Hybrid modeling of atmospheric pollutant spread

Aljaž Pavšek [*,**] **Juš Kocijan** [*,***]

[*] *Jozef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia*
[**] *Jozef Stefan International Postgraduate School, Jamova cesta 39, 1000 Ljubljana, Slovenia (e-mail: aljaz.pavsek@ijs.si)*
[***] *University of Nova Gorica, 5000 Nova Gorica, Slovenia*

Abstract:
Accurate atmospheric pollutant dispersion often relies on advanced physics-based models, but they can be too computationally expensive for real-time use in Seveso-type accidents. In contrast, more efficient physical models may lack sufficient accuracy over complex terrain. Our approach combines a simple, physics-based Gaussian plume model—suitable for flat regions—with a deep neural network that corrects the Gaussian model's prediction errors. The resulting surrogate effectively replaces the more sophisticated and computationally expensive Lagrangian particle dispersion model, retaining part of its predictive accuracy while substantially reducing the overall computational cost.

Keywords: Hybrid model, Surrogate model, Deep Neural Networks, Gaussian plume

1. INTRODUCTION

This work reports on our integration of a physics-based model (PBM) with a data-driven model (DDM) within a hybrid modeling framework to improve the predictions of a standalone physics-based approach. The method is developed and tested on time-series data that track the evolution of selected atmospheric pollutants.

Combining physics-based and data-driven approaches has gained attention in recent years due to their potential to tackle diverse challenges in engineering and physical systems (Willard et al., 2022). Each approach has its strengths and weaknesses, and thoughtfully combining them can lead to more accurate, reliable, and computationally efficient solutions to complex numerical problems. In contrast to the now-popular techniques that embed physical knowledge directly into the DDM (e.g., Willard et al. (2022)), the traditional hybrid paradigm (e.g., Thompson and Kramer (1994)) treats the PBM and DDM as separate modules, whose inputs and outputs are explicitly linked. In this arrangement, the DDM can predict the residuals produced by the PBM with respect to true future system states directly, provided that a sufficient amount of data is available.

Problem statement Industrial sites regulated under the Seveso Directive (European Commission, 2020) pose a risk of high-impact chemical releases. In an emergency, decision-makers need pollutant-dispersion forecasts within minutes. Standard tools, such as the Lagrangian particle-dispersion model (Castelli et al., 2018), though accurate, are too computationally demanding for real-time use. Purely data-driven models, on the other hand, can exhibit unpredictable behaviour. Hence, there is a clear need for surrogate models that can predict pollutant spread both quickly and reliably.

Contribution We combine a simple legacy Gaussian plume model (Pasquill, 1961) with a deep neural network in a residual hybrid-modeling scheme to simulate pollutant dispersion over complex terrain. The resulting model remains rooted in physics yet delivers rapid predictions suitable for time-critical applications.

2. DATA AND METHODS

2.1 Data description

The data used in this study represent ground-level sulfur dioxide (SO_2) concentrations in the vicinity of the Velenje basin in Slovenia, primarily originating from emissions at the Šoštanj thermal power plant. The Velenje basin (Figure 1) has complex topography—the region features multiple rivers, lakes, and surrounding hills—posing a considerable challenge for accurately modeling pollutant dispersion. Such a simulation was performed by Mlakar et al. (2015), who assumed a continuous, constant chimney-like SO_2 source and combined observational data with weather forecasting programs to compute ground-level concentrations across a 15 km $\times$ 15 km region. Their model includes the Lagrangian particle dispersion model SPRAY (Castelli et al., 2018), the MINERVE diagnostic mass-consistent wind-field model (Finardi et al., 1998), and the SURFPRO meteorological preprocessor (Finardi et al., 1997). The SO_2 concentration fields used here were taken from a follow-up study (Mlakar et al., 2019) and are given as relative concentrations (s/m^3)—the ratio of absolute pollutant concentration (µg/m^3) to emission rate (kg/s). We use simplified meteorological variables, adopted from Kocijan et al. (2022), i.e., wind speed and direction (at 10 m and 500 m), global solar radiation, and temperature. All data were recorded at half-hour intervals. To reduce

Fig. 1. Location of the Šoštanj Thermal Power Plant (TPP) in the Velenje basin, a region of complex topography; the red rectangle outlines an area measuring 15 km × 15 km.

computational complexity, the original 100 × 100 grid was averaged to 25 × 25.

