On-line algorithm for ground-level ozone prediction with a mobile station

Juš Kocijan[a,b], Dejan Gradišar[a], Marija Zlata Božnar[c], Boštjan Grašič[c], Primož Mlakar[c]

[a] *Jožef Stefan Institute, Jamova cesta 39, SI-1000 Ljubljana, Slovenia*
[b] *University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica, Slovenia*
[c] *MEIS d.o.o., Mali Vrh pri Šmarju 78, SI-1293 Šmarje-Sap, Slovenia*

Abstract

It is important to be able to predict high concentrations of tropospheric ozone and to inform the population about any violations of air-quality standards, as defined by international regulations. Although first-principle models that cover large geographical regions and different atmospheric layers are improving constantly, they typically still only cover geographical regions with a relatively low resolution. Such model predictions can be problematic for the micro-locations of a complex terrain, i.e., a terrain with a large geographical diversity or urban terrain. For such micro-locations, statistical models can be utilised. This paper presents a modelling and prediction algorithm that can be used in, or in accordance with, a mobile air-quality measurement station. Such a mobile station would enable the set-up of a statistical model and a relatively rapid access to the model's predictions for a specific geographical micro-location without a large quantity of historical of measurements. Uncertainty information about the model's predictions is also usually required. In addition, such a model can adapt to long-term changes, such as climate changes. In the paper we propose Gaussian-process models for the described modelling and prediction. In particular, we selected evolving

[☆]This work was supported by the Slovenian Research Agency with Grant Development and Implementation of a Method for On-Line Modelling and Forecasting of Air Pollution, L2-5475 and Grant Systems and Control, P2-0001. The Slovenian Environment Agency provided part of the data.

Gaussian-process models that update on-line with the incoming measurement data. The proposed algorithm for the mobile air-quality measurement and the forecasting station is evaluated on measurements from five locations in Slovenia with different topographical and geographical properties. The obtained evaluation results confirm the feasibility of the concept.

Keywords: Air pollution, Ozone, Prediction of ozone concentration, Mobile air-quality measurement station, Statistical modelling, Gaussian-process model, Evolving model

1. Introduction

An increased ground-level ozone concentration poses a risk to public health, vegetation and materials in a variety of ways. Environmental agencies are interested in providing both the public and experts with air-quality information that can be used for alarm systems as well as to increase public awareness about air quality. The need to analyse and forecast the air quality in Europe has become an obligation under the EU framework Directive on air quality (EU-Commission, 2008). Therefore, predicting the ozone concentration and informing the population when the air-quality standards are not being met are important tasks. The ozone concentration can be predicted with a variety of models (Im et al., 2015), taking into account the topographical and the climatological conditions. However, the forecasting resolution of these models is usually not high enough to account for a complex terrain, e.g., valleys, mountains and the micro-locations of an urban environment. Models obtained directly from measurement data, the so-called statistical or also black-box models, are a viable alternative for insufficiently covered localities. The implementation of a prediction model in a mobile measurement station would enable the setting-up of a statistical model at a specific geographical micro-location and prompt access to the model's predictions after the station's positioning.

Various studies have demonstrated the added value of statistical modelling for the forecasting of regional air quality. In these studies, statistical models for ozone prediction – obtained with a range of linear and nonlinear regression methods from Principal Component Regression to Takagi-Sugeno fuzzy models – are used for different geographical regions and for different objectives of the ozone prediction. It is possible to find models developed with various methods for the prediction of hourly ozone values, e.g., neu-

ral network and principal component regression models in (Al-Alawi et al., 2008), linear regression ARIMA models in (Duenas et al., 2005), neural network and support vector-machine models in (Feng et al., 2011), fuzzy and nonlinear regression models in (Lin and Cobourn, 2007), Gaussian-process models in (Petelin et al., 2013), neural network models in (Solaiman et al., 2008), for the prediction of daily maximum ozone values, e.g., neural network models in (Baawain and Al-Serihi, 2014), support vector machine models in (Chelani, 2010), fuzzy models in (Cheng et al., 2011), neural network models in (Fontes et al., 2014), neural network and Gaussian-process models in (Grašič et al., 2006), linear regression and neural network models in (Moustris et al., 2012), fuzzy models in (Nebot et al., 2008), classification and regression trees in (Sundaramoorthi, 2014), or for the prediction of different average ozone values, e.g., neural network models in (Fontes et al., 2014), ensembles of regression trees in (Garner and Thompson, 2013), hidden Markov models and generalised linear models in (Sun et al., 2013), classification and regression trees in (Sundaramoorthi, 2014), to list only a selection of recent publications. These models use various pollutants and various meteorological variables, together with their lagged values, as the regressors. All these models are developed off-line, i.e., before the prediction takes place, and are not changing with the new data on-line.

Statistical modelling methods can be divided, in general, into parametric methods, where the modelled system is approximated by fitting the parameters of the selected basis functions (Božnar et al., 1993; Mlakar and Božnar, 2011), and into non-parametric models, where the relationships among the measured data are searched directly from the data, e.g., kernel methods. There is considerably less effort necessary for the selection of model's structure for non-parametric models and, in general, you have to optimise only a low number of parameters, if any at all, when modelling non-parametric models. Lu and Wang (2014) make a comparison between a Multilayer-Perceptron Neural Network, as a representative of parametric methods, and Support Vector Machines, as a representative of non-parametric methods proposed for environmental modelers. The drawbacks of both approaches to modelling the ozone concentration are exposed in (Lu and Wang, 2014). These drawbacks are optimisation problems with the local minima, and model overfitting, which can be reduced with a Bayesian approach. On the other hand, non-parametric methods overcome both mentioned optimisation problems, but are computationally intensive in the case of large datasets.

In this paper we propose a modelling and prediction algorithm that can

be used in, or in accordance with, a mobile air-quality measurement station and that overcomes the described problems of statistical methods. A non-parametric Gaussian-process (GP) model is used that circumvents the optimisation problems with local minima and overfitting. In particular, its on-line version is used to avoid the problems associated with large datasets. The GP model has been known for a long time in the field of geostatistics, where the method was named 'kriging' by Krige (1951). A regression problem was first solved with GPs in the late 1970s by O'Hagan (1978). It has become popular within the machine-learning community initially due to the research of Neal (1996), who showed the relationship between GP and neural network models. It continued with the research of Rasmussen (1996), who placed GP modelling within the Bayesian probability framework. The research of many others followed. GP models can be used for modelling of various kinds of models and applied in various domains like (Kocijan, 2016): chemical engineering, biomedical engineering, biological systems, power systems and engineering, etc. Numerous papers, most of which have been published since 2000, describe the use of GP models for modelling of dynamic systems. These publications have explored the use of GP models for various applications like dynamic systems modelling from measurements, e.g., (Kocijan et al., 2005; Shi et al., 2005; Gregorčič and Lightbody, 2007, 2008, 2009) and dynamic systems control, e.g., (Gregorčič and Lightbody, 2012; Kocijan, 2016). The idea of using such models for the prediction of ozone has been initiated in (Petelin et al., 2013), with the comparative evaluation of the models obtained with different statistical methods on data from a single geographical location in (Petelin et al., 2015).

