16th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
8-11 September 2014, Varna, Bulgaria

OZONE FORECASTING USING GAUSSIAN PROCESSES AND PERCEPTRON NEURAL NETWORKS

Dejan Petelin[1] and Primož Mlakar[2] and Marija Zlata Božnar[2] and Boštjan Grašič[2] and Juš Kocijan[1,3]

[1]Jožef Stefan Institute, Ljubljana, Slovenia
[2]MEIS d.o.o., Mali Vrh pri Šmarju, Slovenia
[3]University of Nova Gorica, Nova Gorica, Slovenia

Abstract: Ozone is one of the main air pollutants with harmful influence to human health. Therefore, forecasting the ozone concentration and informing the population, when the air-quality standards are not being met, is an important task. One of the possibilities for forecasting are Gaussian process (GP) models or on the other hand artificial neural networks (ANN) that both can make a forecast using available present time or historical measurements of air pollution or meteorological parameters at the location of automatic air quality measuring stations. In this paper, an on-line updating, or evolving, GP model is evaluated. Its main advantage is the ability of learning from scratch, i.e. with almost no prior knowledge or data. This means that it can be used for modelling various variables shortly after the measurement station is established. To assess the viability of the evolving GP model for the ozone forecasting it is compared to the full GP model and multilayer-perceptron ANN model. The investigation shows that the evolving GP model performs sufficiently good to use it for informing citizens about the possibility of high and alarm concentrations occurring, especially, if the prediction-confidence interval is taken into account.

Key words: *Air pollution forecasting, Gaussian processes, on-line modelling.*

INTRODUCTION

Evolving Gaussian processes (EGP) is an on-line learning method which sequentially updates the model with incoming data. Its main advantage is the ability of learning from scratch, i.e. with almost no prior knowledge or data. That means it can be used for modelling various variables shortly after the measurement station is established.

In this paper a comparison of the EGP model with full GP model and multilayer perceptron neural networks (MPNN) for ozone forecasting is presented. Ozone is one of the main air pollutants with harmful influence to human health. Exposure to ozone can cause serious health problems in plants and people, thus ozone pollution is a major problem in some regions of the world. Therefore, forecasting the ozone concentration and informing the population when the air-quality standards are not being met is an important task.

Due to specific topographical and climatological conditions and the presence of urban environments, for ozone, the most critical locations in Slovenia are those in the western part that is open towards the Adriatic Sea and the Po valley. More than a decade ago an automatic urban air pollution measuring station was installed in Nova Gorica as a part of the national air pollution measuring network (ANAS) operating since 2001. Based on the measurements from the measuring station in Nova Gorica various modelling tools were used to predict the ozone concentrations in the air. In (Božnar et al, 2004) MPNN were used and compared to fuzzy logic models. It was shown that both models give satisfactory results. Later in (Grašič et al, 2006) also GP model was compared to MPNN. The obtained results show that GP model performed very similar to MPNN.

The paper is structured as follows. In the following section the modelling tools, MPNN and GPs, are introduced. Further, a method for ozone forecasting based on an on-line updated EGP model is presented.

After that the conducted experiment is described and the obtained results are given. The concluding remarks end the paper.

MODELLING TOOLS

The relationship between cause and consequence, or system input and system output, can be modelled in many ways. If the physics behind the relationship is known one can use first principle models. If this is not the case, models need to be derived out of the available input and output data. Those approaches are referred to as data-driven.

The most frequent way to relate input $\mathbf{x} \in \Re^D$ and output $y \in \Re$ is to use regression model parameterized by a finite vector of model parameters $\boldsymbol{\theta} \in \Re^D$. For the sake of simplicity we will consider static relationship:

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

where $f : \Re^D \to \Re$ is a known function (e.g. linear, polynomial, radial basis function, neural network etc.) known up to the vector $\boldsymbol{\theta}$, ε is noise term needed to describe model imperfection caused by random disturbances or modelling errors. Usually the noise term is described by a probability density function $\varepsilon \sim p_\vartheta(\varepsilon)$ parameterized by ϑ. Having the data records in terms of pairs $\{\mathbf{x}_i, y_i, i = 1,..., N\}$ the model (1) can be identified from data by estimating the the unknown model parameters $\{\boldsymbol{\theta}, \vartheta\}$. The alternative to the parametric model (1) is to use structure-free non-parametric models, such as GP models.

