.8″N, 22°53′32.4″E, 109m a.s.l.) in a large flat area covered by patches of reeds, high sedges, and rushes, very typical of wetlands of the Biebrza National Park. The measurement period used in this analysis covers the years 2013-2017. The open-

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_44



(a) Years 2013-17



(b) June 2015

Fig. 1: Parameters measured during geographical experiments in Biebrza National park.

path eddy-covariance system consisted of fast-response sensors: CO_2/H_2O gas analyzer (Li7500, Li-cor Inc., USA) and a sonic anemometer (81000, R. M. Young, USA) with the middle of the path at 3.7 m above ground level [9], [10]. The CO_2 fluxes were calculated for 1-hour block averaging (and auxiliary on

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_44

a 5-min basis) with the aid of EddyPro 6.0 (Li-cor Inc., USA) software to ensure compatibility with other studies. In the flux calculations, covariance was maximized within a ±2s window, a double rotation of the natural wind coordinates was performed, the sonic temperature was humidity corrected, the Webb-Pearman-Leuning correction was applied, and spectral losses were corrected. Complementary fourteen hydrometeorological parameters were collected simultaneously using slow response sensors: water table depth (wd), temperature at 2 and 0.5 meters above the ground (T50, T2), the temperature of the ground (Tg), atmospheric pressure (p), wind speed (v), wind direction (vdir), volumetric water content (vwc), incident shortwave (visible) radiation (isw), reflected shortwave (visible) radiation (rsw), incident/reflected longwave radiation (ilw, rlw), incident/reflected photosynthetically active radiation (ipar, rpar). Most of these parameters show a seasonal behavior and a dependence on the time of the day (see Figure 1). Moreover, the Figure 1b shows that high-quality data on CO_2 flux is very sparse in June 2015 and in total is available only in about 25% of the records in the dataset. In the paper, we assume that CO_2 flux (fco2) depends on the time of the measurement and the rest of the fourteen parameters.

A detailed description of the acquired data's measurement equipment and postprocessing can be found in [9]. The postprocessing data quality included three stationarity tests and the friction velocity threshold criterion. In the result, the data was divided into three classes that define the quality of the result, High-quality (HQ), Medium-quality (MQ), and Low-quality (LQ). These groups describe the credibility of the raw data. The HQ data (flagged by EddyPro and accepted by all three additional tests) passed more rigorous criteria than usually used in eddy-covariance data analysis. The MQ data (accepted by at least one of the three additional tests) are similar to those usually analyzed. The remaining data were classified as LQ, and they may be burdened with substantial measurement errors.

3 Neural network approach

The verification and prediction of real data are two areas of data processing. These approaches are similar in many ways. We build, train, and validate the prediction model with historical data. We can use the system to predict new values that will be obtained in the future. Usually, the real result has higher credibility and any differences are used to assess the model. In that way, we verify the model with historical data. If we do not observe any external effects that should result in model modification, we can assume some level of correctness of the model. Such an approved prediction models can be used for verification purposes. The differences between the predictions made by the approved model and the data obtained from the measurements can suggest low-quality measurements. Researchers use artificial neural networks for gap-filling in meteorological for some time [6,16,17]. First, we have to build the model and validate it against the historical dataset.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_44

All changes in time in a meteorological system preserve continuity. Therefore, we can treat the measurements as a time series. The changes between two consecutive measurements should express a level of smoothness. In the paper, we plan to follow the presented approach and assess the measurement data using Long Short Term Memory (LSTM) neural network [11, 13].

3.1 Neural networks

The neural network is a technique based on the neural structures of living animals. A feed-forward artificial network is built from neurons grouped in layers. Each layer receives signals only from the preceding layer. The information flows sequentially from layer to layer. The first layer is usually called the Input Layer. The last layer is denoted as the Output Layer. All layers in between are called hidden ones. Each neuron receives information from neurons in the preceding layer. Each input signal is assigned with a weight. A neuron counts a weighted sum of inputs, adds a bias b, and applies an appropriate activation function. All neurons in a layer have the same activation function, and all of them have individual weights and bias. The networks can be trained in a supervised approach using an optimization method (for example, SGD, Adam, AdaGrad, RMSProp) [12]. The training goal is to find a set of network parameters that minimize a given loss function. Usually, the loss function measures the distance between an expected answer and the one obtained from the network.