The paper is structured as follows. The problem is described in the next section. The proposed modelling method is introduced in Section 3. Section 4 deals with experiments to show the feasibility of the proposed algorithm, with the results presented in Section 5. The conclusions are drawn at the end of the paper.

2. Problem description

The problem considered in this paper is to find and evaluate a possible algorithm for statistical modelling with an ability to learn on-line from incoming or streaming data measurements. The obtained model should be able to learn from scratch or with a small set of initial data so that it can be used in the context of a mobile air-quality measurement station relatively

quickly after it is deployed in the field. The ability to learn on-line makes such a model also able to cover for seasonal meteorological and climatological changes, as well as changes in geographical position.

The model is aimed at predictions of the daily maximum ozone concentrations one-day ahead and predictions of the maximum 8-hour-averaged ozone concentrations one-day ahead. The daily maximum value is, in our case, defined as the maximum value of the hourly average ozone concentrations obtained between 1 and 24 hours on a particular day. An 8-hour-averaged value at the moment of consideration is, in our case, the moving average of eight, equally-weighted, consequent, hourly average ozone concentrations, including the most recent one. The 8-hour-averaged values from which the maximum for a particular day is determined are presented as moving averages, calculated between 1 and 24 hours on the particular day.

The obtained predictions should contain information about their uncertainty. Also, the predictions of both models for the next day are to be made at 24.00 hour, but this can be arbitrarily modified.

3. Methods

3.1. Gaussian Process Models

GP models are probabilistic, non-parametric models based on the principles of Bayesian probability. GPs actually provide a Bayesian interpretation to the kernel methods (Rasmussen and Williams, 2006). This means that with a GP model we do not try to approximate the modelled system by fitting the parameters of the selected basis functions, but rather we search for the relationship among the measured data. The modelling properties of GP models are reviewed in (Rasmussen and Williams, 2006; Kocijan, 2016; Shi and Choi, 2011; Seeger, 2004; MacKay, 1998).

GP models can be used for regression, where the task is to infer a mapping from a set of N D-dimensional regression vectors represented by the regression matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^T$ to a vector of output data $\mathbf{y} = [y_1, y_2, \ldots, y_N]$. The outputs are usually assumed to be noisy realisations of the underlying function $f(\mathbf{x}_i)$. A GP model assumes that the output is a realisation of a GP with a joint probability density function:

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{m}, \mathbf{K}), \tag{1}$$

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

$$\mathbf{K}_{ij} = C(\mathbf{x}_i, \mathbf{x}_j) \tag{2}$$

defines the characteristics of the process to be modelled, i.e., the stationarity, smoothness, etc. The value of the covariance function $C(\mathbf{x}_i, \mathbf{x}_j)$ expresses the correlation between the individual outputs $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ with respect to the inputs $\mathbf{x}_i$ and $\mathbf{x}_j$. The covariance function can be any function that generates a positive, semi-definite covariance matrix. Assuming the stationary data is contaminated with white noise, the most commonly used covariance function is the composition of the square exponential (SE) covariance function with 'automatic relevance determination' (ARD) hyperparameters (MacKay, 1998) and a constant covariance function assuming white noise:

$$C(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left[-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^{\mathrm{T}} \mathbf{\Lambda}^{-1} (\mathbf{x}_i - \mathbf{x}_j)\right] + \delta_{ij}\sigma_n^2, \tag{3}$$

where $\mathbf{\Lambda}^{-1}$ is a diagonal matrix $\mathbf{\Lambda}^{-1} = \mathrm{diag}([l_1^{-2}, \ldots, l_D^{-2}])$ of the ARD hyperparameters, σ_f^2 and σ_n^2 are hyperparameters of the covariance function, and $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. The hyperparameters can be written as a vector $\boldsymbol{\theta} = [l_1^{-2}, \ldots, l_D^{-2}, \sigma_f^2, \sigma_n^2]^T$. The ARD property means that $l_i^{-2}; i = 1, \ldots, D$ indicates the importance of individual inputs. If l_i^{-2} is zero or near zero, it means that the inputs in dimension i contain only little information and could possibly be discarded. Further covariance functions suitable for various applications can be found in, e.g., (Kocijan, 2016).

The common aim of regression is to predict the output y^* in an unobserved test location $\mathbf{x}^*$ given the training data, a known mean function and a known covariance function C. The posterior predictive distribution can be obtained by constructing the joint posterior distribution using the Bayes' rule. Then, the posterior predictive distribution is obtained by marginalising over the function f. Assuming GP prior with zero mean function leads to a Gaussian predictive distribution (Rasmussen and Williams, 2006). In order to perform a full Bayesian inference, the effect of unknown hyperparameters $\boldsymbol{\theta}$ has to be taken into account. The posterior predictive distribution is described with the following equation

$$p(y|\mathbf{X}) = \int\int p\left(y|\mathbf{f}, \mathbf{X}, \boldsymbol{\theta}\right) p\left(\mathbf{f}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right) d\mathbf{f} d\boldsymbol{\theta}. \tag{4}$$

The computation of such integrals can be difficult due to the intractable nature of the non-linear functions. In the case of GP inference a frequently used approximate solution to the problem of intractable integrals is to estimate the hyperparameters with the maximising of the marginal likelihood from Bayes' rule. This is carried out by minimising the following negative log-likelihood function:

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\log(\mid \mathbf{K} \mid) - \frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} - \frac{N}{2}\log(2\pi), \tag{5}$$

where $\mathbf{K}$ is a covariance matrix with elements described using equation (2).

