IL NUOVO CIMENTO Vol. 29 C, N. 6 Novembre-Dicembre 2006
DOI 10.1393/ncc/i2006-10011-5

# Ozone prediction based on neural networks and Gaussian processes

B. Grašič(*), P. Mlakar and M. Z. Božnar
*AMES d.o.o. - Na Lazih 30, SI-1351, Brezovica pri Ljubjani, Slovenia*

(ricevuto l' 8 Marzo 2006; revisionato il 12 Ottobre 2006; approvato il 17 Ottobre 2006; pubblicato online il 29 Novembre 2006)

**Summary.** — The urban environment in Slovenia is confronted with the air pollution problem of harmfully high ozone concentrations. In the last two decades the automatic ozone measuring network was extended and now covers regions where the highest values are expected. Due to topographical and climatological conditions and the presence of extensive urban environments, the most critical locations are the ones in the western part of Slovenia. In the city of Nova Gorica a modern automatic urban air pollution measuring station was installed. Measurements at this station clearly showed that ozone is a considerable pollutant there, especially in the summer time. In this work a perceptron neural-network–based model and a Gaussian-process–based model for ozone concentration forecasting for the city of Nova Gorica was developed and evaluated. The methods of feature determination and pattern selection for the model training process are delineated. The shortcomings of the models and possibilities for improvements are discussed with respect to evaluation of the effectiveness of the methods.

PACS 92.60.Sz – Air quality and air pollution.

## 1. – Introduction

Ozone air pollution is becoming increasingly important in the urban environment in Slovenia. The first automatic measurements of ozone in Slovenia were undertaken more than ten years ago. Since then the measuring network has been extended and presently covers regions where the highest values are expected.

Due to specific topographical and climatological conditions and the presence of larger urban environments, the most critical locations are those in the western part of Slovenia that is open towards the Adriatic Sea and the Po valley. In the city of Nova Gorica a modern automatic urban air pollution measuring station was installed. It is part of the renewed national air pollution measuring network (ANAS) operating since 2001. The

(*) E-mail: bostjan.grasic@ames.si

ANAS network is owned and maintained by the Environmental Agency of the Republic of Slovenia. Measurements at the Nova Gorica station clearly show that ozone is a considerable pollutant there, especially in the summer time. Nova Gorica is situated near the larger Italian city of Gorizia in a Mediterranean-influenced location, characterized by hot summers with a low number of days with precipitation.

Thus in the present work, ozone forecasting based on neural networks and Gaussian processes was developed, and the created models used for ozone forecasting in real situations to protect and inform local inhabitants. Both models were constructed so as to predict the maximal hourly value of the ozone concentration on the following day. The aim of the research work was not testing of several training algorithms but rather feature selection and other important steps in the method of air pollution forecasting using advanced tools. The multilayer perceptron artificial neural network [1] used by many other researchers [2] was selected as the modelling tool. A Gaussian processes modelling tool was selected to compare its capabilities and efficiency against the artificial neural-network approach.

The paper begins with an introduction to the modelling tools used in this research. Multilayer perceptron neural network and Gaussian processes are described and compared to each other. The most crucial point in air pollution modelling represents pattern and feature selection. Sophisticated feature selection procedure outcomes are presented in the next section, where the finally selected input features in the models are listed. The following section describes the final construction and parameters of our two models. In sect. **5** performance indices used as a measure of quality and to compare the two models are determined. Finally the results of simulation runs for validation of both models are presented and evaluated through performance indexes.

## 2. – Modelling tools: multilayer perceptron neural network and Gaussian processes

In the past decade artificial neural networks has become a useful and efficient tool for establishing forecasting models in the field of air pollution. We started our work more than ten years ago by using a multilayer perceptron artificial neural network (MPNN) as a platform for $SO_2$ half hourly averaged concentration forecasting of $SO_2$ levels in the vicinity of the Šoštanj Thermal Power Plant in the North-East of Slovenia [3]. The huge available data base allowed us to establish feature selection and pattern selection techniques that proved to be the most important steps in the model construction phase. The MPNN also proved to be a useful tool in similar meteorological applications such as wind prediction [4] and reconstruction and diffuse solar radiation reconstruction [5]. In recent years other authors reported successful forecasting of air pollution using artificial neural networks. An overview of applications is given in Gardner [2].