2.2 Physical model - Gaussian plume model

Atmospheric pollution models date back nearly a century (Sutton, 1932), with the Gaussian plume model by Pasquill (1961) among the most influential. Developed in an era of limited computational resources, it provides a straightforward analytical solution for pollutant concentration as a function of position relative to a point source, emission rate, and meteorological stability parameters. This model is most applicable to constant point-source emissions over flat terrain, where homogeneous meteorological conditions hold (e.g., Snoun et al. (2023)). Placing the source in the center of a Cartesian coordinate system, and assuming the wind is blowing in the x direction, the ground level concentration of a pollutant at coordinates (x, y) is given by

$$C^{(x)}(x, y, v) = \frac{Q}{\pi v \sigma_y \sigma_z} e^{-\frac{y^2}{2\sigma_y^2}} e^{-\frac{H^2}{2\sigma_z^2}}; \tag{1}$$

where Q is the emission rate, H is the height of the point-like source, and v is the speed of the wind. The spread of the pollutant in the cross-wind directions y is characterized by standard deviation σ_y and σ_z of the plume in y and z direction, which are simple functions of distance x and stability class parameters (e.g. Snoun et al. (2023)). The graphical representation of the Gaussian plume model is shown in Figure 2.

To also allow for a general wind direction ϕ, we rotate the coordinate frame so that the standard Gaussian plume formula still applies along its local x'-axis. Specifically, let

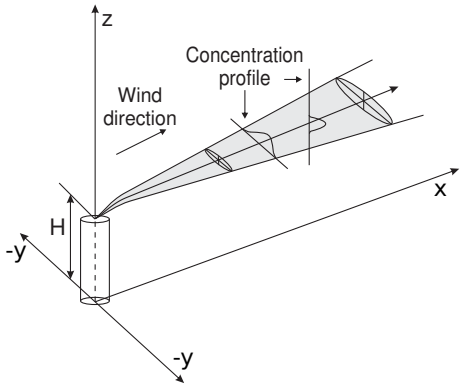

Fig. 2. Schematic of the Gaussian plume model—pollutant released from a point source at height H is advected downwind along the x axis and dispersed laterally and vertically along the perpendicular y and z axes.

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{2}$$

We then define

$$C(x, y, v, \phi) \;:=\; C^{(x)}(x', y', v), \tag{3}$$

which is the same Gaussian plume model as before, but evaluated in the rotated coordinates (x', y').

We recognize the limitations of the simple Gaussian plume model, including the steady-state assumption, the neglect of multiple reflections between parallel boundaries, and the impact of chemical reactions on pollutant deposition, among others. Nevertheless, we chose this model due to its historical significance, analytic simplicity and computational efficiency, which make it a good starting point for hybrid modeling.

In the present study, we employed the Adam optimizer (Kingma and Ba, 2014), a widely used variant of stochastic gradient descent, to find the optimal Gaussian plume parameters with respect to the given data subject to standard mean square error (MSE) loss.

2.3 Residual hybrid modeling approach

To enhance the predictions of the standalone physics-based model (PBM)—the Gaussian plume model—we opt for a hybrid modeling approach (e.g. Thompson and Kramer (1994)), i.e., we account for PBM error with a data-driven model (DDM) to better approximate the underlying data produced by the complex Lagrangian particle dispersion model (see Section 2.1). In particular, we employ a residual correction scheme—the DDM is trained to predict the one-step-ahead residual of the Gaussian plume model with respect to true data, using selected regressors such as concentration profiles at several lagged time instances and relevant meteorological parameters. At the deployment stage, the PBM first predicts the pollutant concentration fields at the next time instance, and then the DDM applies a correction.

