Forecasting air quality by using ANNs
L'uso delle ANNs per la previsione della qualità dell'aria

Annalina Sarra, Adelia Evangelista, Tonio Di Battista and Francesco Bucci

Abstract The artificial neural networks (ANNs) have been extensively used in air pollution prediction because of their flexibility to deal with processes involving non linear and complex data and/or to solve articulate problems in which a priori knowledge is incomplete or noisy. In this study, we trained different ANNs in assessing the capability of models for the prediction of air quality. The air pollution data from two monitoring stations in Pescara (Central Italy), along with some meteorological parameters, were used in forecasting Nitrogen Dioxide (NO_2) levels, one day in advance, in the area of interest. The evaluation of obtained results shows that the degree of success in forecasting NO_2 is promising.

Abstract *Le reti neurali sono state ampiamente utilizzate nella previsione dell'inquinamento atmosferico per la loro flessibilità nel trattare processi che coinvolgono dati non lineari e complessi. In questo studio, diverse reti neurali sono state allenate per la previsione della qualità dell'aria. I dati sull'inquinamento atmosferico, provenienti da due stazioni di monitoraggio di Pescara (Centro Italia), insieme ad alcuni parametri meteorologici, sono stati utilizzati per prevedere, con un giorno di anticipo, le concentrazioni di biossido di azoto nell'area di interesse. La valutazione dei risultati ottenuti mostra che il grado di successo nella previsione di NO_2 è promettente.*

Key words: Air quality, Artificial Neural Network, Multilayer perceptron, Forecast

---

Annalina Sarra
University of Chieti-Pescara, Viale Pindaro, 42, e-mail: annalina.sarra@unich.it

Adelia Evangelista
University of Chieti-Pescara, Viale Pindaro, 42, e-mail: adelia.evangelista@unich.it

Tonio Di Battista
University of Chieti-Pescara, Viale Pindaro, 42, e-mail: tonio.dibattista@unich.it

Francesco Bucci
Independent researcher , e-mail: frabucci@gmail.com

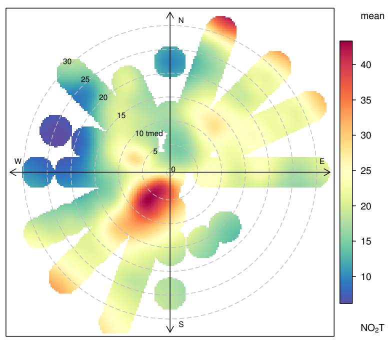
1 Introduction

In recent years, understanding the status and development of air quality levels in urban areas has become of paramount importance worldwide due to the severe effects of atmospheric pollution on human health [4]. In order to obtain accurate and comparable air quality information of a specific area, countries around the world have established air quality monitoring networks. By relying on data retrieved from monitoring stations, many studies are aimed at characterizing the underlying dynamic interaction of pollutant time series in different sites of the monitored areas. Over the years, the implementation of artificial intelligence techniques for air pollution time series modeling and air pollution-concentration forecasting as well, has significantly increased. Among them, Artificial Neural Networks (ANNs) have been widely used in air pollution prediction (see, among others, Hadjiiski and Hopke [3] and Boznar et al. [1]). In this paper, a modeling framework based on ANNs is explored to forecast Nitrogen Dioxide ($\mathrm{NO_2}$) levels, one day in advance, in the urban area of Pescara (Central Italy). Air pollution data from two monitoring stations, along with some meteorological parameters, are included in the analysis. The benefit of using ANNs in our research is twofold. Firstly, ANNs help to establish which meteorological variables have a strong impact on the behavior of the target air pollutant. Secondly, by focusing on prediction of air pollution in each station, ANNs might represent an effective way to investigate redundancy and optimize the layout of air quality monitoring networks. The rest of the paper is arranged as follows: Section 2 illustrates the study area and the available data, Section 3 briefly describes the ANNs framework used and Section 4 is devoted to present the main results and some concluding remarks.

