Improving wind vector predictions for modelling of atmospheric dispersion during Seveso-type accidents

Matija Perne[a,*], Marija Zlata Božnar[b], Boštjan Grašič[b], Primož Mlakar[b], Juš Kocijan[a,c]

[a] *Jožef Stefan Institute, Ljubljana, Slovenia*
[b] *MEIS d.o.o., Mali Vrh, Slovenia*
[c] *University of Nova Gorica, Nova Gorica, Slovenia*

Abstract

In case of a major accident involving airborne emissions of harmful gases, a temporary portable meteorological station may be used to improve atmospheric dispersion modelling (ADM) for protection of people and the environment. While the meteorological station provides signal values in real time, ADM results for the future are of particular interest for planning purposes. It is possible to use the current measured value as a future input to the ADM but it is suboptimal. It is also possible to use model output statistics (MOS) to predict the future local weather information from numerical weather prediction (NWP) models, while available operational NWP models are in general too coarse to be used directly in fine-resolution ADM. MOS models are obtained through machine learning and the training data sets in most traditional uses of MOS are big, which is beneficial for modelling. We envision using MOS in an emergency and for a location of a temporary meteorological station. We use windowing for online data selection to explore its accuracy when the amount of available training data is very limited, which is expected in an emergency situation. We show that MOS for wind vector with 1 day of training data greatly improves on the numerical weather predictions and the persistence model, so its use in such an emergency would be advantageous.

Keywords: Gaussian process, atmospheric dispersion model, industrial accident, system identification, online modelling

*Corresponding author. Department of Systems and Control, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia. E-mail address: matija.perne@ijs.si.

1. Introduction

The European Commission reports that "major accidents involving dangerous chemicals pose a significant threat to humans and the environment" (The European Commission, 2020). In the European Union, such hazards are prevented and controlled through Seveso Directive (The European Commission, 2020), which aims both at "the prevention of major accidents involving dangerous substances" (The European Commission, 2019) and at "limiting the consequences of such accidents not only for human health but also for the environment" (The European Commission, 2019). "The Directive applies to more than 12 000 industrial establishments in the European Union," according to (The European Commission, 2020).

An important component of limiting the consequences of a major accident with dangerous chemicals is atmospheric dispersion modelling. It is thus included in the tool named Accident Damage Assessment Module (ADAM), developed by the Joint Research Centre of the European Commission for assessing the consequences of an industrial accident resulting from an unintended release of a dangerous substance (Fabbri and Wood, 2019). The purpose of ADAM is implementation of the Seveso Directive (Fabbri and Wood, 2019).

Atmospheric dispersion modelling benefits from good quality information on local wind speed and direction (Barratt, 2013; Beelen et al., 2010; Breznik et al., 2003). In certain accidents, installing a portable meteorological station at the accident site is useful for protection of the human health and the environment (O'Mahony et al., 2008), especially if emission into the atmosphere is going on for a longer time. A couple such long duration scenarios, which are covered by ADAM, are pool evaporation and fire (Fabbri and Wood, 2019).

However, the local measurements provided by a portable meteorological station apply to the past. Information on expected winds in the future would enable atmospheric dispersion modelling for the future, providing helpful advice for protecting the human health and the environment.

Numerical weather prediction (NWP) models are predicting future winds and other meteorological parameters in ever increasing spatial and temporal resolution. They can be used for atmospheric dispersion modelling (Grašič et al., 2018; Jones et al., 2007; Sigg et al., 2018). However, predicting local

wind speed and direction in detail is still challenging, especially in complex terrain. Local measurements are thus a more relevant source of data on local winds than NWP model outputs are.

We propose a method for obtaining local wind signal values at the site of a portable meteorological station for the future. By *signal values*, we mean values of a time-dependent physical quantity that carry information about the weather, either measured or obtained as an output of a model. We use the measurements of the portable meteorological stations as the training data set to *train* the model, that is, to optimize its parameters. The model predicts the local horizontal wind vector at the accident site from the NWP model predictions, the current wind measurements from the portable station, and possibly from the measurements of the other meteorological stations in the surrounding area. We only address the scenario in which the portable meteorological station, at the site of which the wind signal values are predicted, is operating from some time after the beginning of the accident on. If the station was already operating before the accident, plenty of data for experimental modelling would have been available and such examples have already been studied (Mori and Kurata, 2008).

The main focus of this work is on evaluating the quality of the proposed models' predictions and determining whether their quality is sufficient for the intended use case. In particular, the predictions are compared to the other available predictions, that is, to the predictions of the NWP model and to the persistence predictions. The wind components at the location of the temporary portable meteorological station are predicted with system identification. Only the data that would be available in an accident is used. Notably, the amount of historical data at the site of the temporary station is severely limited to hours or at most days.

For experimental modelling, we use Gaussian process (GP) models (Kocijan, 2016; Rasmussen and Williams, 2006; Shi and Choi, 2011), mainly because they predict the expected variance of the model output, require few training data points, and require optimization of few parameters. We compare them to linear regression using least squares, which shares the advantages of GP such as predicting the variance. Many mathematical structures that are in use as statistical models, such as GP models (Rasmussen and Williams, 2006), artificial neural networks (Torrontegui and García-Ripoll, 2019), fuzzy models (Pal et al., 2018), Volterra models (Carini et al., 2018), etc. are universal approximators and would deliver similar results.

Experimental modelling of local weather parameters from NWP model

outputs is well established and is called model output statistics (MOS) particularly in the case the experimental model is linear (Bédard et al., 2013; Glahn and Lowry, 1972; Kalnay, 2003). GP has been used for the purpose before, often for wind power forecasting (Chen et al., 2013a,b; Hoolohan et al., 2018; Yan et al., 2016; Zhang et al., 2016, 2019) but also for improving atmospheric dispersion modelling (Kocijan et al., 2019; Perne et al., 2019).

The main original contribution of this work is experimental modelling of meteorological parameters with small training data sets. For the established uses of experimental modelling of local weather variables, big training data sets of months to years tend to be available, increasing the accuracy of the model. In contrast, the available training data set for a temporary meteorological station during an accident with dangerous chemicals is days at most. We show that it is possible to improve on both the numerical weather predictions and the persistence model under this constraint.

To efficiently test the models and evaluate the influence of the training data set size on their performance, we use online modelling method, namely windowing or time-stamp method (Kneale and Brown, 2018). That is, when optimising a model, we use a pre-determined number of most recent data points in system identification.

2. Methods

The scenario we are simulating is an accident with long-term emission of pollutants into the air, for example due to pool evaporation or fire. The accident causes the need for atmospheric dispersion modelling for the future. A temporary meteorological station is established at a relevant site after the beginning of the accident. The study area has a high-resolution numerical weather prediction (NWP) model and several weather stations. Past and current values of the weather parameters are available from both the stations and the NWP model as signals. We are interested in obtaining short-term predictions of the local wind at the location of the temporary station and we do it with an experimental model. Experimental model is a mathematical model that uses training data to learn the relationships between the variables of the system and to predict the output signal. The output mathematically depends on delayed values of the input signals which are called regressors and are assembled in the regressor vector for any given moment. At any time, the training data set of the model is limited to the data collected since the installation of the temporary meteorological station – observational data

at the site is not available for the time before the station was installed. We study the effect of the extent of the available training data set on the accuracy of multi-step predictions.

2.1. Study area and signals

The study is performed in a complex area of a basin and hills, shown in Fig. 1. There are 6 available meteorological stations located up to 30 km apart, listed in Table S1 in the Supporting Material (SM), that are measuring 29 signals in total. All the measurements are taken at ground level – that is, at the standard height of 10 m above ground in the case of wind. We use one of these stations, Stolp at Krško NPP (nuclear power plant), to represent the temporary one, while the others represent the permanent stations. A NWP model for the area is available, providing 7 signals, and 1 additional signal is derived from NWP signals using an artificial neural network (Soares et al., 2004). The NWP wind signal is computed at 10 m level as well. All the signals are sampled every 30 minutes. Snapshot values are used as the NWP signals and 30-minute averages as the measurement signals. We are predicting 30 minute averages as well. Data for the year 2017 is used.

The NWP model is based on WRF-ARW version 3.4.1 (Skamarock et al., 2008) as described in Grašič et al. (2018). The initial and lateral boundary conditions are obtained from GFS. The model is centred on latitude 46° N, longitude 14.5° E, Lambert conformal projection is used. It is composed of two domains, the bigger one has 101×101 grid cells with 12 km horizontal resolution, the smaller one has 82×76 grid cells and 4 km horizontal resolution. 45 vertical layers in η scheme are used. Lin et al. microphysics, Rapid Radiative Transfer Model radiation, Noah Land Surface Model land surface, Yonsei University planetary boundary layer, and Kain-Fritsch cumulus (in the bigger domain) schemes are used.

