



Jul 1st, 12:00 AM

# An Application of Data Mining to PM10 Level Medium-Term Prediction

E. Pasero

A. Montuori

W. Moniaci

G. Raimondo

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

---

Pasero, E.; Montuori, A.; Moniaci, W.; and Raimondo, G., "An Application of Data Mining to PM10 Level Medium-Term Prediction" (2008). *International Congress on Environmental Modelling and Software*. 228.  
<https://scholarsarchive.byu.edu/iemssconference/2008/all/228>

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

# An Application of Data Mining to PM<sub>10</sub> Level Medium-Term Prediction

**E. Pasero<sup>a</sup>, A. Montuori<sup>b</sup>, W. Moniaci<sup>a</sup> and G. Raimondo<sup>a</sup>**

<sup>a</sup>Electronics Department and <sup>b</sup>Department of Architectural and Industrial Design, Polytechnic of Turin, C.so Duca degli Abruzzi, 24 10129 Torino, Italy

([eros.pasero/alfonso.montuori/walter.moniaci/giovanni.raimondo@polito.it](mailto:eros.pasero@alfonso.montuori/walter.moniaci/giovanni.raimondo@polito.it))

INFN, Sez Torino, I-10125 Turin, Italy

**Abstract:** The study described in this paper, analyzed the urban air pollution principal causes and identified the best subset of features (meteorological data and air pollutants concentrations) for each air pollutant in order to predict its medium-term concentration (in particular for the PM<sub>10</sub>). An information theoretic approach to feature selection has been applied in order to determine the best subset of features by means of a proper backward selection algorithm. The final aim of the research is the implementation of a prognostic tool able to reduce the risk for the air pollutants concentrations to be above the alarm thresholds fixed by the law. The implementation of this tool will be carried out using machine learning methods based on some of the most wide-spread statistical data driven techniques (Artificial Neural Networks, ANN, and Support Vector Machines, SVM).

**Keywords:** Machine learning methods; feature selection; air-pollution time series analysis and prediction.

## 1. INTRODUCTION

One of the biggest problems of urban areas is air pollution. Air pollution arises from the adverse effects on the environment of a variety of contaminants emitted into the atmosphere by natural and man-made processes such as industrial emissions, fixed combustions and vehicular traffic. The respect of the European laws concerning urban and suburban air pollution requires the analysis and implementation of automatic operating procedures in order to prevent the risk for the principal air pollutants to be above the alarm thresholds. The aim of the analysis is the medium-term forecasting of the air pollutants mean and maximum values by means of meteorological actual and forecasted data. Critical air pollution events frequently occur where the geographical and meteorological conditions do not permit an easy circulation of air and a large part of the population moves frequently between distant places of a city.

The analysis described in this paper was performed on the hourly data of the principal air pollutants (Sulphur Dioxide SO<sub>2</sub>, Nitrogen Dioxide NO<sub>2</sub>, Nitrogen Oxides NO<sub>x</sub>, Carbon Monoxide CO, Ozone O<sub>3</sub> and Particulate Matter PM<sub>10</sub>) and meteorological parameters (air temperature, relative humidity, wind velocity and direction, atmospheric pressure and solar radiation) measured by a station located in the urban area of the city of Turin (Italy). All the measurements data are relative to the time period 01/06÷10/07, Agenzia Regionale per la Protezione Ambientale del Piemonte (the Piedmont Regional Environmental Protection Agency), ARPA Piemonte, Aria Web.

The Turin urban area is located at the western side of the Po Valley, the most industrialized and populated district in Italy. It is characterized by complex terrains with mountains rising over 2700 m on the north west and a range of hills reaching 700 m located just east of the

