Perceptron Neural Network - Based Model Predicts Air Pollution

Primož Mlakar and Marija Božnar
Jožef Stefan Institute, Jamova 39, SI - 1000 Ljubljana, Slovenia

Abstract

SO_2 air pollution is still an important environmental problem in Slovenia, especially around big thermal power plants. A modern approach a Multilayer Perceptron neural network - based short term air pollution prediction model is explained. Neural network - based models are rarely used in the field of air pollution. The models were developed for the automatic air quality measuring stations around the Šoštanj Thermal Power Plant. An extensive data base is available for this site. It includes meteorological data, ambient concentrations and emission data for a four year period. A model that predicts SO_2 concentration for one averaging interval (half an hour) in advance is explained in detail. Development of the model should solve several problems, including selection of an appropriate structure (number of neurones), selection of features, patterns selection and determination of suitable training algorithm parameters. Results are very encouraging and show that the method is worthy of further research.

KEYWORDS: neural network, SO_2 air pollution, short - term prediction model, feature determination, pattern selection

1: Air Pollution

SO_2 air pollution is still one of the important environmental problems in Slovenia. In past years thermal power plants were built without efficient wet desulphurisation plants. Power plants mostly burn local coal that has a high percentage of sulphur. The resulting pollutant - SO_2 - is emitted into the atmosphere through stacks of different height, some of them very high, but the complex terrain surrounding them (as shown on Figure 1) in some meteorological situations prevents the plume from diluting enough and even brings it to the ground at short distances from the Thermal Power Plant.

Large thermal power plants like the Šoštanj one (as shown on Figure 2) have modern environmental information systems that measure the resulting air pollution. Local authorities compel the Šoštanj Power Plant to reduce power, if the ambient SO_2 or NO_x concentrations are too high. Reducing the power is certainly an economic problem for the Power Plant. Therefore an efficient short - term air pollution prediction model would help the staff to run the Power Plant in such a way to prevent high peaks of ambient concentrations with as little power reduction as possible.

For this purpose we developed a multilayer perceptron neural network - based model for short - term prediction of SO_2 concentrations for the stations around the Šoštanj Thermal Power Plant as an alternative to the stochastic models that were used in this field in the past. Our initial model was the first one published using this method (Božnar[3]).

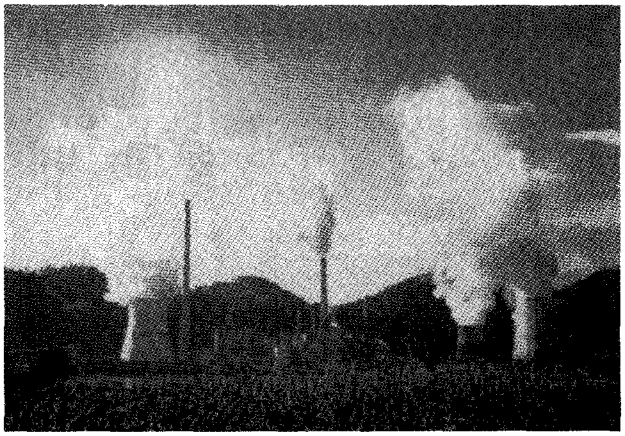

Figure 1. Šoštanj Thermal Power Plant

2: Environmental information system of the Šoštanj Thermal Power Plant and research data base

The Environmental information system of the Šoštanj Thermal Power Plant (EIS ŠTPP) consists of six fixed stations and one mobile automatic station in the TPP surroundings, an emission station in the TPP and a hydrological station. The stations in the surroundings are placed in the basin and on the near - by hills. They continually and fully automatically measure meteorological parameters (such as wind speed and direction on 10m mast (ground level), air temperature, relative humidity, air pressure and global solar radiation) and pollutant concentrations (SO_2, NO, NO_2, NO_x, O_3, CO) (not all the stations have a full configuration). The emission station measures air pollutant concentrations in the exhausted gases, gas velocity and some important power plant parameters. The hydrological station monitors the Paka river which is used for the cooling water supply.

