

Surrogate tree ensemble model representing 2D population doses over complex terrain in the event of a radiological release into the air

Nadja Hvala^{a,*}, Primož Mlakar^b, Boštjan Grašič^b, Marija Zlata Božnar^b, Matija Perne^{a,c}, Juš Kocijan^{a,d}

^a Jožef Stefan Institute, Department of Systems and Control, Jamova cesta 39, SI-1000, Ljubljana, Slovenia

^b MEIS d.o.o., Mali Vrh pri Šmarju 78, SI-1293, Šmarje-Sap, Slovenia

^c Quantectum, prognoza potresov, d.o.o., Dunajska cesta 156, SI-1000, Ljubljana, Slovenia

^d University of Nova Gorica, Vipavska cesta 13, SI-5000, Nova Gorica, Slovenia

ARTICLE INFO

Keywords:

Atmospheric dispersion
Radiological release
Surrogate model
Tree ensembles
Complex terrain

ABSTRACT

Atmospheric dispersion models predict the dispersion of harmful substances in case of accidents at industrial facilities and nuclear power plants (NPPs). However, high computation time limits their usage in an emergency or long-term analyses. This paper reduces the computation time by designing a surrogate data-driven model using a grid of tree ensemble models as a surrogate for the physical model and meteorological station measurements as model regressors. Regression tree modelling provided information for selecting the most important variables for prediction, while model ensembles improved the prediction accuracy. The approach is tested for an NPP in complex terrain to predict spatial (2D) maps of population doses for 24 h after a radiological release. The average performance of 2D maps against the physical model is *SMSE* (Standardized Mean Square Error) < 0.5 and *FMS* (Figure of Merit in Space) > 0.5. The designed model performs very well in predicting the long-term mean and 95th percentile of population doses. The main shortcoming is the underestimation of very high doses. Performance is expected to be further improved by selecting training data using pattern selection techniques and potentially by alternative machine learning algorithms or interconnected models, which we intend to apply in future work.

1. Introduction

Releases of chemical substances due to accidents in industrial facilities and nuclear power plants (NPPs) pose a threat to people and the environment. Some extreme cases of harmful releases into the atmosphere are, for example, the gas leak incident at the pesticide plant in Bhopal, India, or the release of atmospheric radiation in the Fukushima nuclear disaster, to name but a few. Therefore, measures to prevent accidents and ensure adequate preparedness and response are needed to reduce the associated risks.

For nuclear power plants, a model to be used in case of an accidental radiological release is designed in advance. Such models include complex physical phenomena based on atmospheric dispersion and meteorological models. They are computationally intensive, and even for a basic set-up, they require a lot of time and effort for the design. In more advanced cases, model predictions are run online to ensure better preparedness and improve emergency response measures. An example of

such an online Environmental Information System (EIS) for the Krško nuclear power system is described further in this paper. It includes several building blocks for the prognosis of air pollution dispersion for up to 7 days in advance after the incident. It consists of a detailed weather forecast based on the WRF (Weather Research and Forecasting) model, prognostic dispersion coefficients (*X/Q*) or relative concentrations calculation based on the MINERVE/SURFPRO and SPRAY (Lagrangian particle air pollution dispersion model), and the DOZE program for the calculation of population doses (Mlakar et al., 2019a). Such a system requires a lot of effort, i.e. considerable time and appropriately trained personnel for the online monitoring and maintenance of the system. Therefore, to reduce the associated effort and costs, the idea presented in this paper is to prepare in advance a surrogate model for predicting a spatial map of relative radiation doses. In this case, the surrogate model is pre-trained to represent the overall model predictions under different meteorological conditions. Therefore, in the event of an accident, only current meteorological data from SODAR

* Corresponding author.

E-mail address: nadja.hvala@ijs.si (N. Hvala).

<https://doi.org/10.1016/j.pnucene.2023.104594>

Received 12 July 2022; Received in revised form 9 January 2023; Accepted 25 January 2023

0149-1970/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

(sonic detection and ranging meteorological instrument) or LIDAR (light detection and ranging for meteorological applications) or RASS (Radio Acoustic Sounding System) systems and their predictions are needed for the surrogate model to predict the dispersion of harmful substances in the atmosphere (air).

Surrogate modelling is a well-known approach in engineering problems for replacing expensive full-scale high-fidelity simulations by approximating their input-output responses (Jiang et al., 2020; Alizadeh et al., 2020). Surrogate models of atmospheric dispersion models, also called meta-models or model emulation, have been already developed for different purposes and with different modelling approaches. They are most often used for the uncertainty studies of plume prediction. As described by Le et al. (2021), the uncertainties can originate from different sources, i.e. an emission source, meteorological variables, physical model parameters and model approximations. Surrogate modelling was also used for computationally efficient surrogate-based optimization of physical model parameters, where a surrogate model was constructed between model parameters and the physical model output (Le et al., 2019). Another application is in Bayesian inverse modelling (Lucas et al., 2017), where a surrogate model was used to replace weather prediction and dispersion models when relating the model output and field measurements for the estimation of a source term of a nuclear power plant release.

Surrogate modelling techniques are based on different methods. Some of the most often used methods include Response Surface Methods, Radial Basis Functions, Support Vector Machines, Gaussian Process Models, Artificial Neural Networks and Ensemble learning methods (Jiang et al., 2020; Alizadeh et al., 2020). For atmospheric dispersion models, surrogate modelling is very often based on artificial neural networks (ANNs). One of the very first applications of using ANNs for the representation of air pollution dispersion for point location (i.e. 1D) is presented in Boznar et al. (1993), where a new method using ANN was developed for short-term SO₂ prediction around the biggest Slovenian thermal power plant. The method showed promising results and initiated a 30-year-long work done by several scientific groups, dealing also with other and more challenging two spatial dimensions (2D) problems.

Carnevale et al. (2012) developed a surrogate model to optimize air quality measures to reduce the level of PM₁₀ at a regional scale. Therefore, they used ANNs to identify non-linear relations between control variables (emission control measures) and air quality indicators (pollution index) determined from deterministic physics-based model simulations. It was found that the surrogate model can reproduce the results of an air pollution model with a few percent error and requires less than one percent of their computational power.

Girard et al. (2016) used Gaussian process emulation to apply the Sobol' global sensitivity analysis for studying the relative influence of a set of uncertain model inputs and their interactions on the outputs of the atmospheric dispersion for the Fukushima nuclear accident. The emulators were evaluated in predicting the time- and space-aggregated gamma dose rates as well as time-integrated gamma dose rates at 64 measurement stations. It was concluded that local approximations are harder to obtain.

Mallet et al. (2018) constructed a meta-model that reproduces the main features of the air quality model ADMS-Urban for NO₂ and PM₁₀ simulation on an urban scale with street resolution and continuous releases from emission sources. The original model is a static model with low-dimensional inputs and high-dimensional outputs and has no temporal dimension in inputs and outputs. The outputs of the model are projected onto a reduced subspace derived from principal component analysis and then emulated using multiple linear regression and Kriging or Radial basis functions for interpolation between regression residuals.

Lauret et al. (2016) proposed cellular automata coupled with an ANN for forecasting atmospheric dispersion of methane (CH₄) in 2D over complex terrain. The dynamic model is based on wind field data and discretization of the advection-diffusion equation (first and second

concentration derivatives) to provide inputs for the ANN.

Desterro et al. (2020) used a Deep Rectifier Neural Network (DRNN) for 2D dose prediction in a Brazilian NPP. The approach was designed for short-term prediction up to 1 h after the accident and considering five model inputs (wind velocity, wind direction, position x, position y and time after the accident started). The DRNN achieved good accuracy and fast training time.

Gunawardena et al. (2021) proposed a machine-learning emulation to predict the spatial 2D deposition of radioactive materials. The surrogate model is used to emulate FLEXPART-WRF deposition maps obtained from the flexible particle Lagrangian dispersion model (FLEXPART) and meteorological fields generated from the WRF. The aim, in this case, is to obtain an ensemble of predictions due to the uncertainties of meteorological modelling, i.e. for different parametrizations of the WRF model. The surrogate model was designed as a 2D grid of hybrid linear and logistic regression models predicting deposition maps for 48 h after the release. In their case, the surrogate model needs to be first trained for each of the initial meteorological conditions to be able to produce an ensemble of predictions for different WRF parameterizations.

Mendil et al. (2022) developed a Deep Neural Network (DNN) surrogate model for spatial 2D representation of air pollution dispersion for *general cases in urban street canyons*. The DNN model was trained offline using synthetic data generated by a complex atmospheric transport and dispersion Micro-SWIFT-SPRAY model, similar to the model that we used in this study. The DNN model was then used to simulate the spread of hazardous pollution from different source locations. The approach is developed for complex urban environments simulating a 2-h period after the incident. The advantage of the proposed system is that it is using meteorological and topographical input data. Therefore, it can be transferred to similar urban terrains without special DNN model structure adaption.

This paper aims to design a data-driven surrogate model to represent air pollution dispersion for *general cases over non-urban complex terrain* with an NPP as a case study. The task is to emulate the full air-dispersion physical model and represent the 2D maps of radiological doses for 24 h after a hypothetical radiological release. This should be performed for different meteorological conditions using only measured or predicted meteorological station measurements as input data for the surrogate model predictions. This is a challenging task as doses due to the radionuclides in the air are calculated as the integral of the radionuclides concentrations over time multiplied by the time of exposure, in this case, 24 h. Therefore, doses are not simply proportional to a single 2D dispersion pattern at given meteorological conditions but to several consecutive patterns that may differ significantly one from another. That is a significant additional complexity compared to stationary conditions, which could be considered in short-term prediction as addressed, for example, in Desterro et al. (2020) or Mendil et al. (2022). The thus obtained surrogate model can be used for prompt decision-making in case of an accidental radiological release, to analyze different model options, or to perform long-term analyses, for example, evaluating the impact of climate change on the radiological release, which would be otherwise difficult due to high computation times. Data to identify a surrogate model is obtained by simulating a complete physical model. Also, the methodology used is based on a 2D grid of models similar to Gunawardena et al. (2021). However, the modelling in our case includes several specific features and novelties:

- The physical model output is a 2D map of relative doses received by the population at ground level. Hence an additional mapping of the radionuclides concentrations within a limited period of time to relative radiation doses is included in the surrogate model. The analysis is limited to the ground level since people spend most of their time at ground level (we do not analyze skyscraper situations). At the ground level, the modelling is also more complex and challenging because of the impact of the terrain and land use. However,

the methodology would be the same and could be applied to upper levels as well.

- The surrogate dispersion model is developed for the vicinity of the NPP where concentrations are the highest and the computation time is critical since the radioactive plume can disperse in a short time, i.e. within a few hours. Also, in this case, the methodology could be applied for larger distances as well.
- Model predictions are performed for complex terrain with flat and hilly regions using only measured and predicted meteorological station data as surrogate model input regressors.
- Surrogate modelling is based on ensembles of regression tree models. Ensemble modelling is one of the commonly listed modelling techniques, which have favourable properties for surrogate modelling (Alizadeh et al., 2020; Archetti and Candelieri, 2019). The applications in atmospheric dispersion modelling can be found in Lucas et al. (2017) for nuclear power source estimation and Ivatt and Evans (2020) for predicting the bias in tropospheric ozone prediction calculated by an atmospheric chemistry transport model.
- The surrogate model is tested for a wide variety of realistic annual meteorological conditions, and the performance statistics against the physical model are evaluated.

The work presented can also be seen as an attempt at data-driven dispersion modelling as an alternative to complex physical modelling, provided that input-output data are available, i.e. meteorological measurements and information on the process output. In the case presented, the output data were generated by the physical model but can also be obtained from measurement systems or, for example, 2D satellite images, if available for the pollutants of interest.

The paper is organized as follows. In the next section, we first present the atmospheric dispersion model and the case study of the Krško NPP, followed by the description of the main idea, the procedure of designing the surrogate model and the performance metrics used to evaluate the model. In Section 3, we analyze the model regressors and their time span and present some examples of radiation maps as well as statistical results of the model performance. The paper ends with conclusions describing the main results and perspectives for future work.

