Using Preprocessing Techniques in Air Quality forecasting with Artificial Neural Networks

Conference Paper · April 2009

DOI: 10.1007/978-3-540-88351-7_27 · Source: DBLP

3 authors:

Ioannis Kyriakidis
Hellenic Mediterranean University

Kostas Karatzas
Aristotle University of Thessaloniki

Giorgos Papadourakis
Hellenic Mediterranean University

Using Preprocessing Techniques in Air Quality forecasting with Artificial Neural Networks

Ioannis Kyriakidis [1,2], Kostas D. Karatzas [3] and George Papadourakis [1]

[1] Department of Applied Informatics & Multimedia, T.E.I. of Crete, Estavromenos GR-71004, Heraklion, Crete, Greece
(e-mail: kyriakidis@teicrete.gr ; papadour@cs.teicrete.gr)

[2] University of Glamorgan, School of Computing, Pontypridd, Wales, United Kingdom CF37 1DL

[3] Department of Mechanical Engineering, Aristotle University, GR-54124 Thessaloniki, Greece
(e-mail: kkara@eng.auth.gr)

Abstract

Data quality is one of the fundamental issues influencing the performance of any data investigation algorithm. Poor data quality always leads to poor quality results. In the investigation chain, the data selection phase is followed by the preprocessing phase, which results in increased data quality, while in parallel it demands the highest time resources of the overall data investigation chain. The preprocessing phase includes the handling of missing data, handling of the outliers, data de-trending and data smoothing. The methods that are used in the preprocessing phase are usually not sufficiently reported in the literature of environmental data analysis and knowledge extraction. The current paper investigates the performance of several methods in all phases of the preprocessing chain of environmental data, by emphasizing in the use of ICT (Information & Communication Technology) methods for the materialization of such preprocessing tasks, and by making use of the air quality as the environmental domain paradigm.

I.N. Athanasiadis et al., *Information Technologies in Environmental Engineering*,
Environmental Science and Engineering, DOI 10.1007/978-3-540-88351-7_27,

1. Introduction

As air quality has a direct impact on the quality of life and on the environment, it is essential to understand the parameters that contribute to air pollution, the behaviour of air pollutants, their inter-relationships and dependencies, and to be able to extract knowledge and to forecast their future levels. The latter is among the keystones of contemporary air pollution legislation, a paramount example of which is the "Clean Air for Europe" Directive 2008/50/EC.

The need for accurate and operationally effective air quality models has been extensively discussed, along with the complexity of the atmospheric quality from the scientific point of view [1]. Among the categories of models that have been applied in air pollution problems is the one based on data mining algorithms [2]. In data mining applications, most of the efforts are directed towards the data pre-processing phase [3]. Data pre-processing covers all steps that lead to the final dataset which will be explored with the aid of data mining algorithms. The current paper investigates the performance of several pre-processing methods, in an effort to achieve modelling results of better quality. The pre-processing phase includes the following two steps:

- Identifying and removing of the outliers. This step is a prerequisite in order for (a) smoothing functions and (b) replacing missing values functions, not to be affected by outliers.
- Smoothing is the next step because it can eliminate noise. If noise is not removed, replacing missing values functions may be affected.

As it was necessary to evaluate the performance of the various prepossessing methods investigated, it was decided to apply one of the well established methods for AQ modelling. For this reason, Artificial Neural Networks (ANNs) have been employed in the present paper [4] (there are of course other methods that may serve the same modelling purpose like regression trees [4], self organising maps (SOM) [5], K-means clustering [6] and others). An ANN is a mathematical representation of the operation of biological neural networks, in other words it is an emulation of a biological neural system. ANNs are highly adaptive to non-parametric data distributions and make no prior hypotheses about the relationships between the variables. ANNs are also less sensitive to error term assumptions and they can tolerate noise, chaotic components and heavy tails better than most of the others computational methods. An ANN consists of an input layer, one or more layers of neurons (mathematical entities whose behaviour is governed by a predefined function) and an output layer. Each ANN has to be trained first, by using a number of data instances as inputs, and by trying to "fit" to some other data instances (output), commonly in terms of predicting the (future) behaviour of parameters of interest with the aid of their history. In this paper the ANNs models were developed in order to forecast the nitrogen dioxide (NO2) concentration values [7-9].

