ORIGINAL PAPER

An enhanced fourier neural operator surrogate for radioactive plume transport forecasting

Ali Ayoub[1] · Haruko M. Wainwright[1] · Lijing Wang[2] · Giovanni Sansavini[3]

Abstract

Accurate real-time forecasts of atmospheric plume behavior are crucial for effective management of environmental release incidents. However, the computational demands of weather simulations and particle transport codes limit their applicability during emergencies. In this study, we employ a U-Net enhanced Fourier Neural Operator (U-FNO) to statistically emulate the calculations of the WSPEEDI dose forecasting numerical simulator, using pre-calculated ensemble simulations. The developed emulator is capable of effectively simulating any radioactive-release scenario and generating the time series of dose distribution in the environment 4000 times faster than the numerical simulator, while still maintaining high accuracy. It predicts the plume direction, extent, and dose-rate magnitudes using initial- and boundary-condition meteorological data as input. The speed and efficiency of this framework offers a powerful tool for swift decision-making during emergencies, facilitating risk-informed protective actions, evacuation execution, and zone delineation. Its application extends to various contaminant release and transport problems, and can be instrumental in engineering tasks requiring uncertainty quantification (UQ) for environmental risk assessment.

Keywords Deep learning · Fourier neural operator · Atmospheric transport and dispersion · Contaminant releases · Emergency response, WSPEEDI

1 Introduction

Industrial accidents involving the release of toxic materials represent a serious threat to public health, as demonstrated by historical events such as the Bhopal chemical accident (Chernov and Sornette 2020) and the Fukushima Daiichi nuclear power plant (FDNPP) accident (Ayoub et al. 2024; Chernov et al. 2022). The impacts of these accidents on health and the environment are contingent upon various factors, including the quantity of substances discharged, prevailing weather conditions at the time of release, and the response and behavior of the affected population. To effectively safeguard at-risk communities and guide the implementation of appropriate protective measures, we need a comprehensive understanding of the pathways, speed of transport, hotspots, and size of the affected regions (Osborn et al. 2013).

In the nuclear industry, various atmospheric transport and dispersion (ATD) models and codes have been developed to model the behavior of released particles in the environment (Sato et al. 2018). Furthermore, there have been several advancements to address uncertainties and enhance the modeling of the involved physical phenomena (e.g. microphysics, wet and dry deposition, particles size) (Zhuang et al. 2023, 2024). Some of these codes are designed and mainly utilized for risk assessment and licensing purposes, employing historical meteorological data to evaluate potential consequences following hypothetical atmospheric releases. Examples of these codes include OSCAAR in Japan and MAACS in the US (Pascucci-Cahen 2024). Other codes operate as prognostic models to make dose projections and help inform decisions during responses to radiological emergencies. Prominent among these are RASCAL in the US (Bradley 2007), JRODOS (Wengert 2017) and FLEXPART (Stohl et al. 2005) mainly in Europe, and WSPEEDI in Japan (Kadowaki et al. 2017), which necessitate initially running

✉ Ali Ayoub
   aliayoub@mit.edu

1   Massachusetts Institute of Technology, Nuclear Science and Engineering, Cambridge, MA, USA

2   Lawrence Berkeley National Laboratory, Berkeley, CA, USA

3   Institute of Energy and Process Engineering, ETH Zurich, Zurich, Switzerland

numerical weather simulations to forecast the required meteorological fields for the radionuclide-transport calculations.

However, the FDNPP accident has exposed the challenges involved in the radiological emergency response, particularly in the utilization of ATD codes to guide immediate protective actions (Ayoub et al. 2024; Sugawara and Juraku 2018). Despite WSPEEDI (Worldwide System for Prediction of Environmental Emergency Dose Information) providing initial plume predictions during the accident, these forecasts weren't effectively utilized in the evacuation decisions (Nagai et al. 2023). For instance, the main plume during the FDNPP accident was a result of a few hours of release, following a primary containment failure (Ayoub et al. 2024). In fact, major releases generally happen within short periods, as puffs, particularly following containment venting or depressurization events. Hence, swift protective measures become crucial. However, ATD codes capable of guiding these measures necessitate running both fine-resolution numerical weather forecasts and particle-transport models in real-time. Such simulations are computationally demanding, posing a challenge for their prompt utilization in emergency applications.

Machine learning (ML) models offer a promising alternative to emulating computationally expensive physical models. For example, Desterro et al. (Desterro et al. 2020) used a deep rectifier neural network to predict the dose distribution up to 1 h post-release, for a set of postulated accidents. Their model inputs comprised wind velocity, wind direction, x and y positions, and time elapsed since the accident onset. Similarly, Girard et al. (Girard et al. 2016) developed a Gaussian process emulator that reproduces the dose-rate distribution in the environment following a release, using an ensemble of source terms and meteorological data as input. Beyond nuclear applications, numerous studies have explored statistical emulators for atmospheric transport calculations. These studies utilized prevailing or forecasted meteorological variables as inputs to forecast concentration maps of pollutants (Kocijan et al. 2023; Mallet et al. 2018). However, these developments are primarily designed for uncertainty quantification (UQ) applications and are less suitable for real-time forecasts, because they do not produce meteorological forecasts; instead, they depend on the availability of such data as input.

On the weather forecasting side, deep learning methods have shown promising capabilities in effectively replicating numerical simulators (Wang et al. 2019). For example, Höhlein et al. (Höhlein et al. 2020) presented various convolutional neural network (CNN) architectures for short-range wind predictions at near-surface levels. DeepMind and Google developed a graphical neural network that forecasts several meteorological variables (Lam et al. 2022). However, the majority of these ongoing advancements have been confined to short-range forecasts, typically ranging from 1 to 2 h into the future, or operate at relatively low resolutions,

typically between 10 to 30 km in grid size. These limitations render them inadequate for providing the meteorological data forecasts needed as input to ATD surrogates. To address this limitation, Ayoub et al. (2023) developed an autoregressive emulator that forecasts high-resolution meteorological variables at a single point (release location), which can be combined with a simplified ATD Gaussian Plume Model to forecast the steady-state radionuclide distribution. However, the Gaussian Plume Model relies on assumptions such as temporally constant and spatially homogeneous meteorological conditions (Leelőssy et al. 2018), which may not apply during an emergency.

