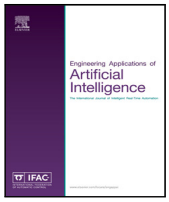
Engineering Applications of Artificial Intelligence 105 (2021) 104423

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Data-driven method for the improving forecasts of local weather dynamics

Tadej Krivec [a,b,*], Juš Kocijan [a,c], Matija Perne [a], Boštjan Grašic [d], Marija Zlata Božnar [d], Primož Mlakar [d]

[a] Jozef Stefan Institute, Jamova 39, Ljubljana, Slovenia
[b] Jožef Stefan International Postgraduate School, Slovenia
[c] University of Nova Gorica, Vipavska 13, Nova Gorica, Slovenia
[d] MEIS d.o.o., Mali Vrh pri Šmarju 78, Šmarje-Sap, Slovenia

ARTICLE INFO

Keywords:
Hybrid model
Atmospheric variables
Numerical weather prediction model
Statistical modeling
Gaussian process model

ABSTRACT

This paper describes the modeling approach for lower atmosphere dynamics in a selected location. The purpose of this model is to provide short-term and long-term forecasts of the weather variables which are used as the input data for the model of the dispersion of radioactive air pollution. The information from this integrated system is important for the implementation of the population safety measures in the case of a nuclear accident with an atmospheric release. We developed a dynamical, probabilistic, and non-parametric model based on Gaussian processes (GPs). GP nonlinear autoregressive model with exogenous inputs and variational training principle was implemented for multi-output training. A Monte Carlo approach to multi-output simulation of the model for long-term forecasts is presented which allows arbitrary prior distributions over function values. The model encompasses all available measurements from the weather stations near the location of interest and combines them with the forecasts from the numerical weather prediction model. The contribution of the developed model is the harvesting of all available information and simultaneously providing interconnected forecasts. The key result of this investigation is the improvement of short-term and long-term weather variable forecasts over those of the numerical weather prediction model. Consequently, we significantly enhance the dispersion forecast of radioactive air pollution for the case study considered. The computationally demanding modeling is accelerated using general-purpose computing on graphics processing units. The proposed method represents a step forward in the assurance of safety in the case of a nuclear accident.

1. Introduction

An accident at a nuclear power plant, which could cause the release of radionuclides into the atmosphere, are dangers that are handled with all seriousness by individual nuclear power plants. We must be prepared for action in advance in case of such events by researching the possible developments of the dispersion of pollution in the air. Appropriate online operating regime models for existing nuclear plants prepare real-time forecasts about the possible development of an event for a day or two ahead (Mlakar et al., 2019). Such a predictive model must spatially and temporally show correctly where the radionuclide cloud would disperse to in the next few hours for the available weather forecast. It should also predict the radionuclide concentrations within a three-dimensional space above the studied area. The input data for the air-pollution dispersion models are the weather variables which describe the condition of the atmosphere during the passage of the radioactive pollution cloud. At present, we can use the Numerical Weather Prediction (NWP) model to collect the forecasts of the weather variables, but these data are not accurate enough for complex terrain.

This motivates the investigation at hand, which is to make a few-hour forecast of the weather variables describing the conditions in the atmosphere surrounding a nuclear facility. We aim to provide predictions that are significantly better than the existing forecasts for complex terrain. A possible approach is to combine the NWP model with a statistical model to provide the input data of sufficient accuracy for the air-pollution dispersion model. The idea for the combination of models of different sorts has been utilized in the past in atmospheric sciences. It is named as statistical post-processing (Worsnop et al., 2018), model-output statistics (Kalnay, 2003) or integrated modeling (Gradišar et al., 2016). It is called hybrid modeling (Ren et al., 2018; Von Stosch et al., 2014) in mathematical modeling and system theory. The claimed advantage of hybrid models is that the physical insight into the atmosphere dynamics and the information about seasonality effects from the theoretical model, i.e., global physics-based model are retained, while

* Corresponding author at: Jozef Stefan Institute, Jamova 39, Ljubljana, Slovenia.
E-mail addresses: tadej.krivec@ijs.si (T. Krivec), jus.kocijan@ijs.si (J. Kocijan), matija.perne@ijs.si (M. Perne), bostjan.grasic@meis.si (B. Grašic), marija.zlata.boznar@meis.si (M.Z. Božnar), primoz.mlakar@meis.si (P. Mlakar).

https://doi.org/10.1016/j.engappai.2021.104423
Received 3 August 2020; Received in revised form 26 July 2021; Accepted 4 August 2021
Available online xxxx

Nomenclature

Acronyms

ARD	Automatic relevance determination
ELBO	Evidence lower bound
GFS	Global forecasting model
GMM	Gaussian mixture model
GP	Gaussian process
KL	Kullback–Leibler
LIN	Linear
MAT	Matérn
MSLL	Mean standardized log loss
NARX	Nonlinear autoregressive model with exogenous inputs
NOE	Nonlinear output-error model
NPP	Nuclear power plant
NRMSE	Normalized root-mean-square error
NWP	Numerical weather prediction
PCC	Pearson correlation coefficient
RBF	Radial basis function
UTC	Universal time coordinated
VFE	Variational free energy
WRF	Weather research and forecast

Abbreviations

B_AH	Brezice absolute humidity measurement
B_AT	Brezice air temperature measurement
C_AH	Cerklje absolute humidity measurement
C_AP	Cerklje air pressure measurement
C_AT	Cerklje air temperature measurement
CA_AH	Cerklje airport absolute humidity measurement
CA_AP	Cerklje airport air pressure measurement
CA_AT	Cerklje airport air temperature measurement
K_AH	Krsko absolute humidity measurement
K_AT	Krsko air temperature measurement
L_AH	Libna absolute humidity measurement
L_AT	Libna air temperature measurement
S_AH2	Stolp absolute humidity measurement at 2 m
S_AP	Stolp air pressure measurement
S_AT10	Stolp air temperature measurement at 10 m
S_AT2	Stolp air temperature measurement at 2 m
S_AT40	Stolp air temperature measurement at 40 m
S_AT70	Stolp air temperature measurement at 70 m
S_GSR	Stolp global solar radiation measurement
S_WX	Stolp wind component x measurement at 10 m
S_WY	Stolp wind component y measurement at 10 m
WRF_AH	WRF model absolute humidity prediction
WRF_AP	WRF model air pressure prediction
WRF_AT	WRF model air temperature prediction
WRF_CC	WRF model cloud cover prediction
WRF_DSR	WRF model diffuse solar radiation prediction
WRF_GSR	WRF model global solar radiation prediction
WRF_WX	WRF model wind component x prediction
WRF_WY	WRF model wind component y prediction

the deviations over the complex terrain are compensated for by the experimental, i.e., statistical model.

Studies of the dispersion of radioactive pollution for various sources are frequent, e.g., Cai et al. (2019), Souissi et al. (2019) and Saeed et al. (2020). Common to such studies is that they need reliable data of atmosphere dynamics, which can be problematic in the case of complex terrain. Many statistical and machine learning methods can be used to model the remaining deviations from the physical model. A popular approach is with artificial neural networks which were previously used in a wide variety of applications (Lee, 2003; Nguyen-Le et al., 2020; Zenzen et al., 2020; Khatir et al., 2020). For modeling the atmosphere dynamics this includes precipitation forecasting (Kuligowski and Barros, 1998; Hall et al., 1999; Pan et al., 2019), temperature forecasting (Hayati and Mohebi, 2007; Al-Shawwa et al., 2018; Khotanzad et al., 1996), geopotential height (Scher, 2018), and wind forecasting (Fan et al., 2008; Barbounis et al., 2006; Karniotakis et al., 1996; More and Deo, 2003; Hewage et al., 2019). An alternative is a Gaussian Process (GP) model, which is a probabilistic, non-parametric, and kernel-method-based model. It achieves state-of-the-art performance in a range of applications (Ko and Fox, 2009; Deisenroth and Rasmussen, 2011; Diggle et al., 1998; Snoek et al., 2012; Guestrin et al., 2005). In atmospheric sciences, they were generally applied to wind forecasting (Chen et al., 2013; Zhang et al., 2016; Hoolohan et al., 2018; Cai et al., 2020). The preliminary results of the probabilistic hybrid modeling for the selected single weather variables for the same case study as in this investigation can be found in Kocijan et al. (2019) for the temperature profile and in Kocijan et al. (2020) for the relative humidity. These studies investigated different sorts of hybrid-model structures and the viability of using different GP models in them.

The limitation of deterministic and highly flexible methods in the listed publications (e.g., artificial neural networks) is that they cannot quantify the uncertainty of the prediction. This poses a significant drawback in decision-making systems, such as in devising the appropriate population safety measures in the case of a nuclear accident with an atmospheric release. Rich deterministic structure requires an appropriate set of meta-parameters that is obtained through cross-validation, which introduces additional computational complexity and abstraction. Bayesian modeling and probabilistic treatment of results is reasonable in such case (Gelman et al., 2004).

GPs define a rich probabilistic model and let the data automatically penalize overly complex models. A review of modeling with GPs can be found in Liu et al. (2020). Generally, GPs can be combined with deep learning (Wilson et al., 2016), but they inherit the pathological problems from the neural networks (optimization over a large set of model parameters). The non-parametric structure of a GP reduces the number of meta-parameters and automatically scales the model complexity with data size. They provide a well-calibrated predictive distribution at a model output, which can be used to quantify the model fidelity systematically (Kocijan, 2016). They are used as a de facto standard for probabilistic regression and are equivalent to an infinite width Bayesian neural network (Matthews et al., 2018). The generative nature of the simulation also provides the samples of future trajectories, which represent the potential scenarios that we have to be prepared for (i.e., in the case of an atmospheric release). The limitation of the preliminary studies of using GPs to model atmospheric variables (for the case study considered) is that they cannot use all the available data in a single integrated model. They either use a full GP model, which has a cubic complexity with respect to the training data size and, therefore, only a subset of the available data is used for model training. Or, they use a multi-input single-output autoregressive structure, which does not permit simulations when all historical measured variables are used in the inputs. This structure can only be used for prediction, since the future measured weather variables are not known. A multi model approach to simulation can resolve this problem, but is inefficient

in handling large models. Existing hybrid GP models used for wind forecasting also use a full GP model and are, therefore, limited with the amount of data they can process. This poses a significant limitation in the case of modeling the atmospheric variables around complex terrain. They also do not consider simulation in the autoregressive manner and only investigate forecasts up to a limited time horizon.

The issue that is investigated in this paper is the integration of the NWP model with the multi-input multi-output variational GP Nonlinear AutoRegressive model with Exogenous inputs (GP-NARX). Resulting single integrated probabilistic hybrid model provides a reliable few-hour forecast of the weather variables that can be used for modeling the dispersion of the radioactive pollution. Consequently, the prevention actions in the case of a nuclear disaster can be devised accordingly. Briefly, the novel contributions of this paper are:

1. We address the computational limitation of the full GPs used in the previous studies with a variational approximation (Titsias, 2009b), extended for multi-input multi-output training with separate likelihood models on the outputs.
2. Monte Carlo multi-input multi-output simulation up to an infinite time horizon (we assume the NWP forecasts are given) of the proposed model in an autoregressive manner for arbitrary prior distributions on function values, which allows us to validate the model also for simulation.
3. General-purpose computing on graphics processing units support for the simulation of the proposed model on large validation/test datasets which are common in atmospheric sciences.
4. The use of all available data for the improvement of short-term (half-hour) and long-term (few-hour) weather variable forecasts over those of the NWP model. Consequently, the improvement of the dispersion model of the radioactive air pollution which could save potential lives in the case of a nuclear disaster.