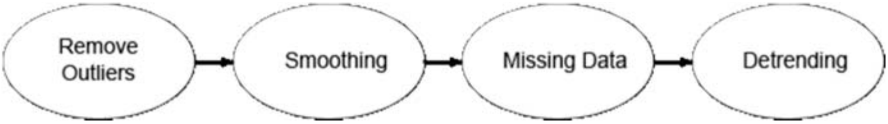

Figure 1: Pre-processing Steps

2. Area and data presentation

The data used in this study were obtained from six monitoring stations operating in the Thessaloniki area, Greece, and were taken by the public air pollution database of the European Environment Agency [10]. These stations are located at the following locations: (1) Eleftherio Kordelio, (2) Kalamaria, (3) Panorama, (4) Sindos, (5) Ag. Sofia's Square and (6) AUTH (Aristotle University main campus). Thessaloniki is the second largest city of Greece, where air emissions come mainly from traffic, while formation and transport of pollutants is heavily influenced by the local meteorological and topographical characteristics (Fig 2.) [11]. Tables I and II present the investigated parameters per monitoring station location. Data used correspond to the years 2001 up to 2003 for all stations. It should be mentioned that the selection of the specific three year period was made taking into account the large number of missing values for the years 2004-2006 which were also investigated.

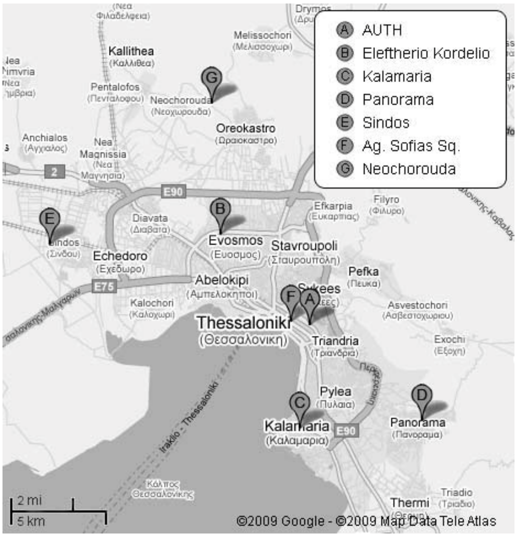

Figure 2. The seven air quality monitoring stations of the Thessaloniki area (Panorama station data are not used in this paper).

Table I. Pollutant concentration values investigated per monitoring site

Location	Eleftherio Kordelio	Kalamaria	Panorama	Sindos	Ag.Sofias	Sq.AUTH
NO_2	✓	✓	✓	✓	✓	✓
CO	✓	✓	-	✓	✓	-
SO_2	✓	✓	-	✓	✓	✓
O_3	-	-	✓	✓	✓	✓
PM_{10}	✓	-	✓	✓	✓	-

The meteorological data used (Table II) are measured at the airport of Thessaloniki and were acquired via the weather underground web site [12] As the majority of these data consisted of nominal values, it was necessary to transform nominal values to numeric (Table II), in order to have a common data category reference for the data pre-processing phase [13],. A parameter that required additional retransformation was wind direction, for which the intention was to replace the cyclic nature of this variable with a linear one (Eq. 1-2). For this reason, the sine transfer function was also applied in this case (Eq. 3) because it may lead to better results [14].

Table II. Meteorological parameters employed

Arithmetic Values	transformed to numeric	Logical Values	transformed to numeric
Temperature C	-	Wind Variable	✓
Dew Point C	-	Fog	✓
Humidity	-	Rain	✓
Sea Level Pressure hPa	-	Snow	✓
Visibility Km	-	Thunderstorm	✓
Wind Direction	-	Tornado	✓
Wind Speed Km/h	-	Shower	✓
Rain Intensity	✓	Freezing	✓
Snow Intensity	✓	Funnel	✓
Thunderstorm Intensity	✓	Unknown	✓
Cloudy	✓		
WD	✓		
sinWD	✓		
cosWD	✓		

$$\sin WD = \frac{\sin(2\pi(\nu - \min(\nu)))}{\max(\nu) - \min(\nu)} \tag{1}$$

$$\cos WD = \frac{\cos(2\pi(\nu - \min(\nu)))}{\max(\nu) - \min(\nu)} \tag{2}$$

$$WD = 1 + \sin(\theta + \pi/4) \tag{3}$$

3. Outliers

The first step in the pre-processing phase is the handling of the outliers. Outliers are data patterns that deviate substantially from the data variation. Outliers result in large errors and consequently large weight updates because of the large deviation from the norm. They may be attributed to measurement error or they may represent a significant feature within the investigated data. Identifying outliers and deciding what to do with them depends on the understanding of the data, information on their source and knowledge concerning the scientific domain they represent and the range of values that the associated parameters are expected to have.

