CAAI Transactions on Intelligence Technology

Research Article

Sparse and hybrid modelling of relative humidity: the Krško basin case study

ISSN 2468-2322
Received on 3rd September 2019
Revised on 3rd September 2019
Accepted on 13th December 2019
doi: 10.1049/trit.2019.0054
www.ietdl.org

Juš Kocijan[1,2] ✉, Matija Perne[1], Boštjan Grašic[3], Marija Zlata Božnar[3], Primož Mlakar[3]

[1]Department of Systems and Control, Jozef Stefan Institute, Jamova 39, Ljubljana, Slovenia
[2]Centre for Information Technologies and Applied Mathematics, University of Nova Gorica, Vipavska 13, Nova Gorica, Slovenia
[3]MEIS d.o.o., Mali Vrh pri Šmarju 78, Šmarje-Sap, Slovenia
✉ E-mail: jus.kocijan@ijs.si

Abstract: This study describes an application of hybrid modelling for an atmospheric variable in the Krško basin. The hybrid model is a combination of a physics-based and data-driven model and has some properties of both modelling approaches. In the authors' case, it is used for the modelling of an atmospheric variable, namely relative humidity in a particular location for the purpose of using the predictions of the model as an input to the air-pollution-dispersion model for radiation exposure. The presented hybrid model is a combination of a physics-based atmospherical model and a Gaussian-process (GP) regression model. The GP model is a probabilistic kernel method that also enables evaluation of prediction confidence. The problem of poor scalability of GP modelling was solved using sparse GP modelling; in particular, the fully independent training conditional method was used. Two different approaches to dataset selection for empirical model training were used and multiple-step-ahead predictions for different horizons were assessed. It is shown in this study that the accuracy of the predicted relative humidity in the Krško basin improved when using hybrid models over using the physics-based model alone and that predictions for a considerable length of horizon can be used.

1 Introduction

Hybrid models combine properties of different kinds of models. In this investigation, the combination of a physics-based model and data-driven model is applied for the modelling of relative humidity in a pre-defined geographical location.

The problem motivating our investigation originates from the need to model and predict the dispersion of radioactive pollutants that could be hypothetically emitted from Krško nuclear power plant (NPP) in Slovenia. Predictions of radiation exposure produced by the air-pollution-dispersion model are necessary for directing the evacuation of inhabitants in the vicinity of the NPP in case of a hazardous event. The air-pollution-dispersion model, which is not the topic of our investigation, relies on information from weather-variables forecasts, which act as the dispersion model inputs. The accuracy of weather-variables forecasts in the range of up to a few hours have a direct effect on the accuracy of the predicted pollutant dispersion [1]. Physics-based numerical weather prediction (NWP) models [2] that are used for weather-variables forecasts refined from global prognostic models and prone to some limitations. When the terrain of interest is complex, standard physics-based models do not provide accurate information enough for further processing with the air-pollution-dispersion model for radiation exposure.

Data-driven or statistical models provide a viable alternative to physics-based models in atmospheric sciences, e.g. [3]. Data-driven modelling of relative humidity in particular can be found in, e.g. [4–7], where a single method or combination of methods are used for relative-humidity modelling. However, the need for the transparency of the physical and chemical background that comes with the physics-based model is apparent. This is the reason that in our study we do not replace the insufficient physics-based model, but rather upgrade it to a hybrid model by combining it with a data-driven model.

Improving the accuracy of weather-variables prediction with statistical modelling and machine-learning methods is the prime motivation of the investigation. Weather is a complex system and each variable can be treated in a slightly different way. Among the many variables with a higher or lower impact on pollution dispersion, our investigation focuses on relative humidity near ground level or more precisely 2 m above ground at the location of Krško NPP.

It is necessary to stress that the investigation described hereafter is not confined to relative humidity only, but can be used for modelling other variables. Furthermore, it is not confined to air-pollution dispersion for radiation exposure only, but can be implemented for improvement of any kind of other point-source emission dispersion that depends on weather and atmospheric variables. Moreover, the investigation method is, in general, not confined to the geographical region of investigation, but can be applied to any other complex terrain.

The aim of this paper is to use the obtained observations from the physics-based model and to process them further with a statistical method to improve the multi-step-ahead predictions of relative humidity. We call this combination of a physics-based and a data-driven model a hybrid model, which will be used for the prediction of relative humidity. Two different approaches of using observations in the multi-step-ahead prediction are investigated. The first approach trains statistical model parameters using measurements as inputs, and then the inputs are changed with physics-based-model forecasts for multi-step-ahead predictions when future measurements are not available. The second approach trains statistical model parameters using physics-based-model predictions from scratch. The obtained results are compared with results of alternative prediction possibilities.

The problems envisaged in the investigation originate from the need to keep the physical transparency of the relative-humidity model in addition to the fact that we are dealing with large amounts of data, and consequently, need a scalable statistical model. The desire for physical transparency prevents the direct use of advanced machine-learning methods and the development of a statistical model only. The problem of relatively large amounts of data will be tackled with the use of sparse methods and high performance computing facilities.

Hybrid modelling is a known method for improving the results of physics-based modelling. It is called statistical post-processing, e.g. [8] in atmospheric sciences, integrated modelling, e.g. [9] or hybrid modelling, e.g. [10, 11] in system theory and mathematical modelling. In our case, the NWP model is upgraded with a Gaussian-process (GP) model as the data-driven model.

GP modelling is a probabilistic, non-parametric and kernel modelling method for regression analysis, but can also be used for classification. It has properties of kernel methods, and due to its Bayesian principles to include prior knowledge the GP models provide predictive distribution, which can be used to quantify the model fidelity in a systematic way [12].

This paper is structured as follows. The modelling method for the data-driven model, the physics-based model and the hybrid model are introduced in Section 2. Section 3 describes the details of the case study of interest. The modelling results are presented and discussed in Section 4. The conclusions are gathered at the end of this paper.