The measurements are averaged every half an hour at the automatic stations and sent to the central unit in the Power Plant where the data base is formed. The data collected are distributed on - line to local authorities, other authorised institutions and to the local TV station. All the data are checked at two levels (at the station and in the central unit) to determine its validity.

The data base in the central unit of the EIS ŠTPP is formed from the information on the last 45 days. It includes all the measured parameters (half hour average values, extreme values with corresponding time, ...). From this data base a research data base was extracted in the period 1990 to 1994. The research data base includes only the average values and the validity of important parameters. For a short period in spring 1991 an extensive data base, suitable for detailed pollution analyses, is available, including SODAR and DIAL measurements (the data were distributed together with an extensive report, Lesjak[6]).

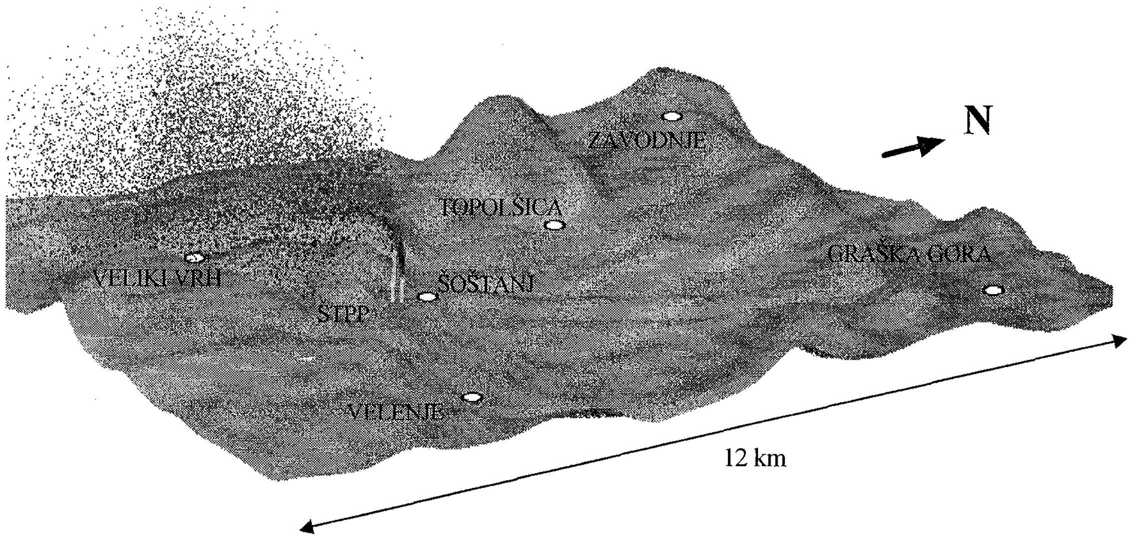

Figure 2. Locations of the automatic measuring stations of the Environmental Information System of the Šoštanj Thermal Power Plant

3: SO_2 air pollution prediction models

SO_2 air pollution at the automatic measuring stations around ŠTPP can be considered as a time series. Every half an hour new measurements are available. The aim of the desired model is to predict the SO_2 concentration for at least one averaging interval in advance for a particular station. As an input for the model we can use not only the SO_2 time series at the particular station, but also other time series can be included (SO_2 time series at other stations, time series of meteorological parameters and emission data).

In the past research in this field started according to the Box - Jenkins time series analysis theory. The first stochastic models were black box ARIMA or similar ones; later exogenous inputs were added to include meteorological or emission data (Finzi[1]).

In the perceptron neural network - based model that we developed earlier (Božnar[3], [5] Mlakar[8], [9], [10]), the data base was not dealt with as a multivariate time series analysis, but as a huge data base of patterns. For each half hour interval in the four year period a corresponding pattern can be constructed. A pattern is defined as an vector compounded of values for the input features (selected pollution and / or measured meteorological parameters for observed and previous already measured half hour intervals; usually not more than 2 or 3 consecutive intervals are taken, because older ones appear to be useless) and values for the output feature (for instance SO_2 concentration for one averaging interval in advance).