The paper is structured as follows. The method for the data-driven part of the hybrid model, the used physics-based model, and the dispersion model are described in Section 2. Section 3 describes the details of the case study of interest and the selected model for solving the meteorological problem. The modeling results are presented and discussed in Section 4. The conclusions are gathered at the end of the paper.

2. Methods

In critical scenarios, as in the case of advising the prevention actions of a potential nuclear disaster, estimation of the predictive uncertainty of the model is critical. Additionally, we want to use all available information (e.g., the measured weather variables) at near geographical locations. The weather variables describing the atmosphere dynamics (preprocessed and extended with additional variables) serve as an input to a Lagrangian particle dispersion model, which forecasts the air pollution dispersion into the near future. To improve the forecasts of the NWP model and provide well-calibrated predictive uncertainty, we propose a probabilistic hybrid model to model the atmosphere dynamics. The high-level modeling approach that we propose is presented on Fig. 1.

2.1. Probabilistic hybrid model

A probabilistic hybrid model is described in this section, which combines all the available data, i.e., the forecasts of the weather variables from the physical NWP model and the historic measured weather variables. This single integrated model aims to improve the weather variable forecast for the complex terrain of interest.

2.1.1. Numerical weather prediction model

Among various NWP models, the Weather Research and Forecast (WRF) model (Advanced Research WRF version 3.4.1) was used in our investigation. The primary purpose of the WRF model is operational forecasting. It enables experimentation with various physical models and different simulations (Skamarock et al., 2008). The WRF model used in our case study, described in Božnar et al. (2012), covers the central part of Europe as well as Slovenia with its surroundings. The domain that covers the central part of Europe consists of 101 by 101 cells, each cell the size of 12 km with a model temporal resolution of 3 h. The domain that covers Slovenia consists of 76 by 76 cells, each cell the size of 4 km with a model temporal resolution of 0.5 h. Initial and boundary conditions of the used WRF model are provided by the Global Forecasting Model (GFS). The model forecasts are produced twice per day, at 00:00 Universal Time Coordinated (UTC) and 12:00 UTC, where the new forecasted data overwrite the previous data for the overlapping time intervals. Nevertheless, the provided spatial resolution is not high enough to encompass all the local weather effects due to the complex terrain of interest.

2.1.2. GP model

A GP model is considered to improve the NWP forecasts in the complex terrain of interest. The problem is defined by

$$\mathbf{y} = f(\mathbf{X}) + \epsilon, \tag{1}$$

where $\mathbf{X} \in \mathcal{R}^{n\times d}$ represents the matrix of observed inputs, f the latent function, and $\mathbf{y} \in \mathcal{R}^n$ is the observed output vector, where n and d are the number of data points and the dimensionality of the input column space respectively. The observed output vector is presumed to be corrupted with independent and identically distributed Gaussian noise defined by $\epsilon_i \sim \mathcal{N}(0, \sigma_n^2)$. Vector $\mathbf{f} = [f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n)]^T$ defines the vector of latent function values where $\mathbf{x}_n$ represents the nth row of $\mathbf{X}$. The prior over the vector of latent function values $\mathbf{f}$ is described with a GP. A GP is defined as a collection of random variables, any finite number of which have a joint Gaussian distribution (Rasmussen and Williams, 2006).

It is specified by a mean function $m(\mathbf{x}_i)$ and a covariance function $k(\mathbf{x}_i, \mathbf{x}_j)$

$$m(\mathbf{x}_i) = \mathbb{E}[f(\mathbf{x}_i)], \tag{2a}$$

$$k(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{E}[(f(\mathbf{x}_i) - m(\mathbf{x}_i))(f(\mathbf{x}_j) - m(\mathbf{x}_j))]. \tag{2b}$$

The distribution over the vector of latent function values $\mathbf{f}$ is defined by

$$p(\mathbf{f}|\mathbf{X}, \theta) = \mathcal{N}(\begin{bmatrix} m(\mathbf{x}_1) \\ \vdots \\ m(\mathbf{x}_n) \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \dots & k(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1) & \dots & k(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix}) = \mathcal{N}(\boldsymbol{\mu}_f, \mathbf{K}_{ff}), \tag{3}$$

where θ represents the set of hyper-parameters, i.e., the parameters of the mean and the covariance function. The likelihood is defined by $p(\mathbf{y}|\mathbf{f}, \sigma_n^2) = \mathcal{N}(\mathbf{f}, \sigma_n^2\mathbf{I})$, where $\mathbf{y}$ is a noisy observation of $\mathbf{f}$ with likelihood variance σ_n^2. Without the loss of generality, the mean function is often selected as $m(\mathbf{X}) = \mathbf{0}$. More important is the selection of the covariance function as it incorporates our prior beliefs and assumptions of the latent function. Some popular covariance functions are defined in the Appendices and many others can be found in Duvenaud (2014) and Kocijan (2016). The parameters of the covariance function, i.e. the hyper-parameters, are denoted by θ.

Learning the hyper-parameters

Let $\mathbf{f}_*$ denote the vector of latent function values at the unobserved input $\mathbf{x}_*$. The joint distribution $p(\mathbf{f}, \mathbf{f}_*|\mathbf{X}, \mathbf{x}_*, \theta)$ is Gaussian and defined by

$$p(\mathbf{f}, \mathbf{f}_*|\mathbf{X}, \mathbf{x}_*, \theta) = \mathcal{N}(\mathbf{0}, \begin{bmatrix} \mathbf{K}_{ff} & \mathbf{K}_{f*} \\ \mathbf{K}_{*f} & \mathbf{K}_{**} \end{bmatrix}), \tag{4}$$

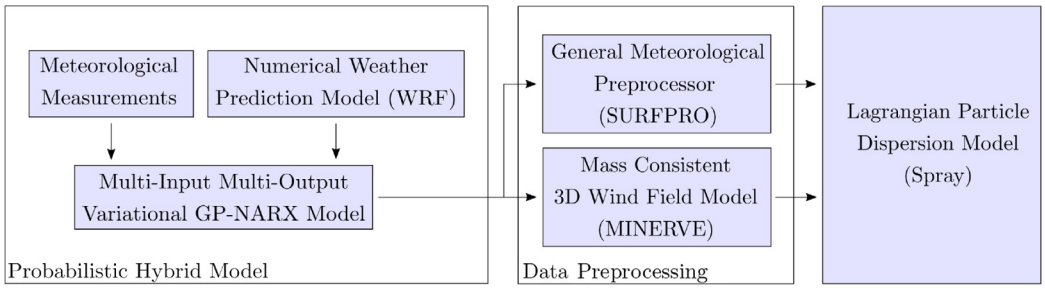


Fig. 1. The proposed modeling approach for the model of the air pollution dispersion at a chosen geographical location. The existing approach uses the forecasts from the NWP model directly as the inputs to the data preprocessing software. The outputs from the data preprocessing software are later used in a Lagrangian particle dispersion model. In our proposed approach the NWP forecasts are enhanced with a data-driven model resulting in a hybrid model. In such case the data preprocessing software uses the forecasts from the hybrid model which are then provided to the Lagrangian particle dispersion model.

where $\mathbf{K}_{ff}, \mathbf{K}_{f*}, \mathbf{K}_{**}$ represent the covariance matrices between the training inputs, training the test inputs, and test inputs respectively. Joint distribution in Eq. (4) defines a prior, which is transformed to the posterior given the observed data

$$p(\mathbf{f}, \mathbf{f}_*|\mathbf{X}, \mathbf{x}_*, \mathbf{y}, \boldsymbol{\theta}) = \frac{p(\mathbf{y}|\mathbf{f}, \sigma_n^2)p(\mathbf{f}, \mathbf{f}_*|\mathbf{X}, \mathbf{x}_*, \boldsymbol{\theta})}{p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}, \sigma_n^2)}. \tag{5}$$

Hyper-parameters can be determined with the maximization of the marginal log-likelihood (Rasmussen and Williams, 2006). Evaluating the marginal log-likelihood has a computational complexity of $\mathcal{O}(n^3)$, which practically limits the use of GPs for large datasets. Hereafter we omit the conditional dependency on $\mathbf{X}, \mathbf{x}_*$, $\boldsymbol{\theta}$, and σ_n^2 in notation for convenience.

Prediction

We obtain the predictive posterior by integrating the joint posterior out of the likelihood at test inputs $\mathbf{x}_*$

$$p(\mathbf{y}_*|\mathbf{y}) = \iint p(\mathbf{y}_*|\mathbf{f}_*)p(\mathbf{f}, \mathbf{f}_*|\mathbf{y})d\mathbf{f}\,d\mathbf{f}_*. \tag{6}$$

The integral can be evaluated in closed-form (Rasmussen and Williams, 2006).

2.1.3. Multi-input multi-output variational GP-NARX model

To reduce the computational complexity of GP models, sparse approximations consider m pseudo-inputs at location $\mathbf{x}_m$ with the corresponding vector of latent function values $\mathbf{u} = [u_1, \dots, u_m]$ that is jointly Gaussian distributed with vectors of latent function values $\mathbf{f}$ and $\mathbf{f}_*$. Assuming conditional independence between $\mathbf{f}$ and $\mathbf{f}_*$ given $\mathbf{u}$, their joint prior is approximated by

$$p(\mathbf{f}, \mathbf{f}_*) \cong \int p(\mathbf{f}|\mathbf{u})p(\mathbf{f}_*|\mathbf{u})p(\mathbf{u})d\mathbf{u}, \tag{7}$$

where the distributions over the vector of latent function values conditioned on pseudo-inputs are given by

$$p(\mathbf{f}|\mathbf{u}) = \mathcal{N}(\mathbf{K}_{fu}\mathbf{K}_{uu}^{-1}\mathbf{u}, \mathbf{K}_{ff} - \mathbf{Q}_{ff}), \tag{8a}$$

$$p(\mathbf{f}_*|\mathbf{u}) = \mathcal{N}(\mathbf{K}_{*u}\mathbf{K}_{uu}^{-1}\mathbf{u}, \mathbf{K}_{**} - \mathbf{Q}_{**}), \tag{8b}$$

and $\mathbf{Q}_{ab} = \mathbf{K}_{au}\mathbf{K}_{uu}^{-1}\mathbf{K}_{ub}$. Given the conditionals, the model is completely specified and has a closed-form solution. The computational complexity of sparse approximations can be reduced to $\mathcal{O}(nm^2)$ with further assumptions. An overview of sparse approximation methods can be found in Quiñonero-Candela and Rasmussen (2005), but these methods do not rigorously minimize the difference between the full and the approximated GP. In the next section we describe a variational method that resolves this problem. Additionally, we introduce a dynamical and multi-output model structure for the problem we consider.

