THE COMPARISON OF LINEAR MODELS FOR PM_{10} AND $PM_{2.5}$ FORECASTING

PIOTR KOWALSKI[1,2] & WIKTOR WARCHALOWSKI[3]
[1]Faculty of Physics and Applied Computer Science, AGH University of Science and Technology, Poland
[2]Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
[3]Airly sp. z o.o. Aleja Pokoju 1a, 31-548 Cracow, Poland

ABSTRACT

Air pollution is a very serious problem in Poland and elsewhere, and it is a factor that significantly affects the quality of human life. However, people are not fully aware of the terrible air quality due to the insufficient number of monitoring stations. This means they have no access to information about the quality of the air they breathe. The aim of this paper is to present and compare some linear procedures for PM_{10} and $PM_{2.5}$ forecasting. Herein, the simulations concerning investigated prediction algorithms are based on real data originating from the Airly company network of pollution measurement stations. Related data, including measurements, were gathered every hour for a period of about one year, moreover, for forecasting purposes, weather data from the Dark Sky portal was additionally used. In this study, several Machine Learning predictive methods are considered. Among these, the results of three are presented. These are: Multiple Linear Regression, Multiple Linear Regression with Regularisation and, finally, Linear Neural Networks. The task for each predictive algorithm was to predict the concentration of PMx dust in the following hours of the next day. As a measure of the prediction task evaluation, several types of error were considered, while, during the research, machine learning group algorithms were utilized as learning models. Via these advanced, efficient and convenient algorithms, a detailed air quality forecast for the next 24 hours is obtainable. The presented algorithms will be implemented into the common air condition prediction system.

Keywords: air pollution prediction (forecasting), particulate matters, machine learning, regression task, linear model, linear neural networks.

1 INTRODUCTION

Human health depends on the purity of the air it breathes. Unfortunately, Poland is not a country that could boast of excellent air quality. This is evidenced by the WHO report which shows that among the 50 cities and towns of the EU with the highest level of pollution, 33 are situated in Poland.

The modelling that is related to the prediction of weather and air pollution levels is a very difficult issue in data analysis. These encounter such problems as Big Data, Missing Data, Unreliable Data, etc. Therefore, finding a convenient formal mathematical model that allows the modelling or the prediction of the studied phenomenon is very difficult [1]–[3]. Such a research task can only be solved by the application of modern methods collectively held under the banner of Machine Learning.

In order to meet this problem, the present paper will present and compare models of PM_{10} and $PM_{2.5}$ air pollution prediction (forecasting), which are based on linear regression, as well as on artificial neural network procedures. Intelligent learning algorithms [4], [5] for individual models will also be proffered. The evaluation of learning process for the investigated models was based on the mean square error criterion, however, during the model validation, a number of other methods of quantitative evaluation were taken into account.

The presented models of pollution prediction have been verified by way of real weather and air pollution data taken from the Airly sensor network. The distributed network of Airly measurement devices enables access to current and archival data on air pollution, $PM_{2.5}$ and PM_{10} suspended particulate matter, CAQI level, temperature, atmospheric pressure and air

WIT Transactions on Ecology and the Environment, Vol 230, © 2018 WIT Press
www.witpress.com, ISSN 1743-3541 (on-line)
doi:10.2495/AIR180161

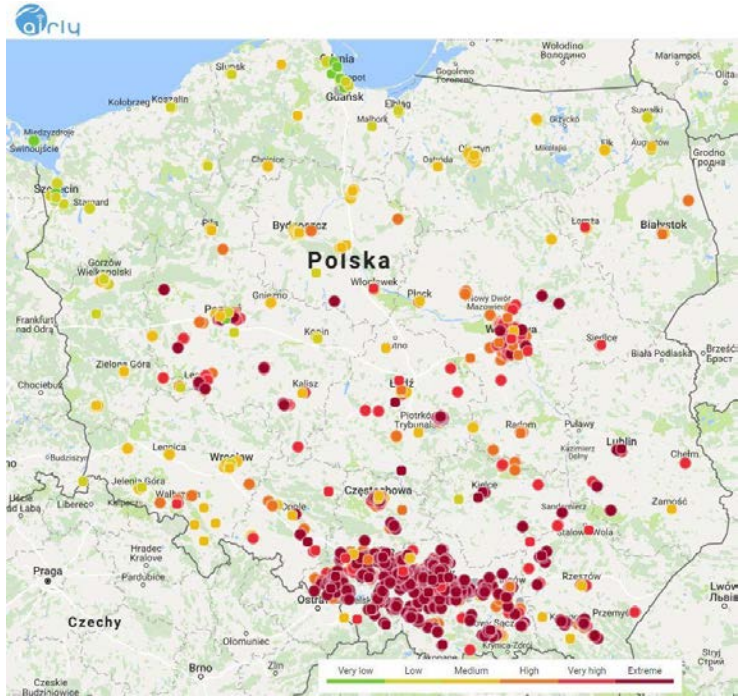


Figure 1: A map with network of Airly's sensors.

humidity. The above parameters are collected from the Airly sensor network by way of more than 1,300 measurement stations located throughout Poland (Fig. 1). Airly is considered the undisputed leader in the domestic market with regard to air pollution measurement and related issues.

