environments

MDPI

Article

A Hybrid CFD–ML Approach for Rapid Assessment of Particle Dispersion in a Port-Industrial Environment

Alejandro González Barberá [1,2], Raheem Nabi [1], Aina Macias [1], Guillem Monrós-Andreu [1] and Sergio Chiva [1,*]

[1] Department of Mechanical Engineering and Construction, Universitat Jaume I, 12006 Castellón de la Plana, Spain; gonzalal@uji.es (A.G.B.); nabi@uji.es (R.N.); amacias@uji.es (A.M.); gmonros@uji.es (G.M.-A.)
[2] Department of Computer Science and Engineering, Universitat Jaume I, 12071 Castellón de la Plana, Spain
* Correspondence: schiva@uji.es

Abstract

Airborne dust emissions from bulk cargo handling in port terminals can degrade local air quality, but traditional dispersion models are often too slow or coarse to support rapid operational decisions. There is thus a pressing need for efficient tools that retain the spatial detail of CFD while enabling near-real-time scenario evaluation. In this work, we develop and test a hybrid framework that couples an RANS-based CFD model of dust dispersion with a neural network surrogate to rapidly predict exposure patterns for a bulk terminal under variable wind and operational conditions. The ML surrogate model, based on a decoder-style Multilayer Perceptron (MLP) architecture, processes two-dimensional slices of dispersion fields across particle diameter classes, enabling predictions in milliseconds with an acceleration factor of approximately 8×10^6 over traditional CFD while preserving high fidelity, as validated by performance metrics such as the F_1 score and precision values exceeding 0.8 and 0.76, respectively. This approach not only addresses computational inefficiencies but also lays the groundwork for real-time air-quality monitoring and sustainable urban planning, potentially integrating with digital twins fed by live weather data.

Keywords: machine learning; CFD; OpenFOAM; particulate matter; port areas

Check for updates

Academic Editors: Evangelos Bakeas and Yonghang Lai

Received: 19 November 2025
Revised: 17 December 2025
Accepted: 24 December 2025
Published: 31 December 2025

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Ports are major contributors to urban air pollution, particularly through the handling of dry bulk commodities like mineral concentrates, coal, and grains at land–sea interfaces, where vessel traffic, cargo operations, and heavy-duty vehicles elevate particulate matter (PM) levels, including mineral dust and trace metals, especially under dry windy conditions in Mediterranean and Atlantic harbors [1,2]. Studies, such as those in Alicante [3], attribute significant PM10 portions to bulk-material activities over shipping emissions, while shipping disproportionately affects ultrafine particles (UFPs) that impact inland areas via sea breezes [4,5]. EU and Spanish policies increasingly integrate air-quality mandates into port governance, promoting techniques like shore power and dust management, yet local controls are essential for UFPs that are not covered by mass-based standards. Recent regulations, including stricter PM limits by 2030 and occupational silica restrictions, demand enhanced monitoring and source identification, but the current fixed sensors and Gaussian dispersion models lack the resolution for short-term peaks and complex urban–port dynamics, hindering compliance and targeted mitigation in port–city settings.

Within bulk material, loading and unloading of trucks are recurrent high-intensity PM sources with complex spatiotemporal behavior. Dust is generated during tipping, hopper discharge, and transfers, where gravitational fall, impact, and induced turbulence entrain fines; subsequent truck accelerations and braking on silt-laden pavements promote vigorous non-exhaust resuspension that can dominate short-term peaks at yards and port–city boundaries, especially under unfavorable meteorology [6,7]. Downwind of mineral bulk material, trace-metal fingerprints in deposited dust and soils (e.g., Pb, Ni, Cd, and As) corroborate a port signature in urban areas, reinforcing the need for control at transfer points and truck loading bays [8]. Engineering controls—encapsulated or aspirated hoppers, covered conveyors or telescopic spouts, fog/mist cannons, targeted wetting of haul roads, and windbreaks—can deliver substantial but context-dependent reductions; for example, controlled experiments over ore piles show that mist generators can reduce PM_{10} by ∼70–80% under favorable operating parameters, while road-dust management reduces non-exhaust contributions but requires sustained maintenance and meteorology-aware operation [9]. At the worker scale, grain loading into holds and hoppers exposes stevedores to high dust and bioaerosol concentrations, adding an occupational dimension to control priorities [10]. Because emissions and exposure depend strongly on wind regimes, terminal geometry, materials, and the choice and placement of abatement technologies, port-specific diagnostic and prognostic tools are needed.

To address these challenges, advanced simulation techniques boosted by real-time data and high-performance computing have become indispensable for devising mitigation strategies and supporting sustainable planning at the urban–port interface [11]. Computational Fluid Dynamics (CFD) provides a robust framework for resolving airflow and pollutant dispersion in complex geometries, capturing the effects of buildings, terminal infrastructure, and local roughness on turbulent transport [12]. In parallel, the rapid evolution of Industry 4.0 and smart-city paradigms is driving the integration of CFD with machine learning (ML) to build surrogate models that are capable of emulating high-fidelity simulations at a fraction of their computational cost [13,14]. Recent studies have demonstrated that deep learning architectures—including three-dimensional convolutional neural networks, residual neural networks, and physics-informed neural networks—can reproduce key flow and dispersion features while achieving orders-of-magnitude acceleration [15–22]. These advances pave the way for near-real-time environmental diagnostics, scenario exploration, and decision support in urban air-quality management [23,24].

Nevertheless, traditional CFD remains computationally expensive for routine or real-time applications, and existing CFD–ML hybrids often lack a specific focus on particulate emissions from port bulk material. In particular, there is limited work on (i) digital twins that integrate high-resolution geometric reconstructions of port infrastructure with (ii) particle-resolved dispersion simulations and (iii) surrogate models optimized for millisecond-scale inference over meteorological and operational scenarios relevant to regulatory diagnostics [25–28]. Data scarcity for robust training and the absence of standardized end-to-end workflows further constrain operational deployment in real port–city environments.

Bulk-material handling operations—such as truck loading, unloading, and on-site stockpiling—are therefore a priority application for hybrid CFD–ML approaches. In the Port of El Grao (Castellón de la Plana, Spain), the materials most frequently handled include ceramic raw materials (such as clays, feldspars, and kaolin), mineral powders, and petroleum coke, all characterized by broad granulometric distributions with a significant fraction of respirable particles below 10 μm and, in certain cases, containing crystalline silica. Their low moisture content and high friability increase their propensity to become airborne under moderate wind speeds, while their density and irregular morphology influ-

ence their settling behavior and potential for long-range transport. These physicochemical properties, combined with the operational dynamics of bulk handling, make these materials particularly relevant for studying particulate dispersion and assessing potential exposure in adjacent urban areas. Under certain meteorological conditions, the resulting particulate plumes can be transported toward nearby residential areas, posing environmental and public-health concerns.

While high-fidelity CFD can resolve the complex aerodynamics of port environments, its computational cost renders it impractical for routine or real-time assessment. The existing hybrid CFD–ML approaches, although promising, seldom address the specific need for particle-resolved dispersion surrogates that integrate high-resolution port geometry and are optimized for millisecond inference across meteorological scenarios relevant to regulatory diagnostics. To address this issue, the present work introduces a hybrid approach that integrates high-fidelity CFD with surrogate ML techniques to evaluate particle dispersion in a port-industrial environment with nearby population exposure. The methodology provides substantial computational acceleration while preserving the physical fidelity necessary for environmental impact assessment. The key contributions of this study are as follows:

- **Port-Industrial Digitalization and Mesh Generation:** A high-resolution 3D digital model of the Port of El Grao (Castellón de la Plana, Spain) was developed using LiDAR data, cadastral information, and CAD processing, enabling a geometrically accurate representation of port infrastructure.
- **Integrated CFD–ML Workflow:** A workflow combining OpenFOAM-based aerodynamic and particulate transport simulations with a decoder-style ML model architecture is proposed. Horizontal (Z-axis) 2D slices of the particle concentration fields are used to train the surrogate model, with hyperparameters optimized using the Optuna framework for efficient convergence.
- **Performance and Accuracy:** While CFD simulations required 432,000 s for steady state and 108,000 s for transient state on 100 cores, the trained ML surrogate performed inference in milliseconds on a GPU, achieving a computational acceleration factor of approximately 8×10^6. The validation results demonstrate that the ML model reliably reproduces CFD predictions across multiple wind scenarios.

All in all, the key novelty of this work lies in the development of an end-to-end hybrid workflow that combines a high-resolution 3D digital port model derived from LiDAR/cadastral data, particle-resolved RANS–Lagrangian CFD simulations for a realistic bulk terminal, and a tailored ML surrogate capable of millisecond-scale inference of binary dispersion fields for variable wind conditions. This framework is a step toward practical digital twins for port air-quality management.

The rest of the paper is organized as follows. Section 2 describes the methodology employed in this study, from the CFD setup to the ML model design and training. Section 3 presents the results, organized into three subsections: aerodynamics, particle dispersion, and machine learning. Section 4 concludes the paper and discusses future work.

2. Methods

This section describes the methodology employed in the present study, which integrates multiple computational techniques to assess atmospheric particle dispersion. To better illustrate the different steps carried out during the research, Figure 1 depicts the digitalization, dataset simulation, data pre-processing, and ML model training.

The study focuses on the port-industrial district of El Grao in Castellón de la Plana (Spain), a densely integrated coastal environment where heavy industrial activity, residential neighborhoods, and transport infrastructure coexist within a relatively compact area (Figure 2). The port hosts intensive bulk-material handling operations, including the load-

ing, unloading, and storage of mineral powders, ceramic raw materials, and petcoke, all of which are prone to generating fine particulate emissions. These facilities are situated only a few hundred meters from urban residential zones, separated by a heterogeneous built environment composed of warehouses, breakwaters, medium-rise buildings, and open yards that shape complex wind-flow patterns. The proximity between industrial sources and populated areas, combined with the prevalence of onshore sea-breeze regimes characteristic of the western Mediterranean coast, frequently facilitate the transport of suspended particulate matter inland. This configuration makes El Grao an especially representative and challenging scenario for assessing air-quality impacts in port–city interfaces, highlighting the need for high-resolution modeling tools capable of reproducing the intricate aerodynamics and dispersion phenomena that govern exposure levels.

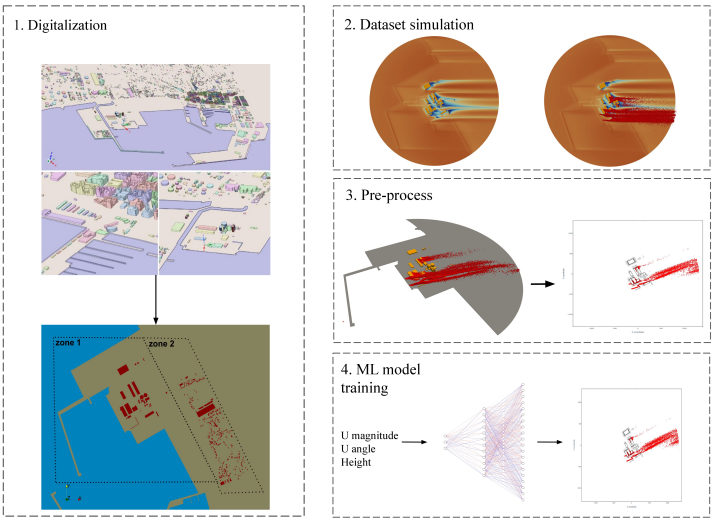


Figure 1. Workflow of the hybrid CFD–ML methodology.

