Artificial neural networks prediction of PM10 in the Milan area

M. Cecchetti, G. Corani [a], G. Guariso[a]

[a]Dipartimento di Elettronica ed Informazione
Politecnico di Milano
e-mail:corani(guariso)@elet.polimi.it

Abstract: PM10 constitutes a major concern for Milan air quality. We presents a series of results obtained applying different neural networks approaches to the PM10 prediction problem. The 1-day ahead prediction shows a satisfactory level of accuracy, which may be further improved if a proper deseasonalization approach is adopted, thus transferring some a priori knowledge in the data pre-processing step. Then, we tackle the problem of the 2-days ahead prediction; in order to optimize the neural network architecture identification procedure, we try a pruning approach besides the usual trial and error one. Prediction performances are very close between the two models, and denote a significant decrease of accuracy with respect to the 1-day case, even though some meteorological improper (i.e. future measures) input is added to the model structure.

Keywords: feedforward neural networks, PM10, time series forecast

1 INTRODUCTION

The Milan urban area is located at the center of the Po Valley, the most industrialised and populated district in Italy. According to the Municipal Report on the State of the Environment (AMA [Municipal Environment and Mobility Agency]), the yearly average of pollutants such as SO2, NOx, CO, TSP has decreased respectively by about 90%, 50%, 65%, 60% during the period 1989-2001; no exceedings of alarm and attention thresholds have been observed since 1997 for SO_2 and *TSP*, and since 1999 for CO. The situation is just slightly worse for NO_x which, though clearly decreasing, showed, on the average, 8 yearly exceedings of the attention threshold (none of the alarm) over the same period. The yearly averages of micropolluttants such as benzene and lead are also largely under the thresholds established for human health protection. These significant results are due to various actions, such as improved formulations of fossil oils for industrial activities, large adoption of methane for residential heatings, and the renewal of the fleet of circulating vehicles. The most severe health issue is now constituted by the high levels of particulate matter (PM10), a category of pollutants including solid and liquid particles having an effective aerodynamic diameter smaller than $10\mu m$. PM10 can be a health hazard for several reasons; it can harm lung tissues and throat, aggravate asthma and increase respiratory illness. Indeed, high PM10 levels have been correlated to increase in hospital admissions for lung and heart disease (Ostro et al. [1999]).

Health-based standards for PM10 have been established by a European directive (99/30/CE). Accordingly, the Regional law (DGR 19/10/2001) fixes at $50\mu g/m^3$ the attention threshold for the daily average; if the threshold is exceeded for 5 consecutive days, the "*attention state*" is declared [1]. Differently from the other pollutants quoted above, the yearly average of PM10 has been substantially stable (about $45\,\mu g/m^3$) since the beginning of monitoring in 1998. On average, PM10 exceeds the attention threshold in Milan for about 100 days/year, and about 20 "*attention state days*" are declared every year (AMA [Municipal Environment and Mobility Agency]).

A system able to predict PM10 concentrations could provide a useful anticipation to Public Authorities, in order to plan an increase in the public transport, warn people to avoid exposures to unhealthy air, or alert them of possible traffic blocks.

[1]The most recent regional law [DGR 13858 29/7/2003] preventively decrees the block of the pre-Euro vehicles for some hours during winter days, leaving the defi nition of other emergency actions to the responsibility of local authorities

Neural networks for air pollution forecast collected a general consensus over the last years, as pointed out by the review of Gardner and Dorling [1998]. In this paper, we train neural networks to forecast the PM10 daily average concentration, assuming to exploit all the data available until 9 a.m. of the current day *t*. As a second modelling step, we extend the prediction horizon to the following day. To this end, we add some further meteorological input variables, measured at ground level over both the day *t* and *t+1*. Although such an approach is, technically speaking, improper since it uses inputs which are unavailable in real time operations, it constitutes an interesting subject of investigation since the obtained performances may be considered as an upper bound of what can be achieved by adding forecasts of such meteorological variables in the model.