LSTM neural networks proved to be adequate for the prediction of time-series data. In short, we can perceive the LSTM cell as a subsystem with a set of fully connected sub-layers and gates. Input signal contains a time series of signals that are fed to four fully connected sub-layers one by one. The sub-layers' signals are joined with three types of gates (forget, input, and output). The so-called long-term state represents the memory of the cell. Forget gate controls how long-term state should incorporate a new signal from a given series. An activation function accompanied each LSTM cell. The LSTM cell processes inputs row by row using the embedded memory feature. Each row contains a single event in a time series processed. Therefore, the cell size depends only on the size of the information stored in one row.

In the classical approach, all time series provided during training have the same credibility. Therefore, all are treated in the same way. We know that the experiment's data has a varied quality and needs to be treated in a special way. In classification approaches, a few solutions to an imbalance in class representation are proposed [19]. For example, each training input is assigned to a weight used during training. We can enforce the underrepresented class signals to appear more often in the training process. In the prediction approach, we can either change the ratio of occurrences of the high-quality signals during training or change the loss function to take the quality into account. In the paper, we use the latter approach and define the weighted mean square error as a loss function. Similar solutions to our approach can be found in [5].

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_44

3.2 Input data

The dataset contains fourteen meteorological parameters measured simultaneously, the date and time of the measurements, and registered CO_2 flux. We decide not to limit ourselves to a subset of experimental data and assume that the CO_2 flux depends on all fourteen meteorological parameters (described in the previous section). It should be mentioned that the authors of the paper [9] limit themselves to nine measured parameters. Additionally, as neural networks treat better signals in the range $(-1, 1)$, all parameters are normalized to this range. We prepare input data representing a set of $n + 1$ consecutive occurrences of measured parameters for our experiments. That means the input of the network has a shape of an $(n+1) \times m$ matrix, where m denotes the size of a single set of simultaneous meteorological parameters accompanied by the measurement time, n denotes the number of consecutive predecessors of the actual measurement. We can interpret $n + 1$ as the size of the time window used to predict the value of CO_2 flux. At each time step t_i, we have a vector of input parameters $\mathbf{v}_i$. Each vector $\mathbf{v}_i$ contains the year, month, day, hour, minute of measurement, as well as all fourteen meteorological parameters. In the preprocessing phase, we create a set of input matrices of the form:

$$I_i = \begin{pmatrix} \mathbf{v}_i \\ \mathbf{v}_{i-1} \\ \cdots \\ \mathbf{v}_{i-n} \end{pmatrix} \tag{1}$$

.

The input matrix I_i has a size $(n + 1) \times m$, with $n + 1$ rows of size m. Each row contains consecutive measurement vectors $\mathbf{v}_i$. The matrix I_i is accompanied by the expected output that contains the value of registered CO_2 flux at the moment t_i ($\mathrm{fco2}_i$). The values $\mathrm{fco2}_i$ are the results the model will be trained to predict.

3.3 LSTM model

LSTM model uses an LSTM layer as well as dense ones (fully connected) as presented in Figure 2. The model contains an LSTM layer of size s and one dense layer with $2s$ neurons. Both layers use tanh activation functions. The output layer is discussed in the next subsection.

3.4 Output layer and loss function

During training, the neural network use loss function and appropriate optimization method to derive modifications of the actual network parameters. In the dataset, we have three types of measurements connected to their quality. High-quality CO_2 flux values, the most reliable data (HQ) accounted for about 25% of all records, Medium-quality (MQ) around (35%), and the rest is a Low-quality