Specifically, suppose we are given true ground-level pollutant concentration data in the form of a three-dimensional tensor $u_{i,j,k}$, where $i \in \{0, 1, \ldots, i_T - 1\}$ indexes consec-

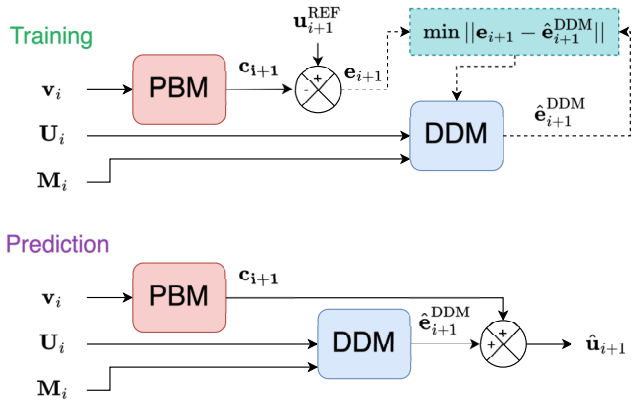

Fig. 3. A block diagram for training and prediction of the residual hybrid model presented in Section 2.3.

utive snapshots at times $t_i = i\Delta t$ up to the maximal sampling time $T = i_T \Delta t$, and $j, k \in \{0, 1, \ldots, N-1\}$ define the spatial locations on a regular $N \times N$ grid with node spacing Δx. We will also refer to $\mathbf{u}(t_i) = \{u_{i,j,k} \mid \forall j, \forall k\}$ as the concentration field at time t_i. The meteorological data, represented by a matrix $\chi_{i,l}$, consist of i_T L-dimensional vectors, one for each time level t_i and $l \in \{0, 1, \ldots, L-1\}$. Furthermore, we denote with $\mathbf{v}_i = (v_i, \phi_i)$ the wind speed and direction at each time instant.

At the training stage, we proceed as follows: first, we use (3) to generate the Gaussian plume profiles from $\mathbf{v}_i$ and store them in a tensor $c_{i,j,k}$; then, we compute the residual tensor $e_{i,j,k} = u_{i,j,k} - c_{i,j,k}$. Finally, the DDM is trained to minimize

$$\min_{\theta} \sum_{i} \frac{1}{i_T} \left\| \mathbf{e}(t_i) \; - \; \hat{\mathbf{e}}^{(\mathrm{DDM})}(t_i) \right\|, \tag{4}$$

where $\mathbf{e}(t_i) = \{e_{i,j,k} \mid \forall j, \forall k\}$ is the field of residuals at time level t_i and $\| \cdot \|$ denotes the chosen metric used to quantify the discrepancy between the predicted and observed states. The DDM estimate at each time level

$$\hat{\mathbf{e}}^{(\mathrm{DDM})}(t_i) = \mathrm{DDM}(U_i, \; M_i, \; \theta) \tag{5}$$

is a vector function of the DDM parameters θ, lagged concentration field values and meteorological data,

$$\begin{aligned} U_i &= \{\, u_{i',j,k} \;\mid\; i - m \le i' \le i - 1, \; 0 \le j, k < N \}, \\ M_i &= \{\, \chi_{i'',l} \;\mid\; i - n + 1 \le i'' \le i, \; 0 \le l < L \}, \end{aligned} \tag{6}$$

where n and m are the corresponding chosen numbers of lags. Each simulation step consists of first using a PBM to predict Gaussian plume field $\mathbf{c}(t_i) = \{\, c_{i,j,k} \mid \forall j, \forall k\}$ at time t_i for all grid points and then correcting it with the DDM correction field $\hat{\mathbf{e}}(t_i)$ to arrive at the estimate of the true pollutant concentration field

$$\hat{\mathbf{u}}(t_i) = \mathbf{c}(t_i) + \hat{\mathbf{e}}(t_i). \tag{7}$$

Both training and prediction procedures are visually represented with block diagrams in Figure 3.

2.4 Data-driven model - A Deep neural network

Our choice for the DDM is a deep neural network (DNN), which is known to be a universal function approximator (Cybenko, 1989). For this reason, deep neural networks are widely used across different fields of academic research and industry, as they can nowadays be trained on GPUs in personal computers. In our case, a DNN is a multi-layer feed-forward network that can be mathematically represented by a function $\mathbf{g} : \mathbb{R}^p \to \mathbb{R}^q$ which maps an input vector from the feature space $\mathbb{R}^p$ through a series of hidden layers to a target space $\mathbb{R}^q$. In other words, it is a composition of layer functions $\mathbf{g}^{(\mu)}(\mathbf{x}) = \sigma^{(\mu)}\big(W^{(\mu)}\,\mathbf{x} + \mathbf{b}^{(\mu)}\big)$, where $W^{(\mu)}$ and $\mathbf{b}^{(\mu)}$ denote the trainable weights and biases of the μ-th layer, respectively, and $\sigma^{(\mu)}$ is the corresponding nonlinear activation function. The DNN parameters θ are determined by Adam stochastic gradient descent (Kingma and Ba, 2014) to minimize the expression (4) where associated metric $\| \cdot \|$ is chosen to be the mean sum of all matrix elements squared $\|\alpha\|^2 = \frac{1}{N^2} \sum_{j,k} \alpha_{j,k} \alpha_{j,k}$. For later use, we also define