Detecting the presence of outliers is very important because they have a strong influence on the estimates of the data model parameters that are been fitted. This could lead to wrong scientific conclusions and inaccurate predictions. Fig. 3 demonstrates how the (fitted) function is pulled towards the outlier in an attempt to reduce the training error. As a result, the generalization capacity of any model being developed to describe the data deteriorates [15].

Visual aids (Dot Plots and Box Plots) may be used to identify outliers. Box plots are useful graphical displays for describing the behavior of the data (Fig. 4), and may help in the qualitative identification of outliers. On the other hand, a common method that may be applied in order to quantitatively identifying outliers is to look for values located more than a certain number of standard deviations away from the mean [16]. Although there is no general rule, a distance of two up to ten standard deviations above the mean may be applied in most AQ data series, as demonstrated in Fig 5.

Alternatively, robust objective functions [17] may be used, as they are not influenced by outliers. Robust objective functions find the best fitting structure concerning the majority of the data. A robust objective function may also be able to identify the outlier's substructures for further treatment [18]. The most common general method of robust objective function is regression where the mostly used is the "m-Estimator" introduced by Huber (1964) [19] (there are additional regression-based algorithms like the bisquare, [20], and the method proposed by Andrews [21]).

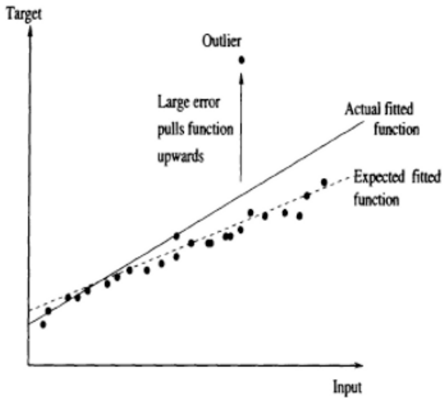

Figure 3. Illustration of the way that outliers affect the fitted function.

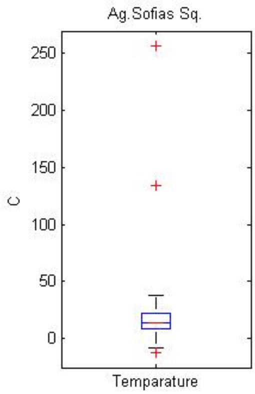

Figure 4. Example of a Box plot for Temperature (at Ag. Sofia's monitoring station)

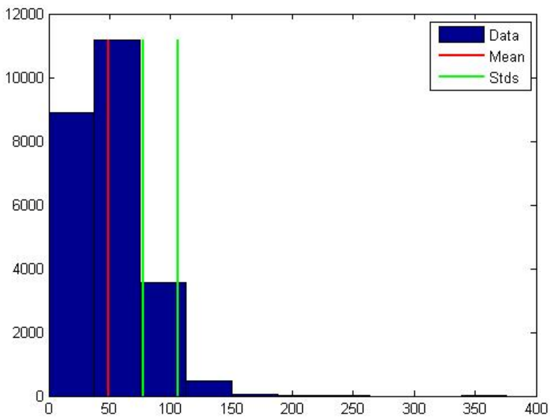

Figure 5. Example of Temperature STD (at Ag. Sofia's monitoring station)

In order to remove the outliers in the non-transformed atmospheric quality data, the bisquare robust function was applied in this paper. The values that do not contribute to the best robust fit have a zero weight. Those are treated as outliers and are removed from the time-series. Fig.6 presents the results obtained by applying the robust bisquare function on the CO time-series from the Ag. Sofia's Square station (original data in Fig. 7). In addition, Fig.8 presents the results obtained for the same pollutant by using the standard deviation criterion, i.e. removing all values that are more that 2 STD away from the mean. The robust function seems to identify outliers better than the specific STD criterion, as it does not cut off a considerable amount of high values. Overall, it should be noted that the removal of outliers requires a deep knowledge of the specific domain of investigation (here air pollution), and in many cases its characteristics for the geographic area of interest (here Thessaloniki), and should be done on the basis of adequate knowledge of the information that it taken our of the data set when outliers are removed.