Introducing dynamic properties and multiple outputs

To model the atmosphere dynamics, the multi-output problem is defined by

$$\mathbf{Y} = f(\mathbf{Z}) + \boldsymbol{\epsilon}, \tag{9}$$

where in general $\mathbf{Y} \in \mathcal{R}^{n\times r}$ represents a multi-dimensional output matrix and the mapping f is modeled with a GP. The input matrix $\mathbf{Z} \in \mathcal{R}^{n\times(r\cdot n_a + d\cdot n_b)}$ is represented with a NARX model. The ith row of the input matrix $\mathbf{Z}$ is defined by

$$\mathbf{z}_i = [\mathbf{y}_{i-n_p}, \dots, \mathbf{y}_{i-n_p-n_a}, \mathbf{x}_{i-n_k}, \dots, \mathbf{x}_{i-n_k-n_b}], \tag{10}$$

where $\mathbf{x}_i \in \mathcal{R}^{1\times d}$ and $\mathbf{y}_i \in \mathcal{R}^{1\times r}$ define the ith rows of the matrix with exogenous inputs $\mathbf{X}$ and the output matrix $\mathbf{Y}$ respectively. Meta-parameters n_p, n_k, n_a, and n_b denote the number of delayed and lagged samples from the multi-dimensional input and output vectors. Eq. (10) defines a GP-NARX model presented on Fig. 2a.

In our case, the exogenous inputs $\mathbf{X}$ represent the forecasts from the physical NWP model and $\mathbf{Y}$ represents a matrix of measured weather variables in the vicinity of the targeted location. They are measured in half-hour intervals. The approach, where the mapping is modeled with a GP, exogenous inputs are selected as the forecasts from a physical model, and the outputs are the measured weather variables, is called hybrid modeling.

Variational learning of hyper-parameters

Introducing a dynamical structure has very little practical implications on learning the hyper-parameters. All the input locations previously denoted as $\mathbf{X}, \mathbf{x}_*, \mathbf{x}_m$ are replaced with $\mathbf{Z}, \mathbf{z}_*, \mathbf{z}_m$ defined in Eq. (10), thus implying a dynamical model structure. The model $p(\mathbf{y}, \mathbf{f}, \mathbf{u})$ defines the marginal log-likelihood by

$$\log\, p(\mathbf{y}) = \log \iint p(\mathbf{f}, \mathbf{u}, \mathbf{y})d\mathbf{f}\,d\mathbf{u}. \tag{11}$$

Lower bounding the marginal log-likelihood gives the following evidence lower bound (ELBO)

$$\log\, p(\mathbf{y}) \geq l(q(\mathbf{u}), \boldsymbol{\theta}, \mathbf{z}_m) = \mathbb{E}_{q(\mathbf{f})}[\log\, p(\mathbf{y}|\mathbf{f})] - \mathrm{KL}[q(\mathbf{u}) \parallel p(\mathbf{u})], \tag{12}$$

where $q(\mathbf{f}) = \int p(\mathbf{f}|\mathbf{u})q(\mathbf{u})d\mathbf{u}$. The optimal variational distribution $q(\mathbf{u}) \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Lambda}^{-1})$ that maximizes $l(q(\mathbf{u}), \boldsymbol{\theta}, \mathbf{z}_m)$ can be evaluated in closed-form and is specified by

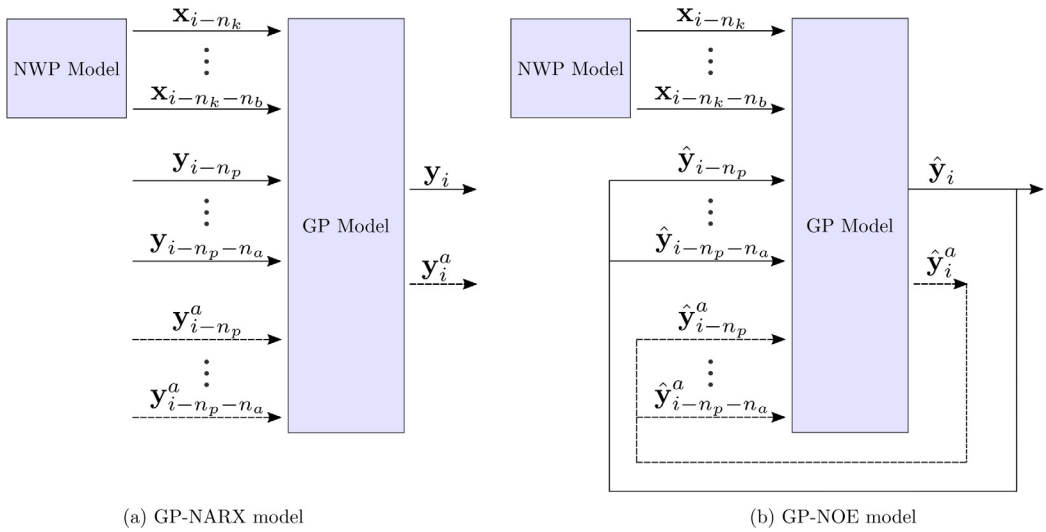
$$\mathbf{m} = \boldsymbol{\Lambda}^{-1}\mathbf{K}_{uu}^{-1}\mathbf{K}_{uf}\mathbf{y}\sigma_n^{-2}, \tag{13a}$$

$$\boldsymbol{\Lambda} = \mathbf{K}_{uu}^{-1} + \mathbf{K}_{uu}^{-1}\mathbf{K}_{uf}\mathbf{K}_{fu}\mathbf{K}_{uu}^{-1}\sigma_n^{-2}. \tag{13b}$$

ELBO after the maximization with respect to $q(\mathbf{u})$ is defined by

$$l(\boldsymbol{\theta}, \mathbf{z}_m) = \log\,[\mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{Q}_{ff} + \sigma_n^2\mathbf{I})] - \frac{1}{2\sigma_n^2}tr(\mathbf{K}_{ff} - \mathbf{Q}_{ff}). \tag{14}$$

The details of the derivation of the optimal variational distribution and the ELBO can be found in Titsias (2009b,a). In the case of a



(a) GP-NARX model (b) GP-NOE model

Fig. 2. Figure (a) is representing a GP-NARX model where the exogenous inputs are in our investigation selected as the forecasts from the physical NWP model. The outputs $\mathbf{y}_i$ describe the weather variables of interest and $\mathbf{y}_i^a$ the additional measurements in the geographical vicinity of the NPP Krško. Figure (b) is representing a GP Nonlinear Output Error (GP-NOE). The model is used to obtain a multiple step-ahead prediction where the outputs $\hat{\mathbf{y}}_i$ and $\hat{\mathbf{y}}_i^a$ are random variables and represent the estimated values.

multi-dimensional output, the bound expands to

$$l_{mo}(\boldsymbol{\theta}, \mathbf{z}_m) = \sum_{i=1}^{r} \Big(\log\left[\mathcal{N}(\mathbf{y}^i|\mathbf{0}, \mathbf{Q}_{ff} + (\sigma_n^i)^2\mathbf{I})\right] - \frac{1}{2(\sigma_n^i)^2} tr(\mathbf{K}_{ff} - \mathbf{Q}_{ff})\Big), \tag{15}$$

where $\mathbf{y}^i$ represents the ith column of the output matrix $\mathbf{Y}$ and σ_n^i is the corresponding standard deviation of the output with the dimension of the output column space r. This ELBO implies that the outputs are independent given the hyper-parameters and the inputs. Marginal log-likelihoods for separate outputs are parametrized by the same pseudo-inputs $\mathbf{z}_m$ and hyper-parameters $\boldsymbol{\theta}$ of the covariance function, but have independent likelihood noise level σ_n^i. Compared to the approximations in Quiñonero-Candela and Rasmussen (2005), the difference between the approximated GP model and the full GP model is rigorously defined. Variational approximation always improves with the addition of new inducing inputs (Bauer et al., 2016). The number of pseudo-inputs m can, therefore, be seen as a true trade-off parameter between the accuracy and the computational complexity of the model.

Prediction

The predictive distribution is obtained by marginalizing over the free variational distribution and the latent function. Following the assumption of independent outputs given the inputs and hyper-parameters, we get the prediction of the ith output by

$$p(\mathbf{y}_*^i) = \iint p(\mathbf{y}_*^i|\mathbf{f}_*^i) p(\mathbf{f}_*^i|\mathbf{u}) q(\mathbf{u}) d\mathbf{f}_*^i d\mathbf{u}. \tag{16}$$

The integral has a closed-from solution. The mean and the variance of the predictive distribution are defined by

$$\mu(p(\mathbf{y}_*^i)) = \mathbf{K}_{*u}\mathbf{K}_{uu}^{-1}\mathbf{m}^i, \tag{17a}$$

$$\sigma^2(p(\mathbf{y}_*^i)) = \mathbf{K}_{**} - \mathbf{K}_{*u}\mathbf{K}_{uu}^{-1}\mathbf{K}_{u*} + \mathbf{K}_{*u}\mathbf{K}_{uu}^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{K}_{uu}^{-1}\mathbf{K}_{u*} + (\sigma_n^i)^2\mathbf{I}, \tag{17b}$$

where $\mathbf{m}^i = \boldsymbol{\Lambda}^{-1}\mathbf{K}_{uu}^{-1}\mathbf{K}_{uf}\mathbf{y}^i(\sigma_n^i)^{-2}$ and $\boldsymbol{\Lambda}$ is defined by Eq. (13b).

Multiple step-ahead prediction

The first step (denoted $t+1$) is identical to the prediction defined with Eqs. (17). Assuming $n_p = 1$, the input vector at the time step $t+2$ is defined by

$$\mathbf{z}_{t+2} = [\mathbf{y}_{t+1}, \dots, \mathbf{y}_{t+1-n_a}, \mathbf{x}_{t+2-n_k}, \dots, \mathbf{x}_{t+2-n_k-n_b}], \tag{18}$$

where $\mathbf{y}_{t+1}$ follows a Gaussian distribution (i.e. the predictive distribution from step $t+1$). The input is therefore no longer deterministic, but rather an uncertain input and makes the integral

$$p(\mathbf{y}_{t+2}) = \iiint p(\mathbf{y}_{t+2}|\mathbf{f}_{t+2}) p(\mathbf{f}_{t+2}|\mathbf{z}_{t+2}, \mathbf{u}) p(\mathbf{z}_{t+2}) q(\mathbf{u}) d\mathbf{f}_{t+2} d\mathbf{z}_{t+2} d\mathbf{u} \tag{19}$$

intractable. The predictive distribution at the time step $t+2$ can be approximated in the form of a Gaussian Mixture Model (GMM) defined by

$$p(\hat{\mathbf{y}}_{t+2}) = \frac{1}{m}\sum_{i=1}^{m} \iint p(\mathbf{y}_{t+2}|\mathbf{f}_{t+2}) p(\mathbf{f}_{t+2}|\hat{\mathbf{z}}_{t+2}^i, \mathbf{u}) q(\mathbf{u}) d\mathbf{f}_{t+2} d\mathbf{u}, \tag{20}$$

where

$$\hat{\mathbf{z}}_{t+2}^i = [\hat{\mathbf{y}}_{t+1}^i, \dots, \mathbf{y}_{t+1-n_a}, \mathbf{x}_{t+2-n_k}, \dots, \mathbf{x}_{t+2-n_k-n_b}], \tag{21}$$

and $\hat{\mathbf{y}}_{t+1}^i$ is a sample from a Gaussian distribution predicted from the previous step. The number of samples is denoted with m. This process can be repeated up to an arbitrary time step into the future where the samples of $\hat{\mathbf{y}}_{t+q}^i$ are drawn from a GMM instead of a Gaussian distribution for all $q \geq 2$. This iterative approach is presented on Fig. 2b with a GP-NOE model.

