A MACHINE LEARNING TOOL TO FORECAST PM_{10} LEVEL

Giovanni Raimondo*, Alfonso Montuori, Walter Moniaci, Eros Pasero and Esben Almkvist
Polytechnic of Turin, Italy and Earth Science Centre of Gothenburg, Sweden

ABSTRACT

The research activity described in this paper concerns the study of the phenomena responsible for the urban and suburban 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 (2004). The study analyzed the 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 PM10). The selection of the best subset of features was implemented by means of a backward selection algorithm which is based on the information theory notion of relative entropy. 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 the most wide-spread statistical data-learning techniques (Artificial Neural Networks, ANN, and Support Vector Machines, SVM).

1. INTRODUCTION

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. 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 not measurement points.

The Artificial Neural Networks (ANN) and the Support Vector Machines (SVM) have been often used as a prognostic tool for air pollution, Benvenuto (2000), Perez (2000), Božnar (2004). In particular SVMs are a recent statistical learning technique, based on the computational learning theory, which implements a simple idea and can be considered as a method to minimize the structural risk, Vapnik (1995). 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 possibile their application to not-stationary situations.

In particular, the combination of the predictions of a set of models to improve the final prediction represents an important research topic, known in the literature as *stacking*. A general formalism that describes such a technique can be found in Wolpert (1992). This approach consists in 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. This technique was used in Canu (2001), where the prediction of the ozone maximum concentration with 24 hours in advance, for the urban area of Lyon (France), was implemented by means of a set of not linear models identified by different SVMs. The choice of the proper model was based on the meteorological conditions (*geopotential label*). The forecasting of ozone mean concentration for a specific day was carried out, for each model, taking as input variables the maximum ozone concentration and the maximum value of the air temperature observed in the previous day together with the maximum forecasted value of the air temperature for that specific day.

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

* *Corresponding author address*: Giovanni Raimondo, Polytechnic of Turin, Dept. of Electronics, C.so Duca Degli Abruzzi, 24, Torino (Italy), e-mail: giovanni.Raimondo@polito.it

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 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 (2003).

Depending on the way the searching phase is combined with the classification, 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.

The analysis, that is described in the following, was performed on the hourly data of the principal air pollutants (SO_2, NO, NO_2, NO_x, CO, O_3, PM_{10}) and meteorological parameters (air temperature, relative humidity, wind velocity and direction, atmospheric pressure, solar radiation and rain) measured by a station located in the urban area of the city of Goteborg (Sweden). All the measurements data are relative to the time period 01/04÷10/05 (Goteborgs Stad Miljo, http://www.miljo.goteborg.se/luftnet/). A brief introduction to the problem of air pollutants and the reasons why such pollutants are dangerous for human health, precedes the description of the data analysis method.

2. AIR POLLUTANTS

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. Due to heavy vehicular traffic and to the presence of possible industrial areas, pollutants can often be found at concentrations higher than the alarm levels fixed by the law in the air of a city urban area. The prediction of an episode of critical pollution is therefore of fundamental importance for the safeguard of citizens' health. A list of the principal air pollutants and their main characteristics follows (Air Quality Archivehttp://www.airquality.co.uk/archive/what_causes.php).

2.1 Sulphur Dioxide (SO_2)

Sulphur dioxide is an acidic gas which combines with water vapour in the atmosphere to produce acid rain. Both wet and dry deposition have been implicated in the damage and destruction of vegetation and in the degradation of soils, building materials and watercourses. SO_2 in ambient air can also affect human health, particularly in those suffering from asthma and chronic lung diseases. The principal source of this gas is power stations burning fossil fuels which contain sulphur. Major SO_2 problems now only tend to occur in cities in which coal is still widely used for domestic heating, in industry and in power stations. As many power stations are now located away from urban areas, SO_2 emissions may affect air quality in both rural and urban areas. The last 40 years have seen a decline in coal burning (domestic, industrial and in power generation) As a result, ambient concentrations of this pollutant in the most developed countries have decreased steadily over this period. Even moderate concentrations may result in a fall in lung function in asthmatics. Tightness in the chest and coughing occur at high levels, and lung function of asthmatics may be impaired to the extent that medical help is required. Sulphur dioxide pollution is considered more harmful when particulate and other pollution concentrations are high.

