Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
ENV2016-2010

Hourly Predictions of PM10 Emissions for Emergency Episode Perception Comparatively using ANN and SVM-R Techniques

Fatih Taspinar
Assistant Professor
Duzce University
Turkey

Zehra Bozkurt
Assistant Professor
Duzce University
Turkey

Fatih Aktas
Assistant Professor
Duzce University
Turkey

An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. This paper has been peer reviewed by at least two academic members of ATINER.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

Taspinar, F., Bozkurt, Z. and Aktas, F. (2016). **"Hourly Predictions of PM10 Emissions for Emergency Episode Perception Comparatively using ANN and SVM-R Techniques",** Athens: ATINER'S Conference Paper Series, No: **ENV2016-2010.**

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209 Email: info@atiner.gr URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm

ISSN: 2241-2891
12/10/2016

Hourly Predictions of PM10 Emissions for Emergency Episode Perception Comparatively using ANN and SVM-R Techniques

Fatih Taspinar

Zehra Bozkurt

Fatih Aktas

Abstract

Predicting short-term levels of air pollutant's emission is an important issue in the air pollution field considering its potential to help local authorities to enact preventative measures. The aim of the present work is to comparatively develop reliable artificial neural network and support vector machine regression models (ANN and SVM-R) to obtain accurate predictions of future hourly PM_{10} concentrations in the atmosphere for one hour ahead. Hourly based air pollution data with meteorological factors have been used in training and testing stages of constructed models. The inputs to models were structured using a combination of time sequenced meteorological and air pollution data from a lagged-dataset. Time series plots of the data showed seasonal air pollution of PM_{10} emissions whose hourly values were up to 891 μg/m^3. Thus, in the pollution monitoring region the PM_{10} levels may adversely affect human health. The obtained models were utilized in establishing an alert system for online PM_{10} emissions based on the threshold values reported in the National Air Quality Assessment and Management Regulation. ANN models optimized with Broyden Fletcher Goldfarb Shanno (BFGS) algorithm and SVM-R parameters were optimized by grid search. The results of the statistical analysis of the models have shown reasonable accuracy in terms of perception of short-term pollution episodes. A simple program for hourly PM_{10} prediction was also developed using the best ANN and SVM-R models.

Keywords: Air pollution, ANN, Hourly forecasting, PM10, SVM-R.

Introduction

The investigation of air pollution in metropolitan centers by researchers demonstrated that there are a lot of risks for urban citizens because of high levels of particulate matter (PM) even in relatively small concentration in ambient area. The human health in urban areas has been affected dramatically by atmospheric particulate pollution (Pope et al., 2006). PM consists of a mixture of liquid and solid matters that enter the atmosphere by natural (e.g., dust storms, volcanoes, forest and grassland fires) and anthropogenic (e.g., construction works, industry, central heating, vehicular traffic, domestic heating) sources (Wang et al., 2014). The levels and composition of air PM may present various physical and chemical patterns depends on the region's climatology and geology. PM_{10}, which has an aerodynamic diameter less than or equal to 10 μm, has primary (e.g., road traffic, tire abrasion and combustion process) (Hooyberghs et al., 2005, Brunelli and Holgate 2007, Amato et al. 2011) and secondary (chemical reactions such as NO_2 to HNO_3 and SO_2 to H_2SO_4, condensation vapors) sources (Keary et al., 1998). According to European union (EU) air quality standards for particulate matter, PM_{10}, annual limit value (40 $\mu g/m^3$) and 24 hour concentration limit (50 $\mu g/m^3$) not to be exceeded more than a specified number of times (35) in a year (Burges, 1998).

The human health still has been exposed to a significant threat, even though PM air quality in Europe has improved in the late years (EEA, 2007). Some significant health problems can be seen via short-term exposure to PM like morbidity and mortality (Brunekreef and Holgate, 2002). The particulate matter has been defined as a possible carcinogen because of so strong effects on health if someone prolonged exposure to it (Dockery et al., 1993). This is a significant concern have to be solved for Turkey. The aim of air quality management is to take precautions before pollution episodes do not to reach acceptable limits, e.g. for PM_{10} is 90 $\mu g/m^3$. In this scope, accurate forecasting of air pollution episodes in many areas has been used for economic and business planning, weather forecasting, signal processing and product controlling. Air quality modelling tools are employed for long term and short term scales. Short-term forecasting systems may be used to constitute warning systems and decreasing effects to prevent serious events. Moreover long-term forecasting systems are the best tools for design and management of transportation networks, residential areas and industrial sites and urban planning (Baklanov et al., 2007).

In the recent last five decades, forecasting and time series analyses were a highlight research area. Researchers thus developed new forecasting models for improving the effectiveness and handling of them. Conventional statistical methods are used to make short-term predictions; however, in the last decade Artificial Neural Networks (ANNs) are preferred often for predictions in time series data analysis, pattern recognition and classification applications. ANNs are simplified mathematical models and working as brain system and can learn functional dependencies, associations and patterns via training data. ANN models have very good results for forecasting of a wide range of pollutants and their concentrations at various time scales. The use of ANNs has been extended for modeling of PM pollutants (e.g. PM_{10}, $PM_{2.5}$) recently (Grivas and Chaloulakou, 2006). ANNs are expected to produce better PM forecasting performance compared to the traditional methods since they have a powerful ability to capture highly nonlinear physical processes. ANNs have also been employed in identifying pollution origins, such as PM_{10} particles, sulphur monoxide (SO), carbon monoxide (CO) (Perez et al., 2004; Perez and Reyes, 2006; Brunelli et al., 2007).