Once the hyperparameter values are obtained, the predictive normal distribution of the output for a new test input can be calculated using:

$$\mu(y^*) = \mathbf{k}(\mathbf{x}^*)^T\mathbf{K}^{-1}\mathbf{y}, \tag{6}$$

$$\sigma^2(y^*) = \kappa(\mathbf{x}^*) - \mathbf{k}(\mathbf{x}^*)^T\mathbf{K}^{-1}\mathbf{k}(\mathbf{x}^*), \tag{7}$$

where $\mathbf{k}(\mathbf{x}^*) = [C(\mathbf{x}_1, \mathbf{x}^*), \ldots, C(\mathbf{x}_N, \mathbf{x}^*)]^{\mathrm{T}}$ is the $N \times 1$ vector of covariances between the test and the training cases, and $\kappa(x^*) = C(\mathbf{x}^*, \mathbf{x}^*)$ is the covariance between the test input itself.

A prediction of the GP model, in addition to the mean value (6), also provides information about the confidence of the prediction using the prediction variance (7). Usually, the confidence in the prediction is interpreted with a 2σ interval, which corresponds to about 95% of the confidence interval. The confidence interval highlights the 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.

A known drawback of GP modelling with a large training dataset is the computational load that increases with the third power of the amount of input data due to the calculation of the inverse of the covariance matrix. To overcome the computational-limitation issues and consequently to make the method viable for large-scale dataset applications, various sparse-approximation methods have been suggested. A common property of all these sparse-approximation methods is that they try 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. Usually, this subset of the training data is called the active set. For more details see (Quinonero-Candela et al., 2007). The majority of the methods for the active set selection are off-line methods, which means

they need all the training data available at once. There are a few on-line sparse methods for GP modelling, which adapts the active set sequentially, e.g. (Csató and Opper, 2002; Seeger et al., 2003; Ranganathan et al., 2011). However, they do not adjust the hyperparameter values in the on-line mode or have some other limitations, e.g., they have possible computational issues, etc. The subsequently described method overcomes these limitations.

3.2. Evolving Gaussian process modelling

The Evolving GP (EGP) model is inspired by Evolving Systems (Angelov et al., 2010), which are self-developing systems, adapting on-line both the structure and parameter values of the model from incoming data. The term Evolving GP models is used in the sense of the sequential adapting of elements of the GP model, including the hyperparameter values.

The EGP processes every new piece of data sequentially and adapts all the influential parts of the GP model in an on-line fashion. This enables the fast and efficient adaptation of the GP model to changes. The EGP concept was proposed in (Petelin and Kocijan, 2011) and further developed in (Petelin et al., 2013). This concept considers the adaptation of four main elements of the GP model: the active set, the hyperparameter values, the covariance function and the regressors. To simplify the concept we decided, like with (Petelin et al., 2013), to use the fixed covariance function SE with ARD as we assume the smoothness and stationarity of the system. The ARD functionality is able to find influential regressors. With the optimization of the hyperparameter values, the uninfluential regressors have a smaller weight and, as a consequence, have a smaller influence on the result. Therefore, all the available regressors can be used and, consequently, only the active set and the hyperparameter values are adapted sequentially.

The proposed method consists of roughly three main steps to adapt the GP model sequentially. In the first step the new data is processed in the sense of including the incoming data in the active set $\mathcal{X}$. In the following step the hyperparameter values $\boldsymbol{\theta}$ are optimized, while in the last step the covariance matrix $\mathbf{K}$ and its inversion $\mathbf{K}^{-1}$ are updated in accordance with the changes from the first two steps.

Processing incoming data. For every new piece of incoming data, first the novelty of the data is verified by predicting the output based on the incoming data and scoring the novelty based on the prediction's mean and variance. If

either of these two values is above the pre-set thresholds, the incoming data is included in the active set.

If the inclusion of the incoming data causes the exceeding of the the pre-set maximum size of the active set, the least informative data is excluded. The least informative data is scored according to the Euclid distance between the regression vectors, extended with the corresponding model prediction, combined with the exponential forgetting factor. The Euclid distance among the data is calculated and the datapoint with the smallest distance is eliminated. Thus, the most diverse, i.e., the most informative, data is preserved. The exponential forgetting factor is implemented as the multiplier of the distance-based score.

Optimising hyperparameter values. After updating the active set, the hyperparameter values should be re-optimised as described beforehand by minimising the negative log-likelihood function (5). The optimisation in each EGP processing step has a limited number of iterations. To avoid the situation when the optimisation becomes stuck in a local optimum, it is begun with the random hyperparameter values and the best set of hyperparameter values is selected. These steps are repeated until the EGP processing is completed.

Updating covariance matrix. At the end, the inverse of the covariance matrix is updated according to the new active set $\mathcal{X}$ and the new hyperparameter values $\boldsymbol{\theta}$. It is updated twice in two steps. In the first step the low rank updates for the Cholesky decomposition (Seeger, 2008) are used as only one rank of the covariance matrix is changed when calculating the marginal likelihood for each subset of the active set. These updates cannot be used in the second step as the whole covariance matrix is changed in the case of new hyperparameter values, so the inverse of the covariance matrix is updated again.

The algorithm is given in Figure 1 (Petelin and Kocijan, 2014).

4. Experiments

The algorithm for a mobile station will be tuned in such a way that the mobile station can be placed in different locations in Slovenia, which is very diverse in terms of geography and climate. It is expected that the algorithm would behave comparably well in different environments. Consequently, it was tested at five selected locations with the different properties shown in Figure 2. Nova Gorica has a Mediterranean climate with a strong influence

```
procedure EGP
    global: 𝒳, K, Q
    Q ← K^{-1}
    repeat
        x*, y* ← GetIncomingData
        𝒳⁺, K⁺, Q⁺ ← add(x*, y*)
        if length(𝒳⁺) > max_𝒳 then
            Scores ← CalculateScores(𝒳⁺, K⁺, Q⁺)
            𝒳, K, Q ← RemoveWorst(Scores)
        end if
        θ, K, Q ← OptimizeHyperparameters(θ, K, Q)
    until incoming data available
end procedure
```

Figure 1: Pseudo code of the EGP method.

from the river Po and the industrial Friuli region in Italy. Koper is an industrial and port town on the Adriatic coast with a Mediterranean climate. Ljubljana is the most populated region in Slovenia and has an unfavourable geographical location in a wider basin with a continental climate, where industrial air pollution is combined with the air pollution from traffic and domestic heating. Similar characteristics apply for the Celje region. A different situation applies for Zagorje, located in a region with highly complex orography and consequently very complex micrometeorological conditions of a generally continental climate, which means that the polluted industrial air stays trapped in a basin and causes problems for the inhabitants.

The meteorological and air-quality variables at these locations are measured every half an hour and are stored in an internal database. The measured data were acquired for all the available variables, as listed in Table 1, for each location for a period of 3 years (from the beginning of 2012 to the end of 2014).