For comparison another model was constructed on the basis of the Gaussian processes (GP)—non-linear regression method tool that performs well on a suite of real-world problems. In brief, according to papers [6] and [7] the Gaussian processes idea is based on placing a prior directly over the function values instead of parameterizing unknown function $f(\boldsymbol{x})$. Consider the system

$$y(k) = f(\boldsymbol{x}(k)) + \varepsilon(k), \tag{1}$$

where $\varepsilon(k)$ is white noise with variance $\upsilon_0$ and $\boldsymbol{x}(k)$ is the vector of system's inputs. To model this system, the Gaussian processes prior with covariance function

$$C(\boldsymbol{x}_i, \boldsymbol{x}_j) = \upsilon \cdot \exp\left[-\frac{1}{2}\sum_{d=1}^{D} w_d \left(x_i^d - x_j^d\right)^2\right] + \upsilon_0 \cdot \delta_{ij}, \tag{2}$$

with unknown hyperparameters $\boldsymbol{\theta} = [w_1, \ldots, w_D, \upsilon, \upsilon_0]^T$ is applied on space of functions $f(\cdot)$, where $\delta_{ij}$ is the Kronecker operator. When a stationary system with a smooth output is assumed, this covariance function is a common choice. To find the predictive distribution of output $y^*$ corresponding to a new given input $\boldsymbol{x}^*$, a set of $N$ training data pairs is gathered in $\mathcal{D} = \boldsymbol{X}|\boldsymbol{y}$. For this collection of (presumably normal) random variables $(y_1, \ldots, y_N, y^*)$ we can write: $(\boldsymbol{y}, y^*) \sim \mathcal{N}(0, \boldsymbol{K}_{N+1})$, where $\boldsymbol{K}_{N+1}$ is the covariance matrix of the process generating outputs $(\boldsymbol{y}, y^*)$. The elements of the covariance matrix $\boldsymbol{K}_{N+1}$ are the covariances between the values of the functions $f(\boldsymbol{x}_i)$ and $f(\boldsymbol{x}_j)$, calculated using the covariance function $C(\cdot,\cdot)$. The covariance matrix of the process is

$$\boldsymbol{K}_{N+1} = \begin{vmatrix} |\boldsymbol{K}| & |\boldsymbol{k}(\boldsymbol{x}^*)| \\ |\boldsymbol{k}(\boldsymbol{x}^*)^T| & |k(\boldsymbol{x}^*)| \end{vmatrix}. \tag{3}$$

The joint probability can be divided into marginal and conditional parts. Hyperparameters of the covariance function can be determined by two methods: evidence maximization [8] or the Monte Carlo approach [9]. The conditional part provides the (Gaussian) output distribution of the GP model with a mean $\mu(\boldsymbol{x}^*)$ and variance $\sigma^2(\boldsymbol{x}^*)$:

$$\mu(\boldsymbol{x}^*) = \boldsymbol{k}(\boldsymbol{x}^*)^T \cdot \boldsymbol{K}^{-1} \cdot \boldsymbol{y}, \tag{4}$$

$$\sigma^2(\boldsymbol{x}^*) = k(\boldsymbol{x}^*) - \boldsymbol{k}(\boldsymbol{x}^*)^T \boldsymbol{K}^{-1} \cdot \boldsymbol{k}(\boldsymbol{x}^*), \tag{5}$$

where $\boldsymbol{k}(\boldsymbol{x}^*) = [C(\boldsymbol{x_1}, \boldsymbol{x}^*), \ldots, C(\boldsymbol{x_N}, \boldsymbol{x}^*)]^T$ is the vector of covariances between training inputs and the test input and $k(\boldsymbol{x}^*) = C(\boldsymbol{x}^*, \boldsymbol{x}^*)$ is the autocovariance of the test input. For a more detailed introduction to Gaussian processes see, *e.g.*, [10] or [9].