2.2. Mass consistent wind field model and meteorological pre-processor

In the complex terrain like the Krško basin, the hybrid model forecasts cannot be used directly for the dispersion modeling because their spatial resolution is not high enough. To obtain the required spatial resolution, a mass consistent wind field model (i.e. MINERVE (Desiato et al., 1998; Geai, 1987)) was used to calculate a three-dimensional wind field needed for dispersion modeling. The weather variables were augmented with additional variables (stability and turbulent diffusivity fields) using SURFPRO (Silibello et al., 2006) diagnostic meteorological preprocessor for the complex terrain.

2.3. Lagrangian particle dispersion model

The extended and the preprocessed weather variables from the mass consistent wind field model and meteorological preprocessor serve as the inputs to a Lagrangian particle dispersion model. SPRAY (Tinarelli et al., 2000; Tinarelli, 2007) is a program package containing a Lagrangian particle dispersion model developed to perform dispersion simulations in complex terrain. The model is based on a

Table 1
The data used in the case study with their respective means and standard deviations. Exogenous inputs are represented with **x**, the modeled outputs with **y**, and additional weather variable measurements in the near geographical location with $\mathbf{y}^a$.

x		**y**		$\mathbf{y}^a$	
Abbr.	$\mu \pm \sigma$	Abbr.	$\mu \pm \sigma$	Abbr.	$\mu \pm \sigma$
WRF_AH	7.82 ± 3.41 g/m^3	S_WX	−0.01 ± 1.03 m/s	B_AT	11.55 ± 9.20 °C
WRF_WX	0.02 ± 1.46 m/s	S_WY	−0.37 ± 1.68 m/s	B_AH	8.30 ± 3.58 mbar
WRF_WY	−0.40 ± 2.21 m/s	S_AT2	11.08 ± 9.32 °C	C_AT	11.12 ± 9.10 °C
WRF_CC	30.18 ± 40.57%	S_AT10	11.34 ± 9.11 °C	C_AH	8.50 ± 3.80 g/m^3
WRF_AT	10.56 ± 8.60 °C	S_AT40	11.73 ± 9.03 °C	C_AP	999.55 ± 7.97 mbar
WRF_GSR	176.98 ± 271.03 W/m^2	S_AT70	11.73 ± 8.91 °C	CA_AT	10.78 ± 9.32 °C
WRF_DSR	57.10 ± 81.70 W/m^2	S_AH2	8.23 ± 3.64 g/m^3	CA_AH	8.04 ± 3.61 g/m^3
WRF_AP	997.11 ± 7.80 mbar	S_AP	998.98 ± 7.93 mbar	CA_AP	998.27 ± 7.93 mbar
				K_AT	11.42 ± 9.15 °C
				K_AH	8.40 ± 3.67 g/m^3
				L_AT	12.05 ± 9.04 °C
				L_AH	8.11 ± 3.53 g/m^3
				S_GSR	998.98 ± 7.93 W/m^2

three-dimensional form of the Langevin equation for the random velocity with coupled nongaussian random forcing. The extended and preprocessed weather variables described in Section 2.2 are used as inputs to such a model.

The atmospheric dispersion modeled with Lagrangian particle dispersion model is simulated by the motion of fictitious particles. The motion is split in two parts: A mean core due to the mean wind flow and a stochastic fluctuation related to the statistical characteristics of the turbulent flow. The accuracy of the simulation strongly depends on the number of emitted particles. The number of emitted particles is, however, constrained with the processing ability of the used computer system.

It was validated in Tinarelli et al. (2000) that SPRAY satisfactorily reproduces local to regional scale dispersion over both the flat and the complex terrain from a single or multiple emission sources.

3. Case study

The example in our investigation deals with the dispersion of radioactive pollutants hypothetically emitted from the Krško Nuclear Power Plant (NPP), located in the East of Slovenia. The modeling problem is of pronounced interest because of the surrounding terrain which is considered to be complex. It is surrounded by hills, valleys, a river, and has different land use, e.g. urban, fields, forests, water bodies, etc. The investigated area covers 25 square kilometers around the Krško NPP. Data used in the investigation are composed of the measurements of the historic weather variables and the NWP model forecasts. The measurements of the weather variables are collected by automatic measurement stations at five different locations. One measurement station is located at Krško NPP and the other five distributed around in the settlements of Brežice, Cerklje, Krško, Libna, and at Cerklje airport. Listed automatic measuring stations take real-time measurements and are averaged over 30-minute intervals. Characteristics of the measurement equipment can be found in Mlakar et al. (2019). The available measured weather variables are: wind speed and direction, temperature, humidity, air pressure, global solar radiation. An additional signal, diffuse solar radiation, is derived from the NWP model variables using a soft sensor (Božnar et al., 2017).

The same weather variables are also provided as the NWP forecasts for this region. The datasets used in our investigation consist of data from the years 2015 through 2017. The data from the year 2015 and the first half of 2016 are used for training and the data from the year 2017 as the test data. Some data gaps in the training data are omitted, which does not influence the results considerably. This resulted in the training set size of $n = 8955$ samples, and the test set size of $n = 17\,519$ samples. The training and test splits are reasonable for atmospheric sciences. It is especially important that the size of the test dataset contains a large amount of data from different seasons with different weather patterns. All the available measured weather variables and the NWP model forecasts are described in Table 1.

Meta-parameters were obtained with 10-fold cross-validation over the training dataset using mean standardized log loss (MSLL) as a selection metric. MSLL is suitable for validation in the form of random variables, weighting the error by the predicted standard deviation. When modeling with the variational GP-NARX models the meta-parameters of the autoregressive model, i.e., the parameters of a NARX model presented with Eq. (10), were firstly identified. The model capability was not improved significantly with an increasing number of lags. A second order system was found as an optimal trade-off between the modeling capability and the computational complexity. The meta-parameters of the NARX model were selected as $(n_p, n_a, n_k, n_b) = (0, 2, 1, 2)$.

Secondly, the meta-parameters (i.e, the covariance function and the number of pseudo-inputs) and the hyper-parameters of the GP model were identified. Table 2 shows the proposed covariance functions and the number of pseudo-inputs used in this empirical process where the results are averaged over 10-fold cross-validation. All of the data described in Table 1 was used in combination with the Automatic Relevance Determination (ARD) property of the covariance function that is able to automatically weight the input data. All covariance functions used in this process are defined in the Appendices. Hyper-parameters were obtained with the maximization of marginal log-likelihood using Adam (Kingma and Ba, 2014), where the initial locations of pseudo-inputs were initialized with the k-means clustering algorithm as suggested in Bauer et al. (2016). Table 2 shows that the best covariance function according to the MSLL for the half-hour prediction was a combination of a linear and a Matérn kernel. Table 2 also demonstrates how increasing the number of pseudo-inputs improved the MSLL for the half-hour prediction.

Lastly, two possible choices of the output vector (defined in Table 1) were considered:

1. Only the weather variables of interest **y**.
2. An augmented version of the outputs, where the measurements of the weather variables in the near geographical vicinity $\mathbf{y}^a$ are also included (they can be seen as additional latent variables to the model).

The augmented version allows the use of all available information on the cost of estimating a higher dimensional input/output vector. When validating the augmented model, only the weather variables of interest were considered. According to the MSLL provided in Table 3, the model was improved with the augmented weather variables $\mathbf{y}^a$.

4. Results and discussion

In this section, the results of modeling the atmospheric weather variables are presented on the test dataset. Firstly, the model was tested for a short-term and a long-term forecast, i.e., a half-hour and

Table 2
Selection of the number of pseudo-inputs and the covariance function for modeling the weather variables of interest, where the best results and the overall best choice are emphasized. The measure used is the mean standardized log loss for the half-hour prediction averaged over 10-fold cross-validation.

	m	S_WX	S_WY	S_AT2	S_AT10	S_AT40	S_AT70	S_AH2	S_AP
LIN	50	−0.645	−1.446	−4.640	−4.705	−4.538	−4.468	−2.990	−7.052
	100	−0.645	−1.445	−4.639	−4.703	−4.536	−4.465	−2.989	**−7.079**
	200	−0.645	−1.445	−4.639	−4.703	−4.536	−4.465	−2.989	**−7.079**
RBF	50	−0.646	−1.443	−4.629	−4.700	−4.532	−4.464	−2.976	−6.868
	100	−0.648	−1.448	−4.669	−4.740	−4.565	−4.497	−3.000	−7.004
	200	−0.648	−1.448	−4.670	−4.741	−4.566	−4.499	−3.001	−7.036
LIN + RBF	50	−0.645	−1.445	−4.641	−4.706	−4.538	−4.468	−2.990	−7.052
	100	−0.645	−1.445	−4.639	−4.703	−4.536	−4.465	−2.989	**−7.079**
	200	−0.645	−1.445	−4.639	−4.703	−4.536	−4.465	−2.989	**−7.079**
LIN + RBF (ARD)	50	−0.645	−1.445	−4.640	−4.705	−4.538	−4.467	−2.989	−7.052
	100	−0.645	−1.445	−4.639	−4.703	−4.536	−4.465	−2.989	**−7.079**
	200	−0.645	−1.445	−4.639	−4.703	−4.536	−4.465	−2.989	**−7.079**
MAT52	50	−0.645	−1.442	−4.601	−4.670	−4.508	−4.441	−2.963	−6.721
	100	−0.648	−1.447	−4.663	−4.733	−4.559	−4.492	−2.995	−6.963
	200	**−0.649**	−1.449	−4.676	−4.748	−4.569	−4.503	−3.007	−6.990
LIN + MAT52	50	−0.645	−1.445	−4.640	−4.704	−4.537	−4.467	−2.989	−7.043
	100	−0.646	−1.447	−4.660	−4.727	−4.554	−4.486	−2.996	−7.063
	200	−0.648	**−1.450**	**−4.679**	**−4.750**	**−4.570**	**−4.505**	**−3.008**	−7.058
LIN + MAT52 (ARD)	50	−0.645	−1.445	−4.640	−4.705	−4.538	−4.467	−2.990	−7.052
	100	−0.645	−1.445	−4.639	−4.703	−4.536	−4.465	−2.989	**−7.079**
	200	−0.645	−1.445	−4.639	−4.703	−4.536	−4.465	−2.989	**−7.079**

Table 3
Comparison between a regular and an augmented model, where the augmented model represents the model with additional weather variables near the NPP Krško. The best results and the overall best choice are emphasized. The measure used is the mean standardized log loss for the half-hour prediction averaged over 10-fold cross-validation.

	m	S_WX	S_WY	S_AT2	S_AT10	S_AT40	S_AT70	S_AH2	S_AP
LIN + MAT52	50	−0.645	−1.445	−4.640	−4.704	−4.537	−4.467	−2.989	−7.043
	100	−0.646	−1.447	−4.660	−4.727	−4.554	−4.486	−2.996	−7.063
	200	−0.648	**−1.450**	−4.679	−4.750	−4.570	−4.505	−3.008	−7.058
LIN + MAT52 augmented	50	−0.650	−1.447	−4.659	−4.729	−4.547	−4.470	−3.026	−6.955
	100	−0.651	−1.447	−4.699	−4.755	−4.577	−4.489	−3.035	**−7.090**
	200	**−0.652**	−1.449	**−4.728**	**−4.791**	**−4.601**	**−4.519**	**−3.045**	−7.072

Table 4
NRMSE, PCC, and MSLL metrics for the half-hour prediction of the weather variables of interest. The GP-NARX approach is compared to the NWP prediction and popular neural network architectures for modeling dynamic systems. Best results are shown in bold.