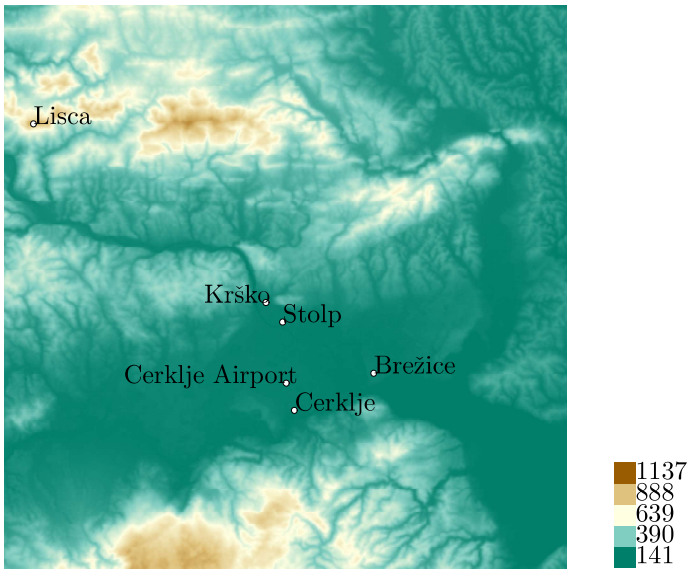
2.2. Mathematical tools

As the experimental models we use autoregressive models with exogenous input (ARX), sometimes nonlinear (NARX), and describe the mathematical form of the modelled relationships with GP. GP is a stochastic process containing random variables $f(\mathbf{z})$ with a normal probability distribution (Kocijan, 2016),

$$p\left(f\left(\mathbf{z}_1\right), \ldots, f\left(\mathbf{z}_N\right) | \mathbf{z}_1, \ldots, \mathbf{z}_N\right) = \mathcal{N}\left(\mathbf{m}, \mathbf{\Sigma}\right). \tag{1}$$

The vectors $\mathbf{z}_i$ are regressor vectors, f is the GP, $\mathbf{m}$ is the mean vector and $\mathbf{\Sigma}$ is the covariance matrix of the Gaussian distribution $\mathcal{N}$. In GP modelling,

(a) A panoramic view of the study area. Photo by Samo Grašič.



(b) Digital elevation model (©European Union, 2019) of the study area (40 km by 40 km) with the meteorological stations marked.

Figure 1: Overview of the study area.

we describe the GP with a mean function and a covariance function,

$$\mathbf{m}_i = m\left(\mathbf{z}_i\right), \qquad \mathbf{\Sigma}_{ij} = C\left(\mathbf{z}_i, \mathbf{z}_j\right), \tag{2}$$

where $m\left(\mathbf{z}_i\right)$ is the mean function and $C\left(\mathbf{z}_i, \mathbf{z}_j\right)$ is the covariance function.

Functioning of a GP-NARX is described by the equation (Kocijan, 2016; Nelles, 2001)

$$\hat{y}\left(t\right) = f\left(y\left(t-1\right), y\left(t-2\right) \ldots, y\left(t-n\right), \mathbf{u}\left(t\right), \mathbf{u}\left(t-1\right), \ldots, \mathbf{u}\left(t-m\right)\right) + \nu, \tag{3}$$

where y is the output value, $\mathbf{u}$ is the input value, $\hat{y}$ is the output prediction, t is the time index, n is the maximum lag in the output values, m is the maximum lag in the input values, and ν is Gaussian noise. The equation is illustrated in Fig. 3.

The covariance functions used depend on parameters called *hyperparameters*. For example, in the case of linear covariance function $C\left(\mathbf{z}_i, \mathbf{z}_j\right) = \mathbf{z}_i^T \Lambda^{-1} \mathbf{z}_j$ where Λ is a diagonal matrix with diagonal elements $\lambda_1^2, \ldots, \lambda_D^2$, we use $\log\left(\lambda_1\right), \ldots, \log\left(\lambda_D\right)$ as hyperparameters. We choose hyperparameters through likelihood optimization with the toolbox of Rasmussen and Nickisch (2010).

The regressors to be used are selected by identifying the significant terms for a linear-in-the-parameters model (Li and Peng, 2007) as implemented in ProOpter IVS (Gradišar et al., 2015).

2.3. Modelling choices

We are interested in the quality of predictions as a function of two design parameters. One is the number of the training data points, which reflects the time since the installation of the temporary meteorological station. The other one is the number of time steps in advance from the time of prediction, reflecting how much in advance the prediction is available. To analyse the influence of these factors, we generate 672 multi-step predictions for the first 14 days of January 2017.

We use windowing for training data selection with windows of 12, 24, 48, 96, 144, 240, 336 time steps for the training data, imitating 7 different amounts of the training data. The model is trained on the active dataset of the chosen number of most recent data points (Kocijan, 2016) available at the start of prediction. The start of prediction keeps moving throughout the test period by one time step. At each start of prediction, a new model is trained on the new active set. The new active set differs from the preceding

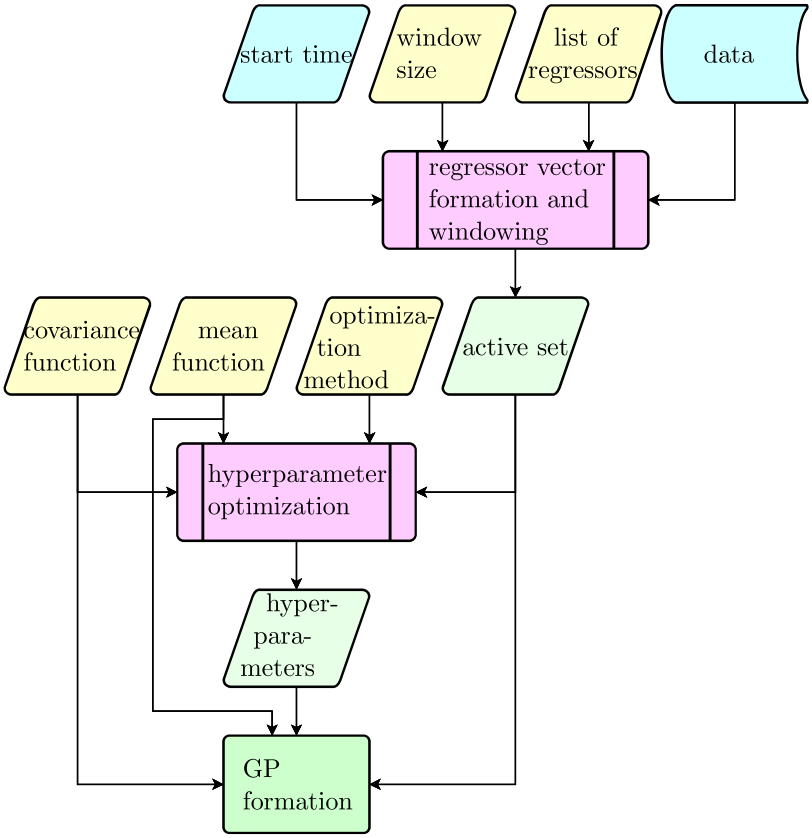


Figure 2: Constructing the Gaussian process model. The yellow boxes represent the choices regarding the modelling, and the green ones are the results. The modelled data are provided and the start time sweeps through the test period. In a real event, the start time would be the real time and the training data window size would match the time since the start of the data collection.

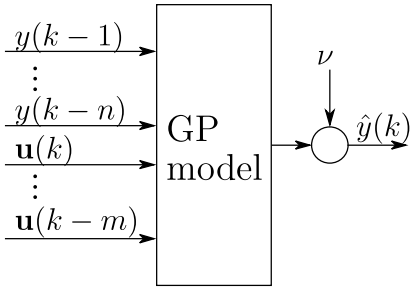


Figure 3: Formula of GP-NARX (Eq. 3) schematically after Kocijan (2016).

one in only one data point, so the new model is similar to the previous one. We thus use the previous model as the initial approximation in training the new model, achieving important savings in computation time.

We study predictions for 1 to 5 time steps, that is, for 30 to 150 minutes, in advance.

We use linear function as the covariance function in GP modelling. Other stationary (Kocijan, 2016; Rasmussen and Nickisch, 2010), e.g., squared exponential, and nonstationary (Kocijan, 2016; Rasmussen and Nickisch, 2010), e.g., sum of squared exponential and linear, covariance functions were tested, but have not provided improved results. Constant mean function and exact inference with Gaussian likelihood are used when constructing models. For multi-step predictions of ARX models, we use naive method, that is, we feed the predicted mean value to the model as the regressor.

Hyperparameter optimization is limited to 5000 steps. When the training data window is moved for one time step, hyperparameter optimization uses the previous hyperparameter values as the initial values.

We explore two different approaches to experimental modelling:

Type 1 models. We use the surrounding weather station measurements in addition to the NWP signals as inputs to the model. We use them with a delay of at least 1 time step. The first prediction step could thus be calculated in real time. For further time steps, we approximate the missing measurements with NWP-predicted values that would also be available in real time.

Type 2 models. We do not use the surrounding weather station measurements as inputs to the model, only the NWP signals. The model has all the necessary inputs available to do multi-step predictions without any modification.

