OPTIMIZATION OF NEURAL NETWORK PERFORMANCES BY MEANS OF EXOGENOUS INPUT VARIABLES FOR THE FORECAST OF OZONE POLLUTANT IN ROME URBAN AREA

A. Pelliccioni*, F. Pungì*, S. Lucidi**, V. Latorre**
* Ispesl-Dipia, Via Fontana Candida 1, 00040, Monteporzio Catone (RM), Italy
** University of Rome "La Sapienza"- DIS Via Ariosto 25 00185 (RM), Italy

INTRODUCTION

Air quality problems produced by high levels of ozone affect human health and are related to respiratory problems. Ozone is a reactive gas and presents concentrations which are dependent both from the meteorological conditions and seasonal effects. The time forecasting of Ozone levels is very complicated to obtain as described in different studies [2] [5]. For Ozone models the most difficult problems to deal with are the simulation of chemical reactions that occur in atmosphere, the contribution due to long range transport and the turbulence conditions [3]. Among the complex systems, an important tool in order to forecast air pollution data is the neural network (NN) [8] that can be used in assessing the non linear dynamics of such systems.
Another tool we used to forecast ozone is the support vector machine (SVM).
Both models (NN and SVM) have been used to forecast ozone using data at different temporal lags, and utilizing different input during training phase.

1. DATASET DESCRIPTION AND METHODOLOGY

In our work, NN and SVM methods have been developed to forecast hourly ozone levels using data from one to ten days in advance (T1-T10). We have analyzed data recorded by monitoring stations for the city of Rome for the calendar year 2005.
The objective of our work concerns the study of various benefits when considering in addition to conventional variables some exogenous variables as inputs for the NN and the SVM models. The role of exogenous variables is to optimize the convergence of mathematical models and to reproduce the ozone at different temporal lags.
As a consequence, as input variables we considered two sets of simulations, the first using only conventional data as pollutants and meteorological measurements (Conventional Data Set - CDS), and the second including some external data (e.g. time of the day, Julian day, day of the week, month of the year) in addition to the other conventional variables (Extended Data Set - EDS).
The data used in our simulations came from the monitoring stations of the ARPA LAZIO (Regional Agency for Environmental Protection in Lazio) network in the urban centre of Rome (Largo Magna Grecia), which recorded hourly data throughout the calendar year 2005.
The conventional variables (CDS) used for the simulations are:

1. monitored pollutants variables:
   - Carbon monoxide (mg/m^3) – CO
   - Nitrogen oxide ($\mu g/m^3$) – NO
   - Nitrogen dioxide ($\mu g/m^3$) – NO_2
   - Ozone ($\mu g/m^3$) – O_3 – (Input/Output variable)

2. meteorological variables:
   - Temperature (C°) – T
   - Global Solar Radiation (W/m^2) – GSR
   - Relative Humidity (%) – RH
   - Pressure (mbar) – P

The additional external variables (EDS) used for the second simulation set (e.g. time of the day) take into account seasonal effects and periodical turbulence conditions, and are to be considered as exogenous variables.
The inclusion of these variables:

a) takes account of hourly and seasonal average conditions
b) takes into consideration a simple periodic mathematical formulation as well as the trend of conventional variables
c) assists the conventional variables during the training of NN and SVM

1.1 Statistical considerations on input variables

Before carrying out the simulations using Neural Networks and SVM, we performed basic statistical analysis to examine in more detail the characteristics of pollution and meteorological input variables. We calculated the correlation matrix of the dataset at our disposal in order to assess the relationships between physical quantities.

Table 1 Correlation matrix

	CO-mg/m³	NO-µg/m³	NO2-µg/m³	O3-µg/m³	TMed-C°	URMed-%	PressMed-mbar	RadSol-W/m²
CO-mg/m³	1.00							
NO-µg/m³	0.83	1.00						
NO2-µg/m³	0.76	0.65	1.00					
O3-µg/m³	**-0.45**	**-0.48**	**-0.52**	**1.00**				
TMed-C°	-0.27	-0.34	-0.25	**0.53**	1.00			
URMed-%	0.14	0.19	0.11	**-0.52**	-0.33	1.00		
PressMed-mbar	0.24	0.26	0.28	**-0.19**	0.01	-0.04	1.00	
RadSol-W/m²	-0.11	-0.15	-0.19	**0.49**	0.53	-0.54	-0.02	1.00

By a preliminary analysis of the correlation matrix it can be seen that ozone in absolute value is correlated with other quantities at equal values (≈ 0.50), except for the mean pressure, which has a much lower level.
The values in Table 1 indicate specifically a positive correlation of ozone with the average temperature and solar radiation, and a negative correlation between ozone and the remaining quantities.

