Hybrid model of the near-ground temperature profile

Juš Kocijan · Matija Perne · Primož Mlakar · Boštjan Grašič · Marija Zlata Božnar

J. Kocijan
Jožef Stefan Institute,
Jamova cesta 39, SI-1000 Ljubljana, Slovenia
Tel.: +386-1-4773661
Fax.: +386-1-4773994
E-mail: jus.kocijan@ijs.si
http://orcid.org/0000-0002-1221-946X
and
University of Nova Gorica,
Vipavska 13, SI-5000 Nova Gorica, Slovenia

M. Perne
Jožef Stefan Institute,
Jamova cesta 39, SI-1000 Ljubljana, Slovenia

P. Mlakar, B. Grašič, M. Z. Božnar
MEIS d.o.o.,
Mali Vrh pri Šmarju 78, SI-1293 Šmarje-Sap, Slovenia

Abstract The topic of the paper is modelling and prediction of atmospheric variables that are further used for prediction of the consequences of radioactive-material release to the atmosphere. Physics-based models of atmospheric dynamics provide an approximate description of the true nature of a dynamic system. However, the accuracy of the model's short-term predictions and long-term forecasts, especially over complex terrain, decreases when the information at a micro-location is sought. Integration of a physics-based model with a statistical model for enhancing the prediction power is proposed in the paper. Gaussian Processes models can be used to identify the mapping between the system input and output measured values. With the given mapping function, we can provide one-step ahead prediction of the system output values together with its uncertainty, which can be used advantageously. In this paper, we combine a physics-based model with a Gaussian-process model to identify air temperature from measurements at different atmospheric surface layers as a dynamic system and to make short-term predictions as well as long-term forecasts.

Keywords Hybrid model · Vertical temperature profile · Physics-based model · Statistical modelling · Gaussian-process model

1 Introduction

Modelling and prediction of vertical temperature profile using a hybrid model are investigated. The modelled atmospheric variable is further used for prediction of the consequences of a radioactive-material release to the atmosphere. Efficient evacuation of the local inhabitants in the case of an accident of a nuclear power plant requires prediction of air-pollution dispersion in the radius of at least ten kilometres. A necessary input to the expert system, i.e., the air-pollution-dispersion modelling system (Breznik et al., 2003) predicting the dynamics of the cloud of radionuclides, is the prediction of the 3D state of the atmosphere above the study area.

Two fundamentally different ways of modelling atmospheric variables are physics-based modelling and statistical modelling. A physics-based model, also called a deterministic, theoretical or first-principles model (Zhang et al., 2012a,b) depending on the professional field, is derived from physics-based relations among the physical and chemical atmospheric variables. It provides an insight into the atmospheric and air-quality dynamics for various typical and atypical scenarios. Nevertheless, this kind of model does not provide information in sufficiently fine spatial resolution. Statistical models, also known as experimental or empirical models, provide strictly local information on atmospheric and air-quality conditions. They rely only on measurements of various variables. Gaussian Processes (GPs) modelling (Rasmussen and Williams, 2006; Shi and Choi, 2011; Kocijan, 2016) is one of the statistical methods that can be used for empirical modelling. The utility of GP modelling in environmental and atmospheric modelling has been demonstrated (Kocijan et al., 2015, 2016, 2018). Nevertheless, such a model is confined to the vicinity or region from where the data to obtain the model originates.

An evident solution avoiding the drawbacks of physics-based and statistical models is combining both methods. This is not a new approach and is called integrated modelling (Gradišar et al., 2016) or statistical postprocessing (Worsnop

et al., 2018) in atmospheric sciences and hybrid modelling (Von Stosch et al., 2014; Ren et al., 2018) in system theory and mathematical modelling. Through combining the models, one retains the physical insight into the atmospheric dynamics and the information about seasonality effects from the global physics-based model as well as a model of deviations addressing a particular location from the statistical model.

The focus of the study is modelling and prediction with hybrid models of the vertical temperature profile at the location of the Nuclear Power Plant (NPP) in Krško, Slovenia. For the study area, a fine-resolution physics-based model of the temperature profile validated with profile measurements exists (Grašič et al., 2019). The study of Grašič et al. (2019) identified several problems that are most likely caused by atmospheric processes not modelled properly due to the spatial resolution not being fine enough. These problems, suitable to be addressed by hybrid modelling, are, e.g., thermal inversion, nocturnal local pools of cold air, and ground-level overheating due to local land-use driven effects. The proposed method is not constrained to the local region of the case study but can be used to improve modelling in locations where a physics-based model alone cannot meet the modelling criteria, e.g., in places surrounded by complex geographical terrain.

The temperature profile is one of the important inputs to the air-pollution dispersion model. It provides information about the air-mass movement (Ramaswamy et al., 2006) and is often investigated with various methods for temperature data retrieval (Emeis et al., 2012; Gangwar et al., 2014; Fochesatto, 2015; Rekhy et al., 2018). Our intention is to develop a system integrated from several models predicting several different atmospheric variables, e.g., the wind speed and direction, with the goal of ensuring a high-resolution forecast of the local atmospheric conditions, which will enable a very good air-pollution forecast and consequently allow for efficient evacuation actions to take place.

In the paper, *prediction* denotes short-term, usually one-step-ahead prediction, while *forecast* denotes long-term, multistep-ahead predictions.

2 Site and measurements

The domain of our investigation extends over the area of 25 km $\times$ 25 km around the Krško NPP, which is located in the East part of Slovenia close to Croatian border. Nevertheless, the model that we are developing and consequently the corresponding variables are placed particularly on the location of the nuclear power plant. The terrain around the power plant can be considered as a complex terrain (Figure 1).

The vertical temperature profile would ideally mean a continuous distribution of the atmospheric temperature versus altitude, but data retrieval using sensors at the Krško NPP can be achieved only with a finite resolution, i.e., only 4 temperatures at heights of 2 m, 10 m, 40 m and 70 m can be retrieved. The vertical temperature profile prediction at the selected geographical point shall be obtained based on as many relevant predicted and measured meteorological variables as possible from the wider domain of interest. However, we are constrained by available observations and it is necessary to gain maximal information about the temperature profile from what is available.
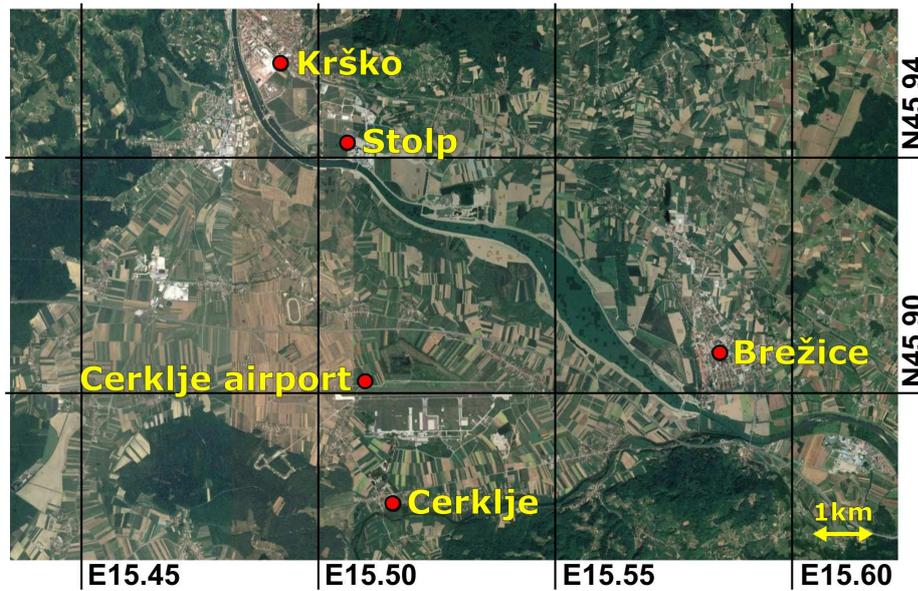
Fig. 1 The geographical features of the surrounding terrain and the measurement station. The plant and its measurement station (marked as 'Stolp') is situated in the basin surrounded by hills and valleys, which influence micro-climate conditions. The vertices of the grid represent points in space where physics-based model predictions are available.