It should be noted that the modeling of the urban area is out of the scope of the current research. The overall approach comprises the following phases: computational model, computational domain and boundary conditions, validation of the aerodynamics, data pre-processing and I/O depiction, and ML model.



Figure 2. Satellite image of the study area showing the port district of El Grao and the adjacent residential and industrial zones.

2.1. Computational Model

In CFD simulations of urban environments, large-eddy simulation (LES) resolves the energy-containing eddies while modeling the subgrid scales, capturing unsteady and intermittent features of urban flows and street-canyon dispersion more faithfully than averaged approaches, but at computational costs that are often orders of magnitude higher—impractical for large parametric datasets across many wind scenarios [29–31]. By contrast, Reynolds-Averaged Navier–Stokes (RANS) models statistically average the equations and represent all turbulence via closures (e.g., k-ϵ), yielding a robust and computationally efficient option widely used in wind engineering [32].

In this study, CFD techniques were employed using OpenFOAM v2212, an open-source CFD toolbox, to model particle dispersion under a neutral atmospheric boundary layer (ABL). The wind flow was resolved using the steady RANS equations with the k-ϵ turbulence model, owing to its validated performance in atmospheric dispersion modeling [33]. This choice enables the simulation of the multiple cases required for dataset construction while remaining feasible within the available computational resources.

For wind flow, the steady-state incompressible turbulent flow solver *simpleFoam* was used. This solver has been designed to solve RANS equations using the SIMPLE (Semi-Implicit Method for Pressure-Linked Equations) algorithm for pressure–velocity coupling.

The simulation of particle transport was carried out using the *icoUncoupledKinematicParcelFoam* solver in OpenFOAM, which is specifically designed for transient modeling of kinematic particle clouds advected by a precomputed flow field. The adopted methodology follows a Euler–Lagrangian approach, where the continuous phase (air) is first resolved using the *simpleFoam* solver under steady-state conditions with the k-ϵ turbulence model within the RANS framework to obtain a converged aerodynamic field. Convergence criteria were set to 10^{-4} for all residuals, a value frequently reported in the literature [34,35]. Subsequently, the discrete phase (particles) is tracked in time using one-way coupling, assuming negligible feedback on the carrier flow. The governing equations account for dominant forces, such as drag, gravity, and pressure gradients, while neglecting inter-particle collisions due to low concentration. This two-step procedure enables a realistic assessment of particle dispersion in complex port environments, capturing the influence of obstacles and turbulence on transport dynamics.

2.2. Computational Domain and Boundary Conditions

To construct the geometry model of the study domain, LiDAR point clouds were obtained from the National Plan for Aerial Orthophotography (PNOA-LiDAR (https://pnoa.ign.es/pnoa-lidar/presentacion (accessed on 23 December 2025))) via the National Geographic Information Center (CNIG). This data was used for classifying the key features of the domain, including buildings, vegetation, and terrain. To further enhance the accuracy of the digital model, cadastral maps from the Spanish Cadastre and Mapping Agency (https://www.sedecatastro.gob.es (accessed on 23 December 2025)) were incorporated to provide detailed parcel and road boundaries.

The integrated processing yielded three raster digital models (DEM, DSM, and nDSM) and two Geopackage layers containing both elevation and height data. A 3D polygonal model was then generated using ArcGIS v3.5.1 Pro®, employing extrusion techniques based on the derived 3D values. Further refinements were accomplished in CAD software (e.g., Blender v4.2.2®) to optimize both the level of detail and computational efficiency (see Figure 3 for further information).

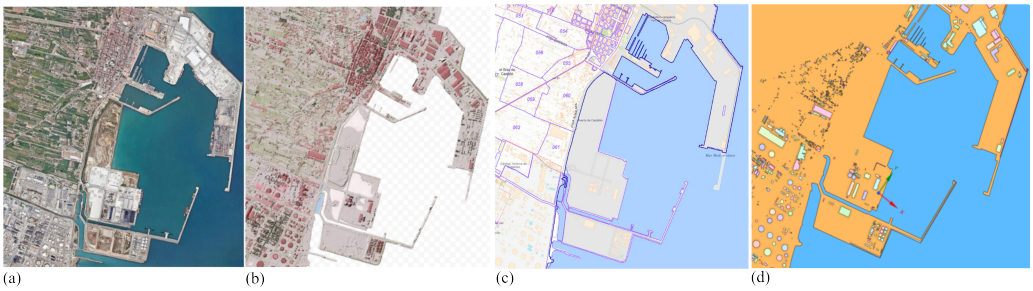


Figure 3. Three-dimensional geometric model generation: (**a**) aerial view of the study area; (**b**) data point LiDAR; (**c**) cadastral data; (**d**) corresponding computational 3D geometry.

The final model comprises three primary geometries: sea, ground, and industrial buildings. Figure 4 schematically depicts the computational domain, where the domain height is $H = 130$ m and the diameter is $D = 2600$ m. Since $H_{max} = 26$ m is the height of the tallest building in the domain, this corresponds to a domain height of $5H_{max}$ and a horizontal extent exceeding $15H_{max}$ in all directions around the built-up area, in line with best-practice guidelines for CFD simulations of flows in urban environments [36,37].

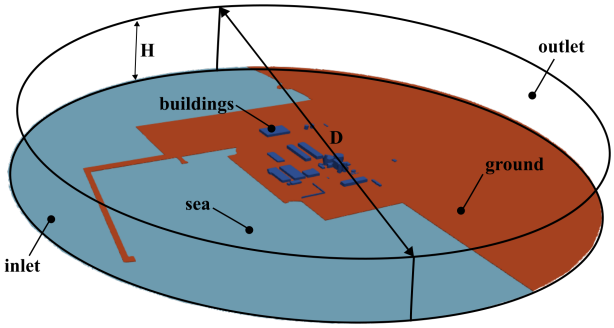


Figure 4. Computational domain used to model the wind flow and particle dispersion. The domain is circular and includes labeled patches for the inlet and outlet boundaries, as well as surfaces representing buildings, sea, and ground.

Figure 4 also illustrates the boundary conditions applied in the simulations. To ensure a homogeneous neutral atmospheric wind profile in the horizontal plane, inflow conditions were implemented in OpenFOAM using the *atmBoundaryLayer* class [38] for streamwise wind velocity (U), turbulent kinetic energy (k), and turbulent dissipation rate (ε). In this context, *atmBoundaryLayerInletVelocity* boundary condition provides a log-law-type inlet boundary condition for the flow speed profile expression:

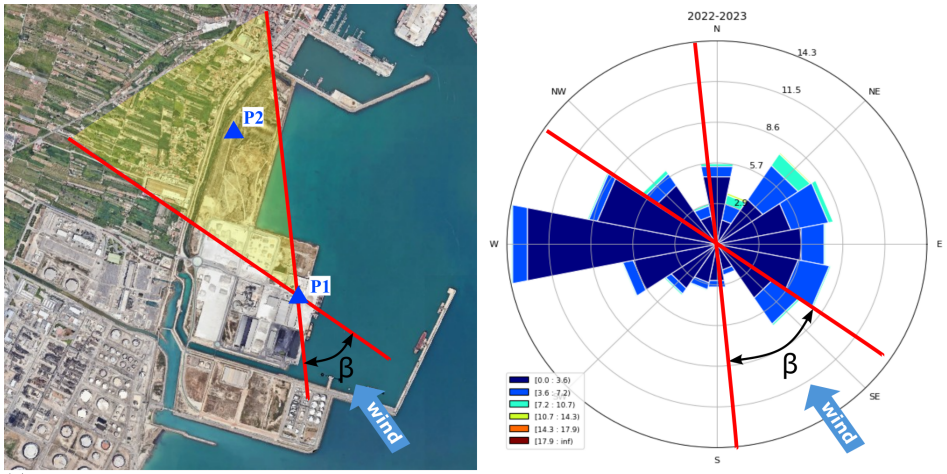
$$U(z) = \frac{u^*}{\kappa} \ln\left(\frac{z + z_0}{z_0}\right) \quad (1)$$

with the following parameters: von Kármán constant $\kappa = 0.40$ (-) and aerodynamic roughness length (z_0) equal to 0.01 m for ground and 0.0002 m for sea according to the classification proposed by [39]. The friction velocity u^* is given by

$$u^* = \frac{\kappa U_{ref}}{\ln\left(\frac{z_{ref}+z_0}{z_0}\right)} \quad (2)$$

where U_{ref} (m/s) is the reference streamwise wind speed at a reference height z_{ref}. To select the U_{ref} values, data from one of the meteorological stations located within the study area and closest to the inlet boundary of the computational domain were used (Figure 5a, P1). This station is located 14 m above the ground, a value that was therefore assigned to z_{ref}.

The wind rose was generated using accumulated data from two consecutive years (2022–2023), (Figure 5b). We selected the wind-direction angles β whose trajectories advect particles toward the urban area (Figure 5a, yellow-shaded sector). In this case, $\beta \in [130°–170°]$, corresponding to winds originating from the south–southeast sector.



(a) (b)

Figure 5. (**a**) Study area showing wind directions (β) that affect the adjacent urban area and the locations of two meteorological stations within the computational domain (P1 and P2); (**b**) corresponding wind-direction sector in the wind rose generated from observations collected during 2022–2023 at meteorological station P1.

To model the turbulent viscosity, wall function was applied to the ground and building surfaces using *atmNutkWallFunction* and *nutkWallFunction*, respectively. As has been commented, the air flow aerodynamics distribution was resolved in the first step. Table 1 presents a summary of all the boundary conditions applied.

Table 1. Details of the boundary conditions imposed in the aerodynamic simulations. Nomenclature: Cc = *calculated*, emp = *empty*, fSS = *fixedShearStress*, fV = *fixedValue*, iF = *inletFunction*, sP = *symmetryPlane*, wF = *wall-Function*, and zG = *zeroGradient*.

	U	p	k	ε	ν_t
inlet	iF	zG	iF	iF	Cc
outlet	zG	fV	zG	zG	Cc
top	fSS	zG	zG	zG	Cc
buildings	fV	zG	wF	wF	wF
ground and sea	fV	zG	wF	wF	wF

The particulate phase in the simulations was modeled using physical properties representative of dust generated during bulk-material handling operations. The dispersion of particulate matter was simulated using a Lagrangian approach based on the *cloud* class in OpenFOAM, with the configuration defined in the file *kinematicCloudProperties*. Particles were treated as inert solid spheres and evolved in a transient one-way coupled manner under the action of gravity and spherical drag. Turbulent dispersion was represented using the *gradientDispersionRAS* model, consistent with the RANS turbulence closure of the carrier flow. Seven independent emission sources (S1–S7) were implemented through the *patchInjection* model, each associated with a specific boundary patch representing an emission area of different bulk-material handling operations (Figure 6).