We hypothesise that the models of the type 1 will have better available information on the initial weather situation and will perform better in the initial time steps, while the suboptimal use of NWP signals in place of measurements will harm it when making predictions for later times. In contrast, the models of the type 2 will optimally use the available information that is relevant for the more distant time steps and thus perform better for them, while it will lack the information on the initial conditions compared to the model that uses measurements and thus not perform as good in the first time steps.

The regressors are selected offline on downsampled data for the years 2012 through 2017. 15 best-ranking regressors are used in each model. For the models of type 1, the regressor candidates are delays of 1 and 2 time steps of the measured signals, and delays of 2, 1 and 0 time steps of NWP signals, adding up to 70 candidates in total. For the models of type 2, the NWP signals are delayed for 4, 3, 2, 1, 0 time steps, resulting in 44 candidates. The number of data points used is 4440 for the type 1 models and 4546 in the type 2 models. The selected regressors are listed in Tables S2 and S3 (see the SM).

Fig. 2 summarizes how the GP models are generated. We do a completely separate computation for each combination of:

- the window size from one of the 7 options of 12 to 336 time steps,
- the Cartesian wind component we are predicting,
- the type of model, 1 or 2, determining the regressor list, and
- the covariance function,

while the mean function and the optimization method are constant throughout the numerical experiments. At the start time, we construct the active set the size of the chosen window from the most recent data. We determine the hyperparameters from the training data using optimization. With the training data and the hyperparameters, the GP is fully defined.

Experimental modelling tools are designed for predicting scalar quantities. As horizontal wind velocity is a 2D vector, we model it by component. The standard polar representation of wind velocity as speed and direction has a discontinuity in direction that is challenging to model, which we avoid by modelling the Cartesian components.

2.4. Evaluation

In addition to qualitative estimates of the match between the measurement and the prediction, we use normalized root-mean-square error (NRMSE), mean standardised log-loss (MSLL), and Pearson correlation coefficient (PCC) as figures of merit, they are defined in Appendix A. For NRMSE and PCC, higher value is better, while lower MSLL value is better. For each number of training data points and each type of the model, they are calculated for each number of steps ahead of the prediction from all the multi-step predictions.

A useful benchmark is persistence model: using the current measured value as a prediction for a certain number of steps in advance. It is equivalent to the assumption that the wind is not going to change in the modelled time period. The goal of our modelling is to do better than both persistence and NWP.

3. Results

Examples of predictions of the models are presented in Figs. 4 and 5. Fig. 4 shows 5 steps ahead predictions of a model trained on 48 data points, that is, 1 day of training data. Fig. 5 shows one step ahead predictions of a model trained on 48 data points. The first graph is for a NWP-only type 2 model, the second one is for a type 1 using the measurements as inputs.

The predictions of wind speed and direction obtained from the 5 steps ahead predictions of the NWP-only type 2 model in Fig. 4 and its S-N equivalent are shown in Figs. S1 and S2 (see the SM). The direction is obtained from the ratio of the components' expected values and the estimate for the speed used is $v = \sqrt{v_x^2 + v_y^2 + \sigma_x^2 + \sigma_y^2}$, where $v_{x,y}$ are the predicted component expected values and $\sigma_{x,y}^2$ are the predicted component variances.

The NRMSE value as a function of how many time steps in advance the prediction is made are shown in Fig. 6a for type 1 models and in Fig. 6b for type 2 models.

These and other numbers are also tabulated. For type 1 models, Table 1 shows the dependence of MSLL values for both components on the number of training data points and the number of time steps in advance for which the prediction is made. Table 2 shows the same for PCC. As these figures of merit operate on scalar values, they are computed for each modelled component separately. Table 3 gives NRMSE for horizontal wind velocity treated as a vector. We notice a strong contrast in model quality between the window sizes of 24 and 48 so we mark it with a dashed line for easier readability.

For type 2 models, the same results are listed in Tables S4 to S6 (see the SM).

The available figures of merit for the persistence and NWP models are collected in Table 4.

We have settled on GP models with linear covariance function because this choice leads to best results. Some results using different choices are given in SM for comparison. Results of GP models with squared exponential covariance function are provided in Tables S7 to S9. Results of GP models

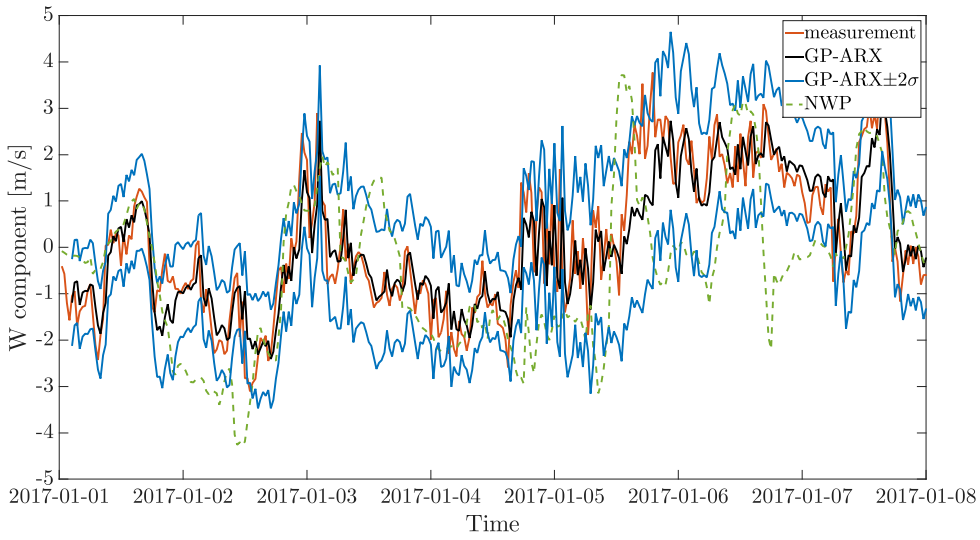


Figure 4: 5 step ahead predictions of type 2 models not using measurements as inputs with 48 training data points for W-E wind component.

with sum of squared exponential and linear covariance function are provided in Tables S10 to S12. Results of linear models based on least squares are provided in Tables S13 to S15.

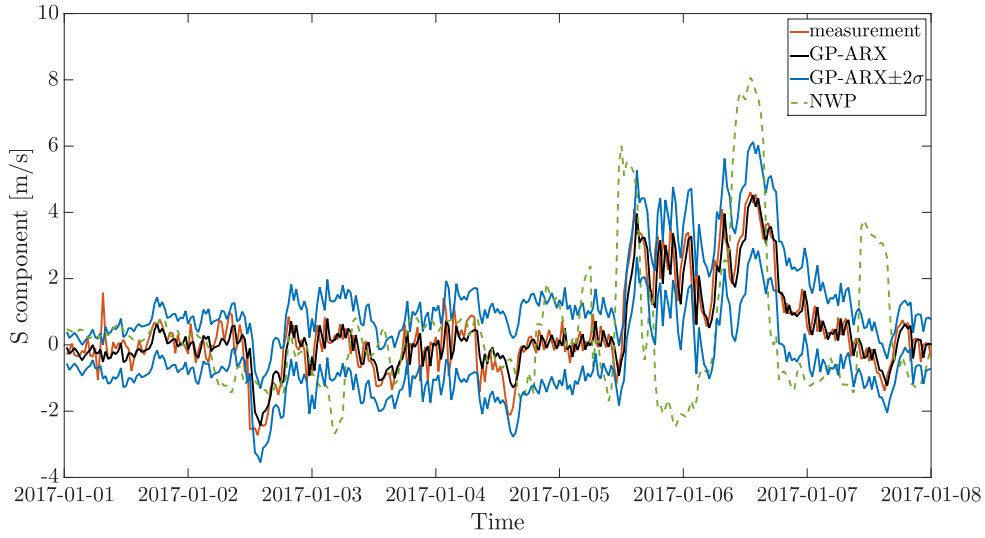
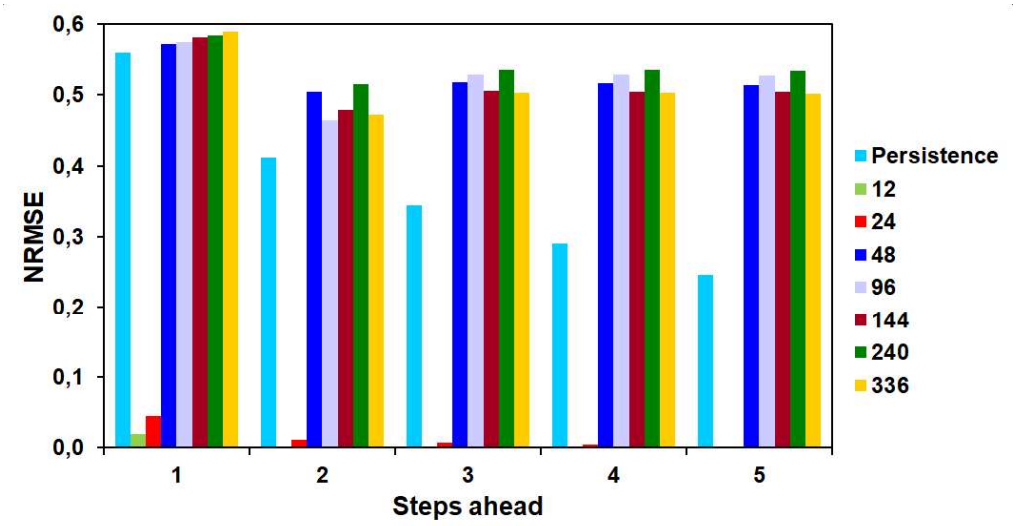
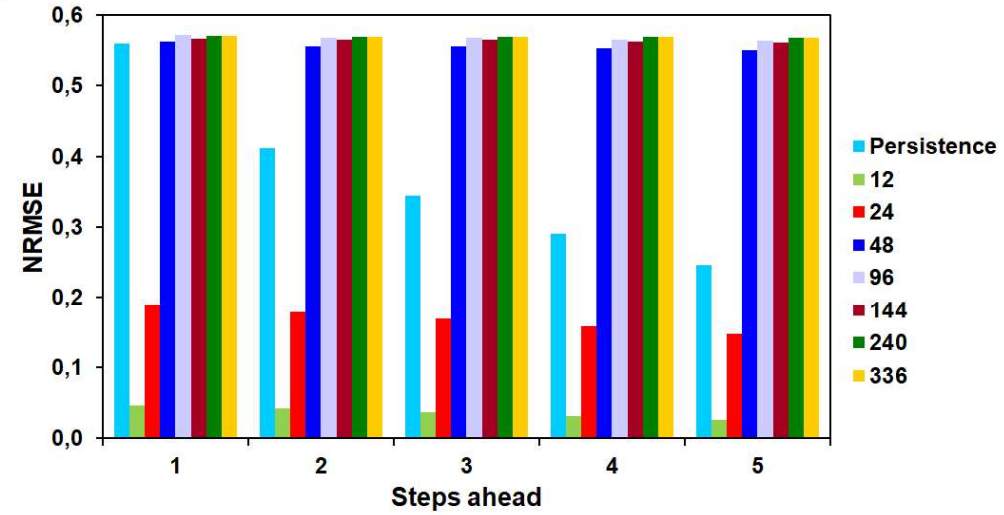