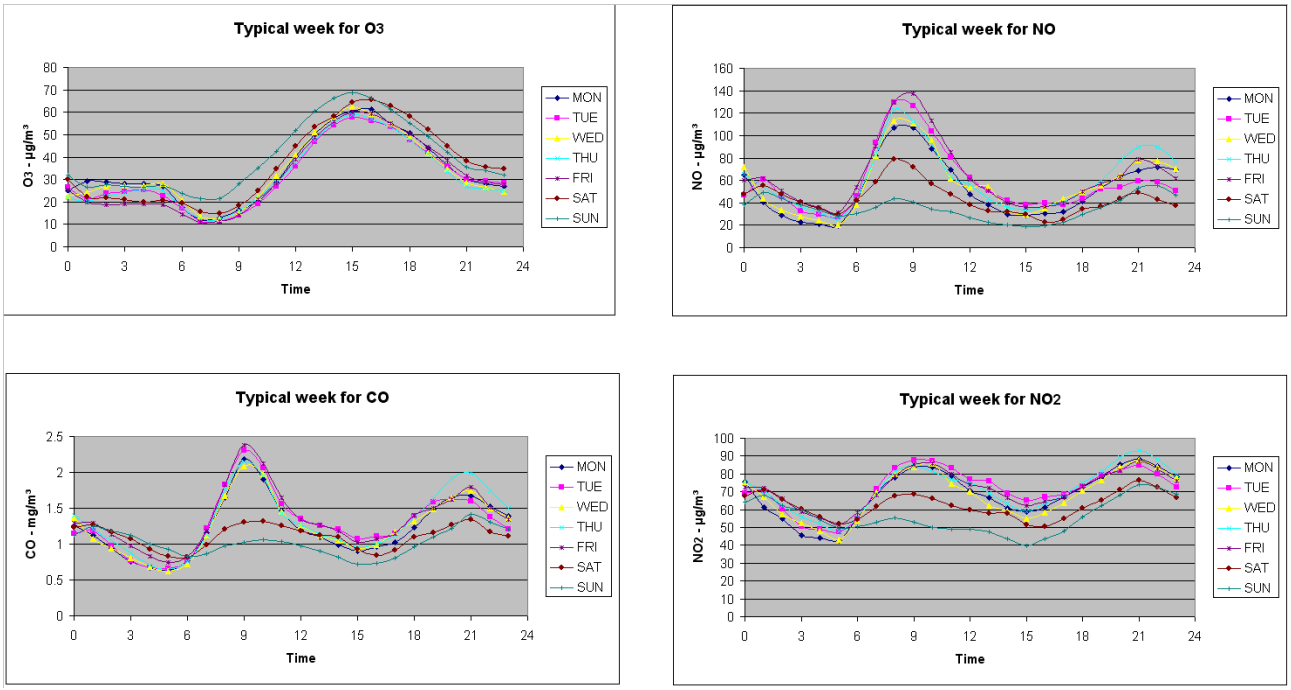

Figure 1 Typical week for the pollution variables

Figure 1 shows the levels of mean pollutants expressed in µg/m^3 (CO in mg/m^3) at different times during different day of week.
These trends represent a typical week for each pollutant. From the figure we can see a different trend by type of pollutant, i.e. whether it is a primary pollutant (CO, NO) or secondary (O_3).
For primary pollutants type, we may notice an increase in the levels at the hours when there is a peak of the emission, i.e. the time slot in the morning (7- 9) during which the contribution is tied to the traffic source.
We can also see from the figure that the trends for the primary pollutants that have lower levels correspond to Saturday and Sunday, where typically there are less traffic sources.
Ozone (O_3) instead is a pollutant that is not produced by man and his activities, and has to be considered a secondary pure.
Its presence is linked to the reactivity of the atmosphere. It is a marker of photochemical activity.
Its background levels are high in unpolluted atmospheres, and are low in those polluted.

This is mainly due to the reactivity with nitrogen oxides (NO, NO_2) and the OH radicals produced by pollution, leading to lower concentrations due to the photochemical activity (at high temperatures and sunlight, typical of the Mediterranean climate).
From Figure 1 it can be observed that its concentration level decreases in correspondence of the peaks of the oxides of nitrogen, namely in the morning time, it increases during the hours of highest insolation in which is lower the emission-related component of these oxides and again it changes during evening hours where there is a new increase of the traffic source.
Table 2 shows the corresponding values of the pollutants that confirm what has previously been observed in the Figure 1.

Table 2 Typical week: µg/m^3 of O_3, NO, NO_2 (average), mg/m^3 of CO (average)

Average O3-µg/m³	Time																							
Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
MON	24.93	29.32	29.03	27.94	27.94	26.84	17.65	12.08	12.88	16.36	20.16	28.53	40.04	50.35	57.33	60.68	61.29	54.63	50.57	44.00	36.44	29.92	28.22	26.93
TUE	26.68	21.60	23.63	24.39	25.18	22.56	17.16	11.30	11.44	14.14	19.15	26.98	35.76	46.92	54.43	57.83	56.18	53.59	47.59	41.79	35.80	30.19	29.31	28.64
WED	22.52	24.29	27.06	27.15	27.25	28.07	20.75	13.72	12.52	15.71	22.26	31.49	41.11	51.05	57.86	62.59	59.22	55.18	49.18	41.61	34.73	28.43	26.44	24.48
THU	22.04	20.44	22.82	24.15	25.25	24.93	17.53	12.33	12.18	15.98	21.96	28.85	37.44	47.43	55.46	59.09	57.02	53.33	47.62	41.32	33.74	26.81	25.66	25.04
FRI	25.45	19.77	18.80	18.98	19.16	18.93	14.54	11.09	11.36	14.13	21.18	28.31	39.13	48.68	55.75	60.09	59.21	55.16	50.17	44.65	38.47	31.55	28.76	27.68
SAT	30.16	22.10	22.02	21.04	20.07	20.74	19.41	15.73	14.84	18.26	25.12	34.54	44.93	53.37	58.09	64.44	65.48	62.83	58.16	52.26	44.82	38.13	35.62	34.60
SUN	32.02	26.56	27.53	27.00	26.47	27.32	23.63	21.31	21.47	27.99	35.29	42.65	51.73	60.38	66.37	68.50	66.15	61.16	55.13	49.22	41.97	35.66	34.08	32.06