In this paper, to address the computational issues associated with running the high-resolution numerical weather simulations and the radioactive transport codes, we develop a data-driven emulator that statistically replicates these calculations based on pre-calculated ensemble simulations. Our approach involves leveraging a specific class of neural operator learning (Li et al. 2020), known as the U-Net enhanced Fourier Neural Operator (U-FNO) (Wen et al. 2022), to construct a surrogate statistical model for the WSPEEDI physical simulator. This simulator includes both a weather forecasting model and a Lagrangian particle transport and dispersion model (Kadowaki et al. 2017). The neural operator utilizes a neural network to learn mesh-independent, resolution-invariant solution operators for Partial Differential Equations (PDEs), without knowing the exact governing equation (Li et al. 2020). U-FNO has been employed to solve various flow-dynamics problems by emulating the ensemble simulations of spatiotemporal image-like outputs (Meray et al. 2024; Zhang et al. 2022). To the best of the authors' knowledge, this study is the first to build and test a novel ML-based surrogate that can effectively simulate any release scenario, forecast the associated plume behavior, and generate the time series of dose distribution in the environment in real-time.

2 Methodology

The framework consists of two parts: (1) ensemble simulations and U-FNO training and (2) emulator prediction. We first run a large number of ATD simulations that capture the variability in the meteorological conditions, thereby creating the ensemble data for U-FNO training (Fig. 1A). Once such an emulator is trained, we can efficiently simulate any release scenario and predict the associated plume behavior and dose distribution (Fig. 1B).

2.1 Physical modeling

In this study, we use the Worldwide version of System for Prediction of Environmental Emergency Dose Information,
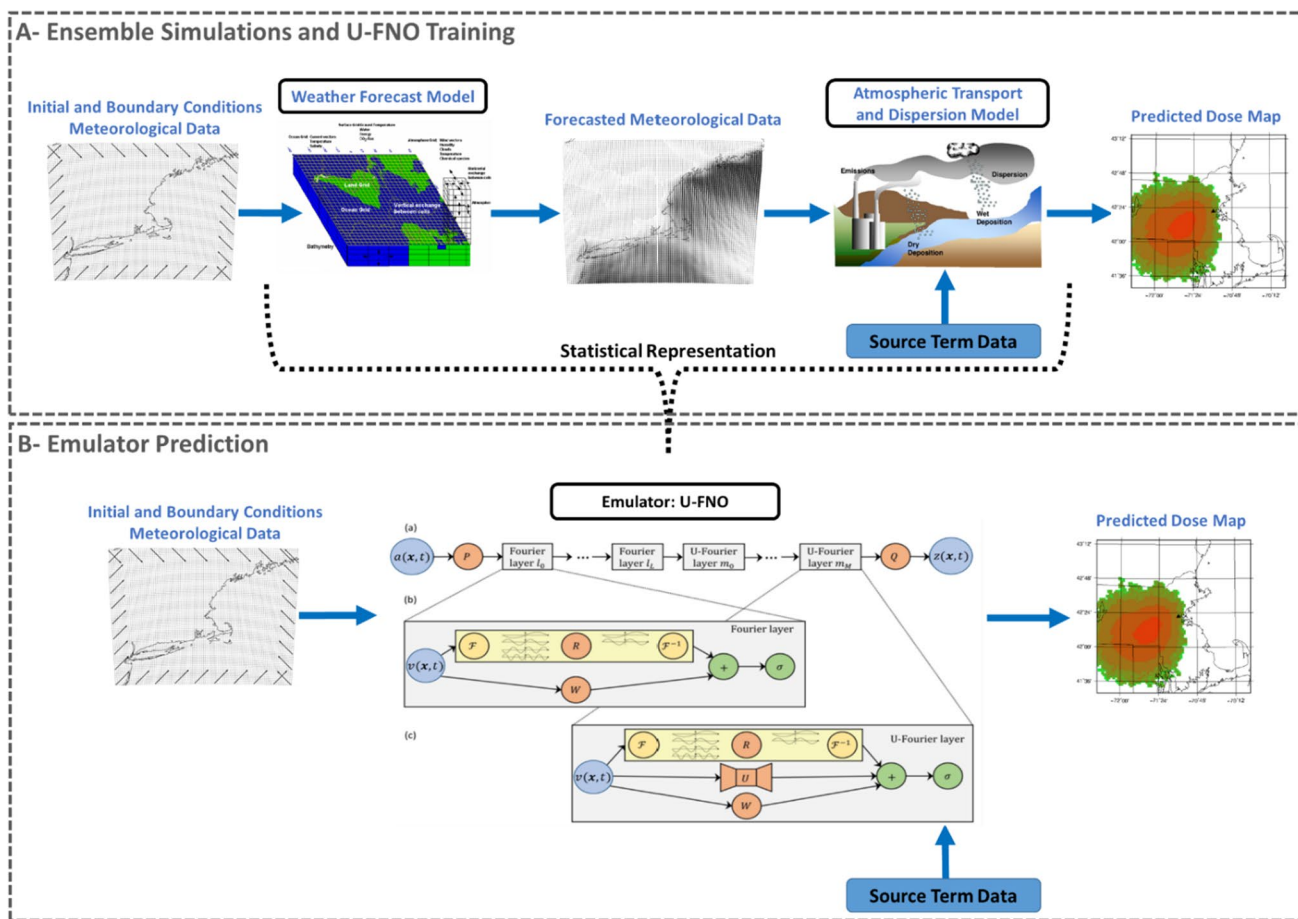
Fig. 1 Computational framework for emulator development: (**A**) Generation of weather and atmospheric dispersion training data (ensemble simulations), (**B**) On-line use of the developed emulator for plume prediction. The middle inset figure in B is the U-FNO architecture retrieved and edited from Wen et al. (2022)

WSPEEDI version 1.1.4 (Kadowaki et al. 2017). WSPEEDI is composed of two computational models, the Weather Research and Forecasting Model (WRF) and a Lagrangian particle transport and dispersion model, GEARN.

Developed by the National Center for Atmospheric Research (Skamarock et al. 2005), WRF is a fully compressible, non-hydrostatic, regional numerical prediction system. The WRF model forecasts the three-dimensional meteorological fields by solving the governing equations of the atmospheric dynamics. To initiate these calculations, WRF requires a set of static geographic data representing the area under consideration. Additionally, to establish the initial and boundary conditions for the calculations, WRF relies on gridded meteorological data, typically available at coarse spatial and temporal resolutions, and contains historical meteorological information as well as real-time forecasts (up to several hours or days in the future) derived from global weather prediction models (Pu and Kalnay 2019). The data come in packed binary format and include temperature, pressure, humidity, wind components, and cloud cover. WRF calculates the forecasted three-dimensional meteorological fields, defined with high spatial and temporal resolution, as specified by the user over the simulated domain. The outputs are provided at 35 hybrid vertical levels, starting from the surface and rising up to approximately 100 hPa (~ 16 km) (Beck et al. 2020).