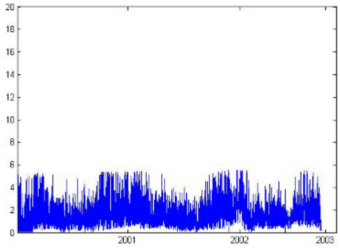

Figure 6. Results by using the robust bisquare function of the CO time-series.

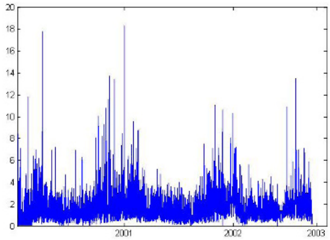

Figure 7. Original CO time-series (including outliers) of Ag. Sofia's Sq. Station.

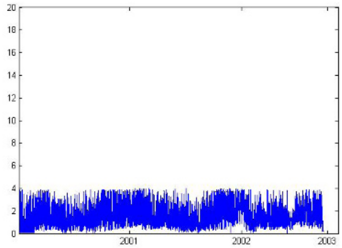

Figure 8. Results by using the standard deviation function of the CO time-series.

4. Smoothing

The next step after removing the outliers from the studied data set is "Smoothing". Data smoothing techniques are used to eliminate "noise" and extract real trends and patterns. They provide with a clearer view of the behaviour of the time series studied. Thus for example, economists use smoothing techniques to reveal economic trends in data [22]. Smoothing may also deal with missing values, in cases that they are not of high percentage.

In some cases, seasonal variation are so strong that do not allow for any trend or periodicity indications, aspects of high importance for the understanding of the process being studied. Smoothing can remove seasonality and may help in revealing lagged fluctuations within the data set [23]. The most common type of smoothing technique is Moving Average smoothing. Another smoothing method applied is the Savitzky-Golay filter method as well as a robust version of a local regression method [24].The robust local regression is using weighted linear least squares and a 1st degree polynomial mode. The application of a robust version of local regression may help in dealing with outliers and smoothing at the same time. The Savitzky-Golay filter method performs a local polynomial regression around each point, and creates a new, smoothed value for each data point. Specifically for each point fi, a least-square fit of a polynomial of given degree (2nd degree is used here) within all points in the smoothing window is generated. The smoothed value, gi is the value of the smoothing polynomial at positions i.

Figures 9 and 10 present a sample of the data after the removal of the outliers (original) and the smoothed data that were obtained by applying the Moving Average, the Savitzky-Golay filter and robust local regression methods. Fig. 9 presents the smoothed data where the STD function was used to remove outliers while Fig 10 presents the smoothed data resulting after the application of the robust objective function in order to remove outliers. The outliers are identified as gaps in the figures. It is evident that the robust local regression method is not capable of dealing with missing values. If this method is going to be selected for data smoothing, missing values should be addressed before. All three smoothing methods provide with better smoothing results when the data was first cleared by the outliers by applying the robust function.

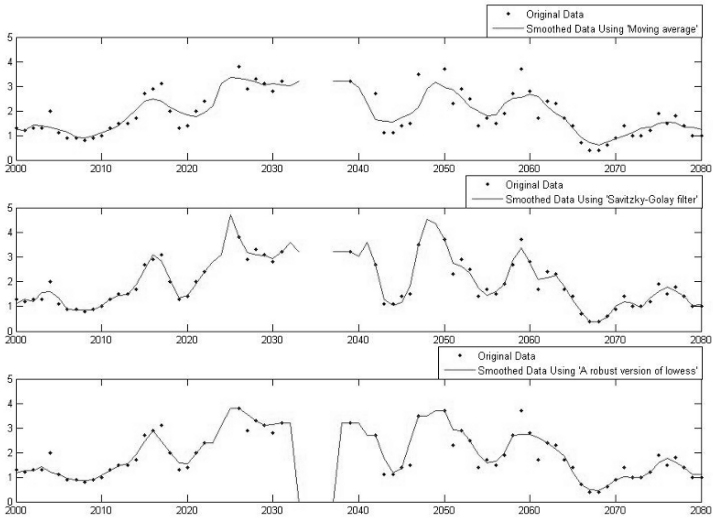

Figure 9. Smoothed sample data. Outliers have already been removed with the aid of the standard deviation function.