Based on the collected half-hour measurements for each variable, their 1-hour and 8-hour averages are calculated. Only the maximum daily values are selected as the regressors used for both, i.e., the 1- and 8-hour, ozone-prediction model. The wind-direction measurements ($WindDir$) are not calculated as an 8-hour average.

As the ozone concentration depends on the present, and not only on the

Figure 2: Geographical location of the selected locations.

past, conditions, the forecasts of the variables were added, as is common practise in this type of investigation. To avoid the forecasts' uncertainty we applied the measurements of these variables, which, in our opinion, provides a more accurate picture of the regressors' relevance. Therefore, no numerical forecasts from meteorological and air-quality models are used for regressors in this paper. The use of numerical forecasts from other models as regressors is certainly a feasible option, but it is beyond the scope of this paper.

With a number of available variables and its lagged values, the size of the regression vector and, consequently, of the model, increases noticeably. For this reason it is only necessary to select the regressors that add the most information to the prediction. Various regressor-selection methods are available. Kocijan et al. (2015) used a method-selection strategy where various methods were tested, and the one with best result was used. Glavan et al. (2013) suggested the use of various algorithms, and the resulting regression selection is achieved based on the average-weighted method. To support the regressor-selection process, Gradišar et al. (2015) built up a ProOpter platform, in which various regressor-selection algorithms are implemented. In

Table 1: Available variables' measurements.

Nova Gorica	Koper	Ljubljana	Celje	Zagorje
$O3$	$O3$	$O3$	$O3$	$O3$
$GlSolRad$	$GlSolRad$	$GlSolRad$	$GlSolRad$	/
$AirTemp$	$AirTemp$	$AirTemp$	$AirTemp$	$AirTemp$
$RelHum$	$RelHum$	$RelHum$	$RelHum$	$RelHum$
$WindSpd$	$WindSpd$	$WindSpd$	/	$WindSpd$
$WindDir$	$WindDir$	$WindDir$	/	$WindDir$
NOx	NOx	NOx	NOx	NOx
$NO2$	$NO2$	$NO2$	$NO2$	$NO2$
/	/	$SO2$	$SO2$	$SO2$
$Dust$	$Dust$	$Dust$	$Dust$	$Dust$
$Precip$	$Precip$	$Precip$	$Precip$	$Precip$
$DifSolRad$	$DifSolRad$	$DifSolRad$	$DifSolRad$	/
$Pressure$	$Pressure$	/	/	/
/	/	CO	/	/

this paper we have applied (i) the distance Correlation – dCorr, (ii) the Partial Mutual Information – PMI and (iii) the model Linear in the Parameters – LIP algorithms to rank the relevance of the regressors. Later, the results were grouped in two stages. In the first stage, the rankings based on statistical measures achieved with all three mentioned methods were averaged for every location. In the subsequent stage, the rankings from the first stage (for every location) were averaged (over locations) again to obtain the final sequence of regressors, ordered in terms of their importance. This procedure made it possible to obtain an averaged set of regressors that encompassed the significant regressors for all the involved locations. The rationale behind this task is to obtain a single uniform regression vector for a lager area, in our case Slovenia, and to avoid a regressor selection every time the mobile station is moved around. The regressors that have measurements only for a specific location were omitted from the general set of regressors.

In the next stage, we determine how many of regressors, out of those selected in the first stage, should be used in order to produce the best prediction.

A Gaussian-process model was used to obtain the best set of regressors, where SE with the ARD covariance function and a constant mean value was

applied. We used all the available data and divided them into 11 subsets. A 10-fold cross-validation was used for the prediction validation on the 10 subsets, while the remaining, larger subset is used for testing the prediction. To score the performance of the models we chose the following statistical measures:

- The root-mean-square error - RMSE:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(E(\hat{y}_i) - y_i)^2}, \tag{8}$$

  where y_i and $\hat{y}_i$ are the observation and the prediction in the i-th step, respectively, $E(\cdot)$ denotes the expectation, i.e., the mean value, of the random variable, and N is the number of used observations.

- The standardised mean-squared error - SMSE (Rasmussen and Williams, 2006):

$$\mathrm{SMSE} = \frac{1}{N}\frac{\sum_{i=1}^{N}(E(\hat{y}_i) - y_i)^2}{\sigma_y^2}, \tag{9}$$

  where σ_y^2 is the variance of the observations.

- The mean standardised log loss - MSLL (Rasmussen and Williams, 2006):

$$\begin{aligned}\mathrm{MSLL} &= \frac{1}{2N}\sum_{i=1}^{N}\left[\log(\sigma_i^2) + \frac{(E(\hat{y}_i) - y_i)^2}{\sigma_i^2}\right] \\ &- \frac{1}{2N}\sum_{i=1}^{N}\left[\log(\sigma_y^2) + \frac{(y_i - E(\mathbf{y}))^2}{\sigma_y^2}\right],\end{aligned} \tag{10}$$

  where σ_i^2 is the prediction variance in the i-th step, and $E(\mathbf{y})$ is the expectation, i.e., the mean value, of the vector of the observations.

- The Pearson's correlation coefficient - PCC:

$$\mathrm{PCC} = \frac{\sum_{i=1}^{N}(E(\hat{y}_i) - E(\hat{\mathbf{y}}))(y_i - E(\mathbf{y}))}{N\sigma_y\sigma_{\hat{y}}}, \tag{11}$$

  where $E(\hat{\mathbf{y}})$ is the expectation, i.e., the mean value, of the vector of predictions, and σ_y,$\sigma_{\hat{y}}$ are the standard deviations of the observations and the predictions, respectively.

- The mean fractional bias - MFB:

$$\mathrm{MFB} = \frac{1}{N}\sum_{i=1}^{N}\frac{E(\hat{y}_i) - y_i}{\frac{1}{2}(E(\hat{y}_i) + y_i)}. \tag{12}$$

- The factor of the modelled values within a factor of two of the observations - FAC2:

$$\mathrm{FAC2} = \frac{1}{N}\sum_{i=1}^{N} n_i \quad \text{with} \quad n_i = \begin{cases} 1 & \text{for } 0.5 \leq |\frac{E(\hat{y}_i)}{y_i}| \leq 2, \\ 0 & \text{else.} \end{cases} \tag{13}$$

RMSE and SMSE are frequently used measures for the accuracy of the predictions' mean values, which are 0 in the case of perfect model. SMSE is the standardised measure with values between 0 and 1. MSLL is a standardised measure suited to predictions in the form of random variables. It weights the prediction error more heavily when it is accompanied by a smaller prediction variance. The MSLL is approximately zero for the simple models and negative for the better ones. PCC is a measure of associativity and is not sensitive to bias. Its value is between -1 and +1, with ideally linearly correlated values resulting in a value 1. MFB is the measure that bounds the maximum bias and gives additional weight to underestimations and less weight to overestimations. Its value is between -2 and +2, with the value 0 in the case of a perfect model. FAC2 indicates the fraction of the data that satisfies the condition from Equation (13). Its value is between 0 and 1, with the perfect model resulting in a value of 1.