Using the meteorological fields predicted by WRF as input, GEARN simulates the three-dimensional atmospheric dispersion of radionuclides discharged from a point source at the regional scale, given a specified release amount and release time (source term). The simulation involves calculating advection, driven by the mean wind field, and diffusion resulting from turbulence on the subgrid scale. This is achieved by tracing the trajectories of a large number of marker particles representing the discharged radioactive material (Terada and Chino 2008). The resulting particle distribution is then converted into air concentration of radioactivity in the three-dimensional calculation cell (measured in Bq/m^3). Additionally, GEARN computes the surface deposition amount (measured in Bq/

m^2) resulting from dry deposition, wet deposition, and fog water deposition (Terada and Chino 2008). Air dose rates (cloudshine, groundshine, and total) are then determined in (μGy/h) using the respective dose conversion factor for each radionuclide.

2.2 U-Net Enhanced Fourier Neural Operator (U-FNO)

2.2.1 Architecture

The Fourier Neural Operator (FNO) architecture is designed to learn the infinite-dimensional-space mapping from a finite number of input $a(\mathbf{x}, t)$ and output $z(\mathbf{x}, t)$ pairs, defined in a 2D spatial domain $\mathbf{x} = (x_1, x_2)$ and time t, using integral kernel operators in the Fourier space (Li et al. 2020). The key difference between FNO and other convolutional neural networks is that instead of learning weights and biases in the spatial or temporal domain, FNO learns the weights in the frequency domain and uses inverse Fourier transform to project back to our domain of interest. Wen et al. (Wen et al. 2022) proposes a new U-FNO architecture (shown in Fig. 1B) where it adds another U-Net architecture to increase the accuracy of training. Specifically, U-FNO contains the following four steps (Wen et al. 2022):

1. Transform the input $a(\mathbf{x}, t)$ to a higher dimensional space $v_{l_0}(\mathbf{x}, t) = P(a(x, t))$ through a fully connected neural network P.
2. Apply a sequence of L iterative Fourier layers $v_{l_0}, v_{l_1}, \ldots, v_{l_L}$, such that: $v_{l_{j+1}}(\mathbf{x}, t) = \sigma\left(F^{-1}\left(R \bullet F\left(v_{l_j}\right)\right)(\mathbf{x}, t) + W v_{l_j}(\mathbf{x}, t)\right)$, for $j = 0, 1, ..., L-1$, where F denotes the Fourier transform from the spatiotemporal domain to the Fourier frequency domain, R is a linear transformation in the Fourier domain, F^{-1} is the inverse Fourier transform, W is a linear bias term, and σ is the activation function.
3. Apply a sequence of M iterative U-Fourier layers $v_{m_0}, v_{m_1}, \ldots, v_{m_M}$, such that : $v_{m_{k+1}}(\mathbf{x}, t) = \sigma\left(F^{-1}\left(R \bullet F\left(v_{m_k}\right)\right)(\mathbf{x}, t) + U\left(v_{m_k}\right)(\mathbf{x}, t) + W v_{m_k}(\mathbf{x}, t)\right)$, for $k = 0, 1, ..., M-1$, where U denotes a U-Net convolutional neural network (CNN) operator, and the other notations have the same meaning as in the Fourier layer (Step 2 above).
4. Project v_{m_M} back to the original space of the output $z(\mathbf{x}, t) = Q(v_{m_M}(x, t))$ using a fully connected network Q.

2.2.2 Inputs and outputs for surrogate model training

In our problem, we have two types of input (predictor) variables: field variables and scalar variables. The field input variables correspond to the interpolated initial and boundary

condition data for the wind components: the zonal U and meridional V. They are computed by the WRF Preprocessing System (WPS) before running the WRF forecast simulations. Specifically, we use the surface pressure level 2D U and V fields interpolated into the higher resolution calculation domain of interest defined by $n_{x1,1} \times n_{x2,1}$, where $n_{x1,1}$ and $n_{x2,1}$ are the number of grids in the x_1 and x_2 directions, at a specified resolution. As such, each of $U(\mathbf{x}, t)$ and $V(\mathbf{x}, t)$ forms a 3D spatiotemporal tensor of shape $(n_{x1,1}, n_{x2,1}, n_t)$, as shown in Fig. 2, where n_t is the number of time steps necessary to cover the forecast period of interest. n_t varies depending on the frequency of the used initial and boundary condition data.

The scalar input variable concerns the time of the radioactive release, T_r, given as an integer representing the hour of the release within the simulation period of interest. To ensure the uniformity in the shape of the input variables for effective U-FNO model training, the scalar input variable (T_r) is broadcasted into a tensor $T_r(\mathbf{x}, t)$ of shape $(n_{x1,1}, n_{x2,1}, n_t)$, where all the elements have the same scalar value.

In our modeling of the source term, we adopt a unit release condition, specifically assuming 1 Bq/h of ^{131}I and 1 Bq/h of ^{137}Cs s for a one-hour release segment. By applying scaling, we can compute the actual results by adjusting the calculated outcomes based on this unit release to match the actual release amount. Moreover, the concentrations resulting from a time-varying release or a series of releases at one location can be summed up through superposition to represent any desired source term, following the approach developed by Terada et al. (Terada et al. 2020). This simplification enables us to streamline the calculations, allowing us to exclude the source term as an input variable.

As the output of interest (target variables), we extract the 2D spatial map of the total air dose rates at ground level, from the results of the GEARN simulations. Denoted as $D(\mathbf{x}, t)$, these extracted dose rates are defined over the domain $n_{x1,2} \times n_{x2,2}$, where $n_{x1,2}$ and $n_{x2,2}$ are the number of grids—of some specific resolution—in the x and y directions respectively, and such that $n_{x2} \times n_{y2} \subseteq n_{x1} \times n_{y1}$. As such, $D(\mathbf{x}, t)$ forms a 3D spatiotemporal tensor of shape (n_{x2}, n_{y2}, n_{t2}), where n_{t2} is the number of dose-map-forecast time steps of interest.

3 Application

3.1 Description of the simulations setup

To demonstrate our approach, we assume a hypothetical release at the location of the authors' institution (Massachusetts Institute of Technology). For the WRF calculations, we simulate an area of about 580 km by 580 km centered at MIT (Domain 1 in Fig. 3). Moreover, to achieve a higher accuracy within a particular area where we want to simulate

