ISEIS

Journal of Environmental Informatics 41(2) 88-103 (2023)

Journal of Environmental Informatics

www.iseis.org/jei

Hybrid Forecasting of Wind for Air Pollution Dispersion over Complex Terrain

M. Perne[1, 2 *], J. Kocijan[1, 3], M. Z. Božnar[4], B. Grašič[4], and P. Mlakar[4]

[1] *Jožef Stefan Institute, Ljubljana 1000, Slovenia*
[2] *Quantectum, prognoza potresov, d.o.o., Ljubljana 1000, Slovenia*
[3] *University of Nova Gorica, Nova Gorica 5000, Slovenia*
[4] *MEIS d.o.o., Mali Vrh pri Šmarju, Šmarje-Sap 1293, Slovenia*

Received 15 August 2019; revised 23 November 2021; accepted 10 May 2022; published online 15 February 2023

ABSTRACT. In case of an unplanned emission event from a nuclear power plant, the local population can be protected more efficiently when valid atmospheric dispersion model results are available. Atmospheric dispersion models use local meteorological variables as inputs. When atmospheric dispersion in the future is being predicted, a forecast of the local meteorological variables is needed. The most important variable in atmospheric dispersion modelling is wind, and accurately predicting ground level winds presents a challenge to numerical weather prediction models. We therefore develop hybrid models for forecasting local ground level wind at a single location where the terrain is complex and the average wind is weak and fluctuating. Wind speed and direction are modelled as west-east and south-north wind components. Each model is composed of a numerical weather prediction model and a Gaussian process statistical model that uses numerical weather predictions as some of its inputs and is trained on historical data to predict the output component. The most advanced Gaussian process models studied are of Gaussian process nonlinear autoregressive model with exogenous input (GP-NARX) type. In addition to numerical weather predictions, they also use local meteorological variables, including the output variable, as their inputs. Numerical weather prediction results based on large scale information and fundamental knowledge of the system are thus supplemented by local measurements that better reflect the effects of local topography and land use. The models are tested by prediction and simulation. The wind components predicted by more advanced models are more accurate than raw or post-processed numerical weather prediction results. As an example, a model predicting 2D wind vector 1.5 h in advance achieves a NRMSE of 0.214 if it uses all the immediately available information. This is better than both the persistence model with NRMSE of 0.188 and post-processed NWP with NRMSE of 0.164. This demonstrates that hybrid modelling provides the best weather information for short-term and medium-term atmospheric dispersion forecasting. While the method is motivated by nuclear emission sources, it could also be applied to other pollution.

Keywords: atmospheric dispersion model, dynamic systems, Gaussian process, hybrid model, system identification, wind forecasting

1. Introduction

Atmospheric dispersion models calculate pollutant concentration fields as a function of time based on topography, land use, emission source, and weather information. They require local meteorological variables that can be either measured or forecasted. Dispersion of radioactive pollutants hypothetically emitted from Krško Nuclear Power Plant (NPP) is studied before the emission. The goal is to facilitate protection of the inhabitants in the surroundings by having the radioactive plume dispersion results available as soon as an unplanned emission from a nuclear power plant happens. For the purpose, knowledge of certain future local meteorological variables is necessary as they are required for atmospheric dispersion modelling. In particular, short-term (30 min to 6 h ahead) and medium-term (6 h to 1 day ahead) (Wang et al., 2016) forecasts are desired. Very short-term forecasts (under 30 min ahead) do not give enough time to react, while long-term ones (over 1 day ahead) would be less accurate so recalculating at a later time would be beneficial.

Studies on dispersion of emissions with measured meteorological conditions and prescribed atmospheric dispersion models are required for environmental impact assessments prior to construction. As specified in the EU Directive 2008/1/EC, Council Directive 84/360/EEC of 28 June 1984 on the combating of air pollution from industrial plants (6) introduced a general framework requiring authorisation prior to any operation or substantial modification of industrial installations which may cause air pollution (EuropeWan Union, 2008). These assessments are performed for past measured local meteorological variables and thus do not need weather forecasting. Their results do not apply to the immediate future in case of an accident. Predictions of local meteorological variables which are not required for permits and licenses are therefore highly beneficial in case of an accident.

There are several ways of forecasting the local meteorological variables. One is persistence method – using the current measured values as the forecast. This method of prediction is good for short term forecasting on flat terrain (Potter and Neg-

* Corresponding author. Tel.: +386 1-4773800; fax: +386 5-6205200.
E-mail address: matija.perne@ijs.si (M. Perne).

ISSN: 1726-2135 print/1684-8799 online
 doi:10.3808/jei.202300489.

nevitsky, 2006) because the weather is changing gradually. For intermediate term forecasts several hours to a day in advance, it will be demonstrated that physics-based, statistical, or hybrid models that give better results can be constructed.

Physics-based numerical weather prediction (NWP) models that are used in forecasting are refined from global prognostic models. In our case, the NWP model for the region of interest is run by WRF-ARW version 3.4.1 (Skamarock et al., 2008). NWP forecasts can be used in place of predictions of local meteorological variables for atmospheric dispersion modelling. Models of this kind have been developed for Switzerland (Testa et al., 2013), Germany (Von Arx et al., 2014) and Slovenia (Mlakar et al., 2019) and validated on field data in Slovenia (Grašič et al., 2019). However, NWP models have certain limitations. Their spatial resolution is limited, and the details of topography on a scale smaller than the resolution are neglected (Wagenbrenner et al., 2016; Yao et al., 2017). In forecasting local meteorological variables, there is thus room for improvement on NWP results, in particular on smaller local scales.

To obtain a statistical model, patterns are identified using machine learning, enabling one to predict the value of the output variable from the available input information. An example from weather forecasting is a model that forecasts future local meteorological variables based on the current and past measurements (Božnar and Mlakar, 1995; Božnar et al., 2017). No explicit knowledge about the dynamical system that generates the observed signals is used in building such a model. The dynamics is instead identified from training data that is composed of historical measured signals.

A hybrid, integrated, or statistical post-processing model is formed by using outputs of a NWP model as some of the inputs of a statistical model (Neal et al., 2014). In comparison with pure NWP models, the advantages of a hybrid model for our use case are that it predicts local meteorological variables and that it utilizes the available local measurements that reflect small-scale topography by using them as inputs. It applies the results of the numerical weather prediction model that is based on physics knowledge and on the information on the large scale state of the weather system, giving it an edge over pure statistical models to which this information is not available.

Since wind is the most influential variable in atmospheric dispersion modelling (Breznik et al., 2003; Beelen et al., 2010; Barratt, 2013), and it is challenging to predict with NWP (Wagenbrenner et al., 2016; Yao et al., 2017), we focus on wind speed and direction forecasting. The methods used are universal and can be applied to other local meteorological variables as well.

Hybrid methods for wind forecasting have often been used in wind power forecasting (Okumus and Dinler, 2016). Gaussian process (GP) models have been used in wind power forecasting as well (Mori and Kurata, 2008; Wang and Hu, 2015; Hoolohan et al., 2018). However, our study significantly differs from those in the character of the field site and in the goal of modelling. Wind power forecasting is concerned with the wind speed, while in atmospheric dispersion modelling, wind direction is as important as wind speed. Wind power forecasting is mostly used in flat terrain with strong persistent winds (Lei et al., 2009). In contrast, we are investigating an example of a complex terrain with changeable wind and low average wind speed, see Figures 2 and 3. For pollution dispersion modelling, changeable and low wind speed conditions are more demanding than persistent strong wind conditions (Carvalho and de Vilhena, 2005).

We choose to model the wind with GP models including Gaussian process nonlinear autoregressive models with exogenous input (GP-NARX) (Kocijan, 2016). Compared to methods such as artificial neural networks and support vector machines, which have also been used in hybrid wind forecasting (Okumus and Dinler, 2016), GP is another universal approximator (Deisenroth et al., 2012) distinguished by providing the output probability distribution. The article is focused on the statistical part of the wind predicting hybrid model. Regressors used in wind prediction are selected, where a regressor is a model input and is either a signal or a time-shifted signal to encompass the system dynamics. The system is identified on a subset of available data and the model is tested on a subset that is disjunct from the identification subset. Different figures of merit are evaluated. Several variations of the model are compared to each other and to models with different structures.

In summary, the knowledge gap addressed in this study is hybrid modelling of wind outside of flat terrain with persistent wind. Unlike in the wind power forecasting, in atmospheric dispersion modelling the model output is of great interest also when the wind speed is low. The presented methods are not limited to radiological pollution. They are also applicable to other sites with different types of emission sources. Our results are intended to be used with Lagrangian particle tracing dispersion models (Tinarelli et al., 2000).

2. Methods

2.1. Overview

WRF-ARW version 3.4.1-based NWP model is available for the area of interest. The terrain is complex as can be seen in Figure 1. In the area of 30 km by 30 km centred on the Stolp weather station, the minimum elevation is 131 m and the maximum elevation is 1019 m. CORINE Land Cover (European Union, 2019) shows that 35% of this area is covered by broad-leaved forest, 21% by complex cultivation patterns, 11% is land principally occupied by agriculture with significant areas of natural vegetation, 9% non-irrigated arable land, 8% mixed forest, 5% vineyards, and the rest is other land covers such as urban areas, water, other natural vegetation and other agricultural uses. There are 6 meteorological stations: Stolp at Krško NPP, Brežice, Cerklje, Cerklje Airport, Krško, and Lisca, their locations are listed in Table 1 and shown in Figure 1(a). At Stolp weather station, vertical profile measurements obtained using sensors on a tower, radio acoustic sounding system (RASS), and sonic detection and ranging (SODAR), are available. There is also an atmospheric dispersion model that requires local meteorological variables as inputs.

Hybrid models predicting local meteorological variables are developed. Each hybrid model consists of the NWP model

and a statistical part. The inputs of the statistical part, the regressors, are output signal samples of the NWP and measurements of the local meteorological variables, both possibly time-shifted to include the system dynamics. Its output is a prediction of a local meteorological variable. The prerequisite for building such a model is the availability of training data, that is, of historical measurements and NWP-produced signals that represent the inputs and the outputs of the statistical part of the model. A hybrid model uses the knowledge of the physical state of the atmosphere and the effects of large scale atmospheric circulation from the NWP model, the influence of small-scale local topography deduced statistically from the historical measurements, and the local weather information from real-time measurements at the meteorological stations.