Air pollution forecast modelling may include many parameters and take into account several diverse premises. However, it is always quite a difficult task and very individualized [6]–[8]. One solution is to apply Fuzzy Logic techniques [24], [25]. Another is to apply a variety of neural network procedures [11], [12]; among the most useful being Radial Basis Function and Elman Networks [13], as well as the latest recurrent types [14]–[16]. In article [17], the accuracy of prediction of PM_{10} pollution by wavelet transformation and an ensemble of neural predictors is considered.

The paper is organized as follows. In Section II, we briefly inform the reader about the data considered for the presented investigation. In Section III, linear models such as Multiple Linear Regression, Multiple Linear Regression with Regularisation, and, finally, Linear Neural Networks are introduced and investigated. In Section IV, two procedures for the learning perdition model i.e. Gradient Descent Optimisation algorithm [4], as well as Adaptive Moment Estimation algorithm [5], are advanced. In Section V, all results of numerical verification are presented and discussed. Finally, Section VI contains concluding remarks and plans for future investigations.

2 DATA SET

In this paper, real data from Airly's air sensors are used to investigate the state of air pollution with regard to PM_{10} and $PM_{2.5}$ dust. All the studies presented in this article are built upon

data obtained from a measurement station located in one of the villages nearby Cracow. Additionally, forecasts of meteorological phenomena from the DarkSky portal are utilized for prediction purposes of the dust pollution condition.

In order to carry out the simulation, a time series were used. This included a period of about one year, i.e. from September 2016 to July 2017. The above data has been recorded as hourly average measurements of pollutant and weather conditions. In this work, in each measurement (at time t) of the dust concentration PM_{10} [µg/m^3] (Fig. 2(a)) and $PM_{2.5}$ [µg/m^3] (Fig. 2(b)), temperature [°C] (Fig. 2(c)) and the average wind force [m/s] (Fig. 2(d)) were registered (labeled as x_{10}^t, $x_{2.5}^t$, x_{temp}^t and x_w^t respectively).

In total, the data include 5969-hour measurements. Due to the specificity of the modelling process in the Machine Learning domain, these data were divided into two separate sets. The first one is a training set and is employed in model building, including refinement of its particular parameters. The second is the test set, the task of which is to evaluate the finished model. In the present case, 5242 elements of the data set are set aside for the training data (x_{learn}) and the remaining 727 elements are applied as test data (x_{test}). It should be emphasized here that the figures in Fig. 2 are of non-continuous character. This is a consequence of the use of data from real measuring sensors, and these discontinuities result from minor faults, service breaks or breakdowns of the external computer networks transmitting data to servers aggregating them.

Fig. 3 presents an analysis of the metrological variables used in the proposed models. Of significance is the relationships between the daily concentrations of PM_{10} and $PM_{2.5}$. In Fig. 3(a), one can see a near linear relationship between these types of pollution, hence, models and results describing the prediction of PM_{10} dust abatement can be generated [16], [17]. From Fig. 3(b) we can see that the high intensity of wind results in an increased