city. Despite its position at mountain feet, the city of Turin is frequently affected by low winds, making its climatology and exposure to severe pollution episodes similar to other Po Valley sites. Apart from terrain features, the large urban area with its micrometeorological footprint gives rise to complex local atmospheric flow patterns that, particularly in the lowest layers, show relevant differences from large scale circulation (Calori et al. [2006]). From the beginning of the 70's policies for the reduction of chemical agents release in the air have been adopted. According to the Piedmont Regional Environmental Protection Agency such policies (as for example improved formulations of fossil oils for industrial activities, large adoption of methane for residential heatings, and the renewal of the fleet of circulating vehicles) gave good results allowing significant decreases in SO<sub>2</sub>, lead (Pb) and CO emissions. Nowadays air pollution problems arise from NO<sub>2</sub>, PM<sub>10</sub>, and O<sub>3</sub> the critical impact on public health of which was recognized only recently. The most severe health issue is now constituted by the high levels of PM<sub>10</sub>, a category of pollutants including solid and liquid particles having an aerodynamic diameter of up to 10 μm. In 2000 the yearly average of PM<sub>10</sub> in Turin was equal to 71 μg/m<sup>3</sup> and the number of exceedances of the daily limit value for the protection of human health (fixed by the law at 50 μg/m<sup>3</sup> according to EU Directive 99/30/CE) was equal to 214 (while it should be less than 35 per civil year). PM<sub>10</sub> can be a health hazard for several reasons: it can harm lung tissues and throat, aggravate asthma and increase respiratory illness. Indeed, high PM<sub>10</sub> levels have been correlated to increase in hospital admissions for lung and heart disease (Ostro et al. [1999]). These events require drastic measures such as the closing of the schools and factories and the restriction of vehicular traffic. The forecasting of such phenomena with up to two days in advance would allow to take more efficient countermeasures to safeguard citizens health.

In all the cases in which we can assume that the air pollutants emission and dispersion processes are stationary, it is possible to solve this problem by means of statistical learning algorithms that do not require the use of an explicit prediction model. The definition of a prognostic dispersion model is necessary when the stationarity conditions are not verified. It may happen for example when it is needed to forecast the air pollutant concentration variation due to a large variation of the emission of a source or to the presence of a new source, or when it is needed to evaluate a prediction in an area where there are no measurement points.

The research activity described in this paper concerns the feasibility of applying machine learning methods to forecast air pollution. The analysis carries on the work already developed by the NeMeFo (Neural Meteo Forecasting) research project for meteorological data short-term forecasting, Pasero et al. [2004]. The Artificial Neural Networks (ANN) and the Support Vector Machines (SVM) have been often used as a prognostic tool for air pollution, Benvenuto and Marani [2000], Perez et al. [2000], Canu and Rakotomamonjy [2001], Božnar et al. [2004], Cecchetti et al. [2004], Slini et al. [2004]. Even if we refer to these approaches as black-box methods, in as much as they are not based on an explicit model, they have generalization capabilities that make possible their application to non-stationary situations.

The first step for the implementation of a prognostic neural network or SVM is the selection of the best subset of features that are going to be used as the input to the forecasting tool. The potential benefits of the features selection process are many: facilitating data visualization and understanding, reducing the measurement and storage requirements, reducing training and utilization times, defying the curse of dimensionality to improve prediction or classification performance. It is important to highlight that the selection of the best subset of features useful for the design of a good predictor is not equivalent to the problem of ranking all the potentially relevant features. In fact the problem of features ranking, on the basis of the correlation between each of them and the target, is sub-optimum with respect to features selection especially if some features are redundant or unnecessary. On the contrary a subset of variables useful for the prediction can count out a certain number of relevant features because they are redundant, Guyon and Elisseeff [2003]. Depending on the way the searching phase is combined with the prediction, there are three main classes of feature selection algorithms: filters, wrappers and embedded. A filter is defined as a feature selection algorithm using a performance metric based entirely on the training data, without reference to the prediction algorithm for which

the features are to be selected. Wrapper algorithms include the prediction algorithm in the performance metric. The name is derived from the notion that the feature selection algorithm is inextricable from the end prediction system, and is wrapped around it. Finally, embedded methods, perform the selection of the features during the training procedure and they are specific of the particular learning algorithm.

In this work the method used for features selection is a filter. More precisely a selection algorithm with backward eliminations was used. The criterion used to eliminate the features is based on the notion of relative entropy (also known as the Kullback-Leibler divergence), inferred by the information theory.