All sources injected particles following the same size distribution. The choice of a normal particle size distribution centered at 3 µm is motivated by the granulometric

characteristics of the bulk materials handled in the study area. Typical fine fractions of clays (≈2.8 µm) and feldspars (≈3.8 µm) fall within this range, and airborne emissions from these materials are dominated by particles between 2 and 4 µm. To avoid material-specific biases and to retain computational tractability, a representative mean diameter of 3 µm was adopted as a physically grounded surrogate for the fine transport-efficient fraction most relevant to dispersion and exposure. The particle density was assumed to be 2600 kg m^{-3}, consistent with silicate-based minerals.

Particle–wall interactions were handled using a *localInteraction* model: solid surfaces were assigned rebound behavior, the sea surface was treated as perfectly sticking, and open boundaries—including inlet, outlet, lateral faces, and the source baffles—were specified as escape patches, allowing particles to leave the computational domain. In the present configuration, both terrain and building surfaces were treated as rigid walls and assigned the same rebound interaction with high normal and tangential restitution coefficients ($e = 0.97$ and $\mu = 0.97$), reflecting the predominance of paved and hard structural materials in the study area. For the micron-sized particles considered here (mean diameter ≈ 3 µm), size-resolved dry deposition studies indicate that rebound effects mainly influence coarse particles larger than about 5–10 µm, while fine particles are governed by turbulent diffusion, impaction, and gravitational settling and are only weakly sensitive to the exact restitution values [40,41]. Consequently, using uniform rebound coefficients on all solid surfaces provides a computationally efficient and physically reasonable approximation for the fine particulate fraction relevant to this study.

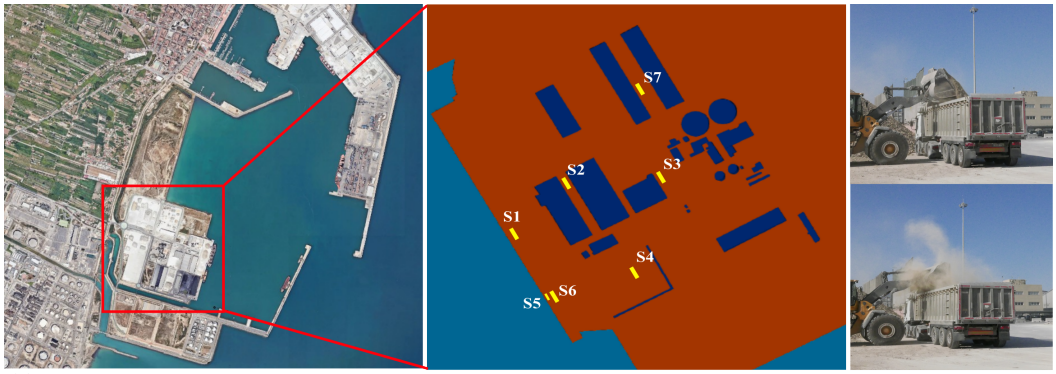


Figure 6. (**a**) Satellite view of the locations of particle sources in study area; (**b**) locations of particle-emission sources (S1–S7) in the computational model; (**c**) example of truck loading and unloading operations.

Regarding meshing, the computational domain was discretized using OpenFOAM's *snappyHexMesh* utility [42]. The mesh was refined around the buildings, resulting in 43 million cells (Figure 7a). The near-wall resolution results yield y^+ values within the logarithmic layer in the range of $30 < y^+ < 300$, ensuring accurate predictions of near-wall turbulence when using wall functions [43,44]. To situate the domain size and problem complexity within the state of the art, previous studies have employed meshes with 1,195,547 elements [23] or domains spanning 100 × 100 m [24], whereas the current domain extends over 1 × 2 km with 43 million cells.

To ensure that the aerodynamic predictions were not sensitive to grid resolution, a mesh-independence analysis was performed prior to the final simulations. Three systematically refined meshes were evaluated (28 M, 43 M, and 60 M cells) by comparing vertical velocity profiles at a reference location upstream of the first buildings. Figure 7b shows that the vertical velocity profiles are consistently reproduced by all three mesh configurations; therefore, the medium mesh was selected for the remainder of the study.

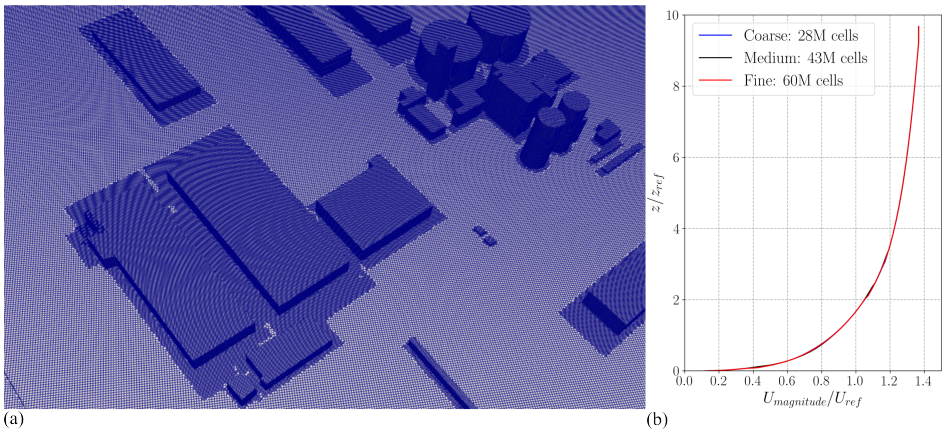


Figure 7. (**a**) Computational grid on the industrial building surfaces and adjacent ground. The minimum cubic edge length is 4.4 cm (due to local refinement), while the maximum cubic edge length is 9.3 m; (**b**) comparison of vertical velocity profiles at a reference location upstream of the first buildings for the three mesh configurations.

2.3. Validation of the Aerodynamics

For aerodynamic validation, data from the meteorological stations located within the study area (Figure 5a) were used for the year 2022. Meteorological station P1 was designated as the reference station as the inlet velocity profiles of the simulations, based on U_{ref}, were selected according to the analysis of the data from this station. Conversely, meteorological station P2 was established as the validation station since it lies within the simulated computational domain but is not in close proximity to the reference station (approximately 1 km away).

Figure 8 shows the complete 2022 year wind data (speed and direction) from both stations. As observed, the temporal evolution of wind speed and direction exhibits a correlated trend at both sites despite the approximately 1 km separation between them, indicating that both stations experience comparable atmospheric forcing.

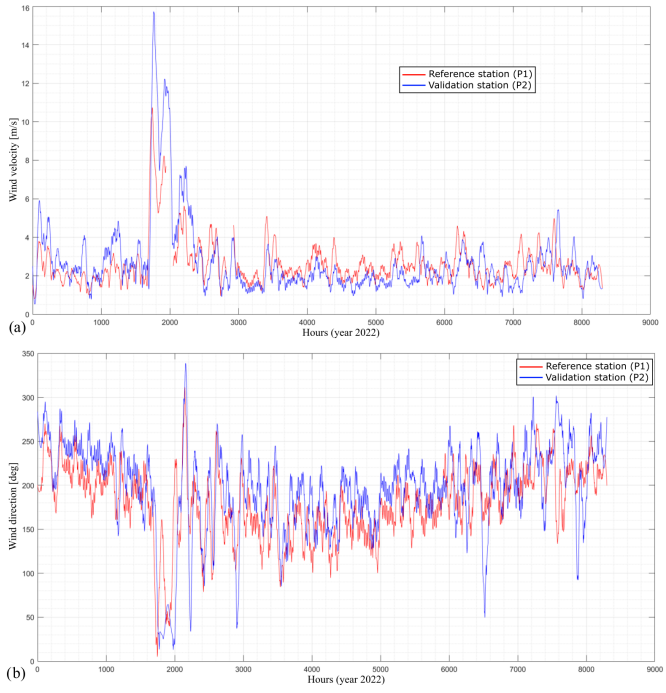


Figure 8. Hourly wind measurements at the reference (P1) and validation (P2) meteorological stations: (**a**) wind speed time series; (**b**) wind direction time series, illustrating the temporal consistency and similarity of atmospheric conditions at both sites.

The simulation used for validation was performed with inlet conditions of U_{ref} = 2 m/s and β = 150°. The corresponding wind velocities at the locations of the two stations (averaged over the nodes adjacent to their exact positions) are presented in Table 2.

Table 2. Input and validation data employed in the aerodynamic CFD simulation, including inlet wind conditions and the corresponding wind speed and direction measurements at the reference (P1) and validation (P2) stations.

	Inlet Wind Condition	Reference Station (P1)	Validation Station (P2)
Velocity (m/s)	2	$\mu = 2.41; \sigma = 0.16$ m/s	$\mu = 1.89; \sigma = 0.08$ m/s
Direction (°)	150	$\mu = 149; \sigma = 0.01$ m/s	$\mu = 147; \sigma = 0.13$ m/s

Based on the inlet wind boundary conditions of the simulation, the corresponding velocity and direction values were selected at the reference station (P1), considering an uncertainty of 10% in both parameters. Subsequently, the corresponding velocity and direction values were also selected at the validation station (P2). Validation was performed by comparing the velocity values obtained at station P1, selected for the same hourly intervals corresponding to station P2 (Figure 9). The results indicate that the average velocity values measured at the validation station and those obtained from the CFD simulation exhibit reasonable agreement for the presented case study.

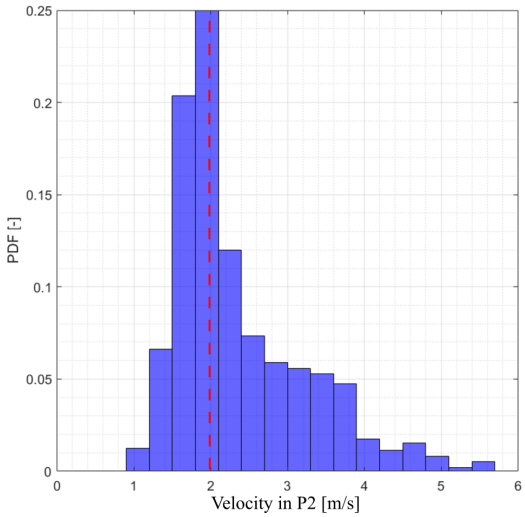


Figure 9. Histogram of wind speed at the validation station (P2). The red dotted line marks the median value (2.06 m/s), with a standard deviation of σ = 0.85 m/s.