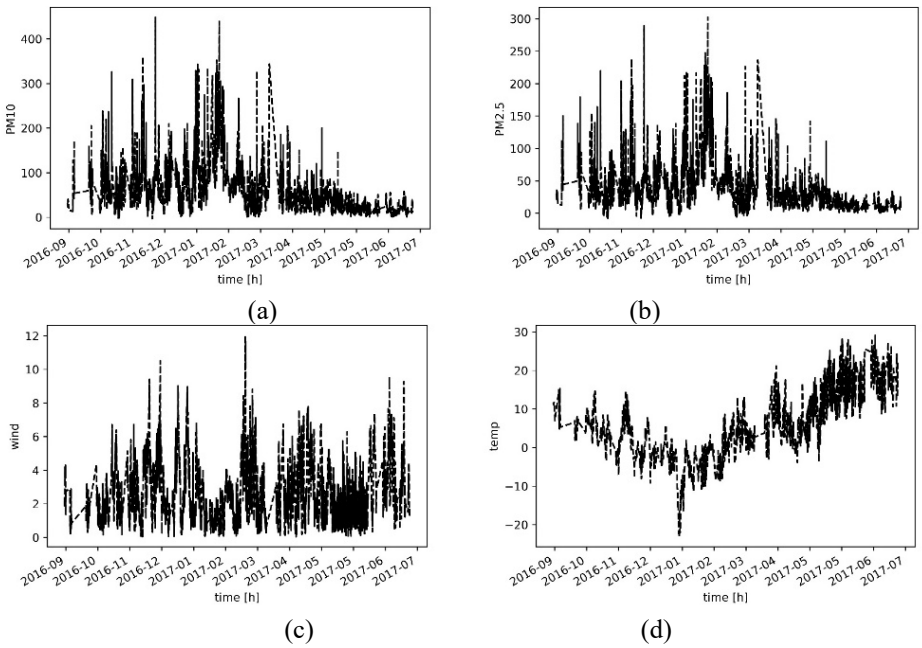


Figure 2: The daily average value of investigated quantities. (a) PM_{10}; (b) $PM_{2.5}$; (c) Wind; (d) Temperature.

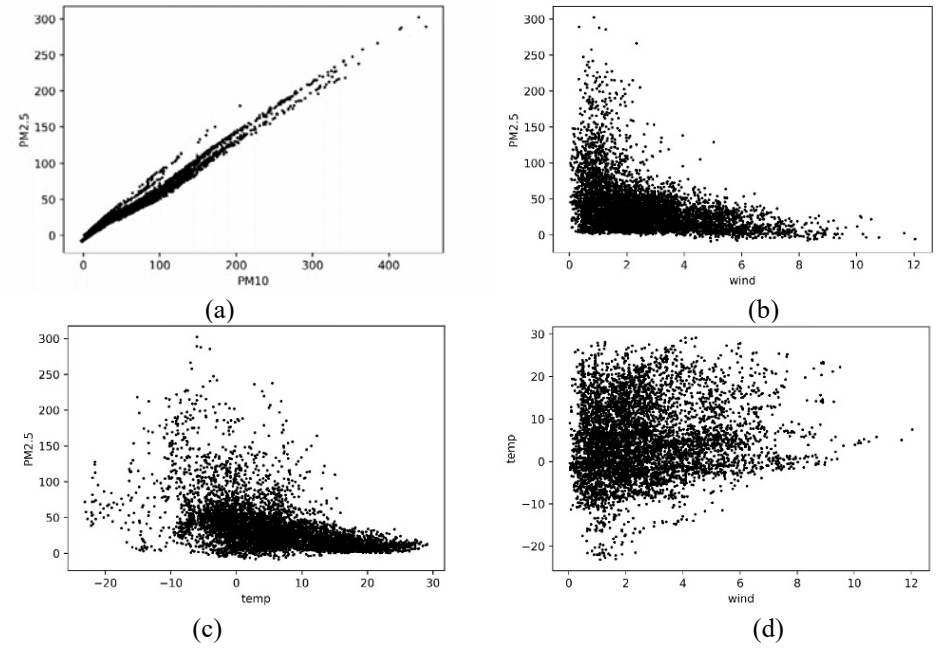


Figure 3: The graphical relationships between the daily concentration of studied quantities. (a) PM_{10} vs $PM_{2.5}$; (b) Wind vs $PM_{2.5}$; (c) Temperature vs $PM_{2.5}$; (d) Wind vs temperature.

concentration of dust. The next Fig. 3(c) shows the dependence of the dust concentration on the temperature. It is noticeable here that there is a certain constant value of dust concentration (about 25 [μg/m^3]), which does not depend on temperature. Furthermore, at temperatures falling below 12 [°C], a growing trend in $PM_{2.5}$ pollution is evident. However, a drop-in temperature below 8 [°C] results in a significant increase in the concentration of dust in the air. Fig. 3(d) indicates the lack of a special connection between the wind and the temperature.

3 INVESTIGATED MODELS

In this work, 24 predictors were considered, the task of which is to predict the average hourly PM_{10} contamination ($\widehat{y^{\tau}}$) for the next 24 hours (i.e. for τ = 1, ..., 24). All used models belonged to the class of linear predictors. In each of the comments considered, the main measure allowing the model to be evaluated was the mean squared error (MSE). This is expressed as:

$$MSE = \frac{1}{P}\sum_{i=1}^{n}(\widehat{y_i^{\tau}} - y_i^{\tau})^2, \quad (1)$$

where $\widehat{y_i^{\tau}}$ is n predictions of investigated forecast (for τ time), and y_i^{τ} consists of n observed values of the predicted variable. What is more, the correlation measurement R is used for the completed construction process and trained model. This can be defined as follows:

$$R = \frac{Cov(y^{\tau}, \widehat{y^{\tau}})}{std(y^{\tau})\ std(\widehat{y^{\tau}})}, \quad (2)$$

where $std(y^{\tau})$ and $std(\widehat{y^{\tau}})$ are the standard deviations of observed value and the investigated forecast, respectively, and $Cov(y^{\tau}, \widehat{y^{\tau}})$ denotes the covariance coefficient.