$$\|\alpha\|_2^2 = \sqrt{\frac{1}{N^2} \sum_{j,k} \alpha_{j,k} \alpha_{j,k}}. \tag{8}$$

3. WORKFLOW AND RESULTS

In this section, we detail the specifics of the workflow, primarily based on the methods discussed above, subject to one important caveat introduced here. The data presented in Section 2.1 consists of 52588 time samples and is divided into the first 41588 training samples, followed by 10000 validation and 1000 test samples. This dataset exhibits high variability in terms of the average pollutant concentration. In particular, the average values of $\mathbf{u}(t_i)$ can fluctuate by more than three orders of magnitude. Through practical experiments, we found it beneficial to model both the scale of the data and its spatial distribution separately. In particular, we proceed as follows. We start by computing the normalization factors $\beta(t_i) = \frac{1}{N^2} \sum_{j,k} u_{i,j,k}$ and define the normalized concentration profiles $\tilde{\mathbf{u}}(t_i)$ through the relation $\mathbf{u}(t_i) = \beta(t_i)\tilde{\mathbf{u}}(t_i)$. The first model we construct is purely data-driven, i.e., a deep neural network that is trained to predict the normalization factor $\beta(t_i)$ based on the lagged meteorological data and previously observed normalization factors. Concretely, let M_i be the lagged meteorological data as defined in equation (6) and further define the lagged normalization factors as $B_i = \big[\beta(t_{i-m}), \, \beta(t_{i-m+1}), \, \ldots, \, \beta(t_{i-1})\big]$. The first model is then defined by

$$\hat{\beta}(t_i) \;=\; \mathrm{DNN}_1(B_i, \, M_i; \, \theta_1). \tag{9}$$

The second model, that is independent of the first, is conceptually the same as the one presented in Section 2.3, i.e., it is expressed by equation (7). The only two subtle differences are, firstly, the part of the model that is used to correct the residual (equation (5)) is a function of lagged normalized concentration fields, i.e.,

$$\hat{\tilde{\mathbf{e}}}^{(\mathrm{DDM})}(t_i) = \mathrm{DDM}(\tilde{U}_i,\ M_i,\ \theta) := \mathrm{DNN}_2(\tilde{U}_i,\ M_i;\ \theta_2) \tag{10}$$

where $\tilde{U}_i$ is defined analogously to U_i in equation (6), using normalized concentration field values $\tilde{\mathbf{u}}(t_i)$. Secondly, we use a modified Gaussian plume model

$$\begin{aligned} \tilde{C}^{(x)}(x,y) =& \frac{Q}{\pi\sigma_y\sigma_z} e^{-\frac{y^2}{2\sigma_y^2}} e^{-\frac{H^2}{2\sigma_z^2}} \\ \tilde{C}(x,y,\phi) =& \tilde{C}^{(x)}\big(x',y'\big) \end{aligned} \tag{11}$$

that is independent of the wind speed v that acts only as a scaling factor in equation (1). The parameters of the model (11) are determined with respect to the normalized concentration field values $\tilde{\mathbf{u}}(t_i)$ as discussed in Section 2.2. Analogously to equation (7) we define the full second model as

$$\hat{\tilde{\mathbf{u}}}(t_i) = \big[\tilde{\mathbf{c}}(t_i) + \hat{\tilde{\mathbf{e}}}(t_i)\big]_+, \tag{12}$$

with $\tilde{\mathbf{c}}(t_i) = \{\, \tilde{c}_{i,j,k} \mid \forall j, \forall k\}$ where $\tilde{c}_{i,j,k}$ is a tensor of modified Gaussian plume profiles $\tilde{c}_{i,j,k}$ (notation congruent with $c_{i,j,k}$). Since the DDM correction might result in nonphysical negative values, we perform an extra step of element-wise ReLU operation: $[x]_+ = \max(x, 0)$, i.e., we correct all negative field values to zero. The combined model can now be expressed as a product of the outputs of both models as

$$\hat{\mathbf{u}}(t_i) = \hat{\beta}(t_i)\hat{\tilde{\mathbf{u}}}(t_i). \tag{13}$$

Given that all components of the model are computationally efficient, the simulation over the test set can be performed within seconds.