The final ranking of regressors was obtained based on statistical measures calculated for different models' predictions with 10-fold cross-validation. The models for this stage were built from the set of potential regressors determined in the first stage.

The first nine regressors from the final ranking give the best results, on average, for all the locations and measures, and are presented in Table 2, separately for *Daily maximum* and for the maximum of *8-hour averages*. Again, the one-hour-averaged daily maximum values of the regressors and the 8-hour-averaged daily maximum values of the regressors are used for the daily maximum and the 8-hour-averaged daily maximum predictions, respectively. If there are no measured values for some of the variables in the final selection at a particular location, the prediction is made without them.

Table 2: Regressors for the final models. k denotes consecutive time instants, where $k+1$ means the present-day maximum values, i.e., at the prediction time, and k the recent maximum values, i.e., from yesterday. Instead of forecasts, measurements are used for the present-day values, as explained earlier in the text.

	Daily maximum	8-hour average
1	$O3(k)$	$O3(k)$
2	$GlSolRad(k+1)$	$GlSolRad(k+1)$
3	$AirTemp(k+1)$	$AirTemp(k+1)$
4	$AirTemp(k)$	$AirTemp(k)$
5	$GlSolRad(k)$	$NOx(k+1)$
6	$RelHum(k+1)$	$RelHum(k+1)$
7	$NOx(k+1)$	$Pressure(k)$
8	$Pressure(k+1)$	$Pressure(k+1)$
9	$Pressure(k)$	$GlSolRad(k)$

Finally, we used selected regressors for a one-day prediction of the ozone based on the EGP. First, 60 days are needed to initialise the GP model. After that, the evolving GP model is applied. Also, in this case we applied SE with the ARD covariance function and a constant mean value. The number of optimisation iterations in each EGP step is limited to 70 and the maximum size of the active set is determined as 90 datapoints. Lower values for any of these design parameters resulted in a worse prediction model, while higher values increased the computational time for each iteration step.

The information gain is defined by the maximum Euclid distance of the i-th element to any other. Exponential forgetting, with a forgetting factor of 0.9995, was determined empirically.

5. Results

This section provides the results for the on-line model predictions. As already mentioned, we expect from the developed model approximately the same prediction accuracy at all the considered locations.

The results of the performance measures for the test data are given in Table 3 for daily maximum concentrations and in Table 4 for the maximum 8-hour-averaged concentrations.

Comparable values of the performance measures for all the considered locations with very different geographical and meteorological conditions are observed. This confirms the adaptivity of the selected algorithm. The same can

Table 3: Performance measures at different locations for predictions of the daily maximum concentrations.

Performance measure	Nova Gorica	Koper	Ljubljana	Celje	Zagorje
RMSE	17.456	14.616	15.028	16.678	14.962
SMSE	0.19	0.19	0.16	0.21	0.21
MSLL	-0.81	-0.79	-0.86	-0.76	-0.74
PCC	0.904	0.900	0.918	0.890	0.898
MFB	0.069	0.025	0.057	0.082	0.108
FAC2	0.95	0.98	0.93	0.93	0.93

Table 4: Performance measures at different locations for predictions of the maximum 8-hour-averaged concentrations.

Performance measure	Nova Gorica	Koper	Ljubljana	Celje	Zagorje
RMSE	15.490	13.406	14.915	15.940	14.316
SMSE	0.16	0.17	0.16	0.20	0.20
MSLL	-0.91	-0.85	-0.88	-0.80	-0.78
PCC	0.927	0.914	0.924	0.900	0.901
MFB	0.110	0.033	0.159	0.105	0.121
FAC2	0.93	0.98	0.89	0.91	0.91

be concluded for the daily maximum and for the maximum 8-hour-averaged predictions of the ozone concentrations.

A visual presentation of the model's predictions with scatter plots and time responses is given in Figures 3 - 8 for one of the considered locations, i.e, for Ljubljana, which is the capital city of Slovenia.

It is clear from Figures 3 - 8 that the model predictions follow the values of the test measurements. Moreover, most of the measurements are contained within the 95 % confidence interval provided by the model.

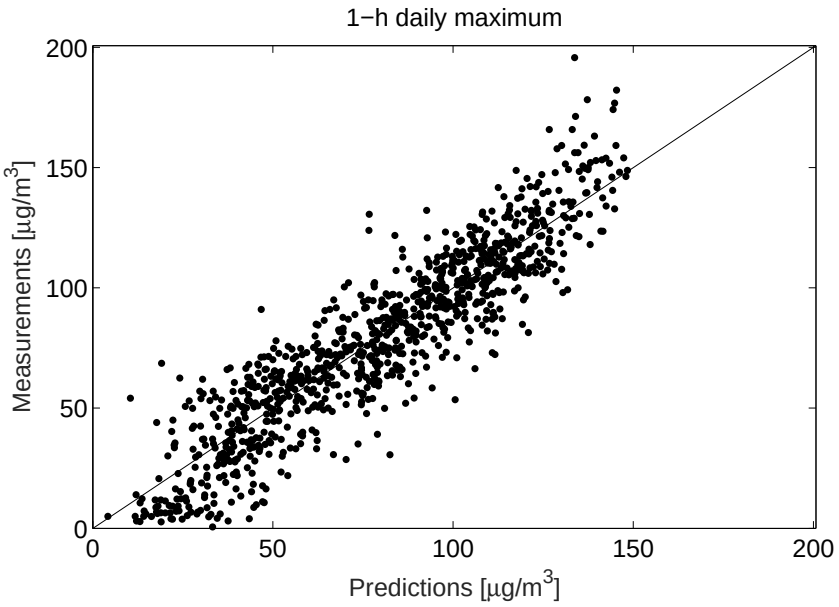

Figure 3: Predicted values versus observation values for 1-h daily maximum ozone concentrations for Ljubljana.