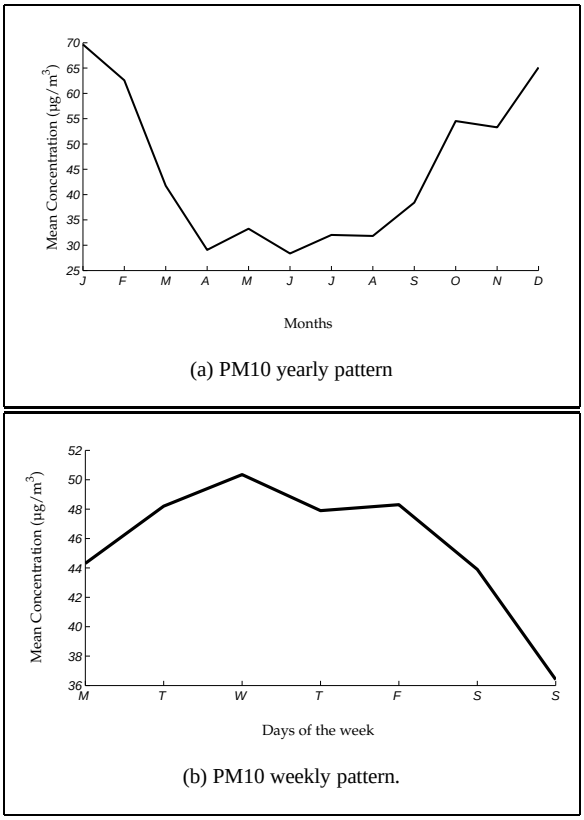
Since PM10 time series underlies clear periodicities at yearly and weekly level, we also investigate different deseasonalization approaches in the data preprocessing step.

A final issue addressed by this study regards the methodology adopted for the identification of the optimal neural network architecture. For the 1-day prediction, we found satisfactory results by selecting the architecture by trial and error, as usually done in neural networks applications. For the 2-days prediction, indeed much more difficult, we try also an alternative approach, based on the Optimal Brain Surgeon pruning algorithm (Hassibi and Stork [1993]).

2 TIME SERIES ANALYSIS

The data used in this work have been collected from a monitoring station located in a residential area of the city and refer to the years 1999-2002. The dataset is constituted by hourly time series; missing values range between 5% and 10% depending on the considered monitor. Data are splitted into training (1999-2000), validation (2001) and testing sets (2002).

To be used as input variable for the predictor, each monitored parameter has to be grouped from the original hourly series to a daily time series; this has been accomplished selecting the most suitable grouping operator (mean over 24 hours, mean over a certain time window, maximum) for any given input, by means of an extensive correlation analysis. The input variables set finally adopted comprises an autoregressive PM10 term, NO_x (used to track the road traffic emissions), SO_2 (a proxy for heatings emissions), and a wide set of meteorological variables, such as temperature, humidity, wind velocity, solar radiation, atmospheric pressure and Pasquill stability class.



(a) PM10 yearly pattern

(b) PM10 weekly pattern.

Figure 1: PM10 average profiles.

Simply plotting the PM10 monthly averages, it is possible to notice cyclic patterns with high peaks in winter and much lower concentrations during summer (Figure 1a). Such an effect is due to the combined action of the unfavorable dispersion conditions which are encountered during winter (e.g., reduced mixing layer height) and to the increase of anthropic emissions (heavier traffic volumes and residential heatings). The average PM10 weekly profile also shows a typical pattern, with a decrease of about 25-30% during the weekends (Figure 1b). Indeed, the spectral analysis performed on the daily values clearly supports such evidence, as one can see from the periodogram shown in Figure 2: two peaks can be clearly detected, corresponding respectively to the yearly and weekly frequencies. The same kind of periodicities have been detected (analyzing both time and frequency domains) also on SO_2 and NO_x.

Such recognized periodicities suggest that a proper

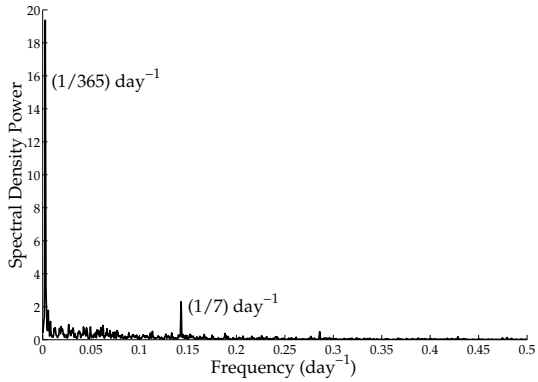


Figure 2: PM10 periodogram

data-deseasonalization approach may be useful in order to introduce some *a priori* knowledge in the data pre-processing step.

3 Prediction methodologies and results

We identify the predictor as a traditional feed forward neural network (Bishop [1995]). The network is constituted by a hidden layer, collecting a series of neurons having hyperbolic tangent as transfer function. The output layer contains just one neuron, having an identity transfer function, which actually returns the prediction. The optimal network architecture is found via trial and error, training 20 times each candidate architecture. The Levenberg-Marquardt algorithm is adopted as training method and early stopping is applied in order to prevent overfitting; hence, the training algorithm is stopped once the squared error on the validation dataset begins to increase. The architecture showing the lowest squared error on the validation set, which is assumed to have the best generalization ability, is finally selected as optimal and then simulated on the testing set, in order to quantitatively evaluate its performances.