2.2 Nitrogen Oxides (NO and NO_2)

Nitrogen oxides are formed during high temperature combustion processes from the oxidation of nitrogen in the air or fuel. The principal source of nitrogen oxides - nitric oxide (NO) and nitrogen dioxide (NO_2), collectively known as NOx - is road traffic, which is responsible for approximately half the emissions in Europe. NO and NO_2 concentrations are therefore greatest in urban areas where traffic is heaviest. Other important sources are power stations, heating plants and industrial processes. Nitrogen dioxide can irritate the lungs and lower resistance to respiratory infections such as influenza. Continued or frequent exposure to concentrations that are typically much higher than those normally found in the ambient air may cause increased incidence of acute respiratory illness in children.

2.3 Carbon Monoxide (CO)

Carbon monoxide (CO) is a toxic gas which is emitted into the atmosphere as a result of combustion processes, and is also formed by the oxidation of hydrocarbons and other organic compounds. In European urban areas, CO is produced almost entirely (90%) from road traffic emissions. It survives in the atmosphere for a period of approximately one month but is eventually oxidised to carbon dioxide(CO2). This gas prevents the normal transport of oxygen by the blood. This can lead to a significant reduction in the supply of oxygen to the heart, particularly in people suffering from heart disease.

2.4 Ozone (O_3)

Ground-level ozone (O_3), unlike other pollutants mentioned , is not emitted directly into the atmosphere, but is a secondary pollutant produced by reaction between nitrogen dioxide (NO_2), hydrocarbons and sunlight. Ozone levels are not as high in urban areas (where high levels of NO are emitted from vehicles) as in rural areas. Sunlight provides the energy to initiate ozone formation; consequently, high levels of ozone are generally observed during hot, still sunny, summertime weather. Ozone irritates the airways of the lungs, increasing the symptoms of those suffering from asthma and lung diseases.

2.5 Particulate (PM_{10})

Airborne particulate matter varies widely in its physical and chemical composition, source and particle size. PM_{10} particles (the fraction of particulates in air of very small size (<10 μm)) are of major current concern, as they are small enough to penetrate deep into the lungs and so potentially pose significant health risks. Larger particles meanwhile, are not readily inhaled, and are removed relatively efficiently from the air by sedimentation. The principal source of airborne PM_{10} matter in European cities is road traffic emissions, particularly from diesel vehicles. Fine particles can be carried deep into the lungs where they can cause inflammation and a worsening of the condition of people with heart and lung diseases. In addition, they may carry surface-absorbed carcinogenic compounds into the lungs.

In the following table the primary and secondary sources of air pollution are shown for each pollutant, ARPA Piemonte (2004-2005). The limit values fixed by the law are reported in Appendix 1.

Air Pollutant	Primary Sources	Secondary Sources
Sulphur Dioxide (SO_2)	Fixed combustions with solid or liquid fuel	Vehicular Traffic (diesel vehicles)
Carbon Monoxide (CO)	Vehicular Traffic (petrol vehicles)	Vehicular Traffic (diesel vehicles); Fixed combustions with solid, liquid fuel or burner gas
Carbon Dioxide (CO_2)	Vehicular Traffic (petrol vehicles)	Vehicular Traffic (diesel vehicles); Fixed combustions with solid, liquid fuel or burner gas
Nitrogen Oxide (NO)	Vehicular Traffic (petrol and diesel vehicles)	Fixed combustions with solid, liquid fuel or burner gas
Nitrogen Dioxide (NO_2)	Vehicular Traffic (petrol and diesel vehicles)	Fixed combustions with solid, liquid fuel or burner gas
Particulate (PM_{10})	Vehicular Traffic (diesel vehicles); Fixed combustions with solid or liquid fuel.	Industrial emissions.

Table 1: Primary and Secondary sources of the principal air pollutants.