Another reliable and effective technique is the support vector machines (SVMs) for regression (SVM-R) and classification analysis. As a recent statistical learning technique, SVMs were used for forecasting the pollutant levels in different time series. Both techniques are the best candidate for best testing and forecasting performances on the most data set. The risk minimization principle is the main difference between ANNs and SVM. As a learning algorithm that carries out accurate classification learning by prediction of previously undetected data, SVM utilizes for observing indirect patterns in complex data sets. The SVM model has several advantages such as containing fewer free parameters, better prediction than the other conventional neural network model (Burges, 1998; Smola and Schölkopf, 2004). In the recent years, researchers focused on web-based air pollution prediction systems for utilizing to warn health care including environmental and traffic management to prevent negative effects (Kurt et al., 2008).

In this study, the special emphasis was put on to estimate hourly PM_{10} levels in the build environment for one-hour ahead, using ANN and SVM-R models with the inputs from meteorological factors and PM_{10} dataset. A simple graphical user interface (GUI) for obtained models have been also developed for future use (e.g. developing a web interface with the models for emergency perception for air pollutants) and tests are made with data from other regions.

Materials and Methods

Area of Interest and Data Used

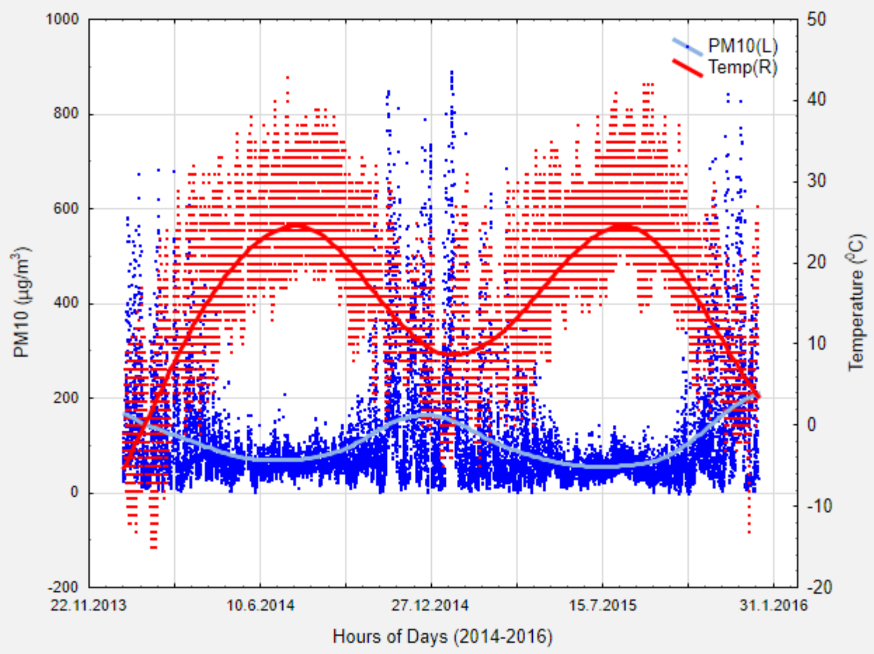
The study area, where the data is collected, Düzce province is located in the Blacksea Region in north western Turkey which is a neighbour of the most industrialized region of Turkey, namely Marmara Region. Düzce have typical terrestrial climates, sharing Black Sea climatic. According to the 2013 census, the population density of Düzce provinces was 351509. The study area covers the most used highways in Turkey. The highway of D100, connecting the western and eastern part of Turkey, passes through the study area and hence vehicle traffic related emissions strongly influence on the area. In addition, despite the widespread use of natural gas in the region either in the residential sector or industry, the use of the coal and biomass sources for heating residential purposes still causes particulate air pollution resulting in the elevated PM_{10} levels during cold months. The study area has a hilly to mountainous topography which covers about 85% of the province. Because of the mountainous geography surrounding the area, the prevailing wind flows to the central district of Düzce is blocked, leading to exposure to high air pollution levels. Due to temperature inversion, the weather in Düzce is foggy for about six months during winter time, from October to March. A low inversion layer above the downtown area of the city can be observed in winter and summer. Hence, the pollution episodes are commonly associated with inversion and the smog caused by the use of fossil fuels for heating.

The main objective of this study is to forecast an emergency pollution episode, so a low-resolution data set is needed. An hourly dataset involving in data from meteorological parameters and PM_{10} levels were gained for the years 2014-2016 for Düzce (days between 01.01.2014-13.01.2016 from hours 00:00 to 23:00), covering 17809 data rows. The parameters in the data set consisted of hourly values of wind speed (WS, m/s), relative humidity (RH, %), air temperature (AT, ^{0}C), wind direction (WD, deg.) and PM_{10} hourly concentration in μg/m^3. Here, a forward-lagged set of these variables was constructed for including in the prior data one-hour before. Table 1 gives the descriptive statistics of the variables and temporal variations in PM_{10} levels with AT were plotted in Figure 1. The maximum PM_{10} concentrations can be seen, particularly at least five months from October to March due to residential heating with fossil fuels in heating seasons, in contrast to those levels observed during the summer periods. PM_{10} and air temperature values were ranged in 0-891 μg/m^3 and -15-43 ^{0}C, respectively. However, the mean and 75% percentile of PM_{10} level were 103.44±110.13 and 113 μg/m^3, respectively, which is higher than the acceptable limit of 90 μg/m^3. The statistics showed that the atmosphere over Düzce is highly polluted by the particulate matter and the pollution episodes particularly during winter periods can affect human health adversely.