3.1 Model parameter selection

To obtain the most suitable PBM for use in the hybrid scheme, we first identified Gaussian plume parameters that best fit the data. The initial plume parameters, which were later optimized using the Adam gradient descent algorithm (Kingma and Ba, 2014), were determined by analyzing the profiles generated under various parameter configurations and atmospheric conditions. Based on visual inspection, we found that the data was best represented by Gaussian plume profiles corresponding to moderately unstable atmospheric conditions, as characterized by Pasquill (1961). Since wind speed and direction were available at two altitudes in the dataset, we introduced four additional parameters to define a linear combination of these inputs. All four parameters were initialized to 0.5 prior to optimization.

Regarding the data-driven part, the minimal number of lagged concentration field values and meteorological data, i.e., m and n (see equation (6)), was previously determined in a study using the same dataset (Kocijan et al., 2022). Testing the model across different combinations of m and n, we obtained the best results for $m = n = 5$. Independently, we identified the best-performing deep neural network structure for the prediction of $\hat{\tilde{\mathbf{u}}}(t_i)$ via Bayesian optimization (O'Malley et al., 2019), optimizing hyperparameters such as the number of layers, number of neurons per layer, dropout rates, activation functions, learning rate, and others. However, we observed experimentally that the training score only weakly correlated with the simulation performance on the validation data. Additionally, there appeared to be non-negligible variation in performance between repeated training runs using the same architecture. Nevertheless, we found that network structures with four hidden layers, each comprising 512 neurons, leaky-ReLU activation with a negative slope of 0.03, and uniform dropout rates of 0.01 across all layers, yielded consistent performance. The results suggest that, within this class of architectures, the specific choice of structure did not substantially affect the final simulation outcomes. For predicting the normalization factors $\hat{\beta}(t_i)$, we adopted an analogous DNN architecture with 256 neurons per layer and leaky-ReLU activations using a negative slope of 0.02.

It should be noted that we also experimented with more elaborate DNN architectures—for example, architectures in which the lagged field values U_i and the lagged meteorological data M_i were processed separately in the initial layers and then merged in the final layers. We also tested convolutional DNNs, which are theoretically more suited to the spatial structure of the data. Furthermore, we investigated a completely different DDM—the XGBoost model (XGBoost et al., 2016)—which is known to perform well on temporal data. Unfortunately, these alternative approaches did not yield any significant improvement over the selected DNN architecture.

3.2 Results

To assess the performance of our model, we primarily used the normalized root-mean-squared error (NRMSE), a standard cost function. Specifically, for our estimate $\hat{\mathbf{a}}(t_i)$ of a target field $\mathbf{a}(t_i)$ at time t_i we define $\mathrm{NRMSE}(\hat{\mathbf{a}}(t_i))$ as the root-mean-squared error (RMSE) computed at each time level and normalized by the amplitude of the true field at that time

$$\mathrm{NRMSE}(\hat{\mathbf{a}}(t_i)) = \frac{\|\mathbf{a}(t_i) - \hat{\mathbf{a}}(t_i)\|_2^2}{\|\mathbf{a}(t_i)\|_2^2} \tag{14}$$