MEIS company has been pursuing measurement activities and analysis for the nuclear plant for years. Measurements used in our study are from the years 2016 and 2017 from automatic measurement stations that are spread unevenly in the Krško basin (Figure 1), and are as follows:

– Brežice measurement station: temperature 2 m above the ground, wind speed 10 m above the ground, wind direction 10 m above the ground, relative humidity 2 m above the ground;
– Cerklje measurement station: temperature 2 m above the ground, relative humidity 2 m above the ground, air pressure;
– Cerklje airport measurement station: temperature 2 m above the ground, wind speed 10 m above the ground, wind direction 10 m above the ground, relative humidity 2 m above the ground, air pressure;
– Krško measurement station: temperature 2 m above the ground, wind speed 10 m above the ground, wind direction 10 m above the ground, relative humidity 2 m above the ground;
– Stolp measurement station at Krško NPP: temperature 2 m above the ground, temperature 10 m above the ground, temperature 40 m above the ground, temperature 70 m above the ground, wind speed 10 m above the ground, wind direction 10 m above the ground, relative humidity 2 m above the ground, global solar radiation, air pressure.

Automatic measuring stations take measurements in real-time at 30-minute statistical intervals. The assimilation rate of data is also 30-minute and equal to

the measuring time interval. Characteristics of the measurement equipment are as follows.

– 2D ultrasonic anemometers are used to measure the horizontal component of wind speed and direction. Measuring precision of wind speed is 0.1 m/s and 1° for wind direction. The accuracy of wind speed is ± 0.1 m/s for the range from 0.0 m/s to 5.0 m/s and ± 2 % of measuring value above 5.0 m/s. The accuracy of wind direction is ± 1°. The sampling rate of wind measurements is 1 second.
– Air temperature at 2 m and 10 m levels above the ground are measured using Pt-100 (platinum resistance) probes with precision of 0.1 °C and accuracy of ± 0.1 °C. Aspirated thermo-linear thermistor temperature sensors are used for the air temperature measurements at levels 40 m and 70 m. Their precision is 0.1 °C and accuracy is ± 3 °C. The sampling rate of air-temperature measurements is 10 seconds.
– Relative humidity is measured using a capacitive humidity sensor with the precision of 0.1 % and accuracy of 1 %. The sampling rate of relative-humidity measurements is 10 seconds.
– Global solar radiation is measured using pyranometer with the precision of 1 W/m^2 and accuracy of ± 2.8 %. The sampling rate is 10 seconds.
– Air pressure measurements are made using barometric pressure transducer. Its precision is 0.1 hPa and accuracy is ± 0.3 hPa. The sampling rate is 10 seconds.

Forecasts taken from the physics-based model, which in our case study is the Weather Research & Forecast (WRF) model (Skamarock et al., 2008) and is described in Section 3.1, at the location closest to Krško NPP, are as follows:

– temperature 2 m above the ground, wind speed 10 m above the ground, wind direction 10 m above the ground, cloud cover, relative humidity 2 m above the ground, global solar radiation and air pressure.

Sensor measurements at intervals of 30 min are used in our study. The measured data are divided into data for identification, i.e., training data, and data for model validation, i.e., test data. The measurements and predictions of WRF model, which is described in the next section, from June 16, 2016, until June 30, 2016, are used as identification data, and data from June 1, 2017, until June 30, 2017, are used as validation data.

Validation using data different from the identification data provides an adequate assessment of the obtained models. There are two reasons for the relatively short period of identification data. The first is that seasonal and other long-term variations of meteorological variables are already covered by the WRF model. Therefore, the statistical model needs to model only the differences in the WRF-model forecasts for which a relatively short period will do. This reasoning is supported also by the fact that differences in the WRF-model forecasts are not excessive, but are persistent (Božnar et al., 2012). The second reason is the computational burden due to a large number of data and is kept reasonable with the number of data used.

3 Methods

The core idea of the proposed approach is to upgrade predictions obtained from a physics-based model with a statistical model. In our case, these models are the WRF model as the physics-based model and the GP model as the statistical model. Our rationale to use GP models for the construction of a hybrid model is manifolds. They can highlight areas of the input space where model-prediction quality is poor, due to the lack of data or its complexity, by indicating the higher variance around the predicted mean. This property is very informative in assessing prediction quality. GP models contain noticeably fewer coefficients to be optimised than other statistic models and are therefore convenient for optimisation. Nevertheless, it is important to emphasise that other statistic models can also be used to construct hybrid models.

3.1 The numerical weather prediction model—WRF

The physics-based model, i.e., the WRF model, is used for fine-resolution numerical weather predictions for the studied domain. The WRF model is used for atmospheric research, but its primary purpose is operational forecasting. It is a system for numerical weather prediction.

The WRF software framework consists of two dynamics solvers for the model computation: the Advanced Research WRF (ARW) and the Nonhydrostatic Mesoscale Model (NMM) solver. The ARW solver, version 3.4, by NCAR, is used in our case study and is hereafter referred to as the WRF model. The WRF ARW 3.4 model has the following settings Gradišar et al. (2016):

- it covers a larger and a smaller geographical domain: central part of Europe and Slovenia with surroundings,
- the larger domain consists of 101 by 101 cells, each cell size is 12 km, model temporal resolution is 3 hours,
- the smaller domain consists of 76 by 76 cells, each cell size is 4 km, model temporal resolution is 0.5 hour,
- the prediction is made for two days and three hours in advance,
- the Global Forecasting Model (GFS) data are used for initial and boundary conditions,
- the simulation is made twice per day, with data at 00:00 Universal Time Coordinated (UTC) and 12:00 UTC,
- results are available with a delay of 5.5 hours,
- a new prediction overwrite the previous one for the time intervals that overlap.
- The weather prediction is better than it would be if made only once per day. The predicted meteorological variables are used as regressors for the statistical part of the hybrid model.

Validation of the above-described system is described in Božnar et al. (2012).

3.2 Gaussian Process Models

The GP model is a non-parametric Bayesian model used for regression (Rasmussen and Williams, 2006; Kocijan, 2016). Its task is to infer a mapping given the data

used for model identification $\mathcal{D} = \{\mathbf{Z}, \mathbf{y}\}$, where $y(\mathbf{z}_i) = f(\mathbf{z}_i) + \nu; i = 1 \ldots, N$ with the white noise $\nu \sim \mathcal{N}(0, \sigma_\nu^2)$ and the regression matrix $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N]^T$ represents N D-dimensional regression vectors. The outputs $y(\mathbf{z}_i)$ are usually assumed to be noisy realisations of the underlying function $f(\mathbf{z}_i)$. The underlying function is represented as a Gaussian process:

$$f(\mathbf{z}) \sim \mathcal{GP}(m(\mathbf{z}), k(\mathbf{z}_i, \mathbf{z}_j)); i, j = 1 \ldots, N. \tag{1}$$

The mean function $m(\mathbf{z})$ is often taken as zero, while the covariance or kernel function $k(\mathbf{z}_i, \mathbf{z}_j)$ controls the smoothness of the GP. A selection of covariance functions is given in, e.g., (Kocijan, 2016; Rasmussen and Williams, 2006). Covariance functions contain so called hyperparameters. The values of the hyperparameters depend on the data-at-hand and it is difficult to select their prior distribution. If a uniform prior distribution is selected, which means that any values for the hyperparameters are equally possible a priori, then the hyperparameters' posterior distribution is proportional to the marginal likelihood of the GP posterior

$$p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{y}|\mathbf{f})p(\mathbf{f})d\mathbf{f} = \mathcal{N}(\mathbf{y}|0, \mathbf{K} + \sigma_\nu^2 \mathbf{I}_n), \tag{2}$$

where $\boldsymbol{\theta}$ comprises the hyperparameters. The hyperparameters are obtained with the maximisation of log marginal likelihood

$$\log p(\mathbf{y}) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{K} + \sigma_\nu^2 \mathbf{I}_n| - \frac{1}{2}\mathbf{y}^T (\mathbf{K} + \sigma_\nu^2 \mathbf{I}_n)^{-1} \mathbf{y}. \tag{3}$$

For the sake of clarity, the hyperparameters $\boldsymbol{\theta}$ now are omitted from the conditioning of the distribution.