Average CO-mg/m³	Time																							
Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
MON	1.25	1.13	0.94	0.76	0.67	0.64	0.77	1.15	1.64	2.19	1.90	1.46	1.20	1.09	0.99	0.90	0.97	1.03	1.24	1.48	1.65	1.67	1.52	1.39
TUE	1.14	1.18	0.97	0.77	0.67	0.66	0.76	1.21	1.83	2.30	2.06	1.56	1.35	1.26	1.21	1.08	1.11	1.15	1.39	1.58	1.62	1.59	1.37	1.20
WED	1.36	1.07	0.94	0.81	0.68	0.62	0.72	1.13	1.68	2.09	1.97	1.45	1.25	1.13	1.05	0.94	0.99	1.18	1.32	1.49	1.64	1.75	1.50	1.37
THU	1.41	1.20	1.04	0.88	0.71	0.64	0.75	1.14	1.75	2.15	1.95	1.43	1.27	1.14	1.06	0.95	1.00	1.12	1.38	1.63	1.92	1.99	1.75	1.50
FRI	1.30	1.30	1.13	0.97	0.82	0.74	0.83	1.19	1.83	2.38	2.12	1.65	1.34	1.27	1.19	1.02	1.07	1.15	1.40	1.50	1.62	1.79	1.52	1.34
SAT	1.24	1.27	1.17	1.07	0.93	0.83	0.81	0.99	1.21	1.30	1.32	1.26	1.19	1.12	1.10	0.93	0.84	0.91	1.09	1.16	1.27	1.35	1.17	1.11
SUN	1.12	1.24	1.18	1.12	1.00	0.92	0.83	0.87	0.97	1.02	1.06	1.03	0.98	0.90	0.81	0.72	0.73	0.81	0.97	1.10	1.22	1.41	1.30	1.21

Average NO-µg/m³	Time																							
Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
MON	64.85	40.94	28.98	22.59	20.89	19.19	41.90	81.57	107.02	106.75	88.57	69.35	47.66	38.20	29.47	29.03	30.27	32.32	41.67	52.82	63.55	68.58	72.12	69.71
TUE	46.81	61.03	46.34	32.91	29.92	26.93	46.06	93.83	129.63	126.75	104.16	80.08	62.06	51.49	42.19	38.61	39.84	38.59	43.87	52.46	53.61	59.35	58.16	50.92
WED	71.60	43.60	33.77	28.43	24.17	19.91	38.54	82.03	113.06	110.60	95.94	61.99	53.88	54.42	33.19	29.90	35.00	43.43	49.64	55.74	63.78	78.42	77.42	70.34
THU	65.77	54.67	44.55	36.46	31.05	25.64	43.78	82.50	122.46	112.13	91.89	63.56	55.88	43.34	36.01	33.51	36.96	40.68	50.03	59.63	77.54	89.92	89.89	76.60
FRI	60.35	61.88	50.52	40.92	36.18	31.11	54.13	90.30	129.19	137.60	113.28	84.88	60.97	50.79	39.80	35.86	36.46	39.56	49.84	57.08	63.57	78.77	72.88	61.55
SAT	47.45	55.46	47.93	40.21	34.96	29.71	41.85	58.75	79.02	71.42	57.37	47.66	38.18	33.09	31.96	29.39	22.63	24.66	34.41	37.01	43.68	49.41	42.76	37.16
SUN	38.51	48.96	43.46	38.43	33.37	28.19	30.60	36.21	43.39	40.42	34.03	31.62	26.53	22.74	19.90	18.35	19.40	22.32	29.60	36.08	42.26	53.31	55.23	46.48