3. FEATURES 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 4 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, rain, 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 130 features. From this analysis an apposite set of data was excluded; such set was used as the test set.

The Koller-Sahami algorithm, Koller 1996, was used to select an optimal subset of features from the set of features described in section III. In the following the formalism of the authors to describe the theoretical framework of the algorithm will be used. Let $F=(F_1, F_2,..., F_N)$ be the set of structural features and let $Q=(Q_1, Q_2,..., Q_M)$ be the set of the chosen target. For each assignment of values $f =(f_1, f_2,..., f_N)$ to F we have a probability distribution $P(Q \mid F = f)$ 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 $g=(g_1, g_2,..., g_P)$ to G

we have a probability distribution $P(Q \mid G = g)$ on the different possible classes, Q. Given an instance $f=(f_1, f_2,..., f_N)$ 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 $P(Q \mid F = f)$ is as close as possible to the probability distribution $P(Q \mid G = f_G)$. To select G the algorithm uses a backward elimination procedure, where at each state 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 $P(Q \mid M_i=f_{M_i}, F_i= f_i)$ and $P(Q \mid M_i=f_{M_i})$ is used to understand how close M_i is to being a Markov blanket for F_i:

$$\delta_G(\boldsymbol{F_i} \mid \boldsymbol{M_i}) = \sum_{f_{M_i}, f_i} P(\boldsymbol{M_i} = f_{M_i}, \boldsymbol{F_i} = f_i) \cdot \sum_{\boldsymbol{Q_i} \in \boldsymbol{Q}} P(\boldsymbol{Q_i} \mid \boldsymbol{M_i} = f_{M_i}, \boldsymbol{F_i} = f_i) \cdot log \frac{P(\boldsymbol{Q_i} \mid \boldsymbol{M_i} = f_{M_i}, \boldsymbol{F_i} = f_i)}{P(\boldsymbol{Q_i} \mid \boldsymbol{M_i} = f_{M_i})}. \quad (3.1)$$

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 $P(Q \mid M_i=f_{M_i}, F_i= f_i)$ and $P(Q \mid M_i=f_{M_i})$ 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, Parzen (1962), Costa (2003).

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_{10} in the city of Goteborg. In fact from the data it was observed that this concentration was often above the limit value for the safeguard of human health (50 µg/m^3). The best subset of 16 features turned out to be the following:

- Average concentration of PM_{10} in the previous day.
- Maximum hourly value of the ozone concentration one, two and three days in advance.
- Maximum hourly value of the air temperature one, two and three days in advance.
- Maximum hourly value of the solar radiation one, two and three days in advance.
- Minimum hourly value of SO_2 one and two days in advance.
- Average concentration of the relative humidity in the previous day.
- Maximum and minimum hourly value of the relative humidity in the previous day.
- Average value of the air temperature three days in advance.

The results can be explained considering that PM_{10} is partly primary, directly emitted in the atmosphere, and partly secondary, that is produced by chemical/physical transformations that involve different substances as SOx, NOx, COVs, NH_3 at specific meteoriological conditions, Quaderno Tecnico Arpa (2002).

4. FORECASTING WHEN PM_{10} LEVEL IS ABOVE 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 (2005), was used to implement the neural networks'set.

An SVM with an ε-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 σ of the Gaussian kernel, and the width ε 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 (2005).

The neural networks were trained on a representative subset of the data used for the features selection algorithm. A subset of the first two years of data was used: a measurement sample every three samples after leaving out one sample every five of the original data. In this way the computational time of the adopted machine-learning algorithms was reduced while obtaining a subset of data as representative as that used for the features selection. In fact such a subset included a sufficient number of all the 4 daily intervals in which the measurement data were divided by our analysis. The test set consisted of the data not used for the features selection algorithm. Since the number of the training samples above the maximum threshold for the PM_{10} concentration was much lower than the number of samples under such threshold, the training of the networks was performed weighting more the kind of samples present a fewer number of times.

As we can see from Picture 1 and Picture 2 the ANN performance, both for the samples under the threshold and for the samples above the threshold, increased when the number of input features increased. More precisely the performance increased meaningfully from 2 to 8 input features and tended to flatten when the size of the input vector was greater than 8.