Table 1. *Descriptive Statistics of Hourly Dataset (2014-2016) Used for Investigation*

	Valid N	**% Valid obs.**	**Mean**	**Median**	**Mode**	**Freq. of Mode**	**Min.**	**Max.**	**25% Perc.**	**75% Perc.**	**Range**	**Std. Dev.**
PM10	17325	97	103.44	66.00	53.00	233.00	0.00	891.00	44.00	113.00	891.00	110.13
Temp	17466	98	15.89	17.00	20.00	707.00	-15.00	43.00	8.00	23.00	58.00	10.30
WD	17466	98	202.13	224.00	262.00	289.00	0.00	359.00	152.00	264.00	359.00	87.75
WDI	17466	98	1.00	0.97	0.11	289.00	0.00	2.00	0.29	1.72	2.00	0.71
WS	17466	98	0.95	1.00	1.00	9804.00	0.00	7.00	0.00	1.00	7.00	0.90
RH	17466	98	77.60	84.00	103.00	2252.00	-10.00	103.00	60.00	98.00	113.00	23.35

Figure 1. *Hourly Time Series Plot of PM_{10} over Air Temperature*



Methodology of Artificial Neural Networks

The artificial neural networks, which are inspired from the biological nervous systems, are adaptive nonlinear systems capable to approximate any function. ANNs are used for overcoming the problems faced in regression and classification studies in which the inspired model that does not have a clear relationship between its inputs and outputs in general (Rumelhart et al., 1986). ANN involves a network of simple processing elements, so-called neurons. They are connected to each other in specified layers between the processing elements. Generally, the structure of ANN is made up of a number of layers with neurons, namely multi-layer perceptron (MLP) ANN. A single neuron processes multiple inputs applying an activation function on a linear combination of the inputs as follows:

$$y_i = f\left(\sum_{q=1}^{l} w_{iq}.f\left(\sum_{j=1}^{m}(v_{qj}x_i + b_j)\right) + b_q\right) \quad (1)$$

where x_j is the set of inputs, w_{iq} and v_{qj} ate the synaptic weights connecting the q^{th} input to the j^{th} neuron, b is bias term, f is the activation or transfer function, and y_i is the output of the i^{th} neuron. Weights are the knowledge base of the ANN system, which represents the non-linear properties of the neuron by its activation function. The activation function is usually non-linear, with a sigmoid shape such as logistic or hyperbolic tangent function as follows:

$$sig(x) = \frac{1}{1+e^{-X}} \quad (2)$$

$$tanh(x) = \frac{1-e^{-2X}}{1+e^{-2X}} \tag{3}$$

Generally, feedforward MLP networks are trained using an error back propagation (BP) algorithm (Lahmiri, 2011), which covers heuristic and numerical optimization algorithms. Heuristic techniques include gradient descent and the resilient algorithm (Dong and Zhou, 2008). On the other hand, numerical techniques cover conjugate gradient (Fletcher-Reeves update, Polak-Ribiére update and Powell-Beale restart), Quasi-Newton (Broyden-Fletcher-Goldfarb-Shanno, BFGS method) and Levenberg-Marquardt (LM) algorithms (Morris, 2000).

The MLP neural networks consist of three types of layers including input, hidden and output layers (Figure 2). The first layer corresponds to the input variables to the problem with one node for each input variable. The second layer used to capture nonlinear relationships among variables by interconnections. The third layer provides the predicted values. The weights of neurons are usually started with random values taken from a standard normal distribution. During an iterative training process, ANN calculates an output o(x) for given inputs and current weights. When the training process is completed, the predicted output (o) will be closer to the input (y) with an acceptable error. To measure global training error, an error function E such as the sum of squared errors could be used as follows:

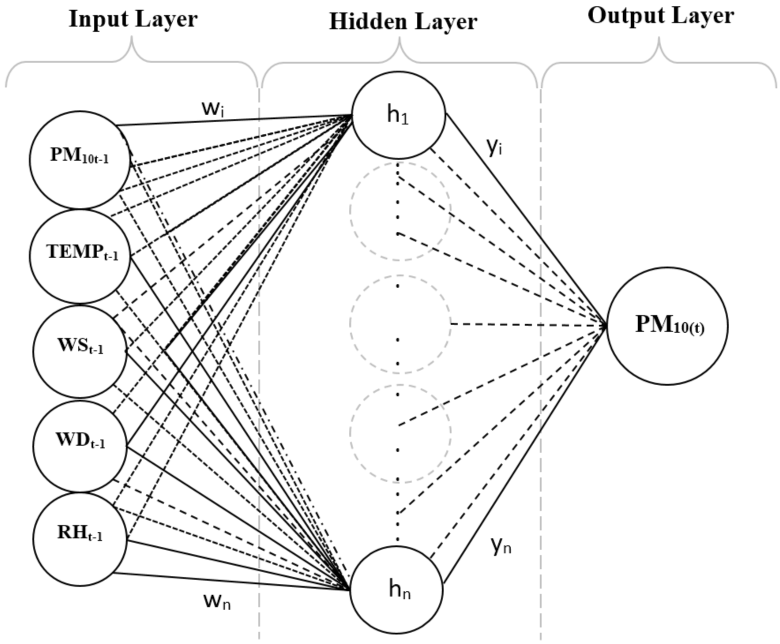
$$E = \frac{1}{L.M}\sum_{i=1}^{L}\sum_{j=1}^{M}\left(o_{ij} - y_{ij}\right)^2 \tag{4}$$

It calculates the average of cumulative differences between predicted and observed. Finally, the process stops if a pre-specified criterion is fulfilled such as checking early stopping conditions by calculating E. In this study, BFGS methods were operated on the data. Here, in the light of the preliminary tests conducted on partial data and of its fast convergence rate and accuracy on the huge time series dataset, the BFGS method is selected for optimization of weights of ANN models in this study.