The first of the investigated predictive models is that based on multiple linear regression (MLR). This model relies on the generation of 24 predictors that are described as:

$$\widehat{y^{\tau}} = Wx. \tag{3}$$

In this study, the vector x contains both "historical" and predictive data. The first group includes 25 actual indications of concentrations of $PM_{2.5}$, PM_{10}, temperature and wind power coming from the Airly station measurements. The second group are temperature and wind forces obtained from the DarkSky portal. Thus, the x vector can be written as:

$$x = \left\{x_{10}^{-25}, \dots, x_{10}^{-1},\ x_{2.5}^{-25}, \dots, x_{2.5}^{-1},\ x_{temp}^{-25}, \dots\ x_{temp}^{24}, x_{w}^{-25}, \dots\ x_{w}^{24}\right\}. \tag{4}$$

With the input data adopted in eqn (3), the W weights matrix contains 150x24 coefficients. At the same time, for each hour τ, 150 individual coefficients are calculated separately.

Another model considered in this article is the MLR modification. This consists of adding a penalty function for too large values of individual elements in the matrix W. This model is called "multiple linear regression with regularization of the L2 type" (MLRR). It is based on modification of the evaluation function (1) in the following way:

$$error_{MLRR} = \frac{1}{P}\sum_{k=1}^{P}(\widehat{y_i^{\tau}} - y_i^{\tau})^2 + \lambda \sum_{i=1}^{\#w} w_i^2\ . \tag{5}$$

In eqn (5), the new element consists of a sum of squares of all weights from model (4). This L2 technique is very often called "Ridge Regression", as it adds "squared magnitude" of coefficient as a penalty to the loss function. This method allows the creation of less complex models especially in situations wherein a large number of features are contained in the considered dataset. It should be also underlined that regularization techniques are used to address the over-fitting problems that may appear during model training.

The last model to be considered is a set of linear neural networks (LNN). The structure of this model consists of 4 layers of neurons: input, two hidden and output, with the number of each being limited to 150, hid1, hid2 and 1, respectively. In each of the layers with the exception of input, a linear activation (transfer) function is used. As in the previous case, for each discrete moment of time τ, a separate neural network is generated.

4 LEARNING PROCEDURES

In this section, certain familiar machine learning procedures will be briefly described. Their task is to adjust the value of weight coefficients of particular methods to the problem under consideration. This process in the Machine Learning nomenclature is often called the "learning" or "training" process.

In the present study, we consider the use of two methods for updating model weight iterates, based on training data. The first of these is the Gradient Descent Optimisation procedure (GDO). This algorithm starts off with initial values for the weight coefficient which could be 0.0 or another small random value. Next, the cost of the model is evaluated by mapping them into the cost function. The derivative of the cost function (a concept that refers to the slope of the function at a given point in the investigated set of weights) is then calculated. As the value of the derivative (in a multidimensional case, a gradient approach is applied) is the slope, the procedure "knows" the direction (sign) to move the weight

coefficient values in order to obtain a lower cost within the next iteration. Hence, from the derivative, the direction of which is downhill, updates of the coefficient values can be made. In this part of the algorithm, a learning rate parameter must be specified. This enables control as to how much the weight coefficients need be changed on each update. This process is repeated iteratively until the stop criteria is not fulfilled. In this research, the stop criteria, defined as a conjunction of the achieving a lower value of the cost function than assumed, and the expiration of the learning time is expressed in so-called "epochs".

Another algorithm used to adapt weight factors in the model is the ADAM algorithm. This procedure is different to classical GDO in that it combines the advantages of other stochastic gradient descent procedures, in particular, Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp). The main advantage is that it enables good control over the learning rate parameter. Herein, the AdaGrad procedure improves performance on problems with sparse gradients, while the RMSProp algorithm lr parameter is adapted based on the average of recent magnitudes of the gradients for the weight. This last results in the algorithm being effective when applied to online and non-stationary problems. The iterative nature of the ADAM algorithm is very similar to the GDO described above.