Fig. 2 Visualization of the broadcasted input variables
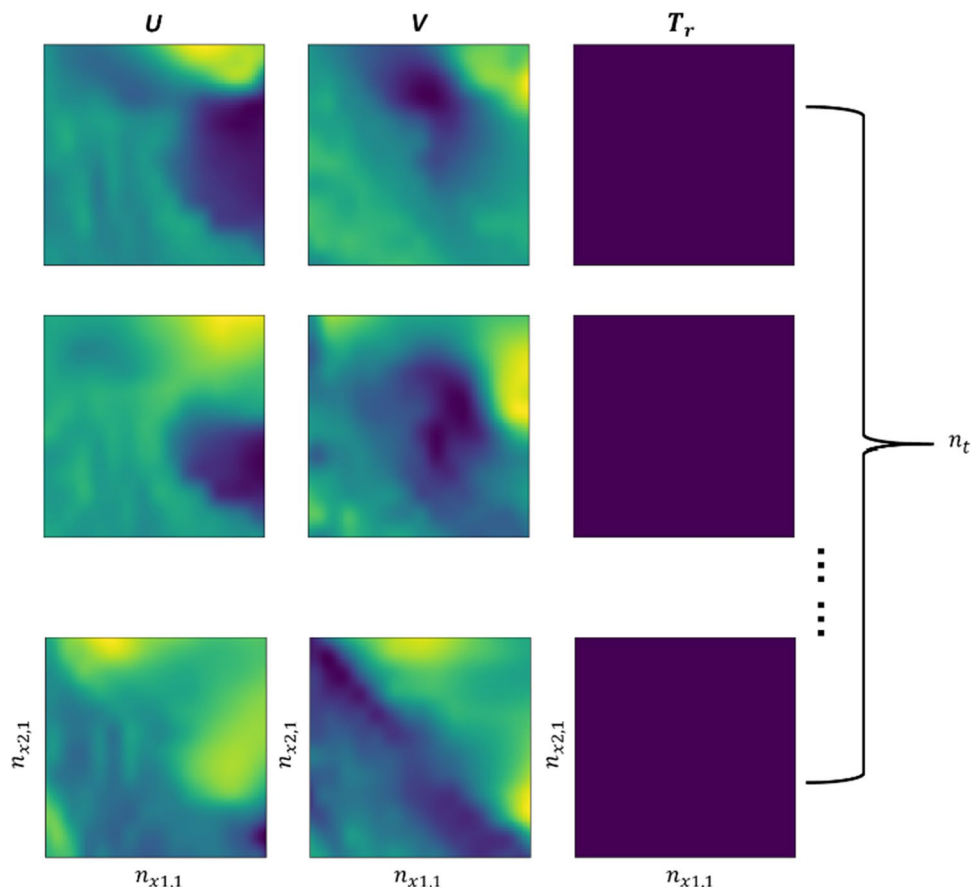
the transport (GEARN calculations), we employ the domain-nesting technique to perform local high-resolution calculations. This technique involves running concurrent WRF simulations in both the fine-resolution domain (Domain 2 in Fig. 3) and a surrounding coarser domain (Domain 1). For the initial- and boundary-condition WRF input data, we use the grid point value (GPV) of the Global Spectral Model (GSM) provided by the Japan Meteorological Agency (JMA) (Amagasa et al. 2007). This input data is available in 6-h frequency, at 20 km horizontal grid resolution, and defined at surface pressure as well as at 17 isobaric vertical levels up to 10 hPa. A summary of the WRF and GEARN calculation settings employed in our simulations is provided in Table 1.

3.2 Datasets

We conducted a total of 616 WSPEEDI simulations, spanning several months of representative weather conditions, with the calculation setup outlined in Section 3.1. We extracted the $U(\boldsymbol{x}, t)$ and $V(\boldsymbol{x}, t)$ input data from the gridded data of the larger domain, Domain 1 (Table 1), where $n_{x1} = n_{y1} = 97$. To cover the 12-h WRF forecast calculations specified in Table 1, and using initial- and boundary-condition meteorological data available at a 6-h frequency, $n_t = 3$ time steps (corresponding to the initial, 6-h, and 12-h
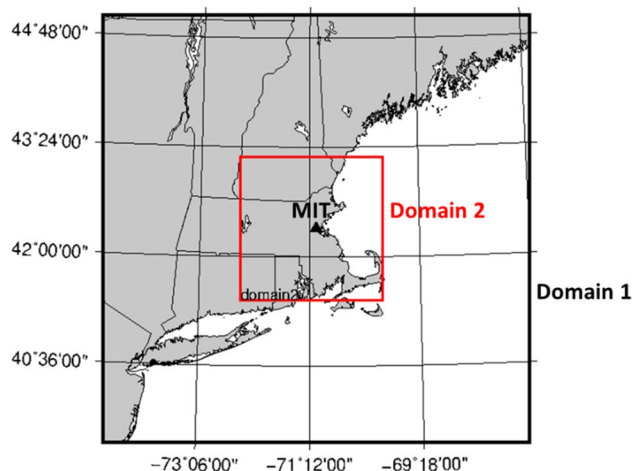

Fig. 3 Computational domains for the WSPEEDI simulations. Parent domain 1 and a high-resolution nested domain 2, centered at MIT with Latitude: 42°21′25″, Longitude: -71°5′34″

forecasted boundary conditions). We considered seven distinct release times (T_r) within the 12-h weather forecasts: at hour 0, 1, 2, ..., 6, resulting in seven unique 6-h GEARN plume transport simulations (Table 1).

Table 1 WSPEEDI simulation settings

	Domain 1	Domain 2
Applied WRF calculations	Yes	Yes
Applied GEARN calculations	No	Yes
WRF Simulation Period	12 h	12 h
GEARN Simulation Period	–	6 h post release
Horizontal grid number	97×97	97×97
Horizontal resolution (km)	6 km	2 km
Vertical levels of WRF	35 hybrid levels from surface up to 100 hPa	
Vertical levels of GEARN	29 levels from surface to 10 km	
WRF Integration time step	36 s	12 s
GEARN dispersion time step	–	12 s
Initial and boundary conditions data	GPV (GSM, 6-h frequency, 20 km horizontal resolution, surface pressure + 17 isobaric vertical levels up to 10 hPa) provided by the Japan Meteorological Agency (Amagasa et al. 2007)	
Physics options (microphysics, cumulus, land surface boundary layer, long-wave and short-wave radiation	Same as in Kadowaki et al. (2017)	
Source Term	• Unit hour release • Release location: Latitude: 42°21′25″ Longitude: -71°5′34″	
Release heights	85 m (to simulate a release from the stack)	

From the GEARN results, we derive our output $D(\boldsymbol{x}, t)$ as the high-resolution dose-rate maps of Domain 2 (Table 1), where $n_{x2} = n_{y2} = 97$. We specifically focus on forecasting the dose-rate maps in 2-h intervals during the 6-h post-release plume simulation. This translates to $n_{t2} = 3$, corresponding to the 2-h, 4-h, and 6-h post-release dose-rate maps.

3.3 U-FNO training configuration

We first standardize the datasets using MinMax scaling to scale $U(\boldsymbol{x}, t)$ and $V(\boldsymbol{x}, t)$ within the 0–1 range. Additionally, we scale the release-time constant tensor $T_r(\boldsymbol{x}, t)$ between 0 and 1 by dividing each value by the largest release time present in the dataset, which is 6. The output $D(\boldsymbol{x}, t)$ is standardized into a zero-mean and unit-variance distribution. We selected this scaling after we explored various scaling techniques such as logarithmic scaling, considering the wide span of dose-rate values across the domain grids (spanning several orders of magnitudes). We adopt a mean square error (MSE)-based loss function in our study.