		S_WX	S_WY	S_AT2	S_AT10	S_AT40	S_AT70	S_AH2	S_AP
GP-NARX	NRMSE	**0.416**	**0.618**	**0.951**	**0.958**	**0.952**	**0.952**	**0.931**	**0.976**
	PCC	**0.812**	**0.924**	**0.999**	**0.999**	**0.999**	**0.999**	**0.998**	**0.999**
	MSLL	0.728	−1.489	−7.401	−7.438	−7.319	−7.257	−4.409	−7.224
NWP	NRMSE	−0.456	0.046	0.737	0.718	0.679	0.672	0.613	0.713
	PCC	0.314	0.687	0.969	0.966	0.960	0.958	0.929	0.993
	MSLL	–	–	–	–	–	–	–	–
LSTM	NRMSE	0.030	0.250	0.768	0.771	0.771	0.773	0.698	0.840
	PCC	0.380	0.673	0.973	0.974	0.974	0.975	0.955	0.987
	MSLL	–	–	–	–	–	–	–	–
DNN-NARX	NRMSE	0.402	0.613	0.943	0.949	0.943	0.943	0.920	0.960
	PCC	0.802	0.922	0.998	**0.999**	0.998	0.998	0.997	**0.999**
	MSLL	–	–	–	–	–	–	–	–
BNN-NARX	NRMSE	0.394	0.614	0.950	0.956	0.947	0.948	0.923	0.968
	PCC	0.796	0.923	**0.999**	**0.999**	**0.999**	**0.999**	0.997	**0.999**
	MSLL	**−0.875**	**−1.840**	**−7.988**	**−8.327**	**−7.753**	**−7.708**	**−5.531**	**−8.203**

a few-hour forecast. Secondly, the model was tested for the limit case of infinite predictive horizon, i.e., simulation. Lastly, the results of modeling the air pollution dispersion are presented for the considered case study.

4.1. Short-term weather variables forecast

Table 3 shows that the augmented model with the additional latent variables performs better than the model using only the weather variables of interest. It is not surprising that the forecasts of the weather variables are improved with the additional measurements spatially distributed around the modeled geographical location. Fig. 3 shows the scatter plot between the measurements and the half-hour prediction of the weather variables.

The largest variability can be seen in the wind component predictions which proved to be a significant modeling challenge. The Cartesian wind components show a slight bias and the modeled noise does not seem to be independent and identically distributed as assumed with modeling with GPs. This is not an unexpected behavior since the wind predictions are complex and heavily depend on the surrounding terrain as well as other predictors that may not be accessible. Other

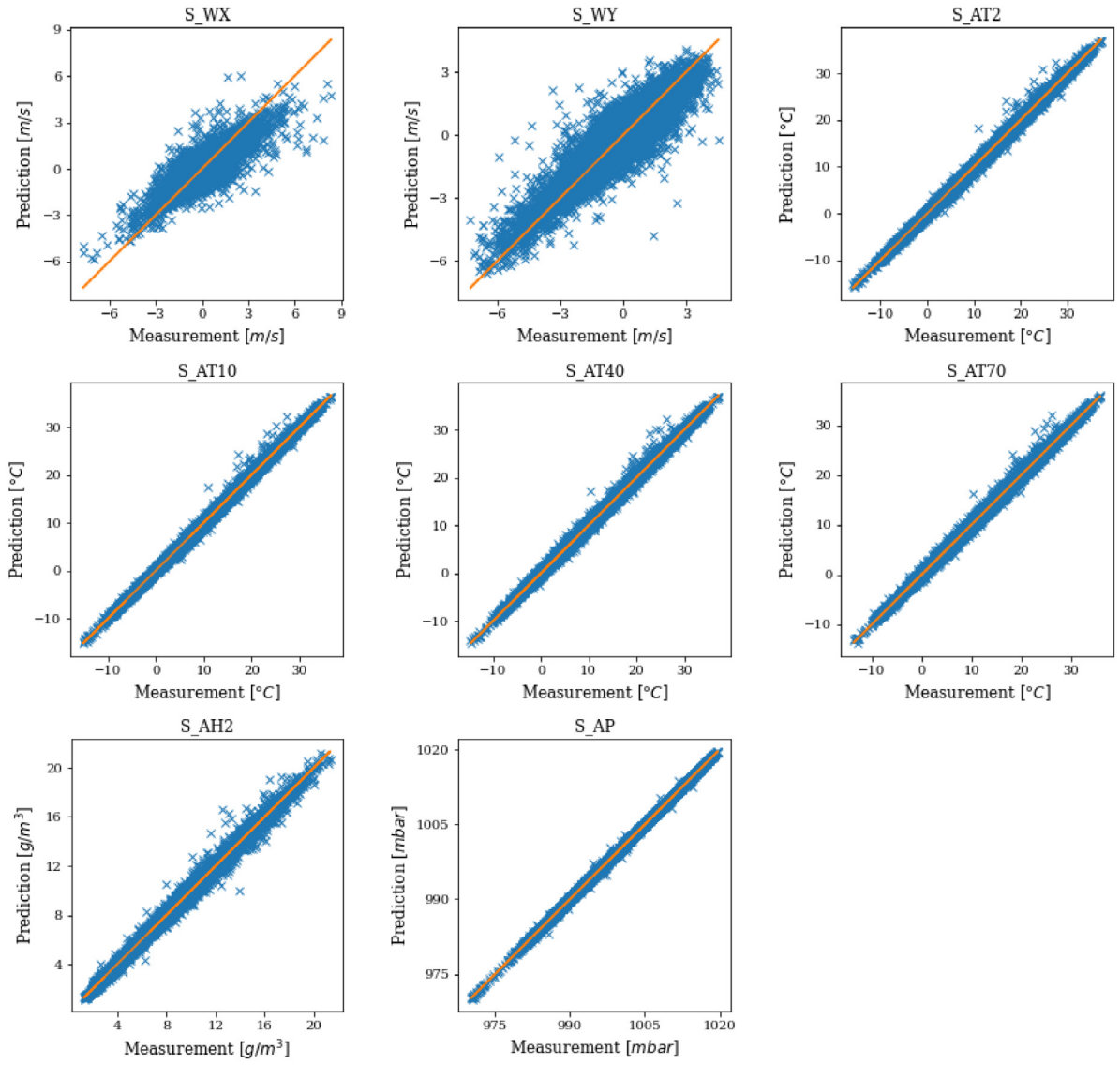


Fig. 3. Scatter plot showing the dependency between the measured weather variables of interest and their corresponding half-hour prediction presented in blue. The orange line shows the ground truth. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

weather variables on Fig. 3 show excellent performance results and the noise is modeled appropriately.

Table 4 presents the results of the validation on the test set for three popular performance metrics and a comparison to a popular approach with neural networks for modeling dynamic systems. The meta-parameters of the neural network approach were determined by random search over the training dataset. For the NARX models the lags were selected identically as with GPs. For the Bayesian neural network NARX (BNN-NARX) model (Neal, 1995) we considered the same architecture as with the deep neural network NARX (DNN-NARX) model (Menezes Jr and Barreto, 2008). The posterior distribution over the network parameters was obtained with mean field variational inference (Zhang et al., 2018). Recurrent long short-term memory (LSTM) architecture was also considered (Hochreiter and Schmidhuber, 1997).

We see that the GP-NARX performs best with respect to the normalized root-mean-square error (NRMSE) and the Pearson correlation coefficient (PCC), exhibiting high correlation between the predicted means and the measured variables. NRMSE is presented as a general performance metric, whereas the PCC is presented since it is widely used in atmospheric sciences. The downside of the aforementioned metrics is that they only take the predicted means into consideration. Therefore, MSLL is also included which is suitable for validation in the form of random variables. BNN-NARX approach outperforms GP-NARX with respect to MSLL, but on the cost of worse predictive means. Both models appropriately model the predictive uncertainty well. The LSTM approach with no lagged output variables underperforms when compared to the NARX models. We want to emphasize that both DNN-NARX and BNN-NARX are a similar class of models as GP-NARX model and we expect similar results given enough time to explore the high dimensional space of meta-parameters well. On the other hand, GP-NARX models have a small number of meta-parameters to determine, where hyperparameters are determined with maximizing the marginal log-likelihood. The advantage is that good solutions are obtained rather quickly without the computationally demanding search over the highly dimensional space of meta-parameters.

4.2. Long-term weather variables forecast

To advise the prevention actions in the case of a nuclear accident, we are interested in longer prediction intervals. Multiple step-ahead prediction for long-term predictions and simulations of the weather variables was obtained with Monte Carlo sampling, defined with Eqs. (20) and (21), with 500 independent samples of future trajectories. Considering careful matrix implementation, this simulation can

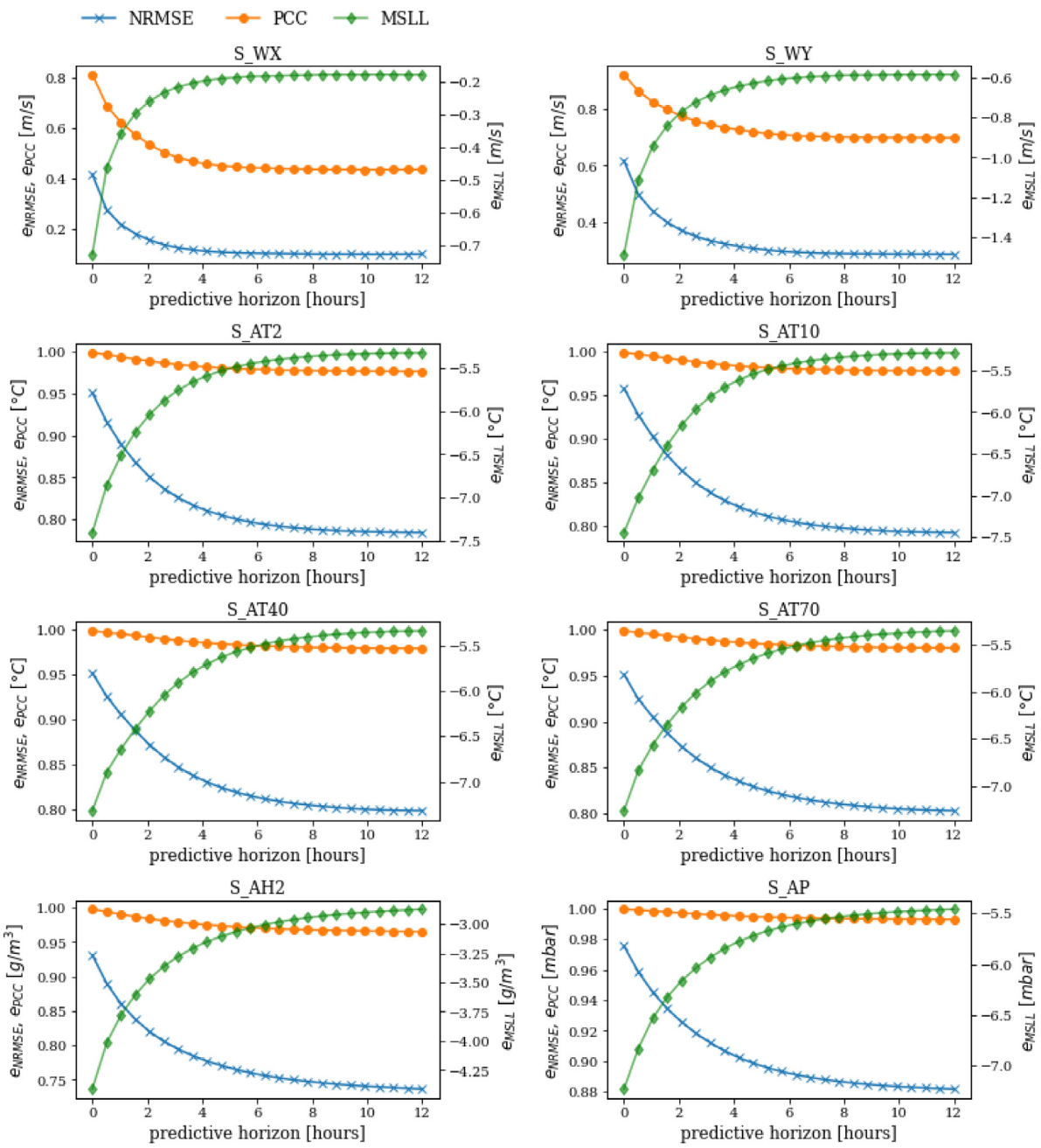


Fig. 4. Multiple step-ahead prediction performance metrics for the weather variables of interest in relationship to the number of predicted steps up to 12 h into the future.

be accelerated using general-purpose computing on graphics processing units (Krivec et al., 2020), since the test dataset is large.