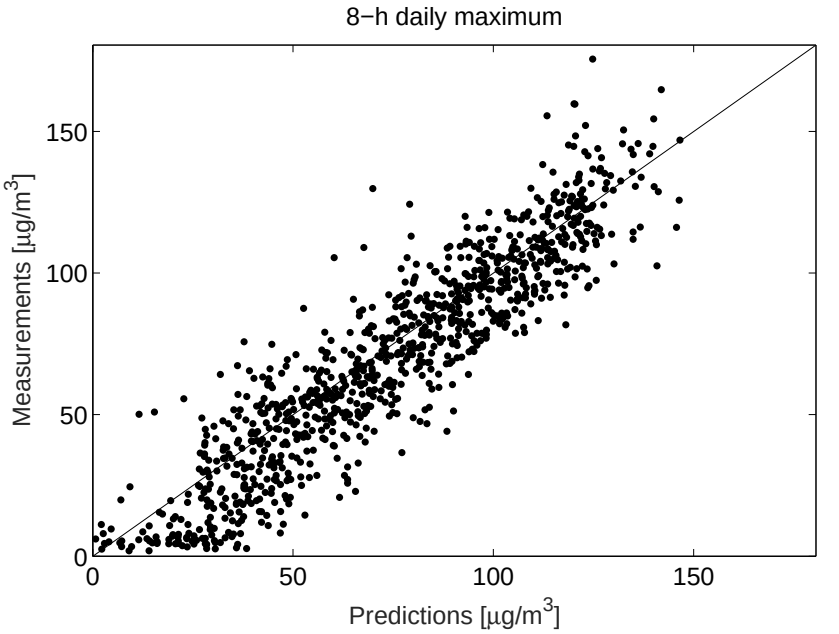

Figure 4: Predicted values versus observation values for maximum 8-hour-averaged ozone concentrations for Ljubljana.

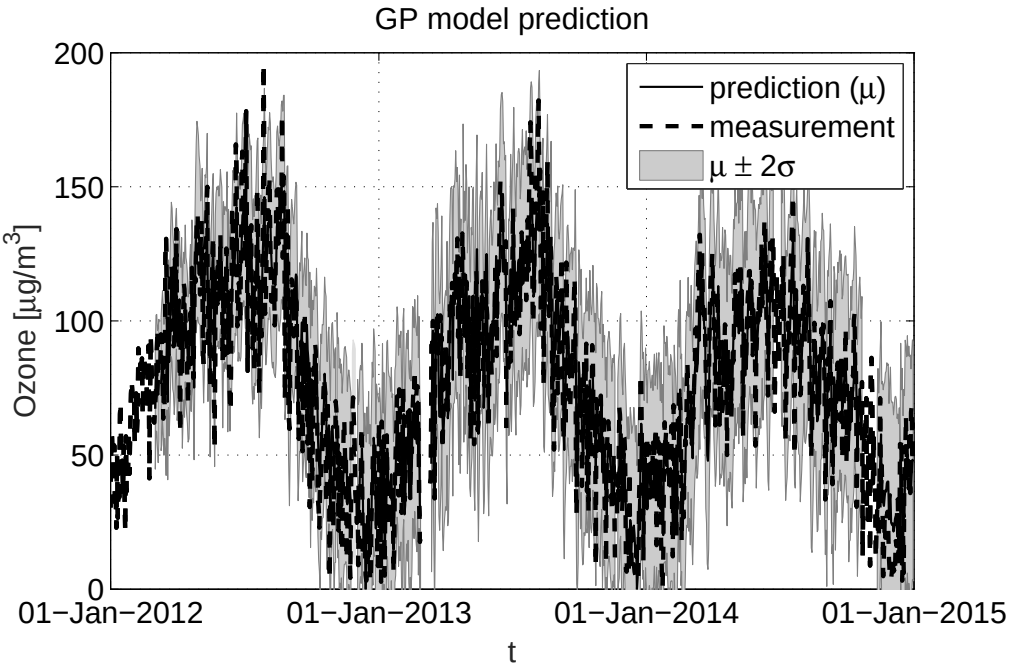

Figure 5: Time-series plot of predictions for 1-h daily maximum ozone concentrations for Ljubljana.

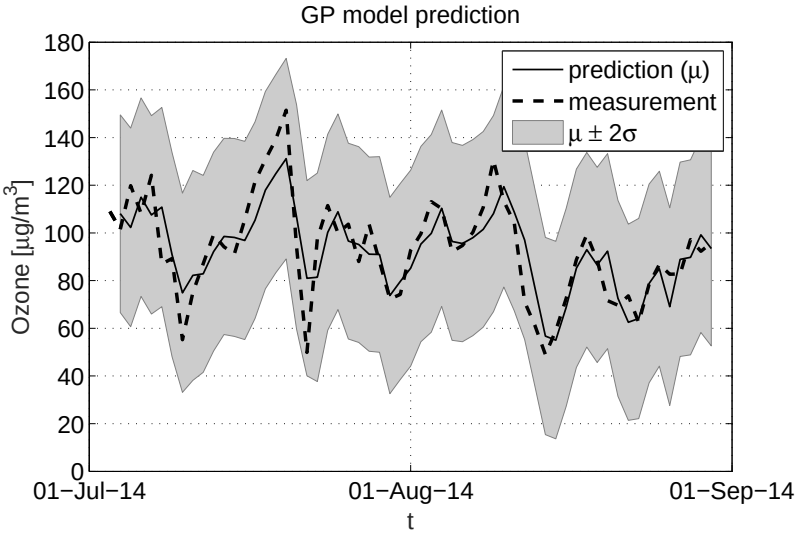

Figure 6: Zoomed part of Figure 5.

