Modeling the parameters of the ambient air: comparison of SVM and NN regression models

Biljana Mileva Boshkoska

VEGA PRESS

Biljana Mileva Boshkoska

Modeling the parameters of the ambient air: comparison of SVM and NN regression models

Vega Press

Reviewers
Ass. Prof. dr. Nadra Damij
Prof. dr. Borut Ročenvić

ISBN: 978-0-9568625-9-4

Acknowledgments

This publication is funded by the Creative Core FISNM-3330-13-500033 'Simulations' project funded by the European Union, the European Regional Development Fund. The operation is carried out within the framework of the Operational Programme for Strengthening Regional Development Potentials for the period 2007–2013, Development Priority 1: Competitiveness and Research Excellence, Priority Guideline 1.1: Improving the Competitive Skills and Research Excellence. The author accknoledges the projects number J1-5454 and P1-0383.

Fakulteta za informacijske štud
Faculty of information s

Kreativno jedro:
Simulacije
Creative core: Simulations

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA IZOBRAŽEVANJE, ZNANOST IN ŠPORT

»Operacijo delno financira Evropska unija in sicer iz Evropskega sklada za regionalni razvoj. Operacija se izvaja v okviru Operativnega programa krepitve regionalnih razvojnih potencialov za obdobje 2007-2013, 1. razvojne prioritete: Konkurenčnost podjetij in raziskovalna odličnost, prednostne usmeritve 1.1: Izboljšanje konkurenčnih sposobnosti podjetij in raziskovalna odličnost.«

"The operation is partially financed by the European Union, mostly from the European Regional Development Fund. Operation is performed in the context of the Operational program for the strengthening regional development potentials for the period 2007-2013, 1st development priorities: Competitiveness of the companies and research excellence, priority aim 1.1: Improvement of the competitive capabilities of the companies and research excellence."

Abstract

This work presents the results obtained from twelve models for prediction of hourly concentrations of nitrogen dioxide and ozone, for one day and one week in August and December.

Two approaches have been used to construct the predictive models. The first approach uses Neural Networks with Radial Basis Functions and the second one employees Support Vector Machines (SVM). SVMs were built with two different kernel functions: a linear kernel and a Gaussian kernel. Three models for prediction of concentrations of ozone for August and three models for prediction of concentrations for December were built. Likewise, three models for prediction of concentrations of nitrogen dioxide for August and three models for prediction of concentrations of nitrogen dioxide for December were built. The modelling includes selection of the best values for the free parameters of the used kernel functions.

The work provides detailed description of three modelling techniques: regression as a tool for solving modelling problems, usage of neural networks with radial basis function and regression with support vector machines. The procedure for obtaining and processing the real data that are used as a training data set as well as the method of generating the models are comprehensively presented. Additionally the work presents discussion on the obtained results from the built models and advantages and disadvantages of each of the used techniques for modelling.

Content

List of Figures

List of tables

1 Introduction

The media reports may lead us to think that the problem of ambient air pollution emerged in the second half of the last century. However, the pollution of ambient air is not a new phenomenon in the history of mankind. Obviously even for the caveman the fire lightning had its consequences (1). There are historical data on destruction of plants as a result of the furnaces since the Roman Empire. However, the type of pollution of ambient air to which people have been exposed have changed throughout history, but the problem has been known for a long time, and caught the interest of the public especially in the 14th century when people first began to use coal for heating in their homes (2).

Until the industrial age, ambient air pollution has been considered as a local problem. Everything that had been done to reduce emissions was with purpose of protecting the health of people. In this respect, the appearance of chimneys was a blessing because that way pollutants were deluded and simply "disappeared" from the ambient air. The science in 1960s has shown that that the atmosphere is a thin layer around the globe and it is far from infinite landfill meaning that sooner or later, most of the pollutants will come back on Earth. This finding was partly prompted by the problem of increased acidification of lakes, which led to the emergence of a sharp decline in the yield of fishing in Scandinavia, and which is the result of long-term movement of pollutants in the air. At the end of the 1980s, the issue of air quality and environment become a world issue, and climate problems and the thinning of ozone layer were placed high on the agenda of politicians. The fact that emissions of mankind can have an effect on climate was known since the end of 19th century. Today, the problems due to air pollution can be found in several areas. Air pollution is a constant environmental problem which introduces significant

costs to the health of the society, the ecosystem (3), and the economy.

The most important and obvious problems due to air pollution are the direct effects on human health. The three pollutants that are recognized to most significantly affect human health are Particulate matter, nitrogen dioxide and ground-level ozone (4), (5). Recent research indicates that small particles ($PM_{2.5}$) in the air caused about 450,000 premature deaths within the 27 EU countries in the year 2005. Another 20,000 premature deaths was caused by ground-level ozone (5), while $PM_{2.5}$ were also responsible for around 100,000 serious hospital admissions in the EU25.

Next in line are the effects and damage to our environment such as the acidification of lakes, including the soil deforestation, eutrophication, ozone at ground level or crop damage.

Finally, the problems of air pollution overlap with other complex environmental issues such as congestion and mobility, landuse and global warming.

Today, science is concerned with modeling of parameters of the ambient air in order to investigate or to protect and improve the environment in which we live (6), (7), (8). Many researches deal with modeling the air quality in order to fill in the gaps of missing data, or to predict the air quality (9), (10), (11), (12), (13), (14), (15), (16), (17), (18). These models are usually based on historical data, they use complex algorithms and are usually understood by a small population of the relevant domain researchers. Here we present modelling of the concentrations of ambient air from the aspect of modelling non-technical systems as a part of the system engineering. The first attempt to develop the proposed models and the obtaind results can be found in (19).

1.1 Problem definition

EU countries are obliged to perform continuous monitoring of the ambient air throughout the whole territory of the country. However, mainly due to financial reasons, and technical problems in the maintenance of the monitoring stations, the data sets from the monitoring stations are not complete. According to the EU directives, the country must fulfil 90% of the measurements for the air quality on the measuring spots during one year. In order to fulfil the gaps in the data sets for air quality, we decided to use appropriate mathematical modelling technique, as a method that is allowed to be used by the EU directives.

Here we present the results obtained from filling in the existing gaps of the measured hourly data for the levels of ozone (O_3) and nitrogen dioxide (NO_2) in the ambient air for two short periods of time in the municipality of Karpos III, in Skopje, Republic of Macedonia. We process two data sets for August and December, 2005 and we build statistical models for hourly predictions of concentrations for one day and for one week. Solution of the problem had to be generated in a simple manner and the used algorithm had to be applicable for similar problems e.g. for prediction of concentrations of other air quality parameters. The predictions of parameters of the air quality for longer time periods is not of interest here, since the background of the problem is fulfilling the data gaps in the measured time series for up to 7 days ahead.

1.2 Methodology

To resolve the problem of prediction of the concentrations of ozone and nitrogen dioxide in the ambient air we regard two approaches. The first approach is through the use of neural networks (NN), while the second one employees Support Vector Machines (SVM).

In particular, we present three models, two SVM models that use linear and Gaussian kernel, and a Radial Basis Function (RBF) neural network model.

Hence three models are built for prediction of each of the parameters ozone and nitrogen dioxide and for each of the periods: daily and weekly predictions in August and December. From the three models for each parameter and time period, we choose the one that provides the best results for prediction of the concentrations.

The modelling follows these steps:

1. Selection of the period for which the prediction should be made
2. Searching for an array of data in which no data are missing or the missing data is minimal. The array of data should be at least one week before the period for which the prediction is performed.
3. The data are transferred to ARRF format.
4. The free parameters are determined for each of the used functions for modelling.
5. Three models using RBF NN and SVMs are generated, which are then used for prediction of concentrations of the ambient air. In each of the modelling procedures, a 10-cross validation is performed in order to select the best model.

Finally the model that performs the best fro each parameter and for each time period is chosen.

1.3 Current developments in the modelling of time series and measurement of the parameters of air quality

The environmental data are very complex for modelling primarily because many relationships that exist between the different variables lead to complicated combinations between

them. In order to simplify the complexity many attempts have been made to model the relations that exist between the data. For example, linear regression methods are applied for decades and they are widely and well known (20), (21), (22). Given that standard statistical techniques sometimes fail to adequately model the complex, nonlinear phenomena and chemical dependencies, new technologies are developed including the classification and regression trees (CART), artificial neural networks, fuzzy logic and SVMs.

NN have the ability to express and to model the data. They can be trained to successfully perform the functions approximation. They are highly adaptable to non-parametric distribution of data and, unlike other statistical methods that seek to meet an entire set of requirements, the NN is not required to do prior hypothesis about the relationship between variables. Other advantages of NN is their resistance to noise, their robustness and adaptability, especially when compared to expert systems. This is a result of the large number of inter-related processing elements (so called neurons) that can be trained to learn new patterns. These features give the NN potential to model complex nonlinear phenomena such as concentrations of ambient air parameters (23), (24), (25), (26).

NN are used in applications to perform short-term weather forecast since the early nineties (27) and for prediction of SO_2 in a polluted industrial area in Slovenia. Sillini T and others. Linear regression with NN was used to build models for forecast of PM_{10} in Thessaloniki, Greece (20). A. Pelliccioni and T. Tirabassi (28) use neural networks to build dispersion model for air as a new perspective for integrated models in simulation of complex situations. The models are used to forecast daily maximum ozone levels in different urban areas. The input data to the models are average values of meteorological data. Regression and multiple layer perceptron

were used to forecast the hourly concentrations of ozone and nitrogen dioxide in Bilbao (29).

NN were applied for forecasting daily maximum ozone concentrations and are compared to the results obtained with regression models (30). A comparison has been made between the models for prediction of concentrations of nitrogen oxides using multiple layer perceptron and other statistical methods. The results have shown that NN are superior compared to the other statistical models (31), (32). They have been used also for prediction of concentrations of $PM_{2,5}$ in Santiago, Chile, and PM_{10} in Helsinki (33).

There are several algorithms for training the NN, such as the back propagation algorithm, radial base function (RBF) etc. Their drawbacks are over-fitting, they get stuck in local minima, a difficult determination of the network architecture and poor generalization which remain unresolved and limit the applicability of NN. Such drawbacks of NN are resolved by introducing new learning algorithms, which lead to the development of SVMs. The models obtained by SVMs lead to a new effective approach for improving the performance of generalization and obtaining global solutions. SVM classificators always find the global maximum, which is a main feature of SVMs At the same time, the training of SVM depends on the following parameters: the penalty factor (C), the speed parameter (σ) and the parameter ε that defines the area of insensitivity (loss function) of the model. It also depends on the training set itself (i.e. its support vectors). SVMs were firstly used for identification (classification) of input-output pairs in a finite number of output classes. More recently, with the introduction of ε - loss function, SVMs are used to solve nonlinear regression problems and for prediction of time series (34), (35), (36), (37), (38), (39), (40).

Weizen LU (41) published a pioneering study using SVM (42) to assess the changes in the concentrations of six pollutants that

are measured in 1999 in Hong Kong every hour. Stephane Canu and Alain Rakotomamonjy (16) forecasted the pollution and the occurrence of peak concentrations of ozone in Lyon, France, and Giovanni Raimondo (43) predicted the PM_{10} levels using SVM.

2 Modelling tools

Many deterministic and statistical models are known which deal with prediction of the concentrations of ambient air parameters. They are all black box models that are based on the principal of the cause-and-effect: the output variable is a linear or nonlinear function of causal variables.

Deterministic models solve mathematical equations to predict air quality by using various data such as the emissions, or meteorological data. Overall, the results of the model are derived from inputs in a deterministic way. These models require accurate data on the emission and transport of substances in the ambient air. The lack of sufficient data is the most common cause of inaccuracy of the deterministic models.

Statistical models can establish relation between the input and output variables without dealing with the causes and consequences of the established relations between pollutants. They are based on statistical or semi-empirical techniques for data analysis for specific time periods, and for analysis of the relationship between air quality and atmospheric measurements and they can predict development of pollution episodes. Different techniques are used to achieve these goals; for example, the frequency distribution analysis, time analysis, Box-Jenkins, spectral analysis, etc.

Statistical models are very useful in situations such as, short forecasts in real time where the available information from the measured concentrations of trends are relevant (for purposes of current forecast). These models are used when there is some uncertainty about the physical and chemical source of data, or when scientific information is not complete. Instead, these models use empirical correlations between the data, for example the concentration of nitrogen oxides (NOx) and wind direction. These models do not attempt to predict the current concentration (which can fluctuate from moment to moment)

but a statistical value (usually the mean). The applications of these models are limited to locations for which they have been developed. A more sophisticated version of this type of model can be found in the approach that uses neural networks (44), (45), (46) that models the unknown variable as a cause and effect dependency.

Statistical models include linear and quadratic (or cubic) regression, and neuronal networks, and more recently the Support Vector machines (SVM) belong to this group (47), (48), (49), (50), (51).