2.4. Data Pre-Processing and I/O Depiction

To create the dataset for feeding the subsequent ML models, real-world conditions were emulated using data from the previously introduced meteorological stations (Figure 5a), which show a predominance of south–southeast winds (130° to 170°) that are critical for particle transport into the adjacent urban area where pedestrians are exposed. The central wind directions were systematically adjusted by applying angular offsets from −20° to +20° in 5° increments (9 angles). Moreover, the wind velocity historical data leverages values from 1 m/s to 10 m/s, mostly in the range [3, 10]. Therefore, three wind velocity values U_{ref} (3, 6, and 10 m/s) were selected. The combination of these adjustments resulted in 27 distinct simulation scenarios (9 angle cases for each of the 3 velocities).

The processing pipeline for each 3D CFD simulation was as follows:

1. Vertical Slicing: The original 3D particle concentration fields (see Figure 10a) were decomposed into 14 discrete horizontal 2D slices along the Z-axis (Figure 10b). These slices were extracted at 0.5 m intervals, spanning the critical 3 m to 10 m height range. The vertical slice range of 3 m to 10 m was selected to represent the human inhalation zone, most relevant for pedestrian and ground-level residential exposure assessment in the closest neighborhood, as previous studies in the area have demonstrated.
2. Spatial Downsampling and Gridding: Each 2D slice was mapped to a standardized 1000 × 1000 pixel grid using the `KDTree` utility from `SciPy` (https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.KDTree.html (accessed on 23 December 2025)), which employs a nearest-neighbor interpolation method. This downsampling (Figure 10c) strikes a balance between preserving the resolution of key dispersion features and maintaining computational tractability for model training.
3. Field Binarization: To simplify the learning task and focus on the primary objective of identifying particle presence, the continuous concentration fields were converted into binary masks. A global threshold was defined as the mean particle count across all non-zero grid cells in the dataset. For each cell, a value of 1 was assigned if its particle count exceeded this threshold, indicating particle presence; otherwise, it was set to 0 (Figure 10c). This transformation converts a complex regression problem into a more stable classification task. A sensitivity analysis has been conducted; varying this threshold by ±25% altered the total classified plume area by less than 5%, confirming the robustness of this method for defining the plume's spatial footprint, which is the primary focus of this study.
4. Particle Size Discretization: To capture the distinct dispersion dynamics governed by aerodynamic diameter, the dataset was partitioned into three discrete classes Figure 10d:
   (a) $d < 2\,\mu m$;
   (b) $2\,\mu m \leq d \leq 3\,\mu m$;
   (c) $d > 3\,\mu m$.

This procedure resulted in a final curated dataset comprising 378 independent 2D binary fields per diameter class (27 simulations × 14 slices). The input features for the ML model were the scalar parameters defining each scenario: wind velocity magnitude, wind direction angle, and vertical slice height. The corresponding output was the 1000 × 1000 binary dispersion field, creating a high-dimensional input–output mapping for the surrogate model to learn.

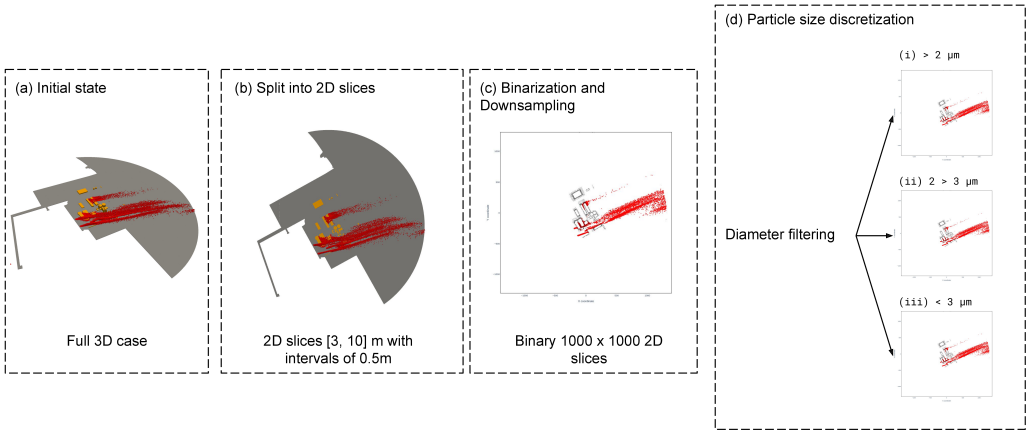


Figure 10. Methodology followed to process the data from the raw 3D CFD simulations to the 2D fields ready to feed the ML model.

2.5. ML Model

To bridge the gap between high-fidelity simulation and real-time assessment, a surrogate modeling framework was developed to emulate the particle dispersion fields derived from CFD. This approach transforms the computationally intensive physics-based problem into a rapid data-driven inference task. The methodology, outlined in Figure 10, encompasses both the data pre-processing from raw CFD and the design of an ML architecture to learn the underlying mapping between input parameters and dispersion patterns.

For each particle diameter category, a surrogate model was developed using a decoder-style MLP architecture. While CNNs are often preferred for spatially structured data, the fully connected MLP was selected for its ability to directly learn the global mapping from a low-dimensional input space (wind speed, direction, and height) to a high-dimensional output (the 1000 × 1000 pixel field), widely used in the literature for the prediction of air quality in urban settings [24]. This task is fundamentally a parameter-to-field regression problem, for which MLPs are well-suited. We acknowledge that CNNs may better capture fine-grained local spatial features (e.g., intricate plume boundaries) but typically require larger training datasets and are more sensitive to input structuring. Given our limited dataset size and the primary objective of predicting the global plume footprint, the MLP offered a more parameter-efficient and stable training pathway. Its efficacy in CFD-related applications has been well documented in prior studies from different fields, such as improving design on latent heat storage tanks [19], studying distribution and pressure drop changes in manifold microchannels [45], researching electric fields on bubble growth in pool boiling [46], surrogate modeling of urban boundary-layer flows [47], and uncertainty-aware surrogate modeling for urban air pollutant dispersion [48]. Recent reviews also highlight MLP and other neural networks as effective surrogates for rapid CFD predictions in built environments [21,49,50]. For further details of the architecture, see Figure 11 for the neural network topology.

Hyperparameter tuning was executed using the Optuna framework [51], which optimized the network by systematically exploring a range of configurations, in this study with the following ranges for the different hyperparameters:

1. Number of hidden layers: [1–5];
2. Number of neurons for each layer: [8–128];
3. Learning rate: [1×10^{-3} down to 5×10^{-5}].

The final configuration of the MLP comprises two hidden layers containing 64 and 128 neurons, respectively; it was trained using the Adam optimizer with a learning rate of 1×10^{-4}, and the loss metric to evaluate the predictions of the model was binary cross-entropy (BCE). Moreover, the dataset was split into 80% for training and 20% for validation. The overall model encompasses approximately 1 million parameters, with a significant portion residing in the final layer responsible for upsampling the latent feature vector to the output grid of 1000 × 1000 pixels. It is noted that the final dense layer, which upsamples the latent vector to the 1 million pixel output, contains the majority of the model's parameters. This design, while effective for the present scope, may pose challenges for generalization and scalability. Indeed, the dataset of 378 samples per diameter class is limited for neural network training and leads to overfitting. To combat overfitting given the high-dimensional output, several measures were employed: a dedicated validation set (20%) including all slices at a completely unseen height (6.5 m), hyperparameter optimization with Optuna, including early stopping based on validation loss, and the use of binary cross-entropy loss, which is robust to class imbalance and architectural simplicity (while it may seem a large model, the hidden layers have only 64 and 128 neurons).

The performance achieved (as detailed in Section 3) validated this architecture for the defined task; however, the authors acknowledge that this could represent a specific solution, and the exploration of more parameter-efficient decoders (e.g., Convolutional or Graph Neural Networks) is a clear direction for future work to enhance generalizability across more diverse urban geometries and flow conditions. A comparative study of MLP vs. CNN architectures for this specific application is a recognized direction for future work to optimize spatial fidelity and generalization.

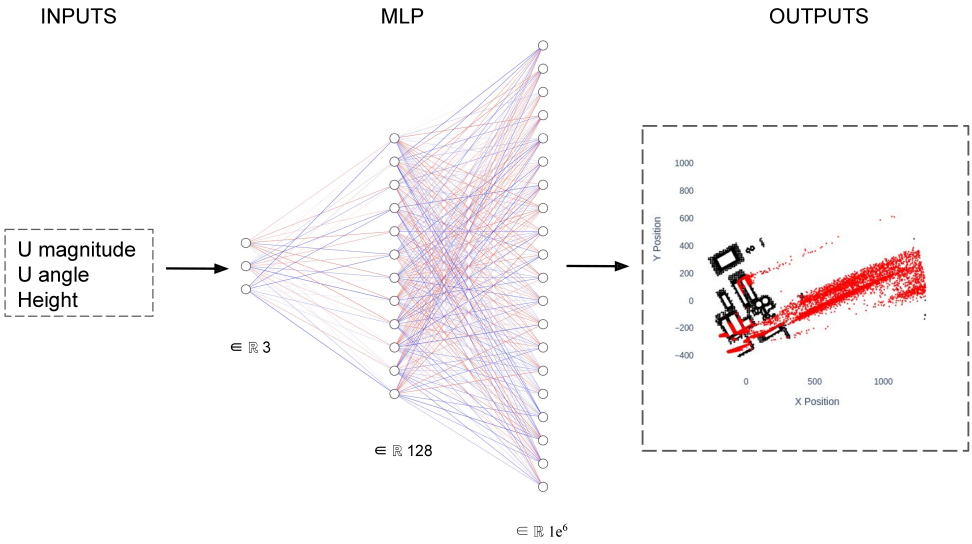


Figure 11. ML model architecture with the expected input and output together with the dimensionality of each of the ML layers.

3. Results

This section describes the outcomes of the integrated computational framework, which is organized into three subsections: Aerodynamics, Particle Dispersion, and Machine Learning.

Experiments were executed on the Marenostrum 5 (MN5) supercomputer (https://www.bsc.es/supportkc/docs/MareNostrum5/overview (accessed on 23 December 2025)) General Purpose Partition (GPP) for CFD simulations and Finisterrae 3 Accelerated Partition (ACC) for ML model training. Experiments on MN5 utilized 2× Intel Xeon Platinum 8480+ 56C 2 GHz, 32× DIMM 64 GB 4800 MHz DDR5, 960 GB NVMe local storage, and ConnectX-7 NDR200 InfiniBand (200 Gb/s bandwidth per node). Experiments on Finisterrae 3 utilized 256 GB RAM (247 GB usable), 960 GB SSD NVMe local storage, 2× NVIDIA A100 GPUs, and 1 Infiniband HDR 100 connection. Software versions include OpenFOAM v2212 and PyTorch 2.6.