4: Multilayer Perceptron neural network - based prediction model

Our idea of using a multilayer perceptron neural network as a tool for air pollution prediction came from its ability to learn from set of historical patterns and its ability to generalise this knowledge to unknown patterns. If the learning set is compounded of a lot many patterns that represent typical possible meteorological and pollution situations in the observed area, the neural network - based model can extract knowledge from them (for learning patterns the values of the output feature should be known - supervised learning). After the training phase is completed, the model can be used for prediction of the output SO_2 concentration of patterns that were not used in the training process.

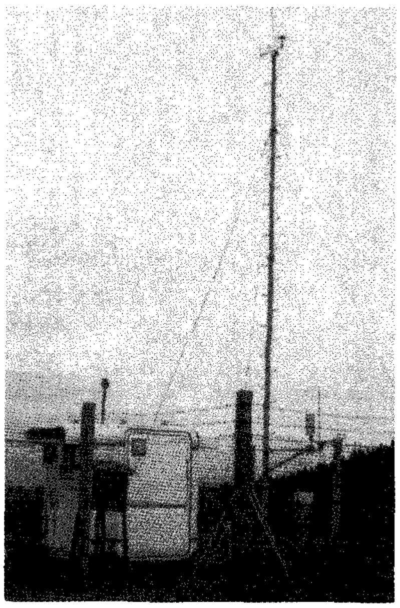

Figure 3. Typical automatic measuring station

4.1: Data base partitioning

The whole set of available patterns (universal set) should be divided into five independent sets, each of them having its own task in the process of training and testing the model. A training set is a set used for adjustment of the neural network's connection weights during the process of training the neural network with the backpropagation algorithm. A testing set is used for periodical testing of the neural network during the process of training to determine the optimal network. A production set of data is the set with which a constructed model is tested in order to determine its performance. If a possibility exists to test the model in real time, then the patterns constructed in this way form the on-line set. Patterns from the universal set of data which do not belong to any of the previously defined sets form the remaining set. The common name for a union of the training and testing set is the learning set. The patterns in the learning set are certainly the ones most responsible for the performance of the model. The production set is responsible for the assessment of the model's performance.

4.2: The structure of the model

For our research we selected a perceptron neural network with two hidden layers (Rumelhart[11]), which has been proven to be a universal aproksimator (Hornik[2]). Some of the experiments were also done with one hidden layer having a similar number of neurones as both hidden layers together. The results did not differ significantly.

For each automatic measuring station a separate model was built, because the air pollution dispersion mechanisms that result in high concentrations are different for different stations. Table 1 shows a typical model for the station in the surroundings of the TPP. We built models for prediction of one averaging interval (half hour) in advance. Prediction for more averaging intervals in advance is certainly a more difficult task.

4.3: Training algorithm

For neural network training, a backpropagation algorithm (Rumelhart[11]) was used. The most suitable learning rate and momentum parameters were determined experimentally. A learning rate of 0.6 and a momentum of 0.9 were mostly used. A relatively high learning rate ensures rapid finding of the error function minimum, and a high momentum prevents too many oscillations of the error function. If lower learning rate is used (0.1), much more time is used for the training process, but the results can be slightly better. The model was usually trained with 100,000 to 300,000 learning events (change of neural network weights according to one pattern error), but the optimal neural network was usually found before or at about 100,000 events.

5: Problems that should be solved to construct good models

The first problem that should be solved is the determination of features. Feature determination is a well known problem in pattern recognition theory. Not all the measurements available from the EIS ŠTPP are suitable for use as an input for the forecasting model. Several feature selection methods were tested for their efficiency on a huge data base.

These methods are (Mlakar[10] - this conference, [7]):

- feature selection using scatter diagrams combined with non-linear regression analysis,
- feature selection using saliency metrics,
- feature extraction using a Kohonen neural network,
- sequential forward selection and sequential backward selection of features.

Some of the methods are our own, and others are taken from the literature and adapted for air pollution prediction problems.

Because of the huge data base (over 70,000 measuring intervals) unique methods for pattern selection were developed that lead to significantly improved results for SO_2 prediction. These methods are (Božnar[5] - this conference):

- pattern selection using meteorological analysis,
- pattern selection using a Kohonen neural network and
- multi - type model strategy.