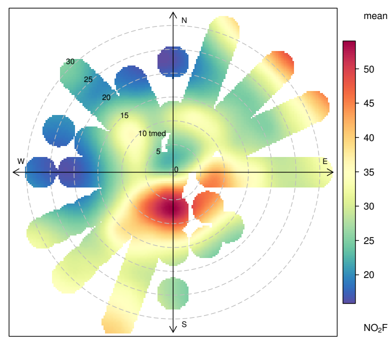
2 Study area and data

Air pollution data for this study consist of measurements of NO_2 obtained from Pescara hourly air quality reporting platform, run by Regional Agency for the Environmental Protection of Abruzzo Region (ARTA). Daily measurements of NO_2 pollutant have been collected from January 1, 2015 to December 31, 2017. Since weather strongly influences pollutants formation and transport, the analyzed data set also includes daily metereological variables, such as wind, rain, temperature.
NO_2 measurements were taken at two monitoring stations: one designed of *urban traffic type* and the other deemed as *urban background station*, representative of the population average exposure. The urban traffic (UT) station is located next to an one-way street (*Via Firenze*), characterized from constant vehicular traffic. The urban background (UB) station, close to *Teatro d'Annunzio*, is located away from urban traffic or other direct pollution sources. A first glance of the trend and variability of the NO_2 concentrations observed in the two monitoring stations reveals that there are, on average, higher values at the urban traffic station rather than in the background one, whereas the average difference between the concentrations measured

in the two stations is less marked in the summer semester (April- September). The highest concentrations of this pollutant are recorded at the urban traffic station, with an overall average of 34.2 $\mu g\ /m^3$. Wind vector data were used to help verifying the effects of synoptic meteorological conditions on NO_2 pollution. The bivariate polar plots of wind-direction and temperature contribution (Fig.1) show that in both stations the highest concentrations occur to a greater extent at low temperatures and this can be ascribed to domestic heating systems and combustion. However, there are higher concentrations observed even at higher temperatures which indicate the influence of road transport throughout the year.



(a) *Urban Background station.*



(b) *Urban Traffic station.*

Fig. 1: The bivariate polar plots with wind direction and temperature presented on a radial scale.

3 Artificial Neural Networks model for Nitrogen Dioxide Prediction

Artificial neural network (ANN) is a computational model inspired by the human brain functioning [5]. ANNs may be defined as structure comprising of a group of interconnected basic processing units, called neurons, associated with a learning rule. The neurons provide a parallel processing of the data. In a commonly used ANN architecture, known as the multilayer perceptron, the neurons are arranged in layers. Here, we used the *Multi-Layer Perceptron Neural Network architecture (MLPNN)*. The chosen architecture is made up of only two layers of multiple neurons (the input and hidden layer) and of a single neuron in the output layer. The elemental structure of the adopted MLPNN is described in Fig.2.

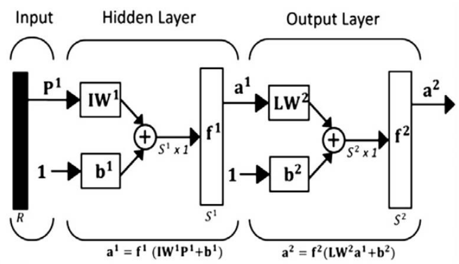


Fig. 2: Sketch of the MultiLayer Perceptron Neural Network (MLPNN) architecture

In this architecture, $\mathbf{P}^1$ is the input data vector, $\mathbf{IW}^1$, $\mathbf{LW}^1$ and $\mathbf{b}^i$ are the input and layer weight matrix and bias vectors that should be calculated by a training and validation procedure; $\mathbf{S}^1$ is the number of neurons in the hidden layer that should be optimized; $\mathbf{a}^i$ are the output vectors and $\mathbf{f}^i$ are the chosen transfer functions (tan-sigmoid and pureline). The input variables chosen for the considered built neural network consist of some meteorological parameters (i.e. average temperature, temperature max, temperature min, rain, wind speed, wind high, wind direction) and NO_2 observed concentrations.

4 Results and conclusions

ANNs are supervised learning techniques involving a training step to create a mathematical model and a prediction step to compute the output for a given set of input values using the model created in the training step. To define the best numbers of neurons and optimize the network we follow the procedure describe below [2].
For each monitoring station, the three years data 2015-2017 are divided into two data sets: the data for the years 2015 and 2016 constitute the set used to train the network whereas the data of the last year were employed in the validation phase. Hourly data have been aggregated into daily observations, and from each dataset missing values for more than 48 hours have been deleted. The Levenberg-Marquardt technique was used to train the chosen net, coupled with the repeated random sample validation procedure. The net structure was identified through an optimization process that provided the most favorable number of neurons in the hidden layer (S) through the