3.1. Aerodynamics

The aerodynamic simulations performed on the GPP partition demonstrated robust convergence behavior. The steady-state (RANS) solver, operating on 100 cores at a velocity of 3 m/s, achieved convergence in approximately 432,000 s. Figure 12 presents the converged CFD solution superimposed on the three-dimensional digital model of the port. The velocity contours (left) reveal the main aerodynamic features of the flow field, including regions of high velocity above open areas and pronounced deceleration in the wake of the larger industrial buildings. The particle distribution (right) exhibits strong correlation with these flow structures, with accumulation and stagnation zones forming immediately downstream of obstacles. The figure highlights the complex interplay between building geometry, flow separation, and particle transport that governs near-field dispersion in the study area.

Indeed, Figure 13 quantitatively characterizes the dispersed particle field. Figure 13b compares particle-diameter distributions for inlet velocities of 3 m/s and 6 m/s, showing that higher wind speed extends the upper tail of the distribution, indicating that larger particles remain entrained for longer distances. Figure 13c details the particle count emitted from each source at 3 m/s, confirming that Source 4 dominates due to its larger effective emission area. Figure 13d illustrates the downstream evolution of particle sizes using violin plots; the gradual narrowing of the distributions demonstrates the preferential removal of coarser particles through gravitational settling and impact with building surfaces. Together, these trends emphasize the aerodynamic filtering effect inherent to the flow regime.



Figure 12. An example of the converged CFD solution (**left**) velocity contours and (**right**) particles, overlayed in a 3D model of the urban environment for the port.

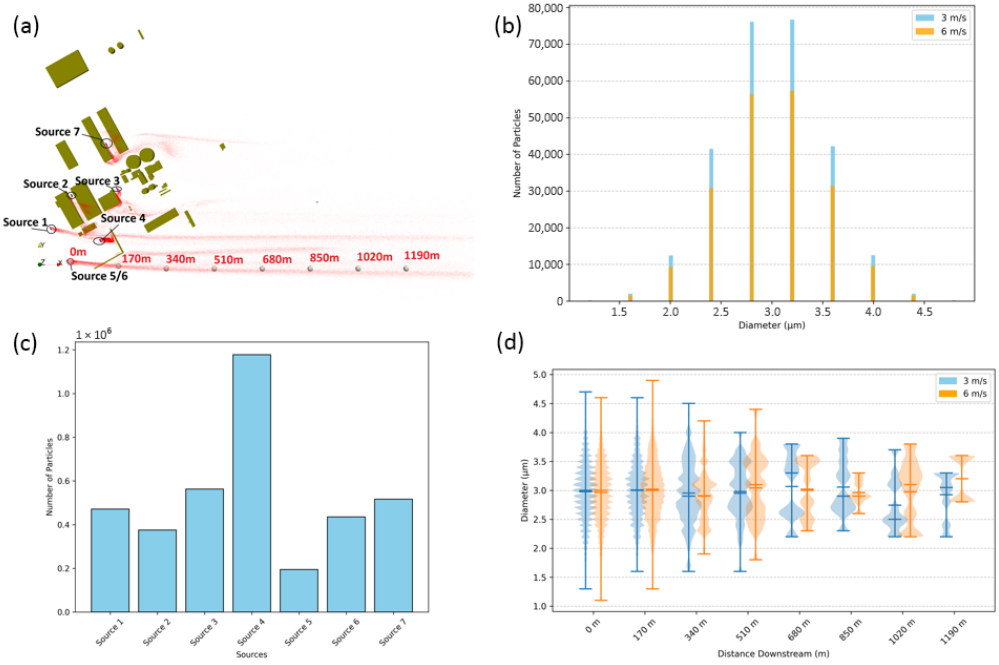


Figure 13. Graphs illustrating (**a**) the different particle source placements; (**b**) the particle source diameter distribution, comparing the differences between velocities of 3 m/s and 6 m/s at point (X = 1190 m, Y = 0 m, and Z = 2 m); (**c**) the particle count at each source location for a velocity of 3 m/s; (**d**) a violin plot depicting the downstream evolution (at positions depicted in (**a**)) of particle size distribution under varying wind speeds.

Furthermore, Figure 14 depicts the detailed velocity-vector field surrounding Sources 3 and 4. The close-up of Source 4 (left) shows a distinct wake region generated by the L-shaped building, where local recirculation traps particles and increases residence time. In contrast, the flow around Source 3 (right) exhibits a combination of shielding and lateral channeling that promotes particle clustering along the building façade. The visualization confirms that localized geometry-induced vortices substantially influence both flow

reattachment zones and subsequent pollutant accumulation patterns within the industrial complex.

These results confirm that the steady RANS approach effectively captures the dominant aerodynamic features governing particle transport in the complex port environment. The agreement with station data, while limited to a single condition, supports the use of this CFD setup for generating training data. However, the inherent averaging in RANS means transient peaks are smoothed, a factor considered when interpreting the ML surrogate's predictions as time-averaged patterns.

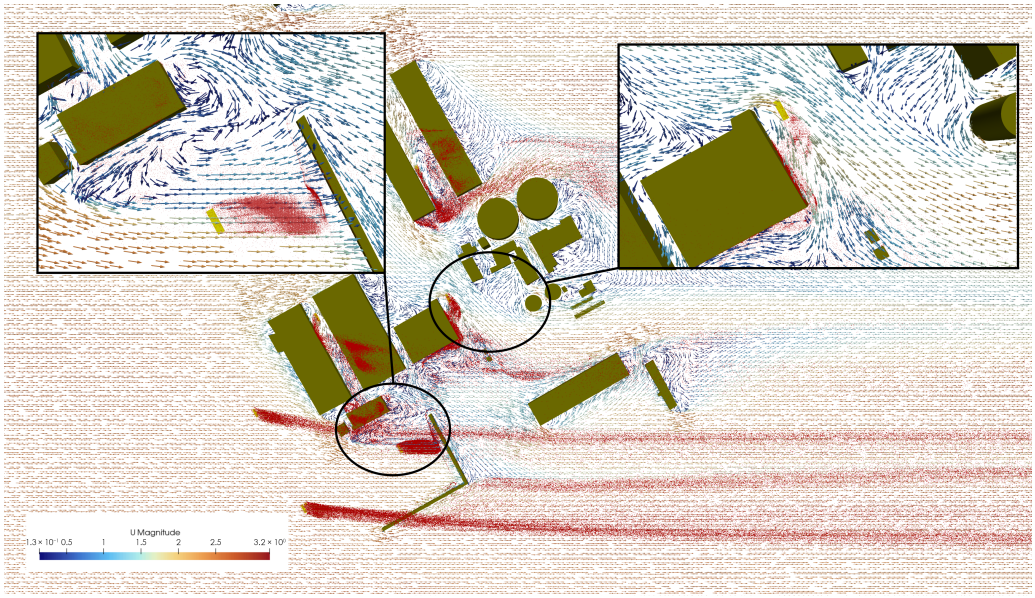


Figure 14. A velocity vector plot of the entire factory site showing close-ups of (upper-left) Source 4 and (upper-right) Source 3, highlighting a combination of effects created by wind shadowing and circulation.

3.2. Particle Dispersion

The particle dispersion simulations, which utilized the converged aerodynamic field as their initial condition, required an average computational time of 108,000 s under similar resource allocations. The lower inlet velocity (3 m/s) yields milder flow separation and comparatively smaller vortex structures, limiting the lateral spread of particles. Conversely, an inlet velocity of 6 m/s amplifies vortex formation, thereby enhancing downstream particle transport and producing a broader dispersion pattern. This trend indicates that a lower wind speed results in a more confined dispersion of particles.

Figure 15 compares the effect of varying wind direction on the near-surface flow field and the associated particle dispersion at a constant inlet velocity of 3 m/s. The top row presents the velocity magnitude contours, whereas the bottom row combines velocity vectors with the corresponding particle trajectories, with building geometries highlighted in orange. The results show that the neutral (0°) inflow produces symmetric wake regions downstream of the main structures, with coherent recirculation cells forming immediately behind the largest buildings. When the inflow is rotated to +15°, the flow deviates toward the leeward side of the domain, generating oblique wake structures and enhanced shear along the lateral façades. In contrast, the −15° case deflects the wake pattern toward the opposite side, producing asymmetric vortices and increased particle accumulation in the sheltered zones on the right-hand portion of the domain. Across all cases, the lateral deflection of the inflow substantially alters the recirculation intensity and the distribution of deposited particles, confirming the strong directional sensitivity of dispersion in dense port-industrial geometries.

The evolution of a representative particle cloud emanating from a source shows that, at the initial positions, both wind speeds exhibit a broad distribution of particle sizes, with a median particle size of approximately 3.0 µm and a range extending from 1.5 µm to 4.8 µm. Between 170 m and 850 m, the distribution narrows; at 170 m, the median particle size remains at 3.0 µm for 3 m/s and increases slightly to 3.1 µm for 6 m/s, with the 6 m/s case showing a wider range (1.6 µm to 4.7 µm) compared to 3 m/s (1.8 µm to 4.5 µm). Downstream between 850 m and 1020 m, the particle size distribution continues to tighten, and, by 1190 m, both wind speeds converge to predominantly smaller particles, with a median size of 2.5 µm for 3 m/s and 2.6 µm for 6 m/s and ranges narrowing to approximately 1.5 µm to 3.5 µm. This indicates an aerodynamic filtering effect, where larger particles settle more quickly while finer particles remain airborne longer. Notably, the faster wind (6 m/s) is capable of sustaining larger particles over a longer distance than the slower wind (3 m/s), as evidenced by the broader distribution observed between 170 m and 850 m for 6 m/s compared to 3 m/s. These findings are summarized in Table 3.

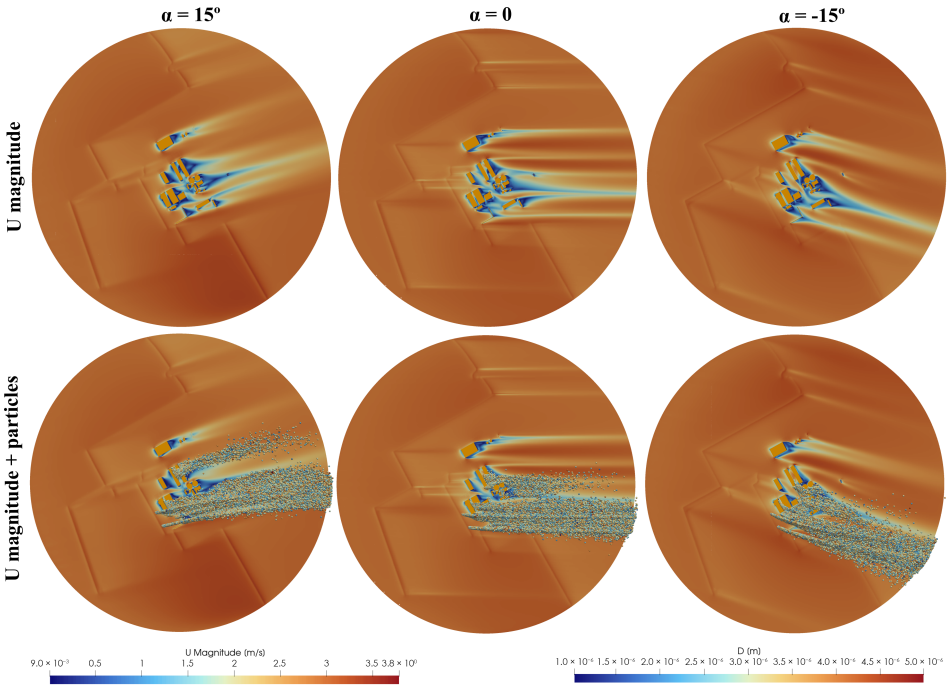


Figure 15. Comparison of flow dynamics and particle dispersion at varying directions ($\alpha = 15°$, $0°$, and $-15°$) and fixed 3 m/s inlet velocity. First row shows the U velocity field magnitude, while second row shows the velocity together with the particle field. Also, the buildings are shown in orange.