2 Methods

2.1 GP models

When GP modelling [12–14] is used for regression, the following system is considered:

$$y = f(z) + \nu, \tag{1}$$

where ν is the white Gaussian noise and z is the vector of regressors from the operating space $\mathbb{R}^D$. The noise is of the form $\nu \sim \mathcal{N}(0,\ \sigma_n^2)$, where σ_n^2 is the variance. Elements of the vector $z \in \mathbb{R}^D$, i.e. $z_i{:}i = 1,\ \ldots,\ D$ are called *regressors* and the vector z is called the *regression vector*.

We look for a non-parametric Bayesian model, which has a GP prior over the function f of

$$f(z) \sim \mathcal{GP}(m(f(z)),\ \mathrm{cov}(f(z_i),\ f(z_j))), \tag{2}$$

where $m(f(z))$ is a mean of function and is frequently set to zero because the mean values can be removed from the function and added later if necessary. Moreover, $\mathrm{cov}(f(z_i),\ f(z_j))$ is the covariance of f.

Mean and covariance define the properties of the process we model. They incorporate the prior knowledge of the process to the system training. For the sake of simplicity, we assume the mean function is selected as 0. The covariance matrix is calculated using covariance functions, i.e. kernel functions, which are characterised with hyperparameters. Some of the possible covariance functions are described in the Appendix. A covariance matrix $\boldsymbol{K}$ is calculated by evaluating the covariance function given all the pairs of measured data. The elements $\boldsymbol{K}_{ij}$ of the covariance matrix $\boldsymbol{K}$ are covariances between the values of the functions $f(z_i)$ and $f(z_j)$ corresponding to the arguments z_i and z_j

$$K_{ij} = \mathrm{cov}(f(z_i),\ f(z_j)) = C(z_i,\ z_j). \tag{3}$$

This means that the covariance between the random variables that represent the outputs, i.e. the functions of the arguments numbers i and j, equals the covariance function C between the arguments numbers i and j.

The data for the training of the model is described as $\mathcal{D} = \{(z_i,\ y_i)|i = 1,\ \ldots,\ N\} = \{(\boldsymbol{Z},\ \boldsymbol{y})\}$. Following the Bayesian modelling framework, we are looking for the posterior distribution over f, which for the given data $\{(\boldsymbol{Z},\ \boldsymbol{y})\}$ and hyperparameters θ is:

$$p(f|\boldsymbol{Z},\ \boldsymbol{y},\ \theta) = \frac{p(y|f,\ \boldsymbol{Z},\ \theta)p(f|\theta)}{p(\boldsymbol{y}|\boldsymbol{Z},\ \theta)}, \tag{4}$$

where $p(\boldsymbol{y}|f,\ \boldsymbol{Z},\ \theta)$ is the likelihood, $p(f|\theta)$ is the function f prior for the given hyperparameters θ, $p(\boldsymbol{y}|\boldsymbol{Z},\ \theta)$ is the evidence and $p(f|\boldsymbol{Z},\ \boldsymbol{y},\ \theta)$ is the posterior distribution over f.

The Bayesian inference of most systems can only be implemented with analytical or numerical approximation. One possible approximation method is the estimation of hyperparameters with the maximisation of the evidence. See [12, 13] for details.

The objective of the modelling is to determine predictive distribution of the latent function values $\boldsymbol{f}^* = f(\boldsymbol{Z}^*)$ at test inputs $\boldsymbol{Z}^*$.

To get the predictive distribution, $f = f(\boldsymbol{Z})$ is marginalised out. The resulting predicted distribution is Gaussian and defined with equation

$$p(\boldsymbol{f}^*|\boldsymbol{y}) = \mathcal{N}(\boldsymbol{K}_*(\boldsymbol{K} + \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{y}, \tag{5}$$

$$\boldsymbol{K}_{**} - \boldsymbol{K}_*(\boldsymbol{K} + \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{K}_*). \tag{6}$$

Therefore

$$E(\boldsymbol{f}^*) = \boldsymbol{K}_*(\boldsymbol{K} + \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{y} \tag{7}$$

$$\mathrm{var}(\boldsymbol{f}^*) = \boldsymbol{K}_{**} - \boldsymbol{K}_*(\boldsymbol{K} + \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{K}_*, \tag{8}$$

where $\boldsymbol{f}^*$ is prediction at $\boldsymbol{Z}^*$, $\boldsymbol{K}_* = [C(z_1,\ \boldsymbol{Z}^*),\ \ldots,\ C(z_N,\ \boldsymbol{Z}^*)]^{\mathrm{T}}$ is the $N \times 1$ vector of covariances between the training input data and the test input data and $\boldsymbol{K}_{**} = C(\boldsymbol{Z}^*,\ \boldsymbol{Z}^*)$ is the autocovariance of the test input data.

The computational demand of the direct implementation of GP regression increases with the third power of the size of training set – $\mathcal{O}(N^3)$, due to the calculation of the covariance matrix inverse. This becomes an issue when we work on problems that involve large quantities of training data. The methods that reduce the computational demand of GP modelling are [12] fast matrix–vector multiplication methods, sparse-matrix methods and direct implementations using parallel processing. In our case, we use sparse-matrix methods, which approximate the covariance matrix. The idea of using the sparse-matrix methods is to reduce the rank of the covariance matrix.

Sparse methods make use of so-called inducing input points and a corresponding set of latent variables $\boldsymbol{u}$ to reduce the computational complexity. The latent or so-called inducing variables are presumed to be drawn from the same GP as $\boldsymbol{f}$ and $\boldsymbol{f}^*$

$$p(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{0},\ \boldsymbol{K}_{\mathrm{uu}}). \tag{9}$$

The joint prior distribution is recovered with marginalising out $\boldsymbol{u}$

$$p(\boldsymbol{f},\ \boldsymbol{f}^*) = \int p(\boldsymbol{f},\ \boldsymbol{f}^*|\boldsymbol{u})p(\boldsymbol{u})\mathrm{d}\boldsymbol{u}. \tag{10}$$