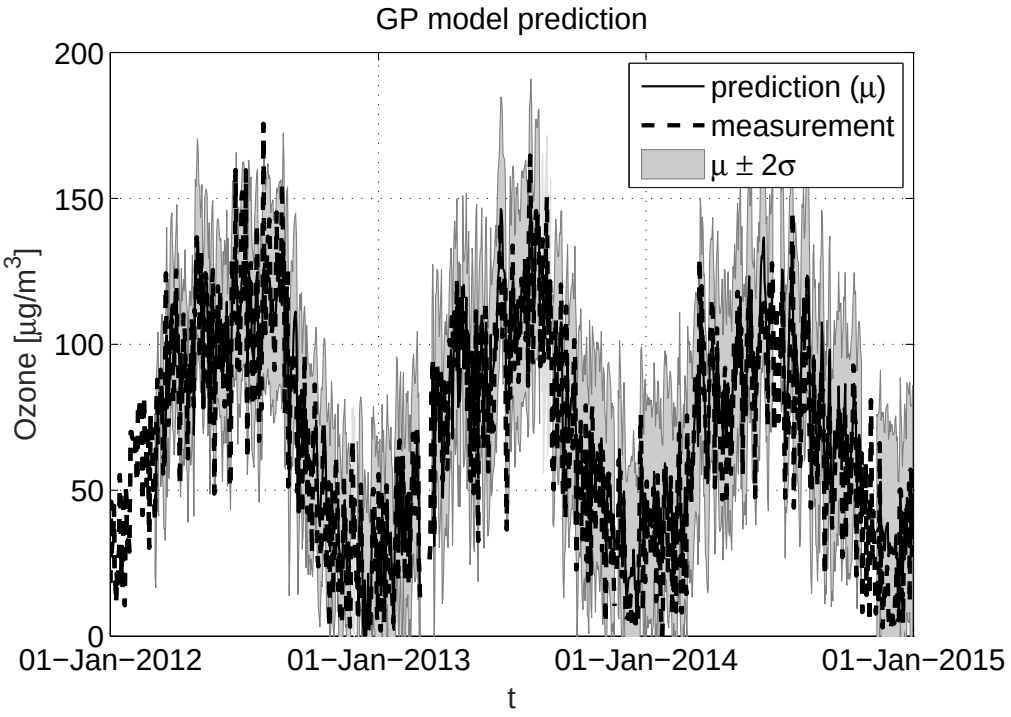

Figure 7: Time-series plot of predictions for maximum 8-hour-averaged ozone concentrations for Ljubljana.

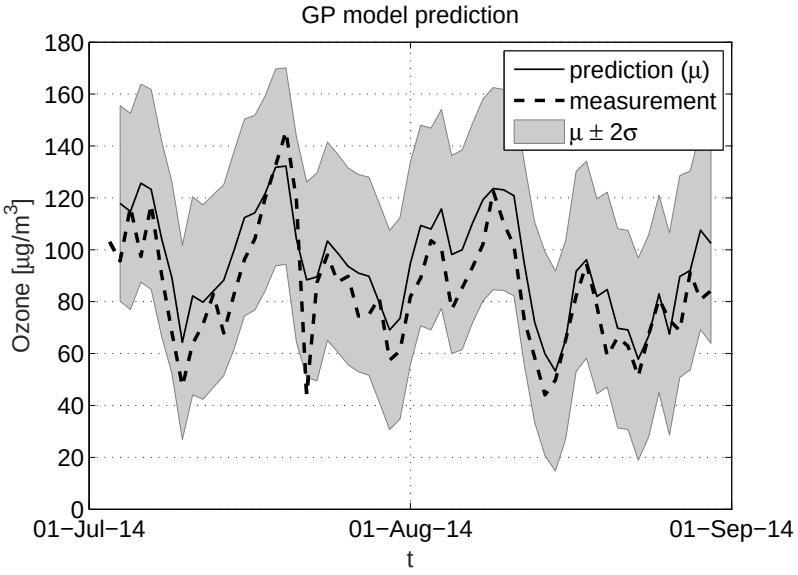

Figure 8: Zoomed part of Figure 7.

6. Conclusions

In this paper we describe an algorithm for on-line statistical modelling that could be used for one-day-ahead predictions of the daily maximum ozone concentrations and the maximum 8-hour-averaged ozone concentrations. The algorithm is designed for a mobile air-quality-measurements station that could be used relatively quickly after being deployed in the field.

The EGP algorithm fulfilled the given objective and the described tests showed that the algorithm behaved comparably well using the data from five different locations in Slovenia with different geographical and meteorological properties. The selection of the common regression vector for all the considered locations enabled a procedure for the regressors' selection that does not need to be repeated every time the station is deployed in the field.

Various field tests, algorithm alternatives and potential improvements of the algorithm are foreseen in the future. Moreover the possibility to include numerical forecasts from meteorological and air-quality models as regressors and the influence of their uncertainty will be investigated.

References

Al-Alawi, S. M., Abdul-Wahab, S. A., Bakheit, C. S., 2008. Combining principal component regression and artificial neural-networks for more accurate predictions of ground-level ozone. Environmental Modelling & Software 23, 396–403.

Angelov, P., Filev, D. P., Kasabov, N., April 2010. Evolving Intelligent Systems: Methodology and Applications. IEEE Press Series on Computational Intelligence. Wiley-IEEE Press.

Baawain, M. S., Al-Serihi, A. S., February 2014. Systematic approach for the prediction of ground-level air pollution (around an industrial port) using an artificial neural network. Aerosol and air quality research 14 (1), 124–134.

Božnar, M., Lesjak, M., Mlakar, P., 1993. A neural network-based method for short-term predictions of ambient so 2 concentrations in highly polluted industrial areas of complex terrain. Atmospheric Environment. Part B. Urban Atmosphere 27 (2), 221–230.

Chelani, A. B., 2010. Prediction of daily maximum ground ozone concentration using support vector machine. Environ Monit Assess 162, 169–176.

Cheng, C.-H., Huang, S.-F., Teoh, H.-J., 2011. Predicting daily ozone concentration maxima using fuzzy time series based on a two-stage linguistic partition method. Computers and Mathematics with Applications, 2016–2028.

Csató, L., Opper, M., 2002. Sparse Online Gaussian Processes. Neural Computation 14 (3), 641–668.

Duenas, C., Fernandez, M., Canete, S., Carretero, J., Liger, E., 2005. Stochastic model to forecast ground-level ozone concentration at urban and rural areas. Chemosphere 61, 1379–1389.

EU-Commission, 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. Official Journal of the European Communities L 152, 1–44.
URL http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:152:0001:0044:EN:PDF

Feng, Y., Zhang, W., Sun, D., Zhang, L., 2011. Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and SVM data classification. Atmospheric Environment 45, 1979–1985.

Fontes, T., Silva, L., Silva, M., Barros, N., Carvalho, A., 2014. Can artificial neural networks be used to predict the origin of ozone episodes? Science of the Total Environment 488–489, 197–207.

Garner, G. G., Thompson, A. M., 2013. Ensemble statistical post-processing of the national air quality forecast capability: Enhancing ozone forecasts in Baltimore, Maryland. Atmospheric Environment 81, 517–522.

Glavan, M., Gradišar, D., Atanasijević-Kunc, M., Strmčnik, S., Mušič, G., 2013. Input variable selection for model-based production control and optimisation. The Int. Journal of Advanced Manufacturing Technology 68 (9–12), 2743–2759.

Gradišar, D., Glavan, M., Strmčnik, S., Mušič, G., Jun. 2015. Proopter. Computers in Industry 70 (C), 102–115.

Grašič, B., Mlakar, P., Božnar, M., 2006. Ozone prediction based on neural networks and Gaussian processes. Nuovo cimento Soc. ital. fis., C Geophys. space phys. 29 (6), 651–661.