(a) Overview map

(b) View from the tower

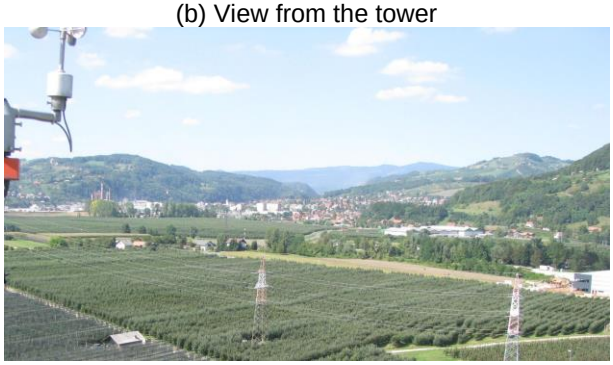

Figure 1. (a) A 40 km × 40 km overview map of the area with the meteorological stations. The Stolp station is in the NW part of the NPP. (b) View from the tower at Stolp weather station towards NW.

This article is focused on the statistical part of the hybrid model, which is in our example a GP model, structured as a GP-NARX in the most advanced example. We predict ground level wind as it is more challenging for NWP to predict than upper level winds and we demonstrate the model for the Stolp weather station location.

Table 1. Locations of Meteorological Stations

Name	UTM grid zone 33T		WGS84	
	East	North	Latitude	Longitude
Stolp at Krško NPP	539776	5087498	45.939900	15.513132
Brežice	546266	5083861	45.906760	15.596502
Cerklje	540614	5081216	45.883312	15.523411
Cerklje Airport	540035	5083159	45.900833	15.516111
Krško	538593	5088899	45.952577	15.497984
Lisca	522034	5101613	46.067735	15.284905

Table 2. Available Signals

Name	2W	T	H	P	Other	N
Stolp	✓	✓	✓	✓	Temperature 10, 40, and 70 m above ground, humidity 10 and 40 m above ground, global solar radiation	11
Brežice	✓	✓	✓			4
Cerklje		✓	✓	✓		3
Cerklje Airport	✓	✓	✓	✓		5
Krško	✓	✓	✓			4
Lisca	✓	✓	✓	✓		5
SODAR					Lowest 5 layers used of 24 total, 3 wind components each	15
NWP+ANN	✓	✓	✓	✓	Global solar radiation, cloudiness, diffuse solar radiation	8
Math					Sin and cos with period 1 day, 1 year	4

Note: A tick in "2W" column means that 2 wind components 10 m above ground are available, "T" stands for air temperature 2 m above ground, "H" is relative humidity 2 m above ground, and "P" is air pressure 2 m above ground. The total number of signals from the station is labelled "N".

2.2. Signals

The local meteorological variable signals that are taken into consideration are listed in Table 2. They are the variables measured at the 6 meteorological stations – 2 horizontal wind components, temperature, relative humidity, air pressure, not all of them at every station, all at ground level except at Stolp station – the SODAR wind components, the numerical weather predictions, and signals corresponding to time of day and season. There are 47 measured signals, of which 32 are measured in situ and 15 are measured remotely as a vertical profile with SODAR, 8 signals from the relevant discretization cell of the NWP model, and 4 time signals, giving a total of 59 signals. Of the NWP signals, 7 are computed directly and 1, diffuse solar radiation, is estimated with an artificial neural network (ANN). The 4 computed signals that represent the time are cos(hour of day × $2\pi/24$), sin(hour of day × $2\pi/24$), cos(day of year × $2\pi/235.25$),

sin(day of year × 2π/365.25). 6 years of the data from the years 2012 ~ 2017 is available.

The quantities we are predicting are west-east and south-north components of ground level horizontal wind, 10 m above ground, at Stolp weather station. A separate model for each one of the components is developed. The predicted components can be converted into horizontal wind speed and direction.

All the meteorological variables used in the model are sampled every Δt = 1/2 h and represent averages over the whole Δt for the measured variables and snapshot values for NWP + ANN and mathematical signals. The ground level wind component signals at Stolp and NWP predictions of the same quantity are shown in Figures 2 and 3.

2.3. Regressor Selection

The statistical model is identified through supervised learning (Bishop, 2006), the relationship between the regressors and the output variable is inferred from the training data. It is then applied to new instances of inputs to predict the output. An additional regressor may in principle give the model more information and improve the prediction. However, there are several reasons for keeping the number of regressors small.

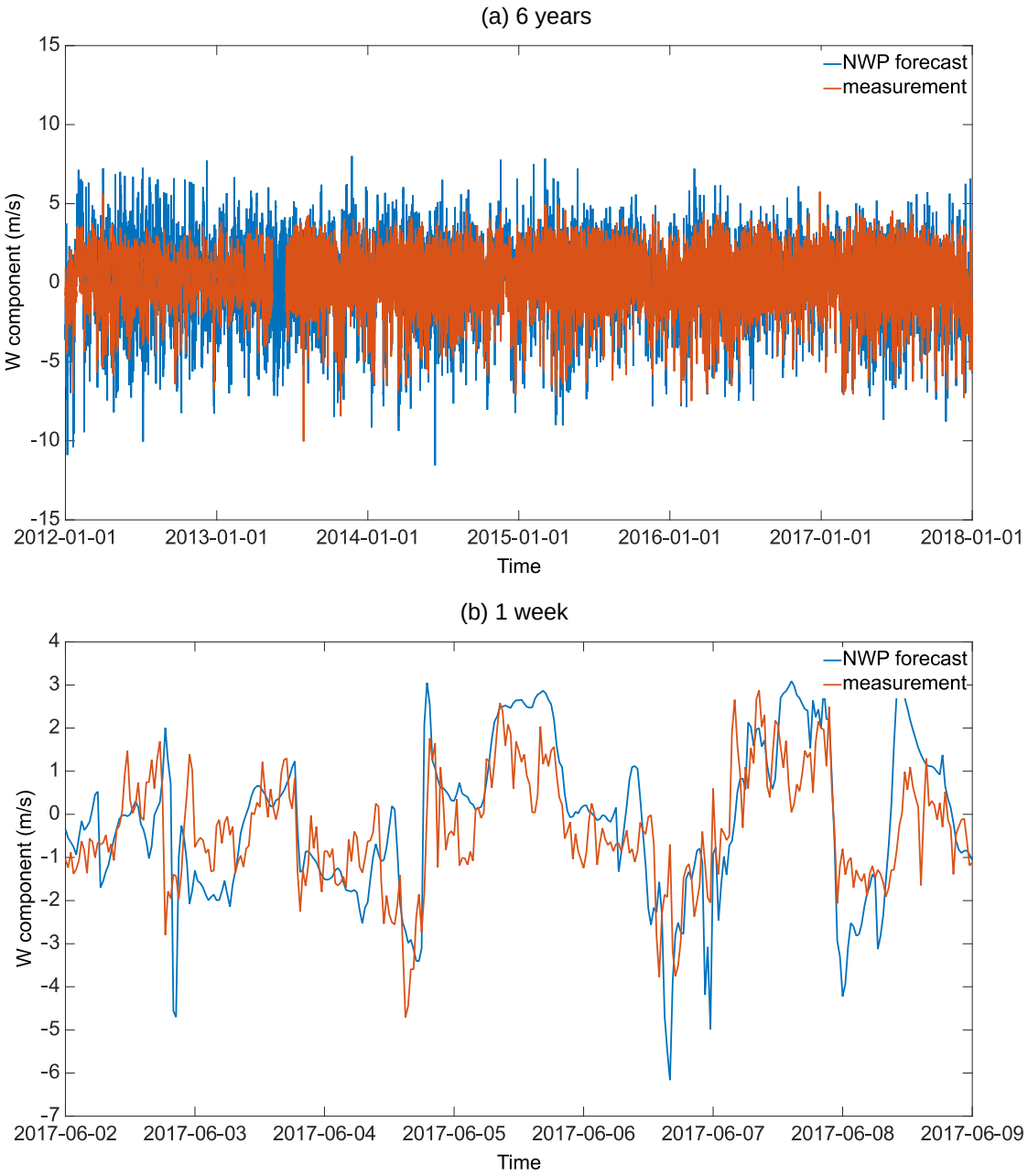

Figure 2. W-E component of the ground level wind. The blue line is NWP forecast, the red one is the measurement at Stolp weather station. Graph (a) is for the whole 6 years of the study, graph (b) is for 1 week of June 2017.