It is assumed that $\boldsymbol{f}$ and $\boldsymbol{f}^*$ are conditionally independent given $\boldsymbol{u}$. The approximation of joint prior is then

$$p(\boldsymbol{f},\ \boldsymbol{f}^*) \simeq q(\boldsymbol{f},\ \boldsymbol{f}^*) = \int p(\boldsymbol{f}|\boldsymbol{u})p(\boldsymbol{f}^*|\boldsymbol{u})p(\boldsymbol{u})\mathrm{d}\boldsymbol{u}. \tag{11}$$

The conditionals are

$$p(\boldsymbol{f}|\boldsymbol{u}) = \mathcal{N}(\boldsymbol{K}_{\mathrm{fu}}\boldsymbol{K}_{\mathrm{uu}}^{-1}\boldsymbol{u},\ \boldsymbol{K}_{\mathrm{ff}} - \boldsymbol{Q}_{\mathrm{ff}}) \tag{12}$$

$$p(\boldsymbol{f}^*|\boldsymbol{u}) = \mathcal{N}(\boldsymbol{K}_{*u}\boldsymbol{K}_{uu}^{-1}\boldsymbol{u},\ \boldsymbol{K}_{**} - \boldsymbol{Q}_{**}), \tag{13}$$

where $\boldsymbol{Q}$ is determined using Nyström approximation

$$\boldsymbol{Q}_{\mathrm{ab}} = \boldsymbol{K}_{\mathrm{au}}\boldsymbol{K}_{\mathrm{uu}}^{-1}\boldsymbol{K}_{\mathrm{ub}}^{\mathrm{T}}. \tag{14}$$

The predictive distribution is recovered with

$$p(\boldsymbol{f}^*|\boldsymbol{y}) = \frac{1}{p(\boldsymbol{y})}\int p(\boldsymbol{y}|\boldsymbol{f})q(\boldsymbol{f},\ \boldsymbol{f}^*)\mathrm{d}\boldsymbol{f}. \tag{15}$$

To apply sparse GP regression, we first find the posterior distribution of the inducing outputs $\boldsymbol{u}$ at the corresponding inducing input points z_u.

The method called *fully independent training conditional (FITC)* was proposed in [15], where it was named the Sparse Gaussian process (SPGP) method. The name FITC comes from the fact that the training set function observations are presumed to be completely independent. The covariances on a diagonal of the covariance matrix are exact. This means that instead of approximated prior variances, the exact prior variances are on the covariance matrix diagonal.

$\boldsymbol{\Lambda}$ is a diagonal matrix of $\boldsymbol{K}_{\mathrm{nn}} - \boldsymbol{Q}_{\mathrm{nn}}$. Then, the effective prior is given with

$$q_{\mathrm{FITC}}(\boldsymbol{f}, \boldsymbol{f}^{*}) \sim \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{Q}_{\mathrm{ff}} + \boldsymbol{\Lambda} & \boldsymbol{Q}_{\mathrm{f}*} \\ \boldsymbol{Q}_{*\mathrm{f}} & \boldsymbol{K}_{**} \end{bmatrix}\right). \qquad (16)$$

The predictive distribution is calculated as

$$p(\boldsymbol{f}^{*}|\boldsymbol{y}) = \mathcal{N}(\boldsymbol{Q}_{*\mathrm{f}}(\boldsymbol{Q}_{\mathrm{ff}} + \boldsymbol{\Lambda} + \sigma_n^2 \boldsymbol{I})^{-1}\boldsymbol{y}, \qquad (17)$$

$$\boldsymbol{K}_{**} - \boldsymbol{Q}_{*\mathrm{f}}(\boldsymbol{Q}_{\mathrm{ff}} + \boldsymbol{\Lambda} + \sigma_n^2 \boldsymbol{I})^{-1}\boldsymbol{Q}_{\mathrm{f}*}). \qquad (18)$$

2.2 NWP model – Weather Research and Forecast

Different NWP models exist based on first-principles modelling. One of them that is considered in this investigation is the Weather Research and Forecast (WRF) model – Advanced Research WRF version 3.4.1. This is a fully compressible model that supports real-time NWP, a wide selection of physics models enabling experimentation, real-data and idealised simulations and various lateral boundary condition options [16].

In our case, the physics-based model is used for fine resolution prediction of relative humidity in the geographical domain of interest. The model covers the domain with cells 4 km in size and temporal resolution of 0.5 h. More details can be found in [17]. Hereafter, this model is referred to as the WRF model. Nevertheless, the provided spatial resolution is not high enough to encompass all the local weather effects due to the complex terrain of interest.

2.3 Hybrid model

We investigate if there is an improvement in the predictions obtained using the existing physics-based model combined with a statistical (or data-driven) model compared with the physics-based model alone. Fig. 1 shows the structure used for humidity modelling. As this structure is the combination of a physics-based model with a concatenated statistical model obtained using different modelling methods, it can be referred to as a hybrid model.

The hybrid model combines WRF predictions of humidity with delayed humidity predictions of the model itself, as depicted in Fig. 1, and with measurements of weather variables from the investigated site. While such a model is directly realisable for one-step-ahead predictions only, a solution for a longer horizon of predictions is envisaged. For the multiple-step-ahead predictions, when measurements of weather variables are not available, they are replaced with long-term WRF predictions for the same weather variables.

Two options for statistical model training are investigated in this paper:

(i) The statistical model is trained using measurements of available historical values of weather variables and WRF predictions of relative humidity. The measurements of relative humidity are the training target. When the model is used for multiple-step-ahead predictions, long-term WRF predictions are used instead of measurements. This is an approach often used in fine weather prediction using hybrid models, e.g. [18]. In [18], Hoolohan *et al.* focus on the wind speed modelling and prediction and use only the mentioned approach while our study investigates also the following alternative for comparison.
(ii) The statistical model is trained using WRF predictions, except for the measurements of relative humidity, which are the training target. This approach provides optimal values of optimised parameters also for long-range predictions.

In both cases, the locally conditioned geographical information is contained only within the relative-humidity variable because the model itself is trained with relative-humidity measurements.

3 Case study

3.1 Overview

The purpose of the investigation is to provide predictions of relative humidity for several steps ahead at the Krško NPP. These predictions shall be better than the available predictions from the physics-based model, namely the WRF model. The Krško NPP is located in the eastern part of Slovenia close to the Croatian border. The terrain is considered to be complex as the NPP is surrounded by hills, valleys, a river and other features. The geographical features are shown schematically in Fig. 2.