2. Materials and methods

2.1. The atmospheric dispersion model

The case study is the Krško NPP, Slovenia, for which the atmospheric dispersion model has been already developed and extensively tested (Mlakar et al., 2019a). The NPP is located in complex terrain with a flat area close to the NPP surrounded by hills reaching up to 450 m in height above the level of the NPP (Fig. 1). The problem to be addressed in this case is the prediction of relative radiation doses received by the population in the vicinity of the NPP after a possible accidental radiological release into the air.

A research environmental information system (R-EIS) for the Krško NPP predicts the relative radiation doses in the event of a radiological release from the NPP. The predictions are made for an area of 25 km × 25 km centred on the Krško NPP. The area is divided into 100 × 100 square cells with sides of 250 m, which proved to be the optimal solution in terms of accuracy of the physical model, and speed and feasibility of the overall modelling process. R-EIS is an upgrade of the online EIS described in Mlakar et al. (2019a). In R-EIS, relative doses (Mlakar et al., 2019b) are used instead of relative concentrations. The input data to R-EIS are meteorological measurements and their WRF forecasts on the premises of the NPP and its vicinity, including wind speed, wind direction, air temperature (all at ground level and vertical profiles) and global solar radiation. The output data of R-EIS are 2D fields of relative radiation doses at a ground level resulting from the radiological release into the air. They are computed by the overall physical transport and dispersion model, consisting of the following building blocks:

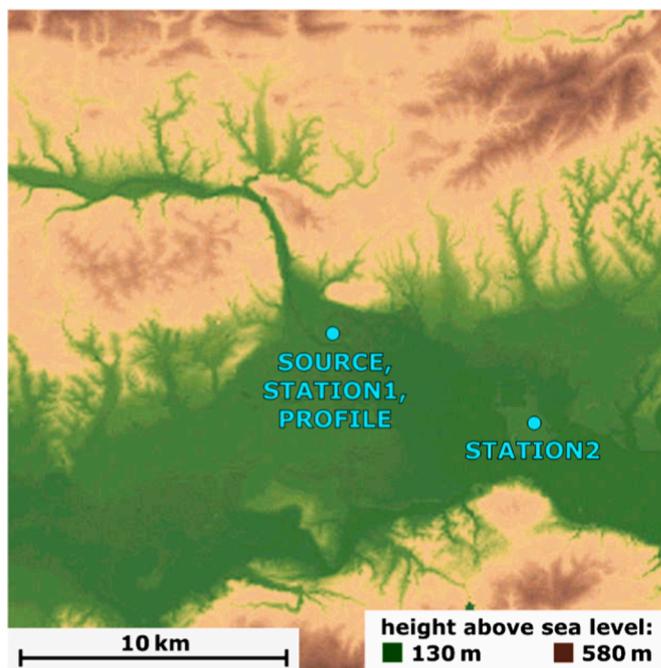


Fig. 1. The complex terrain in the vicinity of the Krško NPP (the source) with indicated locations of the ground-level stations and the vertical profile measurements used in this study. Source of terrain data: Public information of Slovenia, Surveying and Mapping Authority of the Republic of Slovenia.

- a three-dimensional (3D) weather reconstruction performed by the MINERVE and SURFPRO models,
- the Lagrangian particle air pollution dispersion model SPRAY,
- the population relative dose model applied in the DOZE software.

Detailed 3D weather forecast produces spatially and temporally varying meteorological fields with a spatial horizontal resolution of 2 km and a temporal resolution of half an hour for 7 days. It is based on the WRF weather forecast, which is performed once a day based on GFS (Global Forecast System) initial and boundary conditions. The detailed weather reconstruction is then obtained with the mass-consistent wind field model for complex terrain MINERVE (Desiato et al., 1998) and SURFPRO meteorological pre-processor (Arianet, 2011). The high-performance Lagrangian particle model SPRAY (Tinarelli et al., 2000) calculates the radionuclides dispersion based on a 3D numerical description of the atmosphere and a pre-defined hypothetical radionuclide emission from the NPP. The latter is normalized and has a pre-defined time-varying distribution of emission mass over time (Mlakar et al., 2019b). In the R-EIS, such an accidental radiological release is simulated to occur every hour. The Lagrangian particle model calculates accurate relative concentrations, also called “dilution coefficients” X/Q, and is automatically activated and updated every 30 min based on new meteorological data. In EIS (Mlakar et al., 2019a) relative concentrations are then used in the DOZE program to calculate the doses received by the population being exposed through the inhalation of a polluted atmosphere for 24 h after the radiological release. In R-EIS relative doses are calculated directly as a normalized measure of exposure. The WRF set-up was validated to produce a reliable weather prognosis for dispersion (Božnar et al., 2012a), while the air pollution dispersion model was validated nearby in the Šostanj location (Božnar et al., 2012b).

2.2. The main idea

The problem addressed in this paper is whether the described complex and computationally intensive atmospheric dispersion model, also

referred to as the physical model, can be represented by a surrogate data-driven model with reduced complexity and computation time. As shown in Fig. 2, the main idea is that the surrogate model represents the overall physical model behaviour using measured meteorological data and their WRF forecast as inputs and predicting the 2D relative dose maps for the population at ground level. In the surrogate model training phase, the input and output data of the physical model are generated for the training of the data-driven model. The surrogate model is then identified using machine learning methods such as ANN, Gaussian process models, regression trees, ensemble models, etc.

The design of the surrogate model, therefore, requires three main steps (Le et al., 2019):

- 1) Establishment of a design of experiment (DOE) for data collection.
- 2) Computing the response of the physical model for each experiment.
- 3) Training a data-driven model.

An important issue in the design of the surrogate model is to perform experiments with the original model to capture the entire variation space of inputs. There are several algorithms for generating DOEs, some very commonly used ones are, for example, factorial designs, grid search, random search, Latin Hypercube, etc. (Alizadeh et al., 2020). In atmospheric dispersion models, most often the Latin Hypercube sampling is used (Lucas et al., 2017; Mallet et al., 2018; Le et al., 2019). The problem is complicated in cases with a large number of variables, which is one of the biggest challenges. In the case of atmospheric dispersion modelling with meteorological variables as model inputs, an additional issue related to DOEs is the time-dependent perturbation of the amplitude and dynamics of the inputs, since they represent physical quantities with limited dynamical variations (Girard et al., 2020). To overcome these limitations in our study, the first two steps for the input-output data collection were pursued by performing a controlled long-term simulation experiment with the physical model as described in Section 2.3. Long-term modelling, based on two years of meteorological data, in this case, is a practical solution to capture a realistic variety of meteorological inputs into the physical model (we do not need to model cases that would not appear in nature). The main focus is then on the third step, i.e. training the data-driven model as described in Section 2.4.

2.3. Controlled simulation experiment for data collection

Data for the training and validation of the surrogate model was collected by performing a controlled simulation experiment with the physical model. The experiment resembled an actual system set-up at

the Krško NPP. In order to obtain a wide range of different meteorological conditions, meteorological data from July 2019 to July 2021 were used as inputs for the predictions of the physical model. They were obtained from a dedicated fine resolution (2 km, ½ h) weather forecast developed for the Krško NPP site and its surroundings (GFS data, 2023). The weather forecast is based on the WRF model (WRF, 2023) and was, in this case, used to provide information on meteorological data at two locations. The first is at the NPP site, where measurements from ground-level station 1 and RASS measurements with vertical profiles are available. The second site is located some distance from the NPP and provides measurements at ground-level station 2 (Fig. 1). Data from RASS include wind speed, wind direction and temperature at different heights, i.e. 40, 80, 120, 160, 200, 240, 300, 360, 420, and 500 m above the ground level. Data from ground-level stations include wind speed and wind direction at 10 m above the ground, the temperature at 2 m, and global solar radiation. A total of 38 meteorological variables were collected as physical model input variables sampled as half-hour mean values (every second sample was used). All other settings of the physical model were kept constant (DEM and CORINE, 2023). The physical model output data were relative radiation doses at ground level, generated with the full model structure as shown in Fig. 2. A radiological release was simulated to start every hour.

2.4. Surrogate model design

2.4.1. Model grid

The surrogate model for predicting relative radiation doses around the NPP was defined as a grid of models, i.e. a collection of 10,000 models, where each model represents one area in a spatial grid of 100×100 output areas. The individual (i, j) models in a grid, $i = 1, \dots, 100$, $j = 1, \dots, 100$, are defined as static models that map a subset of consecutive meteorological data to relative radiation doses in the corresponding areas of the output grid. For a radiological release initiated at t_k , the input regressors of the surrogate model are meteorological data m in the time interval $[t_{k-p}, \dots, t_{k-1}, t_k, t_{k+1}, \dots, t_{k+r}]$, where p and r denote the limits of the observed time interval relative to t_k . The outputs of the surrogate model $s_{i,j,k}$ that correspond to a radiological release at t_k should resemble the outputs of the physical model $y_{i,j,k}$. As already described, they represent the integrated relative radiation doses at ground level for 24 h after the release at t_k . Identification of the surrogate model, therefore, requires determining the surrogate model mappings $SM_{100 \times 100}$

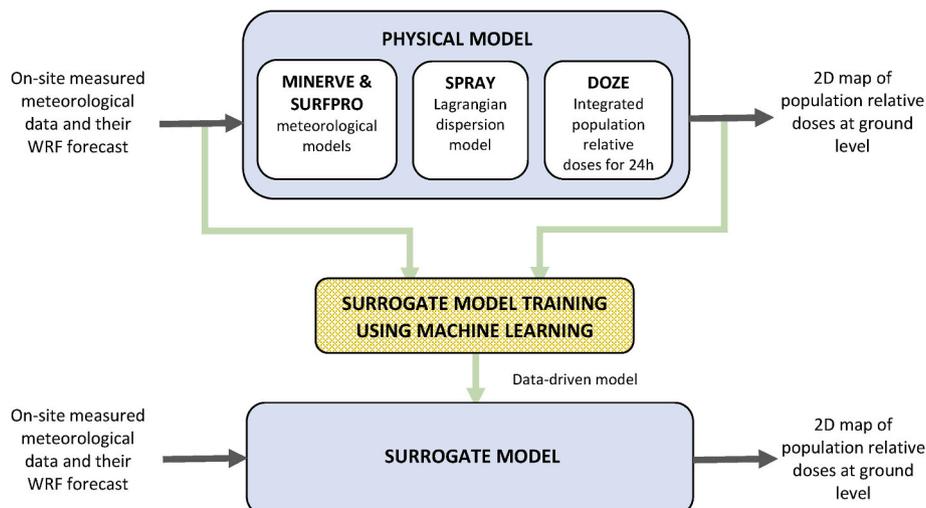


Fig. 2. Surrogate model design from measured meteorological data and physical model predictions using machine learning.

$$\begin{aligned}
 & (m_{1,k-p}, \dots, m_{1,k-1}, m_{1,k}, m_{1,k+1}, \dots, m_{1,k+r}, \\
 & m_{2,k-p}, \dots, m_{2,k-1}, m_{2,k}, m_{2,k+1}, \dots, m_{2,k+r}, \\
 & \dots \\
 & m_{M,k-p}, \dots, m_{M,k-1}, m_{M,k}, m_{M,k+1}, \dots, m_{M,k+r}) \\
 & \xrightarrow{SM_{100 \times 100}} \\
 & \begin{bmatrix} s_{1,1,k} & \dots & s_{1,100,k} \\ \vdots & \vdots & \vdots \\ s_{100,1,k} & \dots & s_{100,100,k} \end{bmatrix}, k = 1, \dots, N
 \end{aligned} \tag{1}$$

so that the difference between the surrogate model and physical model predictions of each model are minimized

$$\min \sum_{k=1}^N |s_{i,j,k} - y_{i,j,k}|, \text{ for each } (i,j) \text{ model, } i,j = 1, \dots, 100, \tag{2}$$

where M is the number of measured meteorological variables and N is the number of samples in the training dataset.

The tasks of the surrogate model design are to select the following:

- the machine learning algorithm for SM mappings,
- the most important regressors among the M meteorological variables,
- the time range of consecutive input meteorological data samples for output prediction, i.e. the time range of input meteorological variables as determined by the r and p parameters.

As presented in the following subsections, the regression tree ensemble was selected for the machine learning algorithm, while the selection of the most important regressors and their time range was performed by evaluating the variable importance. Note that at present, the training pattern selection consists only of the selection of the meteorologically reasonable time period used for model prediction (heuristic selection of several consecutive samples without missing samples, the same time period for all variables).