When modelling by using empirical data, one uses the induction process to build a model of a system, which is expected to deductively determine the output of the observed system. Consequently, the quality and the quantity of observations carry the responsibility for the performance of the empirical model.

By its nature, the observed data are final. The obtained samples are not uniform and are scattered in the input space as a result of their natural multi dimensionality. The traditional approach of NN has difficulties in the process of generalization, producing models that overfit the data. It is a consequence of the optimization algorithms used for the selection of parameters and the statistical measures that are used to select the "best" model.

Support Vector Machines were developed by Vapnik, and they become very popular in the last decade as a result of many promising empirical performances. Their formulation includes the principle of Structural Risk minimization (SRM), which proved to be superior to the traditional principle of empirical risk minimization (ERM) used in NN. SRM performs minimization of the upper bound of the expected risk, as opposed to ERM that performs minimization of the error data for training. Exactly this difference gives greater opportunity

for SVM generalization, which is the goal of statistical learning. SVM is initially developed for solving the problems of classification, but more recently have expanded over the domain of regression problems. SVM term usually used to describe classification with Support Vector Machines and Support Vector Regression is used to describe regression using the methods of support vectors.

2.1 Regression as a modelling tool

Regression is the task of predicting the numerical value of the output variable as a function of the input variable

$$\hat{f}(x) = \hat{y} \qquad (2.1)$$

where the kappa above f and y marks that the function is estimation of the real function (52).

Predictors are commonly applied to samples for which the value of the output variable is not known. However, to assess how much the estimated value of the sample is close to the truth one, it is necessary to determine a set of known output values which will be used for training. To assess the effectiveness of the predictor, a loss function is applied over the predictions made on the set of test samples. The number of the data in the test set is Mtest. A commonly used loss function in regression is the mean square error (MSE).

$$MSE = \sum_{i=1}^{Mtest} \frac{\left(f(x_i) - \hat{f}(x_i)\right)^2}{M_{test}} \qquad (2.2)$$

2.1.1 Networks with Radial Bases Function (RBF)

2.1.2 Radial Basis Functions

The design of supervised neural networks (NN) can be performed in various ways. The back propagation algorithm for

the design of multi-layer perceptron can be understood as an application of an optimization method known in statistics as stochastic approximation. Here we set out with a different approach considering the design of the neuronal network as a problem of curve approximation in multi-dimensional space. From this aspect, the learning task is equivalent to finding a multidimensional space that provides the best curve of the training data, by applying the criterion of "best qualified" measured in a statistical sense. Accordingly, generalization is equivalent to the use of this multidimensional surface interpolation of the test data. Such an approach is the motivation behind the method of RBF in the sense that research entails the traditional interpolation in multidimensional space. When using neuronal networks, the hidden neuron provides a set of "features" that form an arbitrary "basis" for input data when they are expanded in the space of hidden levels. These functions are called radial bases functions.

The construction of the RBF neural network, in its most basic form, includes three completely different levels: an input, a hidden and an output level. The input level consists from genuine nodes (sensory units). The second level is high dimensional hidden level. The output level provides the response of the activation samples applied to the input level of the network. The transformation from the input space to the space of hidden units is nonlinear, while the transformation from the hidden level to the output level is linear. The mathematical explanation of this principle is given by Cover (1965), which states that the problem of classification of samples in nonlinear dimensional space is more likely to be linearly separable than in less dimensional space - hence the reason for making high dimensional space in the hidden level of RBF networks. With careful design, it is possible to reduce the dimension of the hidden level, especially if the centres of hidden units are adaptive.

2.1.3 RBF Architecture

With proper pre-processing, the input-output training data enables the single-layer perceptron to approximate any constrained function. The same role has the hidden layer in neural networks. NN allows us to solve a large class of problems, however they are defined as "black box" models. What we look for is a more comprehensive mapping and at the same time a universality of the solution.

The main idea in RBF is to force each unit of the hidden layer to represent a particular area of the input space. In other words each unit of the hidden layer must contain a prototype of a group of data in the input space. When the input appears in the hidden layer, a new combined entity with most similar group is activated and thus triggered a particular path through the network. The principle of activating the most similar group causes the introduction of the concept of measuring the distance between the centres of the groups and the presented new entry to the network.

RBF functions represent quite successful solution of this problem. They contain hidden layer whose elements pre-process the input data. The pre-processed input data continues to the single layer perceptron (Figure 2-1). Each element k in the hidden layer contains prototype x_k of a particular region of the input space. The corresponding nonlinear activation function Φ_k represents the similarity between the current input x and the prototype x_k through some measure of distance.

With total of H elements in the hidden layer and the p-dimensional output space, the input sample x is transformed from the input space to the i-th dimension of the output space according to the mapping (2.3). The output is given with (2.4) where the weights w_{ik} represent the links between the hidden element k and the output element i, while w_{i0} is a threshold of the output element i. Usually architectures with more than one

hidden layer are not considered. The main reasons are twofold. Firstly, the usage of more hidden layers leads to losing the "transparency" of processing of the network, while the gain of such architecture is linear. Secondly the design of the neuronal network structure becomes more complex.

The elements of the output layer usually have linear or sigmoidal activation functions.

$$\Phi_k(x) = \Phi_k(\|x - x_k\|) \quad k = 1, \dots, H \qquad (2.3)$$

$$o_i = \sum_{k=1}^{H} w_{ik}\Phi_k(x) + w_{i0} \quad i = 1, \dots, p \qquad (2.4)$$

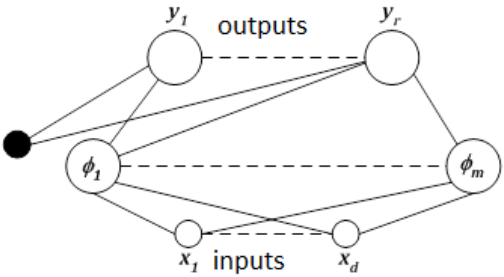

Figure 2-1 Architecture of RBF network

The most frequently used RBF is the Gaussian normalized function (2.5). The center μ_k

$$\Phi_k(x) = e^{-\frac{|x-\mu_k|^2}{2\sigma_k^2}} \qquad (2.5)$$

defines the prototype of the input group k, and the variance σ_k^2 defines its size.

2.2 Support Vector Machines

This section describes the central idea of learning with Support Vector Machines and provides the basic concepts of SVMs.

2.2.1 Data representation and similarity

One of the basic problems of the theory of learning is following: given two classes of objects, one has to assign a new object into one of the two classes. To formalize the problem, one has to start with the given empirical data:

$$(x_1, y_1), \dots, (x_m, y_m) \epsilon \mathcal{X} \times \{\pm 1\}. \qquad (2.6)$$

where $\mathcal{X}$ is not an empty set of samples, from which the samples x_i are drown (known under the names of cases, inputs, observations or instances). It is commonly called domain. The values of y_i are called labels, objectives, outputs or observations. Usualy there are two classes of samples, which are marked with -1 and +1. This situation is very simple, and it is called (binary) recognition samples or (binary) classification (53).

To study the problem of learning, we need a different kind of structure. In learning, we want to be able to generalize the unknown data. It means that for a given new sample $x \epsilon \mathcal{X}$, we want to predict appropriate $y \epsilon \{\pm 1\}$. It means that we choose y so that (x, y) is in some way similar to the training samples. For that reason, we need a concept of similarity in $\mathcal{X}$ and in $\{\pm 1\}$.

Describing the similarity of outputs $\{\pm 1\}$ is simple in the binary classification problem where only two situations occur: two marks may be identical or different. The choice of a measure of similarity in the input set, on the other hand, is a question whose answer lies in the core of machine learning.

Let us consider a measure of similarity that is given in the following form:

$$k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$$
$$(x, x') \mapsto k(x, x') \tag{2.7}$$

In other words, one searches for a function which given two samples x and x' returns a real number which characterizes their similarity. We assume that k is a symmetric function, unless stated otherwise, i.e. $k(x, x') = k(x', x)\ for\ all\ x, x' \epsilon\ \mathcal{X}$. The function k is called a kernel.

To be able to use simple functions, as a similarity measure, such as scalar product, one has to present the input samples as vectors in a space with scalar product H (which does not necessarily coincide with $\mathbb{R}^N$). For this purpose, we use this mapping

$$\Phi: \mathcal{X} \to H$$
$$x \mapsto \vec{x} := \Phi(x). \tag{2.8}$$

Next, we need to consider a general measure of similarity that we obtain by applying the mapping (2.8). In this case, Φ will be a nonlinear mapping.

The space H is called feature space. The incorporation of data into H through Φ leads to three benefits:

1. It allows us to define a measure of similarity of scalar product calculated in H

$$k(x, x') := \langle \vec{x}, \vec{x}' \rangle = \langle \Phi(\vec{x}), \Phi(\vec{x}') \rangle \tag{2.9}$$

2. It allows us to geometricaly deal with the samples, and consequently to study the learning algorithms by using linear algebra and analytic geometry
3. The freedom of choice of the function Φ allows us to design a number of different similarity measures and

algorithms for learning. This also applies to cases where the inputs $\vec{x}_i$ exist in the space of scalar product. In that case, we could directly use the Scalar product as a measure of similarity. However, nothing prevents us from first apply the non-linear mapping Φ to change the data representation to a new one that is more convenient for the given problem.

2.2.2 Support Vector Regression

Unlike the problem of classification with SVM, where the desired output values y_i are discrete, the regression outputs are real values (54). The general problem of SVM regression can be defined as follows: given l training data, the machine is trying to learn an input output relation $f(x)$.

The training set $D = \{[x(i), y(i)] \in \Re^n \times \Re, i = 1, \dots, l\}$ consists of l pairs $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, where the inputs x are n-dimensional vectors $x \in \Re^n$, and the outputs $y \in \Re$ are continuous values. SVM approximates functions from type (52), (53), (55), (56):

$$f(x, w) = \sum_{i=1}^{N} w_i \varphi_i(x) \qquad (2.10)$$

One may notice that the most general model fully complies with RBF models, and to some degree to the fuzzy logical models. It can also be noted that the free parameter b is not explicitly shown. The function $f(x, w)$ is explicitly written as a function of the weights w_i that should be learned.

Support vector machines are based on an algorithm that finds a special kind of linear model: the maximum margin hyper plane. The maximum margin hyper plane is the one that gives the greatest separation between the classes. The instances that are closest to the maximum margin hyper plane and the ones with minimum distance to it are called support vectors. There is

always at least one support vector for each class, and often there are more.

The concept of a maximum margin hyper plane only applies to classification. However, support vector machine algorithms have been developed for numeric prediction that share many of the properties encountered in the classification case: they produce a model that can usually be expressed in terms of a few support vectors and can be applied to non-linear problems using kernel functions.

Similar with linear regression, the basic idea here is to find a function that approximates the training points well by minimizing the prediction error. The crucial difference is that all deviations up to a user-specified parameter ε are simply discarded. Also, when minimizing the error, the risk of over fitting is reduced by simultaneously trying to maximize the flatness of the function. Another difference is that what is minimized is normally the predictions' absolute error instead of the squared error used in linear regression. A user-specified parameter ε defines a tube around the regression function in which errors are ignored.

SVM approximate the learning data set with a function given in a form of:

$$f(x, w) = \sum_{i=1}^{N} w_i \varphi_i(x) + b \qquad (2.11)$$

meaning that the original data $x \rightarrow \phi(x)$ are mapped into high dimensional space and then construct an optimal hyper plane in this space. $\phi(x)$ represents feature of the inputs, while w_i and b are coefficients. These are estimated by minimizing the risk function :

$$R(f) = \int c(x, y, f(x)) dp(x, y) \qquad (2.12)$$

where $c(x, y, f(x))$ is cost function that determines how to penalize estimation errors based on the empirical data X. Given that we do not know the probability measure p(x,y) we can only use X for estimating a function f that minimizes R[f]. A possible approximation consists in replacing the integration by the empirical estimate to get so called empirical risk function

$$R[f] = \frac{1}{n}\sum_{i=1}^{n} c(x_i, y_i, f(x_i)) \qquad (2.13)$$

A first attempt would be to find the function $f_0 = argmin_{f \in H} R_{emp}[H]$ for some hypothesis class H. However if H is very rich, i.e. its capacity is very high as for instance when dealing with few data in very high dimensional spaces, this may be not such a good idea as it will lead to overfitting and thus bad generalization properties. Hence one should add a capacity control term, which in the SV case results to be $\|w\|^2$, which leads to regularized risk function

$$R_{reg} = R_{emp}[f] + \frac{\lambda}{2}\|w\|^2 \qquad (2.14)$$

There are a number of parameters that has to be learned and that can be used in the construction of SVM regression. The most relevant are the insensitivity zone ε and the penalty parameter C. Both parameters should be selected by the user. The increment of ε parameter contributes to the "smoothness" of the results. At the same time this increment leads to a dimensionality reduction of the model because less support vectors are defined.