Multilayer perceptron neural network

The MPNN (Božnar et al, 2004) is a feed-forward neural network with one or more layers between input and output layer. Each layer consists of simple neurons. The neuron computes a single output from multiple real-valued inputs by forming a linear combination according to its input weights and then possibly putting the output through some nonlinear activation function. Mathematically this can be written as:

$$\mathbf{y} = \varphi(\mathbf{w}^T \mathbf{x} + b) \tag{2}$$

where $\mathbf{w}$ denotes the vector of weights, b is the bias and φ is the activation function. Feed-forward means that data flows in one direction from input to output layer. This type of network is trained with the supervised technique called back-propagation, which consists of two steps. In the forward pass, the predicted outputs corresponding to the given inputs are evaluated, while in the backward pass, partial derivatives of the cost function are propagated back through the network. The weights can then be optimised by using any gradient-based optimisation method. The process is repeated until the weights have converged.

Gaussian processes

The GP models are probabilistic, non-parametric models based on the principles of Bayesian probability. They differ from most of the other black-box identification approaches in that they search for relationships among the measured data rather than try to approximate the modelled system by fitting the parameters of the selected basis functions. The output of the GP models is a normal distribution, expressed in terms of the mean and the variance. Their modelling properties are reviewed in (Rasmussen and Williams, 2005) with applications in, e.g. (Bukkapatnam and Cheng, 2010, Huang et al, 2011).

The idea of GP models is rather simple. GP model assumes that the output is realization of a GP with joint probability density function:

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{m}, \mathbf{K}) \tag{3}$$

with mean and covariance being functions of the inputs $\mathbf{X}$. Usually the mean function is defined as $\mathbf{0}$, while the covariance function or kernel:

$$\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) \tag{4}$$

defines the characteristics of the process to be modelled, i.e. stationarity, smoothness etc. Most commonly used is the squared exponential (SE) covariance function with automatic relevance determination (ARD). This covariance function is smooth and stationary and enables to determine the impact on the model for each input by optimization of parameters, called hyperparameters.

Once the covariance matrix $\mathbf{K}$ is calculated, the predictive (normal) distribution for new input is simply calculated by:

$$\mu(y^*) = \mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1}\mathbf{y} \tag{5}$$

$$\sigma^2(y^*) = \kappa(\mathbf{x}^*) - \mathbf{k}(\mathbf{x}^*)^T \mathbf{K}^{-1}\mathbf{k}(\mathbf{x}^*) \tag{6}$$

Where $\mathbf{k}(\mathbf{x}^*)$ is the vector of covariances between new input sample and training input samples, and $\kappa(\mathbf{x}^*)$ is the covariance between the new input sample itself.

As can be seen from equations (6) the GP model, in addition to mean value, also provides information about the confidence in prediction by the variance. Usually, the confidence of the prediction is depicted with 2σ interval which is about 95% confidence interval. It highlights areas of the input space where the prediction quality is poor, due to the lack of data or noisy data, by indicating a wider confidence interval around the predicted mean.

ON-LINE MODELLING

A noticeable drawback of GP model is the computation load that increases with the third power of the amount of input data due to the calculation of the inverse of the covariance matrix. This computational complexity restricts the amount of training data to, at most, a few thousand cases. To overcome the computational limitation issues, only a subset of the most informative data is to be used. In the literature, such a subset is called the *active set*. The basic idea is to retain the bulk of the information contained in the full training dataset, but reduce the size of the covariance matrix so as to facilitate a less computationally demanding implementation of the GP model. As the data is in-streaming the GP model should be adapted continuously. In other words, the on-line learning method should process every new piece of streaming data sequentially.

The basic idea of evolving GP models (Petelin et al. 2013, Petelin and Kocijan, 2014). is that all influential parts of the GP model should be adapted on-line. Although the proposed concept considers all four parts that can evolve, the implementation is somewhat more facile. The SE covariance function with ARD functionality is used presuming smoothness and stationarity of the system. That means the covariance function is fixed and does not need to evolve. Furthermore, with the optimization of the hyperparameter values, due to the ARD functionality uninfluential regressors have smaller values and as a consequence have smaller influence to the result. Therefore, all available regressors can be used. Consequently, only the active set and hyperparameter values have left to be adapted sequentially. More detailed description of the method is available in (Petelin et al, 2013, Petelin and Kocijan, 2014).

EXPERIMENT

To assess the viability of the EGP method for predicting the ozone concentration in the air, the same experiment setup is performed performed as in (Grašič et al, 2006) and is compared with the results obtained by the off-line trained MPNN and full GP model in (Grašič et al, 2006).