Given the identification data (also referred to as training data), $\mathcal{D}$, and the inferred hyperparameters, the predictive distribution
$p(f^*|\mathcal{D}, \mathbf{z}^*) = \mathcal{N}(f^*|\mu(\mathbf{z}^*), \sigma^2(\mathbf{z}^*))$ at a validation point (also referred to as a test point) $\mathbf{z}^*$ has the mean and variance respectively expressed as

$$\mu(\mathbf{z}^*) = \mathbf{k}^T (\mathbf{K} + \sigma_\nu^2 \mathbf{I}_n)^{-1} \mathbf{y}, \tag{4}$$
$$\sigma^2(\mathbf{z}^*) = \kappa - \mathbf{k}^T (\mathbf{K} + \sigma_\nu^2 \mathbf{I}_n)^{-1} \mathbf{k}, \tag{5}$$

where $\mathbf{k} = k(\mathbf{z}^*, \mathbf{Z})$ and $\kappa = k(\mathbf{z}^*, \mathbf{z}^*)$. For y^*, we need to consider the noise such that $p(y^*|\mathcal{D}, \mathbf{z}^*) = \mathcal{N}(y^*|\mu(\mathbf{z}^*), \sigma^2(\mathbf{z}^*) + \sigma_\nu^2)$.

GP models can, like other regression methods, be used to model static nonlinearities and can, therefore, be used for the modelling of dynamic systems as well as time series if delayed samples of the output signals are fed back and used as regressors. A review of recent developments in the modelling of dynamic systems using GP models and their applications can be found in Kocijan (2016). A single-input single-output dynamic GP model is trained as the nonlinear autoregressive model with an exogenous input (NARX) representation, where the output at time instant k depends on the delayed output y and the exogenous control input u:

$$y(k) = f(y(k-1), \ldots, y(k-n), u(k-1), \ldots, u(k-m)) + \nu(k), \tag{6}$$

where f denotes a function, $\nu(k)$ is white noise disturbance with normal distribution, $n, m \in \mathbb{N}$ and the output $y(k)$ depends on the regression vector $\mathbf{z}(k) = [y(k-1), \ldots, y(k-n), u(k-1), \ldots, u(k-m)]$ at time step or time instant k.

The evaluation of the *long-term behaviour* of the dynamic model, i.e., long-term forecasts or model validation is done with a simulation. The simulation is a multistep-ahead prediction when the number of steps in the prediction horizon is infinite or at least as large as the time horizon of interest for the foreseen analysis of the model's behaviour. There are two implementation options for the simulation or for the multistep-ahead prediction (Kocijan, 2016):

- a direct method, where different models are learnt for every perceived horizon or
- an iterative method, where the one-step-ahead prediction is iteratively repeated.

The problem of the direct method is that the horizon needs to be known and fixed in advance. In the case that the horizon is changed, the model has to be learnt again. The second issue with the direct method is that highly nonlinear systems need a large horizon and, consequently, a large amount of learning data.

The iterative method for Gaussian process models of dynamic systems means that the current output estimate depends on previous model estimations and on the measured inputs as described with the following equation:

$$\hat{y}(k) = f(\hat{y}(k-1), \ldots, \hat{y}(k-n), u(k-1), \ldots, u(k-m)) + \nu(k), \qquad (7)$$

where $\hat{y}(k-i); i = 1 \ldots n$ denotes the output estimate i samples or time steps in the past.

In the case of a multistep-ahead prediction, we wish to make a prediction at $\mathbf{z}^*$, but this time the input vector $\mathbf{z}^*$ contains uncertain input values fed back from the outputs. Within a Gaussian approximation, the input values can be described by the normal distribution $\mathbf{z}^* \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{z}^*}, \boldsymbol{\Sigma}_{\mathbf{z}^*})$, where $\boldsymbol{\mu}_{\mathbf{z}^*}$ and $\boldsymbol{\Sigma}_{\mathbf{z}^*}$ are the vector and the matrix of the input mean values and variances, respectively. To obtain a prediction, we need to integrate the predictive distribution $p(y^*|\mathbf{z}^*, \mathcal{D})$ over the input data distribution, that is

$$p(y^*|\boldsymbol{\mu}_{\mathbf{z}^*}, \boldsymbol{\Sigma}_{\mathbf{z}^*}, \mathcal{D}) = \int p(y^*|\mathbf{z}^*, \mathcal{D}) p(\mathbf{z}^*) d\mathbf{z}^*, \qquad (8)$$

where

$$p(y^*|\mathbf{z}^*, \mathcal{D}) = \frac{1}{\sqrt{2\pi\sigma^2(\mathbf{z}^*)}} \exp\left[-\frac{(y^* - \mu(\mathbf{z}^*))^2}{\sigma^2(\mathbf{z}^*)}\right]. \qquad (9)$$

Since $p(y^*|\mathbf{z}^*, \mathcal{D})$ is, in general, a nonlinear function of $\mathbf{z}^*$, the new predictive distribution $p(y^*|(\boldsymbol{\mu}_{y^*}, \boldsymbol{\Sigma}_{y^*}, \mathcal{D}))$ is not Gaussian and this integral cannot be solved without using an approximation. In other words, when the Gaussian distribution is propagated through a nonlinear model, it is not a Gaussian distribution at the output of the model. One of the possible approximations that can be used for simulation is Monte Carlo simulation, where response is calculated based on a large number of model responses utilising sampling of stochastic variables from the model (Korn, 2007). The Monte Carlo simulation is, in our case, implemented as the large number of deterministic realisations of the simulation, where the values of ν and f are obtained using random sampling from their predicted Gaussian distributions based on deterministic inputs at each time step. Mean value and variance of the output variable as a function of time are then estimated from the sample of realisations. Other approximation methods are reviewed in (Kocijan, 2016).

3.3 Hybrid models

Three different sorts of hybrid models incorporating the WRF ARW model and GP model that are evaluated in this paper are as follows:

– Models that combine WRF predictions of temperature with delayed temperature predictions of the model itself, as described with Equation (7), and with measurement updates from Krško NPP except for temperature measurements and measurement updates from other stations (Figure 2) in each time step. This kind of model is realisable when only half-hour-ahead predictions are necessary. Exogenous variables for the GP part of this hybrid model are, therefore, WRF prediction of temperature, measurement updates from Krško NPP except temperature measurements, and measurement updates from other stations. For prediction horizons beyond the interval of sampling, one needs a parallel model for each of the input variables and, consequently, models for each of the parallel models' inputs. This means that models for every measured variable are necessary, which is rather inconvenient but realisable. With this model, the locally conditioned geographical information is contained within the temperature variable and within measurements from the measurement stations, which comprises identification data for the GP part of the model. We name this model as Hybrid 1.
– Models that combine WRF predictions of temperature with delayed temperature predictions of the model itself, as described with Equation (7), and WRF predictions of other necessary variables (Figure 3) are used as a replacement for measurements used in the previous hybrid model. Exogenous variables for the GP part of this hybrid model are, therefore, WRF prediction of temperature and WRF predictions of other necessary variables. With such a model, the locally conditioned geographical information is contained only within the temperature variable because the model itself is trained with temperature measurements. We name this model as Hybrid 2.
– Models that use only WRF predictions of temperature and delayed temperature predictions of the model itself, as described with Equation (7) (Figure 4). The exogenous variable for the GP part of this hybrid model is only WRF prediction of temperature. This model is the simplest form of a hybrid model that enables long-term forecasts, but it possesses the least amount of information about the complex environment. With such a model, the locally conditioned geographical information is contained only within the temperature variable because the model itself is trained with temperature measurements. We name this model as Hybrid 3.