The analysis of flow dynamics and particle dispersion shows that, for a velocity of 3 m/s, the flow is observed to be smooth and more laminar around buildings. Wake regions and recirculation zones are seen behind structures, particularly downstream of dense building clusters. The effect of $-15°$ yaw shifts the wake zones slightly rightward, creating asymmetry in the urban shielding effect. The higher velocity (6 m/s) intensifies shear layers and recirculation zones. Stronger gradients appear near building edges, and wakes are more elongated. The $-15°$ yaw shows skews in the recirculation pattern, with stronger flow channeling and lateral turbulence. The particles follow the flow direction closely, showing increased dispersion for a lower velocity of 3 m/s due to lateral diffusion. A more skewed dispersion is seen due to the change in wind direction, with particles spreading more to the right side. The higher velocity (6 m/s) shows a much narrower particle plume due to stronger momentum, reducing lateral spread. The higher momentum also prevents deposition and less particle accumulation in sheltered zones behind buildings. Wake regions and reduced velocity zones show

directional skew in the flow lines due to the -15° yaw. Much stronger and denser vectors are observed at higher velocities with larger vortices and flow channeling in areas where building geometry interacts with shifted wind direction.

Table 3. Quantitative insights from particle size analysis at various locations and wind speeds. The table compares median particle sizes, size ranges, and key observations between 3 m/s and 6 m/s wind conditions.

Location	Parameter	3 m/s	6 m/s
Initial (0 m)	Median Particle Size (μm)	≈3.0	≈3.0
	Range of Particle Sizes (μm)	1.5–4.8	1.5–4.8
	Insight	Similar particle size distribution at the source for both wind speeds.	
Early Downstream (170 m)	Median Particle Size (μm)	≈3.0	≈3.1
	Range of Particle Sizes (μm)	1.8–4.5	1.6–4.7
	Insight	6 m/s wind retains slightly larger particles, with a wider range.	
Midstream (510 m)	Median Particle Size (μm)	≈2.8	≈2.9
	Range of Particle Sizes (μm)	1.5–4.2	1.5–4.5
	Insight	Both wind speeds show removal of larger particles, with 6 m/s retaining them longer.	
Midstream (850 m)	Median Particle Size (μm)	≈2.7	≈2.8
	Range of Particle Sizes (μm)	1.5–3.8	1.5–4.0
	Insight	Larger particles are progressively removed.	
Long-Range Transport (1020–1190 m)	Median Particle Size (μm)	≈2.5	≈2.6
	Range of Particle Sizes (μm)	1.5–3.5	1.5–3.7
	Insight	Both wind speeds carry primarily fine particles (≈2.5 μm). Aerodynamic filtering removes larger particles, stabilizing smaller ones.	

Overall, it can be seen that doubling the wind direction intensifies flow structures, wakes, and particle momentum. Higher wind speeds reduce local accumulation of particles but increase reach and penetration into urban spaces. Furthermore, skewed inflow as a result of a -15° yaw leads to asymmetric wakes and dispersion paths.

Collectively, these results underscore the capacity of the RANS-based CFD framework to capture key aerodynamic features—including shear layers, wakes, and vortex interactions—and their direct influence on contaminant transport in a complex urban environment. The demonstrated variability across wind velocities and directional changes reflects real-world conditions, thereby validating the suitability of this approach for large-scale urban air-quality studies.

The dispersion simulations highlight the strong coupling between wind conditions, building geometry, and particle fate. The observed aerodynamic filtering effect—where larger particles settle out closer to sources—has direct implications for exposure assessment: finer particles penetrate further into urban areas. This detailed physical insight, captured by CFD, provides the high-fidelity ground truth necessary for training a reliable ML surrogate.

3.3. ML Inference

To evaluate the efficacy of the surrogate ML approach, three distinct models—each targeting a specific particle diameter range—were trained for 1.6 h per model on a single GPU. Post-training, these models are capable of performing inferences within milliseconds, offering an approximate acceleration factor of 8×10^6 relative to the mean time of the CFD simulations. This reduction in computational cost transitions the assessment of particulate dispersion from a multi-day high-performance computing task to a near-instantaneous operation, unlocking the potential for real-time scenario analysis and emergency-response planning. It is important to note that this 'real-time' capability applies to inference based on pre-trained models for the specific port geometry and source configuration. Retraining the model for new geometries or major operational changes would require offline CFD data generation.

Before training, as previously introduced, data was split into 80% training and 20% validation. Moreover, all slices corresponding to a Z-height of 6.5 m were exclusively reserved for validation, ensuring that model performance was assessed against previously unseen data. Figure 16 shows the ground truth (which is represented from the processed CFD cases), the prediction of the ML model, and the field of true positives. The evaluation includes three particle dispersion fields at 6.5 m/s, α = 20, 0, and -20 degrees and at 6.5 m height with particle diameter between 2 µm and 3 µm. Across these scenarios, the deviation between prediction and ground truth in total particle count remains below 2% (the number of predicted particles can be seen in the title of each of the plots in Figure 16), underscoring the model's accuracy in reproducing the dispersion patterns.

3.4. ML Evaluation

To quantitatively assess the predictive performance of the ML classifier, a set of confusion matrix-based metrics was computed for each tested scenario, corresponding to different wind-angle offsets (α) with respect to the reference wind direction. Let TP, TN, FP, and FN denote, respectively, the numbers of true positives, true negatives, false positives, and false negatives, where the positive class corresponds to particle presence and the negative class to background conditions.

Based on these quantities, metrics presented in Table 4 were used to evaluate the surrogate models' performance.

Furthermore, accuracy quantifies overall agreement between predictions and labels, while TNR and NPV describe the reliability of background (negative class) predictions. The FPR and FNR provide complementary information on the propensity of the classifier to generate false alarms or to miss particle events, respectively.

Table 5 summarizes the main classification metrics (F_1 and P) for all wind-angle offsets α. The reported values correspond to means across the three surrogate models and all validation cases.

Additionally, a confusion matrix study was conducted (see Figure 17), which provides a detailed breakdown of the model performance. Out of 10^6 evaluated grid points (1000×1000 slices), 187,586 were correctly identified as containing particles (true positive, TP), and 732,413 were correctly classified as particle-free (true negative, TN). The numbers of misclassifications remain limited, with 46,896 false negatives and 33,103 false positives. The derived values in Table 6 quantify these trends: the model exhibits high accuracy ≈ 0.92,

strong specificity ≈ 0.957, and a large NPV ≈ 0.94, demonstrating reliable identification of background regions. Meanwhile, the moderate FNR ≈ 0.20 emphasizes the importance of capturing small particle clusters, which may be more challenging for the model.

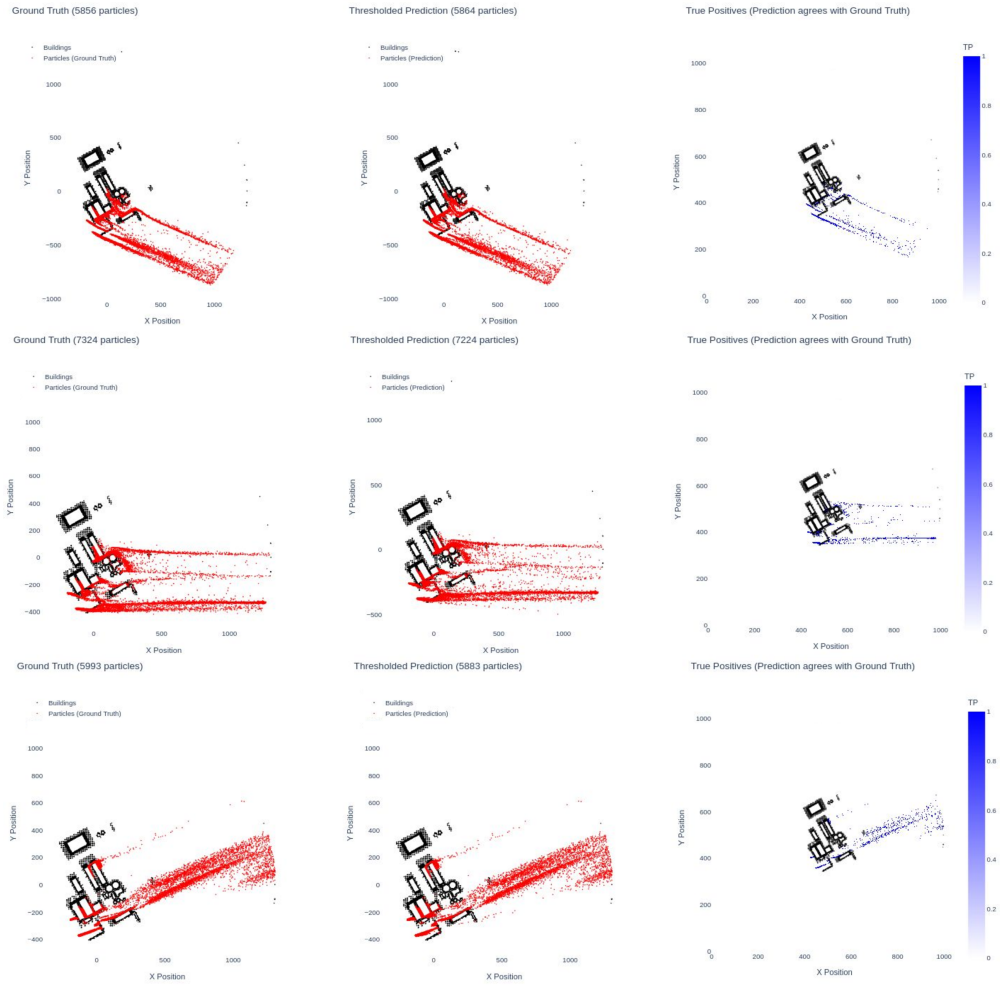


Figure 16. Ground- truth CFD vs. prediction of ML at 6.5 m height slice, 6 m/s input velocity, and angles [20, 0, −20] degrees together with the true positives. The particle diameter represented is between 2 µm and 3 µm.

Table 4. Confusion matrix-derived classification metrics used for performance evaluation.