We partition our 616 data points into training and testing datasets utilizing an 85–15 train-test split. The applied U-FNO architecture features $L = 3$ Fourier and $M = 3$ U-Fourier layers, and uses rectified linear layer (ReLu) activation functions. During training, we set an initial learning rate of 0.001, gradually reducing it with a constant step and reduction rate. The training setup consists of the following specifications: optimizer: Adam, epochs: 140 (we observed diminishing loss improvement beyond 100 epochs), and a batch size: 4. The Python code for the U-FNO model architecture, training, testing, and evaluation, as well as the used datasets, are available at https://github.com/aliayoub123/UFNO_ATD.git.

4 Results

4.1 Plume forecasts

Figure 4 shows the results of the trained U-FNO model predictions for five random test samples. Specifically, it compares between the WSPEEDI plume forecasts (ground truth) and the U-FNO forecasts at 2-h and 6-h post release. Our U-FNO forecasts are found to be in close agreement with the WSPEEDI physical simulations across all test samples, effectively capturing the behavior of the plume, irrespective of its direction, width/spread, or extent. In Test samples 1 and 4, where the plume has a well-defined direction and large extent (size)—typically associated with high wind speed scenarios favoring more transport and less dispersion—the U-FNO result accurately predicts the distinct plume direction and extent. On the other hand, in the scenarios where the plume extent is smaller, and the high dose regions primarily concentrate around the release location—typically arising from lower wind speeds leading to dispersion-dominance—the U-FNO model is equally capable of predicting this dose accumulation behavior and capturing the plume shape (Test samples 2, 3, and 5).

In addition to the shown "Difference" map, which displays a spatial distribution of the absolute dose-rate errors,
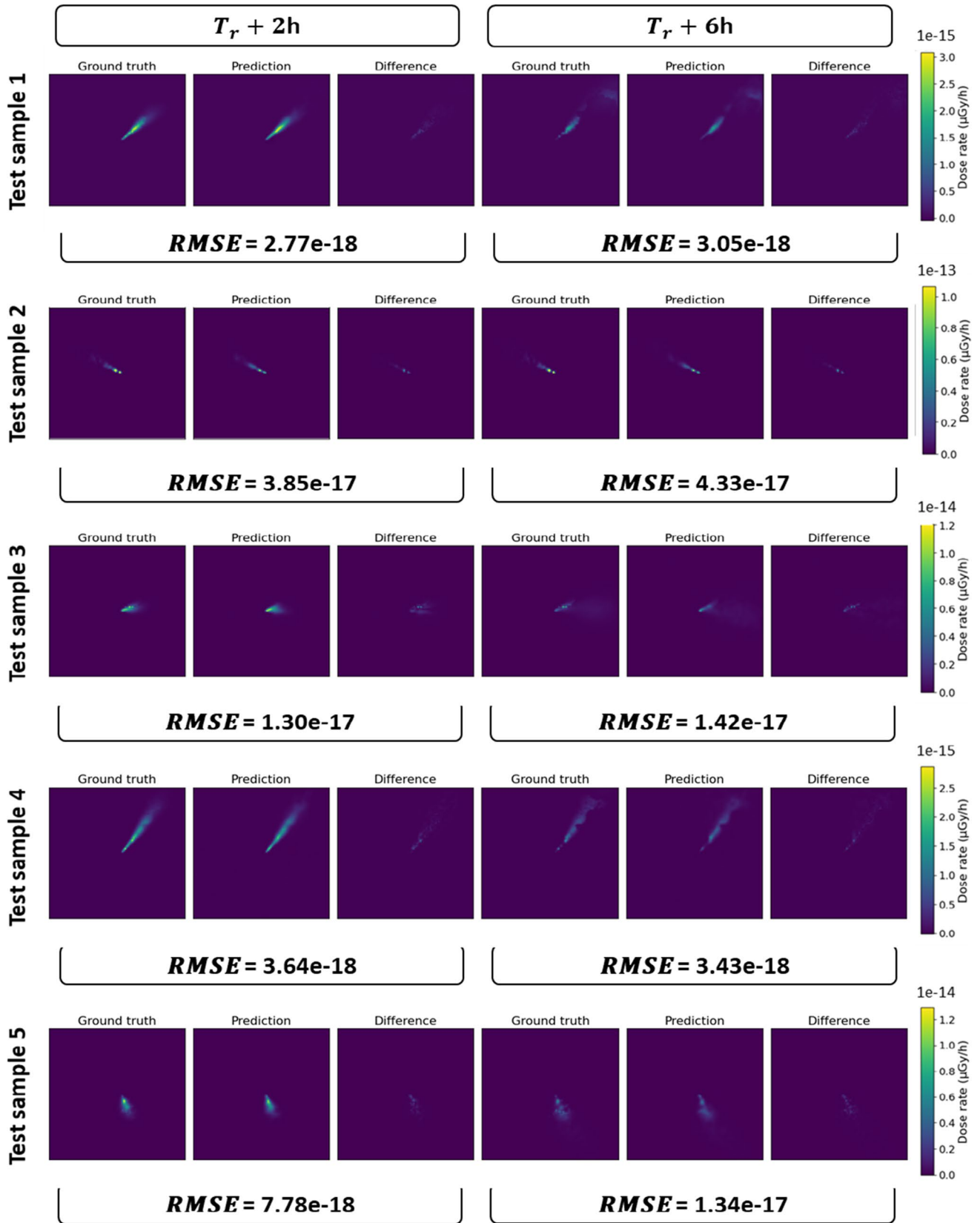
Fig. 4 Comparative analysis of the predicted dose-rate maps. Each test sample depicts the WSPEEDI dose-rate map (Ground truth), the U-FNO predicted dose-rate map (Prediction), and the absolute error map (Difference), at 2-h and 6-h post release. The RMSE between the true and the predicted map averaged over all the grids is shown for each time step

we compute the root mean square error (RMSE) at every time step. The RMSE values consistently indicate that the average discrepancies between our predictions and the physical simulations are typically about three orders of magnitude lower than the respective dose-rate values (color bar), which roughly translates to relative errors of about 0.1%. For instance, in Test Samples 1 and 4, characterized by a large plume extent and low doses (10^{-15} μGy/h (color bar)), the RMSEs are on the order of 10^{-18} μGy/h. Conversely, in Test Samples 2, 3 and 5, where the majority of the dose remains concentrated near the release, the RMSE is on the order of 10^{-17} μGy/h, while the corresponding dose rates (color bar) are on the order of 10^{-13}—10^{-14} μGy/h.