2.4.2. Regression trees

Regression trees are interpretable models, appropriate also for modelling nonlinear systems and systems with interacting inputs (Molnar, 2022). They are representatives of decision trees, i.e. hierarchical models of supervised learning in which the model classifies instances by querying them down the tree from the root to the leaf node (Abraham et al., 2020). Each node represents a test over an attribute, and each branch denotes its outcome. The root node is the beginning node, and the leaf nodes are the outcome. Decision trees can be used for both classification and regression problems. The outcomes of regression trees are continuous values, typically real numbers.

There are various algorithms that can grow a tree. One of the most common algorithms for creating decision trees is the CART (Classification And Regression Tree) algorithm using binary splits (Breiman et al., 1984). It is based on recursive partitioning of the regressor space with p inputs X_1, \dots, X_p into M distinct non-overlapping regions R_1, \dots, R_M . For a system with input-output data (x_i, y_i) with N observations, $i = 1, \dots, N$, where $x_i = (x_{i,1}, \dots, x_{i,p})$, the output at x is predicted as (Hastie et al., 2009)

$$\hat{y} = \hat{f}(x) = \sum_{m=1}^M c_m I(x \in R_m), \tag{3}$$

where $I(x \in R_m)$ is the identity function that returns 1 if x is in the subset R_m and 0 otherwise, and $c_m = \text{ave}(y_i | x_i \in R_m)$ is the average of all training instances in R_m .

2.4.3. Tree ensembles

Although decision trees are simple and intuitive, they also present some shortcomings. For example, the results might be unstable since a small change to the input data could alter the series of splits, which in

turn might result in a completely different prediction. Besides, individual decision trees tend to overfit. Therefore, *ensemble learning* methods have been used. They combine and average over multiple decision trees using the bagging (bootstrap aggregating) technique (Breiman, 1996). The idea is to bootstrap, i.e. random sample with replacement, B training sets from the original training data. For each bootstrap sample, a decision tree is grown and the B trees are then aggregated. For regression tasks, the mean or average prediction of the individual trees is returned

$$\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^b(x). \tag{4}$$

Combining the results of many decision trees reduces the effects of overfitting and improves generalization. Since all decision trees are trained on bootstrap samples of the original training data, the variance decreases as the number of trees B increases. A further extension is random forests (Breiman, 2001), which select a random subset of features to train every tree of the forest. This decreases the correlation between individual predictors. Both properties, i.e. the bagging and random feature selection, result in one of the most effective and computationally efficient methods in machine learning (Archetti and Candelieri, 2019).

2.4.4. Variable importance

An important property of decision trees is that they provide an inherent property of variable importance, which enables, in a computationally efficient way, the selection of a smaller number of important features. Similarly, the bagged decision trees allow for empirically accessing the variable importance based on the *permutation importance* measure introduced by Breiman (2001). To measure the importance of variable X_j , the idea is to permute all values of this variable, and the variable importance measure is defined as the decrease in prediction accuracy caused by the permutation. If the variable consists of purely random noise, the prediction accuracy will likely not be affected by permuting the values. Conversely, if the variable is associated with the response, the prediction accuracy decreases substantially.

Formally, the variable importance VI is computed as follows (Strobl et al., 2008; Hjerpe, 2016). Let β^t denote the out-of-bag samples, i.e. observations not contained in a bootstrap sample of tree t , with $t \in \{1, \dots, ntree\}$, and let L denote the prediction accuracy at the i -th training example computed from y_i and the output of the tree T_t . The importance of variable X_j in tree t is defined as

$$VI^{(t)}(X_j) = \sum_{i \in \beta^t} L(T_t(x_i), y_i) - L(T_t(x_{i,\pi_j}), y_i), \tag{5}$$

where $x_{i,\pi_j} = (x_{i,1}, \dots, x_{i,\pi_j(j)}, x_{i,j+1}, \dots, x_{i,p})$, and π_j is a random permutation of X_j . In regression, the prediction accuracy L is defined as the RMSE (root mean square error). The variable importance measure for X_j is computed as the mean importance over all trees in the ensemble

$$VI(X_j) = \frac{\sum_{t \in B} VI^{(t)}(X_j)}{ntree}. \tag{6}$$

2.4.5. Model implementation

The whole procedure of the surrogate model design consists of several steps as shown in Fig. 3.

The surrogate model was designed using Matlab (Mathworks, 2020) Regression Learner application that is part of the Statistics and Machine Learning toolbox. Before applying the machine learning algorithms, the collected input-output data were normalized using Z-score statistics (Matlab function *zscore*) to obtain the mean value 0 and standard deviation 1 for each model input and model output.

The Regression Learner enabled the comparison of the performance of different algorithms for SM mappings, e.g. linear regression, regression trees, support vector machines, Gaussian Process Regression and ensemble models. The best-performing algorithms with respect to the 4-

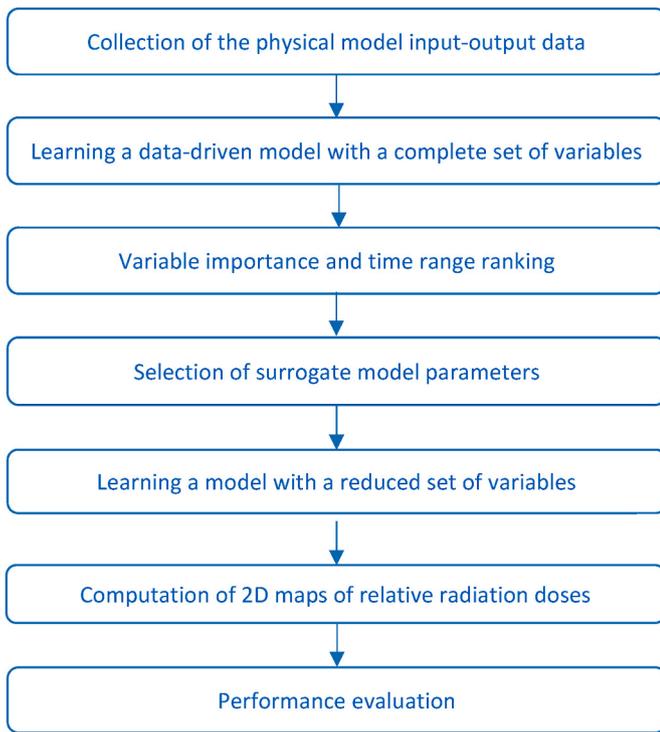


Fig. 3. The steps of the surrogate model design.

fold RMSE criterion in validation, computed from the outputs of the physical and surrogate model, were regression tree ensembles.

The machine learning algorithm chosen to grow a fitted regression tree ensemble was the Matlab function *TreeBagger*. It grows the decision trees in the ensemble using bootstrap samples of the data. Also, *TreeBagger* selects a random subset of predictors to use at each decision split as in the random forest algorithm. The adjusted parameters of the algorithm were the number of trees in the ensemble and the minimum number of observations per tree leaf. These two parameters were chosen experimentally by testing different settings and observing the out-of-bag error in model learning, calculated as the mean squared error (MSE) for the out-of-bag samples in the training data. A higher number of trees improves the model quality but also increases the computation time. A sufficiently high number of observations per tree leaf prevents over-training the model.

The selection of the most important regressors in the tree ensemble models was evaluated using the Matlab predictor importance function *OOBPredictorImportance*, which determines out-of-bag estimates of feature importance in the ensemble. The ten most important regressors of each (i, j) model were stored for the evaluation of the most important meteorological variables and their p and r parameters for the surrogate model prediction.

Finally, after the model training, the 2D maps of relative radiation doses were constructed and evaluated by different performance criteria.

2.5. Performance metrics

The following criteria were used for model evaluation:

- *Standardized Mean Square Error (SMSE)*

SMSE is the mean squared error of the true y_i and predicted \hat{y}_i values divided by the variance σ_y^2 of the true values

$$SMSE = \frac{1}{N} \sum_{i=1}^N \frac{(y_i - \hat{y}_i)^2}{\sigma_y^2}. \quad (7)$$

SMSE is dimensionless due to the standardisation on the variance. For a perfectly accurate prediction, it equals 0. A value of 1 indicates a naive model that predicts the mean of the true values. A value greater than 1 means predictions that are more erroneous than the prediction of the mean of the true values (Kocijan, 2016).

In the case of the surrogate model design, the true values correspond to the prediction of the physical model and predicted values correspond to the prediction of the surrogate model. The *SMSE* was used to evaluate the trained models. The evaluation criterion was the average *SMSE* of all 10,000 models for the training or test data. The *SMSE* was also used to evaluate 2D maps, i.e. the outputs of all 10,000 models in a given time instance t_k . The evaluation criterion was the average *SMSE* of the 2D maps for all time instances in the training or test data.

- *Figure of Merit in Space (FMS)*

FMS is a spatial error metric for evaluating the plume area in 2D maps. It is based on a Jaccard index, which is a statistic used for gauging the similarity and diversity of sample sets. When evaluating the predicted and actual plumes in dispersion, it is defined as the intersection of the area of the predicted A_p and actual A_o plumes divided by the union of the area of the predicted and actual plumes (Maurer et al., 2018)

$$FMS = \frac{A_p \cap A_o}{A_p \cup A_o}. \quad (8)$$

In our case, A_p and A_o correspond to the plume area predicted by the surrogate model and the physical model, respectively.

This metric depends only on the absence or presence of radiation dose, not the magnitude. It is implemented by counting the cells in 2D maps. It varies between 0 and 1. Values of 0.8 and above are generally considered good for atmospheric models (Gunawardena et al., 2021). Note that the *FMS* is not an absolute measure for assessing performance, as it depends on the threshold for relative radiation dose which needs to be selected. In our case, this value was determined as $2 \times 10^{-8} \text{ s/m}^3$.

- *Mean absolute percentage error of maximum prediction (MAPE_{max})*

The *MAPE_{max}* measure was used to evaluate the model accuracy in forecasting the maximum relative radiation dose. It was determined as the relative error between the maximum of the physical model and the maximum of the surrogate model in a 2D map. The final *MAPE_{max}* was calculated as the average for all samples in the training or test dataset

$$MAPE_{max} = \frac{1}{N} \sum_{k=1}^N \left| \frac{\max\{y_{i,j,k}\}_{i,j=1,\dots,100} - \max\{s_{i,j,k}\}_{i,j=1,\dots,100}}{\max\{y_{i,j,k}\}_{i,j=1,\dots,100}} \right| \times 100\%. \quad (9)$$

- *Mean, 95th percentile and maximum statistics*

We used these measures to statistically evaluate 2D maps obtained by the physical and surrogate models for a selected period of different meteorological conditions and to compare, point-by-point in a grid, each grid cell's mean, 95th percentile and maximum values of all available cases in a selected evaluation period.

3. Results

3.1. Training and validation data

Three-quarters of the data obtained in the controlled simulation

experiment were used to train the surrogate model and one-quarter for testing. This corresponds to approximately 13,150 and 4380 data samples in the training and test datasets, respectively. The number of samples with a relative radiation dose greater than 0 is smaller and depends on the terrain and the spread of the radiation plume under different meteorological conditions. Depending on the position in the grid, it ranged from 783 to 9589 samples, with an average value of 3384 samples. The highest relative radiation doses were between 3.68×10^{-7} and $1.74 \times 10^{-5} \text{ s/m}^3$, with an average maximum value of $2.6 \times 10^{-6} \text{ s/m}^3$. The unit for relative radiation doses is not the same as for doses (the unit for radiation doses is [Sv]) since we use the approach of normalized emission. This approach leads to the unit for relative radiation doses in $[\text{s/m}^3]$ and is derived in [Mlakar et al. \(2019b\)](#).

3.2. Learning the model with a complete set of variables

Training the grid of 100×100 models was initially performed for all meteorological variables as model input regressors and for the time interval -1 to 5 h with respect to the start of the radiological release (sampled every 1 h , i.e. seven consecutive samples). These limits were selected based on trial simulations, which showed that weather conditions before or far after the release do not significantly affect the performance of the model in the observed area around the NPP. Satisfactory model performance was obtained with 30 trees in the ensemble and 5 as a minimum number of observations per tree leaf. The so obtained set of models in the grid was used to optimize the input variables and their time range as well as tree ensemble parameters.