Procedures for modelling and prediction using a hybrid model are as follows.

Modelling

1. Selection of the geographical point of interest.
2. Acquisition of the history of measurements from the weather-measurement station at the location of interest and measurements from stations distributed in surroundings to capture data the characterise the complexity of the terrain.
3. Acquisition of the history of atmospheric-variables forecasts from physics-based model for the location of interest.
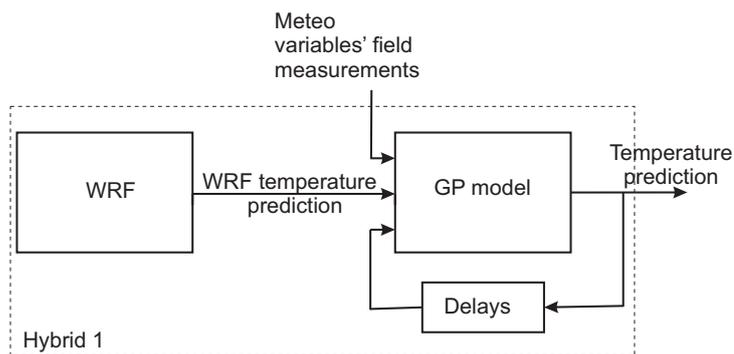
Fig. 2 Hybrid model of the WRF ARW model and statistical model trained with measurements from available measurement stations – Hybrid 1. This hybrid model requires measurements of meteorological variables for making forecasts in each time step.
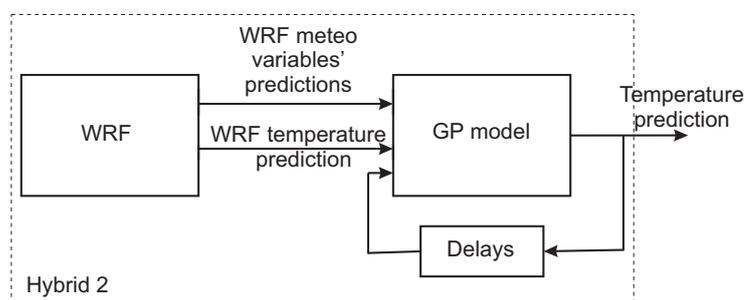
Fig. 3 Hybrid model of the WRF ARW model and statistical model where the statistical part of the model uses WRF predictions to replace measurements from measurement stations in each time step – Hybrid 2. The statistical part of this hybrid model is trained with temperature measurements and WRF predictions of all relevant meteorological variables.

4. Selecting the structure, e.g., regressors, etc., and training the statistical part of the hybrid model to improve forecasts of physics-based model to be as close to measured values.

Prediction

1. Acquisition of present and for the hybrid-model-relevant past measurements from weather stations in the location of interest and surroundings to act as some of the inputs of statistic-part of the hybrid model.
2. Acquisition of present values and forecasts from physics-based model to act like the rest of the inputs of statistic-part of the hybrid model.
3. Prediction of the hybrid-model output over a horizon of interest using acquired inputs.

Hybrid models, described in Section 3.3 for temperatures at heights of 2 m, 10 m, 40 m and 70 m are trained for one-step-ahead predictions on identification
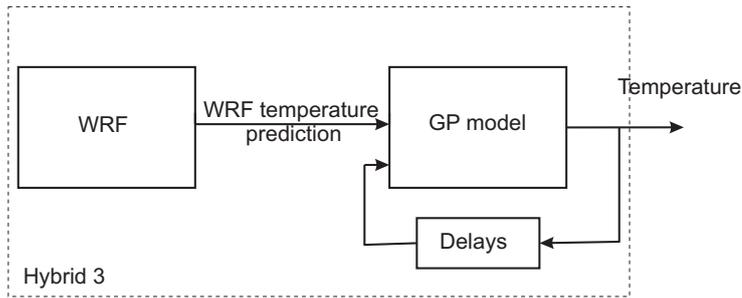
Fig. 4 Hybrid model of the WRF ARW model and statistical model using temperature-prediction time series and WRF temperature predictions only – Hybrid 3. The statistical part of this hybrid model is trained with temperature measurements and WRF predictions of temperature.

data, that is half-hour-ahead predictions, and are consequently also evaluated on validation data for the predictions. These models can be validated and provide enough information that a vertical temperature profile for low heights can be formed. They are compared with the WRF-model predictions as well as with predictions of the so-called trivial model, where predictions equal predecessor.

While these short-term predictions are useful for the selection of statistical-model regressors and to test the identified model, they do not fulfil the main purpose. The long-term models' forecasts are required to be used as inputs to the air-pollution dispersion model. The WRF model's long-term forecasts are composed of 12-hour-lasting piecewise parts as described in Section 3.1. Therefore, an option for validating a long-term forecast would be to compare one of the 12-hour-lasting WRF predictions with the hybrid models' predictions. Having in mind, however, that assessment in a longer period would reveal a more realistic picture of the hybrid models' ability and that the WRF model is also updated periodically, the decision to take a longer forecasting horizon was made. We are aware of the fact that in this way the composed WRF forecasts themselves are better than the WRF forecasts of a single, but longer run. Nevertheless, this will not influence the purpose of the assessment, which is to test whether or not the hybrid models improve the forecasts of the WRF model locally.

Consequently, long-term forecasts of hybrid models for up-to 5 days are assessed besides short-term predictions. These long-term forecasts are obtained as the hybrid-models' simulations by an iterative method, as described in Section 3.2, and used as the ultimate validation for the models of temperature at the selected heights. This means that the up-to-5-days simulation is realised so that the model is fed with input signals only, e.g., WRF forecasts in each time step for the Hybrid 3 model from the start of the validation signal until the end of the 5-day period, while output estimates are fed back instead of future observations. This way, the model output in each time step is independent of the observations of the output variables from the beginning of the simulation until the end of the 5-day period. The simulation run is, therefore, a single run from the beginning to the end of the 5-day period, but the uncertainty of the model response due to stochastic output is realised with Monte Carlo simulation as described in Section 3.2.

Our division to short-term predictions and long-term forecasts is conditioned by the available data sampling interval and the purpose of the models and can be selected differently in other applications.

4 Results

This section provides results that show a comparison between the three different hybrid models between the WRF model and the GP model for short-term predictions and long-term forecasts as explained in Section 3.2.

One of the main issues with designing a statistical model is finding relevant regressors. The model depends on relevant regressors, which should not be redundant to avoid conditioning problems. There exist several regressor-selection methods that are classified differently. One known type is the classification on wrappers or wrapper methods, embedded methods and filter methods (May et al., 2011). Several methods have been tested and, in the end, two methods were selected with a cut-and-try approach. For temperatures at heights of 2 m and 10 m, the best set of regressors is found with an exhaustive search based on log marginal likelihood as described with Equation (3), which can be classified as a wrapper method. Surprisingly, for heights of 40 m and 70 m, the best set of regressors is obtained with the Relieff method (Kononenko et al., 1997). The Relieff method is a filter method. It selects regressors based on the identification of regression value differences between nearest-neighbour instance pairs.

The final regression vector for each height level is shown in Table 1.