Both tools, the multilayer perceptron neural network and the Gaussian processes, construct a model on the basis of learning information from the data log of measurements using an appropriate training algorithm. This is the most important capability of the tools presented. Interest in Gaussian processes was initiated by the work of Neal [11] on priors for infinite networks. Neal showed that the properties of a neural network with one hidden layer converge to those of a Gaussian process as the number of hidden neurons tends to infinity if standard "weight decay" priors are assumed. The covariance function of this Gaussian process depends on the details of the priors assumed for the weights in the network and the activation functions of the hidden units. This observation motivated the idea of replacing supervised neural networks by Gaussian processes, a research direction explored by Williams and Rasmussen [9] and Neal [12]. Gaussian process models have already been successfully used for modelling of dynamic systems [7] and subsequently for prediction, *e.g.* [6].

The aim of this work was not testing a number of training algorithms but feature selection and other important steps in the method of air pollution forecasting using these advanced tools. We assumed according to our previous work [3,4,13,5] with a back propagation training algorithm that other training algorithms should give similar results. The crucial point is the features and patterns used. Both models tested used the same

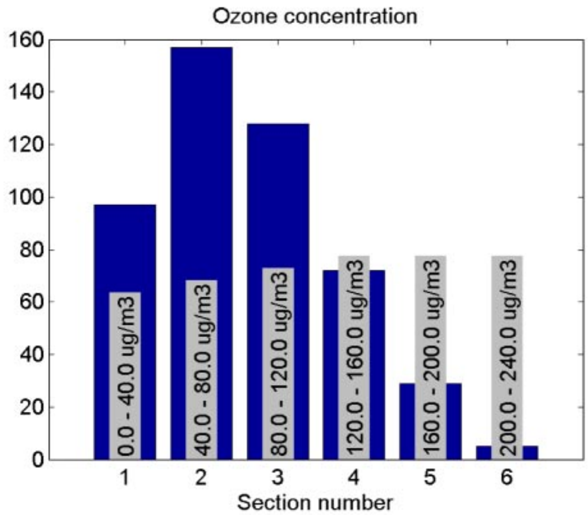


Fig. 1. – Histogram output feature "ozone concentration".

input features and the same set of learning patterns. When the models were constructed they were tested on the same independent validation set of patterns. In this paper only results on the validation set are presented.

## 3. – Pattern selection

The ANAS automatic station in Nova Gorica has been operating only for a short period. The model was constructed to predict the maximal hourly value of the ozone concentration for the following day—therefore only one pattern per day is available. Data logs from the start of 2002 until the end of 2004 were used. August 2003 (high concentrations), January 2004 (low concentrations) and September 2004 (medium concentrations) were excluded and used for the validation set, and all the rest of the patterns were used for the learning set. The learning set was divided randomly into a 10% set used for optimization and the rest used for the training algorithm during the model construction. No other pattern selection was performed because of the relatively small set of data logs available.

## 4. – Feature selection

Feature selection was a sophisticated procedure. The ANAS automatic measuring station at Nova Gorica measures basic meteorological parameters (wind, temperature, relative humidity, air pressure, global solar radiation), pollution by gases ($SO_2$, $O_3$, NO, $NO_2$, $NO_X$) and dust ($PM_{10}$). A forecast of the basic meteorological parameters for the next day is available for the city of Nova Gorica from the ALADIN meteorological prognostic model. As these ALADIN prognostic values were not available in our data logs, we tested the model construction by using the values as forecasted by ALADIN as measured ones. Therefore with real ALADIN values the performance of the ozone prediction model may not be as good as the prognosis is not perfect. But at least we could test the upper limits of modelling. All the measured parameters were available as half-hour averaged values or any longer interval averaged values. Features were selected in two steps. Firstly a wide range of features that could influence the following day's ozone concentration were selected according to the modeller's knowledge about the phenomenon. As an example, fig. 1 shows learning patterns divided into 6 groups according to the output feature—ozone prediction.