4.2 Performance metrics

To study the overall performance of the U-FNO model, we calculate two key performance metrics: the RMSE and the R^2 score, for both the training and testing datasets. Table 2 displays the average metrics calculated across all samples for the dose rates over all grids and time steps. The trained U-FNO model demonstrates excellent performance in forecasting the plume transport time series, exhibiting high accuracy (low RMSE) and a strong goodness of fit (high R^2 score), for both the training and testing data.

Notably, the RMSEs of the test samples shown in Fig. 4 did not exhibit any clear trend across the three time steps, remaining similar—within the same order of magnitude. In order to investigate the potential temporal dependencies or error propagation over time, we utilize a boxplot representation (Fig. 5) to visualize the RMSE distribution across all test samples vs time. Figure 5 shows that the RMSEs exhibit similar distributions over time, with similar quartile estimates and interquartile ranges. This suggests the absence of significant temporal dependencies or systematic error propagation.

4.3 In-plume results

Given that the plume typically occupies only a small portion of the domain, as illustrated in the maps of Fig. 4, the majority of the grids have zero dose rates. We evaluate the performance within the plume by masking all the grids that have zero dose rates and extracting the "ground truth plume domain." We then compare the values in these grids to their respective ones predicted by the U-FNO model. We find that the model's performance inside the plume domain is comparable to the
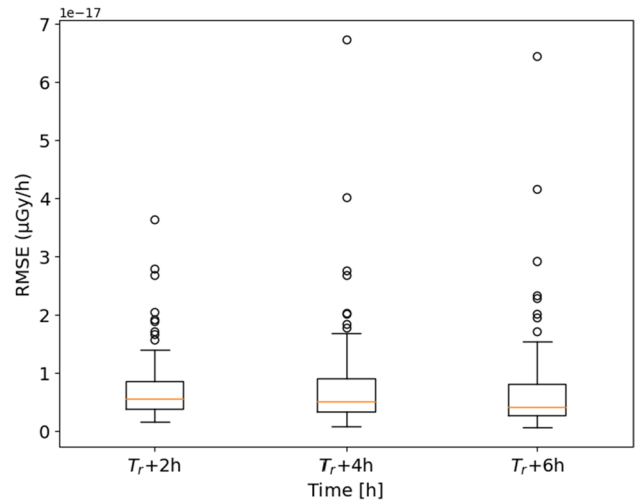

Fig. 5 Boxplot representing the distribution of RMSEs across the test samples for each time step

full domain, with an average RMSE of 2.92e-17 μGy/h and an average R^2 score of 0.78.

To further assess the one-to-one resemblance between the WSPEEDI and U-FNO predictions within each grid of the WSPEEDI plume domain, we present Fig. 6, which displays a scatter plot of WSPEEDI vs U-FNO predicted dose rates across all test samples. Given the considerable volume of data points, we use normalized density contours to visualize the number of data points in each contour level. Notably, the high-density region with less than 0.2e-13 μGy/hr is along the one-to-one line. In the high-dose region with smaller density, the U-FNO model exhibits a tendency to underestimate some dose rates, particularly within the mid to high dose range (4e-14—1e-13 μGy/h). Such high dose regions are a result of wet deposition phenomena, suggesting that the underestimations are prevalent in scenarios involving significant wet deposition or rainfall around the release time, resulting in more localized deposition and reduced plume dispersion. For instance, Test Sample 2 (Fig. 4) visually illustrates this discrepancy, particularly in capturing near-release hotspots where the U-FNO prediction falls short.

In addition, we evaluate the plume size between the simulations and the U-FNO predictions. The plume size is characterized here by the count of grids with non-zero dose rates. Figure 7 shows that the trained U-FNO model tends to overestimate the size of the plume in the majority of the test cases. Specifically, this indicates that the U-FNO model sometimes predicts non-zero dose rates for the grids that, in reality, reside outside the true plume, where the "true" dose rates are zero. When examining the U-FNO predicted values in these grids, we find that the majority correspond to very low dose rates, resembling those typically found at the periphery of the plume.
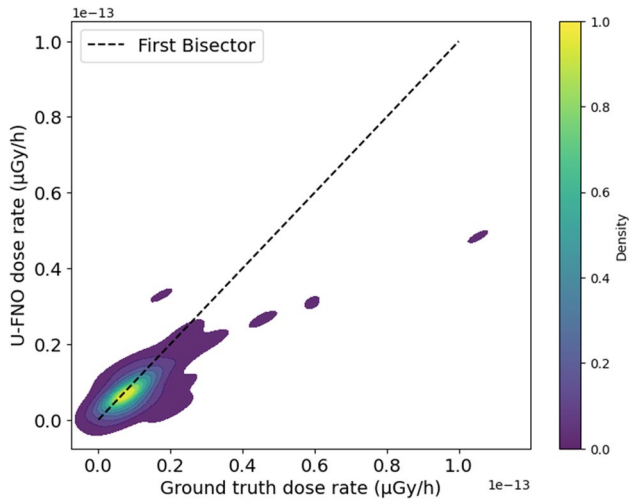
Table 2 Performance metrics over the training and testing datasets

	Training Dataset	Testing Dataset
Average RMSE (μGy/h)	5.06e-18	7.43e-18
Average R^2 score	0.95	0.79

Fig. 6 Scatter plot of the WSPEEDI (Ground truth) vs. U-FNO predicted dose rates across all grids and time steps, inside the plume domain, for all test samples. For better visualization, we use contours to represent the density of the data points
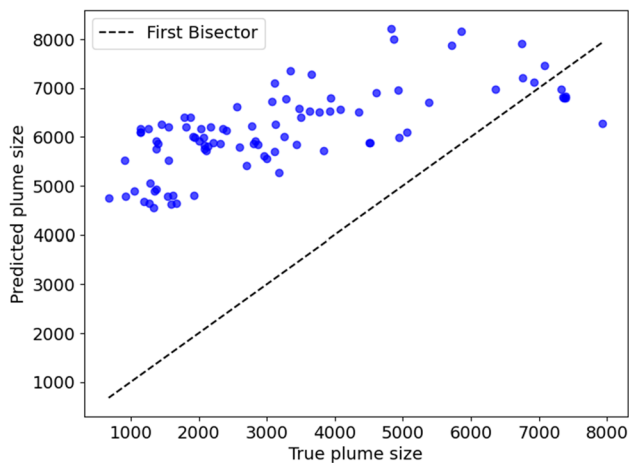

Fig. 7 Scatter plot of the WSPEEDI (true plume size) vs the U-FNO (predicted plume size) for all test samples 6 h after the release ($T_r + 6h$)

5 Computational efficiency

The first steps to set up the emulator framework necessitate an upfront investment in computational resources for simulation, data preparation, and model training. However, the true payoff becomes apparent when considering the significantly increased speed of simulation reproductions once these preparatory stages are done. Training the U-FNO architecture using the 524 data points takes about 17 min for 140 epochs on an Nvidia A100-SXM GPU.