3.3. Importance of different meteorological variables

The importance of different meteorological variables as model regressors was evaluated by observing the ten most important regressors of the identified models. The frequency of each meteorological variable among the most important regressors is shown in [Fig. 4](#). It can be seen that wind direction is the most important for the relative radiation dose prediction. Very high importance is obtained for wind direction at both ground-level stations as well as for RASS measurements at lower heights. High importance is obtained also for global solar radiation. Wind speed has lower importance, but also in this case higher importance can be

seen for ground-level station measurements and RASS measurements at lower heights. Finally, the temperature has the lowest importance for the prediction of relative radiation doses.

The importance of variables was analysed also with regard to the position in the grid. For that purpose, the most important regressors of the identified set of models were grouped into four groups, i.e. wind direction, wind speed, temperature and solar radiation. [Fig. 5](#) shows 2D maps indicating whether the regressors from the four groups are among the most important regressors. It can be seen that wind direction is among the most important regressors of all grid models, followed by wind speed, solar radiation and temperature. If we compare the obtained results with the terrain in [Fig. 1](#), we can see that wind speed is more important in the hilly area and solar radiation in the flat area, while temperature does not seem to be very important for the model prediction.

The results show that although some meteorological variables have lower overall importance, they might be important for the relative radiation dose prediction in certain parts of the grid. Therefore, they should not be excluded from the set of regressors. They should be retained as model inputs if showing importance in certain grid areas. Based on these results, a smaller set of meteorological measurements was selected. Temperature signals have not been included as they do not show high importance. Solar radiation was included due to its overall and grid importance. Wind direction and wind speed signals were also included but at a smaller number of heights. A detailed analysis of individual variables showed that wind direction at lower heights is important for the prediction of relative radiation dose in the middle flat area of the grid, wind direction at higher heights and wind speed at all heights are important for the prediction in distant hilly areas. Finally, the reduced set of input regressors included 17 variables (see [Table 1](#)).

3.4. Time range of input meteorological data samples

The most appropriate time range of the meteorological variables was also determined by observing the time lag of the ten most important regressors of the identified set of tree ensemble models. [Fig. 6](#) (left) shows the percentage of regressors for each time lag. It can be seen that meteorological data within $0-3 \text{ h}$ with respect to the radiological release have the largest impact on the prediction of relative radiation doses.

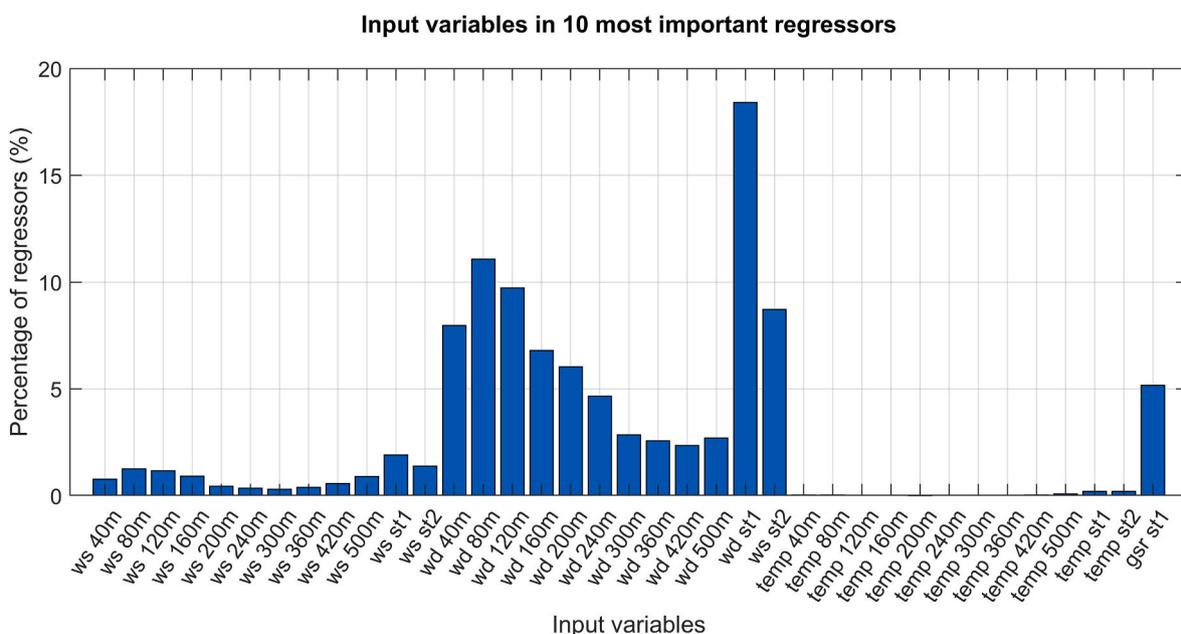


Fig. 4. Frequency of meteorological variables in the ten most important model regressors of the identified tree ensemble models (ws – wind speed, wd – wind direction, temp – temperature, gsr – global solar radiation, st1 – first station, st2 – second station).

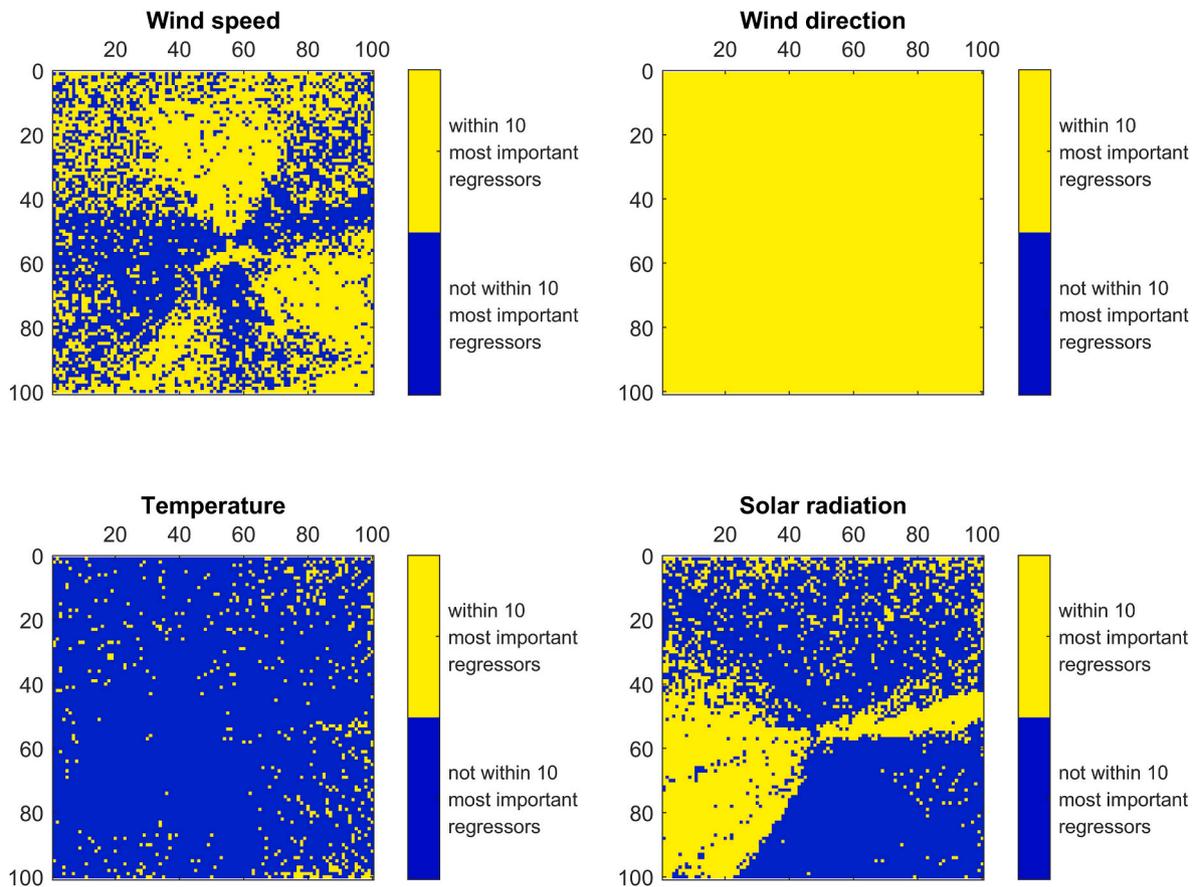


Fig. 5. The presence of different meteorological variables among the ten most important regressors of the identified tree ensemble models.

Table 1

Initial and finally selected surrogate model variables and their time period with respect to radiological release.

	Initial model variables	Important model variables
Wind speed	Time range $\{-1, 0, 1, 2, 3, 4, 5\}$ hours RASS at 40, 80, 120, 160, 200, 240, 300, 360, 420 and 500 m Ground-level stations 1 and 2	Time range $\{0, 1, 2, 3, 4\}$ hours RASS at 40, 80, 160, 200, 360 and 500 m Ground-level stations 1 and 2
Wind direction	RASS at 40, 80, 120, 160, 200, 240, 300, 360, 420 and 500 m Ground-level stations 1 and 2	RASS at 40, 80, 160, 200, 360 and 500 m Ground-level stations 1 and 2
Temperature	RASS at 40, 80, 120, 160, 200, 240, 300, 360, 420 and 500 m Ground-level stations 1 and 2	-
Global solar radiation	Ground-level station 1	Ground level-station 1

Fig. 6 (right) also shows the average time lag of the most important regressors of the individual models presented in a 2D map. It can be seen that the average time lag is smaller in the flat area closer to the radiation release, while it is larger in more distant and hilly areas. This is an expected result since long-term meteorological conditions with larger delay are important for the spread of the plume in the surrounding hilly area. Based on these results, the selected time lags of meteorological variables were $\{0, 1, 2, 3, 4\}$ hours for all variables.

3.5. Selection of tree ensemble parameters

The reduced set of model regressors includes 17 variables and 5 time lags for each variable (Table 1). Therefore, the total number of model inputs is 17×5 , i.e. 85 input signals. For this set of variables, different settings of tree ensemble parameters were evaluated to optimize the performance.

To select the appropriate number of trees in *TreeBagger*, a grid with a reduced number of models with $i, j \in \{5, 20, 35, 50, 65, 80, 95\}$ was

evaluated for different numbers of trees while observing the out-of-bag error in model learning. As can be seen in Fig. 7, a larger number of trees improves the performance but also increases the computation time for model learning. A value of 100 trees was selected for further model evaluation.

The number of observations per tree leaf was also evaluated. A sufficiently large number reduces the overfitting of the model on the training data, but a further increase gives a similar or lower performance. Based on the results in Fig. 7, the value of 10 was selected. Other parameters of *TreeBagger* were left at default values.

3.6. Learning a model with a reduced set of variables

For the chosen model structure and parameters, a grid of 100×100 models with a reduced number of variables was trained and evaluated using the *SMSE* metric. The *SMSE*, computed as the average of all models in the grid, was 0.337 and 0.647 for the training and test data, respectively. We can see that the performance is much better for the training

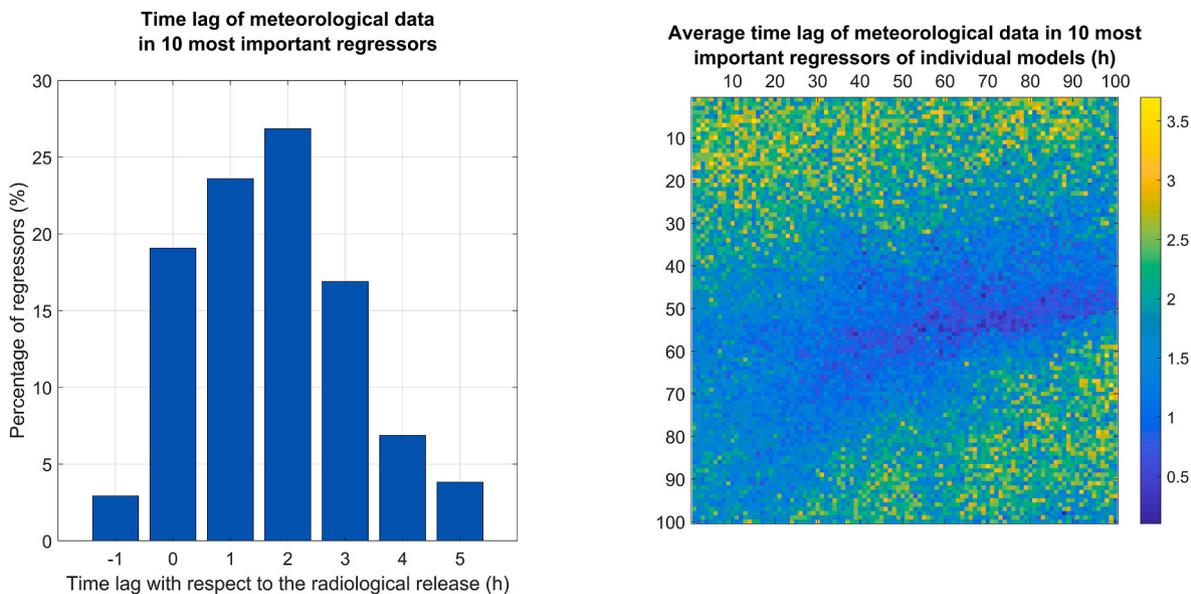


Fig. 6. Time lag of meteorological data in the ten most important model regressors of the identified tree ensemble models: percentage of most important regressors with a given time lag (left), average time lag of individual models (right).