After preliminary feature selection both models were trained with all the available patterns and the pre-selected features. Then contribution factors and Saliency metrics were calculated [13]. Both give a useful measure of the non-linear relationship between a particular input feature and the output. On the basis of the score a particular parameter was rejected or not.

The finally selected input features were:

1) air temperature (24 h average),

2) global solar radiation (24 h average),

3) NO (24 h average),

4) $NO_2$ (24 h average),

5) $O_3$ (24 h average),

6) maximum air temperature (prognostic),

7) north-south direction wind speed (prognostic),

8) east-west direction wind speed (prognostic).

Here the 24 h average values 3) to 5) were calculated by taking the one-hourly measured average values for the previous day up to 19 h today.

Here the prognostic values 6) to 8) were those forecasts for the following day from the data base.

## 5. – Modelling

After feature selection was made, the model was constructed using the training and optimizing sets of patterns.

The MPNN consists of one hidden layer constructed of 14 neurons and one output layer constructed of 1 neuron. All neurons use a tangent-sigmoid transfer function to generate their output:

$$\mathrm{tansig}\,(x) = \frac{2}{1+e^{-2x}} - 1. \tag{6}$$

The MPNN was trained using the Levenberg-Marquardt method with the learning rate set to 0.2 and momentum set to 0.3. MNPP was implemented using Matlab's Neural Network Toolbox [14].

The GP was implemented using Rasmussen's Gaussian Processes Software Code for Matlab [15]. Since there were 8 inputs in the model, there were 10 hyperparameters to define. The hyperparameters of GP were determined using the evidence maximization method. The values of hyperparameters determined are shown in fig. 2.

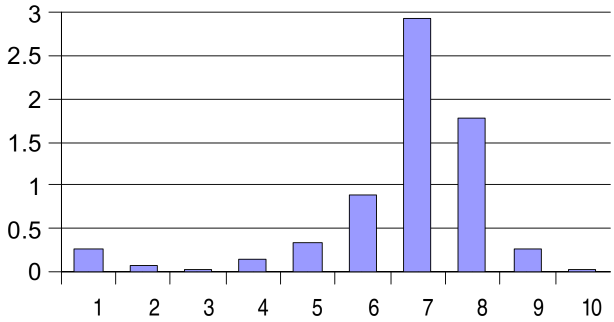


Fig. 2. – Determined hyperparameter values.

The first 8 hyperparameters refer to each input. A larger hyperparameter value signifies that the predictions made using GP would have less dependence on the corresponding input. The $\Theta 9$ hyperparameter gives the overall vertical scale relative to the mean of GP in output space. The $\Theta 10$ hyperparameter gives the vertical uncertainty. This reflects how far we expect the mean of the function to fluctuate from the mean of GP.

## 6. – Results

After the model was constructed it was tested on an independent validation set that was not used during the learning period. The results of a one-step ahead prediction are shown in fig. 3 and fig. 4. The figures show that both MPNN and GP models give results