Mean Square Error (MSE) minimization procedure. For the training of the ANNs we used the Matlab Neural Networks Toolbox. In order to detect the optimum number of neurons for the hidden layer, we have built as many networks as those obtained by varying the number of neurons. In this respect, it worth noting that the optimum number of interior neurons was searched between 1 and 30. Each network has been trained 300 times and among them we have chosen the best one by looking at the minimum MSE. Subsequently, the selected structure has been trained to optimize the number of input variables, stopping the convergence process again in the minimum MSE. We built all the networks resulting from the all possible combinations of neurons and input variables. The better MLPNN networks in term of performance metrics are displayed in Table 1. According to the considered model indicators, the ANN with one neuron and all variables as input (model 1) has resulted the best architecture for the air quality prediction in the urban background (UB) monitoring station. As shown in Fig.3, this network has a very high accuracy in forecasting NO_2 concentrations. Looking at the values for the urban traffic (UT) monitoring station, from the aspect of absolute error (measured by RMSE and MAE) and min MSE, the most accurate prediction is achieved through the architecture of model 3. However, from Fig.3 it is evident that this model exhibits a poor performance in mimic the observed data at the beginning of time window analysed. Probably, the higher variability in the traffic volumes and the limited wind exposition of this monitoring site are at the basis of the initial poor accuracy of ANN that achieve better results for smooth data. Nonetheless, overall, the correlation coefficient for measured versus predicted NO_2 concentrations for both selected ANN models was shown to be greater of 0.70 (0.81 in the UB *vs* 0.71 in UT) over the span of 1 year.

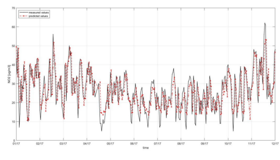
Table 1: Models performance main results

Test days	Station	Model	Inputs	Neurons	minMSE	RMSE	CVRMSE	nMAE	CORR
707	UT	1	all	5	85.64	9.25	26.91	19.83	0.75
707	UT	2	NO_2, tmed, tmax, ws	16	84.73	9.5	27.99	21.72	0.69
707	UT	3	NO_2, tmed, tmax, ws, wdir	23	73.82	8.6	25	18.74	0.79
707	UT	4	NO_2, tmed, rain, wh, wdir	4	84.42	8	29.55	23.12	0.7
707	UT	5	NO_2, tmed, tmin, rain, wh, wdir	9	74	10.04	29.6	22.72	0.65
709	UB	1	all	1	53.85	7.34	30.76	22.29	0.81
709	UB	2	NO_2, tmax, ws, wh	12	51.68	7.8	28.76	22.52	0.71
709	UB	3	NO_2, tmed, tmax, tmin, rain, ws, wd, wdir	16	43.81	7.58	27.97	22.24	0.71
709	UB	4	NO_2, tmax, tmin, ws, wh, wdir	9	43.91	8	26.55	23.12	0.7

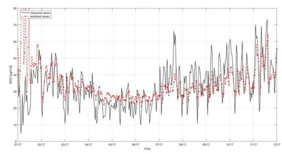
Legend:

minMSE = minimum Mean Square Error;
RMSE = Root Mean Square Error;
CVRMSE = Coefficient of Variation of RMSE;
nMAE= normalized Minimum Absolute Error;
CORR =Correlation coefficient

These are promising results: managers, authorities of urban air quality, practitioners and decision makers could efficiently exploit the toolkit of ANN modeling to estimate the temporal profile of pollutants and air quality indices. Additionally, the different factors involved in the input stages in the two monitoring sites could provide useful insights in ascertaining which variables affect the spatial distributions of Nitrogen Dioxide (NO_2) concentrations and to what extent redundancy information can be detected in the monitoring stations.



(a) *Urban Background station*.



(b) *Urban Traffic station*.

Fig. 3: Measured and predicted NO_2 concentrations at urban background and traffic stations.

References

[1] M. Boznar, M. Lesjak, and P. Mlakar. A neural network-based method for short-term predictions of ambient SO_2 concentration in highly polluted industrial areas of complex terrain. *Atmos. Environ.*, 27:221–230, 1993.

[2] C. Cornaro, F. Bucci, M. Pierro, F. Del Frate, S. Peronaci, and A. Taravat. Solar radiation forecast using neural networks for the prediction of grid connected PV plants energy production (DSP project). *Proceedings of 28th European Photovoltaic Solar Energy Conference and Exibition*, pages 3992–3999, Sept 30–Oct 4, 2013.

[3] L. Hadjiiski and P. Hopke. Application of artificial neural networks to modeling and prediction of ambient ozone concentrations. *J. Air Waste Manag. Assoc.*, 50(5):894–901, 2000.

[4] R. Kelishadi and P. Poursafa. Air pollution and non-respiratory health hazards for children. *Arch. Med. Sci.*, 6:483–495, 2010.

[5] R. Lippman. An introduction to computing with neural nets. *IEEE ASSP Mag.*, 4(2):4–22, 1987.