Metric	Equation	Description
Precision (P)	$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$	Fraction of predicted positives that are correct.
Recall (R)/Sensitivity	$R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$	Fraction of actual positives correctly identified.
F_1 score	$\mathrm{F}_1 = \frac{2\,\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$	Harmonic mean of precision and recall.
Accuracy	$\mathrm{Accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$	Overall agreement between predictions and true labels.
Specificity (TNR)	$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$	Ability to correctly identify negative cases.
Negative Predictive Value (NPV)	$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}}$	Fraction of predicted negatives that are truly negative.
False Positive Rate (FPR)	$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} = 1-\mathrm{TNR}$	Tendency to incorrectly label negatives as positives.
False Negative Rate (FNR)	$\mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{FN}+\mathrm{TP}} = 1-R$	Tendency to miss actual positives.

Table 5. Mean classification metrics (F_1 score and precision) for the predictions of all three surrogate models across different wind angles α. Values are averaged over all validation cases.

α	F_1	Precision
$-20°$	0.82	0.76
$-10°$	0.82	0.85
$0°$	0.84	0.83
$10°$	0.80	0.85
$20°$	0.83	0.85

Table 6. Metrics computed from the aggregated (mean) confusion matrix: TP = 187,586; TN = 732,413; FN = 46,896; FP = 33,103.

Metric	Value
Accuracy	0.9200
True Negative Rate (TNR)	0.9568
Negative Predictive Value (NPV)	0.9398
False Positive Rate (FPR)	0.0432
False Negative Rate (FNR)	0.2

These aggregate results align with the per-angle performance reported in Table 5. The best-performing offsets are observed for $\alpha = -10°, 0°$, and $10°$, whereas the lowest performance occurs at $-20°$ and $20°$. This trend suggests a dependency of predictive fidelity on the relative wind-angle configuration, with mid-range offsets leading to more stable and consistent particle-pattern predictions. Overall, the combination of confusion-matrix metrics and mean per-angle statistics demonstrates that the surrogate ML models effectively capture the essential features of particle dispersion, enabling rapid and reliable assessments for urban air-quality applications.

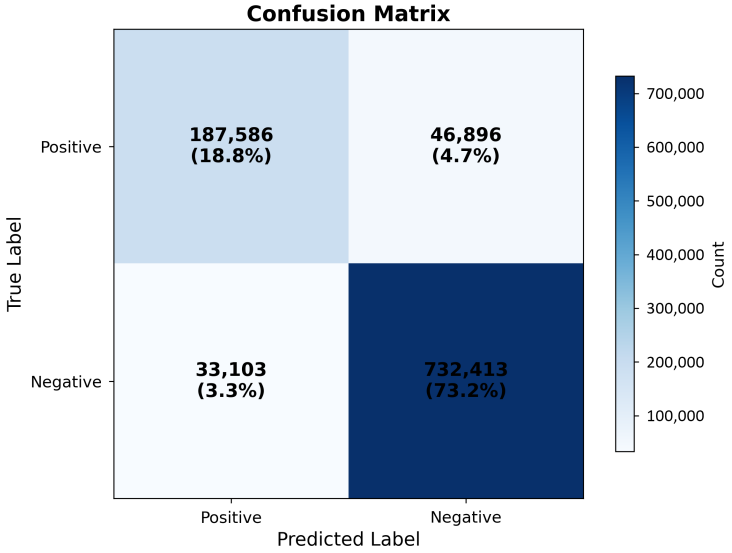


Figure 17. Confusion matrix for the ML model. The results shown represent the mean across the different validation cases for all three surrogate models.

To provide deeper spatial insight into the model's predictive behavior, confusion matrix-derived metrics were computed sectorwise across a 100 × 100 grid for the case with $\alpha = 0°$, wind speed 6.5 m/s, and particle diameters in the range 2–3 µm (middle row of Figure 16). Figure 18 shows the resulting spatial distributions of (a) TNR, (b) NPV, (c) FPR,

and (d) FNR. Both TNR and NPV remain high 0.9 over the vast majority of the domain and only decrease within the particle plume and near the emission sources, where the relative scarcity of true negative instances naturally reduces the number of correct background classifications, yet even there the values typically stay above 0.6–0.8, demonstrating robustness. The FPR is uniformly close to zero across the entire domain, confirming the near absence of spurious (false-alarm) particle detections. In contrast, the FNR being roughly 0.2 suggests a slightly conservative prediction (potentially underestimating the spatial extent of very diffuse contamination). However, the model's high precision (>0.76) and high accuracy in capturing the core plume mean that major exposure hotspots are reliably identified. For emergency response prioritizing acute exposure zones, this performance is adequate. Future models aiming for precise dose–response assessment may require strategies to reduce FNR, such as loss functions weighted towards false negatives. Overall, this sectorwise analysis underscores the surrogate model's reliability in correctly identifying both clean air regions and the core of the plume, with residual discrepancies confined to the most challenging low-density transition discrepancies that are minor in the context of practical urban air-quality and emergency-response applications.

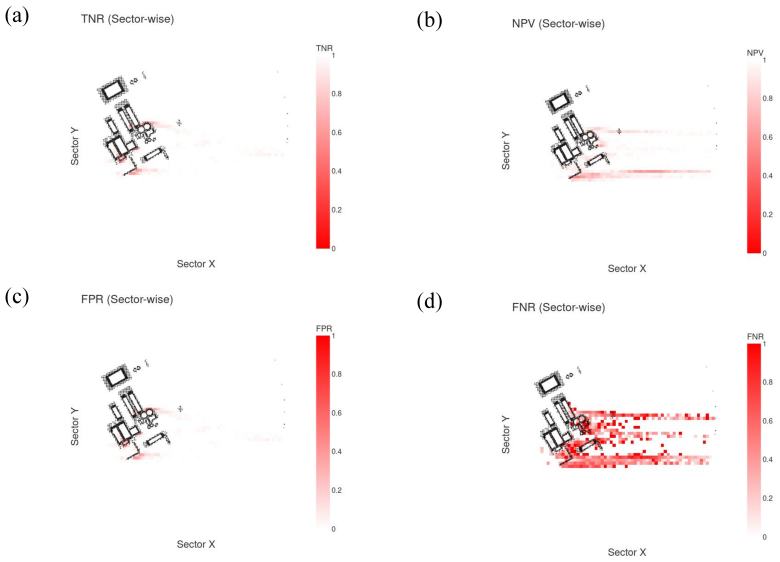


Figure 18. Spatial distribution of confusion matrix-derived metrics for the evaluation case ($\alpha = 0°$, wind speed 6.5 m/s, particle diameter 2–3 µm, and height slice 6.5 m: (**a**) TNR; (**b**) NPV; (**c**) FPR; (**d**) FNR.

The millisecond inference capability represents a paradigm shift from offline simulation to near-real-time diagnostic support. The model's high precision and specificity make it suitable for identifying probable exposure hotspots and clean zones. The elevated FNR at plume fringes suggests a conservative bias, which may be acceptable for emergency response prioritizing core contaminated areas but indicates an avenue for improvement if precise delineation of low-concentration zones is required for regulatory purposes.

4. Conclusions and Future Work

This work introduces a hybrid protocol based on ML for the assessment of particle dispersion in port-industrial environments. The main contribution of this study lies in the implementation of an ML-based prediction methodology that is capable of reproducing particle dispersion patterns with high accuracy and near-instantaneous inference. The developed models, based on a decoder-style MLP architecture, were trained to map input parameters (wind speed, wind direction, and slice height) to high-resolution binary concentration fields (1000 × 1000 pixels). The results demonstrate the following:

- The model achieves robust performance metrics on scenarios unseen during training, confirming its generalization capability.
- The computational cost reduction is remarkable: an acceleration factor of approximately 8×10^6 compared to CFD simulations, reducing multi-day HPC computations to millisecond-scale inference on a GPU.
- This efficiency enables new applications, such as real-time analysis, rapid response to critical pollution episodes, and integration into digital twin systems with live meteorological data.

Despite these promising results, several limitations constrain the current framework and define priorities for future work. First, the CFD setup is based on steady-state RANS with a standard turbulence closure, which cannot fully capture transient intermittent plume dynamics and the associated short-term concentration peaks. Second, the surrogate model is trained for a single port geometry, a fixed layout of emission sources, and a limited range of wind speeds and directions, so its validity is restricted to this specific configuration and cannot be directly generalized to other terminals or operational settings. Third, the training dataset is relatively small compared with the capacity of the neural network, reflecting the high computational cost of generating CFD realizations and raising the possibility of overfitting. Fourth, the binarization of concentration fields, although operationally convenient for diagnosing the presence or absence of plume influence, discards information on intensity and may underestimate uncertainties around regulatory thresholds. Finally, the experimental validation is limited to a small number of field measurements, which, while encouraging, is insufficient to fully characterize model performance across the diversity of meteorological and operational conditions encountered in practice.

Addressing these limitations will guide the next development steps. On the CFD side, extending the framework to unsteady RANS or large-eddy simulations for selected reference cases would improve representations of transient peaks and enable more stringent validation of the surrogate model. Increasing the diversity and size of the training dataset—through additional CFD runs spanning a broader range of wind regimes, emission scenarios, and abatement configurations—would support more robust and transferable ML models. From the ML perspective, exploring the use of Convolutional Neural Networks [52] alongside advancements in Graph Neural Networks (GNNs) [53] could improve knowledge transfer across different urban geometries, thereby increasing the scalability and generalizability of surrogate models. Additionally, physics-informed neural networks (PINNs) [54] will be investigated to ensure that conservation laws are respected, enhancing the physical consistency of predictions beyond a purely data-driven approach.

For port authorities and environmental managers, this workflow could be integrated into a decision support dashboard. By feeding live wind data from port meteorological stations into the trained surrogate model, operators could visualize predicted high exposure zones in near real time, optimize the scheduling of dust-intensive operations (e.g., postponing loading during winds toward residential areas), and dynamically assess the effectiveness of mitigation measures like water spraying. In summary, the proposed CFD–ML workflow demonstrates that high-resolution digital twins of port–city environments can be combined with data-driven surrogates to deliver fast spatially explicit diagnostics of particulate dispersion from bulk-handling operations. With further refinement of turbulence modeling, expansion of the training dataset, and tighter integration with observational networks, such tools can support proactive management of air quality and occupational exposure in port-industrial areas under increasingly stringent regulatory frameworks.

Author Contributions: Conceptualization, S.C. and R.N.; methodology, S.C. and A.G.B.; software, A.G.B. and A.M.; validation, A.M. and G.M.-A.; formal analysis, S.C. and R.N.; investigation, A.G.B., G.M.-A. and A.M.; resources, S.C.; data curation, A.G.B. and R.N.; writing—original draft preparation, R.N., A.M. and A.G.B.; writing—review and editing, S.C. and G.M.-A.; visualization, R.N. and A.M.; supervision, S.C. and G.M.-A.; project administration, S.C.; funding acquisition, S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data available in a publicly accessible repository at https://doi.org/10.1017/jfm.2024.432 (accessed on 23 December 2025).