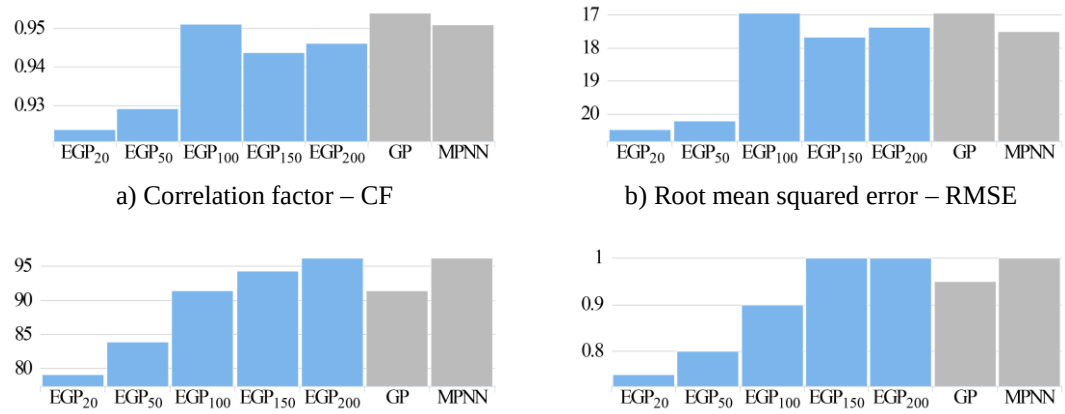
The aim of the experiment is to predict the maximal hourly value of the ozone concentration for the following day – therefore only one sample per day is available. The data logs from the ANAS automatic station in Nova Gorica, measured from the start of 2002 until the end of 2004 are used. Samples from August 2003, presenting high concentrations, from January 2004, presenting low concentrations, and from September 2004, presenting medium concentrations, are used for the validation (68 samples), while all the rest of the samples are used for the training (488 samples).

The data logs contain various measurements, such as basic meteorological parameters: wind, temperature, relative humidity, air pressure, global solar radiation and pollution gases: SO_2, O_3, NO, NO_2, NO_X, dust (PM_{10}). Even though the forecasts of the basic meteorological parameters are available[1], the measured values as forecasts are used to avoid introducing additional uncertainty in the model. However, we use the same input features as those obtained by sophisticated feature selection procedure performed in (Grašič et al, 2006): air temperature (24 h average), global solar radiation (24 h average), NO (24 h average), NO_2 (24 h average), O_3 (24 h average), maximum air temperature (prognostic), north-south direction wind speed (prognostic), east-west direction wind speed (prognostic). The 24 h average values are calculated by taking the hourly measured average values for the previous day up to 19:00 of the present day. The prognostic values are forecasts for the following day.

The EGP model is validated with various active set sizes: 20, 50, 100, 150 and 200. In all cases we use the same initial hyperparameter values as used in (Grašič et al, 2006) for off-line trained GP model. Note that the EGP models are started from scratch, i.e. with no training data. The EGP models are updated sequentially with every data available in data logs. At the end, only predictions from validation set are used to guarantee relevant comparison to the results obtained in (Grašič et al, 2006).

RESULTS

To measure the performance of the models we choose the following error measures: correlation factor – CF, root mean square error - RMSE, success index – SI and performance index – p^6 (Grašič et al, 2006). The values of all selected error measures of EGP models with various active set sizes are presented in Fig. 1. For comparison also the values of selected error measures obtained by full GP model and MPNN are given.



a) Correlation factor – CF

b) Root mean squared error – RMSE

c) Success index – SI

d) Performance index – p^6

Figure 1. Selected error measures for EGP models with various active set sizes compared to full GP model and MPNN model, both trained by using full training set of 488 samples.

It can be seen that EGP models with bigger active sets (100, 150 and 200) perform very similar to full GP and MPNN models. The EGP model with active set size 100 performs very well regarding the error measures CF and RMSE, which penalise only the error, while the EGP with active set size 200 performs better regarding the error measures SI and p^6, which additionally penalise false alarms.

Predictions obtained by the EGP model with active set size 200 are depicted in Fig. 2. It can be seen that despite smaller training set[2] the predictions are sufficiently accurate in most cases to use the model for informing citizens about the possibility of high and alarm concentrations occurring, especially, if the prediction-confidence interval is taken into account.