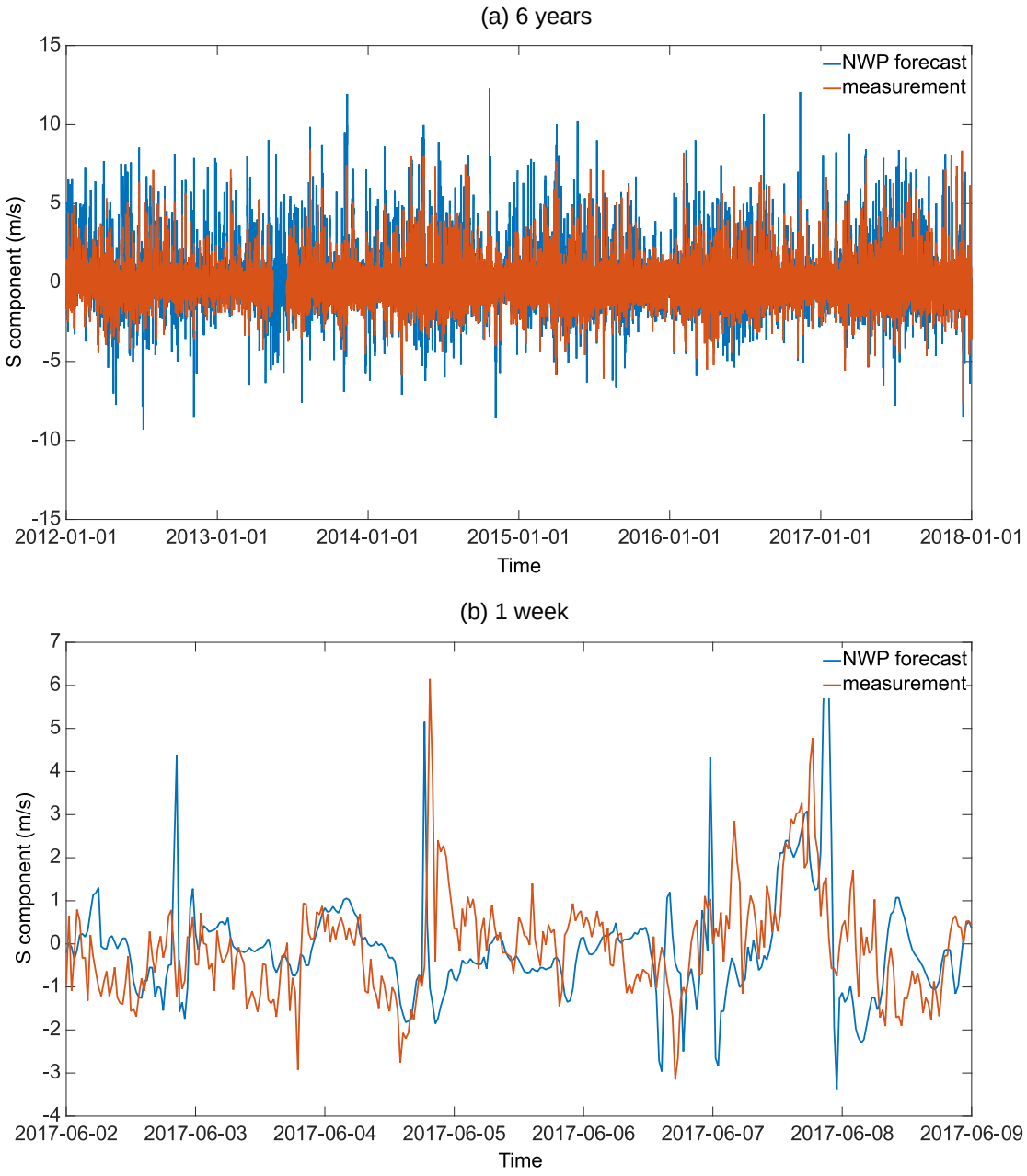

Figure 3. S-N component of the ground level wind. The blue line is NWP forecast, the red one is the measurement at Stolp weather station. Graph (a) is for the whole 6 years of the study, graph (b) is for 1 week of June 2017.

One reason is the curse of dimensionality (Sun et al., 2014). Each regressor adds a dimension to the model space and higher dimensionality leads to data sparsity (Jing et al., 2007). Additional regressors, in particular if they are not relevant (Cichosz, 2015), increase the necessary amount of training data, which is important because the training data set is finite. Having a big ratio of the number of training samples to the number of model parameters prevents overfitting (Theodoridis and Koutroumbas, 2006). Having too many regressors thus leads to worse generalization because regressors introduce parameters.

The data is noisy. More regressors introduce more noise that also adversely affects the algorithm's performance (Elhamifar and Vidal, 2013). A separate issue is computational complexity. More dimensions require more computation both for learning and for predicting.

For these reasons, it is preferable to use few regressors and not to use any unnecessary ones. At the same time, the regressors used should contain enough information for the wind model to make good predictions.

Several methods of regressor selections that can be used to achieve these goals are available. To select the regressors, we use linear in parameters (LIP) method as implemented in Pro-

Opter IVS (Gradišar et al., 2015) to rank the regressors. The regressors are ranked from best to worst by the amount of the useful information they carry, so it is reasonable to use a chosen number of the best ranking ones when building a model, which is what we do.

2.4. Training Data Reduction

In principle, a bigger amount of training data results in a better statistical model. However, excessive amount of training data causes ill-conditioning, unreasonable computer memory requirement, or too long computation times. The available quantity of training data is excessive for the GP models we use, so we reduce it to a reasonably sized sample.

One straightforward way of selecting the training data points to use is decimation. Selecting every s-th sample from the training data multiplies the training data set size with a factor of $1/s$. Another method we try is a smart selection method based on Euclidean distance between training points. Every data point is treated as a vector with (normalized) regressor and measured output values as its coordinates. For every point in the initial training data, the distance to its nearest neighbour is computed. The points whose distances to their nearest neighbours are large are kept in the output training set, the ones with neighbours nearby are rejected. The procedure is done iteratively: in each step, 95% of the points are kept and 5% are discarded, and the obtained output training set is reduced again until the desired number of training points is reached (Perne et al., 2019).

2.5. Gaussian Process Model

GP is a stochastic process $f(\mathbf{z})$ where any finite set of function values $\{f(\mathbf{z}_1), f(\mathbf{z}_2), \ldots, f(\mathbf{z}_M)\}$ is jointly normally distributed. The set is described by its mean vector $\boldsymbol{m}$ and covariance matrix Σ as (Rasmussen and Williams, 2006; Kocijan, 2016):

$$p(f(\mathbf{z}_1), \ldots, f(\mathbf{z}_M) \,|\, \mathbf{z}_1, \ldots, \mathbf{z}_M) = \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma}) \tag{1}$$

Symbol $p(f(\mathbf{z}_1), \ldots, f(\mathbf{z}_M) \,|\, \mathbf{z}_1, \ldots, \mathbf{z}_M)$ stands for the joint probability density function of $f(\mathbf{z}_1), \ldots, f(\mathbf{z}_M)$ for a given choice of $\mathbf{z}_1, \ldots, \mathbf{z}_M$. A GP can be constructed by setting the elements m_i of the mean vector $\boldsymbol{m}$ to the values $m(\mathbf{z}_i)$ of a mean function $m(\mathbf{z})$, and the matrix elements Σ_{ij} to the values $k(\mathbf{z}_i, \mathbf{z}_j)$ of a kernel function or covariance function $k(\mathbf{z}, \mathbf{z}')$. Any function that generates a positive, semi-definite covariance matrix $\boldsymbol{\Sigma}$ can be used as covariance function (Kocijan, 2016).

When modelling a dynamic system, one is interested in a relation between the input $\mathbf{z}$ and the output y of the form $y(\mathbf{z}) = g(\mathbf{z}) + v$, where $g(\mathbf{z})$ is an underlying function and v is the measurement and model error. We choose to model the error as white noise $v \sim \mathcal{N}(0, \sigma_{\mathbf{z}}^2)$.

The goal of GP modelling is constructing a GP that predicts the probability density $p(y(\mathbf{z}))$. A prior guess of the mean function is made, often $m(\mathbf{z}_i) \equiv 0$, and a covariance function that reflects our knowledge of the dynamic system is chosen. The training data is then used to obtain posterior distribution $p(y(\mathbf{z}))$. The training data $\mathcal{D}$ is structured as $\mathcal{D} = \{\mathbf{Z}, \boldsymbol{y}\}$, where the regression matrix $\mathbf{Z}$ contains the regression vectors, $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N]$, and the training outputs y_i from the training vector $\boldsymbol{y} = [y_1, \ldots, y_N]^{\mathrm{T}}$ are taken to be noisy realizations of the GP, $f(\mathbf{z}_i) = y(\mathbf{z}_i) + v_i$, $p(v_1, \ldots, v_N) = \mathcal{N}(0, \boldsymbol{\Sigma}_v)$. The noise covariance matrix $\boldsymbol{\Sigma}_v$ is diagonal since the noise is uncorrelated. The "full" covariance matrix is defined as $\boldsymbol{K} = \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_v$. The predictive distribution at a test point $\mathbf{z}^*$ is (Rasmussen and Williams, 2006; Kocijan, 2016):

$$p(f(\mathbf{z}^*) \,|\, \mathcal{D}, \mathbf{z}^*) = \mathcal{N}(\mu(\mathbf{z}^*), \sigma^2(\mathbf{z}^*)) \tag{2}$$

where

$$\mu(\mathbf{z}^*) = \boldsymbol{k}^{\mathrm{T}} \boldsymbol{K}^{-1} \boldsymbol{y} \tag{3}$$

$$\sigma^2(\mathbf{z}^*) = \kappa(\mathbf{z}^*) - \boldsymbol{k}^{\mathrm{T}} \boldsymbol{K}^{-1} \boldsymbol{k} \tag{4}$$

The $\boldsymbol{k}$ and scalar κ are defined using covariance function to equal $\boldsymbol{k} = k(\mathbf{z}^*, \mathbf{Z})^{\mathrm{T}}$ and $\kappa = k(\mathbf{z}^*, \mathbf{z}^*)$. Taking noise into account, the expression for the probability density of the measured output y^* at the test point is $p(y^* | \mathcal{D}, \mathbf{z}^*) = \mathcal{N}(\mu(\mathbf{z}^*), \sigma^2(\mathbf{z}^*) + \sigma_{\mathbf{z}^*}^2)$.

The covariance function typically has some free parameters called hyperparameters Θ, a good choice of which is typically not known in advance, so they are chosen with optimization based on the training data. The values of hyperparameters that maximize the likelihood $p(\Theta | \mathbf{Z}, \boldsymbol{y})$ and are thus the most probable given measurement data are chosen. Assuming a prior distribution of hyperparameters that is uniform – every value of each hyperparameter is equally likely – it is (Rasmussen and Williams, 2006):

$$p(\Theta \,|\, \mathbf{Z}, \boldsymbol{y}) \propto p(\boldsymbol{y} \,|\, \mathbf{Z}, \Theta) \tag{5}$$

The right-hand-side distribution is multivariate normal, and taking its logarithm, it follows:

$$\log p(\boldsymbol{y} \,|\, \mathbf{Z}, \Theta) = -\frac{N}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{K}| - \frac{1}{2} \boldsymbol{y}^{\mathrm{T}} \boldsymbol{K}^{-1} \boldsymbol{y} \tag{6}$$

This expression is used as the objective function for the optimisation to determine Θ.

2.5.1. Model Choices and Assumptions

We assume $\sigma_{\mathbf{z}}$ is constant, independent of $\mathbf{z}$, $\sigma_{\mathbf{z}} = \sigma_n$. It follows that $\boldsymbol{\Sigma}_v = \sigma_n^2 \boldsymbol{I}$. We choose squared exponential covariance function, expressed in component notation as:

$$k(\mathbf{z}, \mathbf{z}') = \sigma_f^2 \exp\left[-\frac{1}{2} \sum_{d=1}^{D} \frac{(z_d - z_d')^2}{l_d^2}\right] \tag{7}$$

There are $D + 1$ hyperparameters that have to be chosen, consisting of σ_f and the l_d's for the values of d from 1 to D. A choice of σ_n has to be made as well.