Relative humidity is one of the weather variables that is required as an input to the air-pollution-dispersion model for radiation exposure, which is being developed in case of an accident at the NPP. However, this investigation can also be used for modelling or improving a physics-based model in the case of any other kind of pollution from a point source. The investigation of relative humidity is just one in a series of investigations of different weather variables that share a common purpose – to be an input to the air-pollution-dispersion model for radiation exposure. Each variable and investigation has special features that are worthwhile to be considered separately. As already mentioned, the WRF model in our case covers weather conditions with a 4 km × 4 km resolution, which is, in our case, not enough to provide a satisfactory weather-variable input for the air-pollution-dispersion model for radiation exposure at the exact location of the Krško NPP.

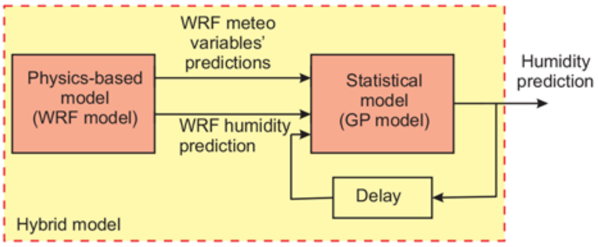
WRF meteo variables' predictions
Physics-based model (WRF model)
WRF humidity prediction
Statistical model (GP model)
Humidity prediction
Delay
Hybrid model

Fig. 1 *Hybrid model of the WRF model and statistical model, where the statistical part of the model uses WRF predictions to replace measurements from measurement stations in each time step. The statistical part of this hybrid model is trained with relative-humidity measurements and WRF predictions of all relevant meteorological variables or with WRF predictions to replace measurements*

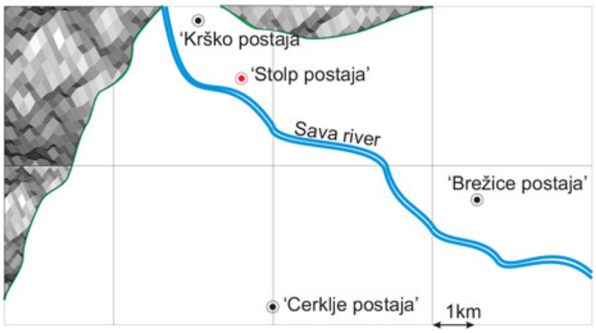

Fig. 2 *Grid showing the available observations from the physics-based model, measurement locations and the location of the plant and its measurement station, which is marked as 'Stolp postaja'*

3.2 Dataset description

Data used in the investigation are composed of measurements of weather variables and WRF predictions. The measurements of various variables are collected by automatic measurement stations at five different locations: Brežice measurement station, Cerklje measurement station, Cerklje airport measurement station, Krško measurement station and ‘Stolp postaja’ measurement station at Krško NPP, all distributed as shown in Fig. 2. The same weather variables are provided also as WRF predictions. Measurements used in this paper are from the years 2015 to 2017 and are sampled with a 30 min interval.

The data from year 2015 and the first half of 2016 are used as training data, the data from the second half of 2016 as validation data and the data from year 2017 as the test data. The datasets have some data gaps omitted in training, but this has a negligible influence on the results. The data periods are sufficiently long that the datasets contain a large amount of data from different seasons with different weather patterns.

3.3 Statistical model structure

The GP model used for improving the WRF-model results is developed using a Gaussian likelihood function, exact inference and constant mean function [19]. The regressors, covariance function and number of inducing points are selected based on one or both of the statistical measures for the assessment. These are as follows:

- The normalised root-mean-square error (NRMSE) is

$$\mathrm{NRMSE} = 1 - \frac{\|\boldsymbol{y} - \boldsymbol{\mu}\|^2}{\|\boldsymbol{y} - E(\boldsymbol{y})\|^2}, \qquad (19)$$

where $\boldsymbol{y}$ is the vector of validation values, $\boldsymbol{\mu}$ is the vector of mean predicted values, $E(\boldsymbol{y})$ is the mean value of $\boldsymbol{y}$.

NRMSE is 1 for a perfect match and $-\infty$ for a very bad match of the validation and mean predicted values.

- The mean standardised log loss (MSLL) [13] is

$$\begin{aligned}\mathrm{MSLL} = {} & \frac{1}{2N}\sum_{i=1}^{N}\left[\ln(\sigma_i^2) + \frac{(E(\hat{y}_i) - y_i)^2}{\sigma_i^2}\right] \\ & - \frac{1}{2N}\sum_{i=1}^{N}\left[\ln(\sigma_y^2) + \frac{(y_i - E(\boldsymbol{y}))^2}{\sigma_y^2}\right],\end{aligned} \qquad (20)$$

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

MSLL is a standardised measure suited to predictions in the form of random variables. It weighs the prediction error more heavily when it is accompanied by a smaller prediction variance. The MSLL is approximately zero for the not very good models and negative for the better ones.

The systematic selection of the covariance function, regressors and the number of inducing points was done using a training and validation dataset. However, it has to be kept in mind that for the selection of each of these elements, other structure elements had to be fixed. The exhaustive selection of various combinations, but not all possible ones, was at least in part done using high-performance computing, a part of the Slovenian national supercomputing network [20]. The investigation was done using Octave open source software [21], except for regressor selection, which was done with MATLAB software [22]. In this way, the relatively large amount of data and relatively high computational load were handled.

What follows are some of the results from which the final structure was selected.