Average NO2-µg/m³	Time																							
Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
MON	75.43	61.22	54.84	45.73	44.25	42.77	55.38	69.59	78.12	84.01	84.06	78.72	70.38	66.47	60.89	58.60	61.31	67.25	72.56	78.40	85.07	88.23	84.46	78.23
TUE	68.87	70.71	60.44	50.26	48.90	47.54	54.97	71.55	83.50	87.92	87.34	83.28	76.83	75.78	69.09	65.14	66.93	68.45	72.86	79.18	81.67	84.74	79.79	72.76
WED	75.16	66.99	57.26	52.27	47.82	43.37	53.97	69.37	79.90	83.63	86.00	74.47	70.20	62.26	59.10	55.14	58.16	63.78	70.68	76.52	83.60	87.99	83.96	77.90
THU	75.54	70.43	62.04	56.03	51.92	47.81	55.66	69.35	81.29	84.35	81.40	77.06	75.05	69.31	63.01	59.05	63.59	68.18	75.18	81.19	89.12	93.34	88.00	79.50
FRI	72.48	72.24	66.03	58.91	55.45	51.53	58.26	68.06	78.75	85.48	85.55	79.95	74.05	72.05	66.50	62.47	64.66	67.33	73.01	78.51	82.02	87.49	82.63	75.88
SAT	67.40	70.89	65.65	60.09	55.94	51.79	54.37	61.63	67.75	68.63	66.01	62.06	59.81	57.79	57.75	51.56	50.65	55.09	61.02	65.39	70.96	76.24	72.73	66.53
SUN	64.00	68.18	62.18	59.64	54.21	49.73	50.35	52.73	55.40	53.14	50.21	49.10	48.87	47.77	43.73	39.81	43.63	48.19	55.76	62.13	68.15	74.19	72.63	68.60

Figure 2 shows the values of measured variables in according to time of day and month of the year. Time of the day usually is linked to the main turbulence conditions related to solar elevation, geographical positions, and seasonal effects and so on.
Synthesizing, each hour can be strictly connected with turbulence distributions that are typical of each site.
The hour of the day and the month have to be considered as exogenous variables and they allow to increase the performance of neural networks and support vector machines during training, as allow to discriminate different situations for the meteorological and pollutants linked to seasonally and hourly variations.
In Figure 2 are evident the effects of these situations, that happen when high gradient of the variables appear on the maps. The maps suggests us how to optimize the NN and SVM models by an additional information respect to the conventional variables.
We could observe that the exogenous variable "time of the day" induces a discriminating level stronger in the daytime than at night.