We follow the procedure of Stepančič and Kocijan (2017) and substitute the parameter σ_n with a new parameter λ, defined as $\lambda = \sigma_n^2 / \sigma_f^2$. It enables us to express the optimum value of σ_f

analytically as:

$$\sigma_f = \left(\frac{N}{\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{\Sigma} + \lambda \boldsymbol{I})^{-1}\boldsymbol{y}}\right)^{\frac{1}{2}} \quad (8)$$

We determine the values of the other hyperparameters and λ through optimization with the criterion (6) using conjugate gradient method (Rasmussen and Williams, 2006; Rasmussen and Nickisch, 2010).

2.6. GP-NARX

NARX stands for nonlinear autoregressive exogenous model. Its structure is shown in Figure 4 and it can be described with a nonlinear stochastic recurrent equation of the form (Nelles, 2001):

$$y(t) = f(y(t - 1\Delta t), \ldots, y(t - n\Delta t), \boldsymbol{u}(t - 1\Delta t), \ldots, \boldsymbol{u}(t - m\Delta t)) + v \quad (9)$$

The system inputs are assembled into the vector $\boldsymbol{u}$ and the model output signal is labelled as y. The parameter t represents the time for which the prediction is made. The model output distribution $p(y(t))$ depends on the previous distributions of the modelled variable $p(y(t - i\Delta t))$ and on the other regressors $\boldsymbol{u}$. The noise v represents measurement and process noise. In the case of GP-NARX, the nonlinear function f is a GP and the variance of the GP is useful in determining the variance of the distribution of model output.

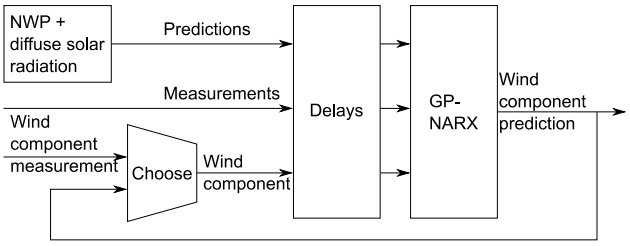

Figure 4. Hybrid wind component prediction model with GP-NARX. The inputs of the GP model are both the values of its output variable in the previous time steps and the values of other external variables. In simulation, predicted values of the output variable are used as model inputs as soon as they are available, while at the beginning of the simulation, measured values are used. Some inputs of the GP-NARX are obtained as outputs of the NWP model and the ANN model for diffuse solar radiation, making the whole a hybrid model.

As a result of nonlinearity, the output distribution $p(y(t))$ is in general non-Gaussian and is hard to describe analytically even if v is Gaussian. It is thus not practical to perform simulations and multi-step predictions (see 2.8) analytically. Instead, Monte Carlo method is used. Many deterministic realizations of simulation are performed, where the values of v and f at every t are determined by random sampling from their Gaussian distributions – distribution for f is Gaussian when the inputs are deterministic, which they are in a realization. Expected value and variance of the modelled variable as a function of time are then estimated from the sample of realizations.

2.7. GP-NFIR

NFIR stands for nonlinear finite impulse response. The structure of a NFIR model is shown in Figure 5 and it can be described with a nonlinear stochastic recurrent equation of the form (Nelles, 2001):

$$y(t) = f(\boldsymbol{u}(t - 1\Delta t), \ldots, \boldsymbol{u}(t - m\Delta t)) + v \quad (10)$$

We see that GP-NFIR is a special case of GP-NARX where output values do not enter the model as regressors. As a result, simulating the model is more straightforward and computationally significantly less demanding. However, the necessary number of regressors is significantly larger than in GP-NARX, which may cause problems with training (Nelles, 2001). With GP-NFIR, there is no distinction between prediction and simulation (see 2.8) as the variable predicted by the model is not used in model inputs.

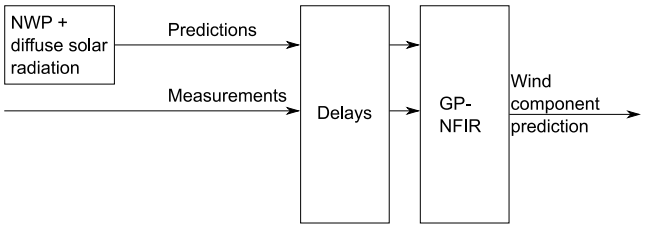

Figure 5. In contrast with GP-NARX, GP-NFIR model does not use the output signal in any of the regressors.

2.8. Prediction Versus Simulation

Statistical GP models compute the predicted value distribution of the output variable based on the provided input values. In the case of GP-NARX, certain input values are values of the output variable. There are thus multiple choices for the input value, either measurement or a model output can be used. If one always uses the necessary measured value when calculating the output for a certain time, we speak of prediction in the modelling sense.

Another approach is multi-step prediction – starting with all measurements when calculating the model output for time t and using no measurements of the output variable from time t on, only the model predictions, when calculating the output at time $t + 1\Delta t$, $t + 2\Delta t$, …, $t + n\Delta t$. The positive integer n is the number of prediction steps. Multi-step prediction can then be restarted at $t + 1\Delta t$ with all measurements and used to obtain the predicted value at $t + (n + 1)\Delta t$., again with n steps.

The third possibility is simulation – starting with all the measurements when calculating the output for time t and then continuing the same way as in multi-step prediction without further measurements of the output variable all the way to the end of the time period of interest.

2.9. Testing

We use the data from the years 2012 ~ 2017, consisting of 105217 data points in total. Points from 94945 to 95616 and from 103729 to 104400, corresponding to the first 14 days of June 2017 and to the first 14 days of December 2017, are used for testing the models, while the rest is available for training. Test data is separate from the training data so as to verify the generalization ability of the model. The reason for using such a small share of data in testing is that certain simulations require substantial computation time proportional with the number of testing data points and that many models have to be tested when developing the modelling methods. For example, GP-NARX simulations with smartly selected training data points for both wind components for two 14-day periods take 7618 seconds on Intel® Core™ i7-3770K CPU @ 3.50GHz × 8 with 16 GB RAM, Ubuntu 20.04.2 LTS, and MATLAB R2015a. The required computation time does not, however, hinder the intended model use. Simulating the GP-NARX models for 1 day ahead only takes a few minutes. It should also be noted that one does not have to wait for an accident to happen before running the models. The models can be run on a regular schedule so that the predictions are already available in the case of an accident.

To test the results, we use normalised root-mean-square error (NRMSE), Pearson's correlation coefficient (PCC), and mean standardised log loss (MSLL) values when applicable as figures of merit. NRMSE is defined as:

$$\mathrm{NRMSE} = 1 - \frac{\|\boldsymbol{y} - \boldsymbol{m}\|}{\|\boldsymbol{y} - E(\boldsymbol{y})\|'} \tag{11}$$

where $\boldsymbol{y}$ is the vector of measured values, $E(\boldsymbol{y})$ is the mean of the measured value, and $\boldsymbol{\mu}$ is the vector of predicted values. NRMSE varies between negative infinity and 1, where 1 corresponds to perfect fit and 0 is the value achieved if the prediction is the mean of the measured value. A convenience of NRMSE is that no assumption that the modelled quantity is a scalar is necessary, it can be calculated for vector quantities. We use it to evaluate predictions of horizontal wind as a 2D vector.

The definition of PCC is:

$$\mathrm{PCC} = \frac{(\boldsymbol{y} - E(\boldsymbol{y}) \times (\boldsymbol{\mu} - E(\boldsymbol{\mu}))}{\sqrt{\|\boldsymbol{y} - E(\boldsymbol{y})\|^2 \|\boldsymbol{\mu} - E(\boldsymbol{\mu})\|^2}}, \tag{12}$$

where $E(\boldsymbol{\mu})$ is the mean predicted value. The expression $\|\boldsymbol{y}-E(\boldsymbol{y})\|^2$ is variance.

MSLL is defined as (Rasmussen and Williams, 2006):

$$\mathrm{MSLL} = \frac{1}{2N}\sum_{i=1}^{N}\left[\log(\sigma_i^2) - \log(\sigma_y^2) + \frac{(E(\hat{y}_i) - y_i)^2}{\sigma_i^2} - \frac{1}{2N}\sum_{i=1}^{N}\frac{(y_i - E(\boldsymbol{y}))^2}{\sigma_y^2}\right] \tag{13}$$

where y_i is the measured value, σ_y^2 is the variance of the measured value, $E(\hat{y}_i)$ is the mean prediction, and σ_i^2 is the predictive variance. The summation includes all the test samples and the index *i* corresponds to the sample. MSLL takes into account the predictive variance and we use it to evaluate the models for wind components that predict a distribution. Lower MSLL value corresponds to a better model, the values are typically negative.

3. Results

Several models are trained, tested on the same data, and compared in the figures of merit. The main distinction between different models is in the regressors they require. The main distinction between different GP-NARX results is in how the regressor values are obtained. When all the necessary local meteorological variables are measured, we speak of prediction. When the predicted value is used in regressors in further steps, this is multi-step prediction if the model is restarted at every time step and run for several time steps, or simulation if the model is started once and run for the whole test period. When measured values that are assumed to not be available are substituted with equivalent NWP signals, we name the case NWP-substituted model.

3.1. NWP

We are interested in the benefits of various different statistical parts of the hybrid model. We thus compare it to the model without a statistical part, that is, to the pure NWP model. The NRMSE value for the 2D wind in the 4 weeks of testing as predicted by WRF NWP is -0.082, PCC for W-E component is 0.577, and PCC for S-N component is 0.313. MSLL values are not available because variance is not predicted.

Table 3. Figures of Merit for Simulation Results from Sections 3.1 to 3.4

Section	Model	MSLL		NRMSE	PCC	
		W-E	S-N		W-E	S-N
3.1	NWP	--	--	-0.082	0.577	0.313
	Hybrid models					
3.2	NWP-based GP-NFIR	-0.258	-0.044	0.164	0.660	0.341
3.3	GP-NFIR 15 decimate	-0.734	-0.509	0.405	0.877	0.672
3.3	GP-NFIR 30 decimate	-0.785	-0.508	0.445	0.877	0.770
3.3	GP-NFIR 30 smart	-0.664	-0.308	0.445	0.882	0.776
3.4	GP-NARX decimate	-0.466	-0.335	0.323	0.806	0.648
3.4	GP-NARX smart	-0.562	-0.151	0.357	0.834	0.691

Note: The simulations are performed for 14 days both in June and in December 2017. The "Section" column refers to the section of the article describing the model for easier navigation.