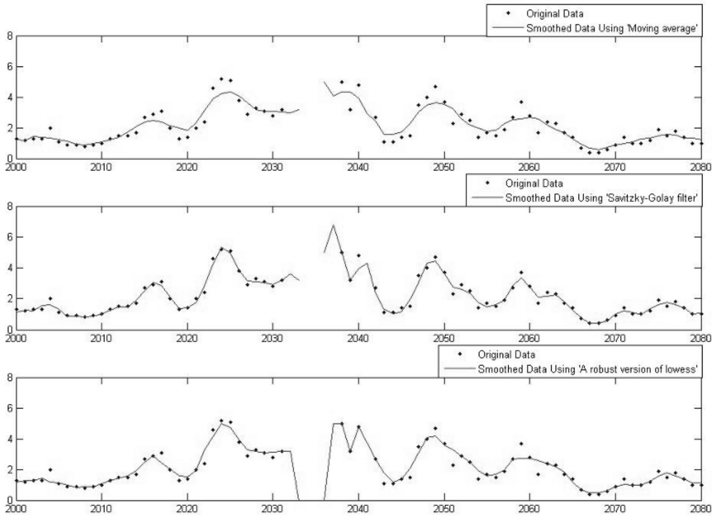

Figure 10. Smoothed sample data Outliers have already been removed with the aid of the robust objective function.

5. Missing Values

It is common in air quality data to have missing values, because of a failure of the monitoring station, or other exogenous reasons. Some algorithms (e.g. Naïve Bayes, CART, Bayes Tree, CN2) are capable of dealing with missing values [25], whereas others require that the missing values should be replaced or removed. While pattern removal solves the missing values problem, important information may be lost and the available information for training is reduced, which may be a real problem if the data is already of limited volume.

Missing or erroneous data complicates the modeling of large data sets, because of the fact that most data processing methods require for complete data metrics before performing any further analysis [26].

In order to get comparable results in the present study, the rule for replacing all missing values was followed. For this reason, interpolation algorithms were employed. There are several interpolation methods that deal with missing values (Linear, Piecewise cubic Hermite, Cubic spline, Nearest neighbor, etc.). As the aim was to investigate the performance of those methods in air quality data, a small experiment was performed. A time series with *no* missing values was used, comprising of hourly values. In this series, 8.5% of the data in the first step, and 12.5% of the data in a next step were removed to imitate a missing values case. The missing values were imputed to the data set, on the basis of the missing values of an air quality time series. The result was that missing values in some cases were spread within the whole data set, while in some other cases they occupied a block within the studied data, with more that 100 hourly missing records.

For the purposes of this study four interpolation methods were evaluated concerning their performance and efficiency in replacing missing values. Table III consolidates the performance results of these interpolation methods. These results suggest that linear interpolation methods have the best performance, in terms of minimum error and higher correlation. On this basis, the linear interpolation method was selected to be applied to replace the missing values in the studied data set. To keep this phase less complex from the computational point of view, no multivariate methods are used [27], since the interpolation method (univariate) by itself lead to 98.9% correlation in the data where the percentage of the actual missing values was 12.5%.

Table III. The performance results of the different interpolation methods.

	Method	Correl. Coef.	MAE	MAPE	MBE	RMSE	RMSE Syst.	RMSE Unsys.	IA
8.5 %	Linear	0.996	0.204	0.009	-0.08	1.464	0.011	0.29	0.995
	Cubic	0.995	0.223	0.01	-0.107	1.755	0.013	0.29	0.997
	Spline	0.813	0.929	0.045	-0.816	11.892	0.091	0.303	0.892
	Nearest	0.994	0.318	0.012	-0.073	1.835	0.014	0.289	0.997
12.5 %	Linear	0.989	0.470	0.019	-0.097	2.508	0.019	0.286	0.995
	Cubic	0.988	0.488	0.019	-0.094	2.645	0.02	0.286	0.994
	Spline	0.792	1.473	0.065	-0.709	12.687	0.097	0.299	0.878
	Nearest	0.985	0.627	0.022	-0.09	2.968	0.023	0.285	0.992

6. Detrending

Trend in a time series is a slow, gradual change in some property of the data over the whole time interval under investigation. Trend is sometimes loosely defined as a long term change in the mean, but can also refer to a change in other statistical properties. Detrending is the mathematical operation of removing trends from the data under investigation. Detrending is often applied to remove a feature thought to distort or obscure the relationships of interest. Thus for example, a temperature trend due to urban warming might obscure a relationship between cloudiness and air temperature. Detrending is also used as a pre-processing step to prepare time series for analysis by methods that assume stationarity [28]. Detrending might improve computational accuracy when the signals vary around a large signal level. In order to detrend, constant and straight like detrending methods may be employed. Constant detrending removes the mean of the data to create zero mean data. Straight line detrending finds linear trends (in the least-squared sense) and the removes them.