5 RESULTS

This section is devoted to the description of the conducted research and to a discussion of the received results. Herein, as a criterion describing the quality of the solution obtained, the MSE error described in eqn (1) was adopted during learning. Moreover, after completing the adaptation procedure of the scales, the model was verified both with the help of the MSE error and the correlation measure, i.e. the R (2) coefficient, as each of the verifications is applicable for the teaching and the testing sample. In the case of the MSE error, a better model is characterized by a lower value of this error, while for the second measure, the ideal model is characterized by a value of the coefficient R = 1. We first assessed the behaviour of the choice of the learning method. In all cases of the proposed models, the use of the GDO procedure showed much better properties than did the ADAM procedure. The aforementioned was made evident through taking into account the MSE error value, the R measure and the convergence of the learning procedure. For example, for the MLR model, for $\tau = 1$, after completion of learning, a model was obtained that characterized MSE = 0.9316 and R = 0.9256 when using the GDO algorithm. With regard to the ADAM procedure, for the same data (without the division into training and testing data), a model is created with MSE = 3350.64 and R = 0.19. For the other models considered here, the results of the comparison between the teaching methods were similar. In connection with the above, only the student's method of GDO was used for further research.

The second study consisted in creating an MLR model that would allow PM10 dust forecasting for the subsequent 24-hour period (with an hourly step). In this case, the data from the weather station was first divided into training and testing data in the proportion of 85% to 15%. Table 1 shows the results of individual simulations, with data related to individual predictors inserted in subsequent columns. In turn, the first two lines present the results of the measure R for the given learners and testers, while the next two lines reveal the value of the error MSE for both sets of data.

From the results presented in Table 1, one can easily see that the predictions of the degree of contamination in subsequent time periods are burdened with increasingly lower values of the measure R, which means a worsening match of the model to the data. The results differ, however, in the case of MSE error values. In this case, the MSE error decreases from the value of 0.9460 for the first prediction, down to the value 0.8159 for the last. This rather strange 6ehaviour is reflected in the lack of consideration of strong pollutant peaks (this effect

will be presented graphically in the following section). Beyond the aforementioned, it is evident that the R measure for the training data is more favourable than that for the test data – a Machine Learning domain algorithm characteristic.

In the next two tables (Table 2 and Table 3), the results of MLR model utilization are presented, but with the application of the modified cost function. This modification consists of adding a penalty fraction for too large weighting factors, hence the influence of the lambda parameter on the quality of prediction can be seen in this part of the research. Due to the limited space of the article, MLRR results are presented for the first and last time moments only, i.e. the first and 24th hours of prediction.

In Tables 2 and 3, the results are presented in the same convention as in Table 1. The first column of results is a repetition of that obtained for the MLR model and is a reference point in the comparative studies. From the presented data, it is notable that adding a penalty for too large model weights yielded desirable results. The choice of the lambda coefficient proved to be the key. From the third and fourth columns (i.e. for λ equal to 0.01 and 0.001) we see that the value of R is smaller than that of the reference in the second column (for the MLR model). What is more, the reduction of this parameter to 0.00001 has resulted in the best results of measure R. Similarly, to Table 1, we observe that the MSE error measure behaves abnormally. In Table 2, the MSE error for the 2, 6 and 7 columns are comparable, and are in fact much worse than that for the results from columns 3 and 4. The explanation of this phenomena can be seen in Fig. 4. The MSE error in the latter cases hold minor values due to the average nature of the forecast (Fig. 4(a)). In cases more favourable, because of the measure R, we observe a much better match to the peaks of time (Fig. 4(b)).

Table 1: Results of PM_{10} prediction for τ =1, 2, 3, 5 ,6, 9, 12, 18, 24 h base on MLR.

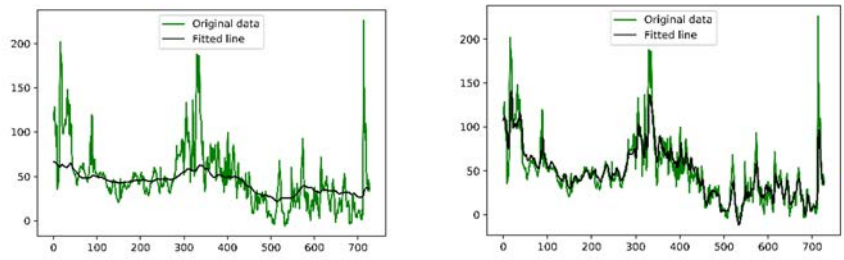
T	1	2	3	4	5	6	9	12	18	24
R l	0.85	0.87	0.84	0.79	0.80	0.79	0.75	0.75	0.73	0.73
R t	0.84	0.85	0.81	0.76	0.77	0.76	0.69	0.70	0.70	0.68
MSE l	0.8670	0.8468	0.8407	0.8342	0.8228	0.8724	0.8171	0.8356	0.8159	0.8670
MSE t	0.7510	0.7263	0.7538	0.7070	0.7235	0.7331	0.6919	0.7428	0.6245	0.7510

Table 2: Results of PM_{10} prediction for τ =1h base on MLR and MLRR.