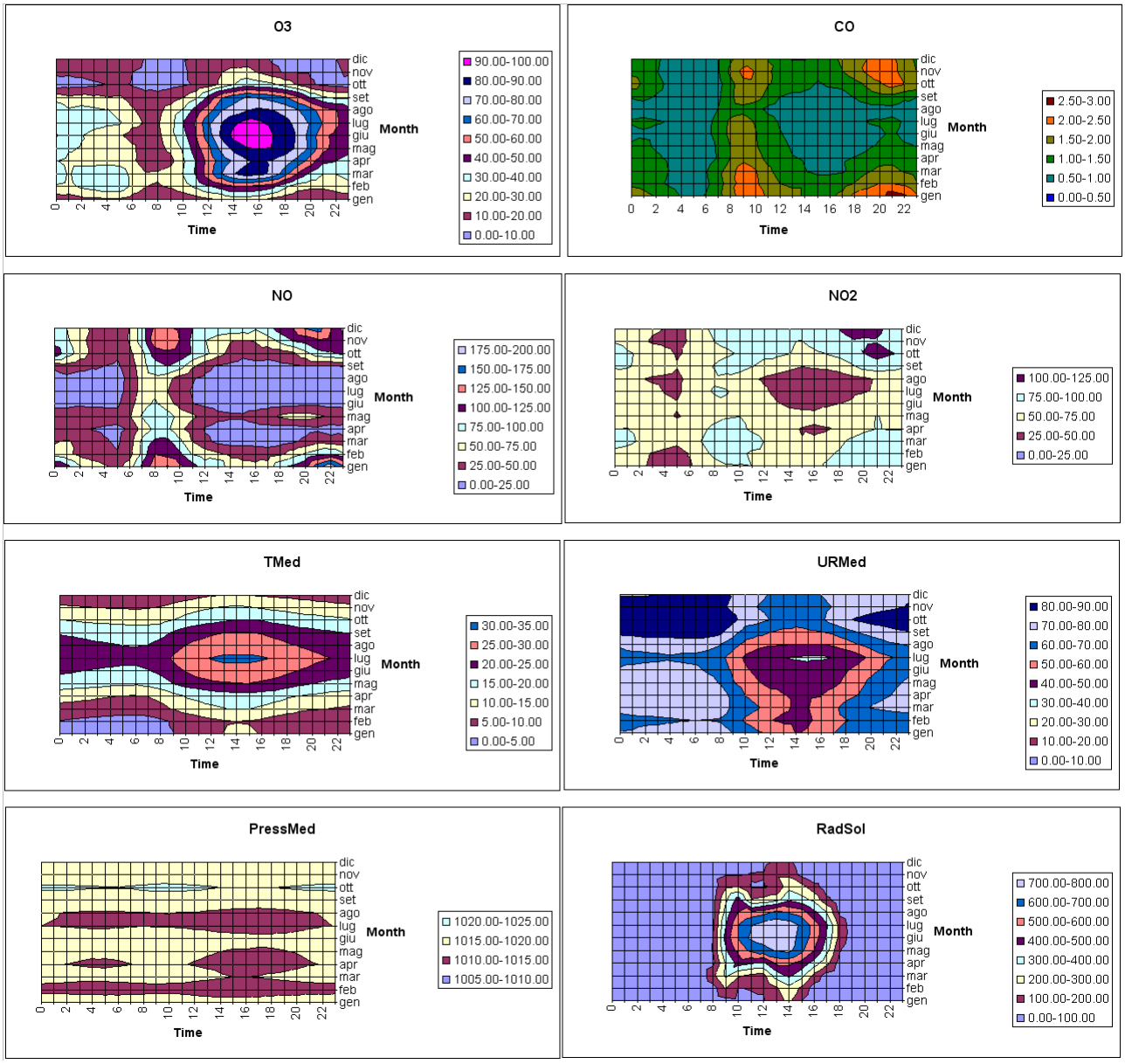

Figure 2 Surface maps

All variables that are constant in the map are not discriminating variables.
Among all, the variable whose map varies more is the ozone that has an absolute maximum of about 90 $\mu g/m^3$ for the summer months from 14 p.m. to 16 p.m.
For the same hours, it can instead be observed the minimum contributions of NO_2, which are evidently connected with the photochemical origin of the ozone itself, also evidenced by the absolute maximum temperature and solar radiation.
CO map shows instead some minima at the same hours during the summer months, which indicates its primary nature and the influence of turbulence on the levels of pollution.
Primary pollutants roughly have a limited variability, while for ozone is observed a marked seasonality in the data.
In short we can say that from the observation of these maps, the exogenous variables are very important to relate ozone with seasons and turbulence conditions during typical day.
In particular we can see from maps that the time of the day has a greater intrinsic variability than monthly (tied to the seasons).
This consideration justifies in the present work, taking time of the day as the main and only external variable to be considered in the simulations.
The above allows us to say that this variable may be very important to improve the performance of our nonlinear models.

1.2 Fundamentals considerations on the use of NN and SVM

The Multi Layer Perceptron (MLP) is the most commonly used neural network in the field of air quality prediction [4]. Figure 3 shows schematically a typical MLP network where O_3 (T) is the variable to predict at time T, while I (T-$d_{1...10}$), M (T-$d_{1...10}$), Ex (T-$d_{1...10}$) and O_3 (T-$d_{1...10}$) are the input variables to the network, from 1 to 10 days before.
It may lead to different results in accordance with the choice of activation function and number of neurons of hidden layer. As activation function we use the standard sigmoid.
A different choice for the activation function could improve the network performance, but given the complexity of our task, we focused essentially on patterns and variables of the net rather than on the algorithm optimization itself.
We trained the network with 5, 10, 15, 20, 25 and 30 neurons for hidden layer, and finally we chose 20 that gives the best performances in terms of minimizing the error function and computational efficiency.

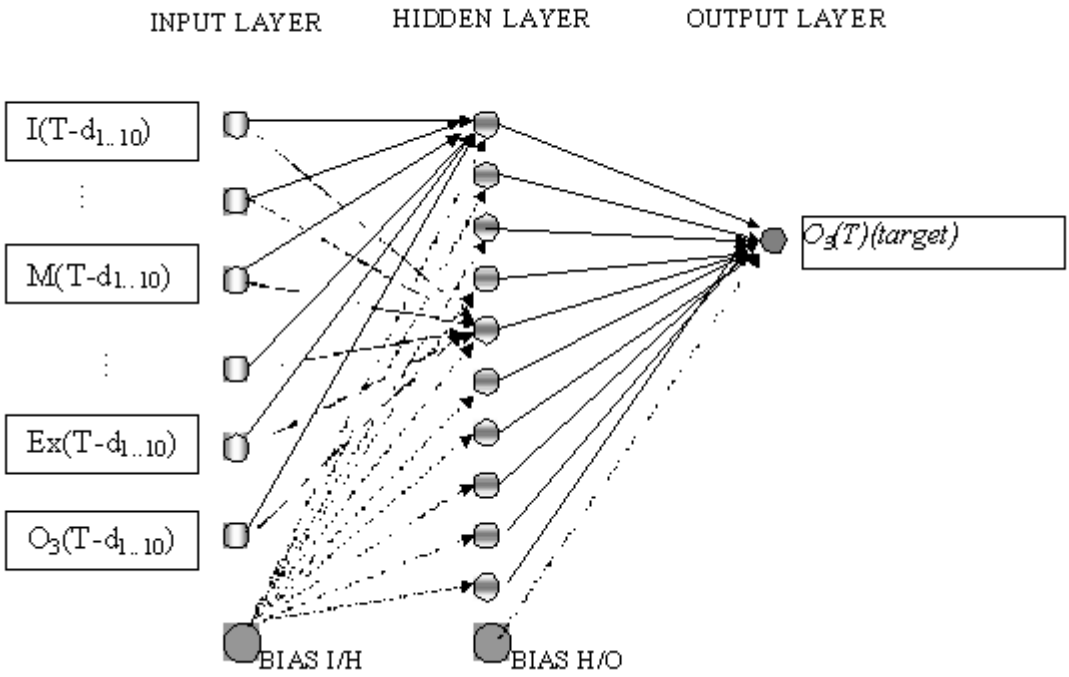

Figure 3 MLP architecture for Ozone forecasting

We presented here the results obtained using a classical architecture for a neural network consisting of a single MLP with one hidden layer of 20 neurons and an output layer with 1 neuron.
For all sets of simulations we used a 3-Layer Perceptron model, which is considered capable of approximating any measurable function [1]. The first layer contains the input variables of the neural network related to all relevant physical parameters, as well as the exogenous variable in the case of the second set.
The second layer consists of neurons of the hidden layer. The third layer is the output layer, which consists of the target variable to be reproduced, i.e. the hourly Ozone concentration.
The NN parameters were obtained by a training procedure based on the use of an efficient unconstrained minimization algorithm.

