

## USE OF ARTIFICIAL NEURAL NETWORKS TO PREDICT GREENHOUSE GASES EMISSIONS

**Ksenia Kononenko, Dmitry Demidov**

Moscow Polytechnic University, Russia

kononenkoxenia@gmail.com, ddemidich@gmail.com

**Abstract.** Agriculture has a greater impact on the environment than any other economic sector in the world. There is a problem of forecasting greenhouse gas emissions depending on various factors of agricultural activity. It is proved that the ability to make forecasts of the amount of greenhouse gas emissions would significantly increase the rational use of agricultural resources. The use of an artificial neural network as a forecasting method in the field of agriculture can allow predicting the environmental impact of greenhouse gases in advance. In addition, this method of data analysis can be an alternative solution of agroecological problems. The article clarifies the features of constructing artificial neural networks for the analysis of dynamic data in the field of agricultural ecology. The approach to solving the problem of forecasting greenhouse gases emissions – an artificial neural network is considered. The problem of forecasting greenhouse gases is considered as a problem of forecasting time series. As an example, an artificial neural network has been designed to predict CO<sub>2</sub>eq emissions from agriculture in the Russian Federation. The data for training the neural network were taken from the Food and Agriculture Organization open database (FAOSTAT). The test results of the developed neural network are presented. The conclusion about the possibility of further application of the development (the model of an artificial neural network for predicting CO<sub>2</sub>eq emissions in agriculture) is given.

**Keywords:** neural network, greenhouse gases emission, agricultural, information technology.

### Introduction

Now the world community is more than ever concerned about environmental issues. More and more countries are beginning to implement state environmental conservation programs. The main problem is an increase in the level of greenhouse gases in the atmosphere. In turn, agriculture is one of the largest sources of greenhouse gases. Agriculture emits methane, carbon dioxide and nitrous oxide [5]. Obviously, predicting possible emissions of these gases would help rationalize the consumption of agricultural resources.

Currently, information technology has a predictive analytics tool – artificial neural networks (ANN). Artificial neural networks have learned to perform complex tasks that are not subject to classical statistics and analytics. Nevertheless, at the moment there is no such artificial neural network that we could predict greenhouse gas emissions from agriculture.

Thus, in our research work, we decided to develop an artificial neural network that allows us to predict the level of CO<sub>2</sub>eq for the year ahead. According to FAOSTAT (the Food and Agriculture Organization open database), CO<sub>2</sub> eq – is a metric used to compare different greenhouse gases, so we use it.

The main goal of the article is to test the feasibility of using an artificial neural network to predict agricultural emissions of various greenhouse gases.

### Material and methods

For training the neural network we used open data from the FAO database called FAOSTAT [4]. For the experiment, we chose the Russian Federation as the region for which the neural network should predict the level of CO<sub>2</sub> eq from agricultural sector. Observations of the level of CO<sub>2</sub> eq in the Russian Federation were carried out from 1992 to 2017 (according to FAOSTAT). The factors affecting CO<sub>2</sub> eq in agriculture, which we took into account when creating the training dataset, are the following.

1. Burning – Crop residues;
2. Burning – Savanna
3. Crop Residues;
4. Cultivation of Organic Soils;
5. Enteric Fermentation;
6. Manure Management;
7. Manure applied to Soils;

8. Manure left on Pasture;
9. Rice Cultivation;
10. Synthetic Fertilizers.

The CO<sub>2</sub> eq emissions by each of the factors were summed over the years. Thus, for training the neural network, data pairs were generated: the volume of CO<sub>2</sub> eq in gigagrams and the year of observation. It turned out 26 pairs.

We normalized the CO<sub>2</sub> eq values placing their values in the range from 0 to 1.

According to the principles of machine learning, the dataset was divided into three parts - a data set for training, a data set for validation, and a data set for testing. The composition of the kits is described in more detail in the “Development of ANN” section.

The CO<sub>2</sub>eq emission represents discrete statistics. So, the studied process, the emission of CO<sub>2</sub> eq at different points in time (years), which have the same intervals, forms a time series [1]. The task of predicting the value of CO<sub>2</sub> eq for the next year can be considered as a task of forecasting time series.