3.3.1 Covariance-function selection: The results given in Table 1 are for various covariance functions at regressors that are shown in the next section and with 185 uniformly selected inducing points of the FITC model. A more detailed description of each covariance function is given in the Appendix. The results of performance measures for validation data show that squared exponential covariance function with automatic relevance determination (ARD) [12, 13] (21), sum of squared exponential (21) and linear covariance function (24) with ARD option, Matérn covariance with hyperparameter $d = (3/2)$ and ARD option (22) and sum of Matérn and linear covariance function with ARD options have comparably good performance results for validation data. It is difficult to argue which of them is best in our case, especially because these results vary slightly with the selected hyperparameters’ initial conditions, number of inducing points and other modelling parameters. We finally selected Matérn covariance function with hyperparameter $d = (3/2)$ and ARD option (22) as our choice for covariance function. Our covariance-function final selection is based on performance measures on validation data that were not used for training, and because Matérn covariance function is known for modelling, less smooth mapping between regression inputs and better outputs [13]. Once the covariance function was fixed, we did not challenge the selection with possible variations of covariance functions on test data. Other covariance functions and their combinations are possible, but we have arbitrarily limited our selection to the ones that are listed.

Table 1 Covariance-function-selection table for validation data with the best measure values in bold

Covariance function	NRMSE	MSLL
squared exponential + ARD	0.8534	−1.9143
squared exponential + linear + ARD	0.8433	**−2.1598**
neural network	0.6217	855.73
Matérn $d = 1.5$ + ARD	**0.8535**	−1.8713
linear + ARD	0.8343	497.10
Matérn $d = 1.5$ + linear + ARD	0.8421	−1.9142
Matérn $d = 2.5$ + linear + ARD	0.8414	−1.9102
Matérn $d = 1.5$ + linear	0.8416	−1.9104

3.3.2 Regressor selection: Regressors were selected with a sequential forward selection method as an example of a wrapping method [23] with four-fold cross-validation on the training dataset. The performance measure used was the logarithm of likelihood. The covariance function for this selection was Matérn covariance with hyperparameter $d = (3/2)$ and an automatic relevance detection option and 185 inducing points of the FITC model.

The final selection of regressors is given as follows:

- relative humidity at ‘Stolp postaja’, delayed for one sample interval,
- global solar radiation at ‘Stolp postaja’, delayed for one sample interval and
- relative humidity from WRF model, prediction for present time.

Relative-humidity values delayed for one sample interval mean that the model output also depends on delayed values, which indicate that we are dealing with a dynamic model [12].

3.3.3 Inducing points selection: The selection of the number of inducing points for the FITC algorithm is shown in Table 2 for

Table 2 Performance measures for validation data of models using a different number of inducing points for Matérn covariance function with hyperparameter $d = (3/2)$ and the ARD option with the best measure values in bold

Number of inducing points	Random NRMSE	Random MSLL	Uniform NRMSE	Uniform MSLL
50	0.8500	**−1.8931**	0.8519	−0.9483
100	**0.8501**	−0.2073	0.8525	0.0764
150	0.0195	2.3922	0.8534	0.2885
185	0.0153	0.0731	**0.8535**	**−1.8713**
200	−0.0781	0.2315	0.8534	0.0042

the Matérn covariance function with hyperparameter $d = (3/2)$ and the ARD [12] option (22). Moreover, options when the inducing points were selected randomly or uniformly among the training dataset are presented. In both cases, regressors as listed in the previous paragraph are used.

Results from Table 2 show that 185 inducing points selected uniformly from the training set with Matérn covariance with hyperparameter $d = (3/2)$ and the automatic relevance detection option result in relatively good results.

Throughout the structure selection, we obtained results that are comparable when Matérn covariance function with hyperparameter $d = (3/2)$ (22) or when squared exponential covariance function (21) with or without linear covariance function (24) are used. Finally, we decided on the following structure: FITC GP model with Matérn covariance with hyperparameter $d = (3/2)$ and ARD option, 185 inducing points uniformly sampled from the training dataset and regressors listed in the paragraph entitled ‘Regressor selection’. It can be claimed that a slightly different choice might be better regarding this or other performance measures, but the one selected provides satisfactory results.

4 Results

The primary goal of this investigation is to obtain better prediction results with a hybrid model than are available with a physics-based model. Since relative humidity is one of the influential input variables to the air-pollution-dispersion model for radiation exposure, it is important to consider improved shorter- and longer-term predictions and to determine which model performs better for different horizons.

As has been already mentioned, two training options are investigated. The first one is to train the statistical part of our hybrid model with measurement data and the measurements of relative humidity as the training target, which is hereafter denoted as the Train 1 option. Moreover, the second one is to train it with forecasts from the WRF model acting as replacements and the measurements of relative humidity as the training target, which is hereafter denoted as the Train 2 option. In particular, this means that global solar radiation measurements were replaced with global solar radiation observations from the WRF model. The two scenarios investigated are necessary because measurements of variables in the future are not available for more than one-step-ahead predictions.

Fig. 3 shows the scatter diagram for the WRF model. It can be seen from Fig. 3 that the predictions of the physics-based model are not optimal. Fig. 4 shows the scatter diagram for the hybrid model with the Train 1 and Train 2 options. It is clear that one-step-ahead predictions are significantly better than those of the WRF model only. This is also confirmed with the performance measures collected in Table 3. The performance of the WRF model can be compared only with the NRMSE measure in our case because we have no information on uncertainties from the WRF-model predictions. The obtained model is also compared with predictions of the so-called persistence model. The persistence model predicts the value in the next time step to be equal to the current value. This kind of model is useful for comparing the one-step-ahead prediction, while assuming that the values staying the same over a larger horizon is not realistic.