Figure 5: 1 step ahead predictions of type 1 models using measurements as inputs with 48 training data points for S-N wind component.



(a)



(b)

Figure 6: Graphical representation of the dependence of NRMSE values for wind vector predictions on the number of training data points and the number of time steps ahead for (a) type 1 models using measurements as inputs (data from Table 3) and (b) type 2 models not using measurements as inputs (data from Table S6).

Table 1: Dependence of MSLL values for each component on the number of training data points and the number of time steps ahead for type 1 models using measurements as inputs.

Component	Window size	Prediction for time step 1	2	3	4	5
W-E	12	-0.113	0.507	0.604	0.667	0.643
	24	-0.202	0.473	0.604	0.681	0.739
	48	-0.941	-0.644	-0.674	-0.670	-0.663
	96	-0.983	-0.521	-0.722	-0.720	-0.714
	144	-0.989	-0.576	-0.646	-0.643	-0.648
	240	-0.960	-0.671	-0.783	-0.782	-0.780
	336	-0.971	-0.452	-0.627	-0.624	-0.622
S-N	12	-0.163	0.179	0.333	0.383	0.418
	24	-0.130	0.065	0.183	0.254	0.311
	48	-0.673	-0.518	-0.610	-0.594	-0.574
	96	-0.646	-0.432	-0.577	-0.568	-0.559
	144	-0.679	-0.407	-0.544	-0.525	-0.513
	240	-0.681	-0.625	-0.615	-0.611	-0.609
	336	-0.688	-0.556	-0.550	-0.546	-0.542

Table 2: Dependence of PCC values for each component on the number of training data points and the number of time steps ahead for type 1 models using measurements as inputs.

Component	Window size	Prediction for time step 1	2	3	4	5
W-E	12	0.258	0.129	0.092	0.088	0.089
	24	0.344	0.283	0.227	0.197	0.177
	48	0.922	0.935	0.884	0.843	0.814
	96	0.927	0.897	0.896	0.855	0.826
	144	0.929	0.929	0.884	0.841	0.810
	240	0.928	0.941	0.895	0.849	0.815
	336	0.931	0.898	0.869	0.830	0.799
S-N	12	-0.001	-0.003	0.026	0.020	0.009
	24	0.240	0.296	0.243	0.214	0.206
	48	0.872	0.968	0.879	0.775	0.727
	96	0.865	0.952	0.867	0.753	0.703
	144	0.872	0.923	0.855	0.743	0.693
	240	0.876	0.986	0.866	0.755	0.709
	336	0.879	0.961	0.845	0.738	0.692

Table 3: Dependence of NRMSE values (higher value corresponds to a better model) for wind vector predictions on the number of training data points and the number of time steps ahead for type 1 models using measurements as inputs.

Window size	Prediction for time step 1	2	3	4	5
12	0.019	-0.008	-0.009	-0.008	-0.005
24	0.046	0.012	0.007	0.004	0.001
48	0.572	0.505	0.518	0.516	0.515
96	0.574	0.465	0.529	0.529	0.527
144	0.582	0.480	0.506	0.505	0.505
240	0.584	0.516	0.536	0.535	0.535
336	0.590	0.472	0.504	0.503	0.502

Table 4: Available figures of merit for the persistence and NWP models.

Model	Figure of merit	Prediction for time step 1	2	3	4	5
Persistence	NRMSE	0.560	0.412	0.344	0.291	0.245
	PCC W-E	0.922	0.870	0.831	0.797	0.767
	PCC S-N	0.870	0.751	0.704	0.663	0.625
NWP	NRMSE	-0.093	-0.093	-0.093	-0.093	-0.093
	PCC W-E	0.637	0.637	0.637	0.637	0.637
	PCC S-N	0.353	0.353	0.353	0.353	0.353

4. Discussion

We demonstrate that experimental modelling is useful for predicting wind speed and direction at a given location even when the available training data set from the location is small and covers a short time period. Such modelling could be applied in major accidents involving dangerous chemicals when a temporary meteorological station is installed at a site and atmospheric dispersion modelling for near-term future is necessary. Models with 48 or more training data points, corresponding to 1 day or more of training data, perform well, while window size of 24 or fewer data points is not sufficient. For example, for predictions 1 time step in advance, NRMSE = 0.572 for type 1 model with 48 training data points and NRMSE = 0.046 with 24 points. The models with 48 points perform better than the persistence model over the whole tested span of 1 to 5 time steps in advance, the advantage is particularly pronounced for the later time steps – in step 1, NRMSE = 0.572 for type 1 model and NRMSE = 0.560 for persistence model, while in step 5, NRMSE = 0.515 for type 1 and NRMSE = 0.245 for persistence. The NWP model is worse than persistence with NRMSE = −0.093, reflecting the fact that it is predicting a cell average and not the local value.

We only refer to NRMSE values because they are in agreement with MSLL and PCC. We prefer NRMSE as it is computed for wind as a 2D vector and not for a single Cartesian component obtained from a single model.

The demonstration of the benefits of system identification can, strictly speaking, only apply to the study location. However, as the study area is a fairly typical industrialized area, similar tests at other relevant locations should give similar results:

- NWP models are available globally. For atmospheric dispersion modelling in the case of an accident, one uses the best available NWP model at the location as input. We use a state-of-the-art fine resolution model in the study so that operational models at typical accident locations may be comparable to it for some time to come.
- A dense meteorological station network is beneficial but not crucial, type 2 models do not require permanent stations.
- Less complex terrain would benefit both the proposed method and its competitors, i.e., NWP and persistence models.

While all the meteorological stations used measure wind at the same height of 10 m and the NWP model predicts it at the same height, one does not have to assume that all the heights are the same. For example, if the portable temporary meteorological station is measuring wind at a lower height, the method is applicable without any modification and the model will be predicting the wind velocity at the height of the temporary station.

For 2 to 5 time steps in advance, type 2 models that do not use measurements from the surrounding weather stations are better in NRMSE than type 1 models using measurements from the surrounding weather stations as inputs – for example, in step 5 with 48 training data points, NRMSE = 0.515 with type 1 model and NRMSE = 0.551 with type 2 model. In the first time step, type 1 models are marginally better with NRMSE = 0.572 than type 2 models with NRMSE = 0.563. The finding confirms the hypothesis that type 1 models should be better than type 2 models for the initial time steps because they have better information on the initial weather situation, and that type 2 models should be better for later time steps because they better use the information relevant for predicting those.

In the steps from 2 on, type 1 models would benefit from use of MOS instead of the raw NWP data for substituting the missing measurements of the input signals. It would be doable for the use case of the model, because historic data for training of the MOS models is available for the permanent meteorological stations. We do not attempt it as the potential for improving on type 2 models in the further time steps is proportional to the advantage that type 1 models have over type 2 models in time step 1, which is very small in our example. Nevertheless, we cannot say that type 2 models are better than type 1 models in general. Firstly, type 1 models are better in the shortest-term prediction, which is the most important one in the intended use of the model. Secondly, the performance of type 1 models that do use measurements from the surrounding weather stations depends on the quantity and quality of these stations, while the performance of type 2 models that do not use these measurements is more dependent on the quality of NWP predictions, so instrumenting the site differently would lead to a different conclusion.