Support Vector Machines for Regression

Support vector machines (SVMs) are supervised learning methods used for classification and regression (Vapnik, 1995). SVMs have been developed by Vapnik (Vapnik et al., 1997) and gained popularity due to promising features such as better accuracy and empirical performance. They are generalized linear classifiers with learning ability. SVMs can be used a classification and regression tool, running based on the machine learning theory. This algorithm maximizes predictive accuracy by automatically avoiding over-fit to the data (Burges, 1998). It is also used in many applications like pattern classification and regression applications.

Figure 2. *The Structure of ANN Models Used*



A tutorial prepared by Smola and Schoelkopf gives a good overview of the regression problem using SVMs (Smola and Schölkopf, 2004). In regression with SVM, the input (x) is first mapped onto an *n*-dimensional feature space using some fixed (nonlinear) mapping and a linear model is then constructed in this feature space. The linear model (in the feature space) $f(x, \omega)$ using a mathematical notation is given as follows:

$$f(x,\omega) = \sum_{i=1}^{n} \omega_i . g_i(x) + b \tag{5}$$

where b is the "bias" term and $g_i(x)$ $(i=1,\ldots,n)$ stands for a set of nonlinear transformations. The data are often assumed to be zero mean which can be achieved by data preprocessing, so the bias term is removed. A loss function $L(y, f(x, \omega))$ is defined to measure the quality of estimation. In the SVM regression a new type of loss function called as ε-insensitive loss function is proposed by Vapnik, 1997:

$$L_\varepsilon(y, f(x,\omega) = \begin{cases} 0, & , if \ |y - f(x,\omega)| \le \varepsilon \\ |y - f(x,\omega)| - \varepsilon & , otherwise \end{cases} \tag{6}$$

The empirical risk, showing the average loss of an estimator for a finite set of data is given by:

$$R_{emp}(\omega) = \frac{1}{n} \sum_{i=1}^{n} L_\varepsilon(y_i, f(x_i, \omega) \tag{7}$$

The idea of risk minimization does not only measure the performance of an estimator by its risk, but to actually search for the estimator that minimizes risk over distribution. Later, SVM regression performs a linear regression in the high-dimension feature space using ε-insensitive loss. Also, it tries to reduce model complexity by minimizing $\|\omega\|^2$ which can be described by introducing (non-negative) slack variables ζ_i, ζ_i^* $i=1,..,n$. These variables measure the deviation of training samples outside the ε-insensitive zone. Then SVM regression can be written as follows which minimizes this functional:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ &\textit{subject to} \quad \begin{Bmatrix} y_i - f(x_i,\omega) \le \varepsilon + \xi_i \\ f(x_i,\omega) - y_i \le \varepsilon + \xi_i^* \\ \xi_i\ ,\xi_i^* \ge 0, i = 1,\ldots,n \end{Bmatrix} \end{aligned} \tag{8}$$

where the constant $C > 0$ determines the tradeoff between the flatness of *f()* and the number of deviations higher than *ε*. Actually, this is the dual problem of which solution is given as follows:

$$f(x) = \sum_{i=1}^{n_{SV}}(\alpha_{\mathrm{i}} - \alpha_l^*).K(x_i,x) + b \quad \textit{subject to } 0 \le \alpha_l^* \le C, 0 \le \alpha_i \le C \tag{9}$$

where n_{sv} stands for the number of Support Vectors (SVs) and the kernel function $K(x, x_i)$ has been defined as a linear dot product of the nonlinear mapping, i.e.,

$$K(x,x_i) \equiv \varphi(x_i)^T \varphi(x_j) \tag{10}$$

The Gaussian Radial Basis Function Kernel (RBF) is by far one of the most versatile kernels. Radial basis function is used here and is the most preferred kernel for non-linear transformations when not much is known about the data to being modelled as follows:

$$K(x,x_i) = \exp(-\gamma.\left\|x - x_i\right\|^2), \quad \textit{where} \quad \gamma = (\frac{1}{2\sigma^2}) \tag{11}$$

where σ indicates the width of radial basis function.

Data Feature Extraction and Pre-processing

In order to have an applicable time series dataset, data pre-processing were applied to the entire dataset, involving in the variables PM_{10}, AT, WD, WS and RH. Firstly, we applied to a list-wise local linear regression to fill the missing values up to six cells by columns, but, the bigger missing areas remained. Thus the average valid data was about 91% of the entire dataset. The parameter WD is converted to a wind direction index (WDI) to avoid the discontinuity according to the following expression:

$$WDI = 1 + \sin\left(\mathrm{WD} + \frac{\pi}{4}\right) \tag{12}$$

Later, the variables were normalized in the range of 0.05-0.95 using min-max normalization given in Eq. (6) as follows:

$$y^{'} = 0.05 + \frac{(y - y_{\min})}{(y_{\max} - y_{\min})} * 0.95 \tag{13}$$

where $y^{'}$ is the normalized value, y_{min} is the minimum value, y_{max} is the maximum value and y is the actual value.