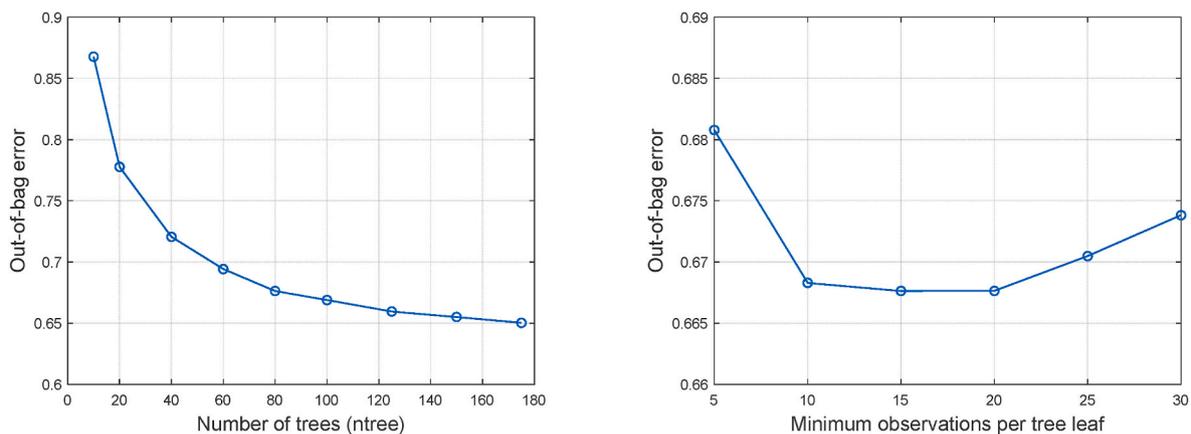


Fig. 7. Out-of-bag error for different numbers of trees and different numbers of observations per tree leaf.

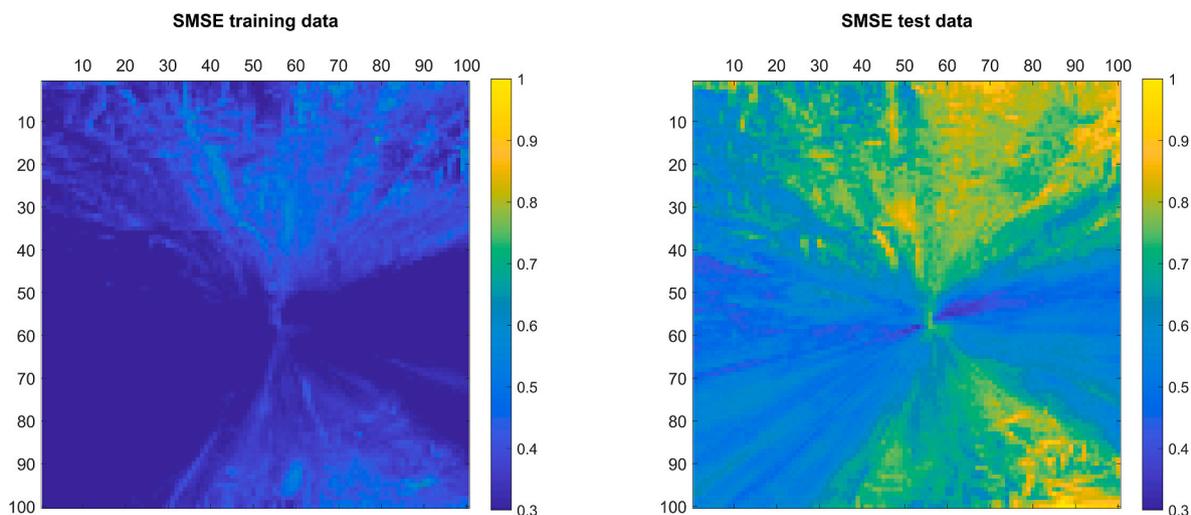


Fig. 8. The SMSE measures of the identified tree ensemble models for the training and test datasets presented in a grid.

data. However, also for the test data, the average performance is relatively good. It is significantly better than the naive prediction, which predicts average values and has an *SMSE* equal to 1.

Fig. 8 also shows the *SMSE* of individual models in the grid for the training and test data. It can be seen that individual models perform consistently better or worse in some parts of the observation area. With respect to the terrain around the NPP (Fig. 1), better performance is achieved in the flat area in the middle part, while worse performance is achieved in the hilly parts. A possible reason for the worse prediction in the hilly areas is the number of informative data available for training the model. Namely, the plume is less likely to spread over the hilly terrain. Therefore, in the available dataset for model training, relative radiation doses in the hilly areas are often zero. Fig. 9 shows the number of data points with relative radiation doses greater than zero. It can be seen that the number is lower in the hilly parts. Another possible reason could be also the meteorological conditions, which are more complex in the hilly parts. Therefore, it is more difficult here to model relative radiation doses with station measurements alone.

3.7. Examples of 2D maps of relative radiation doses

Using the grid of identified tree ensemble models, 2D radiation maps were constructed for the time instances in the training and test datasets. The performance of the trained models has been inspected for different meteorological conditions. A video showing 10 days of consecutive 2D maps for the test data is available in Appendix A (Supplementary material).

Figs. 10–13 show some examples of good model performance on test data. The meteorological parameters of the presented examples are shown in Table 2 for the two most influential measurements, i.e. RASS at 80 m and ground-level station 1. For better visualization, the scale is the same in all figures, i.e. between 10^{-8} and 10^{-6} s/m³. The upper limit was chosen because the area of very high relative radiation doses is usually small. It can be seen from the figures that most of the plume typically has a relative radiation dose of less than 0.35×10^{-7} s/m³ for a unit radiological release.

Supplementary video related to this article can be found at <https://doi.org/10.1016/j.pnucene.2023.104594>

Fig. 10 shows an example of a very good prediction of the surrogate model with small differences between the physical and surrogate

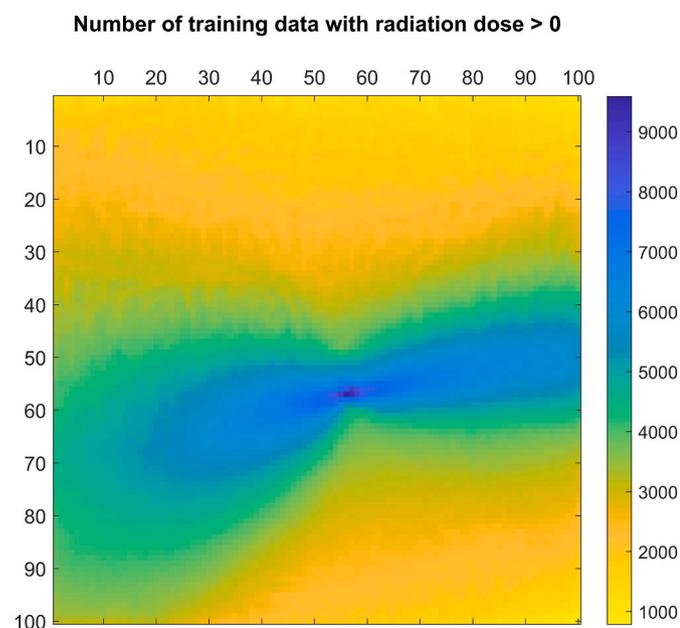


Fig. 9. Number of training data in a grid with radiation relative doses greater than zero.

models, and a good prediction of the plume. Such results are obtained for moderate or high wind speed and southwest wind direction. Similar good results are usually obtained also for northeast or east winds.

Fig. 11 shows an example of a northwest wind spreading the plume towards the southeast hilly area. Because of higher instability of the atmosphere and differences in wind direction, in this case, the spreading of the plume is wider. Nevertheless, the prediction of relative radiation doses is good, as confirmed by the low *SMSE* and high *FMS* values.

Fig. 12 shows a different case at low wind speed and nighttime conditions. In this case, the plume appears to spread over a much larger area. The prediction of relative radiation doses is less accurate. This can be seen from the larger *SMSE* value and the larger area of significant differences between the predictions of the physical model and the surrogate model. However, also in this more challenging case, the spread of the plume is well predicted.

Fig. 13 also shows an example of the high maximum relative radiation doses. In this case, the surrogate model also predicted the relative radiation doses quite well. However, during model training, it was observed that very high doses are underpredicted in both the training and test datasets.

This can be seen in Fig. 14, where the obtained grid of models was statistically evaluated with regard to mean, maximum and 95th percentile of relative radiation doses in comparison to the physical model predictions for the test data (around 4380 data in each grid cell). A comparison shows that the surrogate model is very reliable in predicting the mean relative radiation doses over a longer period of data. Also, the 95th percentile is well predicted although slightly lower, which indicates the underestimation of very high doses by the surrogate model. The greatest difference can be seen in maximum prediction. The maximum values of the surrogate model are low and around 3 times lower than those of the physical model. It could be concluded that the surrogate model is well predicting the long-term mean and 95th percentile statistics, while the maximum values are underpredicted.

3.8. Performance metrics of 2D maps

The average performance metrics of 2D maps are shown in Table 3. It can be seen that the performance on the training data is better than on the test data. The biggest difference can be observed for the *SMSE* measure, but its value of 0.470 for the 2D maps is significantly lower, i.e. better compared to 0.647 for the *SMSE* of the trained models on test data. This is because of the higher variance of the relative radiation doses in the spatial than in the temporal dimension. It suggests that while the prediction error of a single model in the grid may be relatively large, the significance of this error decreases when the predictions of all models in the grid at a certain time instance are considered. A visual inspection of the radiation maps showed that an *SMSE* of 0.5 or less gives a moderate or good radiation map. These results were obtained for 97% and 61% of training and test data, respectively. Better *SMSE* values are achieved at higher wind speeds, especially when measured at both ground-level stations.

The mean values of the *FMS* criterion were 0.639 and 0.547 for the training and test data, respectively. These values are relatively low compared to the 0.8 suggested by Gunawardena et al. (2021) for atmospheric models. In our case, 0.8 and above provided very good plume prediction (see Figs. 10–13, bottom left graph). It should be noted that, in our case, terrain and atmospheric conditions are challenging due to the basin and very often weak or no wind conditions. This is a difficult case for predicting atmospheric dispersion, as wind fluctuations and, in particular, wind direction could be almost random. Therefore, *FMS* values of 0.5 or higher are considered acceptable and represent a moderate or good radiation map. They were achieved for 81% and 61% of the training and test data, respectively. Better results were achieved at higher wind speeds and higher solar radiation.

Table 3 also shows the mean absolute percentage error $MAPE_{max}$ in predicting the maximum relative radiation doses, calculated as the

Table 2

Meteorological conditions at the time of hypothetical radiological release for RASS 80 m and Station 1 for the four cases of radiation maps presented in Figs. 10–13.