7. The Artificial Neural Network application

In the frame of the current study, Artificial Neural Networks (ANNs) were employed in order to develop an AQ model for the forecasting of nitrogen dioxide (NO2) concentration levels values and thus to investigate which combinations of the previous pre-processing methods lead to better forecasting results. The type of the ANN applied was the multilayer perceptron using the backpropagation training algorithm, as this is a common method that leads to good forecasting performance in similar air quality related applications [29].

Data were firstly normalized by the hyperbolic tangent sigmoid transfer function in order to be imported to the ANN. The same transfer function was used for the hidden layer. On the output layer the linear transfer function was used. This is a common structure in function approximation problems [30]. For the training phase, the Levenberg-Marquardt backpropagation algorithm [31] was implemented to the fifty percent of the data while the other fifty percent was split to the validation and testing phases.

The combinations of the all methods led to 72 different models that were implemented and tested during the training phase in order to avoid local minima [32]. For reasons of comparability, 6 additional models are implemented were the data received no pre-processing. In all of those models one hidden layer was used. The number of neurons in the hidden layer was the same as the number of the parameters per monitoring station to which the models refer to. An increase of the hidden layers or the neurons used was proven to be unnecessary, as the purpose of this paper is to investigate the performance of different pre-processing methods and their combination, rather than the absolute forecasting performance of the ANN.

8. Results and Discussion

Table IV represents the best three results concerning the performance of the ANNs per station location, as quantified with the aid of the correlation coefficient and the index of agreement [33]. It should be noted that the linear interpolation method (used to replace missing values) is not mentioned, as it was applied to all models.

According to these results, the robust objective function leads to a better forecasting performance when applied to remove outliers. In the smoothing phase it is obvious that the moving average leads to better results (72.2% of the top results used this method). In the detrending phase constant detrending provided with the best results in the majority of stations comparing the results obtained after the data are being pre-processed (Table IV), with those obtained without applying any pre-processing to the initial data (Table V) is evident that pre-processing leads to better AQ forecasting results and to better data modelling performance.. On the basis of these results, a general structure for the pre-processing of air quality data is being proposed (Fig 11). This pre-processing structure lead to satisfactory results in four out of the six examined AQ monitoring site time series, and it was also among the structures that gave the top results per location.

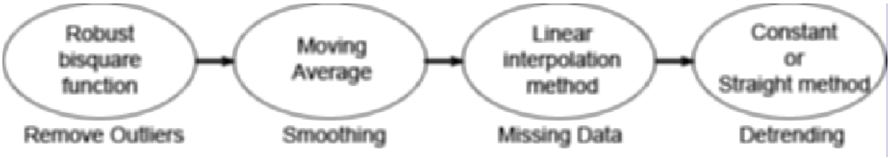

Figure 11. Pre-processing structure proposed for air quality data investigations

Table IV. Best three AQ forecasting results per station

	Remove Outliers	Smoothing	Detrending	Cor. Coef.	IA	CO	SO_2	O_3	PM10
Ag. Sofia's Sq.	RST	MAV	CNT	864	911				
	RST	SAG	LNR	869	915	✓	✓	✓	✓
	STD	**MAV**	**CNT**	**883**	**93**				
AUTH	RST	MAV	LNR	786	847				
	STD	MAV	CNT	774	823		✓	✓	
	STD	**SAG**	**CNT**	**79**	**866**				
KAL	**RST**	**MAV**	**LNR**	**836**	**895**				
	RST	SAG	LNR	807	884	✓	✓		✓
	STD	MAV	LNR	767	855				
KOR	**RST**	**MAV**	**CNT**	**881**	**931**				
	STD	MAV	CNT	866	922	✓	✓		✓
	STD	MAV	LNR	873	925				
PAN	RST	MAV	CNT	705	795				
	RST	**MAV**	**LNR**	**731**	**826**			✓	✓
	RST	SAG	LNR	706	805				
SIN	RST	MAV	CNT	**819**	871				
	RST	MAV	LNR	813	854	✓	✓	✓	✓
	RST	RST	LNR	813	**886**				

RST: Robust Bisquare Function, STD: Standard Deviation, MAV: Moving Average Smoothing, SAG: Savitzky-Golay Smoothing, CNT: Constant detrending, LNR: Straight line detrending.