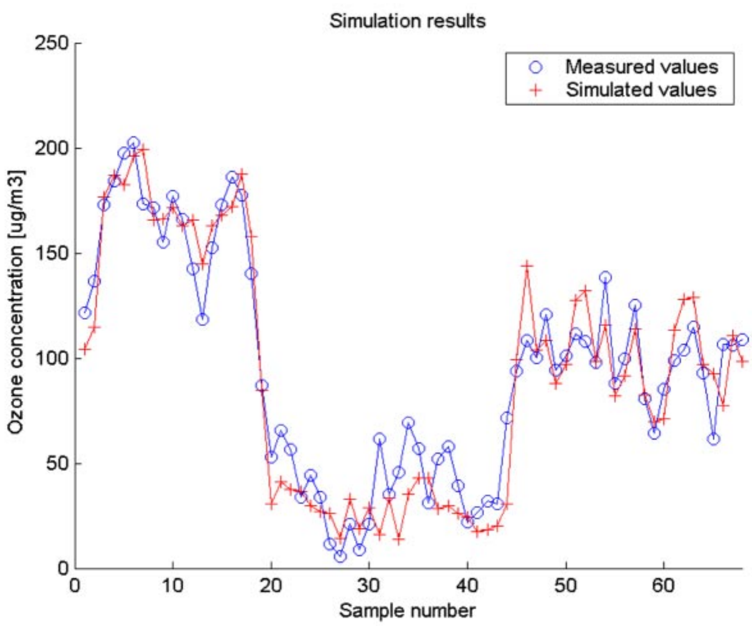


Fig. 3. – MPNN simulation results.

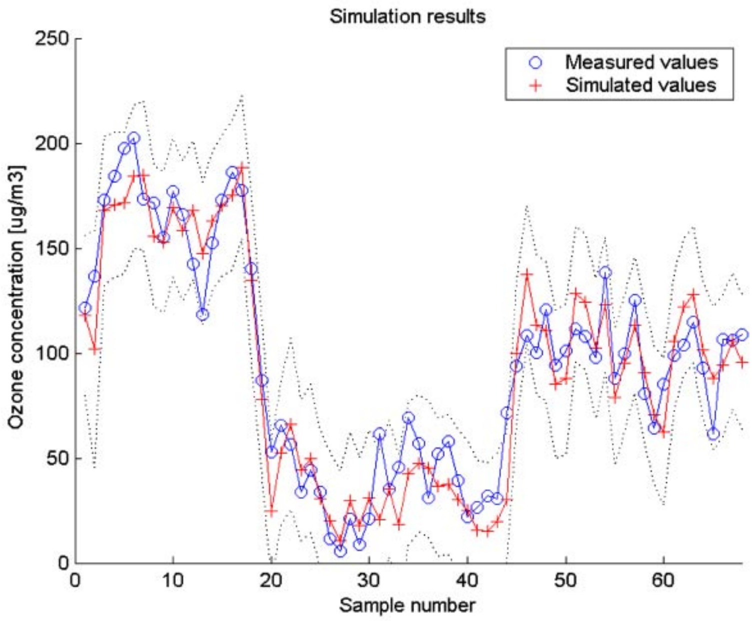


Fig. 4. – GP simulation results.

that can be used for ozone forecasting in real situations and the capabilities shown on the figures are already good enough to use the model for informing citizens about the possibility of high and alarm concentrations occurring.

In order to have a measure of quality and to compare the two models, the following four performance indices have been determined:

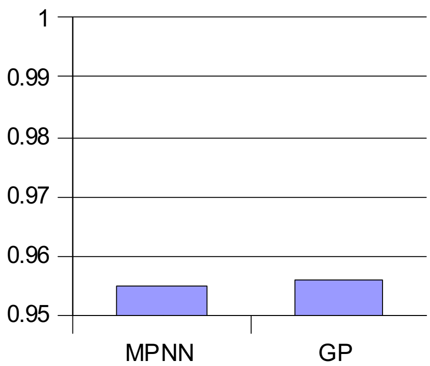


Fig. 5. – Correlation factor.

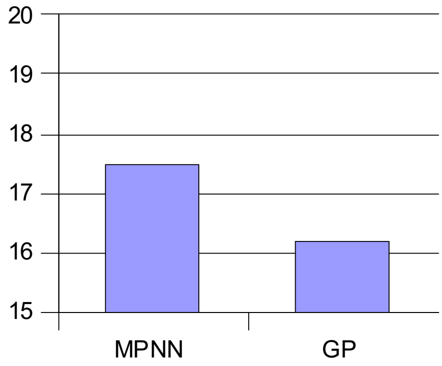


Fig. 6. – Root mean square error—RMSE.