The covariance or kernel function used for the calculation of the covariance matrix $\mathbf{K}$ and other covariances in our model, selected due to achieving the best results on the identification data set, is the composition of the square exponential (SE) covariance function and linear covariance function with 'automatic relevance determination' (ARD) hyperparameters (MacKay, 1998; Kocijan, 2016):

$$k(\mathbf{z}_i, \mathbf{z}_j) = \sigma_f^2 \exp\left[-\frac{1}{2}(\mathbf{z}_i - \mathbf{z}_j)^{\mathrm{T}} \boldsymbol{\Lambda}_{\mathrm{SE}}^{-1}(\mathbf{z}_i - \mathbf{z}_j)\right] + \mathbf{z}_i^{\mathrm{T}} \boldsymbol{\Lambda}_{\mathrm{LIN}}^{-1} \mathbf{z}_j, \qquad (10)$$

where $\boldsymbol{\Lambda}_i^{-1}; i \in \{\mathrm{SE}, \mathrm{LIN}\}$ denotes $\boldsymbol{\Lambda}_i^{-1} = \mathrm{diag}([l_1^{-2}, \ldots, l_D^{-2}])$ of the ARD hyperparameters and σ_f^2 is the scaling factor. All hyperparameters can be written as a vector $\boldsymbol{\theta} = [l_1^{-2}, \ldots, l_{2D}^{-2}, \sigma_f^2]^T$. Other parameters used for the GP model are constant mean function and the Exact inference method with Gaussian likelihood (Rasmussen and Nickisch, 2015). Log-marginal likelihood of Equation (3) is used as the cost function for model identification.

The evaluation criterions for all experiments are normalised root-mean-square error (NRMSE), standardised mean square error (SMSE), and Pearson correlation coefficient (PCC) for evaluating the most likely values of the predictions, and mean standardised log-loss (MSLL) for evaluating the complete GP model output, i.e., prediction distributions in every time instant with their mean values and variances. The explanation of the used evaluation criterions is given in the Appendix.

Both kinds of experiments, for half-hour-predictions and up-to-5-day forecasts, have to be compared to output observations of the WRF model alone. The comparison gives information about the improvement of the WRF-model output, which is the main purpose of our investigation.

Table 1 The final regression vectors for each height level. The delay in samples explains the lag of observations and value +1 means prediction. These four regression vectors are used for training and prediction of the statistical-model part of the hybrid models for each of the investigated heights.

2 m level:

Mes. station	Variable	Delay in samples
Stolp	temperature at 2 m	1
Stolp	temperature at 2 m	0
Stolp	global solar radiation	1
Cerklje	temperature at 2 m	0
Stolp	wind speed	0
Stolp	global solar radiation	0
WRF model	temperature at 2 m	'+1'
WRF model	global solar radiation	'+1'

10 m level:

Stolp	temperature at 10 m	1
Stolp	temperature at 10 m	0
Krško	temperature at 2 m	1
Cerklje	temperature at 2 m	0
Krško	temperature at 2 m	0
Krško	wind speed	0
WRF model	temperature at 2 m	'+1'

40 m level:

Stolp	temperature at 40 m	1
Stolp	temperature at 40 m	0
Brežice	relative humidity	1
Cerklje	temperature at 2 m	1
Cerklje	relative humidity	1
Cerklje airport	temperature at 2 m	1
Krško	wind speed	1
Krško	wind direction	1
Libna	temperature at 2 m	1
Stolp	relative humidity	1
WRF model	wind speed	'+1'
WRF model	temperature at 2 m	'+1'

70 m level:

Stolp	temperature at 70 m	1
Stolp	temperature at 70 m	0
Cerklje	temperature at 2 m	1
Stolp	global solar radiation	0
WRF model	temperature at 2 m	'+1'

Figure 5 shows a scatter plot of the WRF model's half-hour-ahead predictions in comparison with the measurements at the location of interest on the data for validation.

Figure 6 shows scatter plots of the hybrid models from Figure 2 for half-hour-ahead predictions at different heights in comparison with the measurements that were not used for the identification of the statistical model at the location of interest. The hybrid model is a combination of the WRF model and selected measurements as regressors. It is clear from Figures 5 and 6 that predictions of hybrid models from Figure 2 for all four heights have substantially less variations than the measurements. The performance evaluation with the criterions used for each type of hybrid model shown in Figures 2 to 4 as presented in the previous section is given in Table 2 and confirms substantial improvements in the quality of the predictions. The performance measures are selected to provide assessment
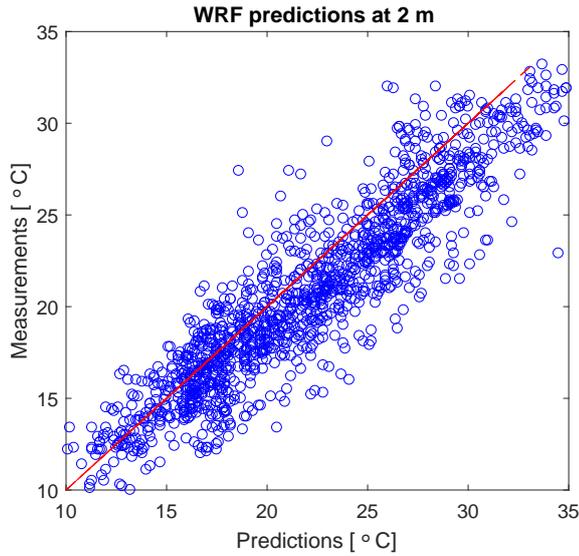
Fig. 5 WRF temperature forecasts (blue circles) vs. temperature measurements for the period of 30 days at the height of 2 m in the location of Krško NPP. The red line shows the line of the ideal match between forecasts and measurements.

of the most likely prediction values, i.e., square-error-based measures, as well as of predicted random variables, i.e., log-loss-based measure.

However, the real test of purpose fulfilment is the forecast for a longer period. In our case, a few-hours-ahead forecast is obligatory for successful evacuation planning, but any further forecasts are preferred as well. The obtained models are evaluated with forecasts for the period up to 5 days, which, at least in our case, can be considered as long-term forecasts. These forecasts are obtained with Monte Carlo simulation (Korn, 2007) based on the iterative method as described in Section 3.2 with feeding back delayed output predictions to prediction model inputs. 1000 samples of prediction distributions are used in the Monte Carlo simulations. Figure 7 shows a time series of WRF forecasts for the period of up to 5 days at a height of 2 m in comparison with the measurements at the location of interest.

Figure 8 shows a time series of forecasts for the hybrid models of Figure 2 at different heights for the period of up to 5 days in comparison with the measurements that were not used for identification, i.e., data for validation. Again, this hybrid model is a combination of the WRF model and selected measurements as regressors. It is expected that the WRF-model forecasts are better at higher altitudes due to the decreased influence of complex terrain. On the other hand, the accuracy of the hybrid models depends on the selection of regressors and the identification data. The best results in our case are demonstrated at 10 m height, otherwise the higher the height, the lower the accuracy. However, the 95 % confidence band provides information about the variance of the distribution. It is clear from Figure 8 that 95 % of the absolute forecast errors, in comparison with measurements not used for model identification, are within the confidence band. This information about forecast accuracy that is inherent to Bayesian modelling meth-
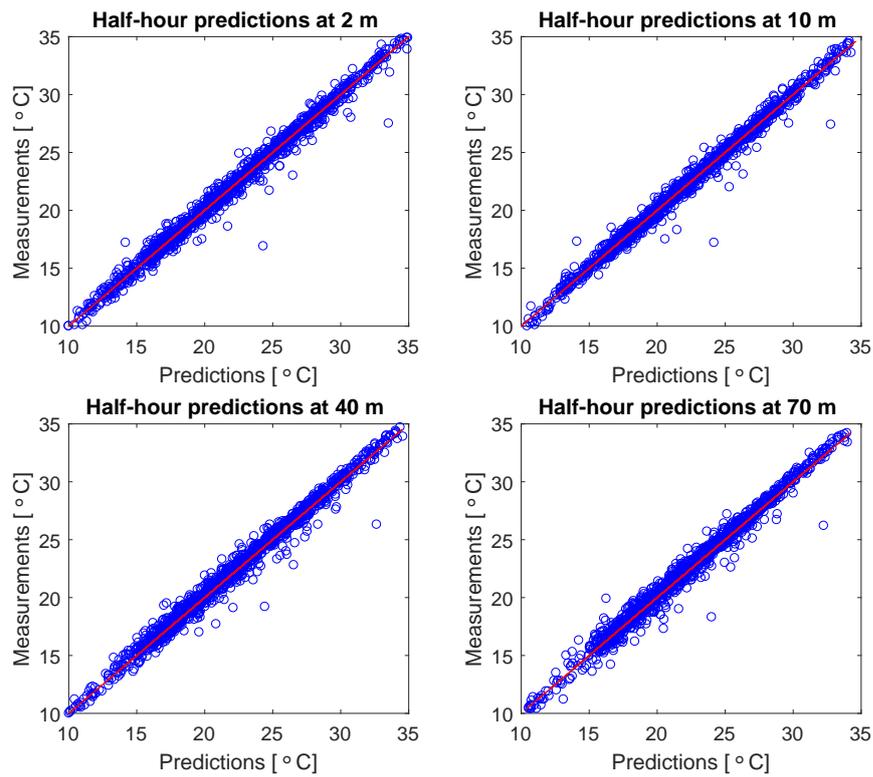
Fig. 6 Comparison of scatter plots for temperature half-hour-ahead predictions vs. temperature measurements for the period of 30 days at heights 2 m, 10 m, 40 m and 70 m, respectively with the hybrid model of Figure 2. The red line shows the line of ideal match between hybrid-model temperature forecasts and temperature measurements.