## 2. FEATURE SELECTION ALGORITHM

The first step of the analysis was the selection of the most useful features for the prediction of each of the targets relative to the air-pollutants concentrations. For each air pollutant the target was chosen to be the mean value over 24 hours, measured every 4 hours (corresponding to 6 daily intervals a day). The complete set of features on which was made the selection, for each of the available parameters (air pollutants, air temperature, relative humidity, atmospheric pressure, solar radiation, wind speed and direction), consisted of the maximum and minimum values and the daily averages of the previous three days to which the measurement hour and the reference to the week day were added. Thus the initial set of features, for each air-pollutant, included 112 features. From this analysis an apposite set of data was excluded; such set was used as the test set.

The algorithm developed by Koller and Sahami [1996] was used to select an optimal subset of features from the set of features described above. In the following the formalism of the authors to describe the theoretical framework of the algorithm will be used. Let  $F=(F_1, F_2, \dots, F_N)$  be the set of structural features and let  $Q=(Q_1, Q_2, \dots, Q_M)$  be the set of the chosen target. For each assignment of values  $f=(f_1, f_2, \dots, f_N)$  to  $F$  we have a probability distribution  $P(Q | F = f)$  on the different possible classes,  $Q$ . We want to select an optimal subset  $G$  of  $F$  which fully determines the appropriate classification. We can use a probability distribution to model the classification function. More precisely, for each assignment of values  $g=(g_1, g_2, \dots, g_P)$  to  $G$  we have a probability distribution  $P(Q | G = g)$  on the different possible classes,  $Q$ . Given an instance  $f=(f_1, f_2, \dots, f_N)$  of  $F$ , let  $f_G$  be the projection of  $f$  onto the variables in  $G$ . The goal of the Koller-Sahami algorithm is to select  $G$  so that the probability distribution  $P(Q | F = f)$  is as close as possible to the probability distribution  $P(Q | G = f_G)$ . To select  $G$  the algorithm uses a backward elimination procedure, where at each step the feature  $F_i$  which has the best Markov blanket approximation  $M_i$  is eliminated, Pearl [1988]. Formally, we say that a subset  $M_i$  of  $F$  which does not contain  $F_i$  is a Markov blanket for  $F_i$  if  $F_i$  is conditionally independent of  $F - M_i - \{F_i\}$  given  $M_i$ . If  $M_i$  is a Markov blanket of  $F_i$  then it is also the case that the classes in  $Q$  are conditionally independent of the feature  $F_i$  given  $M_i$ . The mean value of the relative entropy between the distributions  $P(Q | M_i=f_{M_i}, F_i=f_i)$  and  $P(Q | M_i=f_{M_i})$  is used to understand how close  $M_i$  is to being a Markov blanket for  $F_i$ :

$$\delta_G(F_i | M_i) = \sum_{f_{M_i}, f_i} P(M_i = f_{M_i}, F_i = f_i) \cdot \sum_{Q_i \in Q} P(Q_i | M_i = f_{M_i}, F_i = f_i) \cdot \log \frac{P(Q_i | M_i = f_{M_i}, F_i = f_i)}{P(Q_i | M_i = f_{M_i})} \quad (1)$$

At each step the feature  $F_i$  for which  $\delta_G(F_i | M_i)$  is minimum is eliminated. The computational complexity of this algorithm is exponential only in the size of the Markov blanket, which is small. For the above reason we could quickly estimate the probability distributions  $P(Q | M_i=f_{M_i}, F_i=f_i)$  and  $P(Q | M_i=f_{M_i})$  for each assignment of values  $f_{M_i}$  and  $f_i$  to  $M_i$  and  $F_i$ . The estimate of the probability density was made by using the Parzen method as described in Parzen [1962] and Costa et al. [2003]. We modified the Koller-

Sahami algorithm in the way a candidate Markov blanket  $M_i$  for the feature  $F_i$  is selected. Instead of selecting a candidate Markov blanket  $M_i$  of size  $k$  for the features  $F_i$  by using the set of the  $k$  features most correlated to  $F_i$  we selected the  $k$  features  $F_j$  which minimize the mean value of relative entropy between the distributions  $P(Q | F_i = f_i, F_j = f_j)$  and  $P(Q | F_j = f_j)$ . After a trial and error procedure the Markov Blanket size was set equal to 3 considering the trade-off between the best approximation of a Markov blanket (large sizes) and minimum fragmentation of the training set (small sizes), Koller and Sahami [1996].