Table 3 Algorithms for training neural networks

Training Algorithms	Epochs	Learning Rate	Transfer Function	
			Hidden Layer	*Output Layer*
Gradient Descent Backpropagation	20000	0.05	Sigmoid	Linear
Scaled Conjugate Gradient Backpropagation	1500	–	Sigmoid	Linear
Levenberg-Marquardt Backpropagation	500	–	Sigmoid	Linear

For training the network we used 60% of the original data for each simulation, leaving 40% as the testing phase, to evaluate the performance of generalization of the model.
As confirmed by the literature, this partition is balanced enough to ensure good performance in generalization.
In general, the task of NN training is to find the optimum weights of the NN by means of input/output pattern presentation, thus enabling the Neural Network to simulate chemical reactions and turbulence dispersion of the Ozone levels.
Support vector machines (SVM) are a set of related supervised learning methods used for classification and regression [7]. Viewing input data as two sets of vectors in an n-dimensional space, an SVM will construct a separating hyper plane in that space, one which maximizes the margin between the two dataset. To calculate the margin, two parallel hyper planes are constructed, one on each side of the separating hyper plane, which are "pushed up against" the two datasets.
Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the neighbouring data points of both classes, since in general the larger the margin the lower the generalization error of the classifier.
We compared the results of NN and SVM for the ozone concentrations, considering the CDS and the EDS separately.

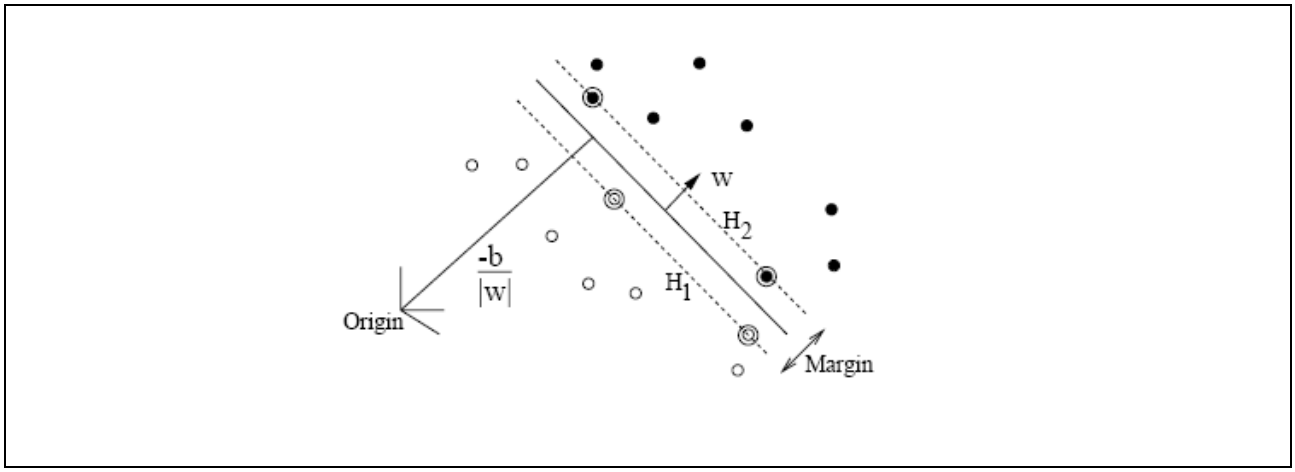

Figure 4 Linear separating hyper planes for the separable case. The support vectors are circled

2. RESULTS AND CONCLUSIONS

Aim of the work is to demonstrate that exogenous variables may improve the NN and SVM models when they are added to the conventional input variables.
To show this, we must distinguish the contribution due to the model, from that due to the variables. For this purpose, we show the performances considering, other than NN and SVM, the conventional statistical model (Multi Regression Model - MR). All models (MR, NN, SVM) are trained with identical data set at different temporal lags. So doing, results can be compared and we can evaluate the contribution both of each models category and input variables choice.
We considered four different types of simulations, two relating to the CDS and two to the EDS.
Simulations on the CDS are:

- I+M, where the variables taken into account are all the meteorological variables and air pollutants except ozone (Background Simulation - BS)
- I+M+O_3 that includes also the Ozone variable besides the variables of the BS

Simulations on the EDS are:

- I+M+Ex, where the variables used for training neural networks and SVM are identical to I+M with the addition of one or more exogenous variables (time of day, day of week, Julian day, month of year). Here we focused our attention on the influence of daily cycle on ozone and so we show results adopting only time of the day as exogenous input.
- I+M+O_3+Ex, which, in addition to having as inputs all the variables of the simulation above, also has the Ozone variable as input.