The particles at an arbitrary time-step are approximated with a Gaussian distribution for the model validation. Fig. 4 shows the performance metrics in relationship with increasing the prediction horizon into the future. A step corresponds to a half-hour increase in time. The prediction error metrics for the forecasts up to 12 h into the future were considered, where the exogenous inputs in each step are defined by the most recent forecast from the NWP model. The model performance is worsened with the increased predictive horizon. Temperature, humidity, and air pressure are still relatively well modeled. Their respective NRMSE and PCC still exhibit high correlation and their corresponding MSLL is negative. In the case of the wind forecasts, the model performance degrades faster with increased predictive horizon. For most of the predicted steps into the future, the corresponding MSLL is still well below 0, which intuitively means that the model still performs better than simple models, i.e., when the predicted values are just the mean values of the observed history.

4.3. Simulation of the weather variables

The limit case of an infinite predictive horizon was also considered. In simulation the known parameters and measurements are used to initialize the model. In our case, the model was initialized with the start of the test data. Then the simulation is obtained with a multiple step-ahead prediction described in Section 4.2. Similarly, we obtain the simulation response for the BNN-NARX, whereas for DNN-NARX only the point estimate is propagated to the future. LSTM model prediction response can also be seen as a simulation response since the lagged outputs do not parametrize the model input space. It is also the only model that was specifically trained for longer forecasting intervals. We want to emphasize that the measured weather variables are unknown

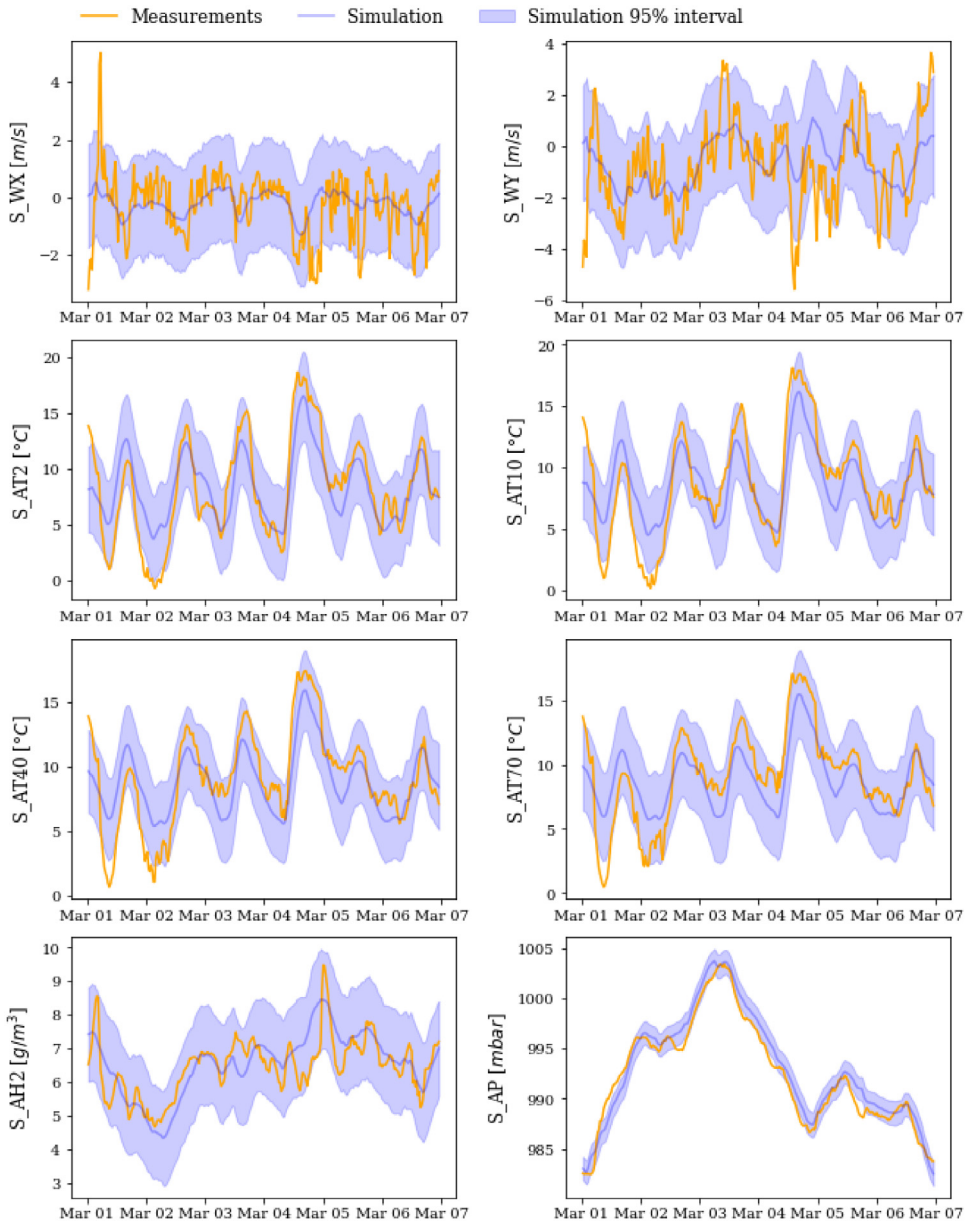


Fig. 5. Simulated response and the corresponding interval of 95% for the weather variables of interest near NPP Krško. The simulation was initialized on the first of January 2017 and the weather variables continuously simulated where the most recent NWP forecasts were used as exogenous inputs at each iteration. The results are shown for the period between the first and the seventh of March 2017.

in the future and are not used when validating the test dataset for multiple step-ahead prediction or simulation. The unknown inputs at an arbitrary time step into the future are replaced with the samples from the previous step. The exogenous inputs in each step are defined by the most recent forecasts from the NWP model, meaning that we assume that the NWP model short-term forecasts are known. In practice, this forecast would have to be replaced with a long-term NWP model forecast if available. The simulation presented in this section is therefore limited with the accuracy of such forecast.

Table 5 shows the validation metrics for the simulation on the test dataset. We observe that the simulation responses of the weather variables for a GP-NARX model are better modeled than the one step-ahead predictions from the NWP model (can also be seen as a simulation model). Temperature, humidity, and air pressure are described well even in the extreme case of simulation. Their respective NRMSE and PCC exhibit high correlation and their corresponding MSLL is negative. We expect similar results with DNN-NARX and BNN-NARX models. However, considering the results in Table 5, GP-NARX model performs better than the approach with neural networks. The GP-NARX model provided a more robust estimation of the simulation response compared to the neural networks and exhibited better stability for long prediction intervals.

As expected, the worst performance is obtained for wind simulations, although the MSLL is still below zero and the simulated wind components are better than the NWP one step-ahead prediction. We observe that the performance metrics on Fig. 4 start to converge to the values in Table 5. Fig. 5 shows the time response of the weather variables and the associated uncertainty for simulating the probabilistic hybrid model. As previously shown with the validation metrics, the temperature, humidity, and air pressure are described well with the model. Even the wind simulations capture the dynamics to some degree, although the variance is higher but well-calibrated. The predicted uncertainty interval of 95% captures most of the predicted variability of the measured signals.

Table 5
NRMSE, PCC, and MSLL metrics for the simulation of the weather variables of interest. The GP-NARX approach is compared to the NWP simulation and popular neural network architectures for modeling dynamic systems. Best results are shown in bold.

		S_WX	S_WY	S_AT2	S_AT10	S_AT40	S_AT70	S_AH2	S_AP
GP-NARX	NRMSE	**0.098**	**0.284**	**0.783**	**0.790**	**0.794**	**0.798**	**0.726**	**0.879**
	PCC	**0.434**	**0.698**	**0.976**	**0.978**	**0.979**	**0.979**	**0.962**	**0.993**
	MSLL	−0.179	−0.583	**−5.316**	**−5.269**	**−5.312**	**−5.316**	**−2.810**	**−5.425**
NWP	NRMSE	−0.456	0.046	0.737	0.718	0.679	0.672	0.613	0.713
	PCC	0.314	0.687	0.969	0.966	0.960	0.958	0.929	**0.993**
	MSLL	–	–	–	–	–	–	–	–
LSTM	NRMSE	0.030	0.250	0.768	0.771	0.771	0.773	0.698	0.840
	PCC	0.380	0.673	0.973	0.974	0.974	0.975	0.955	0.987
	MSLL	–	–	–	–	–	–	–	–
DNN-NARX	NRMSE	−0.023	0.177	0.610	0.601	0.588	0.596	0.642	0.816
	PCC	0.363	0.619	0.940	0.937	0.930	0.932	0.947	0.984
	MSLL	–	–	–	–	–	–	–	–
BNN-NARX	NRMSE	0.075	0.261	0.746	0.737	0.717	0.713	0.702	0.851
	PCC	0.395	0.676	0.973	0.973	0.971	0.971	0.957	0.989
	MSLL	**−0.290**	**−0.626**	−4.116	−3.827	−3.449	−3.427	−2.791	−5.077

4.4. Air pollution dispersion

The presented modeling investigation in Sections 4.1, 4.2, and 4.3 aims to provide the input data for the mass consistent wind field model and meteorological preprocessor. The output from the mass consistent wind field model serves as the input for the model of dispersion of radioactive pollution. A hypothetical pollution cloud is computed using the Lagrangian particle dispersion model. Program package Spray was used with the following inputs: three-dimensional wind field obtained from package MINERVE and other meteorological variables from package SURFPRO. The listed configuration was repeatedly validated to provide good results when used on complex terrain (Mlakar et al., 2015). Some data that characterize the used NWP model:

- Domain: 5 km × 5 km with resolution 50 m.
- Digital elevation model from year 2005 and Corine land cover from year 2000.
- Meteorological input data:
  - Temporal resolution 30 min, average values.
  - Ground meteorological station data: wind at 10 m, temperature, relative air humidity, air pressure, global solar radiation.
  - Meteorological station data: temperature and relative air humidity at 10 m, 40 m and 70 m.
  - SODAR measurements of wind vertical profile: 10 levels up to 500 m.
- Emission input data representing ground level exhaust:
  - Exhaust diameter 2.8 m.
  - Exhaust height 1 m (ground output).
  - The speed of gaseous emissions 0.1 m/s.
  - The temperature of gaseous emissions 25 °C.
  - Emission: normalized to unit.
- Output data: half-hour average relative concentrations in ground cells of dimensions 50 m × 50 m × 10 m. (length, width, height — 100 × 100 × 1 three-dimensional cells).