In particular this method was applied to the selection of the best subset of features useful for the prediction of the average daily concentration of PM<sub>10</sub> in the city of Turin that was often above the limit value (threshold) for the safeguard of human health (50 µg/m<sup>3</sup>). The best subset of 16 features turned out to be the following:

- Average concentration of PM<sub>10</sub> in the previous day.
- Maximum hourly value of the ozone concentration one, two and three days in advance.
- Average concentration of the ozone one, two and three days in advance.
- Minimum hourly value of the air temperature one, two and three days in advance.
- Maximum hourly value of the air temperature one day in advance.
- Average value of the solar radiation one day in advance.
- Average concentration of NO<sub>2</sub> one day in advance.
- Minimum hourly value of the humidity one and two days in advance.
- Minimum hourly value of the wind direction one day in advance.

The results can be explained considering that PM<sub>10</sub> is partly primary, directly emitted in the atmosphere, and partly secondary, that is produced by chemical/physical transformations that involve different substances as SO<sub>2</sub>, NO<sub>x</sub>, VOCs, NH<sub>3</sub> at specific meteorological conditions (Quaderno Tecnico Arpa [2002]). The features most correlated with the target and still present in the best subset of 8 features are the following ones (reported with the corresponding correlations): Average concentration of PM<sub>10</sub> in the previous day (0.77), Average concentration of NO<sub>2</sub> one day in advance (0.49), Average value of the solar radiation one day in advance (0.35) and Maximum hourly value of the air temperature one day in advance (0.34).

### 3. FORECASTING WHEN THE CONCENTRATIONS EXCEED THE LIMIT VALUE FOR THE PROTECTION OF HUMAN HEALTH

A set of feed-forward neural networks with the same topology was used. Each network had three layers with 1 neuron in the output layer and a certain number of neurons in the hidden layer (varying in a range between 3 and 20). The hyperbolic tangent function was used as transfer function.

The back-propagation rule, Werbos [1974], was used to adjust the weights of each network and the Levenberg-Marquardt algorithm, Marquardt [1963], to proceed smoothly between the extremes of the inverse-Hessian method and the steepest descent method. The Matlab Neural Network Toolbox, Demuth and Beale [2005], was used to implement the neural networks' set.

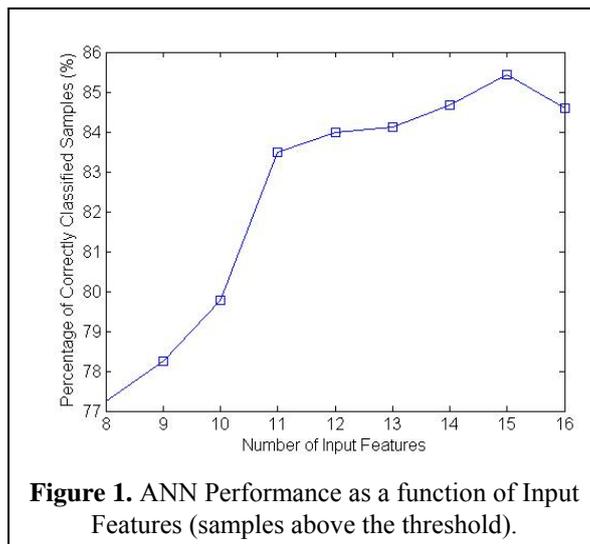


Figure 1. ANN Performance as a function of Input Features (samples above the threshold).

An SVM with an  $\epsilon$ -insensitive loss function, Vapnik [1995], was also used. The Gaussian function was used as kernel function of the SVM. The principal parameters of the SVM