Λ	0.0	0.01	0.001	0.0001	0.00001	0.000001
R l	0.85	0.77	0.82	0.90	**0.91**	**0.91**
R t	0.84	0.72	0.78	0.88	**0.90**	**0.90**
MSE l	0.9460	0.5574	0.7306	0.8468	0.9180	0.9615
MSE t	0.8872	0.4828	0.6259	0.7249	0.7868	0.7978

Table 3: Results obtained of PM_{10} prediction for τ=24h base on MLR and MLRR.

λ	0.0	0.01	0.001	0.0001	0.00001	0.000001
R l	0.73	0.68	0.71	0.75	**0.75**	0.74
R t	0.68	0.64	0.67	0.71	**0.72**	**0.72**
MSE l	0.8159	0.5471	0.6877	0.7644	0.7850	0.8121
MSE t	0.6245	0.3838	0.5076	0.5956	0.6272	0.6465

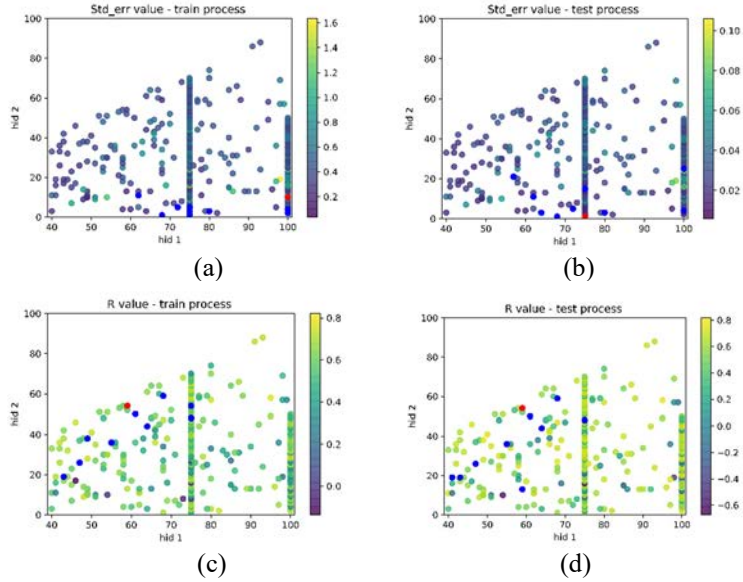


(a) (b)

Figure 4: Results of PM_{10} prediction for τ =1h based on linear multi regression model with L2 regularization (MLRR). (a) For λ=0.01; (b) For λ=0.00001.

The next model considered in this study is a linear neural network. The first step of such research is to construct a convenient neural network topological structure. In this case, we can presuppose that an independent neural network would be used for each time step prediction. Due to the assumed form of the x (4) input vector, as well as the suppositional predictor form, the number of neurons in the input layer is equal to 150 and is equal to 1 in the output.

In this research task, appropriate hid1 and hid2 values (the number of neurons in the hidden layers) were to be determined. This selection was based partly on setting certain hid1 parameters (75 and 100) and then searching for solutions for diverse hid2 values and based partly on a random space search over a number of weights. Fig. 5 shows the search results (Figs 5(a) and 5(c)) using a training (learning) sample, while (Figs 5(b) and 5(d)) reveals the



(a) (b)

(c) (d)

Figure 5: Evaluation of data sets with respect to number of hidden neurons in NN's layers. (a) Based on SEM for learning set; (b) Based on SEM for testing set; (c) Based on R correlation measure for learning set; (d) Based on R correlation measure for testing set.

effect of employing a test sample. The first two illustrations show the dependence of the standard error of the mean (SEM) for any given set. Here, the top 10 topological solutions are marked in blue, and in red, the best SEM result. In Figs 5(c) and 5(d), a similar designation was used, except that, in this case, the best result means the value of measure R close to 1.

Again, one can see that the measure associated with the error, this time SME, indicates that this was not a good solution, because in the case of the training set, a completely different solution was obtained than that for the test set. Thus, it must be underlined that the results of the search for a convenient topological structure with the use of the R measure are completely different. In this case, we can see that the best solution for both training and testing sets is the type structure (150-59-54-1).