A notable strength of the presented models is that they perform well throughout the whole test period, under all circumstances they encounter, as demonstrated by the values of the figures of merit. All the figures shown refer to random weather, we never select the meteorological situations in which the models perform best to evaluate them only on those. We have not

explored the influence of the weather on the performance of the model for a couple of reasons:

- Such higher-order results tend to transfer to other examples less well.
- While NRMSE is sufficient for comparing the models to one another, one would need a dedicated figure of merit based on the requirements for limiting the consequences of the accident to figure out in what weather the model helps more. NRMSE depends only on the vector difference between the measurement and the prediction, while in the accident it may matter whether the error is in speed or in direction, or if it occurs in strong or weak wind.

In experimental modelling, it is known that it is beneficial if the training data is similar to the modelled situation. We can thus infer that the model will perform well under circumstances similar to the ones already encountered since the installation of the temporary meteorological station and less well in different ones. If, for example, the conditions since the beginning of the accident have been calm, it is unlikely to perform well when it encounters its first weather disturbance.

The tested models with very small active data sets may suffer from a poor ratio of the number of training data points to the number of regressors (Anatolyev, 2012). Attempts at remedying this situation may be recommended as a worthwhile further work, particularly because the smallest data sets are the most common – every data set that grows big during the accident starts as a small one.

In this work, we have avoided the effects of regressor selection. We have selected them offline on the basis of data that would not be available in a real use case. For real use, they would have to be selected either on the basis of the available training data or heuristically.

There is a major difference between the results of GP models with linear covariance function and of linear models determined by the least squares method, even though both of these models are linear. The difference results from the different assumptions regarding noise. The least squares method leads to the best linear unbiased estimator if the noise on the output values satisfies the conditions of Gauss–Markov theorem (Puntanen and Styan, 1989), while likelihood optimization makes no such assumption.

If a different universal approximator was used in place of GP, for example an artificial neural network, one would expect similar results. The particular

mathematical structure used to identify the system tends not to have a big influence. A peculiarity of GP that is taken advantage of in this work is that it predicts the output variance, giving an estimate of the model's confidence in its output.

5. Conclusion

We predict local wind speed and direction for the near future with a method adapted for use in emergencies. Unlike experimental models of wind for other purposes, the tested experimental models are based on a small amount of training data. One can identify such models for a site of a portable temporary meteorological station soon after the station is installed.

We find out that such modelling is feasible. We demonstrate that 1 day of training data with a sampling interval of 30 minutes suffices for making predictions that are better than the available alternatives of either using the current measured value to approximate the future or using NWP to predict the local wind. The best results are achieved with GP models using linear functions as covariance functions. As their inputs, the models use numerical weather predictions and in some examples the measurements from the surrounding weather stations.

Such predictions can be used for atmospheric dispersion modelling in case of a major accident involving dangerous chemicals for which a temporary meteorological station is installed. The method is ready to be implemented in a real or simulated accident.

The modelling relies on the model inputs being available throughout the accident. All the models use NWP signal values and some also use observations from meteorological stations in the surrounding area, so a NWP model of the area is necessary and a meteorological station network is beneficial. The method is thus applicable globally, as long as NWP model predictions are available over the whole world.

In the studied location, the tested experimental models fail to produce good predictions with 12 hours of training data or less. Any decrease in the necessary amount of training data would be highly valuable, so we recommend exploring the possibilities for achieving it.

Acknowledgements

We are grateful to the NPP Krško for the measurement data from their automatic meteorological measuring system.

The authors acknowledge the financial support from the Slovenian Research Agency (projects No. L2-8174 and L2-2615, research core funding No. P2-0001).

References

Anatolyev, S., 2012. Inference in regression models with many regressors. J. Econ. 170, 368–382. doi:10.1016/j.jeconom.2012.05.011.

Barratt, R., 2013. Atmospheric dispersion modelling: an introduction to practical applications. Taylor & Francis, Abingdon.

Bédard, J., Yu, W., Gagnon, Y., Masson, C., 2013. Development of a geophysic model output statistics module for improving short-term numerical wind predictions over complex sites. Wind Energy 16, 1131–1147. doi:10.1002/we.1538.

Beelen, R., Voogt, M., Duyzer, J., Zandveld, P., 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. doi:10.1016/j.atmosenv.2010.08.005.

Breznik, B., Božnar, M.Z., Mlakar, P., Tinarelli, G., 2003. Dose projection using dispersion models. Int. J. Environ. Pollut. 20, 278–285. doi:10.1504/IJEP.2003.004291.

Carini, A., Cecchi, S., Orcioni, S., 2018. Chapter 2 - orthogonal LIP nonlinear filters, in: Comminiello, D., Príncipe, J.C. (Eds.), Adaptive Learning Methods for Nonlinear System Modeling. Butterworth-Heinemann, pp. 15–46. doi:10.1016/B978-0-12-812976-0.00003-8.

Chen, N., Qian, Z., Meng, X., 2013a. Multistep wind speed forecasting based on wavelet and Gaussian processes. Math. Probl. Eng. 2013. doi:10.1155/2013/461983.

Chen, N., Qian, Z., Meng, X., Nabney, I., 2013b. Short-term wind power forecasting using gaussian processes, in: International joint conference on Artificial Intelligence IJCAI'13, pp. 1771–1777.

©European Union, 2019. Copernicus Land Monitoring Service 2018. European Environment Agency (EEA).

Fabbri, L., Wood, M.H., 2019. Accident damage analysis module (adam): Novel european commission tool for consequence assessment—scientific evaluation of performance. Process Saf. Environ. Prot. 129, 249 – 263. doi:10.1016/j.psep.2019.07.007.

Glahn, H.R., Lowry, D.A., 1972. The use of model output statistics (MOS) in objective weather forecasting. J. Appl. Meteorol. 11, 1203–1211.

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

Grašič, B., Mlakar, P., Božnar, M.Z., Kocijan, J., 2018. Validation of numerically forecasted vertical temperature profile with measurements for dispersion modelling. Int. J. Environ. Pollut. 64, 22–34. doi:10.1504/IJEP.2018.099143.

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 doi:10.1016/j.renene.2018.04.019.

Jones, A., Thomson, D., Hort, M., Devenish, B., 2007. The U.K. Met Office's Next-Generation Atmospheric Dispersion Model, NAME III, in: Borrego, C., Norman, A.L. (Eds.), Air Pollution Modeling and Its Application XVII, Springer US, Boston, MA. pp. 580–589.

Kalnay, E., 2003. Atmospheric modeling, data assimilation and predictability. Cambridge university press.

Kneale, C., Brown, S.D., 2018. Small moving window calibration models for soft sensing processes with limited history. Chemom. Intell. Lab. Syst. 183, 36–46. doi:10.1016/j.chemolab.2018.10.007.

Kocijan, J., 2016. Modelling and Control of Dynamic Systems Using Gaussian Process Models. Springer International Publishing, Cham. doi:10.1007/978-3-319-21021-6.

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, 2019–2032. doi:10.1007/s00477-019-01736-5.

Li, K., Peng, J.X., 2007. Neural input selection–A fast model-based approach. Neurocomputing 70, 762–769. doi:10.1016/j.neucom.2006.10.011.

Ljung, L., Singh, R., 2012. Version 8 of the MATLAB system identification toolbox. IFAC Proceedings Volumes 45, 1826–1831.

Mori, H., Kurata, E., 2008. Application of Gaussian Process to wind speed forecasting for wind power generation, in: 2008 IEEE International Conference on Sustainable Energy Technologies, IEEE, New York. pp. 956–959. doi:https://doi.org/10.1109/ICSET.2008.4747145.

Nelles, O., 2001. Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models. Engineering online library, Springer-Verlag, Berlin. doi:10.1007/978-3-662-04323-3.

O'Mahony, M.T., Doolan, D., O'Sullivan, A., Hession, M., 2008. Emergency planning and the Control of Major Accident Hazards (COMAH/Seveso II) Directive: An approach to determine the public safety zone for toxic cloud releases. J. Hazard. Mater. 154, 355–365. doi:10.1016/j.jhazmat.2007.10.065.

Pal, A., Karim, I.U., Mondal, B., Raha, S., 2018. A similarity based fuzzy system as a function approximator. Int. J. Intell. Sci. 8, 89–116. doi:10.4236/ijis.2018.84005.

Perne, M., Stepančič, M., Grašič, B., 2019. Handling big datasets in Gaussian processes for statistical wind vector prediction. IFAC-PapersOnLine 52, 110–115. doi:10.1016/j.ifacol.2019.09.126. 5th IFAC Conference on Intelligent Control and Automation Sciences ICONS 2019.

