Ozone forecasting using Gaussian processes and Perceptron neural networks

Dejan Petelin[1], Primož Mlakar[2], Marija Zlata Božnar[2], 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

The aim of the project is forecasting of ozone concentrations at Nova Gorica town where an air quality station operates for several years. Forecasting is made by three types of black box models.

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.

MODELLING TOOLS:

- **Multilayer perceptron neural network**
- **Gaussian processes**
- **Evolving Gaussian processes**

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.

EXPERIMENT

To assess the viability of the EGP method for predicting the ozone concentration in the air, the same experiment setup is 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.

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. 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. Even though the forecasts of the basic meteorological parameters are available (the forecasts for the next day are available from the ALADIN or WRF meteorological prognostic model), 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).

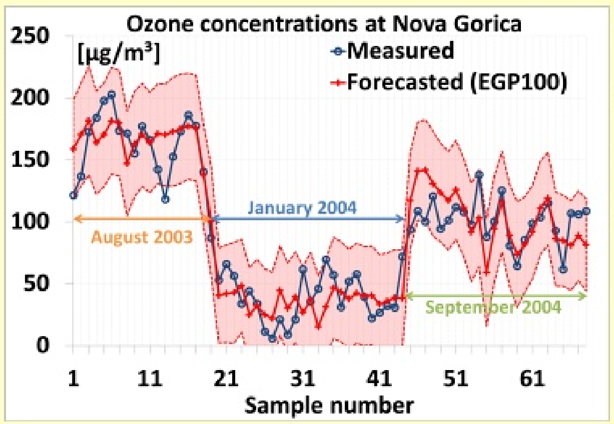
The 24 h average values are calculated by taking the hourly measured average values for the previous day up to 19:00 of the present day. The prognostic values are forecasts for the following day.

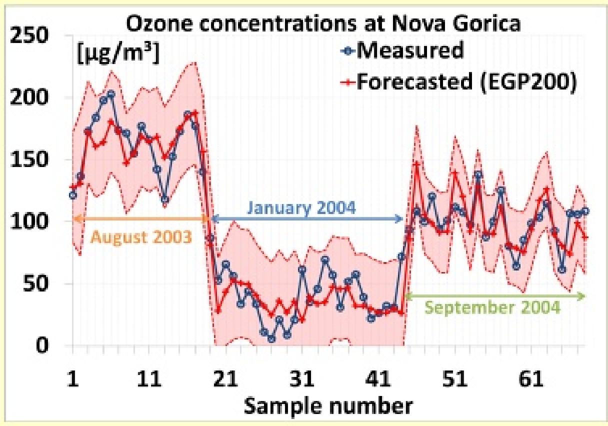
The EGP model is validated with various active set sizes: 20, 50, 100, 150 and 200. In all cases we use the same initial hyperparameter values as used in (Grašič et al, 2006) for off-line trained GP model. Note that the EGP models are started from scratch, i.e. with no training data. The EGP models are 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).

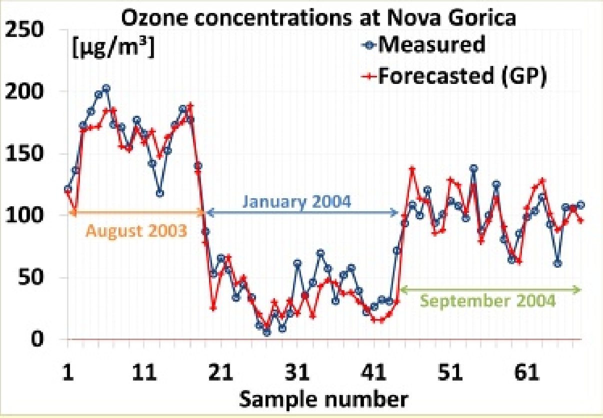
CONCLUSIONS

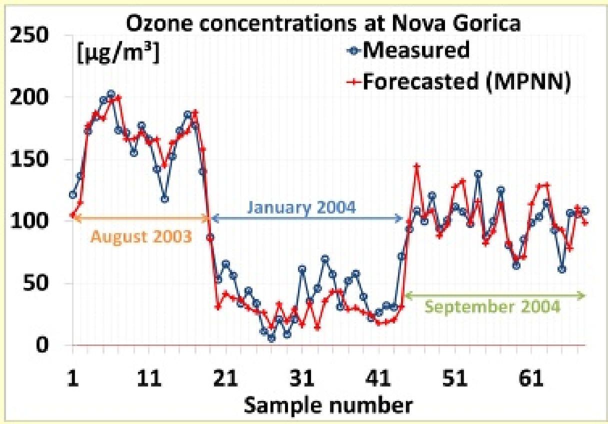
In this paper we compared an on-line learning GP model, called EGP model, with full GP model and MPNN model for the ozone forecasting. In particular, the EGP models with various active set sizes were compared to the full GP model and MPNN model. The investigation shows that the EGP models with bigger 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.

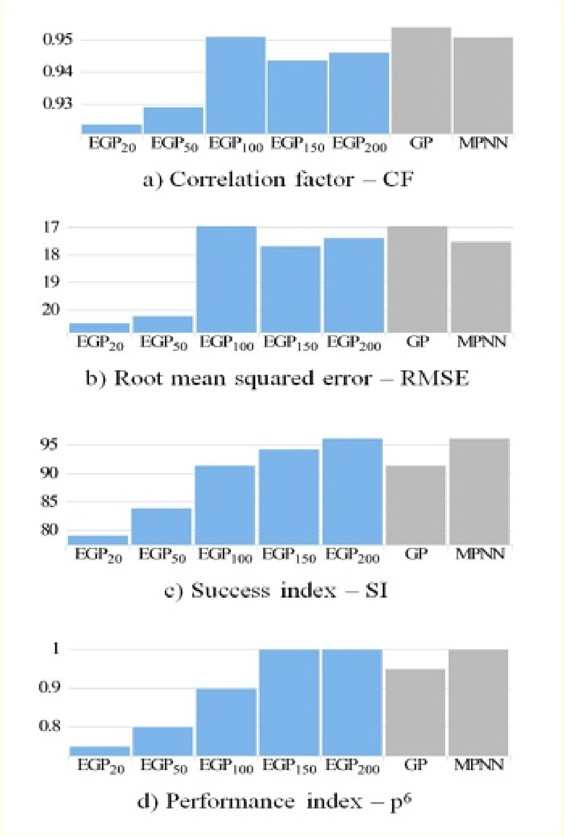
The main advantage of the EGP model is that it is adapted sequentially with every new data. Furthermore, the EGP model can be trained from scratch, i.e. without prior knowledge or data. That means it can be used immediately after the measurement station is established.











a) Correlation factor – CF

b) Root mean squared error – RMSE

c) Success index – SI

d) Performance index – p^6

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.