We present a graph (Fig. 1) reflecting the dependence of the volume of emissions on time.

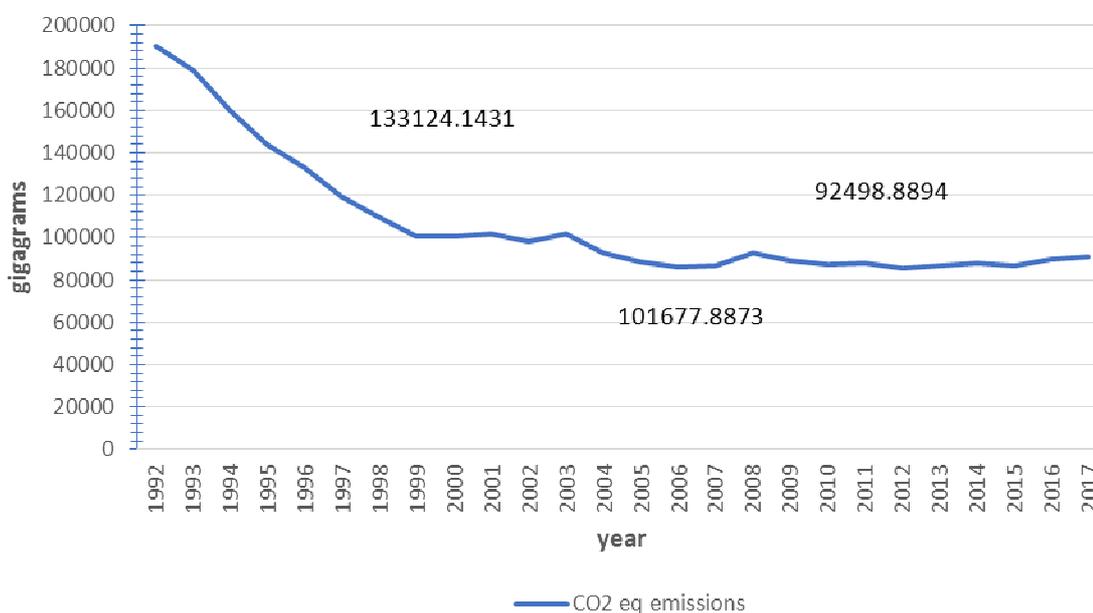


Fig. 1. Emissions CO<sub>2</sub> by agriculture in Russia from 1992 to 2017

Thus, the task of predicting a time series can be understood as predicting future values of a time series from present and past values.

Three values are fed to the input of the ANN – CO<sub>2</sub> eq values for the current year, CO<sub>2</sub> eq values for the previous year and CO<sub>2</sub> eq values two years before the current one. Thus, the data set for training was transformed into triples of values.

The training set consisted of 17 values (CO<sub>2</sub> eq values from 1995 to 2011), the set for validation consisted of 3 values (CO<sub>2</sub> eq values from 2012 to 2014), and the set for test consisted of 3 values (CO<sub>2</sub> eq values from 2015 to 2017).

At the output, we get the value of CO<sub>2</sub> eq next year.

It is important to correctly determine the number of hidden layers. To solve this problem, we used the gradual addition method. First, the number of hidden layers was calculated by the formula (1).

$$N_h = \frac{N_s}{\alpha \cdot N_i \cdot N_o}, \quad (1)$$

where  $N_h$  – number of hidden layers;  
 $N_i$  – number of input neurons;  
 $N_o$  – number of output neurons;

$N_s$ – number of samples in the training data set;  
 $\alpha$ – arbitrary scaling factor.

The initial number of neurons was calculated according to formulas (2), (3) and (4) [3].

$$r = \sqrt[3]{\frac{n}{m}}, \tag{2}$$

$$k_1 = mr^2, \tag{3}$$

$$k_2 = mr, \tag{4}$$

where  $r$  – intermediate ratio;  
 $n$  – number of neurons in the input layer;  
 $m$  – number of neurons in the output layer;  
 $k_1$  – initial number of neurons for the first hidden layer;  
 $k_2$  – initial number of neurons for the second hidden layer.