Table V. AQ forecasting without data pre-processing

Locations	Cor. Coef.	IA
Ag. Sofia's Sq.	786	869
AUTH	556	666
KAL	614	721
KOR	817	893
PAN	402	489
SIN	74	817

9. Conclusions

Preprocessing is very important to environmental data because it influences the results of any data investigation method, and therefore the performance of any data mining algorithm towards problem investigation and parameter forecasting. Thus, it is crucial to know which preprocessing methods lead to the "best" quality of environmental data. Smoothing is used for noisy data to expose its features and to provide with a reasonable starting approach for parametric fitting. In this paper, several smoothing methods have been examined and evaluated: Moving Average, Savitzky-Golay, robust local regression. The performance of the smoothing algorithms improves if it is employed after the preprocessing of outliers. For this reason, the removal of outliers is of major importance and a number of methods have been tested, both qualitative (in the form of graph representations), and quantitative (in the form of robust objective functions and standard deviation functions). In this paper the robust objective functions seem to outperform the standard deviation function. An additional preprocessing process that was addressed was de-trending, where constant as well as straight like de-trending methods were investigated. Neither of these methods seemed to make a difference in terms of forecasting performance.

One of the most influential (in terms of its effects to the data quality) preprocessing tasks is the handling of missing values. Some algorithms (e.g. Naïve Bayer, CART, Bayes Tree, CN2) can deal with missing values, whereas others require that the missing values to be replaced or to remove the entire pattern of the missing values. While pattern removal solves the missing values problem important information may be lost and the available information for training is reduced, which can be a problem if the data is already limited. To replace missing values interpolation methods are used and it was found that linear interpolation methods provide with the best results. It is important to note that the evaluation of the performance of all methods was done on the basis of a number of data sets created, and the performance of an ANN model for AQ parameter forecasting, which took into account all these different data sets in order to be developed. The overall results indicated the importance of preprocessing, and led to the following suggestion concerning the steps to be followed and the methods to be used: Firstly the removal of the outliers should be done with the aid of a method like the robust bisquare function; then smoothing may be applied by employing a simple method like moving average. The next step should be the handling of missing data. In this case the linear interpolation methods are suggested. Last but not least, a method for detrending should also be employed. In terms of performance, the differences between the three best ANN models per monitoring station location varied from 2% to 10 % for the correlation coefficient and from 2% up to 5% for the Index of Agreement. This indicates the importance of preprocessing and reveals the need for further investigation, concerning its influence to environmental-air pollution data modeling.

References

1. Kukkonen J, Partanen L, Karppinen A, Ruuskanen J, Junninen H, Kolehmainen M, Niska H, Dorling S, Foxall R, Cawley G (2003) Extensive evaluation of neural network models for the prediction of NO2 and PM10 concentrations, compared with a deterministic modelling system and measurements in central Helsinki. Atmospheric Environment 37: 4539-4550.
2. Tzima F, Karatzas K, Mitkas P, Karathanasis S (2007) Using data-mining techniques for PM10 forecasting in the metropolitan area of Thessaloniki, Greece. Proceedings of the 20th International Joint Conference on Neural Networks (http://www.ijcnn2007.org):2752–2757
3. Pyle D (1999) Data preparation for data mining. Los Altos.
4. Gardner MW, Dorling SR (1998) Artificial Neural Networks (The Multilayer Perceptron) - a Review of Applications in the Atmospheric Sciences. Atmospheric Environment 32 (14/15):2627-2636
5. Kolehmainen M, Rissanen E, Raatikainen O, Ruuskanen J (2001) Monitoring odorous sulfur emissions using self-organizing maps for handling ion mobility spectrometry data. Journal of Air and Waste Management 51:966-971.
6. Sfetsos A, Siriopoulos C (2004) Time series forecasting with a hybrid clustering scheme and pattern recognition, IEEE Transactions on Systems, Man and Cybernetics, Part A, Vol. 34 (3): 399-405
7. Bianchini M, Di Iorio E, Maggini M, Mocenni C, Pucci A (2006) A Cyclostationary Neural Network Model for the Prediction of the NO_2 Concentration. ESANN 2006:67-72