Case	Date	Meteorological and plume conditions	RASS at 80 m			Station 1		
			Wind speed	Wind direction	Temperature	Wind speed	Wind direction	Solar radiation
			m/s	°	°C	m/s	°	W/m ²
		min	0	0	-7.9	0.03	0	0
		max	20.40	360	32	14.36	360	975
		mean	3.96	166	12.52	2.68	164	176
1	June 05, 2021 10:30	Plume in the middle flat area	3.60	238	23.1	2.91	238	892
2	December 31, 2020 19:30	Plume in the southeast hilly area	4.00	272	3.5	3.06	317	0
3	March 11, 2021 00:30	Plume at low wind speed conditions	1.30	115	2.7	1.02	64	0
4	June 03, 2021 23:30	Plume with high maximum radiation dose	4.70	107	18.3	3.63	66	0

Test data, 05-Jun-2021 10:30:00

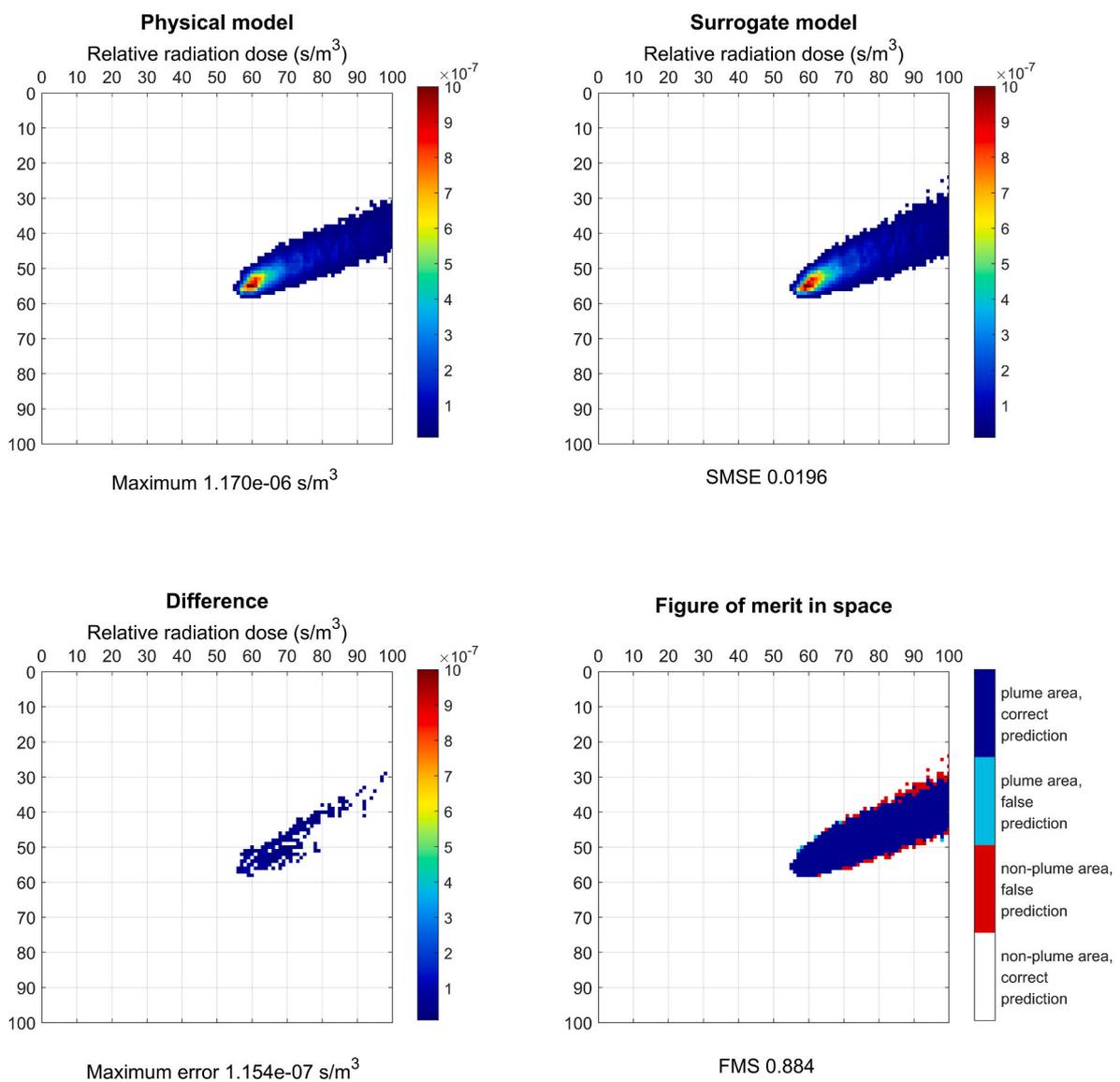


Fig. 10. Plume spreading in the flat middle area (case 1). Predicted relative radiation doses of the physical model (top left) and the surrogate model (top right), the differences between the two models (bottom left), and the figure of merit in space for the predicted plume (bottom right).

Test data, 31-Dec-2020 19:30:00

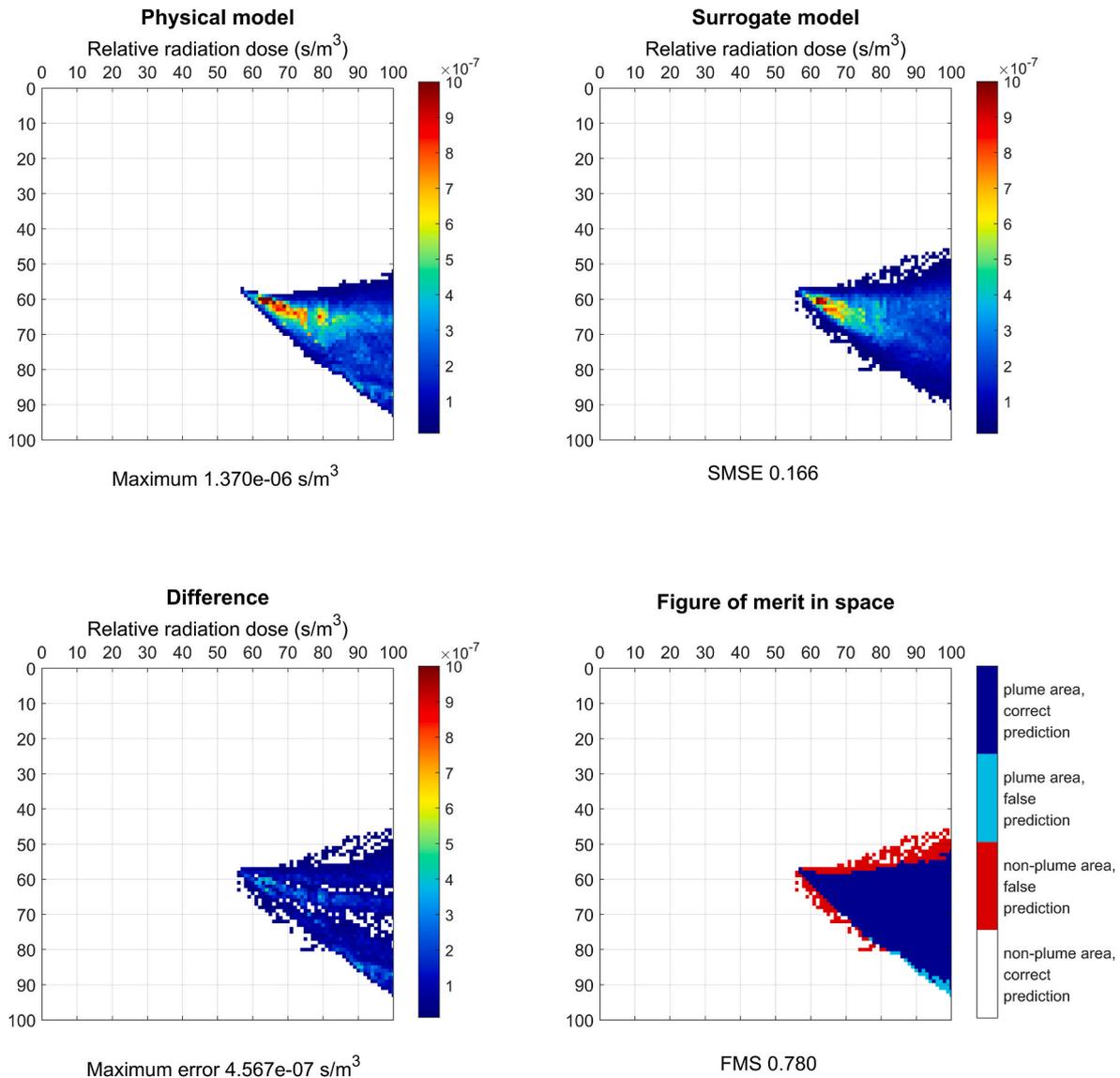


Fig. 11. Plume spreading in the southeast hilly area around the NPP (case 2). Predicted relative radiation doses of the physical model (top left) and the surrogate model (top right), the differences between the two models (bottom left), and the figure of merit in space for the predicted plume (bottom right).

average of all training or test data. The values presented are quite large and are in most cases due to underestimation of the maximum doses by the surrogate model. That shows again the already mentioned under-prediction of the very high doses by the surrogate model. Note that Table 3 shows the average error in predicting the maximum in one 2D map, i.e. the average error in predicting the maximum for one meteorological condition, while Fig. 14 (bottom plot) shows the maximum values in each grid cell as expected over a longer period of time, i.e. the worst case scenario for a given location under all different meteorological conditions.

4. Discussion

Based on statistical results obtained for a wide range of meteorological conditions and a detailed case-by-case examination, we can summarize that the developed surrogate model can predict 2D maps of relative radiation doses quite reliably using measured meteorological data and numerical weather predictions as model inputs. Results show that the central part of the plume is homogenous and well-predicted,

even though it was determined using a set of non-interacting models. This can be considered a very good result as more “outlying” cells or larger differences in neighbouring cells could be present. The direction of the plume expansion is also well predicted. However, the developed surrogate model also has some shortcomings, as seen in a significant share of cases that do not reach the desired SMSE and FMS values. Therefore, some improvements in data collection and the modelling approach are still needed to improve the performance.

As already mentioned, two years of collected input-output data with three quarters for training and one quarter for testing provide a reasonably large dataset to train and test a data-driven surrogate model. However, in some parts of the observed area, because of zero radiation values, it provides a relatively small set of informative data for model training, which reduces the performance of the trained models in these parts. Besides, data was based on a realistic occurrence of actual meteorological conditions. Therefore, some meteorological conditions may be underrepresented and statistically insignificant in machine learning algorithms. More balanced training data could be selected in future research using dimension-reduction techniques and/or pattern-selection