With reference to data deseasonalization, we tried different solutions. The most effective one proved to be the fitting of a periodic regressor of type $f(\omega) = \sum_{k=1}^{k=N} a_k sin(k\omega t) + b_k cos(k\omega t)$, i.e.a linear combination of sine and cosine terms up to a harmonic N. Its parameters a_k and b_k have been estimated, on each pollutant time series, via a standard least square technique.The regressor including both the yearly angular frequency ($\omega_1 = 2\pi/365 days^{-1}$) and the weekly angular frequency ($\omega_2 = 2\pi/7 days^{-1}$) is obtained as:

$$R(t) = c + f(\omega_1) + f(\omega_2) \tag{1}$$

where c is a constant term. Models are then trained -with reference to the pollutants variables- on the standardized residuals time series, while meteorological data are just standardized (i.e. converted to a zero average and a unit standard deviation).

	Net_R	Net_S
Average goodness		
ρ	0.94	0.91
MAE	7.70	8.71
Threshold detection		
n_0	127	127
CPO	0.81	0.87
CPP	0.91	0.83
FA	0.09	0.17

Table 1: 1-days ahead prediction performances.

The most suitable configuration of the regressor for each polluttant (i.e., the number of harmonics to be considered) is investigated up to $N = 2$, thus avoiding higher parametrization of the regressors. Indeed, a similar attempt using periodic regressors (Kolehmainen et al. [2001]) reported a prediction worsening with respect to the use of simply standardized data; this was probably due to the too high number of harmonics they considered, which returned a too poor signal.

The testing set performances for the networks identified using the regressors (Net_R) or the usually standardized data (Net_S) are given in Table 1, and show a satisfactory prediction accuracy for both models. Performances are assessed in terms of average prediction ability by means of the true/predicted correlation ρ and by the mean absolute error MAE. Moreover, the problem of predicting the exceedances of the attention threshold is evaluated through of a series of indicators:

- CPO, the ratio of *C*orrectly *P*redicted exceedances N_{CP} to the number of total *O*bserved exceedances N_O:

$$CPO = \frac{N_{CP}}{N_O} \tag{2}$$

- *CPP*, the ratio between the number N_{CP} of *C*orrectly *P*redicted exceedances to the number N_P of totally *P*redicted exceedances:

$$CPP = \frac{N_{CP}}{N_P} \tag{3}$$

- *FA*, the ratio between the number of *F*alse *A*larms and the number N_P of totally *P*redicted exceedances:

$$FA = \frac{N_{FA}}{N_P} = 1 - CPP \tag{4}$$

The model trained on the residual time series shows an appreciable improvement on the average indicators. With regard to threshold exceedances prediction however, Net_S shows higher percentage (*CPP*: 87% vs 81%) of detection. On the other hand however, an above threshold prediction given by Net_R is far more reliable than by Net_S, with significantly better values of both CPP and FA.

For a comparison of these results with modelling approaches other than neural networks, one can refer to the work by Corani and Barazzetta [2004], who addressed the same case study using a traditional linear predictor (ARX model). They report, for instance, a true/predicted correlation of about $.89$ and a MAE of about $11\mu g/m^3$. The non-linearity of ANN appears therefore to allow a significant improvement of the prediction quality.

Given the satisfactory results on the 1-day prediction, we tried the more ambitious target of 2-days ahead forecast. To this end, we add further meteorological input variables to the model: temperature, rainfall, humidity pressure and wind speed measured over day (*t*) and (*t+1*). Such data constitute a set of improper inputs, since their values are not available at prediction time (9 a.m. of day *t*). However, they provide interesting indications, if one considers that the prediction performances obtained in this way constitute an upper bound of what can be achieved by inserting in the model the actual forecasts of such variables, obtained by means of a meterological model. Polluttant time series have been again deseasonalized by means of periodic regressors, given the valuable contribution of such a technique on the 1-days prediction.