Fig. 6 shows an example, comparing the dispersion results of virtual pollutants for two and a half-hour period on the tenth of June 2017. Graphical comparison between the calculated atmospheric dispersion based on weather measurements at the site, half-hour prediction, simulation, and the forecast from the NWP model, are shown. Taking the measured inputs as the ground truth, we see that the results are favoring the proposed model for both the half-hour prediction and the simulation when compared to the dispersion using only the forecasts from the NWP model as the input.

5. Conclusions

We proposed a modeling approach that provides short-term (half-hour) and long-term (few-hour) forecasts of the atmospheric weather variables from the historic data at the selected location. The purpose of this model is to provide the input data for the model of the dispersion of air pollution from a point source.

The major findings are:

- The weather variables modeling task is solved using hybrid modeling, where NWP forecasts are enhanced with the statistical model trained from the available measured meteorological history from the Krško area.
- Although any function approximator could be used instead of a GP (e.g., deep neural networks), multi-input multi-output GP-NARX model using variational learning of hyper-parameters fits the application-specific requirements (i.e., safety critical application) well.
- Monte Carlo estimation of future trajectories provides well calibrated multiple step-ahead predictions and simulations.
- Half-hour prediction, few-hour forecasts, and simulation, that significantly improve the NWP forecasts, are possible with the hybrid model.
- The modeling and prediction task is solved using a single model and not a set of models each corresponding to a single weather variable. The advantage of a single multi-output model over a set of single-output models is in the model development and handling of large datasets.

The proposed hybrid-modeling strategy is aimed for forecasting the weather variables at an arbitrary selected geographical location, but not for a wider region. It is aimed at complex terrain. While the model is confined to a location, the method itself can be used for any general location having in mind its suitability for complex terrain. Even though, the case study dealt with the potential radioactive pollution, the proposed hybrid-modeling method is not constrained to only model the radioactive dispersion. It can be used for any kind of models where the weather variables serve as the input variables in a higher level modeling approach.

Besides the locality of the forecasts with the proposed method, a further limitation is a computational burden. It requires state-of-the-art computational facilities like a computer with a graphics processing unit for enhanced computational ability.

The investigated method will be suggested for the implementation of the safety-assurance system at NPP Krško. Moreover, other modeling methods for multi-input multi-output systems may be searched for and assessed for computational efficiency in the future with the same modeling aim as in the presented investigation.

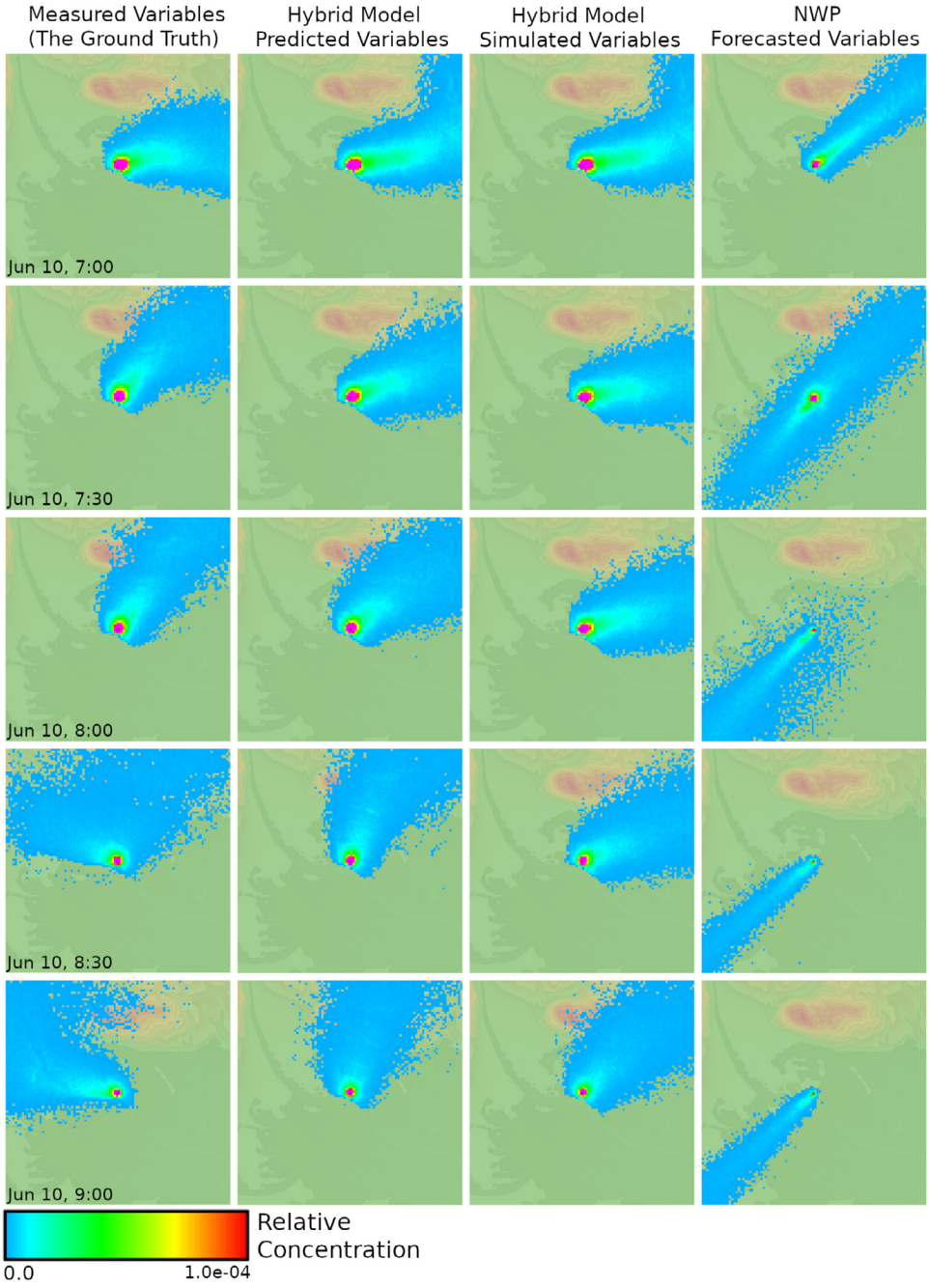


Fig. 6. Comparison of the results of modeling the air pollution dispersion for two and a half-hour period on the tenth of June 2017. The left most column represents the ground truth where the measured inputs are used to model the dispersion. Air pollution dispersion, modeled using the weather variables forecasted from the hybrid model, is presented for the half-hour prediction and simulation in the middle columns. Simulation was initialized on the first of January 2017 and the weather variables continuously simulated where the most recent NWP model forecasts were used as exogenous inputs at each iteration. The right most column represents the results of the modeled dispersion of the current approach using only the forecasts from a physical NWP model.

CRediT authorship contribution statement

Tadej Krivec: Writing – original draft, Methodology, Software, Validation, Formal analysis. **Juš Kocijan:** Methodology, Conceptualization, Supervision, Funding acquisition, Writing – review & editing. **Matija Perne:** Writing – review & editing, Data curation. **Boštjan Grašic:** Investigation, Data curation. **Marija Zlata Božnar:** Validation, Writing – review & editing. **Primož Mlakar:** Investigation, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge projects “Method for the forecasting of local radiological pollution of the atmosphere using Gaussian process models”, ID L2-8174, “Modelling the dynamics of short-term exposure to radiation”, ID L2-2615, research core funding No. P2-0001, and

Ph.D. grant for Tadej Krivec which were all financially supported by the Slovenian Research Agency. We are also grateful for the financial support of Krško Nuclear Power Plant, Slovenia, for the project L2-2615.

Appendix A. Performance metrics

The model was validated with the following performance metrics.

A.1. Normalized root-mean-square error

Normalized root-mean-square error (NRMSE) is defined by

$$\text{NRMSE} = 1 - \frac{\|\mathbf{y} - \boldsymbol{\mu}\|}{\|\mathbf{y} - \mathbb{E}(\mathbf{y})\|}, \tag{22}$$

where $\|\cdot\|$ represents the Euclidean norm, $\mathbf{y}$ the ground truth, and $\boldsymbol{\mu}$ the mean of the predictive distribution. The score for a perfect match would be 1 and $-\infty$ for a bad fit.

A.2. Pearson's correlation coefficient

Pearson's Correlation Coefficient (PCC) is a metric of the correlation between variables and is not sensitive to bias. It is defined by

$$\text{PCC} = \frac{(\boldsymbol{\mu} - \mathbb{E}(\boldsymbol{\mu}))(\mathbf{y} - \mathbb{E}(\mathbf{y}))}{N\sigma_y\sigma_\mu}, \tag{23}$$

where parameters σ_y, σ_μ denote the standard deviation of the ground truth vector and predicted vector respectively, N the number of data points, $\mathbf{y}$ the ground truth, and $\boldsymbol{\mu}$ the mean of the predictive distribution. The metric is ranging from -1 to 1, where 1 would be the best fit (perfect correlation).

A.3. Mean standardized log loss

The mean standardized log loss (MSLL) is a performance metric which is suitable for validation in the form of random variables, weighting the error by the predicted standard deviation. It is defined by

$$\text{MSLL} = \frac{1}{2N}\sum_{i=1}^{N}\left[\ln(\sigma_i^2) + \frac{(y_i - \mu_i)^2}{\sigma_i^2}\right] - \frac{1}{2N}\sum_{i=1}^{N}\left[\ln(\sigma_y^2) + \frac{(y_i - \mathbb{E}(\mathbf{y}))^2}{\sigma_y^2}\right], \tag{24}$$

where $\mathbf{y}$ represents the ground truth, σ_y^2 the variance of the ground truth, μ_i the predicted mean at time step i, σ_i^2 the predictive variance at time step i, and N the number of data points.

Appendix B. Covariance functions

The covariance functions used in our investigation are defined in this section. Vector $\mathbf{x}_i$, used throughout this section, is defined as the ith row of the input regressor $\mathbf{X}$. The static inputs $\mathbf{x}_i$ are without the loss of generality replaced with $\mathbf{z}_i$ defined with a NARX model.

B.1. Linear covariance function

A linear covariance function is defined by

$$k_{LIN}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \mathbf{x}_i \mathbf{x}_j, \tag{25}$$

where σ_f denotes a scaling factor.

B.2. Radial basis covariance function

A radial basis covariance (RBF) function is defined by

$$k_{RBF}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 e^{-\frac{1}{2}r^2}, \tag{26}$$

where $r = \frac{1}{l}\|\mathbf{x}_i - \mathbf{x}_j\|$ and l is represents a lengthscale parameter. Automatic Relevance Determination (ARD) covariance function weights the columns of the input $\mathbf{x}$ with their corresponding lengthscale l_d. It is defined by Eq. (26) and

$$r = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T \boldsymbol{\Lambda}^{-1}(\mathbf{x}_i - \mathbf{x}_j)}, \tag{27}$$

where $\boldsymbol{\Lambda}^{-1} = \text{diag}([l_1^{-2}, \ldots, l_d^{-2}])$ and d is the number of columns in $\mathbf{x}$.