ods, to which GP modelling belongs, is very valuable when pollution dispersion is predicted.
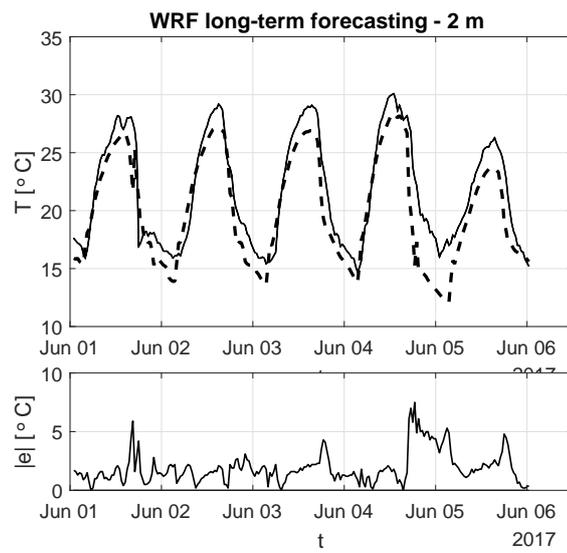
Fig. 7 WRF forecasts for the period of up-to 5 days at the height of 2 m (full line – upper figure) and measurements (dashed line – upper figure). Absolute values of the difference between forecasts and the measurements are shown in the bottom figure.
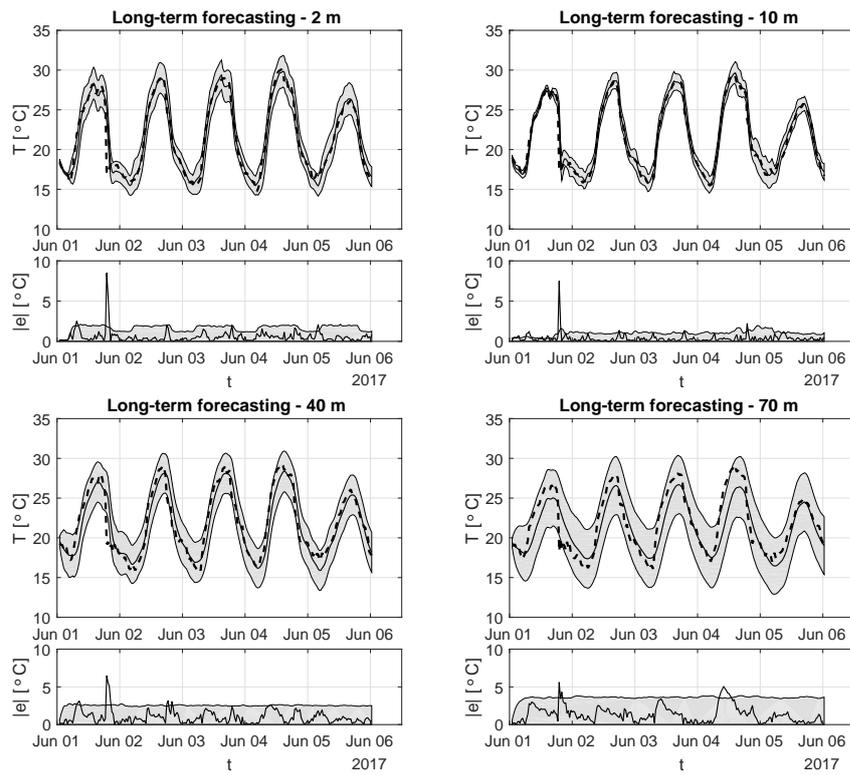
Fig. 8 Comparison of time responses for long-term forecasts of temperature for the period up-to 5 days at heights 2 m, 10 m, 40 m and 70 m, respectively. Legend: mean values of forecasts – full line in upper figure, 95 % confidence interval – grey band in upper figure, measurements – dashed line in upper 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.

The performance evaluation with the criteria used for each type of hybrid model as shown in Figures 2 to 4 is given in Tables 2 and 3. The expressions and explanations of performance measures for evaluation are given in Appendix A.

Table 2 Performance measures, explained in Appendix A, for short-term predictions of temperature are the normalised root-mean-square error (NRMSE), the standardised mean-squared error (SMSE), the Pearson correlation coefficient (PCC), the mean standardised log loss (MSLL) for the following models: WRF model, hybrid model with measurement updates in each step – Hybrid 1 (Figure 2), hybrid model with WRF supplements in each step – Hybrid 2 (Figure 3), hybrid model containing WRF temperature observations – Hybrid 3 (Figure 4), and trivial model.

		Half-hour predictions			
		2 m	10 m	40 m	70 m
NRMSE	WRF model	0.513	0.479	0.380	0.364
	Hybrid 1	**0.895**	**0.898**	**0.882**	0.870
	Hybrid 2	0.888	0.886	0.858	**0.871**
	Hybrid 3	0.890	0.893	0.878	0.868
	Trivial model	0.843	0.850	0.847	0.842
SMSE	WRF model	0.206	0.273	0.415	0.489
	Hybrid 1	**0.011**	**0.010**	**0.014**	**0.016**
	Hybrid 2	0.012	0.013	0.019	**0.016**
	Hybrid 3	0.012	0.011	0.015	0.017
	Trivial model	0.025	0.023	0.023	0.025
PCC	WRF model	0.922	0.900	0.863	0.846
	Hybrid 1	**0.995**	**0.995**	**0.993**	**0.992**
	Hybrid 2	0.994	0.994	0.991	**0.992**
	Hybrid 3	0.994	**0.995**	**0.993**	**0.992**
	Trivial model	0.988	0.989	0.988	0.988
MSLL	WRF model	\	\	\	\
	Hybrid 1	-2.178	**-2.352**	**-2.130**	-2.044
	Hybrid 2	-2.143	-2.165	-1.942	18.528
	Hybrid 3	**-2.263**	-2.266	-2.106	**-2.055**
	Trivial model	\	\	\	\

Table 3 Performance measures for long-term forecasts of temperature are the normalised root-mean-square error (NRMSE), the standardised mean-squared error (SMSE), the Pearson correlation coefficient (PCC), the mean standardised log loss (MSLL) for the following models: WRF model, hybrid model with measurement updates in each step – Hybrid 1 (Figure 2), hybrid model with WRF supplements in each step – Hybrid 2 (Figure 3), hybrid model containing WRF temperature observations – Hybrid 3 (Figure 4), and trivial model.