Gregorčič, G., Lightbody, G., 2007. Local model identification with Gaussian processes. IEEE Transactions on neural networks 18 (5), 1404–1423.

Gregorčič, G., Lightbody, G., 2008. Nonlinear system identification: From multiple-model networks to Gaussian processes. Engineering Applications of Artificial Intelligence 21 (7), 1035–1055.

Gregorčič, G., Lightbody, G., 2009. Gaussian process approach for modelling of nonlinear systems. Engineering Applications of Artificial Intelligence 22 (4-5), 522–533.

Gregorčič, G., Lightbody, G., 2012. Gaussian process internal model control. International Journal of Systems Science 43 (11), 2079–2094.

Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., Bar, R., Bellasio, R., Brunner, D., Chemel, C., Curci, G., Flemming, J., Forkel, R., Giordano, L., Jimnez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Jorba, O., Knote, C., Kuenen, J. J., Makar, P. A., Manders-Groot, A., Neal, L., Prez, J. L., Pirovano, G., Pouliot, G., Jose, R. S., Savage, N., Schroder, W., Sokhi, R. S., Syrakov, D., Torian, A., Tuccella, P., Werhahn, J., Wolke, R., Yahya, K., Zabkar, R., Zhang, Y., Zhang, J., Hogrefe, C., Galmarini, S., 2015. Evaluation of operational on-line-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone. Atmospheric Environment 115, 404–420.

Kocijan, J., 2016. Modelling and Control of Dynamic Systems Using Gaussian Process Models. Springer International Publishing, Cham.

Kocijan, J., Girard, A., Banko, B., Murray-Smith, R., 2005. Dynamic systems identification with Gaussian processes. Mathematical and Computer Modelling of Dynamic Systems 11 (4), 411–424.

Kocijan, J., Hančič, M., Petelin, D., Božnar, M. Z., Mlakar, P., 2015. Regressor selection for ozone prediction. Simulation Modelling Practice and Theory 54, 101–115.

Krige, D. G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of Chemistry, Metal. and Mining Soc. of South Africa 52 (6), 119–139.

Lin, Y., Cobourn, W. G., 2007. Fuzzy system models combined with nonlinear regression for daily ground-level ozone predictions. Atmospheric Environment 41, 3502–3513.

Lu, W.-Z., Wang, D., 2014. Learning machines: Rationale and application in ground-level ozone prediction. Applied Soft Computing 24, 135 – 141.

MacKay, D. J. C., 1998. Introduction to Gaussian processes. NATO ASI Series 168, 133–166.

Mlakar, P., Božnar, M. Z., 2011. Advanced air pollution. InTech, Rijeka, Ch. Artificial neural networks: a useful tool in air pollution and meteorological modelling, pp. 495–508.

Moustris, K. P., Nastos, P. T., Larissi, I. K., Paliatsos, A. G., 2012. Application of multiple linear regression models and artificial neural networks on the surface ozone forecast in the greater Athens area, Greece. Advances in Meteorology 2012, 1–8.

Neal, R. M., 1996. Bayesian learning for neural networks. Vol. 118 of Lecture Notes in Statistics. Springer-Verlag, New York, NY.

Nebot, A., Mugica, V., Escobet, A., 2008. Ozone prediction based on meteorological variables: a fuzzy inductive reasoning approach. Atmospheric Chemistry and Physics Discussions 8, 12343–12370.

O'Hagan, A., 1978. On curve fitting and optimal design for regression (with discussion). Journal of the Royal Statistical Society. Series B (Methodological) 40 (1), 1–42.

Petelin, D., Grancharova, A., Kocijan, J., 2013. Evolving Gaussian process models for the prediction of ozone concentration in the air. Simulation Modelling Practice and Theory 33 (1), 68–80.

Petelin, D., Kocijan, J., 2011. Control system with evolving Gaussian process model. In: Proceedings of IEEE Symposium Series on Computational Intelligence, SSCI 2011. IEEE, Paris.

Petelin, D., Kocijan, J., June 2014. Evolving Gaussian process models for predicting chaotic time-series. In: Evolving and Adaptive Intelligent Systems (EAIS), 2014 IEEE Conference on. pp. 1–8.

Petelin, D., Mlakar, P., Božnar, M. Z., Grašič, B., Kocijan, J., 2015. Ozone forecasting using an on-line updating Gaussian-process model. International Journal of Environment and Pollution 57 (3/4), 111–122.

Quinonero-Candela, J., Rasmussen, C. E., Williams., C. K. I., September 2007. Large-Scale Kernel Machines. Neural Information Processing. The MIT Press, Cambridge, MA, USA, Ch. Approximation methods for Gaussian process regression, pp. 203–223.

Ranganathan, A., Yang, M.-H., Ho, J., 2011. Online Sparse Gaussian Process Regression and Its Applications. IEEE Transactions on Image Processing 20 (2), 391–404.

Rasmussen, C. E., 1996. Evaluation of Gaussian processes and other methods for nonlinear regression. Ph.D. thesis, University of Toronto, Toronto.

Rasmussen, C. E., Williams, C. K. I., 2006. Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA.

Seeger, M., 2004. Gaussian processes for machine learning. International Journal of Neural Systems 14, 2004.

Seeger, M., 2008. Low Rank Updates for the Cholesky Decomposition. Tech. rep., University of California at Berkeley.

Seeger, M., Williams, C. K. I., Lawrence, N. D., 2003. Fast Forward Selection to Speed Up Sparse Gaussian Process Regression. In: Ninth International Workshop on Artificial Intelligence and Statistics. Society for Artificial Intelligence and Statistics.

Shi, J. Q., Choi, T., 2011. Gaussian process regression analysis for functional data. Chapman and Hall/CRC, Taylor & Francis group, Boca Raton, FL.

Shi, J. Q., Murray-Smith, R., Titterington, D. M., 2005. Hierarchical Gaussian process mixtures for regression. Statistics and Computing 15 (1), 31–41.

Solaiman, T. A., Coulibaly, P., Kanaroglou, P., 2008. Ground-level ozone forecasting using data-driven methods. Air Quality, Atmosphere & Health 1, 179–193.

Sun, W., Zhang, H., Palazoglu, A., 2013. Prediction of 8 h-average ozone concentration using a supervised hidden Markov model combined with generalized linear models. Atmospheric Environment 81, 199–208.

Sundaramoorthi, D., 2014. A data-integrated simulation model to forecast ground-level ozone concentration. Annals of Operations Research 216, 53–69.