Results and Discussion

ANN Model Training and Performance Evaluation

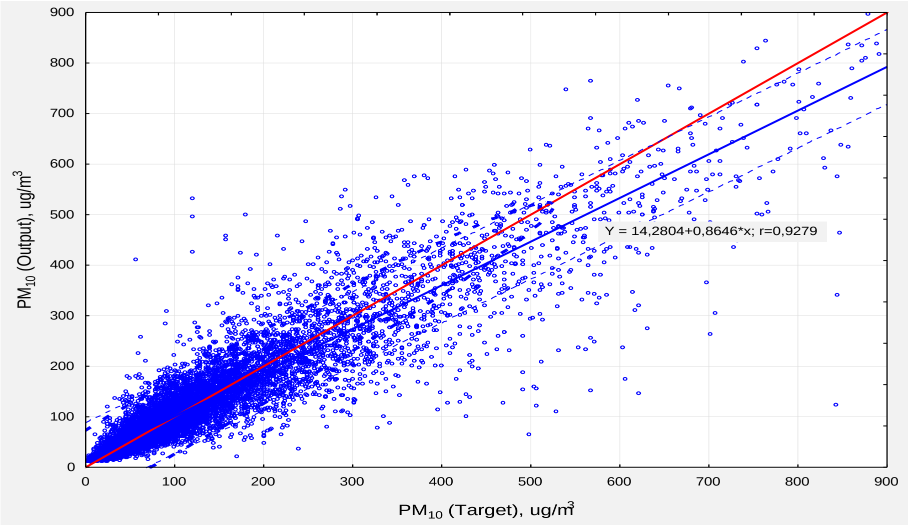
In order to predict one-hour ahead PM_{10} level, numerous ANN models with different layer structures were employed. Firstly, the entire dataset is divided into training (70%), test (20%) and validation sets (10%). ANN model parameters were then set by using logistic, hyperbolic tangent (tanh) and exponential activation functions for hidden and output layers. Input scheme to ANN models were given in Figure 2. Hidden layer neuron counts were changed in a range of [5-20] for BFGS learning method. Maximum number of epochs was set to 5000, applying an early stopping criterion to avoid over fitting setting the validation process at every epoch. In the model evaluations, sum of squared error (SSE), Index-of-Agreement (IA) and co-efficient of determination (R^2) can be used (Taspinar and Bozkurt, 2014). IA values closer to 1.0 indicate the better agreement with the selected model. The smaller values of SSE denote the better model performance and values close to 1.0 for R^2 indicate the greater explained variance. The ANN topologies showing input-hidden-output layer neuron counts with training, testing and validating statistics are given in Table 2.

Table 2. *ANN Model Topologies with Training and Testing Statistics*

ANN Model Topology	Training Set			Test Set			Validating Set		
	SSE	IA	R^2	SSE	IA	R^2	SSE	IA	R^2
5-7-1	1008.15	0.67	0.75	1014.61	0.66	0.74	1016.41	0.65	0.74
5-9-1	996.76	0.69	0.76	1002.40	0.67	0.76	1002.08	0.68	0.77
5-13-1	964.92	0.73	0.79	971.03	0.72	0.76	966.37	0.71	0.78
5-15-1	956.42	0.78	0.86	963.59	0.77	0.85	964.16	0.76	0.84
5-17-1	978.81	0.74	0.83	980.75	0.72	0.82	984.39	0.73	0.83
5-20-1	982.33	0.71	0.81	985.67	0.70	0.79	986.33	0.69	0.80

Generally, all the ANN models were very successful in the predictions comparing IA values ranged in 0.69-0.78 and R^2 values ranged in 0.74-0.86. The best scores for tested ANN models were produced by ANN (5-15-1) model with IA of 0.783, R^2 of 0.861 and SSE of 956.47. However, extreme value analysis for predictions showed that PM_{10} values higher than 550 $\mu g/m^3$ were predicted with a R^2 of 0.61 whereas values lower than 350 $\mu g/m^3$ were predicted with a better rate of R^2 (0.73), so R^2 increased with decreasing PM_{10} levels. These values of R^2 were lower with the other ANN models used. The performance plot of the results of the best ANN model for the whole dataset is visualized on Fig. 3. The red line indicates an exact fit of R^2=1.0 and blue line indicates a linear fitting line of R^2=0.861. 95% confidence bands and elliptic confidence limit are also shown in blue-dotted lines, indicating that most of the data points fall in the confidence bands.

Figure 3. *ANN (5-15-1) Model Performance Plot for Whole Dataset*



SVM Model Design and Performance Evaluation

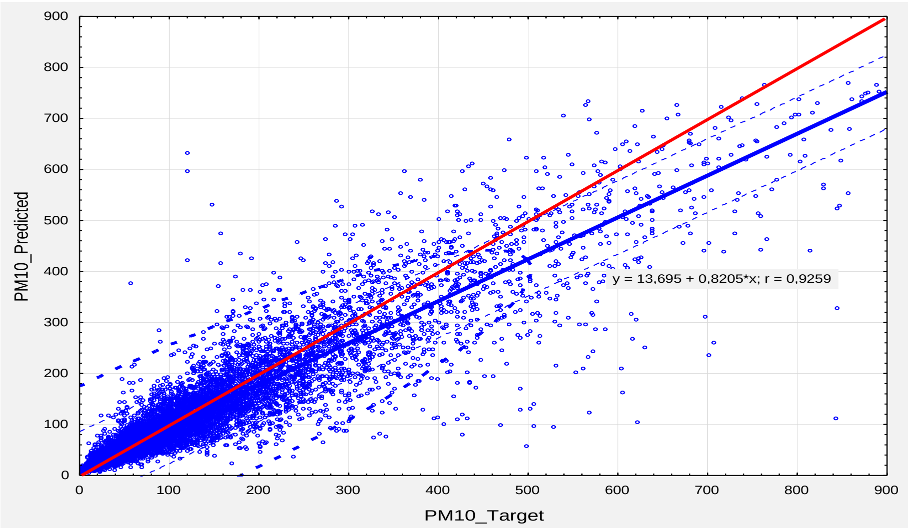
There are several issues needed to consider in the SVR application. Firstly, some parameters must be determined before running the particular algorithm of SVM learning. These parameters are error acceptance ε, capacity C and RBF kernel parameter γ. Theoretically, the spread parameter γ greatly affects the prediction performance. Both too large and too small values may lead to poor predictions. Therefore, in practical applications, for controlling the complexity these parameters should be determined by two iterative techniques cross-validation or grid search on the given datasets (Mitchell, 1997). We used ten-point grid search on the dataset (same as used in ANN models, see Figure 2) by setting ε in [0.001-0.10], C in [0.10-10] and γ in [0.10-1]. The best parameters with the lowest SSE were selected as the SVM-R model for one-hour ahead PM_{10} prediction. Iteratively setting these parameters within a SVM-R engine, the best SVM-R model with parameters of C:0.10, ε:0.01, γ:0.10 (training R^2=0.857, testing R^2=0.841 and SSE of 965.27) were obtained. Table 3 gives the parameter settings of some tried SVM models with related descriptive statistics.