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_44

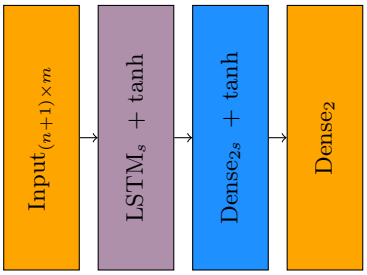


Fig. 2: LSTM neural network model, the subscripts denote size of the Layer

(LQ) (40%). Therefore, during the training, the HQ data need to have the highest priority. The designed network has two neurons in the output layer. The first one y_{fco} predicts the value of CO_2 flux the other y_q expected quality. The expected results, used for training and testing has the same structure:

$$\hat{y} = \left[\hat{y}_{\text{fco}}, \hat{y}_q\right], \qquad \text{where: } \hat{y}_q = \begin{cases} 0 & \text{for result in LQ} \\ 1 & \text{for result in MQ.} \\ 2 & \text{for result in HQ} \end{cases} \tag{2}$$

The results with high quality have higher reliability than the ones for other classes. Thus, we should enforce the network to take the data credibility into account during the training process. We propose to incorporate quality into the loss function. Our self-defined loss function has the form:

$$\text{loss}(\mathbf{y}, \hat{\mathbf{y}}) = \text{mean}\left((y_{\text{fco}} - \hat{y}_{\text{fco}})^2 \cdot 2^{\hat{y}_q}\right), \tag{3}$$

where $\mathbf{y}, \hat{\mathbf{y}}$ denote the batch of predicted and expected results, respectively. The presented loss function is a weighted version of the mean standard error function (WMSE), where HQ results have a weight of 4, the weight of MQ results equals 2, while LQ results have the weight factor set to 1. In this manner, we do not totally neglect the results with low quality but assign them with much lower priority.

4 The numerical experiment

We conducted our numerical experiments using the Tensorflow library [1] in python on a computer with AMD Ryzen 9 3900X processor and 32GB of RAM. In all models, we use the Stochastic Gradient Descent (SGD) optimization algorithm. The dataset was split into two disjoint sets, a training set consisting of 30 766 elements (80% of the total number of measurements) and a test set with 7 666 elements (20%). The ratio 80/20 between training and test sets is typically used in classification and prediction tasks that use neural networks.

The different models have been tested with a few selections of network layer the size parameter s, size of series time window n (Figure 2). The number of

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_44

Trainable parameters for model $\text{LSTM}_{n,s}$			
	$s = 10$	$s = 15$	$s = 20$
$n = 3$	1502	2702	4202
$n = 4$	1502	2702	4202
$n = 5$	1502	2702	4202
$n = 6$	1502	2702	4202

Table 1: The size of the LSTM models trained in numerical experiment.

trainable parameters for each tested model is presented in Table 1. As mentioned before, the number of trainable parameters in the LSTM layer depends on the row's size in an input signal. It does not depend on the number of rows in the input (the input represents a time series, and each row is a single event). Therefore, the increase of n, number of elements in a time series, has no impact on the number of trainable parameters. There are no rules on how to set network layer sizes for data to be processed. We have tried to keep our neural model small. The s parameter used had values $10, 15, 20$. The parameter n represents the size of the time series of measurements taken into account in a single input for our model. The measurements in the dataset in consideration were measured hourly. Thus, the parameter $n = 6$ means we use for prediction the actual results and values measured in the previous six hours. As most of the parameters have a visible daily routine (Table 1), we have decided to experiment with $n \in (2, 3, 4, 5, 6)$. This expresses the continuity and causality property of the measured parameters in a short time frame from two to six hours. For timeframes longer than six hours, many unmeasured parameters can have non-negligible impact.