Test data, 11-Mar-2021 00:30:00

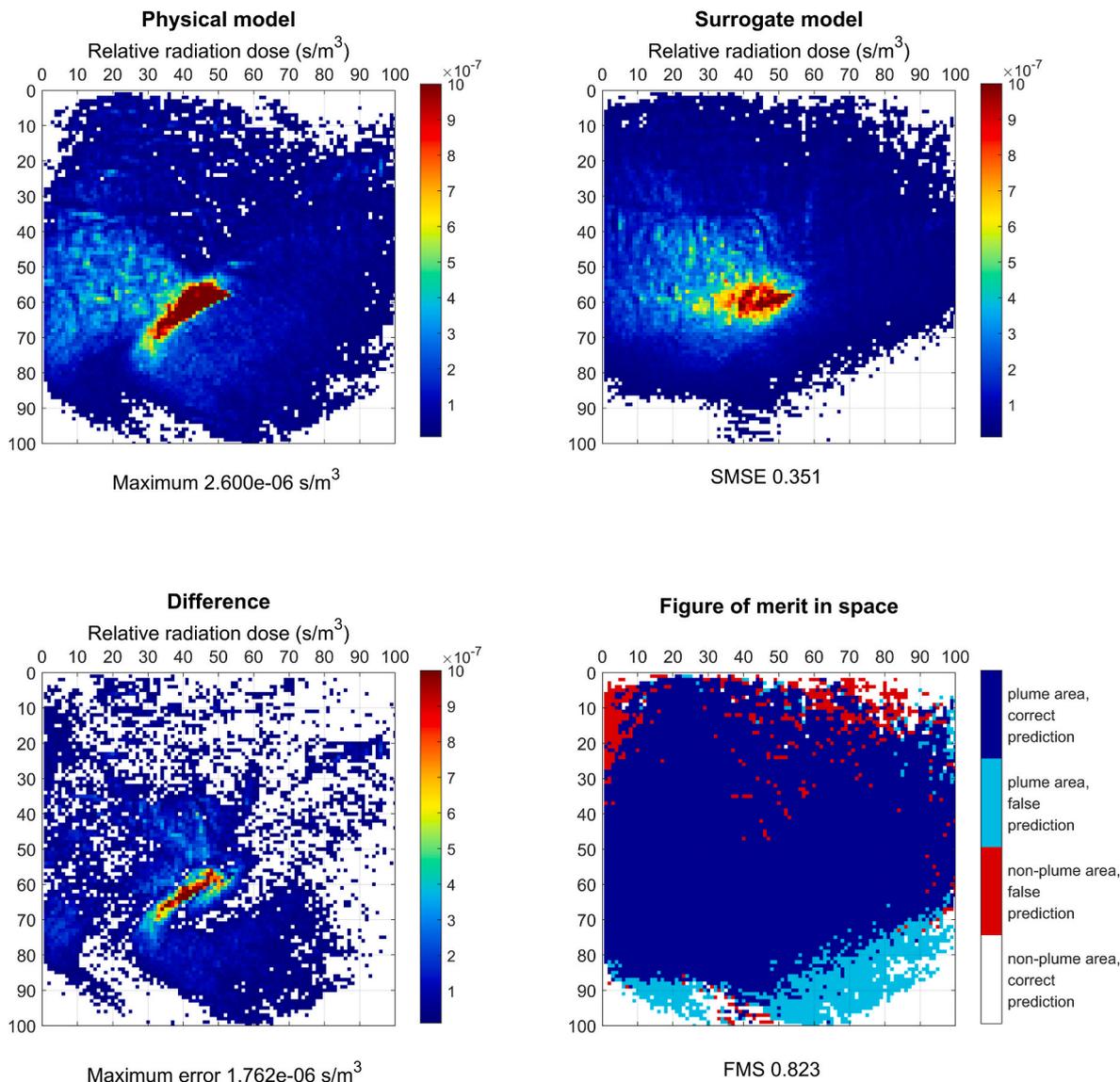


Fig. 12. Plume spreading at low wind speed conditions (case 3). Predicted relative radiation doses of the physical model (top left) and the surrogate model (top right), the differences between the two models (bottom left), and the figure of merit in space for the predicted plume (bottom right).

strategies. An approach of pattern selection strategies that has proven advantageous for improving generalization and successfully predicting high concentrations in air pollution models is described in Božnar (1997).

It should also be noted that the two-year data was split so that the first three quarters were used for training and the last quarter for testing the model. Therefore, data similar to the test data might be under-represented in the training data. Randomly splitting the data samples from the whole period into training and test data improved the performance of the surrogate model. But, in the end, it was not used because it could be too optimistic about the actual quality of the model if the test points were too close to the training points.

Another possibility regarding measurements is to provide additional ground-level station measurements. The importance analysis of meteorological variables showed that station measurements, also those distant from the NPP, are among the most important regressors of the surrogate model. It can be concluded that they partly replace the information that is otherwise provided by the multi-dimensional weather prognosis in the physical model. It is therefore expected that the prediction of the

surrogate model could be improved by additional station measurements at appropriate locations.

An important shortcoming is also in the chosen modelling approach, where a set of independent models was designed without any interactions between the models during the model training or in the construction of a 2D map. It is an excellent result that such a straightforward and relatively simple approach has yielded already quite good and useful results. It is expected that other data-driven methods, e.g. deep learning (neural network) models for multi-output regression or Gaussian Process models with grid structure for large-scale regression, could potentially further improve the performance compared to individual models, which we intend to do as future research.

Finally, the computation time of the surrogate model to predict a 2D map on an Intel Core i7 9750H CPU @ 2.60 GHz, 16 GB RAM is around 300 s. The advantage is that a parallel run for multiple input data does not significantly affect the computation time. Conversely, in the case of the physical model, the computation time increases linearly with the number of 2D maps generated. Hence, the construction of 1000 2D maps using a surrogate model requires around 560 s, while the same task with

Test data, 03-Jun-2021 23:30:00

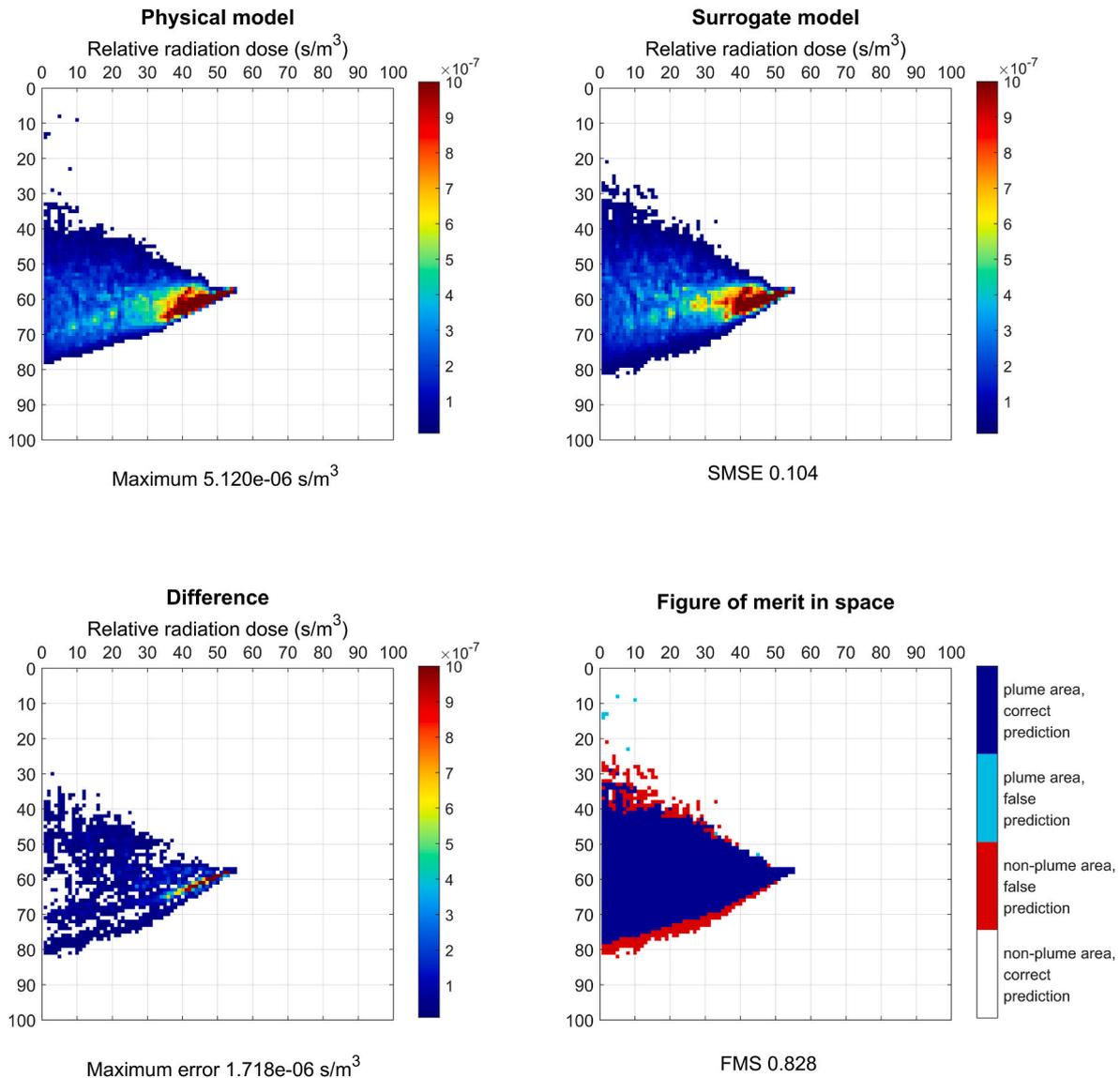


Fig. 13. Plume with a high maximum relative radiation dose (case 4). Predicted relative radiation doses of the physical model (top left) and the surrogate model (top right), the differences between the two models (bottom left), and the figure of merit in space for the predicted plume (bottom right).

the physical model on a similar dedicated computer requires approximately 100,000 s.

5. Conclusions

A surrogate model of an atmospheric dispersion model is needed for various purposes, but its design is a challenging task due to the complexity of the physical processes involved. This paper addresses the question of whether machine learning methods can replace the three demanding and computationally intensive steps of the physical model prediction, i.e. 3D weather reconstruction, Lagrangian particle dispersion and integrated relative radiation dose calculation, to predict relative radiation doses at ground level in two spatial dimensions (2D) in the case of a hypothetical radiological release from the NPP. The specific tasks were to represent the physical model predictions over complex terrain, in all and non-selectively chosen weather conditions and by using meteorological data at selected locations as the only information on weather conditions. Hence, we deal with the most general and challenging cases over complex terrain.

An attempt to represent the physical model by a set of independent tree ensemble models, each model representing a specific non-interacting location in the grid around the NPP, showed that by selectively choosing the regressors and their time range, a model is obtained that provides a suitable prediction of relative radiation doses. Among the most important regressors of the model are meteorological variables at ground-level stations, also those distant to the source of the hypothetical radiological release that provide additional information on the weather conditions. Although the surrogate model prediction in 2D is composed of individual and independent models, it gives a homogenous picture of the radiation plume and adequately predicts the direction of radiation propagation under different meteorological conditions. The homogeneity of the plume is very good with almost no individual false alarms of increased radiation outside the range of dispersion. A shortcoming of the model is the under-prediction of very high relative radiation doses, which we believe is due to the use of realistic meteorological measurements as training data, where the representation of such cases is statistically small. This problem will be further addressed by pattern selection strategies.

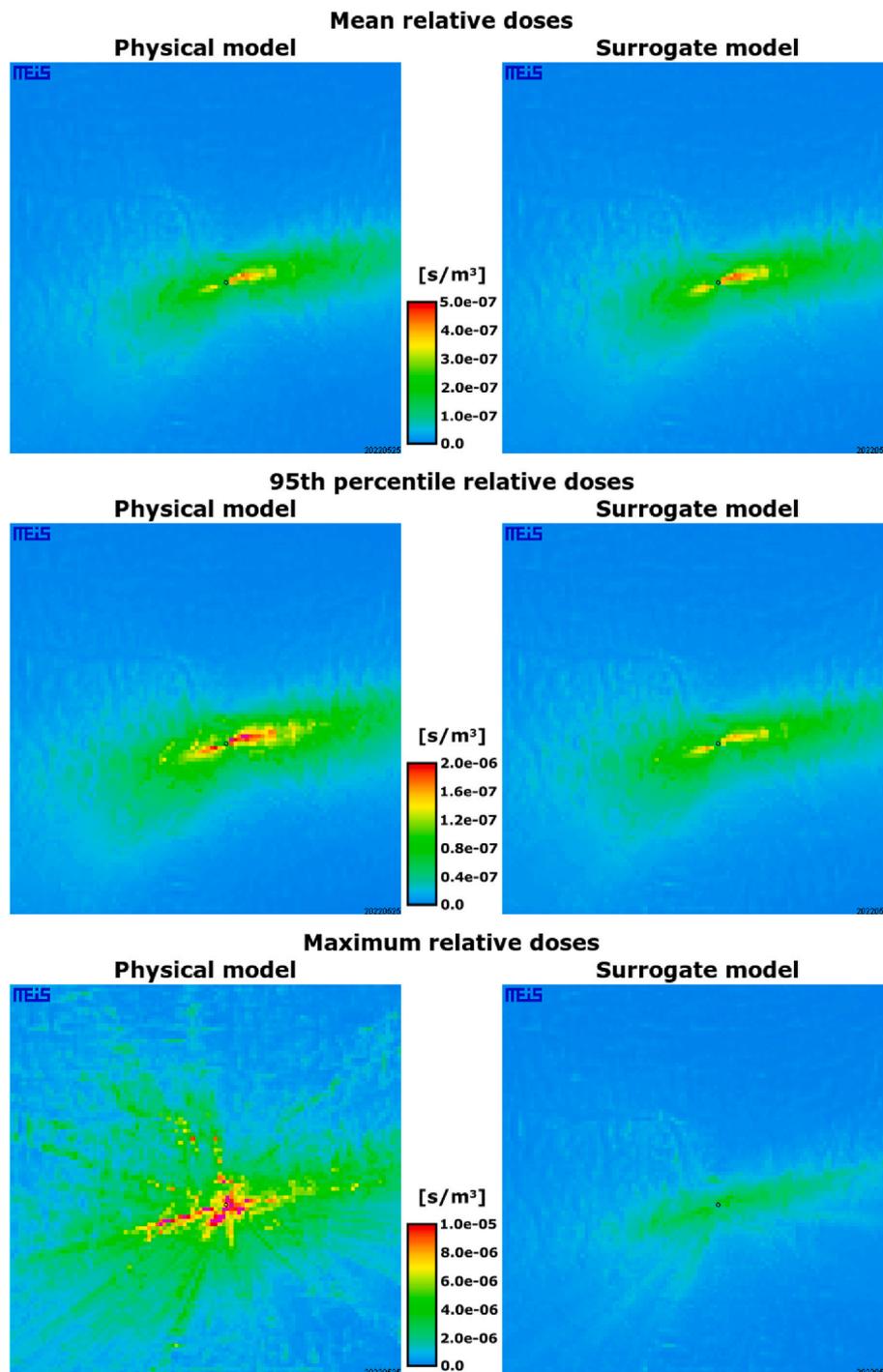


Fig. 14. Comparison of the mean (upper plot), 95th percentile (middle plot) and maximum (bottom plot) relative doses received by the population as predicted with the physical model (left) and surrogate model (right) for the test data, i.e. for approximately 4380 2D maps.