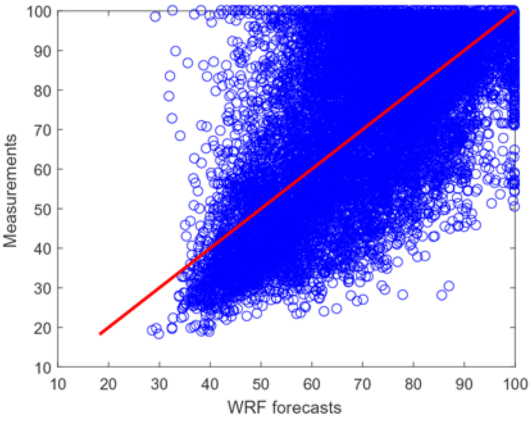

Fig. 3 *Physics-based-model forecasts for the test dataset*

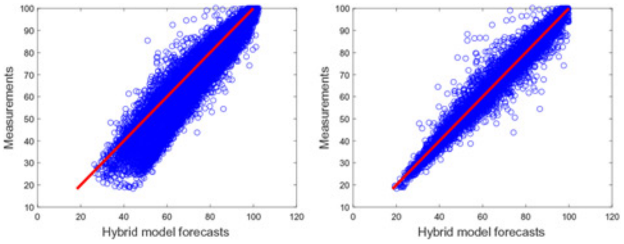

Fig. 4 *Forecasts of the hybrid model for the test dataset. Train 1 option (test inputs are measurements for one-step-ahead predictions and WRF observations for multi-step-ahead predictions) is in the left figure and Train 2 option (test inputs are WRF observations only) is in the right figure*

Table 3 Performance measures

	NRMSE	MSLL
train 1	0.8576	−1.965
train 2	0.8521	−1.8097
WRF on test	−0.0211	/
test 1	0.5588	−0.8054
test 2	0.8528	/
persistence	0.8259	—

A comparison of the performance results for one-step-ahead predictions on the test signal according to the scatter plots and numerical values is in favour of the Train 2 option, where replacements are used for training.

As pollution dispersion, and consequently, in our case, relative-humidity predictions are needed for horizons longer than just one time step, i.e. half an hour, we are interested in more than one-step-ahead predictions as well as in simulation. Simulation in this case means that multi-step-ahead prediction is done iteratively using the most likely predicted value from the model output on the model input for predicting future values until the end of the time period of interest. This is an option that is used when dynamic-systems models are investigated, as it is in our case. More details on options for dynamic-system simulation with GP models can be found in [12]. Model simulation is the most severe test for the developed dynamic model. The NRMSE performance measure for simulation on the test data is 0.2499 and the MSLL performance measure is 2.7715.

The simulation is a multi-step-ahead prediction with an infinite number of prediction steps. A fragment of the simulation response for the hybrid model with the Train 2 option is shown in Fig. 5. The hybrid model with the Train 1 option is not performing well in the simulation. The simulation response completely mismatch the test data response and provides unrealistic response as shown in Fig. 6.

The simulation result in our case serves only as an illustration of the achievable limit in the extreme-prediction case. More interesting is the assessment of how performance degrades with an increasing number of steps in the prediction horizon. Degradation of performance with increasing steps of prediction according to NRMSE and MSLL performance measures for the Train 1 option is shown in Fig. 7. The Train 2 option is shown in Fig. 8 and the NRMSE performance measure for the persistence model is shown in Fig. 9. The investigation is performed for horizons up to 12 h ahead because this is the ultimate period of interest for pollution dispersion.

It is clear that the hybrid model with the Train 2 option performs the best even with an increasing number of prediction steps. While the persistence model can be used for a very-few-steps ahead prediction, it is necessary to exercise caution

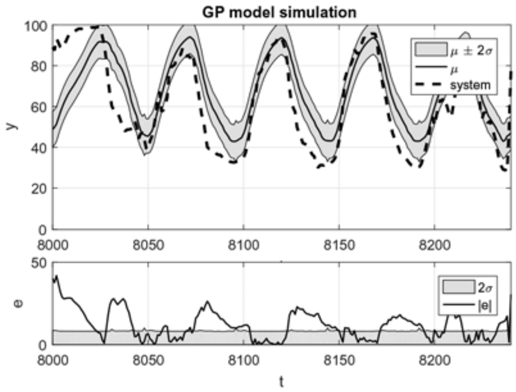

Fig. 5 *Fragment of the comparison of time responses for long-term forecasts of relative humidity on test data for the Train 2 option. Legend: mean values of forecasts – full line in upper figure, 95% confidence interval – grey band in upper figure, measurements – dashed line in upper figure, the absolute value of the difference between forecasts and measurement – full line in the bottom figure, 95% confidence interval – grey band in the bottom figure*

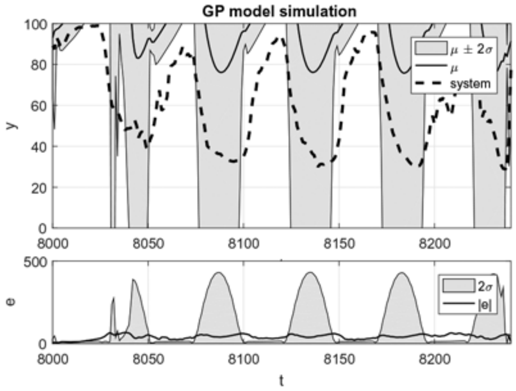

Fig. 6 *Fragment of the comparison of time responses for long-term forecasts of relative humidity on test data for the Train 1 option. Legend: mean values of forecasts – full line in upper figure, 95% confidence interval – grey band in upper figure, measurements – dashed line in upper figure, the absolute value of the difference between forecasts and measurement – full line in the bottom figure, 95% confidence interval – grey band in the bottom figure*

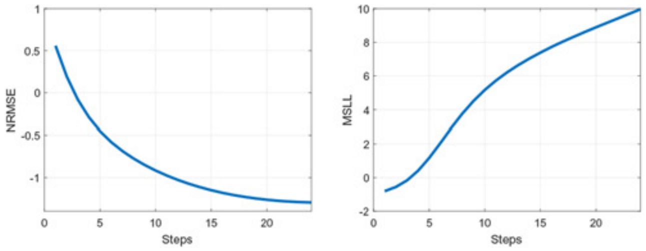

Fig. 7 *NRMSE versus number of prediction steps and MSLL versus number of prediction steps for the Train 1 option*

when using such predictions. The values of the NRMSE and MSLL measures for the Train 2 option converge toward performance measures of the simulation run, which are not really good (NRMSE = 0.2499, MSLL = 2.7715). Nevertheless, it provides predictions as depicted in Fig. 5, which clearly mimic general daily and seasonal periodical trends and do not differ significantly from measurements used as test data. This is important information when using relative humidity as the input to the air-pollution-dispersion model for radiation exposure.