The definition of $\|\cdot\|_2^2$ is given in equation (8). Value of $\mathrm{NRMSE}(\hat{\mathbf{a}}(t_i))$ at a time level t_i represents the relative discrepancy between the predicted and true fields; values close to zero indicate high accuracy, while larger values correspond to greater prediction errors. The $\mathrm{NRMSE}(\hat{\mathbf{u}}(t_i))$ for the 1000 simulated time steps of the test set is shown in Figure 4A. Its value ranges from approximately 0.25 to 1, with an average of about 0.554, but remains stable even for longer simulation times. Although the performance of the employed hybrid model judged solely by $\mathrm{NRMSE}(\hat{\mathbf{u}}(t_i))$ is not ideal, it still substantially outperforms the standalone physics-based Gaussian plume model, whose NRMSE is depicted in Figure 4B. Furthermore, a closer inspection of the visual pollutant-concentration profiles offers additional insight into the simulated data. In Figure 5, two typical simulation scenarios are shown. Under steady winds, the cross-sections consistently resemble a Gaussian plume, but during episodes of weak or rapidly shifting wind, the field becomes noticeably less regular. In such situations, our

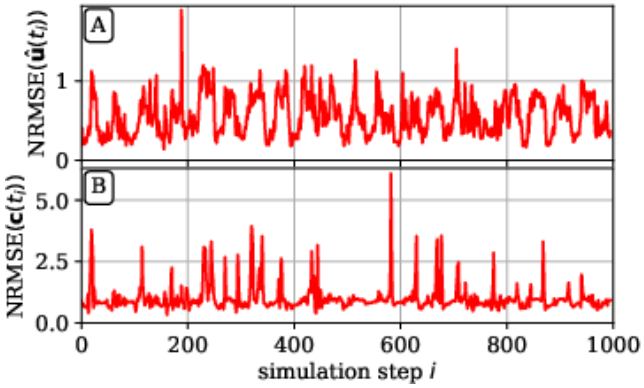

Fig. 4. Model simulation errors. A: The NRMSE of the simulated pollutant concentration fields $\hat{\mathbf{u}}(t_i)$, B: The NRMSE of the standalone physics-based Gaussian plume model prediction (see Section 2.2).

hybrid model substantially improves the predicted pollutant concentration profile. Unpredictable profiles also arise in the case of small wind speeds also indicated by the noticeable negative correlation observed between NRMSE($\hat{\mathbf{u}}(t_i)$) and the scalar wind speed [1] . In such cases, the point-wise fluctuations of the wind vector field across the grid are too large to be represented adequately by a single speed–direction pair. Consequently, such irregular samples cover larger portions of state space (i.e., the set of all possible pollutant profile configurations $\mathbb{R}^{N^2}$) more sparsely, making it more challenging for a DNN to accurately tune its free parameters to produce good results for such cases.

The intricate meteorology of the Velenje basin yields a highly unbalanced training set, leading to a mean spatial pollutant field that is far from uniform (Figure 6B). This imbalance is reflected in the temporally averaged residuals of the simulation (Figure 6A), where zones of small errors correspond to regions that were frequently polluted in the training data, and large errors occur where observations were sparse. Ultimately, the observed mean pollutant pattern mirrors the local wind–direction climatology: a persistent northerly flow dominates the higher layers (Figure 6D), whereas channeling by valley topography favors a westerly wind near the surface (Figure 6C). Because wind is measured only at the power-plant site, this near-surface westerly flow may differ substantially elsewhere, which helps explain the larger time-averaged errors observed to the east of the plant (Figure 6A).

Our model's performance-both the modeling accuracy and simulation time-is comparable to previous similar studies such as (Kocijan et al., 2022) that employ a purely data-driven approach, which were designed to reduce the computational cost of the complex physical model introduced by Mlakar et al. (2015) and Mlakar et al. (2019). While also preserving low computational cost, the main novelty of our approach lies in its partial reliance on the PBM, which confers additional interpretability.

[1] Pearson coefficient roughly -0.44 for NRMSE($\hat{\mathbf{u}}(t_i)$) and wind speed data at 500m (see Section 2.1).