3.2. NWP-Based GP-NFIR

The next model tested is one that does not include any measurements. It is a GP-NFIR with 44 regressors. 40 are derived

from NWP and ANN signals by delaying them from $t - 2\Delta t$ to $t + 2\Delta t$ for each one of 8 signals, where t is defined to be the time for which the output is being calculated. The remaining 4 are the derived signals (sines and cosines of scaled time with periods of 1 day and 1 year) without delay. The training data is decimated for a factor 19 so 2570 training points remain to avoid too big matrices in the learning phase. Prediction for test data is then used, NRMSE, PCC, and MSLL values are obtained: NRMSE is 0.164, PCC for W-E component is 0.660, PCC for S-N component is 0.341. MSLL for W-E component is –0.258, and MSLL for S-N component is -0.044. Judging by NRMSE and PCC, the model is an important improvement over pure NWP results.

Table 4. The Regressors for Calculating W-E Component of the Wind Using GP-NFIR Model as Selected with ProOpter IVS LIP Method

Source	Variable	Delay
Cerklje Airport	W-E wind	1
Krško	W-E wind	1
Brežice	W-E wind	1
NWP + ANN	W-E wind	-2
Time	Cosine 1 day	0
Krško	S-N wind	1
SODAR	S-N wind	1
Stolp	S-N wind	2
Cerklje Airport	W-E wind	2
SODAR	W-E wind	2
Krško	Relative humidity	1
Brežice	S-N wind	3
SODAR	S-N wind	2
Stolp	Global solar radiation	1
NWP+ANN	Global solar radiation	2
SODAR	S-N wind	1
SODAR	S-N wind	4
SODAR	Vertical wind	5
Krško	W-E wind	2
SODAR	W-E wind	1
Cerklje	Air temperature	4
Cerklje Airport	Air pressure	5
NWP+ANN	W-E wind	1
Cerklje Airport	S-N wind	2
SODAR	W-E wind	3
Stolp	Air temperature	1
Stolp	global solar radiation	5
Brežice	S-N wind	1
NWP + ANN	W-E wind	0
Krško	Air temperature	1

Note: Delay is measured in time steps from the time to which the model output corresponds. A positive delay means that the signal value corresponds to a time before the time to which the prediction corresponds, and vice versa. Delay 0 corresponds to the time for which the prediction is made.

3.3. Hybrid Models with GP-NFIR

In the GP-NFIR model, we allow measured regressors as well. We start from 274 regressor candidates: from each one of 46 local meteorological variable signals, 5 delayed signals are used, from $t - 1\Delta t$ to $t - 5\Delta t$. For each one of the 8 NWP and ANN signals, 5 delays are used, $t - 2\Delta t$ to $t + 2\Delta t$, and a single delay for each mathematical signal, $t - 0\Delta t$. We use LIP to rank the regressors. As the total number of data points with all the regressors available is too big for the method, decimation is used and every 10th is kept, resulting in 1673 points. The regressors are listed in Tables 4 and 5 from the best on, the models use either the first 15 or all 30 of them. Decimation with the quotient $s = 19$ results in an acceptable number of training points: for 15 regressors, 2619 points for W-E component and 2596 for S-N component, for 30 regressors, 2543 points for W-E component and 1550 for S-N component. With 30 regressors, we also use the smart selection method to select the same number of training data points as when we use decimation. The results are summarised in Table 3 together with the other simulation results as "GP-NFIR 15 decimate", "GP-NFIR 30 decimate", and "GP-NFIR 30 smart". We see that the results are better than with the model that does not take measurements into account and that the increase of regressor number from 15 to 30 and smart training data selection do not bring much improvement.

3.4. Hybrid Models with GP-NARX

We allow the use of delayed output signal in regressors to obtain GP-NARX models. In theory, GP-NARX models require fewer delays and regressors than GP-NFIR models, so we begin with 122 regressor candidates. From each one of 47 local meteorological variable signals, 2 delayed signals are used, $t - 1\Delta t$ and $t - 2\Delta t$. For each one of the 8 NWP and ANN signals, 3 delays are used, $t - 1\Delta t$ to $t + 1\Delta t$, and a single delay for each mathematical signal, $t - 0\Delta t$. 15 regressors are selected using LIP. As the total number of data points with all the regressors available is too big for the LIP method, decimation is used and every 10th is kept, resulting in 4422 points. The best 15 regressors for each wind component are listed in Table 6 from the most to the least important.

With GP-NARX models, there is a difference between prediction and simulation (see 1.1). In the initial step, all the regressor values are read from the database of the available signal values. The model output is calculated, the predictive distribution for the wind component at time t is obtained. Now two conflicting pieces of information on the wind component at time t are available: the measurement and the model prediction. If the model prediction is used as the regressor value when calculating the component at $t + 1\Delta t$ with the GP-NARX model, and the new prediction is used when calculating the value at the next time step and so on to the end of the data set, we talk of simulation. If only measured values and no model outputs are used, it is called (one step) prediction. We can also do n-step prediction, where in the process of calculating the wind component value at $t + n\Delta t$, the model is initialized with measured values of the component up to time t and predicted values for later times are used.

Two different GP-NARX models are constructed for each wind component. In the first example, the training data is decimated with $s = 19$ to decrease the number of training points to

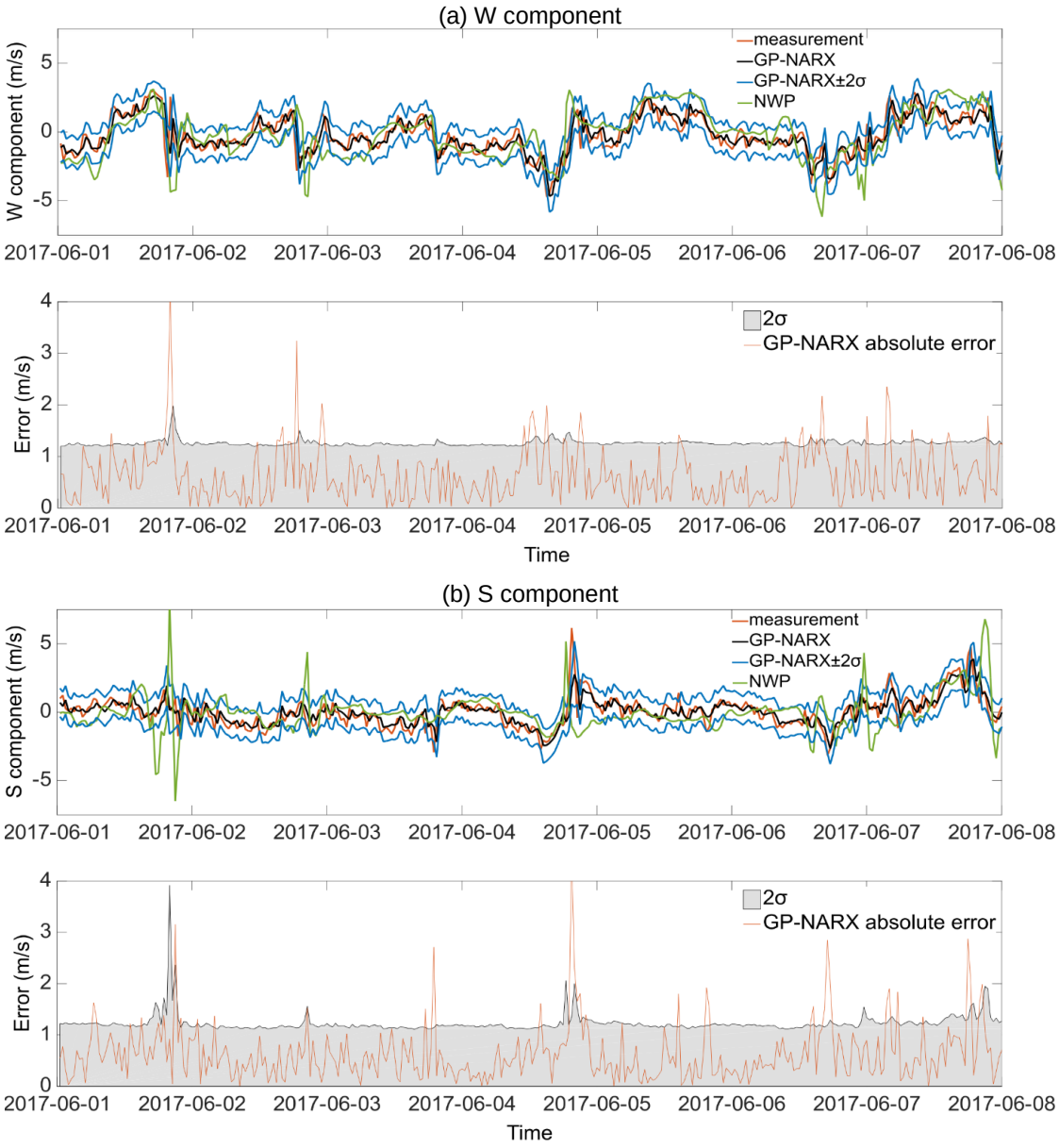

Figure 6. NWP-substituted decimated GP-NARX prediction results 2 steps ahead as an illustration. NWP forecasts are also shown for comparison.

a reasonable level, 2545 for W-E component and 2104 for S-N component. Simulations result in NRMSE of 0.323, MSLL of –0.466 in W-E direction and MSLL of –0.335 in S-N direction. The second example is constructed with the same number of training points for each component, but smart training point selection is used. Simulations of these models result in NRMSE of 0.357, MSLL of –0.562 in W-E direction and MSLL of –0.151 in S-N direction. The results are a huge improvement on NWP results and much better than the NWP-based GP-NFIR results obtained without taking measurements into account. The decrease of performance compared to GP-NFIR arises from the use of the inaccurate forecasted signal in regressors in simulation: in single-step prediction, where there is no such inaccuracy, GP-NARX model outperforms GP-NFIR.