To assess the computational efficiency speed-up, we compare the physical simulation CPU run time with each machine learning model's testing time (i.e., a single model prediction). The run time for a WSPEEDI simulation (WRF and GEARN) on an Intel Xeon Gold 6240 CPU averages about 82 min for a 6-h forecast. On the other hand, when we utilize the pre-trained U-FNO model for inference, the average inference time is just 1.26 s on an Intel Xeon CPU. This represents a 4000-fold speedup, enabling the rapid generation and analysis of a large number of potential release scenarios and their associated consequences.

6 Discussion and conclusions

This study aims to utilize an emulator for rapid forecasting of radionuclide atmospheric transport in case of an accident based on pre-existing simulations and training. Specifically, we utilize a U-Net enhanced Fourier Neural Operator (U-FNO) architecture to emulate the dose forecasting model, WSPEEDI. WSPEEDI includes forecasting the meteorological fields in the domain of interest and simulating particle transport and dispersion using a Lagrangian framework. The developed emulator addresses a significant limitation of current ATD tools by eliminating the necessity for real-time execution of costly and time-consuming physical simulations. This advancement could allow prompt guidance of protective actions during emergencies and the assessment of contaminant impacts on the environment. Although there have been several ML-based developments to emulate computationally intensive physical phenomena, to the best of the authors' knowledge, this is the first study to apply emulators to weather forecasting and particle transport, and at such a high spatial and temporal resolution.

Our study adds to the growing evidence of the effectiveness of U-FNO (Wen et al. 2022; Meray et al. 2024; Li et al. 2023). The results demonstrate that the trained U-FNO model can effectively capture the spatiotemporal behavior of the plume, irrespective of its direction, width/spread, or extent. This is evident in the consistently low average RMSEs and high R^2 scores across the entire map, as well as within the plume area. These metrics affirm the model's capability to capture a significant portion of the variance inherent in the WSPEEDI simulated data, encompassing the spatial distribution, plume directionality, extent, and differentiation between high and low dose zones. Furthermore, the U-FNO model's performance remained consistent across the forecast time steps. This can be attributed to the used training methodology, where the model predicts all time steps concurrently, as opposed to employing a one-step-at-a-time or recurrent prediction approach.

We identify some issues in the results of the emulator, specifically involving the underestimation of dose rates during the wet depositions and overestimating the plume size.

Wet depositions happen rarely in the training sets, since they require the simultaneous occurrence of rain and a release (plume), which happens in very few cases. Expanding the number of training sets and incorporating supplementary predictor variables in the U-FNO training is crucial for encompassing the underlying physical processes associated with deposition. At the same time, rain forecasting is relatively straightforward to obtain in ordinary weather forecasting (Ashok and Pekkat 2022). In the real application, we might be able to advice the evacuation by combining the plume direction from our emulator and the rain forecast. Moreover, we observe some overestimation of the plume size (i.e., the extent of non-zero dose rates), which have been also reported in previous studies (Berendt-Marchel and Wawrzynczak 2021). This overestimation could be attributed to the uncertainty associated with the low number of particles in the plume edge during the particle tracking simulations, as well as to the Fourier transform's tendency to smooth spatial variations, potentially blurring sharp boundaries that distinguish the true plume from non-plume areas. In the real application, the plume extent threshold needs to be determined based on regulatory decision threshold of the dose rates, given the source term. In addition, the slight overestimation could be acceptable and on the safe side during a nuclear accident.

The trained U-FNO model forecasts plume- and dose-distribution time series 4000 times faster than a numerical simulator, while still maintaining a high accuracy. This rapid and reliable forecasting capability can be instrumental in advising the timely initiation of triggered releases (e.g., containment venting) during severe accidents, reducing public exposure (Ayoub et al. 2024). Its speed and efficiency facilitate a swift exploration of a wide array of emergency scenarios in real-time, aiding risk-informed decisions for protective actions, evacuation execution, and zone delineation. Furthermore, this computational advancement enables various engineering tasks that rely on costly repetitive numerical simulations, especially those involving uncertainty quantification (UQ) for radiological risk assessment. It is important to note that while this study focuses on radioactive releases, the emulator's applicability extends to various contaminant atmospheric release scenarios.

Finally, we acknowledge that this paper serves as the proof-of-the-concept study to show the utility of FNO in ATD applications. Future studies should explore different types of neural network architectures as well as different parameters and variables, including the number of simulations, mesh size, and grid resolution. Additionally, while we focused on predicting plume transport over the next 6 h, future research could consider different forecast periods and examine various release modes. Nevertheless, it is worth mentioning that during the Fukushima NPP accident, the primary contamination area from the Unit 2 reactor was created within a few hours only (Ayoub et al. 2024). Furthermore,

the method should be tested and validated using real-world datasets (e.g., (Gowardhan et al. 2021)).

Author contributions A.A, H.M.W, and G.S conceptualized the study. A.A. conducted the literature review, ran the simulations, prepared the datasets, developed the methodology, and wrote the initial manuscript. A.A. and H.M.W. analyzed and discussed the results. L.W helped in adapting the UFNO code architecture to fit the current application. All authors read and approved the final manuscript.

Declarations

References

Amagasa T, Kitagawa H, Komano T (2007) Constructing a web service system for large-scale meteorological grid data. In: Third IEEE International Conference on e-Science and Grid Computing, pp 118–124

Ashok S, Pekkat S (2022) A systematic quantitative review on the performance of some of the recent short-term rainfall forecasting techniques. J Water Clim Chang 13(8):3004–3029

Ayoub A, Wainwright H, Sansavini G (2023) Machine learning-enabled weather forecasting for real-time radioactive transport and contamination prediction. Prog Nucl Energy 173:105255

Ayoub A, Wainwright H, Sansavini G, Gauntt R, Saito K (2024) Resilient design in nuclear energy: critical lessons from a cross-disciplinary review of the Fukushima Dai-ichi nuclear accident. iScience 27(4):109485

Beck J, Brown J, Dudhia J, Gill D, Hertneky T, Klemp J, Wang W, Williams C, Hu M, James E, Kenyon J, Smirnova T, Kim J-H (2020) An evaluation of a hybrid, terrain-following vertical coordinate in the WRF-based RAP and HRRR models. Weather Forecast 35(3):1081–1096

Berendt-Marchel M, Wawrzynczak A (2021) Does the zero carry essential information for artificial neural network learning to simulate the contaminant transport in urban areas? J Phys: Conf Ser 2090(1):012027