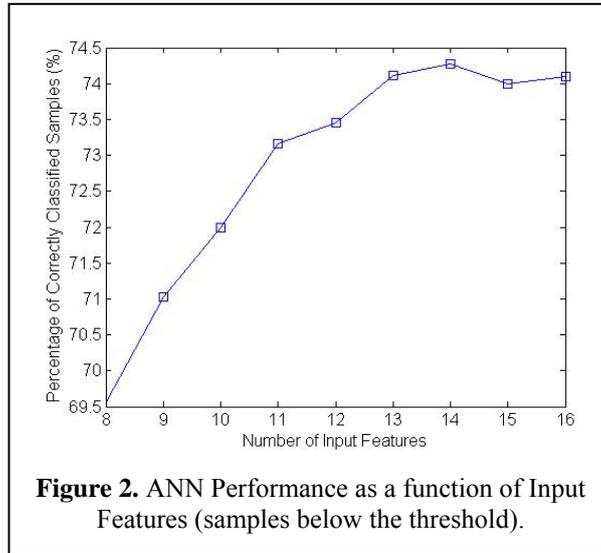
were the regularized constant C determining the trade-off between the training error and model flatness, the width value  $\sigma$  of the Gaussian kernel, and the width  $\varepsilon$  of the tube around the solution. The SVM performance was optimized choosing the proper values for such parameters. An active set method, Fletcher [1987], was used as optimization algorithm for the training of the SVM. The SVM was implemented using the “SVM and Kernel Methods Matlab Toolbox”, Canu et al. [2005].

Furthermore in order to evaluate the global performance of the machine learning methods we calculated the Index of Agreement (IA), following the suggestions of Wilmott et al. [1985]:

$$IA = 1 - \frac{\sum_i |f_i - s_i|^2}{\sum_i (|f_i - \bar{s}| + |s_i - \bar{s}|)^2} \quad (2)$$

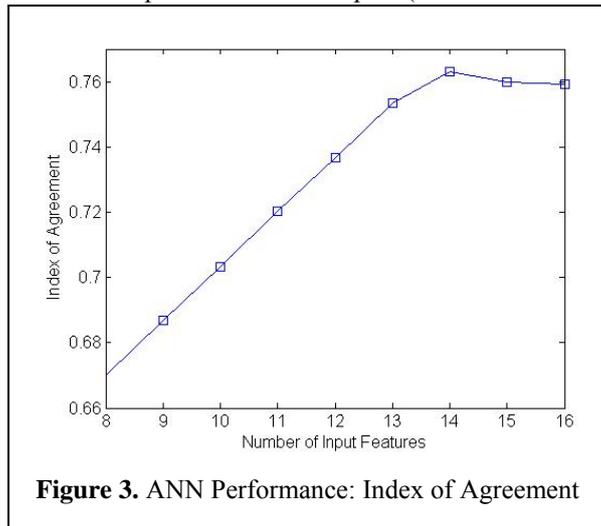
The value of the IA is from 0 to 1 and is a measure of the agreement between forecasts ( $f_i$ ) and observations samples ( $s_i$ ). The best performance corresponds to the value 1.

The neural networks were trained on a subset of the data used for the features selection. The training set was chosen to be adaptive with a constant-width sliding window. The training window width was set equal to half the number of the available data (row1). The window was slid in such a way that to forecast the (row1+1)<sup>st</sup> sample we trained the ANN on the previous row1 samples (from the 1<sup>st</sup> to the row1<sup>st</sup>), to forecast the (row1+2)<sup>nd</sup> sample we used the previous row1 samples (from the 2<sup>nd</sup> to the (row1+1)<sup>st</sup>) and so on. In this way we obtained a subset of data temporally close enough to the sample to be predicted, in order to enhance the forecasting accuracy. The test set consisted of one out of five data not used for the features selection algorithm in order to speed up computation times. Since the numbers of the training samples above and below the maximum threshold for the PM<sub>10</sub> concentration were different, the training of the networks was performed weighting more the kind of samples present a fewer number of times.