		Long-term forecasts up-to 5 days			
		2 m	10 m	40 m	70 m
NRMSE	WRF model	0.500	0.478	0.400	0.401
	Hybrid 1	**0.795**	**0.844**	**0.656**	**0.427**
	Hybrid 2	0.677	0.608	0.562	0.381
	Hybrid 3	0.559	0.525	0.216	-0.374
SMSE	WRF model	0.249	0.320	0.475	0.558
	Hybrid 1	**0.042**	**0.025**	**0.107**	**0.215**
	Hybrid 2	0.106	0.125	0.171	0.231
	Hybrid 3	0.183	0.195	0.407	0.676
PCC	WRF model	0.934	0.916	0.889	0.890
	Hybrid 1	**0.979**	**0.988**	**0.946**	**0.904**
	Hybrid 2	0.958	0.942	0.911	0.897
	Hybrid 3	0.931	0.921	0.849	0.778
MSLL	WRF model	\	\	\	\
	Hybrid 1	**-1.651**	**-1.773**	**-1.131**	**-0.774**
	Hybrid 2	-1.110	-1.011	-0.623	-0.732
	Hybrid 3	6.517	0.630	0.155	1.467

Hybrid models that do not encompass measurements from neighbouring stations are at a loss as it can be seen from Tables 2 and 3. Again, the models are worse at higher altitudes. Nevertheless, it is clear that especially the hybrid models with WRF supplements instead of measurements provide useful utility. The use of hybrid models with supplements is an improvement for short-term predictions as well as for long-term forecasts, except for 70 m heights, over the use of the WRF model's forecasts at the geographical point of interest, i.e., the location of Krško NPP. However, it is clear from Tables 2 and 3 that the hybrid model using WRF supplements (Figure 3) does not overperform the WRF model at 70 m height. Moreover, the hybrid model using the temperature-prediction time series (Figure 4) does not overperform the WRF model at 40 m height, all in the case of forecasts for the period of up-to 5 days.

The graphical presentation of the assessed-models' performance measures depending on height is given in Figure 9 for half-hour predictions and in Figure 10 for long-term forecasts.
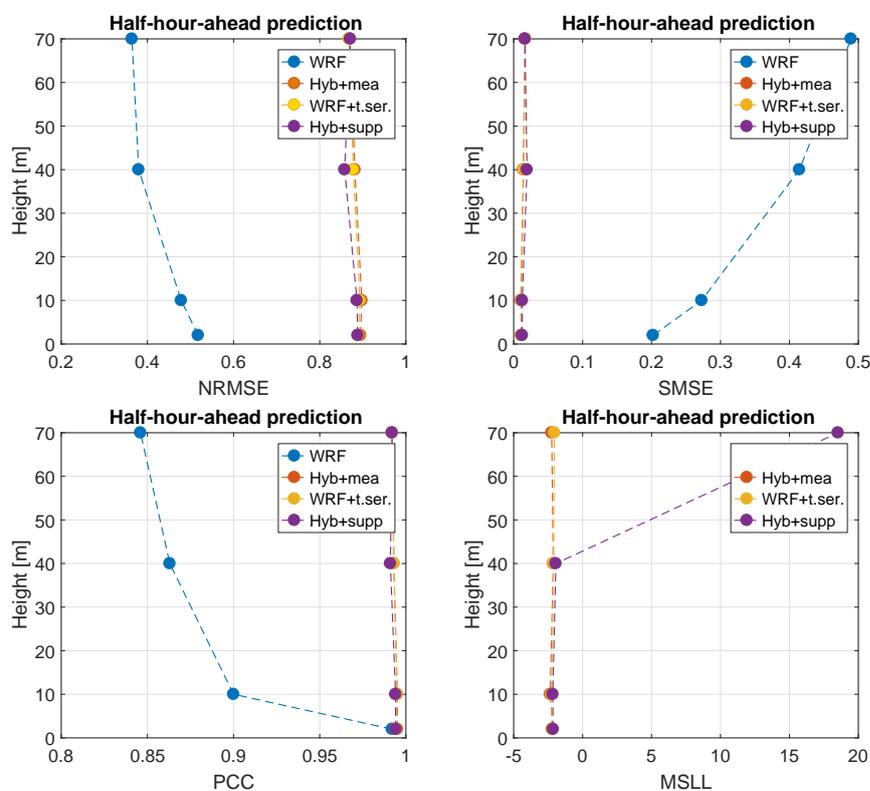
Fig. 9 Graph for comparison of performance-measure values for half-hour-ahead predictions vs. investigated heights. The performance-measure values are depicted as solid dots, while interpolation lines are depicted with dashed line to illustrate the trend of performance-measure value changes with height.
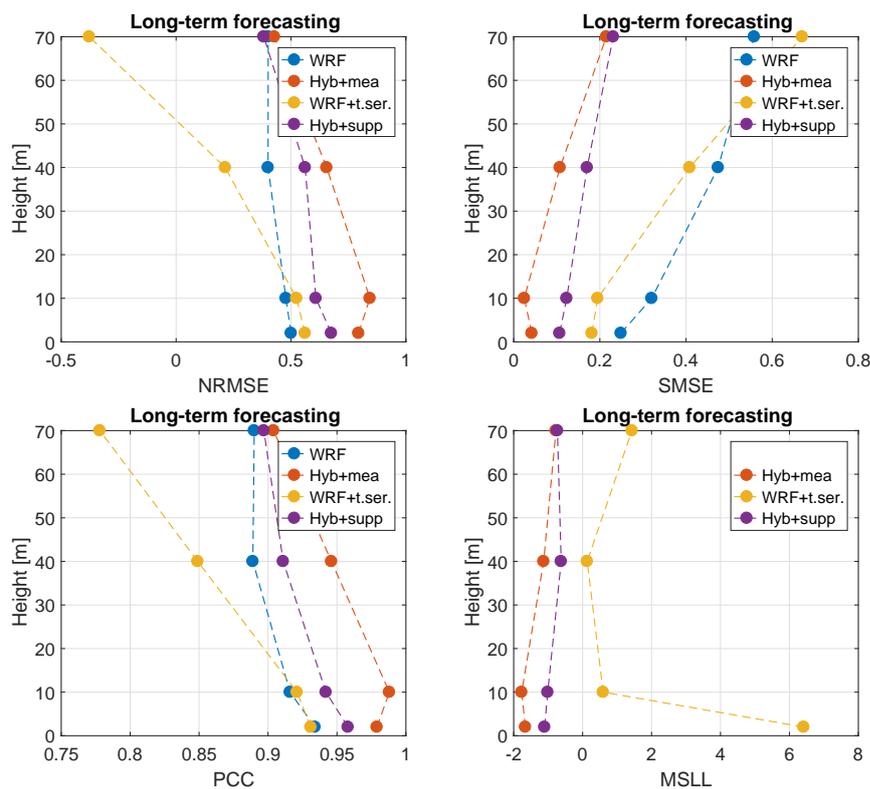
Fig. 10 Graph for comparison of performance measures for long-term forecast up-to 5 days vs. investigated heights. The performance-measure values are depicted as solid dots, while interpolation lines are depicted with dashed line to illustrate the trend of performance-measure value changes with height.

Figures 9 and 10 illustrate data from Tables 2 and 3 and confirm a deterioration of the results with height, which is the expected result.

It is important to emphasise that investigated models predict vertical temperature profile in one geographical point only. As such they cannot be used for the assessment of the spatial distribution of modelling errors in a wider geographical area. This can be done using a set of models each containing model describing the situation at a different location and possibly at different terrain.

5 Conclusions

This paper described an empirical assessment of the integration of a physics-based and a statistical, i.e., first-principles and empirical, model for vertical-temperature-profile modelling at low heights. The shown short-term predictions and long-term forecasts demonstrate a successful use of the integration of a physics-based and a statistical model for vertical-temperature-profile modelling at low heights. In the presented example, a WRF model and a GP model are integrated. The improvements over WRF model predictions alone are large enough that a noticeable

difference is expected when using hybrid-model predictions for radiation dispersion forecasting.

The innovation of the presented hybrid approach is the use of a statistical model to enhance the prediction power of a physics-based model. Such a method is useful in situations where the physics-based model is not accurate enough and its further development would be too expensive while measurement data for the locality is readily available. The use of a dynamic statistical model in a hybrid model with a dynamic physics-based model is a step beyond the traditional static statistical postprocessing of physics-based model forecasts. The described approach is not constrained to the study area and may have a general utility for any complex terrain.