2.2.2.1 Kernels

A kernel is essentially a similarity function with certain mathematical properties, which is possible to define over all sorts of structures such as sets, strings, trees, and probability distributions.

The choice of kernel $K(x_i, x_j)$ influences drastically on the performance of the SVMs depending on the considered problem. Several kernels are available for learning sand they have to satisfy the so-called Mercer's condition (42).

The most commonly used kernels are the Gaussian kernel

$$K(x_i, x_j) = \exp -(\frac{\|x_i - x_j\|}{2\sigma^2}) \qquad (2.15)$$

and the polynomial kernel

$$K(x_i, x_j) = (x_i, x_j + 1)^p \qquad (2.16)$$

which are also used for the purposes of this research.

The kernel functions introduce various parameters that define the function. For example, the polynomial kernel is defined by the parameter d, and the Gaussian kernel is defined by the variance matrix Σ whose elements determine the shape of the kernel function. Another important parameter is the mean μ, which determines the position of the Gaussian kernel. During the design of SVM, the average values are selected by placing the kernels in the data pairs. The choice of parameters d and Σ is experimental. Vapnik proposes to do the following: train the SV machine for different values of the parameters d and Σ, calculate the VC dimension and select the model with lowest VC dimension. Table 2-1 summarizes the steps for training SVM.

Table 2-1 Steps in designing SVM for solving the problem of regression

Step 1	Select the kernel function for regression.
Step 2	Chose the form of the kernel function (degree of polynomial functions or variance of the Gaussian kernel function).

Step 3	Choose the penalty factor C and select the desired accuracy by defining the ε zone.
Step 4	Solve the quadratic problem with $2l$ variables for regression.

3 Modelling of the parameters of the ambient air by using state-of-the-art methods for regression

In this chapter we model the parameters of the ambient air by applying Support Vector Machines and Radial Basis Functions. Both algorithms are implemented in many well-known software packages such as MATLAB, WEKA and Orange.

Both SVMs and RBF are implemented in MATLAB, however the price of this software package is often an obstacle for it usage.

Orange is a software platform for data mining. It implements many pre-processing techniques, methods for predictive and data modelling. It is based on C++ components, which one may use directly, through Python scripts, or through the graphical interface of the software package. Orange allows usage of several built in components or installing new components built by the user.

For the modelling process we chose Weka software package because it is very simple to use, it requires no great prior programming experience (although it allows to more experienced programmers to implement their own code), it is a free software platform, and it implements the algorithms that are of interest here.

The modelling process starts with data pre-processing in order to put them into the ARRF format, required by the software package Weka. Then one has to choose the algorithms that will be used for modelling, and to determine the parameters of the corresponding algorithm. Finally, the model is built using Weka. The whole process of modelling is explained in details below.

3.1 The process of modelling

The modelling of ambient air is a process that consists of six steps which are are graphically presented in Figure 3-1. The modelling steps are following:

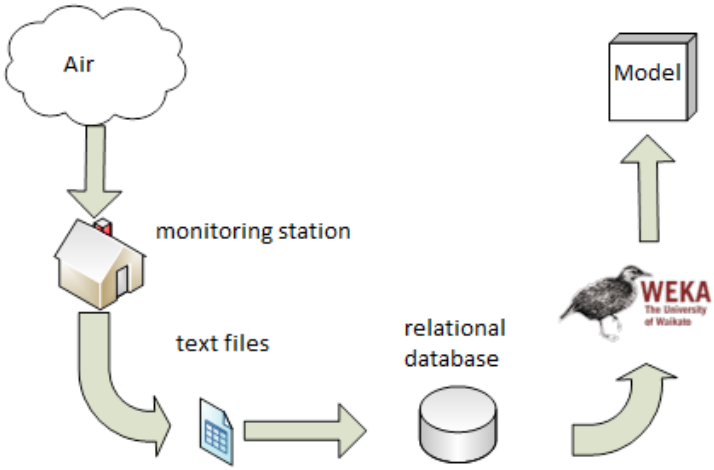

Figure 3-1 The six-phases of the process for ambient air modelling

- measurement of the concentrations of the parameters of the ambient air,
- transfer of the measured data from the monitoring stations to the central database located in the Ministry of Environment and Physical Planning (MEPP)
- data processing and preparation of ARFF files
- selecting tools (software) for modelling
- usage of the software package WEKA,
- comparison of the obtained models of the software package WEKA and selecting the best model for prediction of the concentrations.

3.2 Measurement of the concentrations of the parameters of the ambient air

The frequency of measurement and types of instruments used for collecting the data in the automatic air monitoring stations located in the municipality of Karpos III in Skopje are given in Table 3-1. The monitoring station is connected through a telephone line with the control centre in MEPP. The measured data are transferred from the monitoring station to the central database in MEPP on a daily basis.

The automatic monitoring station measures the concentrations of the following parameters: SO_2, CO, NO, NO_2, NOx, O_3, PM_{10} and the following meteorological parameters: temperature, pressure, humidity, wind direction, wind speed, and global radiation. The monitoring station is given in Figure 3-1.

Figure 3-2 Automatic monitoring station for ambient air located in Skopje, in the municipality of Karpos III

Table 3-1 Frequency and type of instrument for data collection in the measuring station Karpos III

Parameters	Frequency of measurements	Used Instrument Thermo ESM Andersen
Sulfur Dioxide	Measurements are taken each second, however instruments display the data average each 10 seconds. The database in MEPP keeps records of the mean hourly values.	Model 43 C
Nitrogen Oxides		Model 42 C
PM10		FH 62 I-R
Carbon Monoxide		Model 48 C
Ozone		Model 49 C

3.3 The measured dataset

The dataset wes collected by the national network for air monitoring by the Ministry of Environment and Physical Planning (MEPP).

The data were collected through the national grid, and were stored in a database located in the MEPP. They include values of zero's indicating periods for which there are no measurements of no valid registration data. This was the main difficulty during the modelling. Therefore, we decided to choose a small period of time for which we have sufficient data and which includes a minimum number of records that have zero value. Data are saved each hour and are obtained as the average of 60 minutes of data (data are measured each minute). The hourly data that are used for modelling are measured by the automatic monitoring stations in Skopje, in the municipality of Karpos III, for the period 1 - 17 August 2005. The set of training data are those from 1 - 10 August, 2005), which in total leads to 1680 hourly data. They are used to build the models

for prediction of the concentrations of two parameters: NO_2 and O_3.

The prediction of concentrations of NO_2 and O_3 is for two time periods: for the eleventh day of the month, and for one week ahead. The input parametrs for the built models are: SO_2, NO_2, NOx, CO and O_3, temperature and humidity.

Next, three different models are built for each of the two parameters separately. Two of the models are based on SVM, while the third is based on the RBF NN.

3.4 Description of the WEKA software package for modelling of time series

Experience has shown that there is no unique pattern of machine learning that can be applied to all of the data mining problems. In reality, the data sets vary, so to get a precise model, the learning algorithm must match the nature of the problem under consideration.

WEKA is a collection of the latest machine learning algorithms and tools for data processing. It is designed in a way that allows the user to quickly experiment implemented new methods of data sets in a flexible way. It includes a broad support for the entire process of experimental data mining, including preparation of input data, statistical evaluation of schemes for learning and visualization of input sets and learning outcomes. The different learning and data processing tools can be accessed through a common graphical interface that allows the user to compare different methods and identification of the most appropriate machine learning method for the problem at hand.

Weka is developed at Waikato University in New Zealand, and the name is an abbreviation of the English words Waikato Environment for Knowledge Analysis. The package is written in JAVA and is distributed under the terms laid down in the

General Public License. It offers implementation of algorithms for learning that can be easily applied to a given data set. It includes a set of tools to transform the data set, such as an algorithm for data discretization.

It includes the standard methods for the data mining problems: regression, classification, clustering, data mining with association and attributes selection. Understanding the data is an integral part of the work and for that reason Weka provides a tool for data visualization. Algorithms in WEKA work with text files in AFRR format that include a single relational data table (57).

There are several ways of how to use the package. One way is to apply some of the methods for learning the data set, and then to analyse the output of the implemented order to learn something more about the data. Another way is to use already learned models on the real dataset for prediction of new samples. A third way is to apply several different algorithms and compare their performance in order to select the best performing algorithm. Learning methods are called classifiers, and they are accessible through the WEKA graphical interface. Many classifiers have parameters that can be adjusted, and by accessing them through the Properties menu or through the Objects editor. There is a single module for evaluation which is used to measure the performance of all classifiers.

The most valuable resource in Weka is the implementation of current state-of-the-art machine learning methods for pattern recognising. Next to them are filters for data pre-processing.

The simplest way to use Weka is through a graphical interface called Explorer. It provides easy access to data that feeds software, which can be downloaded from the SQL database using JDBC or simply be placed in ARFF, CSV, C4.5, or binary documents. Explorer interface leads the user through their menus to all options that exist in the implemented

algorithms, and through forms that simply need to be filled in by the user. The default starting values for each of the implemented algorithms allow the user with minimal effort to get results - but you need to consider while the user uses the algorithm to fail to understand the results.

There are two other graphical interfaces in Weka. Knowledge Flow interface design allows configurations for targeted data processing. Basic lack of Explorer interface is that it is stored in main memory. The whole data set is is loaded in the main memory, when opened. It means that it can only be used on small to medium-scale projects. Weka contains in itself some incremental algorithms that can be used for processing very large data sets. Knowledge Flow interface allows placement of objects in the workspace representing the learning algorithm and applying it to the desired configuration. It provides the opportunity of specifying the flow of data by connecting components that represent data sources, then pre-processing tools, learning algorithms, methods and evaluation modules for visualization.

The third interface is called Experimenter and it is designed to provide answer to the basic and practical questions such as which method and which value of the parameter best suited for a given problem. Usually there is no way to answer that question apriori, which is the main motivation for the development of this interface. It essentially allows comparisons between different learning algorithms. The same can be achieved with the Explorer, but the Experimenter allows automating the process by using filters, classifiers with different sets of parameters, and provides statistics for performance. Advanced users can use the Experimenter for distributed computing and processing on multiple machines using the JAVA RMI (remote method invocation). This way you can make a large number of statistical experiments on multiple machines.

The basic functionality of Weka is shown through tits interfaces. However, it can be accessed by using text commands, which gives access to all features of the system. When Weka is run, the user can choose with which of the four interfaces s/he will work: Explorer, Experimenter, Knowledge Flow or command line interface. For data processing we used the two interfaces Explorer and Experimenter.

An important resource is the online Weka documentation which is automatically generated from the source code and briefly reflects its structure. The next level of using Weka includes access to a library of user Java programs, and writing and testing learning patterns.

In most applications of data mining, the machine learning component is only a small part of a much larger software system. So if the user wants to program an application, s/he will also need to access the Weka programs from his source code. Hence, the user can solve his/hers problems of machine learning with minimal additional programming.

Here we use the Java programming to access the Weka algorithms.

3.5 Data formats

3.5.1 Text format

The data obtained from the monitoring stations are stored in text files. Each monitoring station generates one file every 24 hours for each parameter separately. A file contains hourly measurements for one measurement unit. Text files has the following structure:

Parameter name	Format

Time of measurement	dd.mm.yyyy hh:mi:ss
Value	#,###.00
Item number	

The text data are parsed and are afterwards included in the database.

3.5.2 ARFF format

WEKA recognise input files in ARFF format. ARFF files have two main parts. The first part is a header, followed by the Data section.

The header of the ARFF file contains the name of the relation, a list of all attributes and their types. An example of an ARFF file is following:

```
% 1. Title: Iris Plants Database
%
% 2. Sources:
%   (a) Creator: R.A. Fisher
%   (b) Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)
%   (c) Date: July, 1988
%
@RELATION iris

@ATTRIBUTE sepallength NUMERIC
@ATTRIBUTE sepalwidth  NUMERIC
```

@ATTRIBUTE petallength NUMERIC

@ATTRIBUTE petalwidth NUMERIC

@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}

The data part of the ARFF file has the following structure:

@DATA

5.1,3.5,1.4,0.2,Iris-setosa

4.9,3.0,1.4,0.2,Iris-setosa

4.7,3.2,1.3,0.2,Iris-setosa

4.6,3.1,1.5,0.2,Iris-setosa

5.0,3.6,1.4,0.2,Iris-setosa

5.4,3.9,1.7,0.4,Iris-setosa

4.6,3.4,1.4,0.3,Iris-setosa

5.0,3.4,1.5,0.2,Iris-setosa

4.4,2.9,1.4,0.2,Iris-setosa

4.9,3.1,1.5,0.1,Iris-setosa

Lines that srart with the symbol % are comments. The spetial words @RELATION, @ATTRIBUTE and @DATA must be written with capital letters.