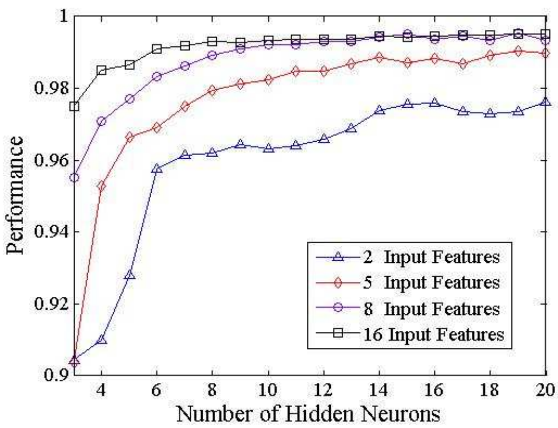
The best subset of 8 features was the following:

- Average concentration of PM_{10} in the previous day.
- Maximum hourly value of the ozone concentration one, two and three days in advance.
- Maximum hourly value of the air temperature in the previous day.
- Maximum hourly value of the solar radiation one, two and three days in advance.

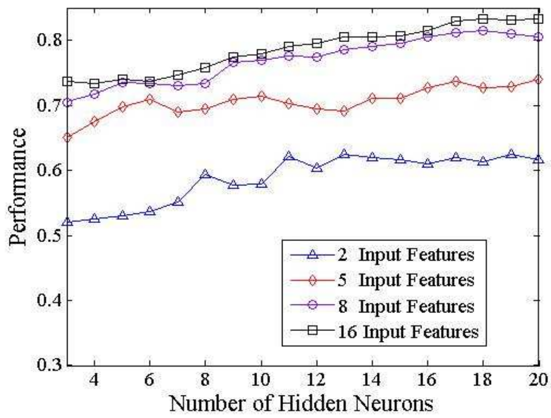
Selecting as input to the ANN such set of 8 features, the best results could be obtained with a neural network having 18 neurons in the hidden layer. In table 2 are displayed the results obtained with 5115 samples of days under the threshold and 61 samples of days above the threshold. It can be noted that the probability to have a false alarm is really low (0.82%) while the capability to forecast when the concentrations are above the threshold is about 80%.

Samples	Correct Forecasting	Incorrect Forecasting
Below the threshold	5073	42
Above the threshold	48	13

Table 2: Neural network performance.

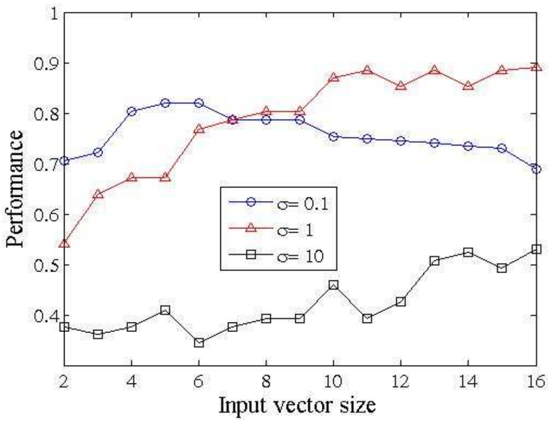


Picture 1: Performance of the ANN as a function of the number of input Features (samples below the threshold).

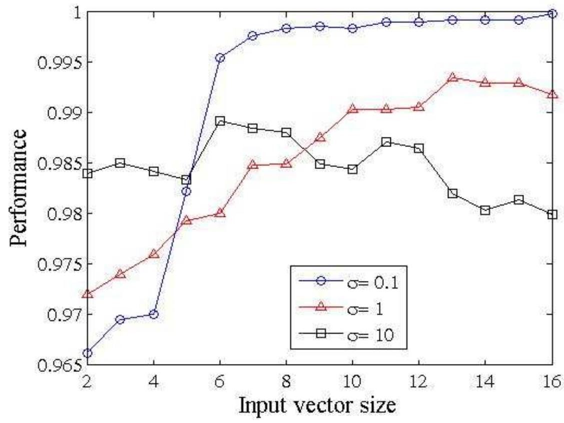


Picture 2: Performance of the ANN as a function of the number of input Features (samples above the threshold).