Puntanen, S., Styan, G.P.H., 1989. The equality of the ordinary least squares estimator and the best linear unbiased estimator. Am. Stat. 43, 153–161. doi:10.1080/00031305.1989.10475644.

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

Shi, J.Q., Choi, T., 2011. Gaussian process regression analysis for functional data. CRC Press.

Sigg, R., Lindgren, P., von Schoenberg, P., Persson, L., Burman, J., Grahn, H., Bränström, N., Björnham, O., 2018. Hazmat risk area assessment by atmospheric dispersion modelling using latin hypercube sampling with weather ensemble. Meteorol. Appl. 25, 575–585. doi:10.1002/met.1722.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Wang, W., Powers, J.G., 2008. A description of the Advanced Research WRF version 3. NCAR Technical Note, NCAR/TN-475+STR. National Center for Atmospheric Research, Boulder.

Soares, J., Oliveira, A.P., Božnar, M.Z., Mlakar, P., Escobedo, J.F., Machado, A.J., 2004. Modeling hourly diffuse solar-radiation in the city of São Paulo using a neural-network technique. Appl. Energy 79, 201 – 214. doi:10.1016/j.apenergy.2003.11.004.

The European Commission, 2019. Major accident hazards: The seveso directive – summary of requirements. URL: `https://ec.europa.eu/environment/seveso/legislation.htm`. [accessed 21 April 2020].

The European Commission, 2020. Major accident hazards: The seveso directive – technological disaster risk reduction. URL: `https://ec.europa.eu/environment/seveso/`. [accessed 21 April 2020].

Torrontegui, E., García-Ripoll, J.J., 2019. Unitary quantum perceptron as efficient universal approximator. EPL Europhys. Lett. 125, 30004. doi:10.1209/0295-5075/125/30004.

Yan, J., Li, K., Bai, E.W., Deng, J., Foley, A.M., 2016. Hybrid probabilistic wind power forecasting using temporally local gaussian process. IEEE Trans. Sustain. Energy 7, 87–95. doi:10.1109/TSTE.2015.2472963.

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. Manag. 126, 1084–1092. doi:10.1016/j.enconman.2016.08.086.

Zhang, Z., Ye, L., Qin, H., Liu, Y., Wang, C., Yu, X., Yin, X., Li, J., 2019. Wind speed prediction method using shared weight long short-term memory network and gaussian process regression. Appl. Energy 247, 270–284. doi:10.1016/j.apenergy.2019.04.047.

Appendix A. Figures of merit

Appendix A.1. NRMSE

NRMSE (Ljung and Singh, 2012) is defined as

$$\mathrm{NRMSE} = 1 - \frac{\|\mathbf{y} - \mu\|}{\|\mathbf{y} - E(\mathbf{y})\|}, \tag{A.1}$$

where $\mathbf{y}$ is the vector of measured values, $E(\mathbf{y})$ is the mean of the measured value, and μ 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. It can be calculated for vector quantities. We use it to evaluate predictions of horizontal wind as a 2D vector.

Appendix A.2. MSLL

MSLL is defined as (Rasmussen and Williams, 2006)

$$\mathrm{MSLL} = \frac{1}{2N} \sum_{i=1}^{N} \left[\ln\left(\sigma_i^2\right) - \ln\left(\sigma_y^2\right) + \frac{\left(E\left(\hat{y}_i\right) - y_i\right)^2}{\sigma_i^2} - \frac{\left(y_i - E\left(\mathbf{y}\right)\right)^2}{\sigma_y^2} \right], \tag{A.2}$$

where y_i is the measured value, σ_y^2 is the variance of the measured value, $E\left(\hat{y}_i\right)$ 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 the predictive variance into account. A lower MSLL value corresponds to a better model, the values are typically negative.

Appendix A.3. PCC

Pearson correlation coefficient is defined as

$$\mathrm{PCC} = \frac{\mathrm{cov}(\mathbf{y}, \mu)}{\sigma_{\mathbf{y}} \sigma_{\mu}}, \tag{A.3}$$

where cov is covariance and σ_μ is the standard deviation of the predicted (mean) value. The value of PCC is 1 for perfect linear correlation, 0 for no correlation, and -1 for perfect negative correlation.

Supporting Material

Improving wind vector predictions for modelling of atmospheric dispersion during Seveso-type accidents

Matija Perne[a,*], Marija Zlata Božnar[b], Boštjan Grašič[b], Primož Mlakar[b], Juš Kocijan[a,c]

[a] *Jožef Stefan Institute, Ljubljana, Slovenia*
[b] *MEIS d.o.o., Mali Vrh, Slovenia*
[c] *University of Nova Gorica, Nova Gorica, Slovenia*

Content

*Corresponding author. Department of Systems and Control, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia. E-mail address: matija.perne@ijs.si.

Table S1: Locations of meteorological stations.

Meteorological station locations

Name	UTM grid zone 33T		WGS84		Height [m]
	east	north	latitude	longitude	
Stolp at Krško NPP	539776	5087498	45.939900	15.513132	10
Brežice	546266	5083861	45.906760	15.596502	10
Cerklje	540614	5081216	45.883312	15.523411	10
Cerklje Airport	540035	5083159	45.900833	15.516111	10
Krško	538593	5088899	45.952577	15.497984	10
Lisca	522034	5101613	46.067735	15.284905	10

Table S2: Regressors used in type 1 models using measurements as inputs. Delay is measured in time steps. The regressors are listed from best to worst as ranked by ProOpter IVS LIP method.

Best regressors for W-E wind

source	variable	delay
Stolp at Krško NPP	W-E wind	1
NWP	W-E wind	0
Cerklje Airport	W-E wind	1
Brežice	W-E wind	1
Krško	W-E wind	1
Stolp at Krško NPP	W-E wind	2
Krško	relative humidity	1
Cerklje Airport	S-N wind	1
Lisca	air pressure	2
NWP	cloudiness	1
Brežice	W-E wind	1
NWP	global solar radiation	0
Cerklje Airport	air pressure	2
Cerklje Airport	temperature	2
Krško	relative humidity	2

Best regressors for S-N wind

source	variable	delay
Stolp at Krško NPP	S-N wind	1
Brežice	S-N wind	1
Cerklje Airport	S-N wind	1
NWP	global solar radiation	0
Cerklje Airport	W-E wind	1
Krško	temperature	1
Cerklje Airport	S-N wind	2
Krško	S-N wind	1
Krško	temperature	2
Cerklje	temperature	2
NWP	S-N wind	0
NWP	diffuse solar radiation	1
NWP	W-E wind	2
Lisca	S-N wind	2
Stolp at Krško NPP	S-N wind	2

Table S3: Regressors used in type 2 models not using measurements as inputs. Delay is measured in time steps. The regressors are listed from best to worst as ranked by ProOpter IVS LIP method.

Best regressors for W-E wind

source	variable	delay
Stolp at Krško NPP	W-E wind	1
NWP	W-E wind	0
Stolp at Krško NPP	S-N wind	1
NWP	air pressure	3
NWP	global solar radiation	2
Stolp at Krško NPP	S-N wind	2
NWP	cloudiness	1
NWP	S-N wind	1
NWP	diffuse solar radiation	1
NWP	S-N wind	0
NWP	global solar radiation	0
NWP	temperature	3
NWP	diffuse solar radiation	0
NWP	air pressure	0
Stolp at Krško NPP	W-E wind	2

Best regressors for S-N wind

source	variable	delay
Stolp at Krško NPP	S-N wind	1
NWP	S-N wind	1
NWP	global solar radiation	0
Stolp at Krško NPP	W-E wind	1
NWP	air pressure	2
NWP	temperature	4
NWP	air pressure	1
Stolp at Krško NPP	W-E wind	2
Stolp at Krško NPP	S-N wind	2
NWP	W-E wind	1
NWP	air pressure	0
NWP	S-N wind	3
NWP	W-E wind	4
NWP	relative humidity	4
NWP	diffuse solar radiation	2

Table S4: Dependence of MSLL values for each component on the number of training data points and the number of time steps ahead for type 2 models not using measurements as inputs.

Component	Window size	Prediction for time step 1	2	3	4	5
W-E	12	-0.207	-0.113	-0.017	0.077	0.155
	24	-0.386	-0.258	-0.126	-0.022	0.075
	48	-0.889	-0.867	-0.857	-0.858	-0.848
	96	-0.957	-0.948	-0.940	-0.938	-0.934
	144	-0.931	-0.925	-0.917	-0.911	-0.905
	240	-0.913	-0.913	-0.911	-0.909	-0.907
	336	-0.909	-0.908	-0.908	-0.906	-0.904
S-N	12	-0.181	0.167	0.299	0.363	0.411
	24	-0.166	0.062	0.222	0.341	0.437
	48	-0.695	-0.678	-0.655	-0.643	-0.632
	96	-0.678	-0.660	-0.623	-0.610	-0.597
	144	-0.666	-0.654	-0.643	-0.630	-0.624
	240	-0.660	-0.656	-0.641	-0.637	-0.635
	336	-0.657	-0.653	-0.641	-0.638	-0.635

Table S5: Dependence of PCC values for W-E component on the number of training data points and the number of time steps ahead for type 2 models not using measurements as inputs.