We went directly to the method of gradual addition – we added 1 neuron, tracked the error and the speed of learning.

Acceptable results – a relatively short training time and a fairly minimal error, were achieved using two hidden layers of 10 neurons in each (Fig. 2).

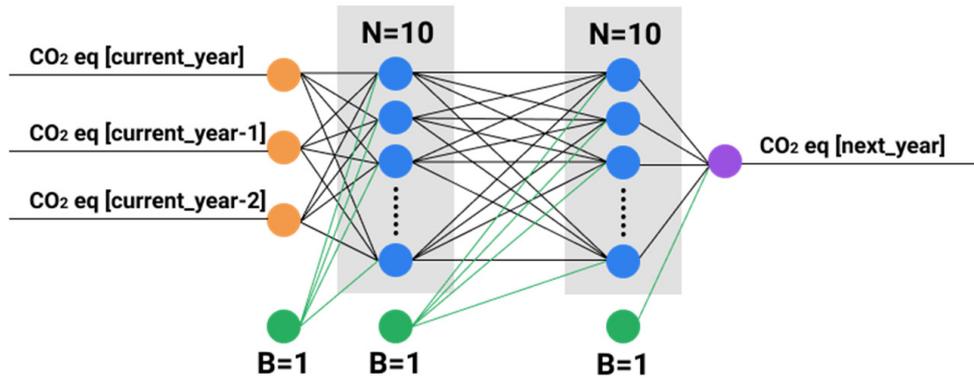


Fig. 2. Architecture of the developed neural network

As a result, this ANN contains 161 parameters.

In Figure 2, the blue color indicates the neurons of the hidden layers, the green color indicates the biases, the orange color indicates the input neurons and the purple color indicates the output neuron.

We chose the Adadelta algorithm as a training method. This algorithm is based on stochastic gradient descent, but has several advantages over it. The algorithm is able to set its learning speed for various parameters. Also, unlike its predecessor, the Adagrad algorithm, it allows to limit the number of stored gradients [2]. The disadvantage of Adadelta is the gradually decreasing learning speed, this proves the graph of the loss function during the training of our ANN, presented below (Fig. 3).

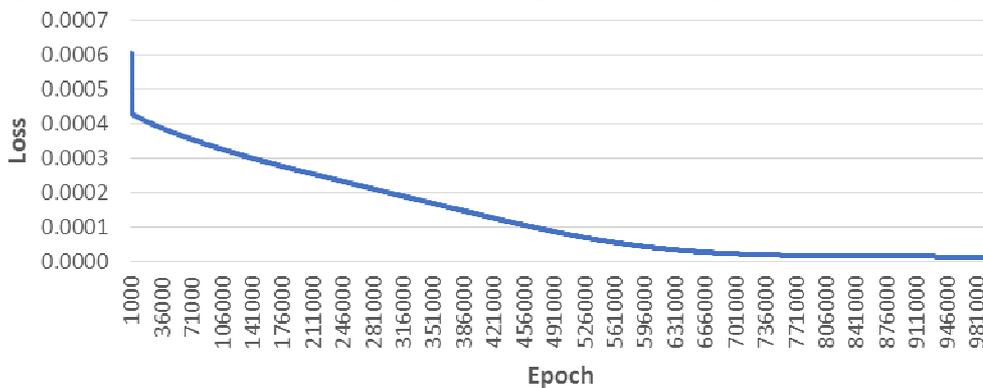


Fig. 3. Value of the loss function during training

Thus, it is difficult to quickly achieve high-precision results. In addition, when training, it is important to consider that the output values should be accurate, and not approximately similar to the standard.

### Result and validation

The result predicted by the model and the goal should be the same, have a minimum error and have a high correlation. The regression curves should show how good a match is between the predictions and the goal.