The performance of the GP-NARX models with smart training point selection is also evaluated on a part of the training data set, in particular on the data from the last 14 days of May and the last 14 days of November 2017. One would expect the performance on this data to be slightly better because a small part of the data points is also used in training. The achieved NRMSE is 0.412, MSLL is –0.609 in W-E direction and –0.003 in S-N direction, PCC is 0.882 in W-E direction and 0.626 in S-N direction. Three of the figures of merit are slightly better than for the test data set, but the two specific to the S-N direction are worse. The explanation is that the weather in these periods is different and the S-N component is more challenging to

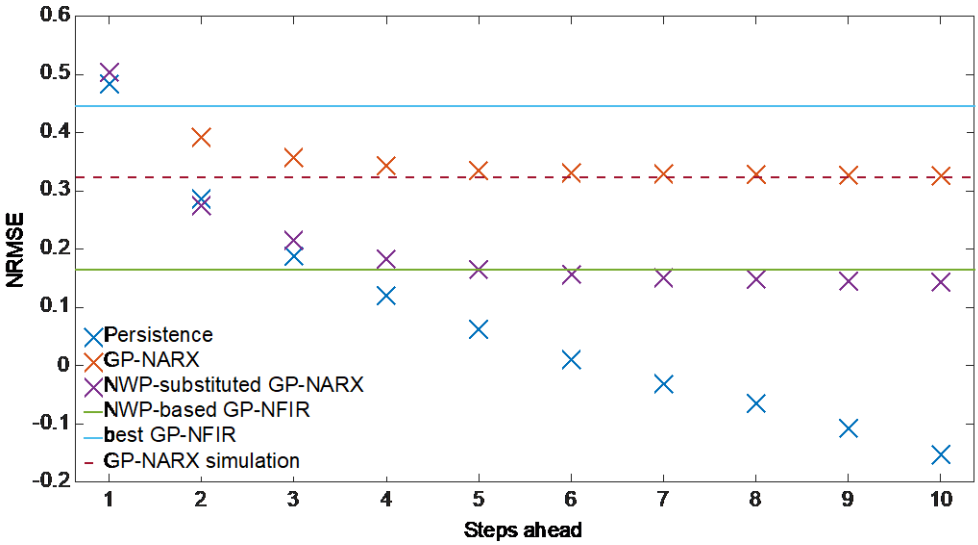

Figure 7. NRMSE decrease along prediction, shown graphically. "Persistence" means using measured wind velocity in step 0 as prediction for steps 1 to 10. "GP-NARX" is the classical decimated GP-NARX model, regressor signals are measured (except the output signal, where the predicted value is used from step 1 on). "NWP-substituted GP-NARX" is the classical decimated GP-NARX model, regressor signals are approximated with NWP forecasts from step 1 on (except as the output signal from step 1 on the predicted value is used). Horizontal lines represent simulation values. NWP-based GP-NFIR model requires no online measurements. Persistence and NWP-substituted GP-NARX only require measurements up to time 0. Best GP-NFIR and GP-NARX rely on measurements taken before the time for which the prediction is made. The NWP-substituted GP-NARX model NRMSE decays below the NWP-based GP-NFIR model one because the latter one uses the NWP information appropriately while the former one treats the NWP information as if it were measurements which is suboptimal.

predict than in the original test period. It is confirmed by the results of the one-step ahead persistence prediction, which achieves NRMSE of 0.543, PCC of 0.932 in W-E direction and 0.795 in S-N direction – the PCC for S-N direction is much worse than for the original test period. This observation suggests that the numerical values of the figures of merit are strongly dependent on the weather in the chosen test period, but the same kind of weather is beneficial for the performance of at least GP-NARX and persistence models.

Multi-step prediction results are evaluated in Table 7. It can be seen that smart training point selection results in better NRMSE and PCC values than decimation. In MSLL, the roles are mostly reversed, decimation mostly results in better values than smart selection. It follows that the model obtained by decimation gives better information on prediction uncertainty.

Figure 6 presents the results of NWP-substituted GP-NARX model prediction. In these calculations, measured local meteorological variables are used as regressors in decimated GP-NARX model up to $t = 0$ and replaced with equivalent NWP signals further on. The exception is the output variable, where the model output is used. Table 9 lists figures of merit for this model. The prediction figures of merit are also shown in Figure 7.

The prediction variance, shown in grey in Figure 6, illustrates the model's confidence. The variance depends on the output noise and the density of the training data points in the part of the input space for which the prediction is made (Kocijan, 2016, Ch. 2.8). We do not explore the effect of this variance on the atmospheric dispersion model predictions in the scope of this paper but it is a possible topic of further research.

3.5. Persistence Model

A trivial way of predicting the wind component in the next time step is predicting that the value will equal the current value. Nevertheless, this kind of model is useful only for one-step prediction, i.e. half an hour in advance. The NRMSE value achieved using this so-called persistence method is 0.484, and PCC is 0.895 in W-E direction and 0.818 in S-N direction. Of our more advanced predictions, only GP-NARX predictions take the wind component measurements into account and they achieve higher NRMSE value than persistence predictions, fulfilling the basic criterion for a reasonable prediction model.

3.6. Influence of Improved Wind Prediction on Dispersion Modelling Results

Lagrangian particle atmospheric dispersion model (Tinarelli et al., 2000) simulations are run for pollutant emitted at ground level and at the temperature of the environment at the location of Krško NPP. The models' resolution is 50 m, domain size 5 km × 5 km centred on the pollutant source, while the other parameters are as described in Mlakar et al. (2019). The weather signals entering the model are temperatures at 2, 10, 40, and 70 m levels, global solar radiation at 2 m, and wind speed and direction at 10 m level at Stolp weather station. The temperatures and global solar radiation are obtained from measurements, while the three simulations differ in the source of the wind signal.

Table 5. The Regressors for Calculating S-N Component of the Wind Using GP-NFIR Model as Selected with ProOpter IVS LIP Method

Source	Variable	Delay
SODAR	S-N wind	1
Cerklje Airport	S-N wind	1
Krško	S-N wind	1
Brežice	W-E wind	1
Brežice	S-N wind	1
Stolp	Global solar radiation	1
SODAR	W-E wind	2
SODAR	S-N wind	2
Stolp	Air temperature	4
Cerklje Airport	S-N wind	2
NWP + ANN	S-N wind	0
Krško	W-E wind	1
SODAR	W-E wind	4
Cerklje Airport	Air pressure	3
SODAR	W-E wind	5
Krško	Relative humidity	1
Cerklje Airport	Relative humidity	3
NWP + ANN	Global solar radiation	1
SODAR	S-N wind	3
SODAR	S-N wind	2
SODAR	W-E wind	1
Stolp	Relative humidity	4
Cerklje Airport	Relative humidity	1
Stolp	Air temperature	1
Lisca	Air pressure	2
NWP + ANN	Relative humidity	2
SODAR	S-N wind	1
Cerklje Airport	Relative humidity	2
Lisca	S-N wind	3
Stolp	Relative humidity	2

Note: Delay is measured in time steps from the time to which the model output corresponds. A positive delay means that the signal value corresponds to a time before the time to which the prediction corresponds, and vice versa. Delay 0 corresponds to the time for which the prediction is made.

One simulation utilizes the measured wind speed and direction. It is known from atmospheric dispersion model validation studies (Grašič et al., 2011) that the agreement between atmospheric dispersion model simulations with measured weather signals and measured pollution dispersion tends to be good. We thus use this result as the reference. The second simulation uses decimated hybrid GP-NARX model one-step-ahead prediction as described in Section 3.4. The third simulation uses the available NWP prognosis of WRF-ARW version 3.4.1, which is the signals this study improves on.

It is shown in Table 7 and Section 3.1 that the decimated hybrid GP-NARX model one-step-ahead prediction is better than the NWP prognosis in NRMSE and PCC figures of merit. The question is whether the improvement in the wind prediction translates into an improvement in the pollutant concentration predictions. Judging by Figure 8, the results based on hybrid model predictions (middle column) are closer to the reference (left column) than the results based on NWP predictions (right column). The improved wind prediction thus helps with the atmospheric dispersion modelling.

Table 6. The Regressors for Calculating Each Component of the Wind Using GP-NARX Model as Selected with ProOpter IVS LIP Method

a)	Best regressors for W-E wind		
	Source	Variable	Delay
	Stolp	W-E wind	1
	NWP+ANN	W-E wind	–1
	Cerklje Airport	W-E wind	1
	Krško	W-E wind	1
	Cerklje Airport	W-E wind	2
	Brežice	W-E wind	1
	Krško	Air temperature	1
	NWP+ANN	Global solar radiation	–1
	Cerklje Airport	S-N wind	1
	NWP+ANN	Global solar radiation	1
	SODAR	W-E wind, layer 1	1
	Cerklje Airport	Air temperature	1
	Cerklje	Air pressure	2
	Stolp	Air pressure	1
	Krško	S-N wind	1
b)	Best regressors for S-N wind		
	Source	Variable	Delay
	Stolp	S-N wind	1
	Brežice	S-N wind	1
	Krško	S-N wind	1
	Cerklje Airport	W-E wind	1
	Cerklje Airport	S-N wind	1
	Cerklje Airport	S-N wind	2
	NWP+ANN	S-N wind	–1
	SODAR	S-N wind, layer 3	1
	NWP+ANN	Global solar radiation	–1
	Stolp	Air temperature, 70 m	2
	Lisca	S-N wind	1
	Krško	Air temperature	1
	NWP+ANN	W-E wind	–1
	Cerklje Airport	Air pressure	1
	Cerklje Airport	Air temperature	2

Note: Delay is measured in time steps from the time to which the model output corresponds. A positive delay means that the signal value corresponds to a time before the time to which the prediction corresponds, and vice versa. Delay 0 corresponds to the time for which the prediction is made. a) Selection for W-E component; b) Selection for S-N component.

4. Discussion

We have shown that important improvements in spot wind forecasting compared to NWP are possible with statistical modelling. The improvement is significant if the statistical model is only post-processing the NWP forecasts, and much better if measured local meteorological variables are also included as the statistical model inputs.