The header of the ARFF file has the following elements:

@RELATION <name of the relation name of the relation >

where <name of the relation> is a string. If the name contains spaces, then it must be placed in quotes.

Attributes are given in a sequence. Each attribute from the data set are indicated in ARFF file with the element @ATTRIBUTE. The format of the element is:

@ATTRIBUTE <Name of the attribute> <type>

where < Name of the attribute > must begin with a letter. The following data types are supported in Weka:

Numeric

String

Date [<date format>]

The data file starts with the keyword @DATA after which all the attributes values are given separated by commas. The order of the values must match the order of the defined attributes in the file header. The missing values are given with the symbol "?".

3.5.3 Data processing

The data processing passes through several phases. The first phase consists of parsing the original data obtained from monitoring stations and their proper storage in a relational database (Figure 3-3).

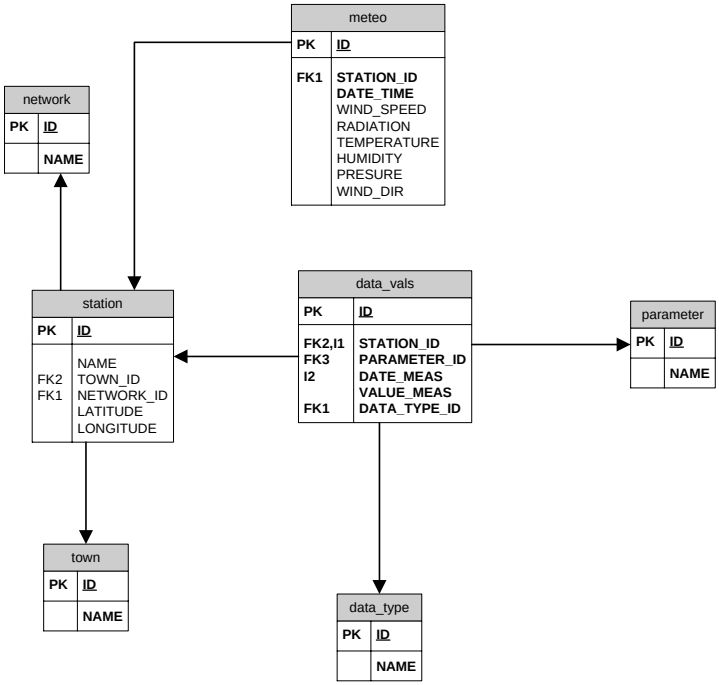

Figure 3-3 Relational database

Data processing means the process of training SVM and then the process of testing the accuracy. The first part, is the process of adjustment is a complex and long-term and hence the distributed processing is used. The overall learning model is divided among more computers that are controlled by a central "master" computer. For this procedure the Experimenter tool in Weka is used. In this way the overall time required to tune the SVM is reduced. The principal scheme of this connection is shown in Figure 3-4.

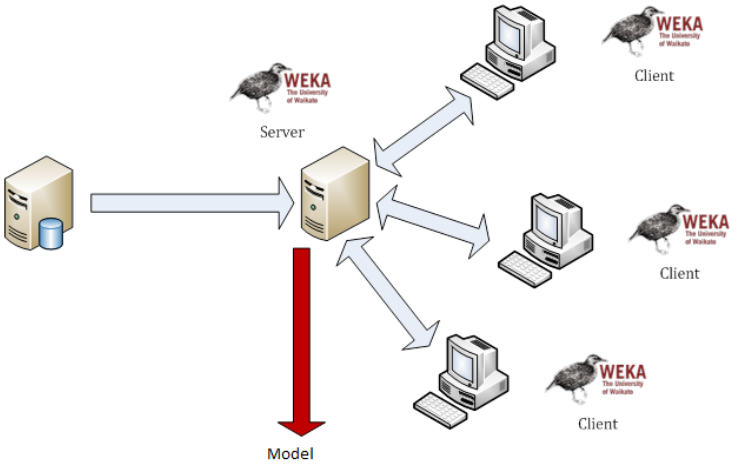

Figure 3-4 Distributive data processing

3.6 Models for prediction of ozone and nitrogen dioxide

3.6.1 Models for predicting the values of NO_2

The prediction of the concentrations of NO_2 are based on three different models: SVMreg with linear kernel, SVMreg with Gaussian kernel and a RBF NN. These are compared in order to select the one with best prediting performances.

The three functions are used both for building models for prediction of concentrations for the eleventh day, and the prediction of concentrations over a period of one week leading to two sets of results. In both cases, for each of the three models, four input parameters are used: SO_2, NOx, CO, and NO, while the output of the model predicted concentrations of NO_2:

$$NO_2(t) = f(NO_x(t-1), SO_2(t-1), NO_2(t-1), NO(t-1), CO(t-1)) \quad (3.1)$$

Figure 3-5 (Figure 3-6) shows the distribution of original data for NO_2 for the one day (for one week) in August. The same figures present the predicted concentrations of nitrogen dioxide obtained using the three models: SVMreg with linear kernel, SVMreg with Gaussian kernel and RBF NN. Here, the entry to the models are five different concentrations of ambient air parameters, while meteorological data are not taken into account. The results show that the best predictions are obtained with SVM model based on linear kernel.

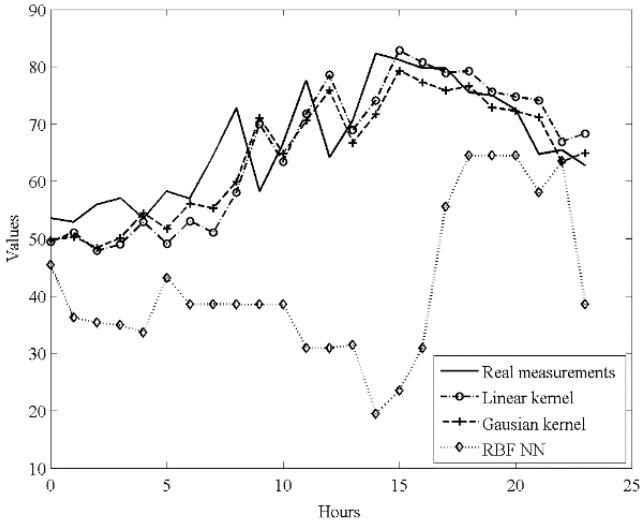

Figure 3-5 Predictions of the concentrations of NO_2 for 24 hours in August, 2005

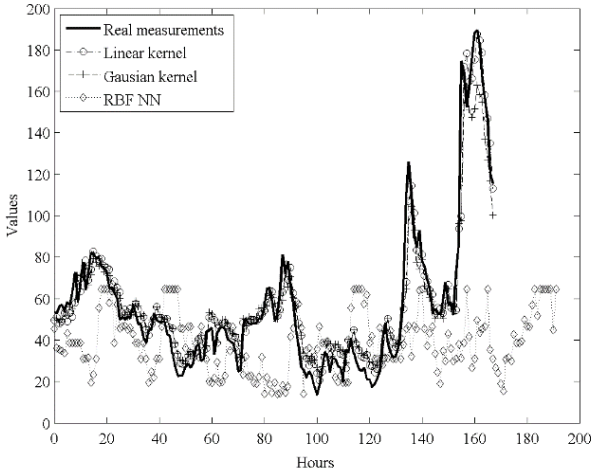

Figure 3-6 Predictions of the concentrations of NO_2 for 7 days in August, 2005

The models are compared based on the value of mean absolute error (MAE), that is given with:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\alpha_i - p_i| \qquad (3.2)$$

where $\boldsymbol{\alpha_i}$ is the predicted value and $\mathbf{p_i}$ is the measured value.

MAE values for both sets of models are given in Table 3-4 and Table 3-5.

Table 3-2 Statistical data for SO_2, NO_x, NO_2, NO and CO for the location Karpos III 1 – 17.8.2005

	$\ln \frac{\mu g}{m^3}$	SO_2	NO_x	NO_2	NO	CO
	Min	1.26	3.50	8.42	1.7	0.46
	Max	26.22	199.40	108.83	78.36	2.34
	Mean	9.91	45.02	36.41	9.74	0.86
	Std dev	3.46	31.09	18.55	10.28	0.26
Percentile	50	9.00	38.47	32.68	6.24	0.80
	75	11.29	58.54	46.3	12.19	0.94
	95	14.51	84.83	63.84	20.13	1.17

Table 3-3 Statistical data for SO_2, NO_x, NO_2, NO and CO for the location Karpos III 1 – 17.12.2005

	$\ln \frac{\mu g}{m^3}$	SO_2	NO_x	NO_2	NO	CO
	Min	9,4	12,95	13,27	0,96	0,39
	Max	365,98	734,36	189,63	395,10	6,07
	Mean	41,06	142,52	56,69	60,45	1,85
	Std dev	47,83	102,67	27,99	56,08	1,06
Percentile	50	24,78	112,16	51,76	42,62	1,54
	75	36,66	169,29	61,04	69,55	2,22
	95	80,17	283,29	82,69	133,39	3,34

Table 3-4 Prediction of concentrations of NO_2 one day ahead in August, 2005 (total of 24 instances)

	SVM with polynomial kernel	RBF	SVM with Gaussian kernel $\sigma = 0.5$
Correlation coefficient	0.8282	0.6723	0.5932
Mean Absolute Error	**8.9335**	13.6988	27.6238
Root Mean Squared Error	13.6003	19.807	38.1648

Table 3-5 Prediction of concentrations of NO_2 one week ahead in August, 2005 (total of 192 instances)

	SVM with polynomial kernel	RBF	SVM with Gaussian kernel $\sigma = 0.5$
Correlation coefficient	0,7981	0,6527	0,5696
Mean Absolute Error	**8,9883**	11,7024	15,877
Root Mean Squared Error	12,5184	15,2343	23,4619

These models can be improved if we determine the optimal values of the free parameters of SVM models. The Gaussian kernel function has three free parameters: σ, C and ε. Given that there are no general rules for determining the values of these free parameters, one has to determine the resulting error of the

model based on the chosen values. Here, we use the mean absolute error (MAE), to evaluate the deviation between the original data and the predicted data. In the general case, we consider that the lower the value of MAE, the better results are achieved with the model.

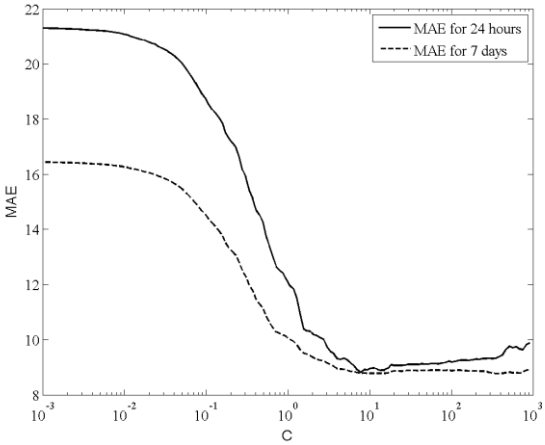

Figure 3-7 MAE variations for NO2 depending on the values of parameter C for 24h and for 7 days, August, 2005

Figure 3-7 shows the variations of MAE for the parameter NO_2, depending on the values of the parameter C. The Figure shows that the parameter C has very little impact on the error and it is sensitive to C only at very small values of the parameter C, for example, If $C \leq 0.001$. By increasing the value of C, the value of the MAE reduces steep and slightly changed values of $C \geq 0.5$. In the general case, to ensure a stable process of learning, the value of the parameter C should get very large values, such as $C = 100$, as is the case here.

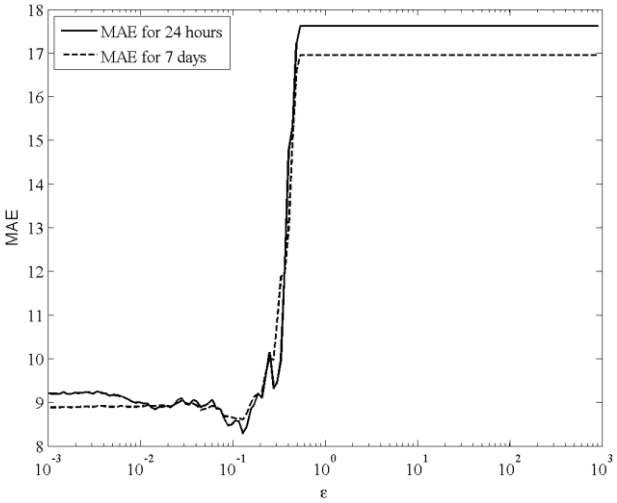

Figure 3-8 MAE variations for NO_2 depending on the values of parameter $\boldsymbol{\varepsilon}$ for 24h and for 7 days, August, 2005

Figure 3-8 shows the variations in the MAE for the parameter NO_2, depending on the values of the parameter ε. Parameter ε has also very little impact on the performance of the model for prediction of concentrations of NO_2. MAE values are almost constant for values of the parameter $\varepsilon < 10^{-2}$ и $\varepsilon > 0.5$. In models that use SVM small values of ε are recommended. In this simulation, the value of is ε = 0.1.