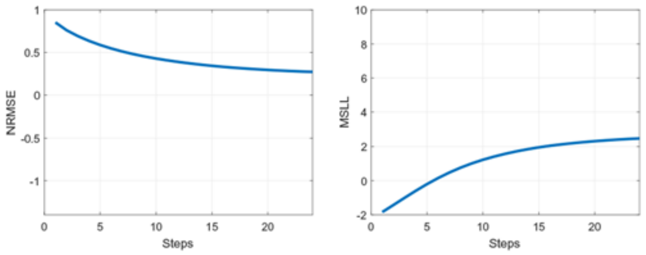

Fig. 8 *NRMSE versus number of prediction steps and MSLL versus number of prediction steps for the Train 2 option*

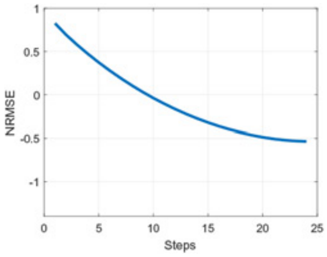

Fig. 9 *NRMSE versus number of prediction steps for the persistence model*

It can be concluded that, for the data at hand, the hybrid model comprised of the physics-based WRF model enhanced with the GP model provides acceptable results only with the Train 2 option. It has to be kept in mind that our main goal is to overperform predictions of the WRF model and we have achieved this goal. The persistence model, on the other hand, might be a viable alternative only for a very small horizon, i.e. one- or two-step horizon, which corresponds to horizons of half an hour and 1 h.

5 Conclusion

An application of the sparse hybrid modelling between physics-based modelling and modelling from data of the Krško basin was presented in this paper. The hybrid modelling was implemented as modelling from data based on a dataset containing observations from a physics-based model and measurements from the field. The main goal of the investigation was to improve the prediction results of a dynamic physics-based model and this was achieved for the investigated case study. The obtained data-based model part of the hybrid system was also a dynamic-system model.

The main contributions of this paper are as follows:

- an improvement in relative-humidity prediction results of the case study, which are required for further processing with the air-pollution-dispersion model for radiation exposure,
- systematic comparison of two training approaches for the regression model with the purpose of multi-step-ahead prediction – one with measurements for training and physics-based-model observations for predictions and the other with physics-based-model observations, except for the target variable, and
- comparison of the hybrid model's responses with the persistence model and evaluation of the usability of the developed models for multi-step-ahead forecasting for the case study.

The main conclusions are as follows:

- Hybrid modelling, where the results of the physics-based model are further processed with statistical methods are a promising solution for the improvement of predictions in the case study.
- The hybrid model with observations from the physics-based model of relative humidity in the Krško basin apart from the target variable that was measured performed notably better than the other assessed models. This is in contrast to what is frequently practised

in that models are developed based on measurements only and replaced with other observations for predictions in the future.

- GP modelling was demonstrated as a suitable method in this case study investigation, especially its capability to quantify the quality of the model prediction in a systematic way, which is useful for the obtained model assessment.
- Sparse and high performance computing (HPC) improved realisability of the modelling from large amounts of data.
- While this paper demonstrated modelling of relative humidity with the purpose of using the predictions for further processing with the air-pollution-dispersion model for radiation exposure in the basin where Krško NPP is located, the method used for hybrid-model development and assessment can be directly transferable to the modelling of other atmospheric variables and for other pollution-dispersion investigations.

Modelling of other atmospheric variables, e.g. wind speed and direction and temperature profile, is envisaged as further steps in the investigation. Large amounts of data and seasonal changes in the data direct our future investigations toward the regular and periodic remodelling of the hybrid model as well as to the utilisation of an online modelling approach.

6 Acknowledgments

The authors acknowledge the project ‘Method for the forecasting of local radiological pollution of atmosphere using Gaussian-process models’, ID L2-8174 and research core Funding no. P2-0001, which were financially supported by the Slovenian Research Agency. We are grateful to the NPP Krško for the measurement data from their automatic measuring system.

7 References

[1] Breznik, B., Božnar, M.Z., Mlakar, P., *et al.*: ‘Dose projection using dispersion models’, *Int. J. Environ. Pollut.*, 2003, **20**, (1–6), pp. 278–285

[2] Lynch, P.: ‘The origins of computer weather prediction and climate modeling’, *J. Comput. Phys.*, 2008, **227**, (7), pp. 3431–3444. predicting weather, climate and extreme events

[3] Alimissis, A., Philippopoulos, K., Tzanis, C.G., *et al.*: ‘Spatial estimation of urban air pollution with the use of artificial neural network models’, *Atmos. Environ.*, 2018, **191**, pp. 205–213

[4] Lazos, D., Sproul, A.B., Kay, M.: ‘Development of hybrid numerical and statistical short term horizon weather prediction models for building energy management optimisation’, *Build. Environ.*, 2015, **90**, pp. 82–95

[5] Philippopoulos, K., Deligiorgi, D., Kouroupetroglou, G.: ‘Artificial neural network modeling of relative humidity and air temperature spatial and temporal distributions over complex terrains’. In: Fred, A., De Marsico, M. (Eds.): ‘Pattern recognition applications and methods’ (Springer International Publishing, Cham, 2015, pp. 171–187

[6] Mba, L., Meukam, P., Kemajou, A.: ‘Application of artificial neural network for predicting hourly indoor air temperature and relative humidity in modern building in humid region’, *Energy Build.*, 2016, **121**, pp. 32–42

[7] Bayatvarkeshi, M., Mohammadi, K., Kisi, O., *et al.*: ‘A new wavelet conjunction approach for estimation of relative humidity: wavelet principal component analysis combined with ANN’, *Neural Comput. Appl.*, 2018

[8] Worsnop, R.P., Scheuerer, M., Hamill, T.M., *et al.*: ‘Generating wind power scenarios for probabilistic ramp event prediction using multivariate statistical post-processing’, *Wind Energy Sci.*, 2018, **3**, (1), pp. 371–393

[9] Gradišar, D., Grašič, B., Božnar, M.Z., *et al.*: ‘Improving of local ozone forecasting by integrated models’, *Environ. Sci. Pollut. Res.*, 2016, **23**, (18), pp. 18439–18450