Table 3
The average values of evaluation criteria for 2D maps.

Performance measure	Unit	Training data	Test data
SMSE	–	0.203	0.470
FMS	–	0.639	0.547
MAPE _{max}	%	31.9	39.3

We can conclude that the idea of using station measurements as regressors for the surrogate model has proven to be feasible and promising also for complex terrain, although it has not yet yielded the best results. It is expected that the use of other data-driven methods, e.g. deep learning (neural network) models for multi-output regression or Gaussian Process models with grid structure for large-scale regression, would improve the performance compared to a set of individual models, which we intend to investigate in future work together with more

elaborated training data selection.

Author contributions

Nadja Hvala: Conceptualization, Methodology, Data curation, Visualization, Writing. **Primož Mlakar:** Conceptualization, Methodology, Data curation. **Boštjan Grašič:** Conceptualization, Methodology, Visualization. **Marija Zlata Božnar:** Funding acquisition, Conceptualization, Methodology, Review. **Matija Perne:** Conceptualization, Methodology, Review. **Juša Kocijan:** Supervision, Project administration, Funding acquisition, Conceptualization, Methodology, Review.

Declaration of competing interest

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

Data availability

We have shared the links to publicly available data in Section 2.3.

Acknowledgement

This work was financially supported by the Slovenian Research Agency, program P2-0001 and research project L2-2615 co-financed by the Nuclear power plant Krško. The authors would like to thank the Nuclear power plant Krško for financial support.

Appendix A. Supplementary data

Supplementary video to this article can be found online at <https://doi.org/10.1016/j.pnucene.2023.104594>.

References

- Abraham, S., Huynh, C., Vu, H., 2020. Classification of soils into hydrologic groups using machine learning. *Data* 5 (1), 2. <https://doi.org/10.3390/data5010002>.
- Alizadeh, R., Allen, J.K., Mistree, F., 2020. Managing computational complexity using surrogate models: a critical review. *Res. Eng. Des.* 31, 275–298. <https://doi.org/10.1007/s00163-020-00336-7>.
- Archetti, F., Candelieri, A., 2019. The surrogate model. In: *Bayesian Optimization and Data Science*. Springer. https://doi.org/10.1007/978-3-030-24494-1_3.
- Arianet, 2011. SURFPRO (SURface-Atmosphere Interface Processor) User's Guide. Arianet, Milano, Italy. <http://doc.aria-net.it/SURFPRO>.
- Božnar, M., Lesjak, M., Mlakar, P., 1993. A neural network-based method for the short-term predictions of ambient SO₂ concentrations in highly polluted industrial areas of complex terrain. *Atmos. Environ.* 27B (2), 221–230. doi: 10.1016/0957-1272(93)90007-S.
- Božnar, M., 1997. Pattern selection strategies for a neural network-based short term air pollution prediction model. *Proceedings Intelligent Information Systems, IIS 97*, 340–344. <https://doi.org/10.1109/IIS.1997.645285>.
- Božnar, M.Z., Mlakar, P., Grašič, B., 2012a. Short-term fine resolution WRF forecast data validation in complex terrain in Slovenia. *Int. J. Environ. Pollut.* 50, 12. <https://doi.org/10.1504/IJEP.2012.051176>.
- Božnar, M.Z., Mlakar, P., Grašič, B., Tinarelli, G., 2012b. Environmental impact assessment of a new thermal power plant Soštanj Block 6 in highly complex terrain. *Int. J. Environ. Pollut.* 48, 136. <https://doi.org/10.1504/IJEP.2012.049660>.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. Chapman & Hall, Boca Raton, FL. <https://doi.org/10.1201/9781315139470>.
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24, 123–140. <https://doi.org/10.1007/BF00058655>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Carnevale, C., Finzi, G., Guariso, G., Pisoni, E., Volta, M., 2012. Surrogate models to compute optimal air quality planning policies at a regional scale. *Environ. Model. Software* 34, 44–50. <https://doi.org/10.1016/j.envsoft.2011.04.007>.
- Dem and Corine, 2023. Digital Elevation Model and Corine Land Cover. <https://land.copernicus.eu>. (Accessed 6 January 2023). Accessed.
- Desiato, F., Finardi, S., Brusasca, G., Morselli, M.G., 1998. TRANSALP 1989 experimental campaign-I. Simulation of 3D flow with diagnostic wind field models. *Atmos. Environ.* 32, 1141–1156. [https://doi.org/10.1016/S1352-2310\(97\)00196-9](https://doi.org/10.1016/S1352-2310(97)00196-9).
- Desterro, F.S.M., Santos, M.C., Gomes, K.J., Heimlich, A., Schirru, R., Pereira, C.M.N.A., 2020. Development of a Deep Rectifier Neural Network for dose prediction in nuclear emergencies with radioactive material releases. *Prog. Nucl. Energy* 118, 103110. <https://doi.org/10.1016/j.pnucene.2019.103110>.
- GFS data, 2023. <https://www.nco.ncep.noaa.gov/pmb/products/gfs/>. (Accessed 6 January 2023). Accessed.
- Girard, S., Mallet, V., Korsakissok, I., Mathieu, A., 2016. Emulation and Sobol' sensitivity analysis of an atmospheric dispersion model applied to the Fukushima nuclear accident. *J. Geophys. Res. Atmos.* 121, 3484–3496. <https://doi.org/10.1002/2015JD023993>.
- Girard, S., Armand, P., Duchenne, C., Yalamas, T., 2020. Stochastic perturbations and dimension reduction for modelling uncertainty of atmospheric dispersion simulations. *Atmos. Environ.* 224, 117313. <https://doi.org/10.1016/j.atmosenv.2020.117313>.
- Gunawardena, N., Pallotta, G., Simpson, M., Lucas, D.D., 2021. Machine learning emulation of spatial deposition from a multi-physics ensemble of weather and atmospheric transport models. *Atmosphere* 12 (8), 953. <https://doi.org/10.3390/atmos12080953>.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning. Data Mining, Inference and Prediction*, second ed. Springer. <https://doi.org/10.1007/978-0-387-84858-7>.
- Hjerpe, A., 2016. *Computing Random Forests Variable Importance Measures (VIM) on Mixed Continuous and Categorical Data*. Thesis at KTH Computer Science and Communication, Stockholm. Corpus ID: 208013322.
- Ivatt, P.D., Evans, M.J., 2020. Improving the prediction of an atmospheric chemistry transport model using gradient-boosted regression trees. *Atmos. Chem. Phys.* 20, 8063–8082. <https://doi.org/10.5194/acp-20-8063-2020>.
- Jiang, P., Zhou, Q., Shao, X., 2020. *Surrogate Model-Based Engineering Design and Optimization*. Springer Nature Singapur. <https://doi.org/10.1007/978-981-15-0731-1>.
- Kocijan, J., 2016. *Modelling and Control of Dynamic Systems Using Gaussian Process Models*. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-21021-6>.
- Lauret, P., Heymes, F., Aprin, L., Johannot, A., 2016. Atmospheric dispersion modeling using Artificial Neural Network based cellular automata. *Environ. Model. Software* 85, 56–69. <https://doi.org/10.1016/j.envsoft.2016.08.001>.
- Le, N.B.T., Mallet, V., Korsakissok, I., Mathieu, A., Périllat, R., 2019. Calibration of a Surrogate Dispersion Model Applied to the Fukushima Nuclear Disaster. 3rd ECCOMAS Thematic Conference UNCECOMP, pp. 215–228. <https://doi.org/10.7712/120219.6337.18843>.
- Le, N.B.T., Korsakissok, I., Mallet, V., Périllat, R., Mathieu, A., 2021. Uncertainty study on atmospheric dispersion simulations using meteorological ensembles with a Monte Carlo approach, applied to the Fukushima nuclear accident. *Atmos. Environ.* X 10, 100112. <https://doi.org/10.1016/j.aeoa.2021.100112>.
- Lucas, D.D., Simpson, M., Cameron-Smith, P., Baskett, R.L., 2017. Bayesian inverse modeling of the atmospheric transport and emissions of a controlled tracer release from a nuclear power plant. *Atmos. Chem. Phys.* 17, 13521–13543. <https://doi.org/10.5194/acp-17-13521-2017>.
- Mallet, V., Tilloy, A., Poulet, D., Girard, S., Brocheton, F., 2018. Meta-modeling of ADMS-Urban by dimension reduction and emulation. *Atmos. Environ.* 184, 37–46. <https://doi.org/10.1016/j.atmosenv.2018.04.009>.
- Mathworks, 2020. *Statistics and Machine Learning Toolbox™ User's Guide R2020a*.
- Maurer, C., Baré, J., Kusmierczyk-Michulec, J., Crawford, A., Eslinger, P.W., Seibert, P., Orr, B., Philipp, A., Ross, O., Generoso, S., et al., 2018. International challenge to model the long-range transport of radionuclides released from medical isotope production to six Comprehensive Nuclear-Test-Ban Treaty monitoring stations. *J. Environ. Radioact.* 192, 667–686. <https://doi.org/10.1016/j.jenvrad.2018.01.030>.
- Mendil, M., Leirens, S., Armand, P., Duchenne, C., 2022. Hazardous atmospheric dispersion in urban areas: a Deep Learning approach for emergency pollution forecast. *Environ. Model. Software* 152, 105387. <https://doi.org/10.1016/j.envsoft.2022.105387>.
- Mlakar, P., Božnar, M.Z., Grašič, B., Breznik, B., 2019a. Integrated system for population dose calculation and decision making on protection measures in case of an accident with air emissions in a nuclear power plant. *Sci. Total Environ.* 666, 789–800. <https://doi.org/10.1016/j.scitotenv.2019.02.309>.
- Mlakar, P., Božnar, M.Z., Grašič, B., 2019b. Relative doses instead of relative concentrations for the determination of the consequences of the radiological atmospheric releases. *J. Environ. Radioact.* 196, 1–8. <https://doi.org/10.1016/j.jenvrad.2018.10.005>.
- Molnar, C., 2022. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, second ed. <https://christophm.github.io/interpretable-ml-book/>.
- Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. *BMC Bioinf.* 9, 307. <https://doi.org/10.1186/1471-2105-9-307>.
- Tinarelli, G., Anfossi, D., Trini Castelli, S., Bider, M., Ferrero, E., 2000. A new high performance version of the Lagrangian particle dispersion model spray, some case studies. In: Gryning, S.E., Batchvarova, E. (Eds.), *Air Pollution Modeling and its Application XIII*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-4153-0_51.
- WRF, 2023. https://www2.mmm.ucar.edu/wrf/users/download/get_sources.html. (Accessed 6 January 2023). Accessed.