The methods show, that the structure of the learning set is much more important than the number of patterns. Even with a relatively small number of carefully selected learning patterns (about 500 to 1000) we obtained better models than with up to 6000 unselected learning patterns.

The number of the hidden neurones was determined experimentally. If it is too large, the model learns training patterns in detail and looses itsgeneralising capabilities; if it is too small, the model does not learn enough. A typical value is seen from Table 1.

INPUTS FEATURES:		HIDDEN NEURONES:	OUTPUT FEATURE:
Šoštanj, wind direct. in deg. (t)	sin(Veliki vrh wind dir. (t))	40 + 38	Šoštanj SO_2 (t+1/2h)
Šoštanj, wind speed (t-1/2h)	Zavodnje, SO_2(t) - SO_2(t-1/2h)		
Šoštanj, wind speed (t)	Zavodnje, wind direct. - TEŠ direction (t)		
Topolšica, SO_2 (t)	daily half hour interval		
Zavodnje, SO_2 (t-1/2h)	cos(Graška gora wind dir. (t))		
Zavodnje, SO_2 (t)	sin(Graška gora wind dir. (t))		
temperature diference (Zavodnje(t)-Graška gora(t))	sin(Graška gora wind direction (t-1/2h))		
cos(Šoštanj wind dir. (t-1/2))	SO_2 emission TEŠ 123 (t)		
sin(Veliki vrh wind dir. (t-1/2))	SO_2 emission TEŠ 123 (t-1/2h)		
cos(Veliki vrh wind dir. (t))	Graška gora, wind speed (t)		

Table 1. Features and structure of the neural network model for the Zavodnje automatic measuring station

6: Results

The performances of all the models were evaluated on independent data (production sets) over a long period. A typical model performance is shown in Figure 4. The high concentrations rise and fall very quickly and usually last for no more than a few half hour intervals (Božnar[4]). Only about 5% of the half hour intervals at the more polluted stations have high concentrations. But because they can be very high (more than 2 mg/m^3), they are also very harmful to people and vegetation. For this reason, it is very important to predict very high concentrations properly and not to make false alarms when the actual concentrations (measured later) are low. Therefore the correlation coefficient, mean square error and similar measures are not suitable for performance evaluation of the model, since they mostly stress the model's performance in the intervals with low ("not interesting") concentrations. Therefore we designed a new measure termed p^6 (Mlakar[10] - this conference) - the probability of successful prediction of high concentrations. It is defined as the number of intervals with successfully predicted high concentrations (actual concentration $\geq$ 0.15 mg/m^3 and the absolute error < 0.1 mg/m^3 or relative error < 0.2) divided by the sum of the number of intervals with high concentrations and the number of intervals with false alarms (a false alarm occurs when the actual concentration is low and the predicted concentration $\geq$ 0.25 mg/m^3). (As an example of the measure: if p^6=0.40, with 44 high concentrations, 6 false alarms, all together 954 production patterns, this means that the concentration was well predicted for 96.85% of all intervals).

The average values of p^6 obtained (several models tested over large production sets from 1000 to almost 7000 production patterns) vary very much from one measuring site to another. For the Šoštanj automatic measuring station, which is situated near the power plant and is effected mostly by the down - wash effect, typical results for p^6 were not greater than 0.3, because the meteorological measurements there are not extensive enough. But for the Zavodnje station, which lies on the slope of a hill about 7km from the TPP, the results for the p^6 were much better, even exceeding 0.65. At both stations the p^6 values can be improved by taking additional more sophisticated meteorological measurements as input features. The vertical wind profile would help a lot, because it would be easier to estimated the proper wind at the effective stack height. Also the vertical temperature profile would be usefull for better knowledge about the stability of the atmosphere. On the other hand models prediction capability can be improved to some level by changes of neural network structure and neurones characteristics. This is both planed for further research.