**Figure 2.** ANN Performance as a function of Input Features (samples below the threshold).

ANN on the previous row1 samples (from the 1<sup>st</sup> to the

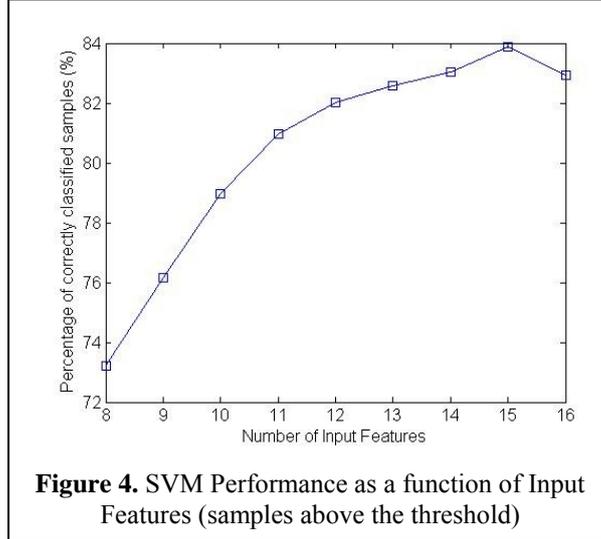


**Figure 3.** ANN Performance: Index of Agreement

The ANN performance, both for the samples under the threshold and for the samples above the threshold, reaches a maximum at a given number of input features (optimal value) and then tends to flatten when the number of input features increase. In Figure 1 and Figure 2

are shown the performances of the ANN with 18 hidden neurons, which gave the best forecasting results, for the samples below and above the threshold. This behaviour shows the effectiveness of the features selection algorithm; in fact the performance flattens adding features above an optimal number because the non-linearities involved in the generation and dispersion of the PM<sub>10</sub> are already tracked by the optimal number of features and adding extra features may not provide further relevant information for the forecasting. More precisely the performance increased meaningfully from 8 to 14/15 input features and tended to flatten when the size of the input vector was greater than 14. The results obtained with 55 samples of days under the threshold and 102 samples of days above the threshold are the following: *FA*, defined as the ratio between the number of false alarms and the

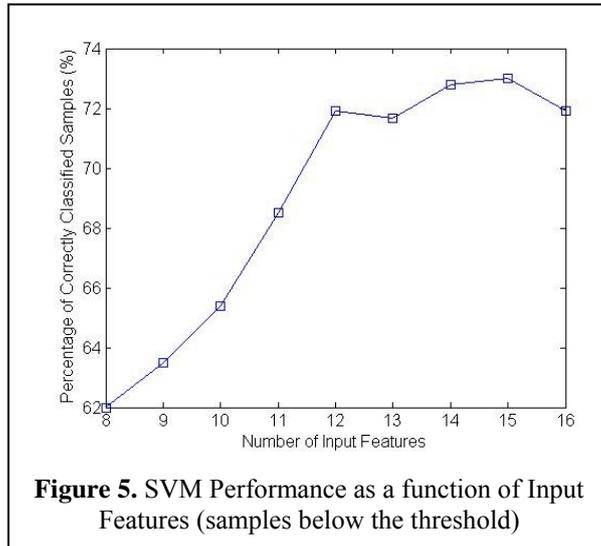
number of totally predicted exceedances, is low (14%) while the capability to forecast when the concentrations are above the threshold is about 85%. These results are comparable to those, relative to similarly polluted areas, that can be found in the literature as in Cecchetti et al. [2004]. The index of agreement, shown in Figure 3, follows the same trend of the performance as a function of the input features and approaches 0.76 as its maximum. In fact models are trained to return a binary classification (exceedance/not exceedance)



**Figure 4.** SVM Performance as a function of Input Features (samples above the threshold)

rather than to forecast the expected concentration; hence, their performances are optimal in terms of exceedances detection but are still good in terms of prediction capability.

Different assignment for SVM parameters  $\epsilon$ ,  $\sigma$  and  $C$ , were tried in order to find the optimum configuration with the highest performance. When  $\epsilon$  and  $C$  were kept constant ( $\epsilon=0.001$  and  $C=1000$ ), the SVM performances depended on  $\sigma$  and reached a maximum when  $\sigma=1$ , corresponding to an optimum trade-off between SVM generalization capability (large values of  $\sigma$ ) and model accuracy with respect to the training data (small values of  $\sigma$ ). When  $\sigma$  and  $C$  were kept constant ( $\sigma=1$  and  $C=1000$ ), the best performances were achieved when  $\epsilon$  was close to 0