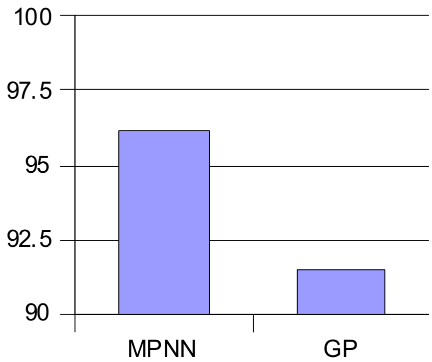


Fig. 7. – Success index—SI.

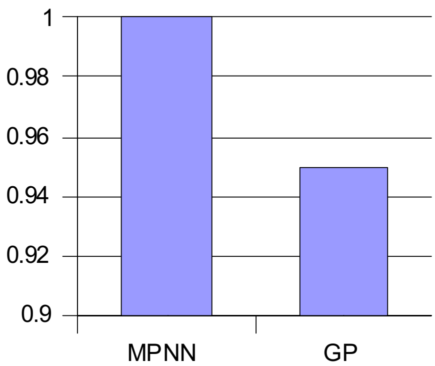


Fig. 8. – Performance index—$p^6$.

– The correlation coefficient (commonly used by many researchers in the field of air pollution modelling; the calculated value for both simulations is presented on fig. 5):

$$r(Cm, Cr) = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(Cm_i - \hat{C}m\right)\left(Cr_i - \hat{C}r\right)}{\sigma_{Cr}\sigma_{Cr}}. \tag{7}$$

– The root-mean-square error (used as a performance function for MPNN training; the calculated value for both simulations is presented on fig. 6):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Cr_i - Cm_i\right)^2}. \tag{8}$$

– The success index (defined by the European Environment Agency following a standard contingency table [16]) with a threshold of $140\,\mu\mathrm{g/m}^3$ (the calculated value for both simulations is presented on fig. 7):

$$SI = \left(\frac{a}{m} + \frac{N + a - m - f}{N - m} - 1\right) \cdot 100\%. \tag{9}$$

– The performance index [13] (calculated value for both simulations is presented on fig. 8):

$$p^6 = \frac{1}{N}\sum_{i=1}^{N} J_i^6, \tag{10}$$

where

$Cm_i$ … $i$-th measured concentration,
$Cr_i$ … $i$-th calculated concentration,
$\hat{C}$ … average concentration,
$\sigma_C$ … concentration standard deviation,
$N$ … number of patterns tested,
$f$ … number of forecasted events when the limit is exceeded,
$m$ … number of measured events when the limit is exceeded,
$a$ … number of correctly forecast events when the limit is exceeded.

The performance index $p^6$ is used to stress the importance of high concentrations [13]. The correct classification is calculated by the cost function $J^6$. The cost function $J^6$ equals "1" (correctly classified) if there is no case of a false alarm and the measured concentration is high, and if at the same time, the absolute error is less than $20\,\mu\mathrm{g/m}^3$ or the relative error is less than 20%.

Correct classifications and classification errors are counted if there is a high ozone concentration and not counted if the measured concentration is less than $120\,\mu\mathrm{g/m}^3$ and at the same time the forecasted ozone concentration is less than $180\,\mu\mathrm{g/m}^3$ (no false alarm).

According to fig. 2 the most important inputs are NO, global solar radiation, $NO_2$, air temperature and $O_3$. Experience gained in this project showed that prognostic inputs had a relatively minor dependence on high ozone concentration predictions and a major dependence on low ozone concentration predictions.

## 7. – Conclusions

The air pollution system and decision support system used in our research work was presented. The multilayer perceptron neural network and Gaussian processes that were used as modelling tools were described and compared. Finally, input features chosen as a result of a sophisticated feature selection procedure were presented and listed: namely, air temperature, global solar radiation, NO, $NO_2$ and $O_3$ concentrations and prognostic maximum air temperature and wind speed and direction. It was also shown that among these, the major role is played by $O_3$, NO and $NO_2$ concentrations, air temperature and global solar radiation.