B.3. Matérn Covariance Function

A Matérn covariance function is defined by

$$k_{MAT52}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2(1 + \sqrt{5}r + \frac{5}{3}r^2)e^{-\sqrt{5}r}, \tag{28}$$

where $r = \frac{1}{l}\|\mathbf{x}_i - \mathbf{x}_j\|$. Matérn covariance function with an ARD property is defined by Eqs. (27) and (28).

B.4. A combination of the covariance functions

A combination of covariance functions is also permitted. For example, the sum of a linear and a RBF function defines a valid covariance function

$$k_{LIN+RBF}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_{f'}^2 \mathbf{x}_i \mathbf{x}_j + \sigma_f^2 e^{-\frac{1}{2}r^2}. \tag{29}$$

where l represents the lengthscale parameter, $\sigma_{f'}$ and σ_f represent the respective scaling factors, and r is defined in Eq. (26).

References

Al-Shawwa, M.O., Al-Absi, A.A.-R., Hassanein, S.A., Baraka, K.A., Abu-Naser, S.S., 2018. Predicting temperature or humidity in the surrounding environment using artificial neural network.

Barbounis, T.G., Theocharis, J.B., Alexiadis, M.C., Dokopoulos, P.S., 2006. Long-term wind speed and power forecasting using local recurrent neural network models. IEEE Trans. Energy Convers. 21 (1), 273–284.

Bauer, M., van der Wilk, M., Rasmussen, C.E., 2016. Understanding probabilistic sparse Gaussian process approximations. In: Advances in Neural Information Processing Systems. pp. 1533–1541.

Božnar, M.Z., Grašič, B., de Oliveira, A.P., Soares, J., 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.

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

Cai, J., Ip, K.F., Eze, C., Zhao, J., Cai, J., Zhang, H., 2019. Dispersion of radionuclides released by nuclear accident and dose assessment in the Greater Bay Area of China. Ann. Nucl. Energy 132, 593–602.

Cai, H., Jia, X., Feng, J., Li, W., Hsu, Y.-M., Lee, J., 2020. Gaussian process regression for numerical wind speed prediction enhancement. Renew. Energy 146, 2112–2123.

Chen, N., Qian, Z., Nabney, I.T., Meng, X., 2013. Wind power forecasts using Gaussian processes and numerical weather prediction. IEEE Trans. Power Syst. 29 (2), 656–665.

Deisenroth, M., Rasmussen, C.E., 2011. PILCO: A model-based and data-efficient approach to policy search. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11). Citeseer, pp. 465–472.

Desiato, F., Finardi, S., Brusasca, G., Morselli, M., 1998. Simulation of 3D flow with diagnostic wind field models. Atmos. Environ. 32 (7), 1141–1156.

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. J. R. Stat. Soc. Ser. C. Appl. Stat. 47 (3), 299–350.

Duvenaud, D., 2014. Automatic Model Construction with Gaussian Processes (Ph.D. thesis). University of Cambridge.

Fan, G.-f., Wang, W.-s., Liu, C., DAI, H.-z., 2008. Wind power prediction based on artificial neural network. Proc. CSEE 28 (34), 118–123.

Geai, P., 1987. Methode d'Interpolation et de Reconstitution Tridimensionnelle d'un Champ de Vent: Filecode d'Analyse Objective MINERVE. Electricite de France Rep. ARD-AID: E34-E11.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, second ed. Chapman and Hall/CRC.

Gradišar, D., Grašič, B., Božnar, M.Z., Mlakar, P., Kocijan, J., 2016. Improving of local ozone forecasting by integrated models. Environ. Sci. Pollut. Res. 23 (18), 18439–18450.

Guestrin, C., Krause, A., Singh, A.P., 2005. Near-optimal sensor placements in Gaussian processes. In: Proceedings of the 22nd International Conference on Machine Learning. pp. 265–272.

Hall, T., Brooks, H.E., Doswell III, C.A., 1999. Precipitation forecasting using a neural network. Weather Forecast. 14 (3), 338–345.

Hayati, M., Mohebi, Z., 2007. Application of artificial neural networks for temperature forecasting. World Acad. Sci. Eng. Technol. 28 (2), 275–279.

Hewage, P., Behera, A., Trovati, M., Pereira, E., 2019. Long-short term memory for an effective short-term weather forecasting model using surface weather data. In: IFIP International Conference on Artificial Intelligence Applications and Innovations. Springer, pp. 382–390.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

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.

Kalnay, E., 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge university press.

Kariniotakis, G., Stavrakakis, G., Nogaret, E., 1996. Wind power forecasting using advanced neural networks models. IEEE Trans. Energy Convers. 11 (4), 762–767.

Khatir, S., Boutchicha, D., Le Thanh, C., Tran-Ngoc, H., Nguyen, T., Abdel-Wahab, M., 2020. Improved ANN technique combined with Jaya algorithm for crack identification in plates using XIGA and experimental analysis. Theor. Appl. Fract. Mech. 107, 102554.

Khotanzad, A., Davis, M.H., Abaye, A., Maratukulam, D.J., 1996. An artificial neural network hourly temperature forecaster with applications in load forecasting. IEEE Trans. Power Syst. 11 (2), 870–876.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Ko, J., Fox, D., 2009. GP-BayesFilters: Bayesian filtering using Gaussian process prediction and observation models. Auton. Robots 27 (1), 75–90.

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

Kocijan, J., Perne, M., Grašic, B., Božnar, M.Z., Mlakar, P., 2020. Sparse and hybrid modelling of relative humidity: the Krško basin case study. CAAI Trans. Intell. Technol. 5 (1), 42–48.

Kocijan, J., Perne, M., Mlakar, P., Grašič, B., Božnar, M.Z., 2019. Hybrid model of the near-ground temperature profile. Stoch. Environ. Res. Risk Assess. 33 (11–12).

Krivec, T., Papa, G., Kocijan, J., 2020. Simulation of variational Gaussian process NARX models with GPGPU. ISA Trans..

Kuligowski, R.J., Barros, A.P., 1998. Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks. Weather Forecast. 13 (4), 1194–1204.

Lee, S.-C., 2003. Prediction of concrete strength using artificial neural networks. Eng. Struct. 25 (7), 849–857.

Liu, H., Ong, Y.-S., Shen, X., Cai, J., 2020. When Gaussian process meets big data: A review of scalable GPs. IEEE Trans. Neural Netw. Learn. Syst. 31 (11), 4405–4423.

Matthews, A.G.d.G., Rowland, M., Hron, J., Turner, R.E., Ghahramani, Z., 2018. Gaussian process behaviour in wide deep neural networks. arXiv preprint arXiv:1804.11271.

Menezes Jr, J.M.P., Barreto, G.A., 2008. Long-term time series prediction with the NARX network: An empirical evaluation. Neurocomputing 71 (16–18), 3335–3343.

Mlakar, P., Božnar, M.Z., Grašič, B., 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.

Mlakar, P., Božnar, M.Z., Grašič, B., Brusasca, G., Tinarelli, G., Morselli, M.G., Finardi, S., 2015. Air pollution dispersion models validation dataset from complex terrain in Šoštanj. Int. J. Environ. Pollut. 57 (3–4), 227–237.

More, A., Deo, M., 2003. Forecasting wind with neural networks. Mar. Struct. 16 (1), 35–49.

Neal, R., 1995. Bayesian Learning for Neural Networks [Ph.D. thesis]. Department of Computer Science, University of Toronto, Toronto, Ontario, Canada.

Nguyen-Le, D.H., Tao, Q., Nguyen, V.-H., Abdel-Wahab, M., Nguyen-Xuan, H., 2020. A data-driven approach based on long short-term memory and hidden Markov model for crack propagation prediction. Eng. Fract. Mech. 235, 107085.

Pan, B., Hsu, K., AghaKouchak, A., Sorooshian, S., 2019. Improving precipitation estimation using convolutional neural network. Water Resour. Res. 55 (3), 2301–2321.

Quiñonero-Candela, J., Rasmussen, C.E., 2005. A unifying view of sparse approximate Gaussian process regression. J. Mach. Learn. Res. 6 (Dec), 1939–1959.

Rasmussen, C.E., Williams, C.K., 2006. Gaussian Processes for Machine Learning. MIT press Cambridge.

Ren, W.W., Yang, T., Huang, C.S., Xu, C.Y., Shao, Q.X., 2018. Improving monthly streamflow prediction in alpine regions: integrating HBV model with Bayesian neural network. Stoch. Environ. Res. Risk Assess. 32 (12), 3381–3396.

Saeed, I.M.M., Saleh, M.A.M., Hashim, S., Hama, Y.M.S., Hamza, K., Al-Shatri, S.H., 2020. The radiological assessment, hazard evaluation, and spatial distribution for a hypothetical nuclear power plant accident at Baiji potential site. Environ. Sci. Eur. 32 (1), 6.

Scher, S., 2018. Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. Geophys. Res. Lett. 45 (22), 12–616.

Silibello, C., Finardi, S., Calori, G., 2006. SURFPRO (SURrface-Atmosphere InterFace PROcessor) - User's Guide. ARIANET report 2006.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, M., Duda, K.G., Huang, X.Y., Wang, W., Powers, J.G., 2008. A Description of the Advanced Research WRF Version 3. Tech. rep., National Center for Atmospheric Research, pp. 1–113.

Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical Bayesian optimization of machine learning algorithms. arXiv preprint arXiv:1206.2944.

Souissi, R., Tawfik, F., Ramadan, A., Reguiguia, N., 2019. A comparative study for the atmospheric dispersion characteristics of two proposed sites for nuclear power plant in Tunisia. Res. Rev.: J. Ecol. 8 (2), 1–13.

Tinarelli, G., 2007. SPRAY 3.1–General Description and User's guide. ARIANET Report 2007.

Tinarelli, G., Anfossi, D., Castelli, S.T., Bider, M., Ferrero, E., 2000. A new high performance version of the Lagrangian particle dispersion model SPRAY, some case studies. In: Air Pollution Modeling and Its Application XIII. Springer, pp. 499–507.

Titsias, M.K., 2009a. Variational Model Selection for Sparse Gaussian Process regression. Report, University of Manchester, UK.

Titsias, M., 2009b. Variational learning of inducing variables in sparse Gaussian processes. In: Artificial Intelligence and Statistics. PMLR, pp. 567–574.

Von Stosch, M., Oliveira, R., Peres, J., de Azevedo, S.F., 2014. Hybrid semi-parametric modeling in process systems engineering: Past, present and future. Comput. Chem. Eng. 60, 86–101.

Wilson, A.G., Hu, Z., Salakhutdinov, R., Xing, E.P., 2016. Deep kernel learning. In: Artificial Intelligence and Statistics. PMLR, pp. 370–378.

Worsnop, R.P., Scheuerer, M., Hamill, T.M., Lundquist, J.K., 2018. Generating wind power scenarios for probabilistic ramp event prediction using multivariate statistical post-processing. Wind Energy Sci. 3 (1), 371–393.

Zenzen, R., Khatir, S., Belaidi, I., Le Thanh, C., Wahab, M.A., 2020. A modified transmissibility indicator and Artificial Neural Network for damage identification and quantification in laminated composite structures. Compos. Struct. 248, 112497.

Zhang, C., Bütepage, J., Kjellström, H., Mandt, S., 2018. Advances in variational inference. IEEE Trans. Pattern Anal. Mach. Intell. 41 (8), 2008–2026.

Zhang, C., Wei, H., Zhao, X., Liu, T., Zhang, K., 2016. A Gaussian process regression based hybrid approach for short-term wind speed prediction. Energy Convers. Manage. 126, 1084–1092.