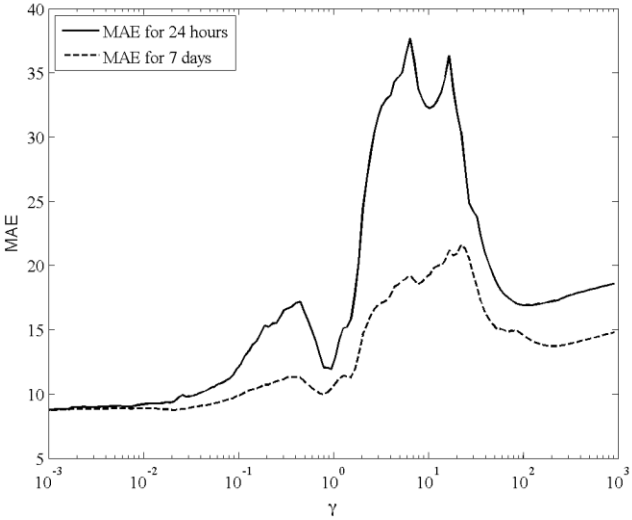

Figure 3-9 MAE variations for NO2 depending on the values of parameter γ for 24h and for 7 days, August, 2005

Figure 3-9 shows variations in the MAE for the parameter NO_2, depending on the values of the parameter γ, which is connected with the parameter of speed σ through the relation $\gamma = \sqrt{\left(\frac{1}{\sigma^2}\right)}$. In theory, the speed parameter σ affects the prediction performances. Very small ($\sigma \to 0$) or very large values ($\sigma \to \infty$) for σ can lead to poor predictions. If $\sigma \to 0$, all training data become support vectors. In this case, when an unknown data input occurs to the SVM model, the model will not be able to offer good computer guidance and achieve good performance predictions. On the other hand, if $\sigma \to \infty$, all training data will be considered as one point. Hence it follows that the SVM model can produce the same results for any new data. Therefore, these two extreme cases should be avoided. It

should be noted that $\sigma \to \infty$ and $\sigma \to 0$ represent approximate two processes. In a real application, if $\sigma \ll \|x_i - x_j\|$ and $\sigma \gg \|x_i - x_j\|$ the extreme cases mentioned above will occur.

From the results shown in Figure 3-9 can be seen that MAE is large (e.g. around 24.7), when σ is small (e.g. $\sigma = 0.001$), then decreased with increasing σ and reaches a minimum (about 10.1) for value of $\gamma = 0.01, \sigma = 0.7$. From Figure 3-9 one can notice that MAE fluctuates in the range of [0.1, 100], when σ is in the range [0.9, 1.1], gradually increases with σ, and finally tends to remain constant for $\sigma \geq 30$. Hence it follows that in practical applications, only the parameter σ of the Gaussian kernel function should be determined during the simulations, while the two parameters C and ε can be set in advance by experience. In this application, the value of σ is set to $\sigma = 0.5$ using the rule of trial and error.

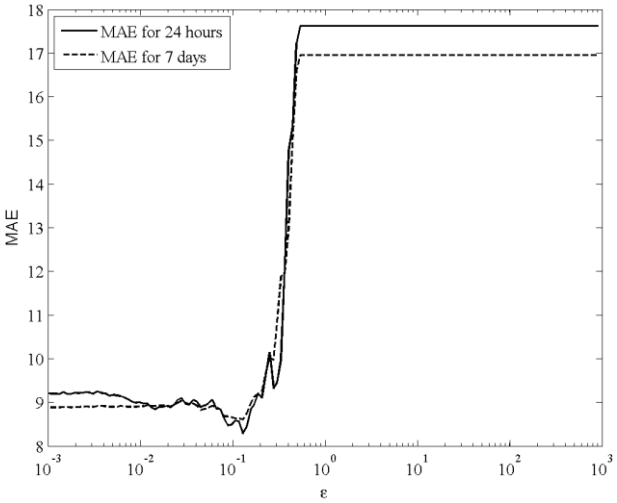

Figure 3-8 MAE variations for NO2 depending on the values of parameter $\boldsymbol{\varepsilon}$ for 24h and for 7 days, August, 2005

Figure 3-7, Figure 3-8 and Figure 3-9 show that the value of the MAE is smaller when the predicting NO_2 concentrations for a period of one week compared with the value of MAE when predicting the concentrations of NO_2 over a period of 24 hours. These values derive from the data distribution, which have very close values to each other as shown in Table 3-2.To show that this is indeed the case, we chose another month of the year, December, 2005, in which the data range is quite higher as can be seen from Table 3-3.

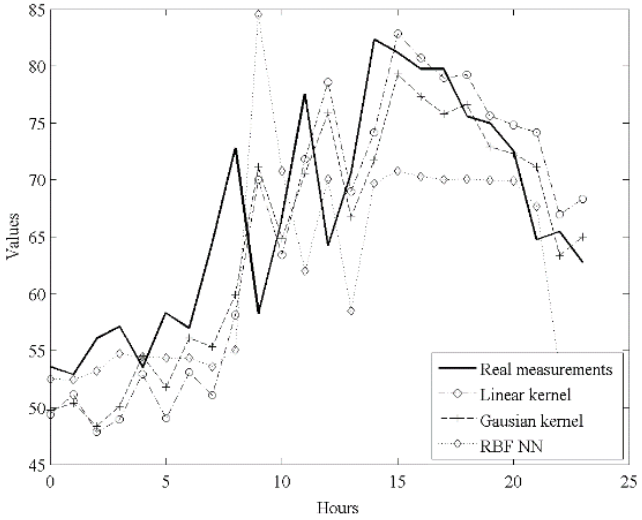

Figure 3-10 Predictions of the concentrations of NO_2 for 24 hours in December, 2005

Modeling results for December 2005 are shown in Figure 3-10 and Figure 3-11. They shows the distribution of original data for **NO_2** for one day and one week, respectively. The same figures presents the distribution of the predicted data obtained with three models built with the SVMreg with linear kernel, SVMreg with Gaussian kernel and the RBF NN. In this case, the free parameters have the following values: C =100, γ=0.01 and ε=0.1.

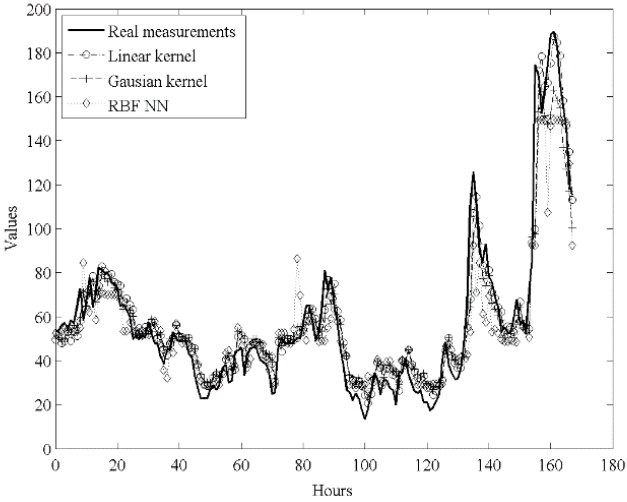

Figure 3-11 Predictions of the concentrations of NO_2 for 7 days in December, 2005

Table 3-6 and Table 3-7 show that the value of the MAE is smaller when models are built with SVM compared to the model built with RBF neural network. Here once again it follows that models built with SVM provide better results than models built with neural networks.

Table 3-6 rediction of concentrations of NO_2 for 24 hours in December, 2005 (total number of instances 24)

	SVM with polynomial kernel	RBF	SVM with Gaussian kernel $\sigma = 0.5$
Correlation coefficient	0.8021	0.5352	0.8106
Mean Absolute Error	5.6376	7.7928	**5.0545**
Root Mean Squared Error	7.2229	9.9271	6.3965

Table 3-7 Prediction of concentrations of NO_2 for 24 hours in December, 2005 (total number of instances 192)

	SVM with polynomial kernel	RBF	SVM with Gaussian kernel $\sigma = 0.5$
Correlation coefficient	0.9622	0.9261	0.9631
Mean Absolute Error	**6.8722**	9.8944	8.1039
Root Mean Squared Error	10.1383	15.2272	11.6752

3.6.2 Models for predicting the values of O_3

For modelling purposes in this case we also used the software package WEKA. To perform regression we used three

functions: SVMreg with linear kernel, SVMreg with Gaussian kernel and RBF neural network. The three functions are used for prediction of concentrations of the eleventh day, and the prediction of concentrations over a week. Thus, three models are available for prediction of concentrations during the eleventh day, and three models for prediction of concentrations over a week. The models in each group are then compared and conclusions are drown regarding their performances. Four parameters are used as inputs to the modelling: NO_2, O_3, temperature and humidity, while the output of the model are the concentrations of O_3. The output function of the model is given by:

$$O_3(t) = f\big(NO_2(t-z), O_3(t-z), NO_2(t), temperature(t-z), humidity(t-z)\big) \quad (3.3)$$

Table 3-8 Statistical data for NO_2, O_3, temperature and humidity for the monitoring station Karpos III for 1 – 17 December, 2005

		O_3	NO_2	Temperature	Humidity
Min		2,06	13,27	-1,3	53,77
Max		55,14	156,16	13,52	98,51
Mean		12,84	52,27	4,08	87,41
Std. dev		11,79	19,74	3,37	10,46
Percentile	50	8,74	49,59	3,14	91,95
	70	12,71	58,79	5,24	95,24
	90	28,57	77,08	9,25	97,22

Table 3-9 Statistical data for NO_2, O_3, temperature and humidity for the monitoring station Karpos III for 1 – 17 August, 2005

		O_3	NO_2	Temperature	Humidity
Min		3,3	8,42	11,68	24
Max		137,72	108,83	37,81	97,04
Mean		50,60	38,15	22,23	64,53
Std. dev		37,80	18,70	5,27	19,25
Percentile	21,28	34,15	21,28	65,65	91,95
	68,38	44,88	25,18	77,99	95,24
	29,45	66,19	29,45	90,71	97,22

Furthermore, we built eight different models for prediction of concentrations of O_3 for $t - z$ hours, where z = 1,2, ..., 8. The results from the prediction of the concentration of O_3 for different values of z = 1,2, ..., 8 are given below. Statistical data for the parameters NO_2, O_3, temperature and humidity for the monitoring station Karpos between 1 – 17 December, 2005 are given in Table 3-8, while for August 2005 are given in Table 3-9.

In continuation we provide the results obtained for z = 3 for two characteristic cases. In the firs case the training data consist of 9 consecutuve days from 1-10.8.20105 ommiting the missing data from 8.8.2005. The second case consists of consecutuve days from 1-11.8.20105 ommiting the missing data from 8.8.2005. The goal was to determine the influence of including more training data to the performances of the obtained models.

3.6.2.1 Model for prediction of $O_3(t)$ when z=3

Figure 3-12 shows the **distribution of the original data for ozone, for 11 August, 2005.**

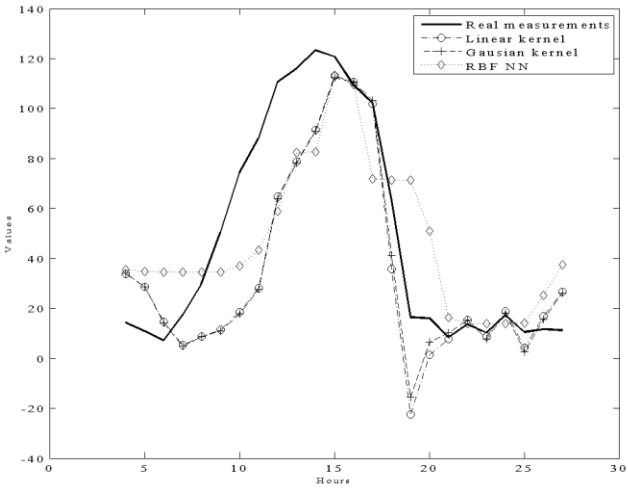

Figure 3-12 Predictions of the concentrations of O_3 for 24 hours in August, 2005 for z=3; training data does not include the missing data of 08.08.2005

The same figure presents the distributions of the predicted data with the SVM model based on linear kernel, the SVM model based on RBF kernel (also known as SVM with Gaussian kernel), and with the model based on RBF neural network. The input data are NO_2 (t-3), O_3 (t-3), NO_2 (t), temperature (t-3), humidity (t-3).

In this case, the training set does not consider the data for August 8, 2005, when data for ozone concentrations are missing and are recorded as zero in the database. They are not considered because they do not represent the real values of ozone concentration on that day.