Bradley MM (2007) NARAC: an emergency response resource for predicting the atmospheric dispersion and assessing the consequences of airborne radionuclides. J Environ Radioact 116–121

Chernov D, Sornette D (2020) Critical risks of different economic sectors. Based on the Analysis of More Than 500 Incidents, Accidents and Disasters

Chernov D, Sornette D, Sansavini G, Ayoub A (2022) Don't tell the boss!: how poor communication on risks within organizations causes major catastrophes. Springer

Desterro F, Santos M, Gomes K, Heimlich A, Schirru R, Pereira C (2020) Development of a Deep Rectifier Neural Network for dose prediction in nuclear emergencies with radioactive material releases. Prog Nucl Energy 118:103110

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(7):3484–3496

Gowardhan A, McGuffin D, Lucas D, Neuscamman S, Alvarez O, Glascoe L (2021) Large Eddy simulations of turbulent and buoyant flows in urban and complex terrain areas using the Aeolus model. Atmosphere 12(9):1107

Höhlein K, Kern M, Hewson T, Westermann R (2020) A comparative study of convolutional neural network models for wind field downscaling. Meteorol Appl 27(6)

Kadowaki M, Nagai H, Terada H, Katata G, Akari S (2017) Improvement of atmospheric dispersion simulation using an advanced meteorological data assimilation method to reconstruct the spatiotemporal distribution of radioactive materials released during the Fukushima Daiichi Nuclear Power Station accident. Energy Procedia 208–215

Kocijan J, Hvala N, Perne M, Mlakar P, Grašič B, Božnar M (2023) Surrogate modelling for the forecast of Seveso-type atmospheric pollutant dispersion. Stoch Environ Res Risk Assess 37(1):275–290

Lam R, Sanchez-Gonzalez A, Willson M, Wirnsberger P, Fortunato M, Pritzel A, Battaglia P (2022) GraphCast: Learning skillful medium-range global weather forecasting. arXiv preprint arXiv:2212.12794

Leelőssy Á, Lagzi I, Kovács A, Mészáros R (2018) A review of numerical models to predict the atmospheric dispersion of radionuclides. J Environ Radioact 20–33

Li Z, Kovachki N, Azizzadenesheli K, Liu B, Bhattacharya K, Stuart A, Anandkumar A (2020) Fourier neural operator for parametric partial differential equations. arXiv preprint arXiv:2010.08895

Li Z, Peng W, Yuan Z, Wang J (2023) Long-term predictions of turbulence by implicit U-Net enhanced Fourier neural operator. Phys Fluids 35(7)

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

Meray A, Wang L, Kurihana T, Mastilovic I, Praveen S, Xu Z, Memarzadeh M, Lavin A, Wainwright H (2024) Physics-informed surrogate modeling for supporting climate resilience at groundwater contamination sites. Comput Geosci 183:105508

Nagai H, Furuta Y, Nakayama H, Satoh D (2023) Quantitative visualization of a radioactive plume with harmonizing gamma-ray imaging spectrometry and real-time atmospheric dispersion simulation based on 3D wind observation. J Nucl Sci Technol, pp 1–16

Osborn DM, Aldemir T, Denning R, Mandelli D (2013) Seamless Level 2/Level 3 dynamic probabilistic risk assessment clustering. In: ANS PSA 2013 International Topical Meeting on Probabilistic Safety Assessment and Analysis

Pascucci-Cahen L (2024) The Economic Costs of a Nuclear Accident. In: Nuclear Economy 2: Nuclear Issues in the Energy Transition, J. Percebois and N. Thiolliere, Eds., Wiley, p 137

Pu Z, Kalnay E (2019) Numerical weather prediction basics: Models, numerical methods, and data assimilation. In: Duan Q, Pappenberger F, Wood A, Cloke H, Schaake J (eds) Handbook of Hydrometeorological Ensemble Forecasting. Springer, Berlin, Heidelberg

Sato Y, Takigawa M, Sekiyama T, Kajino M, Terada H, Nagai H, Kondo H, Uchida J, Goto D, Quélo D, Mathieu A (2018) Model intercomparison of atmospheric 137Cs from the Fukushima Daiichi Nuclear Power Plant accident: Simulations based on identical input data. J Geophys Res: Atmos 123(20):11–748

Skamarock W, Klemp J, Dudhia J, Gill D, Liu Z, Berner J, Wang W, Powers J, Duda M, Barker D, Huang X (2005) A Description of the Advanced Research WRF Version 2

Stohl A, Forster C, Frank A, Seibert P, Wotawa G (2005) The Lagrangian particle dispersion model FLEXPART version 6.2. Atmos Chem Phys 5(9):2461–2474

Sugawara SE, Juraku K (2018) Post-Fukushima controversy on SPEEDI system: contested imaginary of real-time simulation technology for emergency radiation protection. In: The Sociotechnical Constitution of Resilience. Singapore, Palgrave Macmillan, pp 197–224

Terada H, Chino M (2008) Development of an atmospheric dispersion model for accidental discharge of radionuclides with the function of simultaneous prediction for multiple domains and its evaluation by application to the Chernobyl nuclear accident. J Nucl Sci Technol 45(9):920–931

Terada H, Nagai H, Tsuduki K, Furuno A, Kadowaki M, Kakefuda T (2020) Refinement of source term and atmospheric dispersion simulations of radionuclides during the Fukushima Daiichi Nuclear Power Station accident. J Environ Radioact 213:106104

Wang J, Balaprakash P, Kotamarthi R (2019) Fast domain-aware neural network emulation of a planetary boundary layer parameterization in a numerical weather forecast model. Geosci Model Dev 12(10):4261–4274

Wen G, Li Z, Azizzadenesheli K, Anandkumar A, Benson S (2022) U-FNO—An enhanced Fourier neural operator-based deep-learning model for multiphase flow. Adv Water Resour 163:104180

Wengert A (2017) JRodos: an off-site emergency management system for nuclear accidents. Karlsruhe Institute of Technology (KIT), Karlsruhe

Zhang K, Zuo Y, Zhao H, Ma X, Gu J, Wang J, Yang Y, Yao C, Yao J (2022) Fourier neural operator for solving subsurface oil/water two-phase flow partial differential equation. SPE J 27(03):1815–1830

Zhuang S, Fang S, Goto D, Dong X, Xu Y, Sheng L (2023) Model behavior regarding in-and below-cloud 137Cs wet scavenging following the Fukushima accident using 1-km-resolution meteorological field data. Sci Total Environ 872:162165

Zhuang S, Fang S, Xu Y, Goto D, Dong X (2024) Wet scavenging of multi-mode 137Cs aerosols following the Fukushima accident: Size-resolved microphysics modeling with observed diameters. Sci Total Environ 917:170287