Different assignment for SVM parameters ε, σ and C, were tried in order to find the optimum configuration with the highest performance. As we can see from Picture 3, when ε and C were kept constant (ε=0.001 and C=1000), the SVM performances referring to samples above the threshold, for a high number of input features, depended on σ and reached a maximum when σ=1, corresponding to an optimum trade-off between SVM generalization capability (large values of σ) and model accuracy with respect to the training data (small values of σ). The value of σ corresponding to this trade-off decreased to 0.1 for lower values of the input vector size (Picture 3) and for samples below the threshold (Picture 4), reflecting the fact that the generalization capability was less important when the training set was more representative

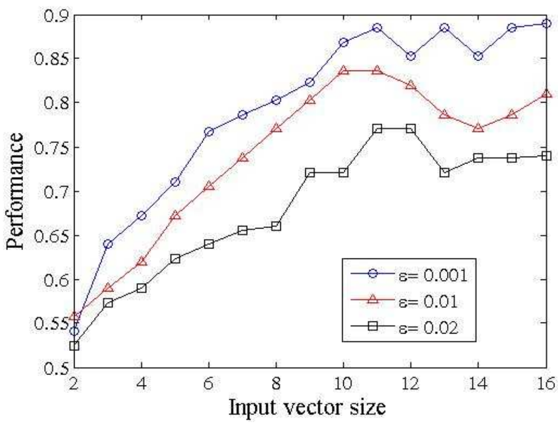


Picture 3: Performances of the SVM as a function of σ (ε=0.001 and C=1000), samples above the threshold.

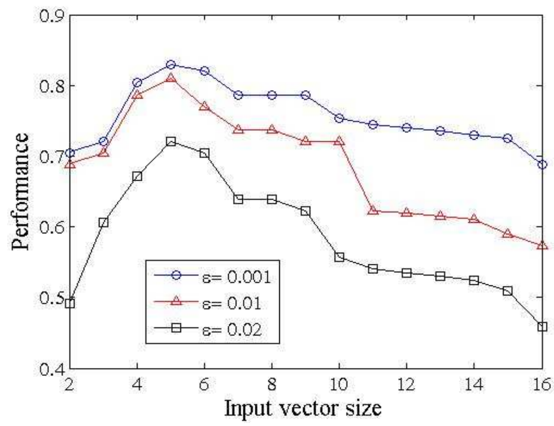


Picture 4: Performances of the SVM as a function of σ (ε=0.001 and C=1000), samples below the threshold.

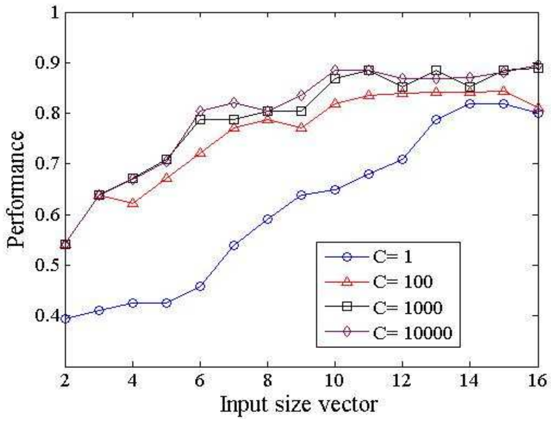
When σ and C were kept constant (Picture 5: σ=1 and C=1000; Picture 6: σ=0.1 and C=1000), the best performances were achieved when **ε** was close to 0 and the allowed training error was minimized. From this observation, by abductive reasoning we could conclude that the input noise level was low. In accordance with such a behavior the performance of the network improved when the parameter C increased from 1 to 1000 (see Picture 7). Since the results tended to flatten for values of C greater than 1000, the parameter C was set equal to 1000.



Picture 5: Performances of the SVM as a function of **ε** (σ=1 and C=1000), samples above the threshold.