Here we built one more model in which the missing data are removed and are replaced with data from the consecutive day in the observed period, i.e. the 11 August.

Table 3-10 Errors for the two SVM models built with linear kernel

SVM with linear kernel	Model with 9 training days	Model with 10 training days
Correlation coefficient	0.8561	0.8283
Mean Absolute Error (MAE)	**19.1424**	**17.5576**
Average Square Error	26.0908	24.7707
Relative Absolute Error	45.2593 %	44.3782 %
Relative Average Square Error	58.2281 %	57.6919 %
Total number of instances	24	24

From Table 3-10 one can see that the error of the model is smaller in case when the training set consists of data from 10 days in comparison to the model in which a training set consists of data from 9 days. In both cases the missing data from the eight day are discarded from the training set. Next models use the first training set as it has smaller error.

Figure 3-13 shows the predicted concentration of O_3 obtained with the model in which the data from August 8, 2005 are discarded from the training set, and instead data from August 11 are included into the data set.

Figure 3-14 presents the **distribution of the measured data for ozone for a period of one week** in August, 2005, taking into account data on 08/08/2005. The same figure presents the predicted distribution of O_3 data for one week obtained with models with SVM with linear kernel, with Gaussian kernel and

RBF neural network. Figure 3-15 shows the **distribution of the measured data for ozone for a period of one week** in August 2005, not taking into account data on 08/08/2005.

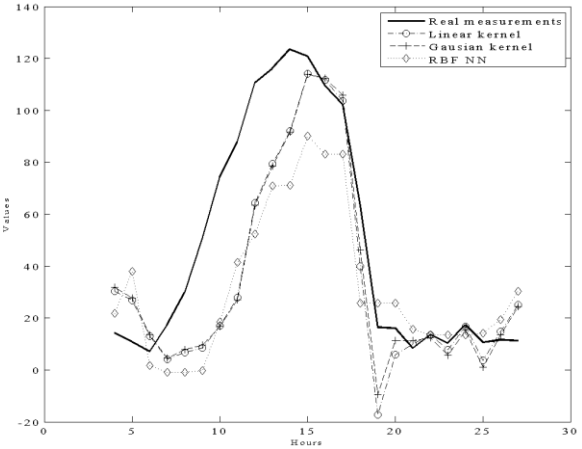

Figure 3-13 Predictions of the concentrations of O_3 for 24 hours in August, 2005 for z=3; training data does not include the missing data of 08.08.2005, but includes the data from 11.08.2005

If you closely examine at the graphs presented in Figure 3-12 and Figure 3-13, one can notice a "shift" of the three graphs representing the prediction of ozone. However, this shift is only a visual effect. If we consider the graphs in detail, you will notice that the "shift" is present only during the day, and that it starts in the morning, it is highest in the middle of the day, and then decreases and disappears completely during the night. This occurrence is understandable if you take into account that the prediction is based on heuristically chosen inputs to the model. The formula for ozone prediction, does not take into

account the insolation which helps in the formation of ozone in nature, and which is present only during the day.

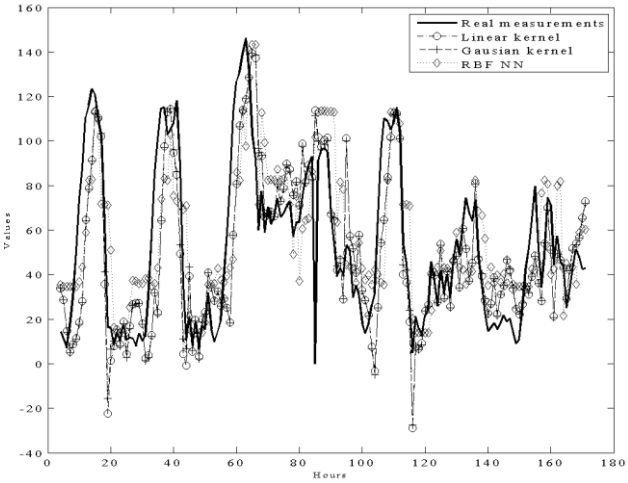

Figure 3-14 Predictions of the concentrations of O_3 for 7 days in August, 2005 for z=3; training data does not include the missing data of 08.08.2005

Insolation data for the formula for ozone certainly would improve results during daylight hours. Such a shift is observed also in other graphs given below. The shift is smallest when models are obtained for prediction of ozone in the next hour (t+1), which is understandable given that the difference in hours between the insolation at time t and t + 1 is the smallest compared to the difference in insolation between t and t + n, here n = 2, .., 8.

Figure 3-14 and Figure 3-15 show the results obtained using the three models for prediction of ozone levels in real time for one week. Similarly to the case of modelling of concentrations of

NO_2, here as well the results shows that the best predictions are obtained if one uses the SVM model with linear kernel, compared with the results obtained with SVM model based on Gaussian kernel or the model based on RBF neural network.

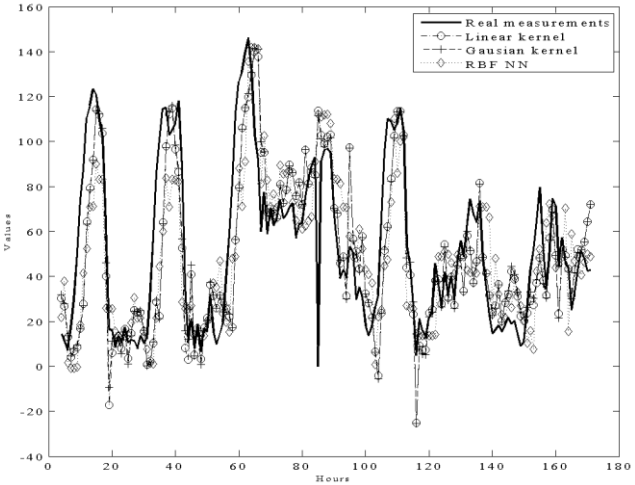

Figure 3-15 Predictions of the concentrations of O_3 for 7 days in August, 2005 for z=3; training data does not include the missing data of 08.08.2005, but includes the data from 11.08.2005

Similarly to the analysis of data for prediction of the concentration of NO2, here we continue with creating a model to predict the concentration of O3 in December 2005. The data analysis results in the following graphs.

Figure 3-16 (Figure 3-17) shows the **distribution of original data for ozone for 11 December (for one week)**. The same

figure presents the distribution of the predicted data for the eleventh day (one week) in the month with SVM model with linear kernel, SVM model with Gaussian kernel, and a model based on RBF neural network. The input to the prediction model are the following data NO_2 (t-3), O_3 (t-3), NO_2 (t), temperature (t-3), humidity (t-3).

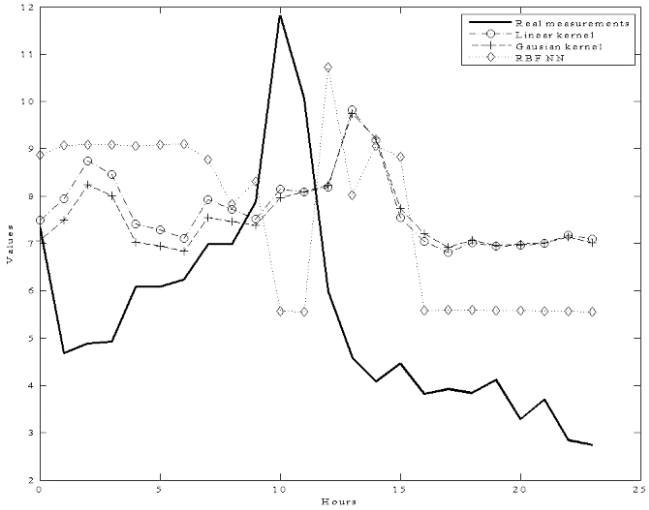

Figure 3-16 Predictions of the concentrations of O_3 for 24 hours in December, 2005 for z=3

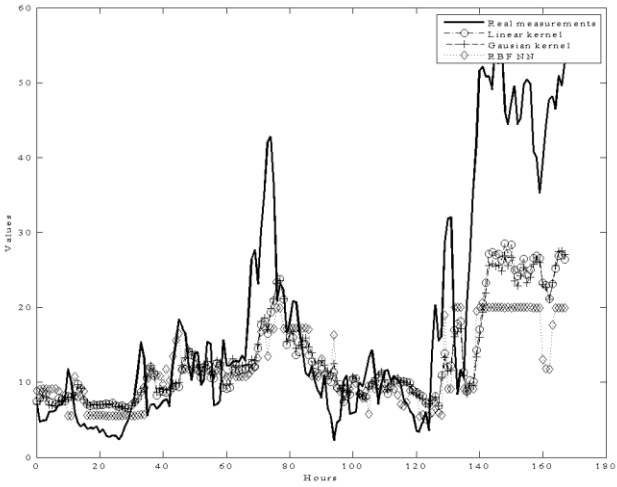

Figure 3-17 Predictions of the concentrations of O_3 for 7 days in December, 2005 for z=3

3.7 Discussion on the obtained models

The measured data on the concentrations of parameters of ambient air in August 2005 are very similar, meaning that the standard deviation is very small. Consequently, it appears that the three models give very similar results for the MAE for this case. All three models give good results for prediction of concentrations. This is not the case when the input data to the model have large standard deviation, such as the case for December 2005. Namely, in case of prediction of ozone concentrations for one week in December, 2005, the features of SVM become prominent. The better generalization property of SVM leads to a model with best results. This is the one marked as LINEAR for z = 1,2,3, or GAUSS for z = 4,5,6,7,8.

We calculated the optimal values for C, ε and γ when z = 3. The same calculated values for C, ε and γ are used as well for cases when $z \neq 3$. One may argue that better results would be obtained if the same procedure of finding the optimal values of C, ε and γ is repeated. Still the results show that even in this cases the value of MAE is smaller for the models obtained with SVM then for the one obtained with NN.

From the Table 3-11 it is clear that the best results are obtained with the model built with SVM model with linear kernel for when z = 1,3,4,6,7,8, while for z = 2 and 5 the best results are obtained with the SVM model built with Gaussian kernel. The best achieved vales for each case are given with bold letters in the Table 3-11 - Table 3-14.

Table 3-11 MAE values for August, 2005, for prediction of ozone concentrations for a period of 7 days (CC – correlation coefficient, MAE - Mean absolute error, MSE - Mean square error, RAE - Relative absolute error, RMSE - Relative mean square error, TNS - Total number of samples)

Z	Error type	RBF NN	GAUSS	LINEAR
z=1	CC	0,8734	0,9362	0,9345
	MAE	12,9727	8,2517	**8,2396**
	MSE	18,1319	12,8925	13,0079
	RAE	42,4004	26,9703	26,9306
	RMSE	50,1584	35,6647	35,9841
	TNS	168	168	168
z=2	CC	0,7765	0,8651	0,863
	MAE	17,8784	**12,6182**	12,8682
	MSE	23,4842	18,4953	19,0927
	RAE	58,7833	41,4881	42,3101

Z	Error type	RBF NN	GAUSS	LINEAR
	RMSE	65,192	51,3428	53,0013
	TNS	168	168	168
z=3	CC	0,6843	0,7802	0,7806
	MAE	21,1517	16,498	**16,3756**
	MSE	27,3691	23,6819	23,6024
	RAE	69,8923	54,515	54,1106
	RMSE	76,1652	65,9041	65,6828
	TNS	168	168	168
z=4	CC	0,5848	0,7047	0,7031
	MAE	23,5529	19,5862	**19,5387**
	MSE	31,4468	27,2885	27,3443
	RAE	78,0977	64,9448	64,7871
	RMSE	87,7237	76,1236	76,2793
	TNS	168	168	168
z=5	CC	0,4794	0,6366	0,6423
	MAE	25,7211	**21,5782**	21,7519
	MSE	33,8384	29,491	29,6215
	RAE	85,4468	71,6839	72,2611
	RMSE	94,518	82,3747	82,7394
	TNS	168	168	168
z=6	CC	0,41	0,5797	0,5943
	MAE	26,6023	22,9743	**22,8666**
	MSE	35,4407	31,3077	31,1326
	RAE	88,1973	76,169	75,8118
	RMSE	98,9066	87,3724	86,8838

Z	Error type	RBF NN	GAUSS	LINEAR
	TNS	168	168	168
z=7	CC	0,3405	0,5599	0,5881
	MAE	26,9367	23,4642	**23,0146**
	MSE	36,0148	31,6561	31,0131
	RAE	89,047	77,5675	76,0814
	RMSE	100,3971	88,2466	86,454
	TNS	168	168	168
z=8	CC	0,4937	0,5792	0,6112
	MAE	24,9491	22,5738	**22,2367**
	MSE	32,1188	31,0132	30,3662
	RAE	82,7025	74,8288	73,7113
	RMSE	89,5527	86,4699	84,6661
	TNS	168	168	168

From Table 3-12 one can tell that in case of prediction of Ozone, for August 2005, the best results are obtained with the SVM model based on linear kernel when z = 3,4,6 and 7, while for z = 1,2,5 and 8 the best results are obtained with the SVM model based on Gaussian kernel.