For all four types of simulations, the inputs were considered at different time lags from 1 to 10 days in advance. For each simulation set, and for each temporal lag within the simulation set, we calculated the coefficient of determination R^2, the bias and the Mean Absolute Error (MAE) for the target ozone levels.
While R^2 is a dimensionless parameter, the bias and MAE are given in μg /m^3. Bias is related to the ozone background reproduced by different models and the optimum value could be zero. MAE index, expresses the averaged absolute error between measured and modelled ozone. If average error is zero, this means that all ozone levels are well reproduced in average. Bias and MAE are related to different meanings. While the bias concerns the bad behaviour of models to reproduce the background, MAE is related to the systematic error on the mean values [6].

Our work was the result of about 150 simulations, where each one is composed of a training phase and a testing phase.
In the Table 4 results coming from the final simulations (120) are given.
In general the results showed that SVM and NN performed better than MR.
We calculated the various performances indices for the target ozone using the conventional statistical regression model (MR), the MLP and the SVM for the CDS and EDS where the additional exogenous variable was the time of the day.

Table 4 R^2, Bias and MAE for different simulations

R^2	I+M			I+M+Ex			I+M+O3			I+M+O3+Ex		
	MR	NN	SVM	MR	NN	SVM	MR	NN	SVM	MR	NN	SVM
T1	**0.44**	**0.62**	**0.63**	**0.53**	**0.69**	**0.70**	**0.67**	**0.71**	**0.72**	**0.67**	**0.74**	**0.74**
T2	0.38	0.56	0.59	0.45	0.64	0.66	0.55	0.62	0.64	0.56	0.68	0.70
T3	0.39	0.53	0.54	0.46	0.57	0.63	0.50	0.56	0.61	0.52	0.64	0.66
T4	0.38	0.53	0.55	0.45	0.60	0.64	0.50	0.57	0.60	0.51	0.64	0.66
T5	0.39	0.54	0.56	0.47	0.62	0.63	0.50	0.58	0.59	0.52	0.66	0.65
T6	0.38	0.54	0.58	0.46	0.62	0.65	0.49	0.61	0.63	0.51	0.64	0.66
T7	0.37	0.54	0.57	0.46	0.62	0.64	0.49	0.59	0.62	0.52	0.64	0.66
T8	0.36	0.52	0.56	0.43	0.61	0.65	0.48	0.57	0.61	0.49	0.64	0.65
T9	0.34	0.52	0.54	0.44	0.63	0.64	0.46	0.54	0.60	0.49	0.65	0.65
T10	0.34	0.50	0.52	0.41	0.60	0.64	0.44	0.54	0.61	0.46	0.62	0.65

Bias (μg/m^3)	I+M			I+M+Ex			I+M+O3			I+M+O3+Ex		
	MR	NN	SVM	MR	NN	SVM	MR	NN	SVM	MR	NN	SVM
T1	**19.43**	**12.05**	**9.56**	**16.55**	**10.10**	**7.91**	**11.81**	**8.89**	**6.94**	**11.60**	**8.12**	**6.83**
T2	20.43	13.26	10.72	17.83	10.62	9.03	14.40	11.34	9.31	14.18	9.72	8.20
T3	20.52	14.30	11.56	17.71	11.44	9.96	15.76	12.48	10.01	15.23	10.18	8.09
T4	20.01	14.02	11.58	17.45	11.76	8.94	16.02	13.04	10.17	15.44	10.65	8.10
T5	21.25	15.17	11.28	18.79	12.13	9.87	17.83	13.86	10.14	17.23	11.00	8.80
T6	21.40	14.97	11.08	18.34	11.47	9.47	17.17	12.46	9.07	16.35	10.65	8.99
T7	21.77	15.02	11.39	18.92	11.92	9.32	17.57	13.25	9.75	16.88	11.17	8.93
T8	21.72	15.42	11.82	19.01	12.95	8.91	18.09	13.80	9.15	17.41	11.45	9.53
T9	22.23	16.21	11.79	19.23	12.43	9.11	18.50	14.76	10.45	17.64	11.76	9.56
T10	22.23	15.20	12.12	19.82	11.89	9.46	18.89	13.80	9.93	18.33	11.05	9.25