8. Zhang Z, San Y (2004) Adaptive Wavelet Neural Network for Prediction of Hourly NO_X and NO_2 Concentrations. Winter Simulation Conference 2004:1170-1778
9. Finardi S, Pellegrini U (2004) Systematic Analysis of Meteorological Conditions Causing Severe Urban Air Episodes in the Central Po Valley”. 9th Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes:250-254
10. Airbase, the European Air quality database: http://air-climate.eionet.europa.eu/databases/airbase (accessed 06 March 2009).
11. Karatzas K, Kaltsatos S (2007) Air pollution modelling with the aid of computational intelligence methods in Thessaloniki, Greece. Simulation Modelling Practice and Theory, vol 15, issue 10:1310-1319
12. Weather Underground web site http://www.wunderground.com/
13. Witten IH, Eibe F (2005) Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, 2nd edition
14. Slini T, Karatzas K, Mousiopoulos N (2004) Correlation of air pollution and meteorological data using neural networks. Int. J. Environment and Pollution, vol 20(1-6):218-229
15. Engelbrecht AP (2002) Computational Intelligence, An Introduction. University of Pretoria, South Africa.
16. Matlab Documentation, Section of Pre-processing Data.
17. Agnew DC, Constable C, Lecture on Total Least Squares and Robust Methods, http://mahi.ucsd.edu/cathy/Classes/SIO223/Part1/sio223.chap8.pdf
18. Ronchetti E (2008) Lectures. Department of Economics, University of Geneva, Switzerland.
19. Huber PJ (1964) Robust Estimation of a Location Parameter. Ann. Of Mathematical Statistics 35(1):73-101
20. Yohai V (2006) The teaching of robust statistics for regression, in Proceedings of the 7th International Conference on Teaching Statistics, http://www.ime.usp.br/~abe/ICOTS7/Proceedings/PDFs/InvitedPapers/3B3_YOHA.pdf (accessed 05 March 2009)
21. Andrews DF (1974) A Robust Method for Multiple Linear Regression. Technometrics (16):523-531
22. Leblebicioğlu A (2008) Financial integration, credit market imperfections and consumption smoothing. North Carolina State University
23. Xiong L, Guo S, O'Connor KM (2005) Smoothing the seasonal means of rainfall and runoff in the linear perturbation model (LPM) using the kernel estimator, Journal of Hydrology 324(1-4):266-282.
24. Savitzky A, Golay MJE (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry 36:1627-1639
25. Michie D, Spiegelhalter DJ, Taylor CC (1994) Machine Learning, Neural and Statistical Classification (eds), (accessed 06 March 2009) http://www.shams.edu.eg/www.maththinking.com/4/whole.pdf
26. Bishop C (1995) Neural Networks for Pattern Recognition. Clarendon Press, Oxford.

27. Kolehmainen M, Junninen H, Niska H, Patama T, Ruuskanen A, Tuppurainen K, Ruuskanen J (2007) Environmental Communication in the Information Society. 16th International Conference Informatics for Environmental Protection, September 25-27, Vienna University of Technology, 2002:445-451
28. Varotsos C, Ondov J, Efstathiou M (2005) Scaling properties of air pollution in Athens, Greece and Baltimore, Maryland, Atmospheric Environment 39(22):4041-4047.
29. Slini T, Karatzas K, Mousiopoulos M (2004) Correlation of air pollution and meteorological data using neural networks. Int. J. Environment and Pollution, vol 20, nos 1-6:218-229
30. Matlab Documentation, Section of Neural Network Toolbox.
31. Saini LM, Soni MK (2002) Artificial neural network based peak load forecasting using Levenberg-Marquardt and quasi-Newton methods. Generation, Transmission and Distribution, IEE Proceedings, vol 149, issue 5:578–584
32. StatSoft, Inc. © Copyright (1984-2003) Neural Networks http://www.statsoft.com/textbook/stneunet.html
33. Willmott CJ, Ackleson SG, Davis RE, Feddema JJ, Klink KM, Legates DR, O'Donnell J, Rowe CM (1985) Statistics for the Evaluation and Comparison of Models, Geophys J (Res), 90(C5):8995–9005.