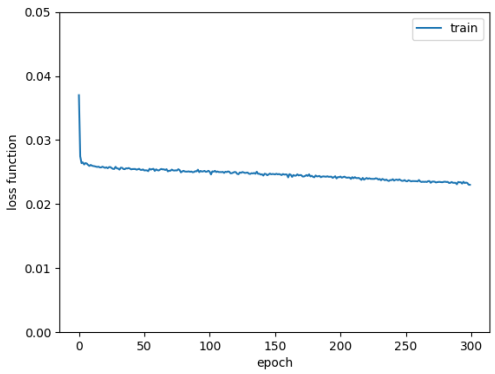
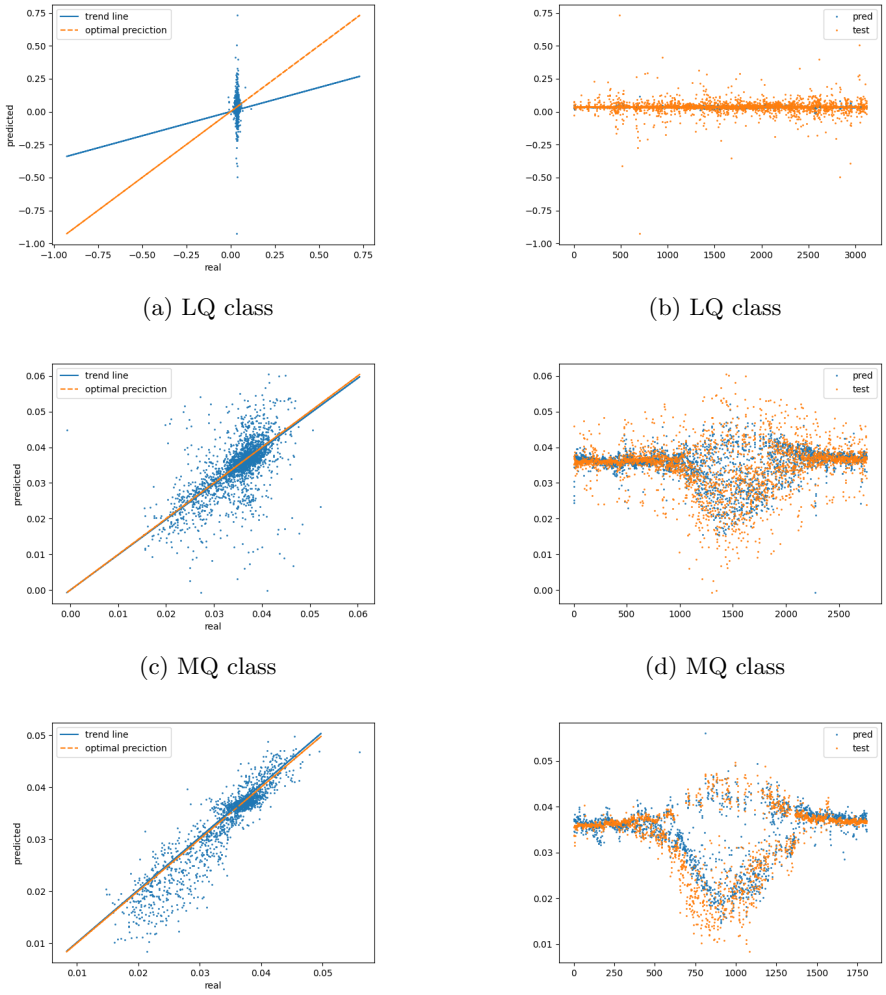


Fig. 3: Training process for LSTM network with $n = 6$ and $s = 20$.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_44

All trained networks achieve stability after around 50 epochs, and the loss function decreases at a stable slow pace. No vital difference has been observed when trained for 300 epochs. For example, we present the training process of the LSTM network with $n = 6$ and $s = 20$ in Figure 3. Therefore, for most of the cases, we have trained the models for 120 epochs.



(a) LQ class

(b) LQ class

(c) MQ class

(d) MQ class

(e) HQ Class

(f) HQ Class

Fig. 4: Prediction given from LSTM network with $s = 20$, $n = 6$ after 120 epochs. On both axes we present normalized predicted CO_2 flux compared to expected(real) CO_2 flux.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_44