MAE (μg/m^3)	I+M			I+M+Ex			I+M+O3			I+M+O3+Ex		
	MR	NN	SVM	MR	NN	SVM	MR	NN	SVM	MR	NN	SVM
T1	**17.10**	**13.52**	**12.55**	**15.58**	**12.42**	**11.63**	**12.55**	**11.81**	**11.26**	**12.47**	**11.30**	**10.93**
T2	17.74	14.53	13.47	16.64	13.18	12.55	14.51	13.34	12.43	14.56	12.25	11.37
T3	17.47	14.93	14.36	16.36	14.29	13.07	15.08	14.26	13.17	15.00	13.06	12.21
T4	17.91	15.22	13.99	16.92	14.04	12.89	15.74	14.52	13.34	15.66	13.28	12.24
T5	17.90	15.25	14.19	16.78	13.80	13.14	15.89	14.51	13.22	15.67	13.15	12.63
T6	17.95	15.47	14.25	16.78	13.82	12.87	15.93	14.10	13.22	15.76	13.40	12.41
T7	18.21	15.43	14.19	16.84	13.85	12.70	15.94	14.32	13.31	15.69	13.78	12.36
T8	18.48	15.60	14.42	17.42	14.15	12.81	16.48	14.68	13.43	16.27	13.50	12.70
T9	19.00	15.73	14.52	17.62	13.96	13.04	16.96	15.07	13.57	16.59	13.53	12.75
T10	18.49	15.77	15.05	17.43	13.90	12.96	16.72	15.00	13.76	16.48	13.59	12.83

As highlighted by the table, NN and SVM perform better (R^2 from 0.50 to 0.74) than the classic statistical linear regression model (R^2 from 0.34 to 0.67). These results confirm that non linear models (NN and SVM) perform better respect to the linear ones (MR).
The most interesting results concern the NN and SVM performances when we add the exogenous variables to the conventional dataset, as we can see from the increasing of the R^2 values.
For that regards the bias, the NN models don't decrease so much when we consider exogenous variables. In fact, we have 14.56 μg/m^3 at I+M simulations up to 10.57 μg/m^3 at I+M+O_3+Ex. The SVM works better

respect to the NN. We calculate for the average ozone background 11.29 μg/m^3 at I+M, and 8.63 μg/m^3 at I+M+O_3+Ex simulations. Bias calculated by MR models presents always greater values respect to the NN and SVM models.
For that regards the value of MAE, we find a similar behaviour obtained with bias coefficient. It has to be observed that the ozone error levels are lower respect to the bias ones for the I+M simulations, and greater when we consider I+M+O_3+Ex using NN and SVM.
In general, we see that the best performances in terms of R^2, bias and MAE are obtained obviously for simulations lag T1. It is noted, however, a clear distinction of individual performance curves according to the models used and the inputs used in the models.
Using R^2 as index, we may notice in fact that the external variable (Ex) offers a contribution in the performance of the multiple regression model (MR) of about 20%, while we note an increase of the performance of neural networks (NN) and SVM with its introduction of about 12% and 13% respectively.
However, it is worth noting that the initial average of R^2 is 0.38, 0.54 and 0.56 for MR, NN and SVM respectively, and therefore the best performances are still related to the NN and SVM with a slight majority of the latter.
For the MR model, the variable that shows a significant improvement compared to the background simulation (I+M) is ozone (O_3), with an increase of about 34 % against an increase of about 9.9% for the other two models (NN - SVM).
Evidence of the importance of the association between ozone and external variable is highlighted by the last simulation in the Table 4 (I+M+O_3+Ex), in which all inputs are considered.
Compared to BS, it can be observed that the MR model increases by about 40% compared to an increase of 21% and 19% for NN and SVM respectively, keeping a sharp improvement in the final R^2 of the latter (0.74) against the MR model (0.67).
This simulation shows that is the combination of the two variables that provides significant performance improvement for artificial intelligence (AI) models compared to classical statistical models.
So the neural network models and support vector machines provide still better performance values.
Regarding all these classes of models, SVM seems to have slightly higher performance than the neural networks for the value of the bias and MAE. The values of the coefficient of determination are very close but with a slight majority in this case for the NN.
Both these models are sensitive to the introduction of both exogenous (Ex) and O_3 variables, and observing the values in the table we see that while the introduction of ozone variable for training helps to improve the shorter time lags (1 days), the exogenous variable contributes to an improvement in the forecast for longer time lags (3-5 days).
In conclusion we can say that the AI models still provide performances superior to those of statistical models, although the latter exhibit the best gains with the addition of ozone and external variables.
For both AI models we can say that they are equivalent with a slight majority for the neural network models against the SVM according to the performance indicators used.
Finally, our work suggests that using the exogenous variables as input significantly improved the results of simulations and suggested a way of optimizing the environmental simulation using NN and SVM models approach.

REFERENCES

1. Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press, Oxford.

2. Comrie R.S, (1997). Comparing neural network and regression models for ozone forecasting. Journal of the Air and Waste Management Association 47, 653-663.

3. Dutot, A.L., Rynkiewicz, J., Steiner, F.E. and J.Rude, 9 September 2007: A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions

4. Gardner M.W, Dorling S.R., (1998). Artificial Neural Networks (the Multilayer Perceptron) - E Review of applications in the atmospheric sciences, Atmos. Environ., 32(14/15), 2627-2636.

5. Gardner, M.W and S.R. Dorling, 2000: Statistical surface ozone models: an improved methodology to account for non-linear behaviour. Atmospheric Environment 34, 21-34.

6. Hanna, S.R.," Confidence limits for air quality models, as estimated by bootstrap and jackknife resampling methods. Atmospheric Environment, 23, 1989, 1385-1395.

7. Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press.

8. Rojas R., (1996): Neural Networks: a systematic introduction, Springer-Verlag, Berlin Heidelberg.