Using the created artificial neural network, we made predictions of CO<sub>2</sub> eq. The result obtained was quite satisfactory to us and we did the same with the set for test. As a result, a graph with real and predicted values (for the entire data set) can be seen below (Fig. 4).

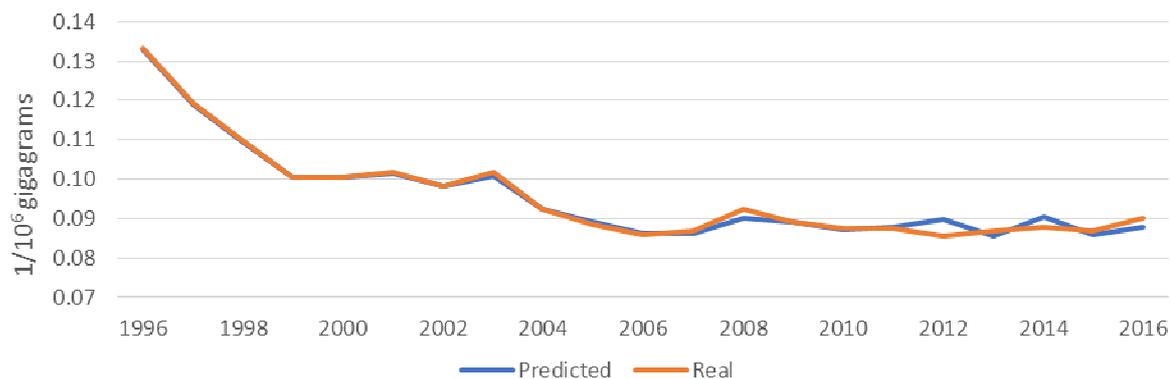


Fig. 4. Comparison of predicted and real CO<sub>2</sub> eq values

Based on the graph, comparing the predicted and real values, we can see that the created network made good forecasts, almost coinciding with the real ones.

To verify the assumption, the Pearson correlation coefficient between the predicted and real values was calculated. The Pearson coefficient is 0.993. This means that there is a strong relationship between the quantities.

Thus, taking into account the fact that training can be continued, the value of the loss function can still decrease.

The model proved its competence in solving the problem of forecasting time series (CO<sub>2</sub> eq emissions from agriculture).

### Conclusions

We proposed a solution to the problem of predicting CO<sub>2</sub>eq emissions – we created and tested our artificial neural network. At the present stage of creation, the model has proved its operability in the framework of the forecasting problem.

This architecture of the neural network and the principles of its construction can be used in similar tasks – the prediction of time series.

The disadvantage of this model is the insufficient number of trainings – model training should be continued to obtain more accurate results.

Another disadvantage of this model is the possibility of random noise. To solve the problem of the appearance of random noise, we plan to use additional methods to eliminate it.

### References

- [1] Стьюарт А., Кендалл М. Дж. Многомерный статистический анализ и временные ряды (Multivariate Statistical Analysis and Time Series). Москва: «Наука» (Moscow: “Nauka”), 1976. 736 p. (In Russian)
- [2] Mohamad H. Hassoun. Fundamentals of Artificial Neural Networks [online] [10.01.2020]. Available at:

<https://books.google.ru/books?id = Otk32Y3QkxQC&lpg = PP1&dq = neural %20network&hl = r u&pg = PP1#v = onepage&q = number %20of %20hidden&f = false>

- [3] Терехов С.А. Лекции по теории и приложениям искусственных нейронных сетей. Лаборатория искусственных нейронных сетей НТО-2, ВНИИТФ, Снежинск. [online] [03.02.2020]. Available at: [http://alife.narod.ru/lectures/neural/Neu\\_ch06.htm](http://alife.narod.ru/lectures/neural/Neu_ch06.htm) (In Russian)
- [4] The Food and Agriculture Organization. Database. [online] [10.12.2019] Available at: <http://www.fao.org/faostat/en/#data/>
- [5] FAO. Greenhouse Gas Emissions from Agriculture, Forestry and Other Land Use. [online] [10.12.2019] Available at: <http://www.fao.org/resources/infographics/infographics-details/en/c/218650/>