Since the prediction target is much more difficult in this case, we try to improve the neural networks identification procedure performing a set of experiments for the identification of the architecture via the Optimal Brain Surgeon pruning algorithm (Hassibi and Stork [1993]), besides the classical trial and error approach. The basic idea of pruning algorithms is to start from a fully connected network, considered large enough to capture the desired input-output relationship. Then, they compute some measure of the contribution of each parameter to the problem solution, and consequently prune the less influential one from the network, to generate a new partially connected model, containing one parameter less. In this way, weights and neurons considered redundant are eliminated, significantly reducing the amount of guesswork needed for model selection. The network showing the lowest validation error between the many pruned architectures generated is finally chosen as predictor, consistently with the model selection criterion adopted for fully connected networks. The selected pruned network architectures may contain one order of magnitude less parameters than the fully connected ones, and are hence very parsimonious, thus providing a greater generalization power.

	Net_{2d}	Net_{2dP}
Average goodness		
ρ	0.76	0.76
MAE	13.08	12.89
Threshold detection		
n_0	127	127
CPO	0.73	0.72
CPP	0.75	0.76
FA	0.25	0.24

Table 2: 2-days ahead prediction performances.

The performances of the network designed by trial and error Net_{2d} and by pruning Net_{2dP} on the testing set are given in Table 2. One can easily notice that the performances of the two networks are very close to each other, and that a strong decrease takes place with respect to the 1-days prediction case. The true-predicted correlation decreases from $.93$ to $.76$, and the threshold exceedance detection indexes becomes significantly worse.

Since a quantitative meteorological forecast was not available, it is impossible to verify the performances which may be achieved with this additional information. However, the results will certainly lie between the upper bound already found and a lower bound constituted by a model trained without any improper meteorological information. Such a model shows a true/predicted correlation of about 0.74, with CPO and FA being respectively 0.72 and 0.29. The gap between the two approaches is not substantial and indicates that it is worth inves-

tigating more advanced descriptions of the meteorology (e.g. at a synoptic level) and more effective ways of exploiting such information.

4 CONCLUSIONS

The results on the prediction computed at 9a.m. for the current day are clearly satisfactory, with a true/predicted correlation of about 0.94 and a index of correctly predicted exceedances higher than 0.80. Data deseasonalization seems a valuable approach to increase of some points the average performances indicators; improvements are however less clear on the threshold exceedances prediction indicators. The 2-days prediction appears as an open problem, and the extension of air quality forecast horizons is likely to require a great research effort. In our opinion, dramatical performances improvements are not to be expected by studying new prediction algorithms: indeed, neural networks constitute a flexible non-linear modelling approach, able to learn very complex relationship from data. On the other hand, the availability of more advanced meteorological data, able to describe the air masses motion in the atmosphere (e.g. vertical profiles of wind speed and temperature, mixing height), can greatly increase the informative content of the input variables set, and may thus allow more significant improvements of air quality predictions.

We remark that, although presently no clear trend is detected on the PM10 time series, the situation may evolve over time, thus requiring a retraining of the predictor. Neural networks cannot be easily updated, and in fact it will be necessary to identify ex novo both the structure and parameters of the network, in order to have an up-to-date predictor. From this point of view, it is worth to mention that lazy learning, a local linear modelling approach, can constitute a viable alternative to neural networks; in fact, according to (Birattari et al. [1999]), this method may provide comparable prediction performances, allowing at the same time a quicker design and an easier update of the predictor.

REFERENCES

AMA (Municipal Environment and Mobility Agency). *Relazione sullo stato dell'ambiente del comune di Milano*. Milan, 2003.

Birattari, M., G. Bontempi, and H. Bersini. Lazy learning for modeling and control design. *Int. J. of Control*, 72:643–658, 1999.

Bishop, C. *Neural networks for pattern recognition*. Oxford University Press, 1995.

Corani, G. and S. Barazzetta. First results in the prediction of particulate matter in the Milan area. In *9th Int. Conf. on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, Garmisch-Partenkirchen, 2004.

Gardner, M. and S. Dorling. Artificial neural network (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmospheric Environment*, 6(32):2627–2636, 1998.

Hassibi, B. and D. Stork. Second order derivatives for network pruning: Optimal brain surgeon. In S.J. Hanson, J.D. Cowan, C. G., editor, *Proceedings of Advances in Neural Information Processing System*, pages 164–171, 1993.

Kolehmainen, M., H. Martikainen, and J. Ruuskanen. Neural networks and periodic components used in air quality forecasting. *Atmospheric Environment*, 35:815–825, 2001.

Ostro, B., L. Chestnut, N. Vichit-Vadakan, and A. Laixuthai. The impact of particulate matter on daily mortality in Bangkok, Thailand. *Journal of Air and Waste Management Association*, 49: 100–107, 1999.