[1] The forecasts for the next day are available from the ALADIN or WRF meteorological prognostic model.

[2] Active set, which is adapted sequentially.

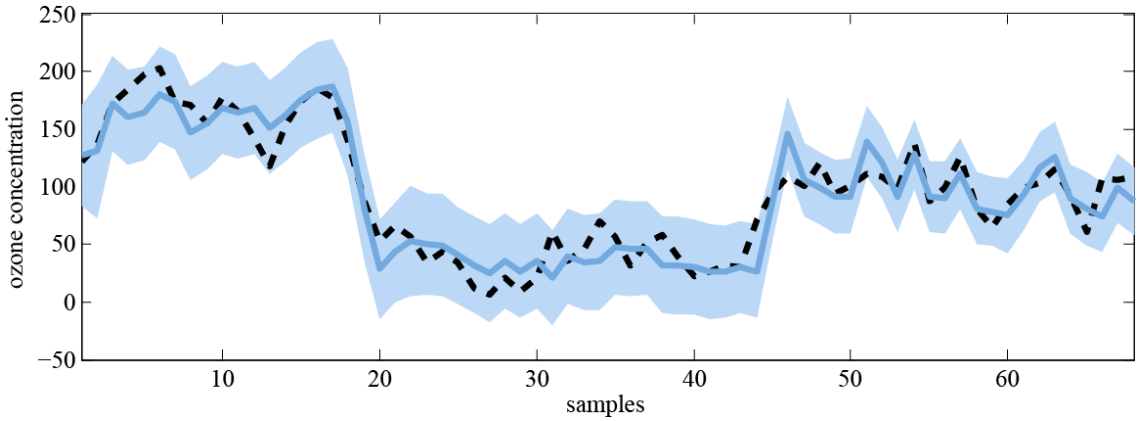


Figure 2. Predictions of the maximal ozone concentration (one day ahead) for validation data set. Predictions are obtained by the EGP model with active set size 200. Black dashed line depicts measured values, while blue solid line depicts the mean value of predictions and blue band depicts the 2nd standard deviation of predictions, which is interpreted as a ~95 % confidence interval.

CONCLUSIONS

In this paper we compared an on-line learning GP model, called EGP model, with full GP model and MPNN model for the ozone forecasting. In particular, the EGP models with various active set sizes were compared to the full GP model and MPNN model. The investigation shows that the EGP models with bigger active set sizes (100, 150 and 200) performed sufficiently good to use them for informing citizens about the possibility of high and alarm concentrations occurring, especially, if the confidence interval is taken into account.

The main advantage of the EGP model is that it is adapted sequentially with every new data. Furthermore, the EGP model can be trained from scratch, i.e. without prior knowledge or data. That means it can be used immediately after the measurement station is established.

ACKNOWLEDGEMENT

We like to acknowledge the support of the Slovenian Research Agency through the Research Programme P2-0001 and the Research Projects L2-5475. We are also very grateful to the Slovenian Environment Agency for providing the data.

REFERENCES

Božnar, M. Z., Mlakar, P. and Grašič, B., 2004: Neural networks based ozone forecasting. *Proceedings of the 9th International conference on Harmonisation with Atmospheric Dispersion Modelling for Regulatory Purposes*, Garmisch-Partenkirchen, Germany, 356-360.

Bukkapatnam, S. T. S. and Cheng, C., 2010: Forecasting the evolution of nonlinear and nonstationary systems using recurrence-based local Gaussian process models. *Physical Review E*, **82 5**, 056206.

Grašič, B., Mlakar, P. and Božnar, M. Z., 2006: Ozone prediction based on neural networks and Gaussian processes. *Il Nuovo Cimento*, **29 C**, 651-661.

Huang, W., Wang, K., Jay Breidt, F. and Davis, R. A., 2011: A class of stochastic volatility models for environmental applications. *Journal of Time Series Analysis*, **32 4**, 364-377.

Petelin, D., Grancharova, A. and Kocijan, J., 2013: Evolving Gaussian process models for prediction of ozone concentration in the air. *Simulation modelling practice and theory*, **33**, 68-80.

Petelin, D. and Kocijan, J., 2014: Evolving Gaussian process models for predicting chaotic time-series. *Proceedings of the IEEE Conference on Evolving and Adaptive Intelligent Systems 2014*, Linz, Austria.

Rasmussen, C. E. and Williams, C. K. I., 2006: Gaussian processes for machine learning, The MIT Press, Cambridge, MA, USA, 266 pp.