Table 3. *SVM-R Model Parameters and Training Statistics*

	C	**ε**	**γ**	**R^2**	**SSE**	**IA**	**SVs**
SVM-R_1	0.1	0.001	0.1	0.77	1002.33	0.72	7893
SVM-R_2	*0.1*	*0.01*	*0.1*	*0.86*	*965.27*	*0.77*	*9334*
SVM-R_3	1	0.05	0.2	0.72	981.39	0.71	9271
SVM-R_4	5	0.1	0.5	0.67	991.65	0.70	8967
SVM-R_5	10	0.1	0.7	0.63	1004.90	0.69	7469

SVM engines calculate support vectors that hold information for discriminating data points by minimizing error for all model inputs. The best SVM-R model (here SVM-R_2) produced 9334 support vectors (SVs) for each input variable, which equals to 46670 SVs in total. The performance plot of the best SVM-R_2 model over the whole dataset is visualized in Figure 4. The red line indicates an exact linear fit of R^2=1.0 and blue line indicates linear fitting line for data with R^2=0.858. 95% confidence bands and elliptic confidence limit are also shown in blue-dotted lines, which indicates the most of the data points were fallen in the confidence bands.

Figure 4. *SVM-R Model (C:0.10, ε:0.01, γ:0.10) Performance Plot for the Whole Dataset*



Experimental Results for Time Series Data

In order to analyze the performance of models over time series data, predictions of the best ANN and SVM-R models were plotted as showed in Fig. 5 (ANN model is on the left and SVM-R model is on the right). Both of the models followed the data pattern due to temperature changing by fitting the peaks during winter periods. However, the ANN model slightly covers more data points than those of the SVM-R model. Here, it's important to investigate the model's behaviors for extreme conditions such as very stable atmospheric conditions with high PM_{10} levels for emergency perception. Therefore, to see the outcomes at extreme conditions, some data points were analyzed and some statistics were summarized in Table 4.

Figure 5. *Time Series Plots for the Predictions of the Best ANN and SVM Models*

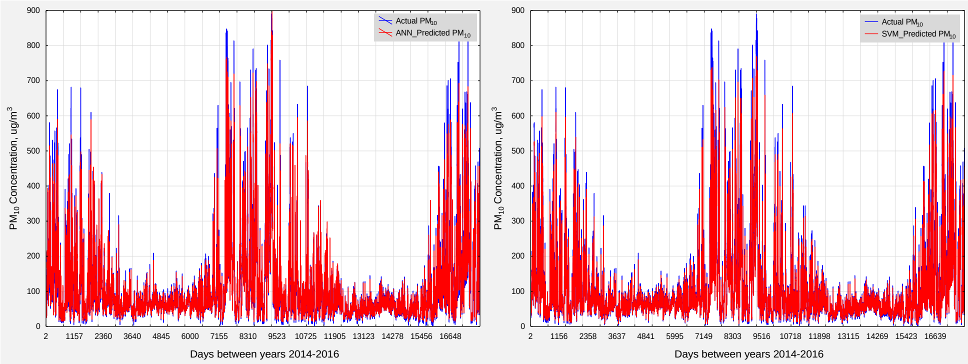
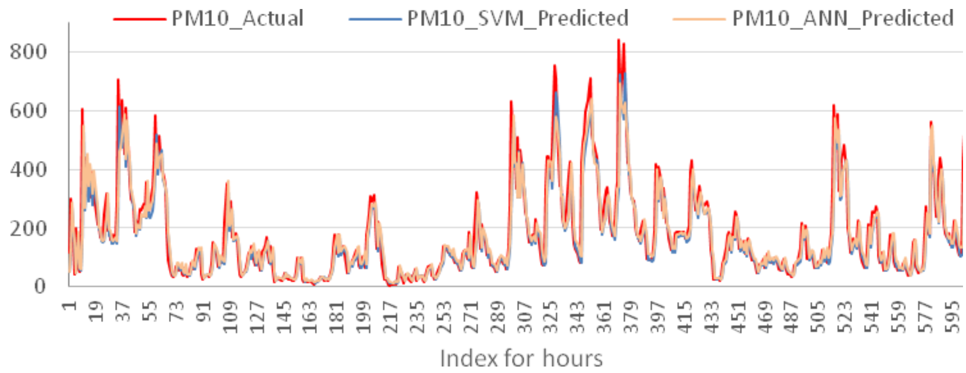