Model	class	R^2	slope	MSE	RMSE
$\text{LSTM}_{n=3,s=10}$	HQ	0.798	0.926	$2.12 \cdot 10^{-5}$	$4.60 \cdot 10^{-3}$
	MQ	0.499	0.973	$3.58 \cdot 10^{-5}$	$5.98 \cdot 10^{-3}$
$\text{LSTM}_{n=3,s=15}$	HQ	0.819	1.006	$2.06 \cdot 10^{-5}$	$4.54 \cdot 10^{-3}$
	MQ	0.529	1.011	$3.96 \cdot 10^{-5}$	$6.29 \cdot 10^{-3}$
$\text{LSTM}_{n=3,s=20}$	HQ	0.794	0.978	$2.10 \cdot 10^{-5}$	$4.58 \cdot 10^{-3}$
	MQ	0.542	0.941	$3.52 \cdot 10^{-5}$	$5.94 \cdot 10^{-3}$
$\text{LSTM}_{n=4,s=10}$	HQ	0.824	1.006	$1.93 \cdot 10^{-5}$	$4.43 \cdot 10^{-3}$
	MQ	0.562	0.972	$3.24 \cdot 10^{-5}$	$5.69 \cdot 10^{-3}$
$\text{LSTM}_{n=4,s=15}$	HQ	0.857	0.978	$1.81 \cdot 10^{-5}$	$4.26 \cdot 10^{-3}$
	MQ	0.536	0.975	$3.37 \cdot 10^{-5}$	$5.81 \cdot 10^{-3}$
$\text{LSTM}_{n=4,s=20}$	HQ	0.848	1.006	$1.59 \cdot 10^{-5}$	$3.99 \cdot 10^{-3}$
	MQ	0.584	1.002	$3.29 \cdot 10^{-5}$	$5.74 \cdot 10^{-3}$
$\text{LSTM}_{n=5,s=10}$	HQ	0.824	0.975	$1.96 \cdot 10^{-5}$	$4.43 \cdot 10^{-3}$
	MQ	0.549	0.967	$3.31 \cdot 10^{-5}$	$5.75 \cdot 10^{-3}$
$\text{LSTM}_{n=5,s=20}$	HQ	0.836	0.986	$1.76 \cdot 10^{-5}$	$4.19 \cdot 10^{-3}$
	MQ	0.569	1.027	$3.36 \cdot 10^{-5}$	$5.79 \cdot 10^{-3}$
$\text{LSTM}_{n=5,s=15}$	HQ	0.835	0.961	$1.76 \cdot 10^{-5}$	$4.19 \cdot 10^{-3}$
	MQ	0.553	0.949	$3.41 \cdot 10^{-5}$	$5.84 \cdot 10^{-3}$
$\text{LSTM}_{n=6,s=10}$	HQ	0.847	0.965	$1.72 \cdot 10^{-5}$	$4.14 \cdot 10^{-3}$
	MQ	0.561	0.954	$3.24 \cdot 10^{-5}$	$5.69 \cdot 10^{-3}$
$\text{LSTM}_{n=6,s=15}$	HQ	0.823	0.976	$1.73 \cdot 10^{-5}$	$4.16 \cdot 10^{-3}$
	MQ	0.541	0.963	$3.33 \cdot 10^{-5}$	$5.77 \cdot 10^{-3}$
$\text{LSTM}_{n=6,s=20}$	HQ	0.804	0.981	$1.85 \cdot 10^{-5}$	$4.31 \cdot 10^{-3}$
	MQ	0.543	0.983	$3.46 \cdot 10^{-5}$	$5.88 \cdot 10^{-3}$

Table 2: Estimation of the quality of prediction of CO_2 flux for considered models. For each model parameters Mean Squared Error (SME), Root Mean Squared Error (RSME) and R^2 as well as the slope of the trend line are presented.

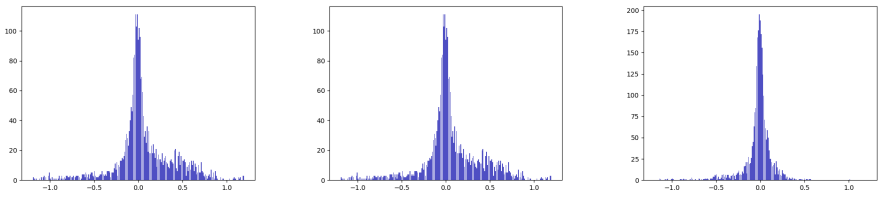
The models are trained to focus on HQ results. Therefore, for a detailed assessment of the training process, we measure each class's effectiveness independently. We present a comparison of the results obtained from a neural network with the expected value of CO_2 flux for each class. The assessment was conducted on the test dataset that was not used in training. As shown in Figure 4, the results predicted from $\text{LSTM}_{n=6,s=20}$ network give results with a high level of agreement with the expected (real) values for HQ and MQ classes. The results for the lowest quality data are not similar to the expected ones. As was mentioned before, these measurements have very low reliability. In Figures 4a, 4c, 4e we present the dependence of predicted CO_2 flux on the real value, this