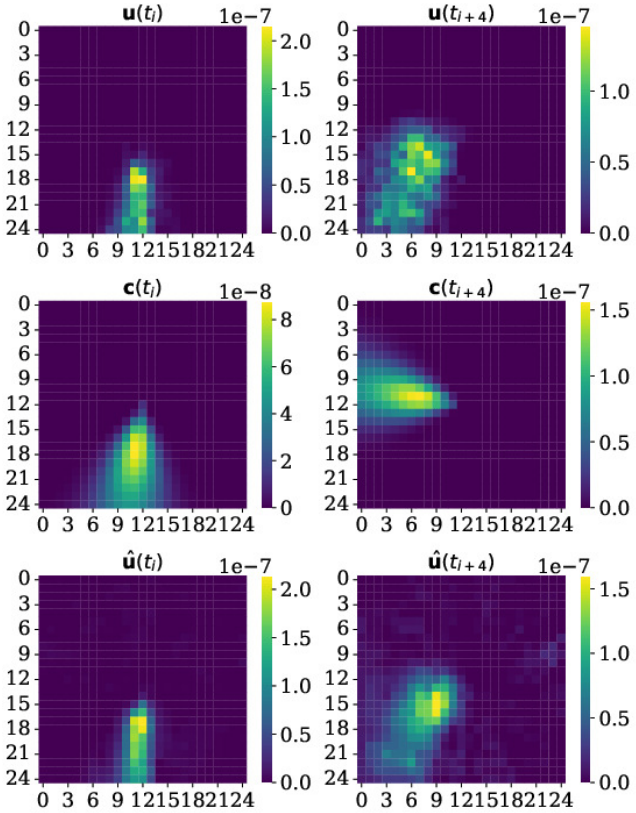

Fig. 5. Two instances of simulation scenarios. Each of 25 by 25 grid cells is proportional to the area of 0.36 km^2, i.e., together they cover the area marked with a square on Figure 1. On the left-hand side a typical situation for a steady wind conditions is shown, corresponding to simulation step 195 (see Fig. 4)-true field $\mathbf{u}(t_i)$ on top, the Gaussian plume prediction $\mathbf{c}(t_i)$ in the middle, and the simulated field $\hat{\mathbf{u}}(t_i)$ bellow, all in units of (s/m^3), as discussed in Section 2.1. The corresponding plots for pollutant concentration profiles after two hours (4 discrete time steps of 0.5h), when the wind direction is changing, are shown on the right-hand side. In this scenario, our model $\hat{\mathbf{u}}(t_{i+4})$ presents a substantial improvement over the baseline PBM (Gaussian plume) prediction $\mathbf{c}(t_{i+4})$.

4. CONCLUSION

In this paper, we propose a method for improving the physics-based prediction of atmospheric pollutant spread—namely, the Gaussian plume model Pasquill (1961) with a data-driven component, a deep neural network, embedded in a residual hybrid-modeling framework. We demonstrate the efficacy of our approach on ground-level sulfur-dioxide (SO_2) measurements from Slovenia's Velenje basin, where the Šoštanj Thermal Power Plant is the principal emission source. The resulting hybrid model enables rapid simulations of SO_2 dispersion based on a small set of meteorological variables. Our model substantially improves upon the standalone Gaussian plume model, which is ill-suited to the complex topography of the Velenje basin. Whereas the Gaussian plume formulation remains semi-applicable under stable atmospheric condi-

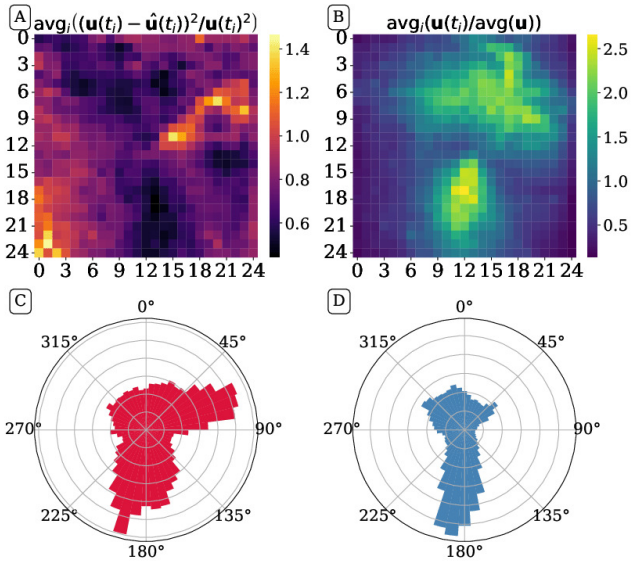

Fig. 6. A: The spatial profile of the temporally averaged residuals of the simulated test data over the terrain depicted in Figure 1. B: The time-averaged pollutant concentration per grid cell, normalized by the average concentration **u** over all cells and times in the training set. C, D: Wind direction distribution at 10m (C) and 500m (D) at the location of Šoštanj thermal power plant. The wind directions are expressed in degrees, and 0° corresponds to north. The area of each bin is proportional to the number of samples with wind pointing in the respective direction, weighted by their corresponding speeds.

tions, it fails when wind speed and direction exhibit strong gradients. In contrast, the hybrid model captures such transients well; although its accuracy varies, the simulation remains stable even over extended time windows.