[10] Von Stosch, M., Oliveira, R., Peres, J., *et al.*: ‘Hybrid semi-parametric modeling in process systems engineering: past, present and future’, *Comput. Chem. Eng.*, 2014, **60**, pp. 86–101

[11] Ren, W.W., Yang, T., Huang, C.S., *et al.*: ‘Improving monthly streamflow prediction in alpine regions: integrating HBV model with Bayesian neural network’, *Stoch. Environ. Res. Risk Assess.*, 2018, **32**, (12), pp. 3381–3396

[12] Kocijan, J.: ‘Modelling and control of dynamic systems using Gaussian process models’ (Springer International Publishing, Cham, 2016)

[13] Rasmussen, C.E., Williams, C.K.I.: ‘Gaussian processes for machine learning’ (MIT press, Cambridge, 2006)

[14] Shi, J.Q., Choi, T.: ‘Gaussian process regression analysis for functional data’ (CRC Press, Boca Raton, 2011)

[15] Snelson, E., Ghahramani, Z.: ‘Sparse Gaussian processes using pseudo-inputs’. In: Weiss, Y., Schölkopf, B., Platt, J. (Eds.): ‘Advances in neural information processing Systems’, vol. 18 (MIT Press, Cambridge, MA, 2006) pp. 1257–1264

[16] Skamarock, W.C., Klemp, J.B., Dudhia, J., *et al.*: ‘A description of the advanced research WRF version 3’ (National Center for Atmospheric Research, Boulder, CO, 2008)

[17] Božnar, M.Z., Mlakar, P., Grašič, B.: ‘Short-term fine resolution WRF forecast data validation in complex terrain in Slovenia’, *Int. J. Environ. Pollut.*, 2012, **50**, (1–4), pp. 12–21

[18] Hoolohan, V., Tomlin, A.S., Cockerill, T.: ‘Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data’, *Renew. Energy*, 2018, **126**, pp. 1043–1054

[19] Rasmussen, C.E., Nickisch, H.: ‘Gaussian processes for machine learning (GPML) toolbox’, *J. Mach. Learn. Res.*, 2010, **11**, (Nov), pp. 3011–3015

[20] ‘Slovenian national supercomputing network’. (SLING). Available at https://www.sling.si, accessed 19 July 2019

[21] ‘Gnu octave’. (Octave). Available at https://www.gnu.org/software/octave/, accessed 19 July 2019

[22] ‘MATLAB’. (Mathworks). Available at https://www.mathworks.com/, accessed 19 July 2019

[23] May, R., Dandy, G., Maier, H.: ‘Review of input variable selection methods for artificial neural networks’. In: Suzuki, K. (Ed.): ‘Artificial neural networks – methodological advances and biomedical Applications’ (InTech, Rijeka, 2011), pp. 19–44

8 Appendix

Covariance functions [12, 13]:

- Squared exponential + ARD

$$C_{\mathrm{f}}(\boldsymbol{z}_i,\ \boldsymbol{z}_j) = \sigma_{\mathrm{f}}^2 \exp\left[-\frac{1}{2}(\boldsymbol{z}_i - \boldsymbol{z}_j)^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} (\boldsymbol{z}_i - \boldsymbol{z}_j)\right], \tag{21}$$

where $\boldsymbol{\Lambda}^{-1} = \mathrm{diag}([l_1^{-2},\ \ldots,\ l_D^{-2}])$ represents different length scales on different regressors and can be used to assess the relative importance of the contributions made by each regressor through comparison of their lengthscale hyperparameters, which is called ARD. The hyperparameter σ_{f}^2 represents the scaling factor of the possible variations of the function or the vertical scaling factor.

- Matérn

$$C_{\mathrm{f}}(\boldsymbol{z}_i,\ \boldsymbol{z}_j) = \sigma_{\mathrm{f}}^2 \left(\frac{2^{1-d}}{\Gamma(d)}\right) \left(\frac{\sqrt{2\mathrm{d}}r}{l}\right)^d K_d\left(\frac{\sqrt{2\mathrm{d}}r}{l}\right), \tag{22}$$

where the hyperparameter l or the horizontal scaling factor determines the relative weight on distance for the input variable $\boldsymbol{z}$, $r = |\boldsymbol{z}_i - \boldsymbol{z}_j|$, Γ is the gamma function and K_d is a modified Bessel function and the hyperparameter d can be seen to control the differentiability of the modelled mapping function. Often, d is fixed to be $d = (3/2)$ or $d = (5/2)$.

- Matérn + ARD
  Same as (21), but with $r = \sqrt{(\boldsymbol{z}_i - \boldsymbol{z}_j)^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} (\boldsymbol{z}_i - \boldsymbol{z}_j)}$
- Linear

$$C_{\mathrm{f}}(\boldsymbol{z}_i,\ \boldsymbol{z}_j) = \sigma_{\mathrm{f}}^2 (\boldsymbol{z}_i\, \boldsymbol{z}_j). \tag{23}$$

- Linear + ARD

$$C_{\mathrm{f}}(\boldsymbol{z}_i,\ \boldsymbol{z}_j) = \boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} \boldsymbol{z}_j. \tag{24}$$

- Neural network

$$C_{\mathrm{f}}(\boldsymbol{z}_i,\ \boldsymbol{z}_j) = \sigma_{\mathrm{f}}^2 \frac{2}{\pi} \sin^{-1}\left(\frac{2\tilde{\boldsymbol{z}}_i^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} \tilde{\boldsymbol{z}}_j}{\sqrt{1 + 2\tilde{\boldsymbol{z}}_i^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} \tilde{\boldsymbol{z}}_i}\sqrt{1 + 2\tilde{\boldsymbol{z}}_j^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} \tilde{\boldsymbol{z}}_j}}\right), \tag{25}$$

where $\tilde{\boldsymbol{z}}_i = [1,\ \boldsymbol{z}_i] = [1,\ z_1,\ \ldots,\ z_D]^{\mathrm{T}}$.