**Figure 5.** SVM Performance as a function of Input Features (samples below the threshold)

and the allowed training error was minimized. From this observation, by abductive reasoning we could conclude that the input noise level was low. In accordance with such a behavior the performance of the network improved when the parameter  $C$  increased from 1 to 1000. Since the results tended to flatten for values of  $C$  greater than 1000, the parameter  $C$  was set equal to 1000. The best performance of the SVM corresponding to  $\epsilon=0.001$ ,  $\sigma=1$  and  $C=1000$  is shown in Figure 4 and Figure 5 for the samples above and below the threshold. *FA* was low (14.7%) while the capability to forecast when the concentrations were above the threshold was about 84%. The best performance of the SVM was achieved using as input features the best subset of 16 features. In Table 1 it is shown a comparison of

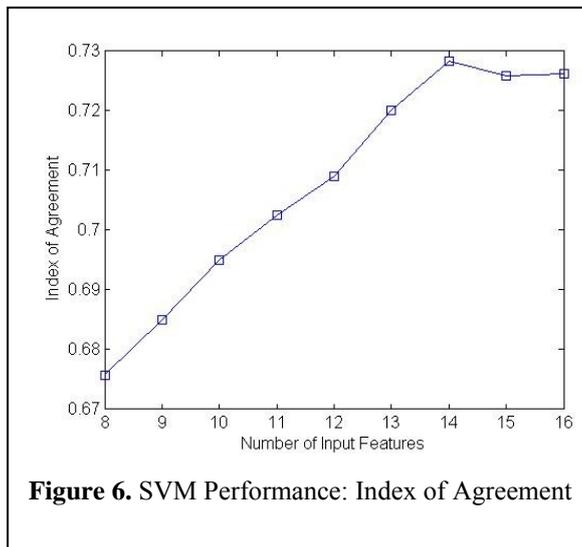
the performances of the SVM ( $\epsilon=0.001$ ,  $C=1000$  and  $\sigma$  equal to 1) and the ANN (18 neurons in the hidden layer): performance indexes include *FA*; furthermore performances are also assessed in terms of average prediction ability by means of *IA*, the true/predicted correlation  $\rho$  and by the mean absolute error *MAE*.

**Table 1.** Best Performances of the ANN and SVM.

Machine Learning Method	% correct forecasting above the threshold	% correct forecasting below the threshold	<i>FA</i>	<i>MAE</i> ( $\mu\text{g}/\text{m}^3$ )	$\rho$	<i>IA</i>
ANN (18 Hidden Neurons; 15 Input Features)	85.4%	74%	14%	10.25	0.8	0.755
SVM ( $\epsilon=0.001$ , $C=1000$ and $\sigma=1$ ; 15 Input Features)	83.8%	73.5%	14.4%	10.75	0.785	0.727

#### 4. CONCLUSIONS

The training of the ANN and SVM will be improved with a proper data-deseasonalization (Cecchetti et al. [2004]) and with stacking techniques (Wolpert [1992]) using the measurements and forecasted values of the selected features as inputs. The stacking approach consists of iterating a procedure that combines measurements data and data which are obtained by means of prediction algorithms, in order to use them all as the input to a new prediction algorithm. Since for some pollutants the meteorological conditions are very important in the generation process, different neural networks will be trained for each geopotential condition, Benichou [1995]. The analysis will be completed extending the forecasting capability of the machine learning algorithm to



**Figure 6.** SVM Performance: Index of Agreement

areas where there are no measurement points, by means of the optimization of a multi-source gaussian dispersion model. Finally it could be interesting to carry out the same kind of analysis described in this paper for PM<sub>10</sub> also for the other air pollutants.

#### ACKNOWLEDGEMENTS

We would like to thank the authors of the programs used for the scientific research related to this paper. We acknowledge S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy for the “SVM and Kernel Methods Matlab Toolbox”.

#### REFERENCES

ARPA Piemonte, Aria Web, <https://secure.regione.piemonte.it/ambiente/aria/rilev/ariaweb/index.htm>  
 ARPA Piemonte, Campagna di Rilevamento della Qualita' dell'aria con Utilizzo del Laboratorio Mobile nel Comune di Villastellone, Relazione Conclusiva. Aprile 2004 – Febbraio 2005.