Table 3-12 MAE values for August, 2005, for prediction of ozone concentrations for a period of 24 hours (CC – correlation coefficient, MAE - Mean absolute error, MSE - Mean square error, RAE - Relative absolute error, RMSE - Relative mean square error, TNS - Total number of samples)

Z	Error type	RBF NN	GAUSS	LINEAR
z=1	CC	0,9518	0,9789	0,9765
	MAE	10,9027	**7,0284**	7,9482

Z	Error type	RBF NN	GAUSS	LINEAR
	MSE	14,2878	9,3278	9,9895
	RAE	27,0327	16,8248	19,0267
	RMSE	32,952	21,0591	22,553
	TNS	24	24	24
z=2	CC	0,8889	0,928	0,9226
	MAE	18,6888	**13,7295**	14,1851
	MSE	23,393	18,3266	18,9994
	RAE	46,0738	32,6571	33,7408
	RMSE	53,6947	41,1313	42,6415
	TNS	24	24	24
z=3	CC	0,8268	0,853	0,8562
	MAE	23,9551	19,5844	**19,1323**
	MSE	30,4058	26,5492	26,1009
	RAE	58,7447	46,3045	45,2356
	RMSE	69,481	59,2512	58,2506
	TNS	24	24	24
z=4	CC	0,7278	0,8119	0,8174
	MAE	24,8629	23,6534	**23,463**
	MSE	32,9661	31,1631	31,2701
	RAE	60,6093	55,5592	55,112
	RMSE	74,9181	69,0949	69,3321
	TNS	24	24	24
z=5	CC	0,6117	0,7909	0,8024
	MAE	30,4538	**26,3758**	27,0605
	MSE	38,4114	34,6681	35,1665

Z	Error type	RBF NN	GAUSS	LINEAR
	RAE	74,4816	62,0568	63,6679
	RMSE	87,5213	76,9959	78,1026
	TNS	24	24	24
z=6	CC	0,4871	0,7803	0,7844
	MAE	32,6497	28,4986	**28,4034**
	MSE	42,4704	36,6613	36,675
	RAE	80,1415	67,1948	66,9703
	RMSE	97,0765	81,6167	81,6473
	TNS	24	24	24
z=7	CC	0,5539	0,7643	0,766
	MAE	31,4257	29,0422	**28,8844**
	MSE	40,2295	36,6948	36,6215
	RAE	76,7887	68,1129	67,7429
	RMSE	91,6035	81,2874	81,1251
	TNS	24	24	24
z=8	CC	0,6254	0,7935	0,7862
	MAE	30,8265	**26,5339**	27,1136
	MSE	36,296	33,3506	34,059
	RAE	75,4477	62,2959	63,6568
	RMSE	82,703	73,9376	75,5081
	TNS	24	24	24

From Table 3-13 one can tell that in case of prediction of Ozone concentrations for one week in December, 2005, the best results are obtained with the SVM model based on linear kernel

when z = 1, 2 and 3, while for z = 4, 5, 6, 7 and 8 the best results are obtained with the SVM model based on Gaussian kernel.

Table 3-13 MAE values for December, 2005, for prediction of ozone concentrations for a period of 7 days (CC – correlation coefficient, MAE - Mean absolute error, MSE - Mean square error, RAE - Relative absolute error, RMSE - Relative mean square error, TNS - Total number of samples)

Z	Error type	RBF NN	GAUSS	LINEAR
z=1	CC	0,8736	0,9716	0,9725
	MAE	12,3772	3,3741	**3,2628**
	MSE	17,7991	4,7442	4,6729
	RAE	39,2617	30,9098	29,8906
	RMSE	48,6145	27,7981	27,3805
	TNS	168	168	168
z=2	CC	0,7948	0,9182	0,9212
	MAE	16,6615	6,4568	**6,3807**
	MSE	22,3034	9,5358	9,463
	RAE	53,1132	57,9121	57,2299
	RMSE	61,128	54,8762	54,4571
	TNS	168	168	168
z=3	CC	0,6775	0,8711	0,8846
	MAE	21,6752	7,7936	**7,7398**
	MSE	27,2322	11,6612	11,5303
	RAE	69,3673	68,258	67,7872
	RMSE	74,8335	65,7475	65,0094
	TNS	168	168	168
z=4	CC	0,6719	0,7318	0,7468

Z	Error type	RBF NN	GAUSS	LINEAR
	MAE	20,855	**9,7972**	9,9421
	MSE	27,301	15,157	15,3812
	RAE	66,9218	83,7787	85,0177
	RMSE	75,2071	83,7466	84,9854
	TNS	168	168	168
z=5	CC	0,5334	0,304	0,1191
	MAE	24,8994	**11,2747**	11,5455
	MSE	30,8784	17,5158	18,0215
	RAE	79,9267	94,4117	96,6793
	RMSE	85,1315	95,0862	97,8316
	TNS	168	168	168
z=6	CC	0,5626	0,1553	-0,0292
	MAE	23,9133	**11,8059**	11,91
	MSE	30,6839	18,3713	18,5911
	RAE	76,504	96,9222	97,777
	RMSE	84,5749	98,1496	99,3239
	TNS	168	168	168
z=7	CC	0,5664	-0,2931	-0,337
	MAE	23,4215	**12,2526**	12,2679
	MSE	30,2468	19,0937	19,1659
	RAE	74,8007	99,2503	99,3743
	RMSE	83,2788	101,1343	101,5168
	TNS	168	168	168
z=8	CC	0,5466	-0,513	-0,587
	MAE	24,6931	**12,6391**	12,8648

Z	Error type	RBF NN	GAUSS	LINEAR
	MSE	31,681	19,8269	20,3094
	RAE	79,1157	101,5451	103,3581
	RMSE	87,322	104,6486	107,1955
	TNS	168	168	168

From Table 3-14 one can tell that in case of prediction of Ozone concentrations for one day in December, 2005, the best results are obtained with the SVM model based on linear kernel when z = 1, 2, 5, 6, 7 and 8, while for z = 3 and 4 the best results are obtained with the SVM model based on Gaussian kernel. It is important to stress once again that the parameters of SVM models are optimal only in case when for z = 3.

Table 3-14 MAE values for December, 2005, for prediction of ozone concentrations for a period of 24 hours (CC – correlation coefficient, MAE - Mean absolute error, MSE - Mean square error, RAE - Relative absolute error, RMSE - Relative mean square error, TNS - Total number of samples)

Z	Error type	RBF NN	GAUSS	LINEAR
z=1	CC	0,9474	0,8024	0,8043
	MAE	10,4434	1,2373	**1,1356**
	MSE	15,1336	1,5032	1,4107
	RAE	24,9999	29,9496	27,488
	RMSE	34,1669	33,8522	31,7696
	TNS	24	24	24
z=2	CC	0,9093	0,3692	0,4075
	MAE	15,8553	2,1481	**1,9521**
	MSE	20,8383	2,5138	2,3224

Z	Error type	RBF NN	GAUSS	LINEAR
	RAE	37,7136	51,2505	46,5752
	RMSE	46,7684	55,6115	51,3776
	TNS	24	24	24
z=3	CC	0,7973	0,1573	0,2287
	MAE	21,3973	**2,6484**	2,718
	MSE	26,9651	3,0411	3,0777
	RAE	50,5907	61,285	62,8958
	RMSE	60,1795	65,5998	66,39
	TNS	24	24	24
z=4	CC	0,7975	0,0842	0,1106
	MAE	20,3196	**2,9814**	3,0967
	MSE	26,6218	3,3044	3,3814
	RAE	47,7284	68,349	70,993
	RMSE	59,0259	70,9719	72,6242
	TNS	24	24	24
z=5	CC	0,6074	0,0818	0,1092
	MAE	30,6804	3,7794	**3,7719**
	MSE	36,907	4,0306	4,0123
	RAE	72,1848	87,5373	87,3638
	RMSE	81,9683	87,2369	86,8407
	TNS	24	24	24
z=6	CC	0,7048	0,3598	0,3235
	MAE	26,6718	3,7617	**3,7232**
	MSE	32,9151	4,0075	3,974
	RAE	62,8875	86,9399	86,0501

Z	Error type	RBF NN	GAUSS	LINEAR
	RMSE	73,2769	86,2481	85,5265
	TNS	24	24	24
z=7	CC	0,7785	0,5874	0,6172
	MAE	24,1607	3,8093	**3,7448**
	MSE	31,6594	4,0478	3,9765
	RAE	56,6643	88,0761	86,5851
	RMSE	70,1328	87,0483	85,515
	TNS	24	24	24
z=8	CC	0,9041	0,7186	0,699
	MAE	18,2458	3,5401	**3,3321**
	MSE	22,7583	3,7455	3,5325
	RAE	42,8373	80,7178	75,9763
	RMSE	50,4547	79,7447	75,2099
	TNS	24	24	24

The results clearly show the advantage of SVM models in comparison to the models built with neural networks. Although the parameters of SVM models are not optimally adjusted in cases when $z \neq 3$, still, in almost all cases, they give MAE values that are smaller than the MAE values obtained models built with neural networks. This is primarily due to the fact that SVM has uses the Structural Risk Minimization Principle which leads to a better generalization than the conventional technique. Also, the typical problem of overfitting and falling into local minimum is eliminated in SVM, and it is the main lack of neuronal networks. It should be also mentioned that the SVM method has a smaller number of free parameters than neuronal networks. Here we presented the parameters of σ, C and ε, which have to be set when the Gaussian kernel is used.

When using NN, the size of the network, the learning parameters and the training of the network play a major role in the performance of the built model for prediction.

Alternatively, the proposed SVM model can be used for prediction of the concentration of other parameters in the air, not only for those that are considered here. More importantly, the results show that SVM model gives excellent results even in an absence of meteorological data as in the case of prediction of concentrations of NO_2.

4 Conclusion and further research

This book presents the theoretical background and the results of the modelling of the concentrations of ozone and nitrogen dioxide in ambient air using neural networks and Support Vector Machines. Different models are built for prediction of the concentrations of two parameters: ozone and nitrogen dioxide, for two seasons: August and December. For each season several models are built which are compared and the best one is chosen. The free parameters of the models are set so that the lowest median absolute error is obtained, which is an indicator of the accuracy of the performed prediction.

The obtained models and the simulations confirmed that SVM and neuronal networks can accurately model the relationship that exists between the various parameters of the ambient air and the meteorological parameters in an urban environment. The models can follow the trend of air emissions without external guidance.

The data that are used for training and testing of the models were firstly pre-processed so that the appropriate text formats are obtained. The pre-processed data are saved into a relational data base that can be used in future to examine different algorithms such as adaptive RBF networks, decision trees etc. Then by using the software tool WEKA different models were built.

Apart from the software package WEKA, the programming language JAVA was used for development of a program for automated testing of the obtained models. Automatic testing accelerates the modelling process.

In total 108 tests on the built models were performed. All models that were compared with each other, were trained and tested on the same training and testing set.

Models for prediction of concentrations of nitrogen dioxide show that better results are obtained by using SVM methods in comparison to the results obtained with the RBF NN.

Models for prediction of ozone concentrations are examined for eight consecutive hours. The best results are achieved with SVM models for predicting the values in t + 1 hour ahead. In three cases the SVM model built with polynomial kernel provided the best results, in particular for the prediction of ozone concentrations for 24 hours ahead in December and August. One model based on SVM with Gaussian kernel provided best results when predicting the concentrations for a week ahead in December. It should be noted that in all cases the SVM models performed better than RBF NN models.

Finally, one may conclude that SVM models have the advantage of time series prediction over RBF NN. The following free parameters were determined in the SVMs: the parameters C, ε and σ and it was shown that that only the parameter σ has a significant impact on the results of the proposed models. As a result of using the principle of structural risk minimization, the models built with SVM provide better prediction results than RBF models. Finally, using the SVM outperforms the disadvantages of neuronal networks to overfit the data or to fall into a local minimum.

Although it is not possible to use exactly the same models at the other measuring points, the presented methodology is general and can be used to build new models for other measuring points, which will be trained using local data from the monitoring stations.

Models for prediction of ozone concentrations can be further extended. The developed model for prediction of ozone uses data for NO_2, O_3, temperature and humidity. It could be expanded with data for NOx, emissions data from transport and other predictors of ozone formation, such as reactive gases,

aldehydes, and additional meteorological data for various micro regions. Similarly, we can extend the model for prediction of concentrations of nitrogen dioxide, mainly with meteorological data. The models may include data of the terrain where the prediction are made. Additionaly models may be enhanced with the time dependence or chemical dependency between parameters, which will lead to the development of new prediction models.