The construction and parameters of both models created were described. Performance indices to compare the two models were introduced and results of a simulation on a validation data set for both models were presented and evaluated by the performance indexes. Evaluation showed that MPNN performed slightly better than GP as measured by the success index SI and performance index $p^6$, while as regards the correlation factor and root-mean-square error evaluation, GP performed slightly better. From this result it can be concluded that neither of the created models is superior and the effort required to be invested in the two modeling techniques was approximately the same.

The main outcome of this study consists of the list of determined input features that are the most important in ozone forecasting. Other outcomes are the constructed models that are suitable and useable in real situations. Very similar results obtained by both modelling techniques show that the most important step in air pollution modelling is feature selection.

REFERENCES


[1] Hornik K., Stinchcombe M. and White H., *Neural Networks*, **2** (1992).

[2] Gardner M. W. and Dorling S. R., *Atmos. Environ.*, **32** (1998) 2627.

[3] Božnar M., Lesjak M. and Mlakar P., *Atmos. Environ.*, **27**B (1993) 221.

[4] Božnar M. and Mlakar P., *Use of neural networks in the field of air pollution modelling*, in *25th NATO/CCMS International Technical Meeting on Air Pollution Modelling and its Application, 15-19 October 2001, Aula Magna, Louvain-la-Neuve, Belgium.* Preprints Louvain: Université Catholique de Louvain, 2002, pp. 265-272.

[5] Soares J., Oliveira A. P., Božnar M. Z., Mlakar P., Escobedo F. J. and Machado J. A., *Appl. Energy*, **79** (2004) 201.

[6] Ažman K. and Kocijan J., *Gaussian process models validation: biotechnological systems case studies*, in *5th MATHMOD Vienna*, edited by Troch I. and Breitenecker F., *Proceedings of 5th Vienna Symposium on Mathematical Modelling, February 8-10, 2006*, Vienna University of Technology, Austria, ARGESIM Report, no. 30 (ARGESIM, Vienna), Vol. **2** (2006), pp. 3-1–3-10.

[7] Kocijan J., Girard A., Banko B. and Murray-Smith R., *Math. Computer Modelling Dyn. Syst.*, **11** (2005) 411.

[8] Mackay D. J. C., *Hyperparameters: Optimize, or integrate out?*, in *Maximum Entropy and Bayesian Methods, Santa Barbara 1993*, 1996, pp. 43-60.

[9] RASMUSSEN C. E., *Evaluation of Gaussian Processes and other Methods for Non-Linear Regression*, in Ph. D. Dissertation, University of Toronto (1996).

[10] GIBBS M. N., *Bayesian Gaussian Processes for Regression and Classification*, in Ph. D. Dissertation, Cambridge University (1997).

[11] NEAL R. M., *Bayesian Learning for Neural Networks*, in *Lecture Notes in Statistics*, No. **118** (Springer, New York) 1996.

[12] NEAL R. M., *Monte Carlo Implementation of Gaussian process Models for Bayesian Regression and Classification*, in Technical Report CRG–TR–97–2, Dept. of Computer Science, University of Toronto (1997).

[13] MLAKAR P., *Determination of features for air pollution forecasting models*, in *Proceedings of Intelligent Information Systems IIS'97*, Grand Bahama Island, Bahamas, December 8-10, 1997, edited by ADELI H. (IEEE Computer Society, Los Alamitos, CA) 1997, pp. 350-354.

[14] THE MATHWORKS, INC., *Neural Network Toolbox*, in Matlab version 6.5 documentation (2002).

[15] RASMUSSEN C. E., *Gaussian Processes for Regression software code*, `http://www.gatsby.ucl.ac.uk/~edward/code/gp/`, 10.10.2004.

[16] VAN AALST R. M. and DE LEEUW F. A. A. M., *National Ozone Forecasting System and International Data Exchange in Northwest Europe*, *European Topic Centre on Air Quality*, `http://reports.eea.eu.int/TEC09/en/tech09.pdf` (1997).