- Benichou, P., Classification automatique de configurations meteorologiques sur l'europe occidentale. Technical report. Meteo-France Monographie, 1995.
- Benvenuto, F., Marani, A., Neural networks for environmental problems: data quality control and air pollution nowcasting. *Global NEST: The International Journal* 2(3), 281-292, 2000.
- Božnar, M. Z., Mlakar, P., Grašič, B., Neural Networks Based Ozone Forecasting. *Proceeding of 9th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Garmisch-Partenkirchen, Germany*, 2004.
- Calori G., Clemente M., De Maria R., Finardia S., Lollobrigida F. and Tinarelli G., Air quality integrated modelling in Turin urban area, *Environmental Modelling & Software*, 21:4, 468-476, April 2006.
- Canu, S., Grandvalet, Y., Guigue, V. and Rakotomamonjy, A., SVM and Kernel Methods Matlab Toolbox. Perception Systèmes et Information, INSA de Rouen, Rouen, France. Available: <http://asi.insarouen.fr/~arakotom/toolbox/index.html>, 2005.
- Canu, S., Rakotomamonjy, A., Ozone peak and pollution forecasting using Support Vectors. *IFAC Workshop on environmental modelling*, Yokohama, 2001.
- Cecchetti M., Corani G., Guariso G., Artificial Neural Networks Prediction of PM<sub>10</sub> in the Milan Area, *Proc. Of IEMSS 2004*, University of Osnabrück, Germany, 14-17 June 2004.
- Costa M., Moniaci W., and Pasero E., INFO: an artificial neural system to forecast ice formation on the road, *Proc. of IEEE International Symposium on Computational Intelligence for Measurement Systems and Applications*, 216–221, 29-31 July 2003.
- Demuth H., Beale M., Neural Network Toolbox User's Guide, The MathWorks, Inc., 2005.
- Fletcher R., Practical Methods of Optimization, John Wiley & Sons, NY, 2nd ed., 1987.
- Guyon I., Elisseeff A., An introduction to variable and feature selection, *The Journal of Machine Learning Research*, 3, 1157-1182, January 2003.
- Koller, D., Sahami, M., Toward optimal feature selection. *Proc. of 13th International Conference on Machine Learning (ICML)*, Bari, Italy, 284-292, 1996.
- Marquardt D., An algorithm for least squares estimation of nonlinear parameters, *SIAM J. Appl. Math.*, vol. 11, 431-441, 1963.
- Ostro, B., Chestnut L., Vichit-Vadakan N., and Laixuthai A.. The impact of particulate matter on daily mortality in Bangkok, Thailand. *Journal of Air and Waste Management Association*, 49, 100–107, 1999.
- Pasero, E., Moniaci, W., Meindl, T., Montuori, A., NEMEFO: NEural MEteorological Forecast. *Proceeding of SIRWEC 2004, 12<sup>th</sup> International Road Weather Conference*, Bingen, 2004.
- Parzen E., On Estimation of a Probability Density Function and Mode, *Annals of Math. Statistics*, vol. 33, 1065-1076, Sept. 1962.
- Pearl J., Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Mateo, CA, 1988.
- Perez, P., Trier, A., Reyes, J., Prediction of PM<sub>2.5</sub> concentrations several hours in advance using neural networks in Santiago, Chile, *Atmospheric Environment*, 34, 1189-1196, 2000.
- Quaderno Tecnico ARPA (Emilia Romagna) - SMR n°10-2002.
- Slini T., Kaprara A., Karatzas K., and Moussiopoulos N., PM<sub>10</sub> forecasting for Thessaloniki, Greece, *Environmental Modelling & Software Volume 21, Issue 4*, 559-565, April 2006.
- Vapnik V., The Nature of Statistical Learning Theory, New York, Springer-Verlag, 1995.
- Werbos P., Beyond regression: New tools for Prediction and Analysis in the Behavioural Sciences, Ph.D. Dissertation, Committee on Appl. Math., Harvard Univ. Cambridge, MA, Nov. 1974.
- Wilmott C.J., Davis R.E., Feddema J.J., Klink K.M., Legates D.R., Rowe C.M., Ackleson S.G. and O'Donell J., Statistics for the evaluation and comparison of models, *Journal Geophysics Research* 90, 8995-9005, 1985.
- Wolpert D., Stacked generalization, *Neural Networks*, 5, 241-259, 1992.