5 Bibliography

1. **P.Elvingson, C.Agren.** *Air and the environment.* Goteborg : Elanders Infologistics AB. Molnlycke, 2004. ISBN 91-973691-7-9.

2. **K.B.Schnelle, C.A.Brown.** *Air Pollution Control – Handbook.* s.l. : CRC Press, 1997.

3. **EEA Technical report, No 11/2014.** http://www.eea.europa.eu/publications/effects-of-air-pollution-on. *Effects of air pollution on European ecosystems , Past and future exposure of European freshwater and terrestrial habitats to acidifying and eutrophying air pollutants ,.* [Online] 12 08 2014.

4. **http://www.eea.europa.eu/themes/air/intro. [Online] 12 08 2014.**

5. http://www.airclim.org/. [Online] 12 08 2014.

6. N.P.Cheremisinoff. *Handbook of Air Pollution Prevention and Control.* s.l. : Butterworth Heinemann, 2006.

7. R.Boubel, D.L.Fox, B.D.Turner, A.C.Stern. *Fundamentals of Air Pollution.* s.l. : Academic Press, 1994.

8. B.D.Turner. ***Atmospheric dispersion estimates.*** **s.l. : Lewis Publishers, 1994.**

9. ***Variable Selection using Support Vector Machines - based Criteria.*** **A.Rakotomamonjy. 1-2, 2003, Neurocomputing, Vol. 55.**

10. ***Traffic pollution modelling and emission data.*** **R.Berkowicz, M.Winther, M.Ketzel. 4, s.l. : Elsevier, 2006, Environmental Modelling & Software, Vol. 21, pp. 454-460.**

11. ***Air quality integrated modelling in Turin urban area.*** **G.Calori, M.Clemente,R.De Maria,S.Finardi, F.Lollobrigida, G.Tinarelli. s.l. : Elsevier, 2005.**

12. M.M.El-Halwagi. ***Pollution Prevention through Process Integration.*** **s.l. : Academic Press, 1997.**

13. ***Relative contributions from traffic and aircraft NOx emissions to exposure in West London.*** **F.Farias, H.ApSimon. 4, s.l. : Elsevier, 2006, Environmental Modelling & Software, Vol. 21, pp. 477-485.**

14. ***Study of the evolution of air pollution in Salamanca (Spain) along a five-year period (1994–1998) using HJ-***

Biplot simultaneous representation analysis. **G.Cabrera, F.Martinez, M.Mateosa, V.Tavera. 1, s.l. : Elsevier, 2006, Environmental Modelling & Software, Vol. 21, pp. 61-68.**

15. *A MACHINE LEARNING TOOL TO FORECAST PM10 LEVEL.* G. Raimondo, Polytechnic Univ., Turin, Italy and and A. Montuori, W. Moniaci, E. Pasero, and E. Almkvist. s.l. : Fifth Conference on Artificial Intelligence Applications to Environmental Science , 2007. 5AI.

16. *Ozone peak and pollution forecasting using support vectors.* S.Canu, A.Rakotomamonjy. Yokohama Japan : IFAC Workshop on Environmental Modelling, 2001. IFAC Workshop on environmental modeling.

17. *Numerical investigation of the pollution dispersion in an urban street canyon.* P.Neofytou, A.G.Venetsanos, C.Rafailidis, J.G.Bartzis. 4, s.l. : Elsevier, 2006, Environmental Modelling & Software, Vol. 21, pp. 525-531.

18. *Prediction of ozone levels in London using the MM5–CMAQ modelling system.* R. S.Sokhi, R.San Joseb, N.Kitwiroon, E.Fragkou,J.L.Perez, D. R.Middleton. 4, s.l. :

Elsevier, 2006, Environmental Modelling & Software, Vol. 21, pp. 566-576.

19. Mileva Boshkoska, B. Sporedba na SVM regresioni modeli za predikcija na dnevni i časovni koncentacii na vo ambientalen vozduh. *Magisterska teza.* Skoje : fakultet za elektroktrotehnika i ing+formaciski tehnologii, 2008.

20. *PM10 forecasting for Thessaloniki, Greece.* T. Slini, A. Kaprara, K. Karatzas, N. Moussiopoulos. (2006) 559–565, s.l. : Elsevier, 2006, Vol. Environmental Modelling & Software 21.

21. *Forecasting severe ozone episodes in the Baltimore metropolitan area.* W.F.Ryan. 17, s.l. : Elsevier, 1995, Vol. Atmospheric Environment 29, pp. 2387–2398.

22. *Regression modelling of hourly NOx and NO2 concentrations in urban air in London.* J.P.Shi, R.M.Harrison. 24, s.l. : Elsevier , 1997, Vol. Atmospheric Environment 31, pp. 4081–4094.

23. *Comparing Neural Networks and Regression Models.* A.C.Comrie. 653–663, s.l. : Elsevier, 1997, Vol. Journal of the Air & Waste Management Association 47.

24. ***Prediction of NO and NO2 concentrations near a street with heavy traffic in Santiago, Chile.*** **P.Perez, A.Trier. 1783-1789, s.l. : Elsevier, 2001, Vol. Atmospheric Environment 35.**

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

26. ***Prediction of maximum daily ozone level using combined neural network and statistical characteristics.*** **W.Wang, W.Lu, X.Wang, A.Y.T.Leung. s.l. : Elsevier, 2003, Vol. Environment International 1049, pp. 1–8.**

27. ***A neural network based method for the short-term predictions of ambient SO2 concentrations in highly polluted industrial areas of complex terrain.*** **M.Boznar, M.Lesjak, P.Mlakar. 2, 221-230, s.l. : Elsevier, 1993, Vol. Atmospheric Environment B 27.**

28. ***"Air dispersion model and neural network: A new perspective for integrated models in the simulation of complex situations".*** **A. Pelliccioni, T. Tirabassi. s.l. :**

Elsevier, 2006, Vols. Environmental Modelling & Software 21 pp. 539–546.

29. ***"Regression and multilayer perceptron-based models to forecast hourly O3 and NO2 levels in the Bilbao area".*** **E.Agirre-Basurko, G.Ibarra-Berastegi and Madariaga, I. 4, s.l. : Elsevier, 2006, Environmental Modelling and Software, Vol. 21.**

30. ***"A neural network model forecasting for prediction of daily maximum ozone concentration in and industrialized urban area".*** **Yi J., Prybutok V.R. 349–357, s.l. : Elsevier, 1996, Vol. Environmental Pollution 92.**

31. ***Neural network modelling and prediction of hourly NOx and NO2 concentrations in urban air in London.*** **M.W.Gardner, S.R.Dorling. 5, s.l. : Elsevier, February 1999, Vol. Atmospheric Environment 33, pp. 709-719.**

32. ***Statistical surface ozone models: an improven methodology to account for non-linear behaviour.*** **M.W.Gardner, S.R.Dorling. 1, s.l. : Elsevier, January 2000, Vol. Atmospheric Environment 34, pp. 21-34.**

33. ***Extensive evaluation of neural network models for the prediction of NO2 and PM10 concentrations,.*** **J.Kukkonen,**

L.Partanen, A.Karppinen, J.Ruuskanen, H.Junninen, M.Kolehmainen, H.Niska, S.Dorling, T.Chatterton, R.Foxall, G.Cawley. 4539-4550, s.l. : Atmospheric Environment, 2003, Vol. 37.

34. S.R.Gunn. *ISIS ISIS-1-98 Technical Report: Support Vector Machines for classification and regression.* Image Speeh & Intelligent System Group, University of Southhampton. s.l. : University of Southampton, 1998.

35. *Genetic Algorithms and Support Vector Machines for Time Series Classification.* D.Eads, D.Hill, S.Davis,S.Perkins,M.Junshui, R.Porter, J.Theiler. 2002. SPIE 4787. pp. 74-85.

36. *Support vector machines experts for time series Forecasting.* L.Cao. s.l. : Elsevier, 2003, Neurocomputing, Vol. 31, pp. 321-339.

37. C.J.C.Burges. *A tutorial on Support Vector Machines for Pattern Recognition.* s.l. : Kluwer Academic Publishers, 1998.

38. *Predicting time series with support vector machines.* K. R.Muller, A.J.Smola, G.Ratshc, B.Scholkopf, J.Kohlmorgen,

V.Vapnik. London, UK : Springer-Verlag , 1997. ICANN. pp. 999-1004.

39. *Support vector method for function estimation, regression estimation and signal processing.* V.Vapnik, S.Golowich, A.Smola. Cambridge : MIT Press, 1997, Neural information processing sytems.

40. A.J.Smola, B.Scöpholf. *A tutorial on Support Vector Regression, Statistics and Computing.* Hingham, MA, USA : Kluwer Academic Publishers, 1998.

41. *Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends.* L.Wei-Zhen, W.Wen-Jian. s.l. : Elsevier, 2005, Chemosphere , Vol. 59, pp. 693-701.

42. *Air pollutant parameter forecasting using support vector machines.* W.Lu, W.Wang, A.Leung, S.M.Lo, R.Yuen, Z.Xu, H.Fan. Hawaii, USA : IEEE/ICJNN International Joint Conference on Neural Networks, 2002. IJCNN. pp. 630-635.

43. *A Machine Learning Tool to Forecast PM10 Level.* G.Raimondo, A.Montuori, W.Moniaci, E.Pasero, E.Almkvist. San Antonio, TX (USA) : WMO publications, 2007. AMS 87th Annual Meeting.

44. D.T.Larose. ***Discovering Knowledge in Data. An Introduction to Data Mining.*** **s.l. : Wiley-Interscience, 2005.**

45. J.Han, M.Kamber. ***Data Mining: Concepts and Techniques.*** **s.l. : Morgan Kaufmann Publishers, 2000.**

46. ***Principles of Data Mining.*** **D.Hand, H.Mannila, P.Smyth. s.l. : MIT Press, 2001.**

47. B.Schlkopf, P.Bartlett, A.Smolax, R.Williamson. ***Shrinking the Tube:A New Support Vector Regression Algorithm.*** **s.l. : MIT Press, 2000.**

48. ***Data Mining via Support Vector Machines.*** **L.Mangasarian, O. s.l. : Kluwer, 2001. 20th International Federation for Information Processing (IFIP) TC7 Conference on System Modeling and Optimization.**

49. ***Information Criteria for Support Vector Machines.*** **K.Kobayashi, K.Fumiyasu. 3, s.l. : IEEE Institute of electrical and electronics, 2006, IEEE Transactions On Neural Networks, Vol. 17, pp. 571-577.**

50. ***Environmental Data Mining and Modelling Based on Machine Learning Algorithms and Geostatistics.***

M.Kanevski, R.Parkin, A.Pozdnukhov, V.Timonin, M.Maignan, B.Yatsalo, S.Canu. s.l. : Elsevier, 2004, Vol. Environmental Modelling and Software 19, p. 414.

51. *"Support vector machines for regional clear-air turbulence prediction".* J.Abernethy and R.Sharman. 2007. Fifth Conference on Artificial Intelligence Applications to Environmental Science.

52. V.Kecman. *Learning And Soft Computing - Support Vector Machines, Neural Networks, And Fuzzy Logic Models.* s.l. : MIT Press, 2001.

53. B.Schlkopf, A.J.Smola. *Engineering - Learning with Kernels - Support Vector Machines, Regularization, Optimization, and Beyond.* s.l. : MIT Press, 2001.

54. N.Cristianini, J.Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* s.l. : Cambridge University Press, 2000.

55. *Support Vector Machines - Training and Applications.* E.E.Osuna, R.Freund, F.Girosi. s.l. : ACM, 1997.

56. *Statistical Learning Theory: a Primer.* T.Evgeniou, M.Pontil, T.Poggio. 1, s.l. : Springer, 2000, International

Journal of Computer Vision, Vol. International Journal of Computer Vision 38, pp. 9-13.

57. M.Kufmann, I. H.Witten. ***Data mining – Practical machine learning tools and techniques, 2nd edition.*** **s.l. : Morgan Kaufmann, 2000.**

58. ***Mean field method for the support vector machine regression.*** **J. B.Gao, S. R.Gunn, S. J.Harris. 2003, Neurocomputing , Vol. 50, pp. 391-405.**

59. L.Ljung, T.Glad. ***Modeling Of Dynamic Systems.*** **1994.**

60. ***An intelligent data analysis system for knowledge discovery and management in environmental databases.*** **K.Gibert, M.Sanchez-Marre, I.Rodriguez-Roda. 1, s.l. : Elsevier, 2006, Environmental Modelling and Software, Vol. 21, pp. 115-120.**

61. ***Analysis of Support Vector Machines regression bounds for variable ranking.*** **A.Rakotomamonjy. 7-9, s.l. : Elsevier , 2007, Neurocomputing , Vol. 70, pp. 1489-1501.**