Picture 6: Performances of the SVM as a function of **ε** (σ=0.1 and C=1000), samples above the threshold.



Picture 7: Performances of the SVM as a function of C (ε=0.001 and σ=1), samples above the threshold.

The best performance of the SVM corresponding to ε=0.001, σ =0.1 and C=1000 was achieved using as input features the best subset of 8 features previously defined. The probability to have a false alarm was really low (0.13%) while the capability to forecast when the concentrations were above the threshold was about 80%. The best performance of the SVM corresponding to ε=0.001, σ =1 and C=1000 was achieved using as input features the best subset of 11 features. In this case the probability to have a false alarm was higher than in the previous one (0.96%) but the capability to forecast when the concentrations were above the threshold was nearly 90%. In Table 3 it is shown a comparison of the performances of the SVM(ε=0.001, C=1000 and σ equal to 0.1 and 1) and the ANN (18 neurons in the hidden layer) as a function of the number of input features.

Samples		SVM (ε=0.001, C=1000 and σ = 0.1)		SVM (ε=0.001, C=1000 and σ = 1)		ANN (18 hidden neurons)	
		Correct Forecasting	Incorrect Forecasting	Correct Forecasting	Incorrect Forecasting	Correct Forecasting	Incorrect Forecasting
8 input features	Below the threshold	5107	8	5038	77	5073	42
	Above the threshold	48	13	49	12	48	13
11input features	Below the threshold	5111	4	5066	49	5070	45
	Above the threshold	42	19	54	7	49	12

Table 3: ANN and SVM performances as a function of the number of input features.

5. FUTURE ACTIVITIES

The training of the ANN and SVM will be improved with stacking techniques using as inputs the measurements and forecasted values of the selected features. Since for some pollutants the meteorological conditions are very important in the generation process, different neural networks will be trained for each different geopotential condition, Benichou (1995). The analysis will be completed extending the forecasting capability of the machine learning algorithm to areas where there are not 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_{10} also for the other air-pollutants.

6. ACKNOWLEDGMENT

Our sincere thanks to the CRT Foundation, with the Lagrange Project, that partly funded this work,Turin,Italy,2006.

7. REFERENCES

[1] Air Quality Archive –The Uk National Air Quality InformationArchive:http://www.airquality.co.uk/archive/what_causes.php
[2] ARPA Piemonte, Aprile 2004 – Febbraio 2005: Campagna di Rilevamento della Qualita' dell'aria con Utilizzo del Laboratorio Mobile nel Comune di Villastellone, Relazione Conclusiva.
[3] Benichou, P. 1995: Classication automatique de congurations mtorologiques sur l'europe occidentale. Technical report. *Meteo-France*