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_44

plot should be as near as possible to a line. In the figures, we draw the orange lines that represents the optimal prediction (a line with slope equal to 1) and the trend line based on data points (in blue). We can see that the HQ data is predicted with high accuracy. The MQ data is not so well concentrated around the optimal prediction line as HQ one is. The last low-quality class is not predicted properly, but as was mentioned before, this was expected, as this data has low credibility. Figures 4b, 4d, 4f compare the behaviour of predictions and expected values for a test set in time. As we can see again, the prediction for HQ class well agrees with the expected values.

The optimal prediction would result in a simple relation: $\hat{y} = 1 * y$ for all $\hat{y}$ that belong to the class in consideration. Therefore, we derive linear regression estimators for every class and all models: R^2, slope, MSE (Mean Squared Error), RMSE (Root Mean Square error) to qualify the prediction. The estimator values are presented in Table 2.

Table 2 presents that most of the LSTM network models have similar quality. It seems that the LSTM with a window length set to 4 gives the best prediction. However, the differences between the qualities of the models are small. The models that have been trained can be used for additional assessment of the quality of experimental data. As we can see in the Figure 4f, the model predicts more often lower values of high quality CO_2 flux than measured. Therefore, we analyze the residuum of prediction (i.e., the difference between prediction and a real value) Figure 5. We can see that the shape of the histogram is not symmetric. As a result, the model predicts more often values lower than the real experimental data. It can be understood as extreme values of CO_2 flux are rare and hard to predict.



(a) LQ class (b) MQ class (c) HQ class

Fig. 5: The histograms of residuum for prediction using LSTM network with $s = 20$, $n = 6$ after 120 epochs on test set. On the horizontal axis we can see the value of residuum on the vertical one the number of occurrences.

5 Conclusions

In the paper, we present LSTM neural network usage for CO_2 flux prediction in the meteorological experiment. The data have three different levels of credibility

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_44

and have to be treated differently during training. We propose to use a weighted means squared error as a loss function during the training process. The method leads to good results with a high level of agreement with the expected values for high-quality results. The network also correctly assesses the quality of the record in consideration. The prepared LSTM network can be used for the automatic verification of experimental raw data. Additionally, we can use the method for gap filling to repair the records with flows. In the future, the usage of the method for gap filling will be analyzed in detail.

Acknowledgements

Funding for this research was partly provided (data collection) by the National Science Centre, Poland under the project UMO-2015/17/B/ST10/02187. The site was established in 2012 under project UMO-2011/01/B/ST10/07550 founded by National Science Centre, Poland. The authors thank the authorities of the Biebrza National Park for allowing continuous measurements in the area of the Park.

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X.: Tensorflow: A system for large-scale machine learning. In: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. p. 265–283. OSDI'16, USENIX Association, USA (2016)
2. Aubinet, M., Vesala, T., Papale, D. (eds.): Eddy Covariance: A Practical Guide to Measurement and Data Analysis. Springer Netherlands (2012). https://doi.org/10.1007/978-94-007-2351-1
3. Baldocchi, D.D.: Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. Global Change Biology **9**(4), 479–492 (2003). https://doi.org/10.1046/j.1365-2486.2003.00629.x
4. Che, Z., Purushotham, S., Cho, K., Sontag, D., Liu, Y.: Recurrent neural networks for multivariate time series with missing values. Scientific Reports **8**(1) (2018). https://doi.org/10.1038/s41598-018-24271-9
5. Christiansen, N.H., Voie, P.E.T., Winther, O., Høgsberg, J.: Comparison of neural network error measures for simulation of slender marine structures. Journal of Applied Mathematics **2014**, 1–11 (2014). https://doi.org/10.1155/2014/759834
6. Dengel, S., Zona, D., Sachs, T., Aurela, M., Jammet, M., Parmentier, F.J.W., Oechel, W., Vesala, T.: Testing the applicability of neural networks as a gap-filling method using CH4 flux data from high latitude wetlands. Biogeosciences **10**(12), 8185–8200 (2013). https://doi.org/10.5194/bg-10-8185-2013
7. Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grünwald, T., Hollinger, D., Jensen, N.O., Katul, G., Keronen, P., Kowalski, A., Lai, C.T., Law, B.E., Meyers, T., Moncrieff, J., Moors, E., Munger, J., Pilegaard, K., Üllar Rannik, Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T.,