Component	Window size	Prediction for time step 1	2	3	4	5
W-E	12	0.396	0.410	0.386	0.367	0.337
	24	0.701	0.728	0.692	0.663	0.640
	48	0.915	0.985	0.928	0.884	0.848
	96	0.922	0.995	0.926	0.878	0.841
	144	0.920	0.993	0.922	0.874	0.837
	240	0.921	0.996	0.922	0.872	0.835
	336	0.922	0.993	0.921	0.875	0.840
S-N	12	0.007	0.047	0.002	-0.020	0.003
	24	0.344	0.341	0.327	0.303	0.282
	48	0.872	0.986	0.882	0.777	0.731
	96	0.869	0.988	0.877	0.765	0.718
	144	0.868	0.985	0.878	0.766	0.716
	240	0.870	0.997	0.873	0.757	0.710
	336	0.871	0.994	0.873	0.760	0.713

Table S6: Dependence of NRMSE values for wind vector predictions on the number of training data points and the number of time steps ahead for type 2 models not using measurements as inputs.

Window size	Prediction for time step 1	2	3	4	5
12	0.047	0.043	0.037	0.031	0.026
24	0.189	0.179	0.168	0.158	0.148
48	0.563	0.557	0.556	0.553	0.551
96	0.571	0.568	0.567	0.566	0.563
144	0.567	0.565	0.565	0.563	0.561
240	0.571	0.569	0.570	0.569	0.568
336	0.570	0.569	0.570	0.569	0.568

Table S7: Results of GP models with squared exponential covariance function. Dependence of MSLL values for each component on the number of training data points and the number of time steps ahead for type 1 models using measurements as inputs and type 2 models not using measurements as inputs.

Type	Component	Window size	Prediction for time step 1	2	3	4	5
1	W-E	12	22.413	222.820	16.588	307.091	23.304
		24	1642.799	3.582	16.211	5.299	1.610
		48	-0.515	-0.020	0.007	0.042	0.083
		96	-0.898	-0.033	-0.050	-0.014	0.008
		144	-0.832	0.671	0.796	0.827	0.849
		240	-0.914	-0.248	-0.263	-0.233	-0.204
		336	-0.916	-0.382	-0.416	-0.399	-0.381
	S-N	12	82186034.977	7.023	0.769	0.659	0.501
		24	-0.071	0.159	0.271	0.352	0.409
		48	-0.043	19.627	0.150	0.229	0.257
		96	-0.400	0.096	0.659	0.759	0.763
		144	-0.499	-0.005	0.033	0.096	0.142
		240	-0.616	-0.409	-0.426	-0.396	-0.364
		336	-0.560	-0.272	-0.277	-0.245	-0.209
2	W-E	12	0.030	0.391	0.493	0.535	0.561
		24	-0.152	-0.004	0.092	0.211	0.269
		48	-0.678	-0.494	-0.368	-0.276	-0.242
		96	-0.834	-0.720	-0.640	-0.596	-0.559
		144	-0.891	-0.863	-0.837	-0.819	-0.800
		240	-0.885	-0.852	-0.822	-0.789	-0.767
		336	-0.874	-0.833	-0.798	-0.766	-0.743
	S-N	12	70371.527	0.245	0.373	0.376	0.409
		24	982808.577	4941.409	70.513	0.697	0.793
		48	-0.380	-0.050	0.061	0.177	0.286
		96	-0.620	-0.514	-0.464	-0.428	-0.386
		144	-0.598	-0.483	-0.421	-0.360	-0.305
		240	-0.583	-0.498	-0.464	-0.454	-0.419
		336	-0.544	-0.481	-0.442	-0.416	-0.374

Table S8: Results of GP models with squared exponential covariance function. Dependence of PCC values for each component on the number of training data points and the number of time steps ahead for type 1 models using measurements as inputs and type 2 models not using measurements as inputs.

Type	Component	Window size	Prediction for time step 1	2	3	4	5
1	W-E	12	0.609	0.467	0.448	0.460	0.450
		24	0.480	0.228	0.199	0.189	0.173
		48	0.848	0.741	0.738	0.723	0.710
		96	0.915	0.733	0.736	0.704	0.675
		144	0.908	0.664	0.652	0.637	0.618
		240	0.919	0.809	0.755	0.723	0.697
		336	0.922	0.834	0.807	0.773	0.746
	S-N	12	0.249	0.229	0.102	0.090	0.059
		24	-0.003	-0.014	0.019	0.016	0.014
		48	0.691	0.586	0.274	0.254	0.255
		96	0.767	0.591	0.435	0.422	0.375
		144	0.810	0.656	0.498	0.390	0.315
		240	0.859	0.876	0.805	0.725	0.677
		336	0.814	0.769	0.746	0.651	0.598
2	W-E	12	0.323	0.311	0.306	0.294	0.286
		24	0.317	0.302	0.284	0.256	0.244
		48	0.887	0.894	0.863	0.833	0.806
		96	0.903	0.939	0.895	0.867	0.841
		144	0.914	0.986	0.922	0.881	0.849
		240	0.916	0.976	0.912	0.868	0.834
		336	0.915	0.973	0.913	0.874	0.843
	S-N	12	0.037	0.031	0.039	0.033	0.020
		24	0.098	0.070	0.054	0.070	0.056
		48	0.791	0.730	0.701	0.655	0.614
		96	0.818	0.858	0.783	0.699	0.654
		144	0.835	0.865	0.754	0.668	0.616
		240	0.845	0.893	0.798	0.707	0.658
		336	0.828	0.891	0.828	0.737	0.706

Table S9: Results of linear models based on least squares. Dependence of NRMSE values for wind vector predictions on the number of training data points and the number of time steps ahead for type 1 models using measurements as inputs and type 2 models not using measurements as inputs.

Model	Window	Prediction for time step				
Type	size	1	2	3	4	5
	12	0.135	0.069	0.058	0.056	0.049
	24	0.072	0.008	0.002	-0.002	-0.004
	48	0.389	0.238	0.169	0.161	0.153
1	96	0.497	0.238	0.191	0.173	0.161
	144	0.512	0.195	0.161	0.146	0.132
	240	0.560	0.376	0.379	0.365	0.352
	336	0.531	0.336	0.368	0.359	0.351
	12	0.031	0.024	0.018	0.014	0.012
	24	0.026	0.020	0.014	0.006	0.003
	48	0.476	0.384	0.337	0.301	0.273
2	96	0.512	0.457	0.423	0.403	0.385
	144	0.537	0.489	0.465	0.439	0.417
	240	0.545	0.507	0.488	0.468	0.450
	336	0.532	0.515	0.486	0.463	0.442

Table S10: Results of GP models with sum of squared exponential and linear covariance function. Dependence of MSLL values for each component on the number of training data points and the number of time steps ahead for type 1 models using measurements as inputs and type 2 models not using measurements as inputs.

Type	Component	Window size	Prediction for time step 1	2	3	4	5
1	W-E	48	-0.921	-0.760	-0.798	-0.802	-0.792
		240	-0.946	-0.525	-0.640	-0.633	-0.633
		336	-0.944	-0.294	-0.447	-0.438	-0.440
	S-N	48	-0.473	-0.196	-0.278	-0.230	-0.211
		240	-0.634	-0.483	-0.487	-0.478	-0.461
		336	-0.655	-0.404	-0.381	-0.382	-0.370
2	W-E	48	-0.902	-0.877	-0.863	-0.861	-0.852
		240	-0.901	-0.885	-0.861	-0.845	-0.834
		336	-0.888	-0.869	-0.852	-0.847	-0.840
	S-N	48	0.463	0.409	0.318	0.463	0.325
		240	-0.562	-0.476	-0.439	-0.427	-0.416
		336	-0.637	-0.601	-0.561	-0.567	-0.551

Table S11: Results of GP models with sum of squared exponential and linear covariance function. Dependence of PCC values for each component on the number of training data points and the number of time steps ahead for type 1 models using measurements as inputs and type 2 models not using measurements as inputs.

Type	Component	Window size	Prediction for time step 1	2	3	4	5
1	W-E	48	0.918	0.969	0.910	0.865	0.830
		240	0.928	0.917	0.877	0.836	0.804
		336	0.927	0.883	0.853	0.817	0.786
	S-N	48	0.807	0.829	0.786	0.720	0.693
		240	0.863	0.923	0.840	0.736	0.693
		336	0.864	0.893	0.810	0.703	0.660
2	W-E	48	0.916	0.988	0.921	0.874	0.838
		240	0.919	0.983	0.922	0.880	0.848
		336	0.915	0.981	0.927	0.886	0.854
	S-N	48	0.705	0.631	0.607	0.566	0.502
		240	0.864	0.933	0.837	0.749	0.701
		336	0.861	0.950	0.868	0.773	0.729

Table S12: Results of GP models with sum of squared exponential and linear covariance function. Dependence of NRMSE values for wind vector predictions on the number of training data points and the number of time steps ahead for type 1 models using measurements as inputs and type 2 models not using measurements as inputs.