Fig. 6 shows the time series representing this solution. The first two Figs 6(a) and 6(b) refer to the results obtained on the basis of the training set, while Figs 6(c) and 6(d) show the aforementioned for the test set. Figs 6(a) and 6(c) reveal a typical regression problem relationship, namely, the dependence of the desired model output and the actual output. The red line indicates the perfect solution, i.e. that for which the MSE error is zero and the measure R = 1. It can be seen from the graphs that the modelled prediction overestimates and not underestimates the PM_{10} dust concentration value. Still, for values up to 150 [$\mu g/m^3$], the prediction stays in the strip in line with the red line, but for values above this, it can be seen that the predictions are significantly lower. The above fact is related to the insensitivity of prediction to the rapid jumps in the concentration of pollutants that are characterized by a

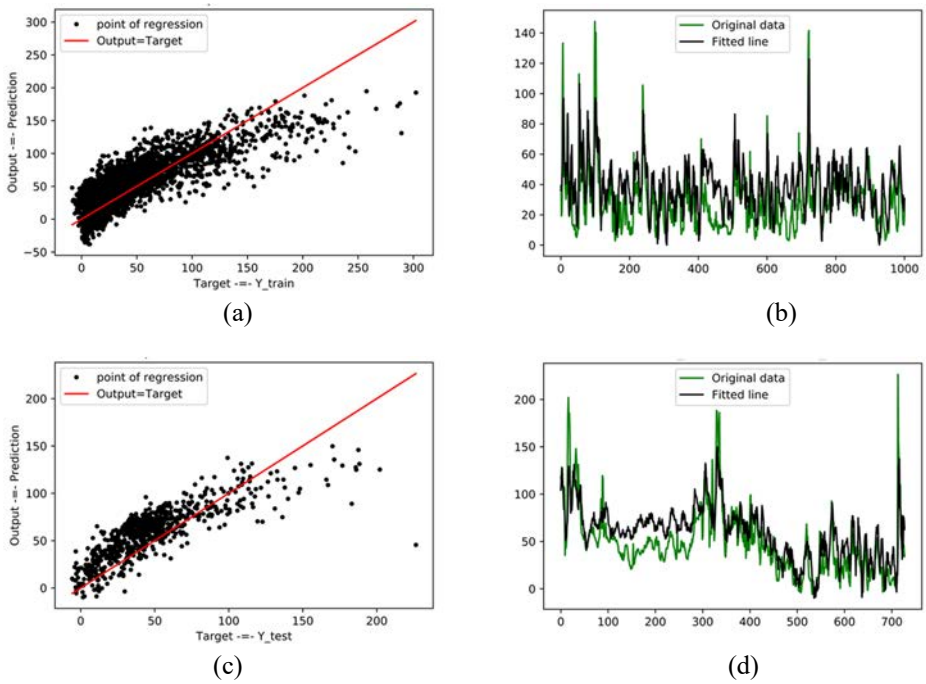


Figure 6: Results of prediction for τ =1, obtained by the way of use linear NN (hid1=59 hid2=54). (a) Results of correlation based on learning set; (b) Prediction for learning data set; (c) Results of correlation based on testing set; (d) Prediction for testing data set.

very high amplitude. The obtained neural solution is notable for having an error MSE = 0.9164, as well as a R = 82.2 for the learning data set, similarly, MSE = 0.7579 and R = 81.8 for the testing data set.

In summarizing this research, it should be noted that in all of the results under consideration, some prediction imperfection can be seen, in particular, in the points for which data continuity was not guaranteed.

6 CONCLUSIONS

The subject of the presented research is the creation and comparison of several algorithms so as to predict dust-type air pollution levels. The proposed models have been trained and tested using real data from the Airly's sensor network. These included measurements made every hour for one year, in one of the villages near Cracow. All procedures proposed in the article have been positively verified numerically. The first of the proposed methods was based on the multiple linear regression model. This model is characterized by the simplicity and lack of any additional internal parameters. In the next considered model, a penalty function for too large weighting factors was applied. In this case, as demonstrated in the research, the algorithm was sensitive enough to determine the appropriate value of the factor, the task of which is to indicate the proportion between the error of MSE and the function of the penalty. The last of the models described here was a linear neural network. In this model, the selection of the appropriate network topology was very important. The obtained results indicate that the method based on the Multiple Linear Regression with Regularization with parameter lambda 0.0001 is the best of the prediction procedures.

Further research will be related to the search for better neural models with non-linear transfer functions [18]–[20]. These studies will also include considerations regarding the significance of particular elements of the feature vector in the prediction issue [21], [22], but also a search for other coveted features that could have a beneficial effect on the quality of prediction of dust-type air pollution [23].

ACKNOWLEDGEMENT

This work was supported by The National Centre for Research and Development (project No. POIR.01.01.01-00-0049/17 NCBiR).

REFERENCES

[1] Domańska, D. & Wojtylak, M., Explorative forecasting of air pollution. *Atmospheric Environment*, **92**, pp. 19–30, 2014.