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_44

Wilson, K., Wofsy, S.: Gap filling strategies for defensible annual sums of net ecosystem exchange. Agricultural and Forest Meteorology **107**(1), 43–69 (2001). https://doi.org/10.1016/s0168-1923(00)00225-2
8. Fathalla, A., Salah, A., Li, K., Li, K., Francesco, P.: Deep end-to-end learning for price prediction of second-hand items. Knowledge and Information Systems **62**(12), 4541–4568 (2020). https://doi.org/10.1007/s10115-020-01495-8
9. Fortuniak, K., Pawlak, W., Bednorz, L., Grygoruk, M., Siedlecki, M., Zielinski, M.: Methane and carbon dioxide fluxes of a temperate mire in central europe. Agricultural and Forest Meteorolog **232**, 306–318 (2017)
10. Fortuniak, K., Pawlak, W., Siedlecki, M., Chambers, S., Bednorz, L.: Temperate mire fluctuations from carbon sink to carbon source following changes in water table. Science of The Total Environment **756**, 144071 (2021). https://doi.org/10.1016/j.scitotenv.2020.144071
11. Gers, F.A., Eck, D., Schmidhuber, J.: Applying LSTM to time series predictable through time-window approaches. In: Artificial Neural Networks — ICANN 2001, pp. 669–676. Springer Berlin Heidelberg (2001). https://doi.org/10.1007/3-540-44668-0_93
12. Goodfellow, I., Bengio, J., Courville, A., Bach, F.: Deep Learning. MIT Press Ltd (2016)
13. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
14. Ke, J., Zheng, H., Yang, H., Chen, X.M.: Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. Transportation Research Part C: Emerging Technologies **85**, 591–608 (2017). https://doi.org/10.1016/j.trc.2017.10.016
15. Kim, Y., Johnson, M.S., Knox, S.H., Black, T.A., Dalmagro, H.J., Kang, M., Kim, J., Baldocchi, D.: Gap-filling approaches for eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal component analysis. Global Change Biology **26**(3), 1499–1518 (2019). https://doi.org/10.1111/gcb.14845
16. Moffat, A.M., Papale, D., Reichstein, M., Hollinger, D.Y., Richardson, A.D., Barr, A.G., Beckstein, C., Braswell, B.H., Churkina, G., Desai, A.R., Falge, E., Gove, J.H., Heimann, M., Hui, D., Jarvis, A.J., Kattge, J., Noormets, A., Stauch, V.J.: Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. Agricultural and Forest Meteorology **147**(3-4), 209–232 (2007). https://doi.org/10.1016/j.agrformet.2007.08.011
17. Papale, D.: Data gap filling. In: Eddy Covariance, pp. 159–172. Springer Netherlands (2011). https://doi.org/10.1007/978-94-007-2351-1_6
18. Selvin, S., Vinayakumar, R., Gopalakrishnan, E.A., Menon, V.K., Soman, K.P.: Stock price prediction using LSTM, RNN and CNN-sliding window model. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE (2017). https://doi.org/10.1109/icacci.2017.8126078
19. Spelmen, V.S., Porkodi, R.: A review on handling imbalanced data. In: 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT). IEEE (2018). https://doi.org/10.1109/icctct.2018.8551020
20. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2. p. 3104–3112. NIPS'14, MIT Press, Cambridge, MA, USA (2014)

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77980-1_44