The main contribution of this study is a novel architecture for predicting atmospheric pollutant spread that combines physics-based and data-driven elements in a hybrid modeling scheme. Specifically, the one-step-ahead prediction residuals of the physics-based model are corrected with two data-driven models, one accounting for the scale and the other for the spatial distribution of the residuals. Our method is transferable to other pollutant dispersion scenarios, including different terrains and types of dispersed materials, provided the model is retrained on relevant data.

Future work will include further modifications to the model structure, especially extending the physics-based component to better capture complex physical phenomena, as well as rebalancing the training data to enhance simulation accuracy.

Acknowledgements The authors acknowledge the research core funding No. P2-0001, project funding "Atmosphere Identification for Protection of Population in Preparation for Accidental Releases - MARIONETTE", ID L2-60149, and Ph.D. grant funding for Aljaž Pavšek, which were all financially supported by the Slovenian Research and Innovation Agency. We also acknowledge MEIS company, Slovenia, for providing us with pollutant-spread data.

REFERENCES

Castelli, S.T., Armand, P., Tinarelli, G., Duchenne, C., and Nibart, M. (2018). Validation of a Lagrangian particle dispersion model with wind tunnel and field experiments in urban environment. *Atmospheric environment*, 193, 273–289.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4), 303–314.

European Commission (2020). Major accident hazards: the Seveso directive – technological disaster risk reduction. https://ec.europa.eu/environment/seveso/. (Accessed 21 April 2020).

Finardi, S., Morselli, M.G., Brusasca, G., and Tinarelli, G. (1997). A 2–D meteorological pre–processor for real–time 3–D ATD models. *International journal of environment and pollution*, 8(3-6), 478–488.

Finardi, S., Tinarelli, G., Faggian, P., and Brusasca, G. (1998). Evaluation of different wind field modeling techniques for wind energy applications over complex topography. *Journal of wind engineering and industrial aerodynamics*, 74, 283–294.

Kingma, D. and Ba, J. (2014). Adam: A method for stochastic optimization. *International Conference on Learning Representations*.

Kocijan, J., Hvala, N., Perne, M., Mlakar, P., Grašič, B., and Božnar, M.Z. (2022). Surrogate modelling for the forecast of Seveso–type atmospheric pollutant dispersion. *Stochastic environmental research and risk assessment*, 1–16.

Mlakar, P., Božnar, M.Z., and Grašič, B. (2019). Relative doses instead of relative concentrations for the determination of the consequences of the radiological atmospheric releases. *Journal of environmental radioactivity*, 196, 1–8.

Mlakar, P., Božnar, M.Z., Grašič, B., Brusasca, G., Tinarelli, G., Morselli, M.G., and Finardi, S. (2015). Air pollution dispersion models validation dataset from complex terrain in Šoštanj. *International journal of environment and pollution*, 57(3-4), 227–237.

O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al. (2019). Kerastuner, github.

Pasquill, F. (1961). The estimation of the dispersion of windborne material. *Meteoro. Mag.*, 90, 20–49.

Snoun, H., Krichen, M., and Chérif, H. (2023). A comprehensive review of gaussian atmospheric dispersion models: current usage and future perspectives. *Euro-mediterranean journal for environmental integration*, 8(1), 219–242.

Sutton, O.G. (1932). A theory of eddy diffusion in the atmosphere. *Proceedings of the royal society of London. Series A, Containing papers of a mathematical and physical character*, 135(826), 143–165.

Thompson, M.L. and Kramer, M.A. (1994). Modeling chemical processes using prior knowledge and neural networks. *AIChE Journal*, 40(8), 1328–1340.

Willard, J., Jia, X., Xu, S., Steinbach, M., and Kumar, V. (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM computing surveys*, 55(4), 1–37.

XGBoost, C.T.G.C. et al. (2016). A scalable tree boosting system. In *Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min*, 785–794.