[2] Osowski, S. & Garanty, K., Forecasting of the daily meteorological pollution using wavelets and support vector machine. *Engineering Applications of Artificial Intelligence*, **20**(6), pp. 745–755, 2007.

[3] Benjamin, N., Sharma, S., Pendharker, U. & Shrivastava, J.K., Air quality prediction using artificial neural network. *International Journal of Chemical Studies*, **2**(4), pp. 7–9, 2014.

[4] Bottou, L., Large-scale machine learning with stochastic gradient descent. *Proceedings of COMPSTAT'2010*, pp. 177–186, 2010.

[5] Kingma, D.P. & Ba, J., Adam: A method for stochastic optimization. *ArXiv,* Preprint arXiv:1412.6980, 2014.

[6] Asghari, M. & Nematzadeh, H., Predicting air pollution in Tehran: Genetic algorithm and back propagation neural network. *Journal of AI and Data Mining*, **4**(1), pp. 49–54, 2016.

[7] Alkasassbeh, M., Sheta, A.F., Faris, H. & Turabieh, H., Prediction of PM_{10} and TSP air pollution parameters using artificial neural network autoregressive, external input models: a case study in salt, Jordan. *Middle-East Journal of Scientific Research*, **14**(7), pp. 999–1009, 2013.
[8] Russo, A., Lind, P.G., Raischel, F., Trigo, R. & Mendes, M., Neural network forecast of daily pollution concentration using optimal meteorological data at synoptic and local scales. *Atmospheric Pollution Research*, **6**(3), pp. 540–549, 2015.
[9] Domańska, D. & Wojtylak, M., Application of fuzzy time series models for forecasting pollution concentrations. *Expert Systems with Applications*, **39**(9), pp. 7673–7679, 2012.
[10] Domanska, D. & Wojtylak, M., Fuzzy weather forecast in forecasting pollution concentrations. *Proceedings of Chaotic Modeling and Simulation International Conference*, pp. 1–8, 2010.
[11] Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L. & Wang, J., Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment*, **107**, pp. 118–128, 2015.
[12] Baawain, M.S. & Al-Serihi, A.S., Systematic approach for the prediction of ground-level air pollution (around an industrial port) using an artificial neural network. *Aerosol and Air Quality Research*, **14**(1), pp. 124–134, 2014.
[13] Khadir, M.T. & Ghazi, S., Radial basis function and Elman networks for pollutant's parameter prediction in the region of Annaba Algeria. *Proceedings of the Fourth International Conference on Advanced Engineering Computing and Applications in Sciences,* Florence, Italy, pp. 26–33, 2010.
[14] Pasero, E. & Mesin, L., Artificial neural networks to forecast air pollution, 2010.
[15] Feng, Q. et al., Improving neural network prediction accuracy for PM_{10} individual air quality index pollution levels. *Environmental Engineering Science*, **30**(12), pp. 725–732, 2013.
[16] Ghazi, S. & Khadir, M.T., Recurrent neural network for multi-steps ahead prediction of PM_{10} concentration. *J. Automation & System Engineering*, **3**(2), pp. 13–21, 2009.
[17] Siwek, K. & Osowski, S., Improving the accuracy of prediction of PM_{10} pollution by the wavelet transformation and an ensemble of neural predictors. *Engineering Applications of Artificial Intelligence*, **25**(6), pp. 1246–1258, 2012.
[18] Barai, S.V., Dikshit, A.K. & Sharma, S., Neural network models for air quality prediction: a comparative study. *Soft Computing in Industrial Applications*, Springer: Berlin, Heidelberg, pp. 290–305, 2007.
[19] Kowalski, P.A. & Kusy, M., Sensitivity analysis for probabilistic neural network structure reduction. *IEEE Transactions on Neural Networks and Learning Systems*, 2017. DOI: 10.1109/TNNLS.2017.2688482.
[20] Kowalski, P.A. & Kulczycki, P., Interval probabilistic neural network. *Neural Computing and Applications*, **28**(4), pp. 817–834, 2017.
[21] Elangasinghe, M.A., Singhal, N., Dirks, K.N. & Salmond, J.A., Development of an ANN–based air pollution forecasting system with explicit knowledge through sensitivity analysis. *Atmospheric Pollution Research*, **5**(4), pp. 696–708, 2014.
[22] Kowalski, P.A. & Kusy, M., Determining significance of input neurons for probabilistic neural network by sensitivity analysis procedure. *Computational Intelligence*, pp. 1–22, 2017. DOI: 10.1111/coin.12149.
[23] Russo, A., Raischel, F. & Lind, P.G., Air quality prediction using optimal neural networks with stochastic variables. *Atmospheric Environment*, **79**, pp. 822–830, 2013.