Model Type	Window size	Prediction for time step 1	2	3	4	5
1	48	0.524	0.463	0.468	0.460	0.451
	240	0.572	0.470	0.494	0.491	0.487
	336	0.573	0.428	0.449	0.448	0.446
2	48	0.464	0.404	0.385	0.368	0.349
	240	0.563	0.543	0.532	0.521	0.514
	336	0.555	0.542	0.535	0.532	0.525

Table S13: Results of linear models based on least squares. Dependence of MSLL values for each component on the number of training data points and the number of time steps ahead for type 1 models using measurements as inputs and type 2 models not using measurements as inputs.

Type	Component	Window size	Prediction for time step 1	2	3	4	5
1	W-E	24	-0.198	0.691	2.084	1.977	2.015
		48	-0.783	0.076	0.971	1.146	1.056
		96	-0.933	0.236	0.576	1.121	1.148
		144	-0.938	0.217	0.387	0.913	1.045
		240	-0.926	0.235	0.609	0.997	1.145
		336	-0.968	0.119	0.517	0.804	0.924
	S-N	24	-0.107	0.575	0.698	0.733	0.760
		48	-0.498	0.100	0.235	0.262	0.312
		96	-0.670	-0.090	0.225	0.367	0.425
		144	-0.686	-0.058	0.375	0.639	0.789
		240	-0.708	-0.074	0.318	0.605	0.775
		336	-0.685	-0.087	0.357	0.666	0.847
2	W-E	48	-0.656	-0.097	0.298	0.559	0.727
		96	-0.865	-0.399	-0.023	0.258	0.509
		144	-0.871	-0.397	0.025	0.382	0.713
		240	-0.894	-0.536	-0.261	-0.034	0.193
		336	-0.905	-0.579	-0.356	-0.180	-0.017
	S-N	24	0.045	0.455	0.638	0.750	0.869
		48	-0.458	0.111	0.389	0.553	0.690
		96	-0.573	0.004	0.283	0.459	0.632
		144	-0.650	-0.129	0.147	0.350	0.524
		240	-0.659	-0.126	0.149	0.360	0.518
		336	-0.660	-0.146	0.078	0.248	0.370

Table S14: Results of linear models based on least squares. Dependence of PCC values for each component on the number of training data points and the number of time steps ahead for type 1 models using measurements as inputs and type 2 models not using measurements as inputs.

Type	Component	Window size	Prediction for time step 1	2	3	4	5
1	W-E	12	0.339	0.029	-0.016	-0.005	-0.003
		24	0.524	0.080	0.089	0.025	0.025
		48	0.862	0.544	0.350	0.334	0.340
		96	0.916	0.611	0.472	0.380	0.389
		144	0.922	0.683	0.636	0.557	0.551
		240	0.925	0.704	0.581	0.491	0.463
		336	0.930	0.727	0.648	0.589	0.559
	S-N	12	0.215	0.081	0.014	-0.002	-0.001
		24	0.526	0.153	0.118	0.075	0.012
		48	0.789	0.467	0.377	0.368	0.327
		96	0.844	0.687	0.632	0.614	0.604
		144	0.863	0.734	0.647	0.581	0.533
		240	0.879	0.776	0.691	0.611	0.558
		336	0.878	0.787	0.684	0.596	0.532
2	W-E	12	0.049	-0.011	0.013	-0.013	0.013
		24	0.051	0.025	0.011	0.005	0.001
		48	0.834	0.663	0.578	0.466	0.310
		96	0.904	0.873	0.839	0.799	0.760
		144	0.909	0.880	0.844	0.806	0.772
		240	0.916	0.897	0.868	0.837	0.806
		336	0.918	0.903	0.881	0.858	0.837
	S-N	12	0.223	0.027	0.013	0.016	-0.010
		24	0.187	0.059	0.029	0.026	0.028
		48	0.777	0.583	0.451	0.374	0.305
		96	0.837	0.770	0.725	0.695	0.666
		144	0.859	0.811	0.766	0.728	0.693
		240	0.866	0.827	0.783	0.742	0.704
		336	0.868	0.841	0.804	0.764	0.727

Table S15: Results of linear models based on least squares. Dependence of NRMSE values for wind vector predictions on the number of training data points and the number of time steps ahead for type 1 models using measurements as inputs and type 2 models not using measurements as inputs.

Model Type	Window size	Prediction for time step 1	2	3	4	5
1	12	-2.134	-19.538	-417.830	-12536.585	-417550.192
	24	-0.466	-6.890	-10.691	-15.663	-19.024
	48	0.416	-0.166	-1.080	-1.248	-1.228
	96	0.537	0.129	-0.189	-0.453	-0.470
	144	0.561	0.238	0.061	-0.150	-0.209
	240	0.581	0.286	0.169	0.074	0.038
	336	0.588	0.310	0.231	0.175	0.147
2	12	-19.695	-4283.499	-976890.794	-223019107.998	-50911751152.568
	24	-11.993	-24.546	-48.804	-62.648	-74.886
	48	0.360	-0.113	-0.432	-0.793	-1.576
	96	0.517	0.334	0.232	0.163	0.109
	144	0.542	0.384	0.297	0.234	0.183
	240	0.558	0.416	0.342	0.288	0.244
	336	0.564	0.429	0.365	0.320	0.286

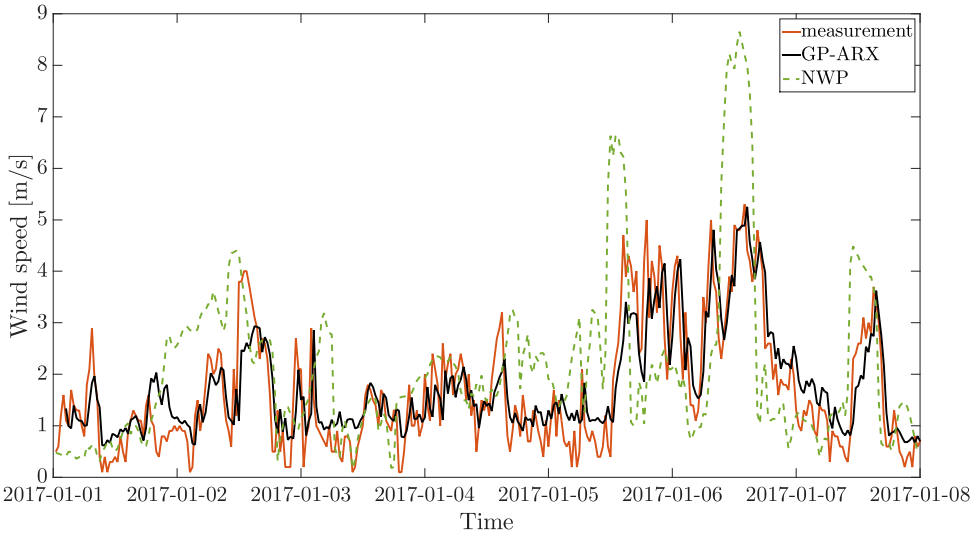


Figure S1: Predicted wind speed obtained from 5 step ahead predictions of type 2 models not using measurements as inputs with 48 training data points.

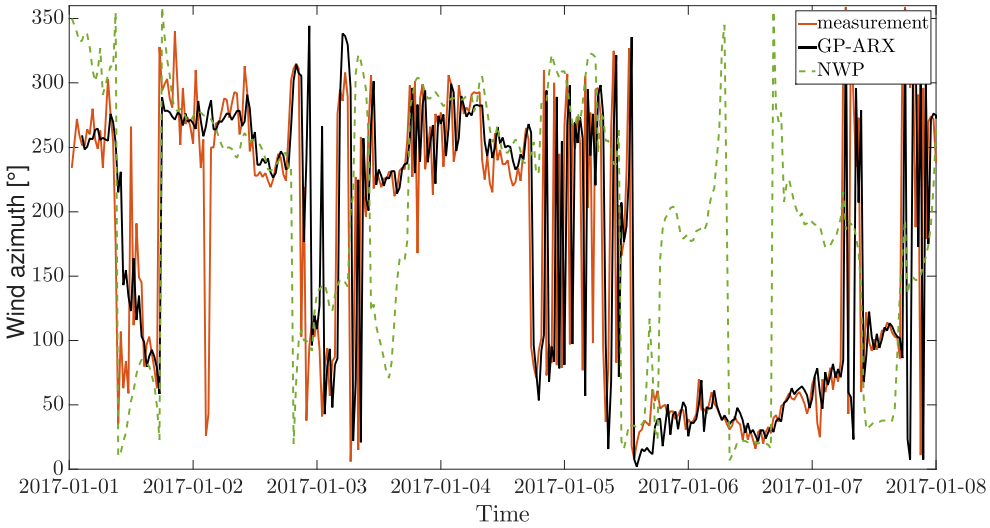


Figure S2: Predicted wind direction obtained from 5 step ahead predictions of type 2 models not using measurements as inputs with 48 training data points.