Monographie.
[4] Benvenuto, F., Marani, A. 2000: Neural networks for environmental problems: data quality control and air pollution nowcasting. *Global NEST: The International Journal 2(3),* 281-292.
[5] Božnar, M. Z., Mlakar, P., Grašič, B., 2004: Neural Networks Based Ozone Forecasting. *Proceeding of 9th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes,* Garmisch-Partenkirchen, Germany.
[6] Canu, S., Grandvalet, Y., Guigue, V. and Rakotomamonjy, A. 2005 : SVM and Kernel Methods Matlab Toolbox. Perception Systèmes et Information, INSA de Rouen, Rouen, France. Available: http://asi.insa-rouen.fr/~arakotom/toolbox/index.html.
[7] Canu, S., Rakotomamonjy, A. 2001: Ozone peak and pollution forecasting using Support Vectors. *IFAC Workshop on environmental modelling*, Yokohama.
[8] Costa, M., Moniaci, W. and Pasero, E. 2003: 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.
[9] Demuth, H., Beale, M. 2005: Neural Network Toolbox User's Guide. The MathWorks, Inc.
[10] Fletcher, R. 1987: Practical Methods of Optimization. John Wiley & Sons, NY, 2nd ed.
[11] Goteborgs Stad Miljo, http://www.miljo.goteborg.se/luftnet/.
[12] Guyon , I., Elisseeff, A. 2003: An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3, 1157-1182.
[13] Koller, D., Sahami, M. 1996: Toward optimal feature selection. *Proc. of 13th International Conference on Machine Learning* (ICML), Bari, Italy, 284-292.
[14] Marquardt, D. 1963: An algorithm for least squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 11, 431-441.
[15] Parzen, E. 1962: On Estimation of a Probability Density Function and Mode. *Annals of Math. Statistics*, 33, 1065-1076.
[16] Pasero, E., Moniaci, W., Meindl, T., Montuori, A. 2004: NEMEFO: NEural MEteorological Forecast. *Proceeding of SIRWEC 2004*, *12th International Road Weather Conference,* Bingen.
[17] Pearl, J. 1988: Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, San Mateo, CA.
[18] Perez, P., Trier, A., Reyes, J. 2000: Prediction of PM2.5 concentrations several hours in advance using neural networks in Santiago, Chile. *Atmospheric Environment,* 34, 1189-1196.
[19] Quaderno Tecnico ARPA (Emilia Romagna) - SMR n°10-2002.
[20] Vapnik, V. 1995: The Nature of Statistical Learning Theory. New York, Springer-Verlag.
[21] Werbos, P. 1974: Beyond regression: New tools for Prediction and Analysis in the Behavioural Sciences. Ph.D. Dissertation, Committee on Appl. Math., Harvard Univ. Cambridge, MA.
[22] Wolpert, D. 1992: Stacked generalization. *Neural Networks*, 5, 241-259.

Appendix 1
European legislation on the air pollutants limit values and alarm thresholds (Italian law references).

Carbon Monoxide (law references: D.M. 2.04.2002 N. 60)

- Limit value for the protection of human health: The maximum daily average over 8 hours calculated every hour and based on the previous eight hours should not be above 10 mg/m^3.

Nitrogen Dioxide (law references: D.M. 2.04.2002 N. 60)

- Hourly limit value for the protection of human health: the hourly average of the NO_2 concentrations should not be above 200 µg/m^3 for more than 18 time per civil year.
- Annual limit value for the protection of human health: The annual average of the NO_2 concentrations should not be above 40µg/m^3.
- Alarm threshold: The alarm threshold is reached when the hourly mean concentrations are above 400 µg/m^3 for 3 consecutive hours.

Sulphur Dioxide (Law references: D.M. 2.04.2002 N. 60)

- Hourly limit value for the protection of human health: the hourly mean value of the SO_2 concentrations should not be above 350 µg/m^3 for more than 24 times per civil year.
- Daily limit value for the protection of human health: the daily mean value of the SO_2 concentrations should not be above 125 µg/m^3 for more than 3 times per civil year.
- Alarm threshold for Sulphur Dioxide: the alarm threshold is reached when the mean hourly concentrations are above 500 µg/m^3 for 3 consecutive hours.

Particulate (PM10) (Law references: D.M. 2.04.2002 N. 60)

- Daily limit value for the protection of human health: the daily mean value of the PM_{10} should not be above 50 µg/m^3 for more than 35 times per civil year.
- Annual limit value for the protection of human health: the annual mean value of the PM_{10} concentrations should not be above 40 µg/m^3.

Benzene (law references: D.M. 2.04.2002 N. 60)

- Annual limit value for the protection of human health: the annual mean value of the benzene concentrations should not be above 10 µg/m^3.

Ozone (law references: D.L.vo n.183 21.05.2004)

- Threshold for the protection of human health (120 µg/m^3) as the mean value over 8 hours calculated every hour and based on the previous 8 hours (D.L.vo n.183 21.05.2004), maximum value of the mean values calculated over 8 hours.
- Threshold for the protection of human health (120 µg/m^3) as the maximum daily mean value over 8 hours calculated every hour and based on the previous eight hours (D.L.vo n.183 21.05.2004), number of times the 8 hours averages are above the limit value during the monitoring day.
- Information threshold (180 µg/m^3) (D.L.vo n.183 21.05.2004)
- Alarm threshold (240 µg/m^3) (D.L.vo n.183 21.05.2004)