Table 4. *Predictions by the Best ANN and SVM Models for Some Conditions*

Date/Time	$PM_{10(t)}$	$PM_{10(t-1)}$	$AT_{(t-1)}$	$WS_{(t-1)}$	$RH_{(t-1)}$	$PM_{10(t)}$ by ANN	$PM_{10(t)}$ by SVM
15.12.2014 22:00	438	594	6	0	100	395.35	384.06
01.01.2015 07:00	61	58	2	0	100	61.47	59.44
22.02.2015 00:00	310	337	-1	0	100	327.10	310.75
25.02.2015 09:00	111	76	9	0	100	78.82	74.68
03.11.2015 08:00	133	107	6	1	100	100.92	111.07
10.11.2015 16:00	72	58	20	1	33	72.23	59.02
11.11.2015 06:00	54	46	11	1	100	48.35	45.83
20.11.2015 00:00	423	462	9	1	100	404.78	415.56
20.11.2015 20:00	620	437	12	1	84	426.22	398.06
23.11.2015 22:00	391	404	17	1	71	418.11	368.64
27.11.2015 00:00	130	121	16	1	70	127.22	114.40

Even though the sudden decreases in hourly PM_{10} levels are generally well handled by both of the models, ANN models gives slightly better prediction performance under these conditions. However, at the situations with increasing hourly PM_{10} levels plus bigger increments as before, SVM-R models yielded slightly better results. In order to visualize a specific time scale from 3.11.2015 00:00 to 28.11.2015 23:00, Figure 6 is rendered for actual $PM_{10(t)}$ vs. predicted $PM_{10(t)}$ for both ANN and SVM models.

Figure 6. *Specific Hourly Time Series Plots for Models within the Days 3.11.2015-28.11.2015*

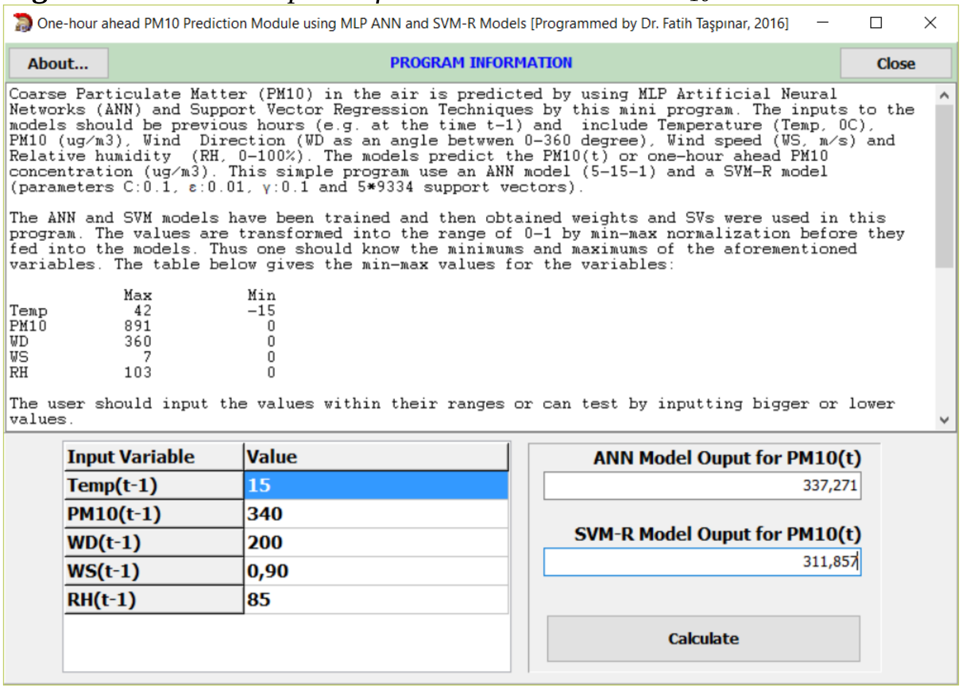


It can be noticed that peak levels were followed by the models but the ANN model produced the bigger values (e.g. at index of 37, 370 and 595) than the actual values mostly at extreme conditions, that is ANN model is overpredicted, however, decreasing PM_{10} levels were very-well predicted by ANN model. On the other hand, the SVM model did not tend to overpredict in general as time series plots indicated, but slightly tended to underpredict at most of the situations, but produced reasonable values at extreme conditions.

GUI Development for ANN and SVM Models

The produced models consisted of coefficients the weights, such as the floating point number, which cannot be used easily. Either to test other values or to provide a test area for other regions, we've developed a simple GUI for one-hour ahead PM_{10} predictions. The ANN and SVM models have been trained and then programmed using obtained model data. The input values are transformed into the range of 0.0-1.0 by min-max normalization before they fed into the models and WS is transformed to WSI internally. Fig. 7 shows the GUI for the PM_{10} predicted. This simple program can predict the PM_{10} levels for the best ANN model (5-15-1) and SVM-R model (parameters C:0.10, ε:0.01, γ:0.10) one-hour ahead. These models will be a base for web-based online prediction system. This will be a part of emergency situation perception system in advance.