The main disadvantage of the presented method is that the developed model is constrained to the location for which it is made.

Three different types of hybrid models, each one having a different amount of measurements available, are evaluated in the investigation. All three provide improvements over the physics-based model.

The investigation taught us several lessons.

- Caution needs to be exercised when such models are used for longer-period forecasts. The improvement depends on the amount of measurement information and decays with the increasing horizon of forecasting.
- Modelling of the vertical temperature profile as described in this investigation is a big-data problem and has to be handled as such. Options for more efficient treatment of the big data are online learning (Kocijan et al., 2016) and the use of High-Performance Computing (Matthews et al., 2017).
- It is worth putting a lot of effort into the selection of regressors.
- On the positive note, GP modelling is a valuable tool due to the information it provides about prediction confidence, which is of particular importance when spatial variability of the forecasts is needed.

Modelling of other variables that are necessary for pollution-dispersion modelling and forecasting is planned for the immediate future.

References

Božnar MZ, Mlakar P, Grašič B (2012) Short-term fine resolution WRF forecast data validation in complex terrain in Slovenia. International journal of environment and pollution 50(1-4):12–21

Breznik B, Božnar MZ, Mlakar P, Tinarelli G (2003) Dose projection using dispersion models. International journal of environment and pollution 20(1-6):278–285

Emeis S, Schäfer K, Münkel C, Friedl R, Suppan P (2012) Evaluation of the interpretation of ceilometer data with RASS and radiosonde data. Boundary-layer meteorology 143(1):25–35

Fochesatto GJ (2015) Methodology for determining multilayered temperature inversions. Atmospheric Measurement Techniques 8(5):2051–2060

Gangwar RK, Mathur AK, Gohil B, Basu S (2014) Neural network based retrieval of atmospheric temperature profile using AMSU-A observations. International Journal of Atmospheric Sciences 2014

Gradišar D, Grašič B, Božnar MZ, Mlakar P, Kocijan J (2016) Improving of local ozone forecasting by integrated models. Environmental Science and Pollution Research 23(18):18439–18450

Grašič B, Božnar MZ, Mlakar P, Kocijan J (2019) Validation of numerically forecasted vertical temperature profile with measurements for dispersion modelling. International journal of environment and pollution To appear

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

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

Kocijan J, Gradišar D, Božnar MZ, Grašič B, Mlakar P (2016) On-line algorithm for ground-level ozone prediction with a mobile station. Atmospheric Environment 131:326–333

Kocijan J, Gradišar D, Stepančič M, Božnar MZ, Grašič B, Mlakar P (2018) Selection of the data time interval for the prediction of maximum ozone concentrations. Stochastic Environmental Research and Risk Assessment 32(6):1759–1770

Kononenko I, Šimec E, Robnik-Šikonja M (1997) Overcoming the myopia of inductive learning algorithms with RELIEFF. Applied Intelligence 7(1):39–55

Korn GA (2007) Advanced dynamic-system simulation: model-replication techniques and Monte Carlo simulation. John Wiley & Sons

MacKay DJ (1998) Introduction to Gaussian processes. NATO ASI Series F Computer and Systems Sciences 168:133–166

Matthews DG, Alexander G, Van Der Wilk M, Nickson T, Fujii K, Boukouvalas A, León-Villagrá P, Ghahramani Z, Hensman J (2017) GPflow: A Gaussian process library using TensorFlow. The Journal of Machine Learning Research 18(1):1299–1304

May R, Dandy G, Maier H (2011) Review of input variable selection methods for artificial neural networks. In: Suzuki K (ed) Artificial Neural Networks - Methodological Advances and Biomedical Applications, InTech, Rijeka, pp 19–44

Ramaswamy V, Hurrell J, Meehl G, Phillips A, Santer B, Schwarzkopf M, Seidel D, Sherwood S, Thorne P, Karl T, et al. (2006) Why do temperatures vary vertically (from the surface to the stratosphere) and what do we understand about why they might vary and change over time? In: Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences, vol Synthesis and Assessment Product 1.1, US Climate Change Science Program//Subcommittee on Global Change Research, Washington, D.C., chap 1, pp 15–28

Rasmussen CE, Nickisch H (2015) The GPML toolbox version 3.6

Rasmussen CE, Williams CK (2006) Gaussian processes for machine learning. MIT press Cambridge

Rekhy A, Shneider MN, Miles RB (2018) Temperature profiling of the atmosphere by filtered rayleigh scattering. In: 2018 Aerodynamic Measurement Technology and Ground Testing Conference, p 3630

Ren WW, Yang T, Huang CS, Xu CY, Shao QX (2018) Improving monthly streamflow prediction in alpine regions: integrating hbv model with bayesian neural network. Stochastic Environmental Research and Risk Assessment 32(12):3381–3396

Shi JQ, Choi T (2011) Gaussian process regression analysis for functional data. CRC Press

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker M, Duda KG, Huang XY, Wang W, Powers JG (2008) A description of the advanced research WRF version 3. Tech. rep., National Center for Atmospheric Research

Von Stosch M, Oliveira R, Peres J, de Azevedo SF (2014) Hybrid semi-parametric modeling in process systems engineering: Past, present and future. Computers & Chemical Engineering 60:86–101

Worsnop RP, Scheuerer M, Hamill TM, Lundquist JK (2018) Generating wind power scenarios for probabilistic ramp event prediction using multivariate statistical post-processing. Wind Energy Science 3(1):371–393

Zhang Y, Bocquet M, Mallet V, Seigneur C, Baklanov A (2012a) Real-time air quality forecasting, part I: History, techniques, and current status. Atmospheric Environment 60:632 – 655

Zhang Y, Bocquet M, Mallet V, Seigneur C, Baklanov A (2012b) Real-time air quality forecasting, part II: State of the science, current research needs, and future prospects. Atmospheric Environment 60:656 – 676

A Performance measures

We use the following statistical measures for the assessment:

- The normalised root-mean-square error – NRMSE

$$\text{NRMSE} = 1 - \frac{\|\mathbf{y} - \mu\|^2}{\|\mathbf{y} - E(\mathbf{y})\|^2}, \tag{11}$$

    where
      $\mathbf{y}$ – the vector of validation values,
      μ – the vector of mean predicted values,
      $E(\mathbf{y})$ – the mean value of $\mathbf{y}$.
    NRMSE is 1 for a perfect match and $-\infty$ for a very bad match of the validation and mean predicted values.

- The standardised mean-squared error – SMSE

$$\text{SMSE} = \frac{1}{N} \frac{\sum_{i=1}^{N} (E(\hat{y}_i) - y_i)^2}{\sigma_y^2}, \tag{12}$$

    where
      σ_y^2 – the variance of the observations.
    SMSE is a frequently used standardised measure for the accuracy of predictions' mean values with values between 0 and 1, where the value 0 is the result of a perfect model.

- The Pearson's correlation coefficient - PCC:

$$\text{PCC} = \frac{\sum_{i=1}^{N} (E(\hat{y}_i) - E(\hat{\mathbf{y}}))(y_i - E(\mathbf{y}))}{N \sigma_y \sigma_{\hat{y}}}, \tag{13}$$

    where
      $E(\hat{\mathbf{y}})$ – the expectation, i.e., the mean value, of the vector of predictions,
      σ_y – the standard deviation of the observations,

$\sigma_{\hat{y}}$ – the standard deviation of the predictions.

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 of 1.

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

$$
\begin{aligned}
\text{MSLL} = {} & \frac{1}{2N} \sum_{i=1}^{N} \left[\ln(\sigma_i^2) + \frac{(E(\hat{y}_i) - y_i)^2}{\sigma_i^2} \right] \\
& - \frac{1}{2N} \sum_{i=1}^{N} \left[\ln(\sigma_y^2) + \frac{(y_i - E(\mathbf{y}))^2}{\sigma_y^2} \right],
\end{aligned}
\tag{14}
$$

where

σ_i^2 – the prediction variance in the i-th step,

$E(\mathbf{y})$ – 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 weights 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.