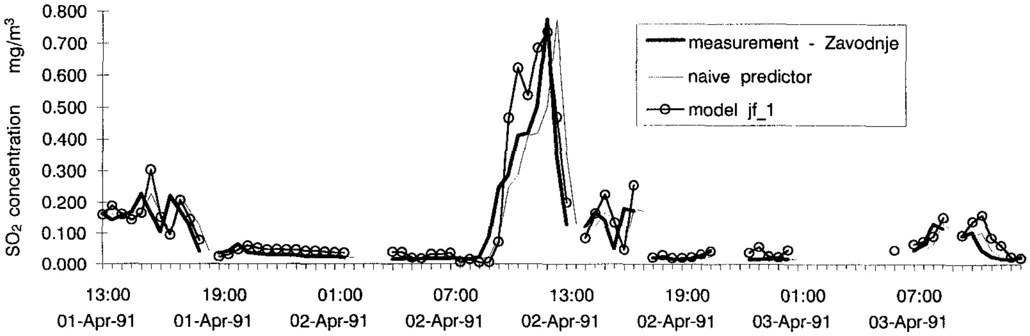

Figure 4. Typical results of the SO_2 short - term prediction model

7: Conclusions

The multilayer perceptron neural network - based short-term air pollution prediction model represents an alternative to ARIMA type stochastic models. Results of the models predicting SO_2 pollution for the automatic measuring stations around the Šoštanj Thermal Power Plant are very encouraging and certainly worth further research. Neural network - based models are capable of learning sophisticated air pollution dispersion mechanisms from basic meteorological and air pollution observations, and use the knowledge obtained to predict air pollution for situations not used in the training process.

8: References

[1] G. Finzi, G. Fronza, A. Spirito, *Multivariate stochastic models of sulphur dioxide pollution in an urban area*, APCA Journal, vol. 30, 1212-1215, (1980).

[2] K. Hornik, *Approximation capabilities of multilayer feedforward networks*, Neural Networks, vol. 4, 251-257, (1991).

[3] M. Božnar, M. Lesjak, P. Mlakar, *A Neural Network-Based Method for Short-Term Predictions of Ambient SO_2 Concentrations in Highly Polluted Industrial Areas of Complex Terrain*, Atmospheric Environment, 27B, pp.221-230, (1993).

[4] M. Božnar, P. Mlakar, *Analysis of ambient SO_2 concentrations in the surroundings of the Šoštanj thermal power plant*, 4th International Conference on Air pollution, Toulouse, 1996, Air Pollution IV, Monitoring, Simulation & Control / eds. B. Caussadle, H. Power, C. A. Brebbia, Southampton, Boston, Computational Mechanics Publ. pp.727-734, (1996).

[5] M. Božnar, *Pattern selection strategies for a neural network-based short term air pollution prediction model*, International Conference on Inteligent Information System, IASTED, Bahamas (this conference)

[6] M. Lesjak, M. Božnar, P. Mlakar, F. Slavič, et al., *Experimental Campaign for the Environmental Impact Evaluation of Šoštanj Thermal Power Plant*, Progress Report, IJS, ENEL, CISE (1992) (can be obtained from: A. Negri, CISE s.p.a., Via Reggio Emilia 39, c.p. 12081, 20134 Milano, Italy)

[7] P. Mlakar, M. Božnar, *Analysis of winds and SO_2 concentrations in complex terrain*, 4th International Conference on Air Pollution, Toulouse, 1996, Air Pollution IV, Monitoring, Simulation & Control / eds. B. Caussadle, H. Power, C. A. Brebbia, Southampton, Boston, Computational Mechanics Publ. pp.455-462, (1996).

[8] P. Mlakar, M. Božnar, M. Lesjak, *Neural Networks Predict Pollution, 20th International Technical Meeting on Air Pollution Modelling and Its Application*, Valencia, 1993, Proceedings.Volume 3, pp.531-532, (1993).

[9] P. Mlakar, M. Božnar, *Short-term air pollution prediction on the basis of artificial neural networks*, 2nd International Conference on Air pollution, Barcelona, 1994, Air Pollution II.Volume 1, Computer Simulation / eds.J.M. Baldasano et al., Southampton, Boston, Computational Mechanics Publ., pp.545-552, (1994).

[10] P. Mlakar, Determination of Features for Air Pollution Forecasting Models (this conference)

[11] Rumelhart D. E., McClelland J. L. *Parallel Distributed Processing 1,2*, MIT Press, Cambridge, MA, (1986).