Figure 7. *GUI Development for One-hour Ahead PM_{10} Prediction*



Conclusions

The main aim of this study was to obtain the ANN and SVM models in the predictions of PM_{10} levels one-hour ahead. The special emphasis is put on developing reasonable ANN and SVM models to capture big variations in PM_{10} levels during heating seasons. The well-optimized models have been produced by numerous trials with different model structures. The models fed with one-hour lagged terms of PM_{10}, AT, WS WD and RH to predict the next hours PM_{10} level. It has been noticed that the most significant lagged inputs were consecutive lagged-terms of PM_{10} and AT. The best ANN model is in the form of 5-15-1 with testing R^2=0.85 and BFGS learning method was the best technique to use a huge dataset such as in this study. The best SVM-R model is obtained for parameters of *C*: 0.10, ε: 0.01, γ: 0.10 with testing R^2=0.84). The ANN model tended to overpredict at extreme conditions but yielded slightly better results in general. On the other hand, the SVM model did not tend to overpredict as time series plots indicated, but slightly tended to underpredict at most of the cases, however, produced reasonable values at extreme conditions. We concluded that an emergency perception strategy can be designed based on these two models considering extreme values and average pollution levels. Hence, under average conditions ANNs can be employed until bigger increments in PM_{10} levels have been detected. It is possible to use these two models in hourly predictions conditionally by setting a lower threshold to run the ANN model. At the higher PM_{10} levels over the given threshold SVM-R model can executed concurrently. This approach appeared to be promising and could have significant applicability in the case of the huge datasets. A GUI application have also been developed to use the models and to make tests on untrained data from other monitoring stations as a real-life application.

References

Amato, F., M. Pandolfi, T. Moreno, M. Furger, J. Pey, A. Alastuey, N. Bukowiecki, A. S. H. Prevot, U. Baltensperger and X. Querol (2011). Sources and variability of inhalable road dust particles in three European cities. Atmospheric Environment 45(37): 6777-6787.

Baklanov, A., O. Hänninen, L. H. Slørdal, J. Kukkonen, N. Bjergene, B. Fay, S. Finardi, S. C. Hoe, M. Jantunen, A. Karppinen, A. Rasmussen, A. Skouloudis, R. S. Sokhi, J. H. Sørensen and V. Ødegaard (2007). Integrated systems for forecasting urban meteorology, air pollution and population exposure. Atmos. Chem. Phys. 7(3): 855-874.

Brunekreef, B. and S. T. Holgate (2002). Air pollution and health. The Lancet 360(9341): 1233-1242.

Brunelli, U., V. Piazza, L. Pignato, F. Sorbello and S. Vitabile (2007). Two-days ahead prediction of daily maximum concentrations of SO2, O3, PM10, NO2, CO in the urban area of Palermo, Italy. Atmospheric Environment 41(14): 2967-2995.

Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. Data Min. Knowl. Discov. 2(2): 121-167.

Dockery , D. W., C. A. Pope , X. Xu , J. D. Spengler , J. H. Ware , M. E. Fay , B. G. J. Ferris and F. E. Speizer (1993). An Association between Air Pollution and Mortality in Six U.S. Cities. New England Journal of Medicine 329(24): 1753-1759.

Dong, X. and D.-X. Zhou (2008). Learning gradients by a gradient descent algorithm. Journal of Mathematical Analysis and Applications 341(2): 1018-1027.

EEA (2007). Air Pollution in Europe1990-2004. European Environment Agency Tech. rep., (EEA Report 2/2007).

Grivas, G. and A. Chaloulakou (2006). Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece. Atmospheric Environment 40(7): 1216-1229.

Hooyberghs, J., C. Mensink, G. Dumont, F. Fierens and O. Brasseur (2005). A neural network forecast for daily average PM10 concentrations in Belgium. Atmospheric Environment 39(18): 3279-3289.

Keary, J., S. G. Jennings, T. C. O'Connor, B. McManus and M. Lee (1998). PM10 Concentration Measurements in Dublin City. Environmental Monitoring and Assessment 52(1): 3-18.

Kurt, A., B. Gulbagci, F. Karaca and O. Alagha (2008). An online air pollution forecasting system using neural networks. Environment International 34(5): 592-598.

Lahmiri, S. (2011). A Comparative Study of Backpropagation algorithms in financial prediction, International Journal of Computer Science, Engineering and Applications (IJCSEA) 1(4).

Mitchell, T. M. (1997). Machine Learning, McGraw-Hill, Inc.

Morris, R. C. S. a. J. B. (2000). A Tutorial on Neural Networks Using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) Training Algorithm and Molecular Descriptors with Application to the Prediction of Dielectric Constants Through the Development of Quantitative Structure Property Relationships (QSPRs)

Perez, P., R. Palacios and A. Castillo (2004). Carbon monoxide concentration forecasting in Santiago, Chile. J Air Waste Manag Assoc 54(8): 908-913.

Perez, P. and J. Reyes (2006). An integrated neral network model for PM10 forecasting. Atmospheric Environment 40(16): 2845-2851.

Pope, C. A., 3rd and D. W. Dockery (2006). Health effects of fine particulate air pollution: lines that connect. J Air Waste Manag Assoc 56(6): 709-742.

Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986). Learning representations by back-propagating errors. Nature 323(6088): 533-536.

Smola, A. J. and B. Schölkopf (2004). A tutorial on support vector regression. Statistics and Computing 14(3): 199-222.

Taspinar F. and Bozkurt Z. (2014). Application of Artificial Neural Networks and Regression Models in The Prediction of Daily Maximum PM10 Concentration in Düzce, Turkey. Fresenius Environmental Bulletin 23(10): 2450-2459.

Vapnik, V. (1995). The Nature of Statistical Learning Theory, Springer.

Vapnik, V., S. G., and A. Smola (1997). Support vector method for function approximation, regression estimation, and signal processing. Cambridge, MIT Press.

Wang, Y., J. Li, X. Cheng, X. Lun, D. Sun and X. Wang (2014). Estimation of PM10 in the traffic-related atmosphere for three road types in Beijing and Guangzhou, China. J Environ Sci (China) 26(1): 197-204.