Validation of air pollution dispersion models is highly challenging (Mlakar et al., 2015) so it was not feasible to directly

Table 7. Figure of Merit Values for Multi-Step Prediction and for Simulation throughout 672 Time Steps

Steps	MSLL				NRMSE		PCC			
	W-E		S-N				W-E		S-N	
	Decimate	Smart	Decimate	Smart	Decimate	Smart	Decimate	Smart	Decimate	Smart
1	**–0.851**	–0.707	**–0.592**	–0.352	0.502	**0.515**	0.905	**0.906**	0.808	**0.830**
2	**–0.599**	–0.566	**–0.411**	–0.223	0.392	**0.415**	0.846	**0.861**	0.714	**0.743**
3	**–0.537**	–0.529	**–0.379**	–0.190	0.356	**0.383**	0.823	**0.846**	0.682	**0.712**
4	–0.511	**–0.516**	**–0.363**	–0.179	0.342	**0.370**	0.815	**0.840**	0.668	**0.701**
5	–0.489	**–0.510**	**–0.349**	–0.175	0.334	**0.365**	0.810	**0.837**	0.659	**0.696**
Simulation	–0.466	**–0.562**	**–0.335**	–0.151	0.323	**0.357**	0.806	**0.834**	0.648	**0.691**

Note: Columns labelled "decimate" correspond to GP-NARX with decimation of training points, and "smart" refers to the GP-NARX with smart selection of training points. The "decimate" and "smart" models are compared to one another in each figure of merit and the better one is shown in bold. We see that the GP-NARX model outperforms the GP-NFIR model in the first prediction step, see Table 3 for comparison. According to MSLL, the "decimated" GP-NARX model is better than the "smart" one in the first 3 prediction steps (1.5 h) for the W-E direction and in all 5 prediction steps (2.5 h) in the S-N direction. According to NRMSE and PCC, the "smart" model is consistently better than the "decimated" model. The two models not taking measurements into account, NWP and NWP-based GP-NFIR, are worse than these predictions.

Table 8. NRMSE and PCC Values for Multi-step Persistence Prediction

Steps	NRMSE	PCC	
		W-E	S-N
1	0.484	0.895	0.818
2	0.286	0.786	0.673
3	0.188	0.712	0.598
4	0.120	0.647	0.552
5	0.062	0.589	0.509

experimentally demonstrate that the proposed hybrid model results in improved atmospheric modelling predictions. However, it can be presumed that the demonstrated improvement in local weather prediction is beneficial for atmospheric dispersion modelling. An ilustration of the benefit of hybrid modelling is in Figure 8. Compared to NWP, hybrid modelling leads to atmospheric dispersion modelling results that are more similar to the ones based on measured meteorological quantities.

The hybrid model results do not reach NRMSE values of much over 0.5. In principle that means that there is still plenty of room for improvement. However, we tested several more models with different parameters and choices of regressors without a significant change in figures of merit. A likely explanation is that not all of the dynamics that influences the wind velocity can be modelled with the available information. In this case, more measurements would be needed for improving predictions. For example, a denser network of measurement stations would better detect the wind phenomena on smaller scales (Schielicke et al., 2019) and better detect local causes of wind (Makarieva et al., 2013). The model performance is also limited by the amount of training data it can handle and by the uncertainty of the observed data.

Another concern with these models is their applicability. To use the model, the input regressor values have to be available. For the NWP and the NWP-based GP-NFIR models, the NWP signals are available several days in advance; however, we use the latest and best NWP forecasts that become available several hours before the time for which the prediction is made. Longer term predictions are possible but would be based on less accurate NWP forecasts and thus perform worse.

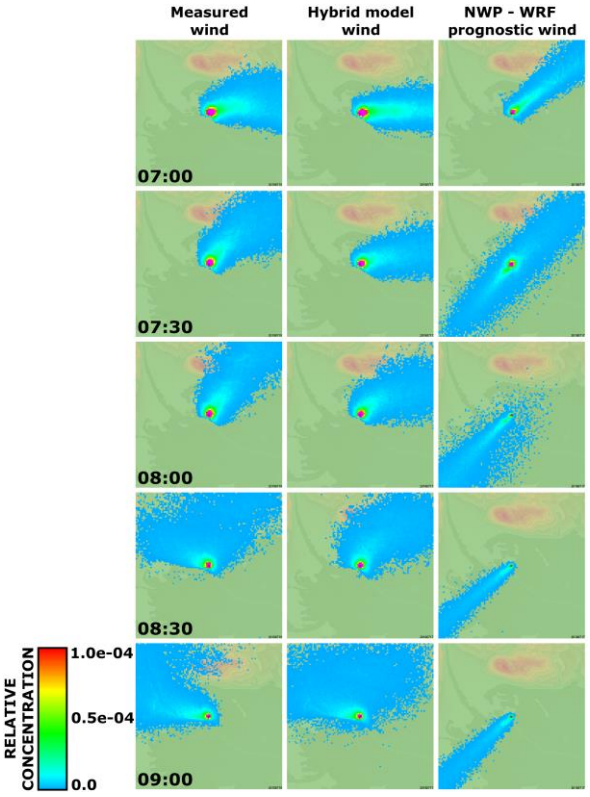

Figure 8. Atmospheric dispersion modelling results given as relative concentrations (Mlakar et al., 2019). Ground level emission at the site of Krško NPP in the center of the field is modelled with measured wind (left column), wind predicted by GP-NARX (middle column), and wind predicted by NWP (right column). We see that the similarity between the reference left column and the GP-NARX middle column is stronger than between the reference and the NWP right column. The results correspond to 2017-06-10.

GP-NFIR and GP-NARX models use measured values of local meteorological variables, some of which are only available one time step (1/2 h) before the time for which the prediction is made. To enable longer term predictions, the input signals

have to be predicted separately. Inaccuracies of those signals impair the model performance. In spite of it, using local meteorological variables as regressors is a promising approach because of the measurements that are available and that improve the prediction. A proof is the NWP-substituted GP-NARX. As an example of replacing measurements with models, NWP forecasts are used in place of local meteorological variables in inputs of decimated GP-NARX model, see Table 9. All 5 presented prediction steps are better than NWP-based GP-NFIR results in NRMSE and 4 or 5 steps in PCC.

Table 9. NWP-Substituted Decimated GP-NARX Multi-Step Prediction

Steps	MSLL		NRMSE	PCC	
	W-E	S-N		W-E	S-N
1	–0.851	–0.592	0.502	0.905	0.808
2	–0.311	–0.005	0.274	0.780	0.504
3	–0.036	0.085	0.214	0.710	0.445
4	0.128	0.146	0.183	0.670	0.412
5	0.235	0.184	0.165	0.646	0.394

Note: Measured meteorological variable signals other than the output signal are replaced with equivalent NWP signal starting with the first step of prediction. One-step prediction is calculated the same way with NWP-substituted decimated GP-NARX and with ordinary decimated GP-NARX so the numbers in the first line of the table are equal to the corresponding numbers in Table 7.

GP-NFIR and GP-NARX models use measured values of local meteorological variables, some of which are only available one time step (1/2 h) before the time for which the prediction is made. To enable longer term predictions, the input signals have to be predicted separately. Inaccuracies of those signals impair the model performance. In spite of it, using local meteorological variables as regressors is a promising approach because of the measurements that are available and that improve the prediction. A proof is the NWP-substituted GP-NARX. As an example of replacing measurements with models, NWP forecasts are used in place of local meteorological variables in inputs of decimated GP-NARX model, see Table 9. All 5 presented prediction steps are better than NWP-based GP-NFIR results in NRMSE and 4 or 5 steps in PCC.

5. Conclusion

We successfully use hybrid modelling for short and medium term local wind forecasting. The novel contribution of the study is that the modelling is done in complex terrain with mostly low wind speeds. Such conditions are of particular interest in atmospheric dispersion modelling.

Post-processing a NWP model forecast with a GP-NFIR model that predicts local wind, forming a NWP-based GP-NFIR model, offers a significant advantage over using crude NWP forecast as local meteorological variable. The method can be applied to the time period for which the NWP signals are available. For shorter term, predictions can be improved by taking available measurements into account. We develop several GP models using both NWP predictions and local meteorological variables as regressors and test them with prediction, multi-step prediction, and simulation. We compare them to one another using different figures of merit. We construct a NWP-substituted GP-NARX model that only depends on information available in real time. It uses local meteorological variables for regressors so it is helped by the available measurements, and NWP forecasts are used in place of measurements for the future. Depending on the selected figure of merit, this model improves on the NWP-based GP-NFIR model for around 5 steps (2.5 h) of prediction.

The hybrid model with NWP-substituted GP-NARX model provides the best short-term local wind forecasts based only on available data. In medium term, the best results are obtained by the NWP-based GP-NFIR model. For atmospheric dispersion modelling, we propose combining the results of the two models, using the hybrid model with NWP-substituted GP-NARX model results in the first several time steps and the NWP-based GP-NFIR model results further on, so that the best available prediction is used in every time step. We also propose recalculating the model whenever new input information is available – measurements for the NWP-substituted GP-NARX model, NWP forecasts for both models.

Modelling of other variables that are used in atmospheric dispersion modelling, such as temperatures at various heights and global solar radiation, is ongoing. Modelling of the local meteorological variables that are used as regressors in the NWP-substituted GP-NARX model is planned. Their predictions will be used to replace the NWP forecasts in the NWP-substituted GP-NARX model to improve the model accuracy through improving the inputs.

Nomenclature

List of symbols

$\mathcal{D}$	Training data
D	Number of regressors, model dimension
E(.)	Mean value, average measurement or prediction
$\boldsymbol{I}$	Identity matrix
MSLL	Mean standardised log loss
$\mathcal{N}$	Gaussian distribution
N	Number of data points
NRMSE	Normalised root-mean-square error
PCC	Pearson's correlation coefficient
$\mathbf{Z}$	Regression matrix, matrix of system input vectors
f(.)	Stochastic process
g(.)	System underlying function
k(.)	Covariance function
$\boldsymbol{k}$	Covariance function of test independent variable and regression matrix
l_d	Scale parameter for regressor d
$\boldsymbol{m}$	Vector of mean values
m(.)	Mean function
m	Number of past inputs
n	Model order, number of past outputs
p(.)	Probability density function
t	Time
$\boldsymbol{u}$	Model, system input

y	Model, system output
$\boldsymbol{y}$	Training vector, vector of system outputs
$\mathbf{z}$	Vector independent variable
$\mathbf{z}^*$	Test independent variable vector
Δt	Time step
Θ	Hyperparameters
$\boldsymbol{\Sigma}$	Covariance matrix
$\boldsymbol{\Sigma}_v$	Noise covariance matrix
κ	Covariance function of test independent variable with itself
λ	Noise variance ratio
$\boldsymbol{\mu}$	Vector of predicted values
μ	Mean of the predictive distribution
v	Noise, error
σ	Standard deviation

Acknowledgments. We are grateful to the Krško NPP for the measurement data from their automatic measuring system. The discussions and technical assistance with data processing by Martin Stepančič are gratefully acknowledged. Funding: This work was supported by the Slovenian Research Agency [grants number: L2-8174, L2-2615, and P2-0001].

References

Barratt, R. (2013). *Atmospheric Dispersion Modelling: An Introduction to Practical Applications*. Taylor & Francis. https://doi.org/10.4324/9781315071527

Beelen, R., Voogt, M., Duyzer, J., Zandveld, P. and Hoek, G. (2010). Comparison of the performances of land use regression modelling and dispersion modelling in estimating small-scale variations in long-term air pollution concentrations in a dutch urban area. *Atmos. Environ.*, 44, 4614-4621. https://doi.org/10.1016/j.atmosenv. 2010.08.005

Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag.

Božnar, M. and Mlakar, P. (1995). Neural networks – a new mathematical tool for air pollution modelling. *WIT Trans. Ecol. Environ.*, 9, 259-266. https://doi.org/10.2495/AIR950301

Božnar, M.Z., Grašič, B., de Oliveira, A.P., Soares, J. and Mlakar, P. (2017). Spatially transferable regional model for half-hourly values of diffuse solar radiation for general sky conditions based on perceptron artificial neural networks. *Renew. Energy*, 103, 794-810. https://doi.org/10.1016/j.renene.2016.11.013

Breznik, B., Božnar, M.Z., Mlakar, P. and Tinarelli, G. (2003). Dose projection using dispersion models. *Int. J. Environ. Pollut.*, 20, 278-285. https://doi.org/10.1504/IJEP.2003.004291

Carvalho, J.C. and de Vilhena, M.T.M.B. (2005). Pollutant dispersion simulation for low wind speed condition by the ILS method. *Atmos. Environ.*, 39(34), 6282-6288. https://doi.org/10.1016/j.atmosenv.2005.07.007

Buttrey, S.E. (2015). Data mining algorithms explained using R. *J. Stat. Softw.*, 66. https://doi.org/10.18637/jss.v066.b02

Deisenroth, M.P., Turner, R.D., Huber, M.F., Hanebeck, U.D. and Rasmussen, C.E. (2012). Robust filtering and smoothing with Gaussian processes. *IEEE Trans. Autom. Control*, 57(7), 1865-1871. https://doi.org/10.1109/TAC.2011.2179426

Elhamifar, E. and Vidal, R. (2013). Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(11), 2765-2781. https://doi.org/10.1109/TPAMI.201 3.57

European Union (2008). Directive 2008/1/EC of the European Parliament and of the Council of 15 January 2008 concerning integrated pollution prevention and control. *Off. J. Eur. Union*, 334, 17-119.

European Union (2019). Copernicus land monitoring service 2018. https://land.copernicus.eu/pan-european/corine-land-cover/clc2018/ (accessed June 20, 2022).

Gradišar, D., Glavan, M., Strmčnik, S. and Mušič, G. (2015). ProOpter: An advanced platform for production analysis and optimization. *Comput. Ind.*, 70, 102-115. https://doi.org/10.1016/j.compind.2015.02.010

Grašič, B., Mlakar, P. and Božnar, M.Z. (2011). Method for validation of lagrangian particle air pollution dispersion model based on experimental field data set from complex terrain. *Advanced Air Pollution*. IntechOpen. https://doi.org/10.5772/17286

Grašič, B., Patryl, L., Mlakar, P., Božnar, M.Z., Kocijan, J. (2019). Quality of weather forecast for modelling air pollution dispersion for nuclear emergency. *19th International Conference on "Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes"*, Bruges, Belgium.

Hoolohan, V., Tomlin, A.S., Cockerill, T. (2018). Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data. *Renew. Energy*, 126, 1043-1054. https://doi.org/10.1016/j.renene.2018.04.019

Jing, L., Ng, M.K. and Huang, J.Z. (2007). An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data. *IEEE Trans. Knowl. Data Eng.*, 19(8), 1026-1041. https://doi.org/10.1109/TKDE.2007.1048

Kocijan, J. (2016). *Modelling and Control of Dynamic Systems Using Gaussian Process Models*. Springer, Cham. https://doi.org/10.1007/978-3-319-21021-6

Ma, L., Luan, S., Jiang, C., Liu, H. and Zhang, Y. (2009). A review on the forecasting of wind speed and generated power. *Renew. Sust. Energ. Rev.*, 13(4), 915-920. https://doi.org/10.1016/j.rser.2008.02.002

Makarieva, A.M., Gorshkov, V.G., Sheil, D., Nobre, A.D. and Li, B.L. (2013). Where do winds come from? A new theory on how water vapor condensation influences atmospheric pressure and dynamics. *Atmos. Chem. Phys.*, 13(2), 1039-1056. https://doi.org/10.5194/acp-13-1039-2013

Mlakar, P., Božnar, M.Z. and Grašič, B. (2019). Relative doses instead of relative concentrations for the determination of the consequences of the radiological atmospheric releases. *J. Environ. Radioact.*, 196, 1-8. https://doi.org/10.1016/j.jenvrad.201 8.10.005

Mlakar, P., Božnar, M.Z., Grašič, B. and Breznik, B. (2019). Integrated system for population dose calculation and decision making on protection measures in case of an accident with air emissions in a nuclear power plant. *Sci. Total Environ.*, 666, 786-800. https://doi.org/10.1016/j.scitotenv.2019.02.309

Mlakar, P., Božnar, M.Z., Grašič, B., Brusasca, G., Tinarelli, G., Morselli, M.G. and Finardi, S. (2015). Air pollution dispersion models validation dataset from complex terrain in Šoštanj. *Int. J. Environ. Pollut.*, 57(3-4), 227-237. https://doi.org/10.1504/IJEP.2015.074507

Mori, H. and Kurata, E. (2008). Application of gaussian process to wind speed forecasting for wind power generation. *2008 IEEE International Conference on Sustainable Energy Technologies*, Singapore, 956-959. https://doi.org/10.1109/ICSET.2008.4747145

Neal, L.S., Agnew, P., Moseley, S., Ordóñez, C., Savage, N.H. and Tilbee, M. (2014). Application of a statistical post-processing technique to a gridded, operational, air quality forecast. *Atmos. Environ.*, 98, 385-393. https://doi.org/10.1016/j.atmosenv.2014.09.004

Nelles, O. (2001). *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Springer. https://doi.org/10.1007/978-3-662-04323-3

Okumus, I. and Dinler, A. (2016). Current status of wind energy forecasting and a hybrid method for hourly predictions. *Energ. Convers. Manage.*, 123, 362-371. https://doi.org/10.1016/j.enconman.2016.06.053

Perne, M., Stepančič, M. and Grašič, B. (2019). Handling big datasets in gaussian processes for statistical wind vector prediction. *IFAC-*

Pap., 52(11), 110-115. https://doi.org/10.1016/j.ifacol.2019. 09.126

Potter, C.W. and Negnevitsky, M. (2006). Very short-term wind forecasting for Tasmanian power generation. *IEEE Trans. Power Syst*., 21(2), 965-972. https://doi.org/10.1109/TPWRS.2006.873421

Rasmussen, C.E. and Nickisch, H. (2010). Gaussian Process Regression and Classification Toolbox version 3.1 for GNU Octave 3.2.x and Matlab 7.x.

Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. Cambridge: MIT Press. https://doi.org/10. 7551/mitpress/3206.001.0001

Schielicke, L., Gatzen, C.P. and Ludwig, P. (2019). Vortex identification across different scales. *Atmosphere*, 10(9), 518. https://doi. org/10.3390/atmos10090518

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Duda, M.G., Huang, X., Wang, W. and Powers, J.G. (2008). *A Description of the Advanced Research WRF version 3*. NCAR Technical Note, NCAR/TN-475+STR. Boulder: National Center for Atmospheric Research. https://doi.org/10.5065/D68S4MVH

Stepančič, M. and Kocijan, J. (2017). On-line identification with regularised evolving Gaussian process. *2017 Evolving and Adaptive Intelligent Systems (EAIS)*, 1-7. https://doi.org/10.1109/EAIS.201 7.7954820

Sun, L., Ji, S. and Ye, J. (2014). *Multi-Label Dimensionality Reduction*. Chapman & Hall, CRC.

Testa, G., Isaak, H., von Arx, C. (2013). Future radiological dispersion calculations in Switzerland with JRODOS and LASAT. *15th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, Madrid.

Theodoridis, S. and Koutroumbas, K. (2006). *Pattern Recognition* (3rd ed.). San Diego: Academic Press.

Tinarelli, G., Anfossi, D., Trini Castelli, S., Bider, M. and Ferrero, E. (2000). A new high performance version of the Lagrangian par- ticle dispersion model spray, some case studies. *Air Pollution Modeling and Its Application XIII*. Springer, Boston, MA, pp 499-507. https://doi.org/10.1007/978-1-4615-4153-051

von Arx, C., Brücher, L., de Donno, V., Gerich, B., Glaab, H., Grimm, C., Martens, R., Päsler-Sauer, J., Scheuermann, W., Schnadt, H., Schumacher, P., Torchiani, S., Walter, H., Wilbois, T. (2014). Comparison of operational atmospheric dispersion models in Germany. *16th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, Varna, Bulgaria, 70-76.

Wagenbrenner, N.S., Forthofer, J.M., Lamb, B.K., Shannon, K.S., and Butler, B.W. (2016). Downscaling surface wind predictions from numerical weather prediction models in complex terrain with WindNinja. *Atmos. Chem. Phys*., 16, 5229-5241. https://doi.org/1 0.5194/acp-16-5229-2016

Wang, J. and Hu, J. (2015). A robust combination approach for short-term wind speed forecasting and analysis – Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy*, 93, 41-56. https://doi.org/10.1 016/j.energy.2015.08.045

Wang, J., Song, Y., Liu, F. and Hou, R. (2016). Analysis and application of forecasting models in wind power integration: A review of multi-step-ahead wind speed forecasting models. *Renew. Sust. Energ. Rev*., 60, 960-981. https://doi.org/10.1016/j.rser.2016.01.114

Yao, R., Sun, K., Liu, F., and Mei, S. (2017). Efficient simulation of temperature evolution of overhead transmission lines based on analytical solution and NWP. *IEEE Trans. Power Deliv*., 33(4), 1576-1588. https://doi.org/10.1109/TPWRD.2017.2751563