Acknowledgments: The authors acknowledge RES resources provided by the Barcelona Supercomputing Center on the GPP partition and by Finisterrae 3 on its ACC partition for the IM-2025-2-0025 activity.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Monteiro, A.; Gama, C.; Baldasano, J.M. Shipping emissions in the Iberian Peninsula and impacts on air quality. *Atmos. Chem. Phys.* **2020**, *20*, 9473–9498. [CrossRef]
2. Viana, M.; Hammingh, P.; Colette, A.; Querol, X.; Degraeuwe, B.; Vlieger, I.; de van Aardenne, J. Impact of maritime transport emissions on coastal air quality in Europe. *Atmos. Environ.* **2014**, *90*, 96–105. [CrossRef]
3. Clemente, Á.; Yubero, E.; Galindo, N.; Crespo, J.; Nicolás, J.F.; Santacatalina, M.; Carratalá, A. Quantification of the impact of port activities on PM10 levels at the port–city boundary of a Mediterranean city. *J. Environ. Manag.* **2021**, *81*, 111842. [CrossRef]
4. Contini, D.; Gambaro, A.; Belosi, F.; De Pieri, S.; Cairns, W.R.L.; Donateo, A.; Zanotto, E.; Citron, M. The direct influence of ship traffic on atmospheric PM2.5, PM10 and PAH in Venice. *J. Environ. Manag.* **2011**, *92*, 2119–2129. [CrossRef]
5. Karl, M.; Ramacher, M.O.P.; Oppo, S.; Lanzi, L.; Majamäki, E.; Jalkanen, J.-P.; Lanzafame, G.M.; Temime-Roussel, B.; Le Berre, L.; D'Anna, B. Measurement and modeling of ship-related ultrafine particles and secondary organic aerosols in a Mediterranean port city. *Toxics* **2021**, *11*, 771. [CrossRef]
6. Amato, F.; Alastuey; A.; de la Rosa, J.; Gonzalez Castanedo, Y.; Sánchez de la Campa, A.M.; Pandolfi, M.; Lozano, A.; Contreras González, J.; Querol, X. Trends of road dust emissions contributions on ambient air particulate levels at rural, urban and industrial sites in southern Spain. *Atmos. Chem. Phys.* **2014**, *14*, 3533–3544. [CrossRef]
7. Karanasiou, A.; Amato, F.; Moreno, T.; Lumbreras, J.; Borge, R.; Linares, C.; Boldo, E.; Alastuey, A.; Querol, X. Road dust emission sources and assessment of street washing effect. *Aerosol Air Qual. Res.* **2014**, *14*, 734–743. [CrossRef]
8. Taylor, M.P. Atmospherically deposited trace metals from bulk mineral concentrate port operations. *Sci. Total Environ.* **2015**, *515–516*, 143–152. [CrossRef]
9. Lee, Y.Y.; Yuan, C.S.; Yen, P.H.; Mutuku, J.K.; Huang, C.E.; Wu, C.C.; Huang, P.J. Suppression efficiency for dust from an iron ore pile using a conventional sprinkler and a water mist generator. *Aerosol Air Qual. Res.* **2020**, *22*, 210320. [CrossRef]
10. Marchand, G.; Gardette, M.; Nguyen, K.; Amano, V.; Neesham-Grenon, E.; Debia, M. Assessment of Workers' Exposure to Grain Dust and Bioaerosols During the Loading of Vessels' Hold: An Example at a Port in the Province of Québec. *Ann. Work Expo. Health* **2017**, *61*, 836–843. [CrossRef] [PubMed]
11. World Health Organization. *Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease*; World Health Organization: Geneva, Switzerland, 2016. Available online: https://www.who.int/docs/default-source/gho-documents/world-health-statistic-reports/world-heatlth-statistics-2016.pdf (accessed on 3 October 2025).
12. Gorlé, C.; van Beeck, J.; Rambaud, P.; Van Tendeloo, G. CFD Modelling of Small Particle Dispersion: The Influence of the Turbulence Kinetic Energy in the Atmospheric Boundary Layer. *Atmos. Environ.* **2009**, *43*, 238–252. [CrossRef]
13. Pan, Y.; Zhang, L. Roles of Artificial Intelligence in Construction Engineering and Management: A Critical Review and Future Trends. *Autom. Constr.* **2021**, *122*, 103517. [CrossRef]
14. Charitonidou, M. Urban Scale Digital Twins in Data-Driven Society: Challenging Digital Universalism in Urban Planning Decision-Making. *Int. J. Archit. Comput.* **2022**, *20*, 238–253. [CrossRef]
15. Cuellar, A.; Güemes, A.; Ianiro, A.; Flores, Ó.; Vinuesa, R.; Discetti, S. Three-dimensional Generative Adversarial Networks for Turbulent Flow Estimation from Wall Measurements. *J. Fluid Mech.* **2024**, *991*, A1. [CrossRef]
16. Haasdonk, B.; Kleikamp, H.; Ohlberger, M.; Schindler, F.; Wenzel, T. A New Certified Hierarchical and Adaptive RB-ML-ROM Surrogate Model for Parametrized PDEs. *SIAM J. Sci. Comput.* **2023**, *45*, A1457–A1489. [CrossRef]

17. Zhu, Q.; Liu, Z.; Yan, J. Machine Learning for Metal Additive Manufacturing: Predicting Temperature and Melt Pool Fluid Dynamics Using Physics-Informed Neural Networks. *Comput. Mech.* **2021**, *67*, 619–635. [CrossRef]
18. Vinuesa, R.; Brunton, S.L. Enhancing Computational Fluid Dynamics with Machine Learning. *Nat. Comput. Sci.* **2022**, *2*, 358–366. [CrossRef]
19. Li, Y.; Huang, X.; Huang, X.; Gao, X.; Hu, R.; Yang, X.; He, Y.-L. Machine Learning and Multilayer Perceptron Enhanced CFD Approach for Improving Design on Latent Heat Storage Tank. *Appl. Energy* **2023**, *347*, 121458. [CrossRef]
20. Mendil, M.; Leirens, S.; Novello, P.; Duchenne, C.; Armand, P. A 3D Discrepancy Modeling Framework for Urban Pollution Prediction in Accelerated Time. *Environ. Model. Softw.* **2025**, *194*, 106662. [CrossRef]
21. Mao, R.; Liu, Y.; Li, L.; Liu, Z.; Ma, M.; Yang, T. Rapid CFD Prediction Based on Machine Learning Surrogate Model in Built Environment: A Review. *Fluids* **2025**, *10*, 193. [CrossRef]
22. Bahman Zadeh, Z. Modeling Spatial Distribution of Particles in Transportation Systems Using Computational Fluid Dynamics and Machine Learning Approaches. Ph.D. Dissertation, Drexel University, Philadelphia, PA, USA, 2024. Available online: https://search.proquest.com/openview/77859b05bd1cc850c56a5eab002349f1 (accessed on 23 December 2025).
23. Issakhov, A.; Sabyrkulova, A.; Rysmambetov, N. Prediction of the Air Pollution from Emissions in Idealized Urban Street Canyons Using Machine Learning and Computational Fluid Dynamics (CFD) Methods. *Environ. Model. Assess.* **2025**, *30*, 36. [CrossRef]
24. Wai, K.-M.; Yu, P.K.N. Application of a Machine Learning Method for Prediction of Urban Neighborhood-Scale Air Pollution. *Int. J. Environ. Res. Public Health* **2023**, **20**, 2412. [CrossRef] [PubMed]
25. Lotrecchiano, N.; Sofia, D.; Giuliano, A.; Barletta, D.; Poletto, M. Real-time On-road Monitoring Network of Air Quality. *Chem. Eng. Trans.* **2019**, *74*, 241–246. [CrossRef]
26. Hashad, K.; Gu, J.; Yang, B.; Rong, M.; Chen, E.; Ma, X.; Zhang, K.M. Designing Roadside Green Infrastructure to Mitigate Traffic-Related Air Pollution Using Machine Learning. *Sci. Total Environ.* **2021**, *773*, 144760. [CrossRef]
27. Kek, H.C.; Mesgarpour, M.; Alizadeh, M.; Wongwises, S.; Doranehgard, M.H.; Jowkar, M.; Karimi, N. Particle dispersion for indoor air quality control considering air change approach: A novel accelerated CFD-DNN prediction. *Energy Build.* **2024**, *306*, 113938. [CrossRef]
28. Van Quang, T.; Doan, D.T.; Yun, G.Y. Recent advances and effectiveness of machine learning models for fluid dynamics in the built environment. *Eng. Appl. Comput. Fluid Mech.* **2024**, **18**, 2371682. [CrossRef]
29. Salim, S.M.; Schlünzen, K.H.; Grawe, S. Numerical simulation of dispersion in urban street canyons with avenue-like tree plantings: Comparison between RANS and LES. *Build. Environ.* **2011**, *46*, 1735–1746. [CrossRef]
30. Ismail, W.H.W.; Mohamad, M.F.; Ikegaya, N.; Chung, J.; Hirose, C.; Abd Razak, A.; Azmi, A.M. Comprehensive comparisons of RANS, LES, and experiments over cross-ventilated building under sheltered conditions. *Build. Environ.* **2024**, *254*, 111402. [CrossRef]
31. Rodríguez Berrio, J.F.; Castaño Usuga, F.A.; Correa, M.A.; Rodríguez Cortes, F.; Saldarriaga, J.C. Comparative CFD Analysis Using RANS and LES Models for NOx Dispersion in Urban Streets with Active Public Interventions in Medellín, Colombia. *Sustainability* **2025**, *17*, 6872. [CrossRef]
32. Menter, F.R. Two-equation eddy-viscosity turbulence models for engineering applications. *AIAA J.* **1994**, *32*, 1598-1605. [CrossRef]
33. Dhunny, A.Z.; Samkhaniani, N.; Lollchund, M.R.; Rughooputh, S.D.D.V. Investigation of Multi-Level Wind Flow Characteristics and Pedestrian Comfort in a Tropical City. *Urban Clim.* **2018**, *24*, 185–204. [CrossRef]
34. Jeanjean, A.P.R.; Hinchliffe, G.; McMullan, W.A.; Monks, P.S.; Leigh, R.J. A CFD study on the effectiveness of trees to disperse road traffic emissions at a city scale. *Atmos. Environ.* **2015**, *120*, 1352–2310. [CrossRef]
35. Takano, Y.; Moonen, P. On the influence of roof shape on flow and dispersion in an urban street canyon. *J. Wind. Eng. Ind. Aerodyn.* **2013**, *123*, 107–120. [CrossRef]
36. Franke, J.; Hellsten, A.; Schlünzen, H.; Carissimo, B. Best Practice Guideline for the CFD Simulation of Flows in the Urban Environment. Technology Report COST Action, 2007. Available online: https://hal.science/hal-04181390 (accessed on 23 December 2025).
37. Tominaga, Y.; Mochida, A.; Yoshie, R.; Kataoka, H.; Nozu, T.; Yoshikawa, M.; Shirasawa, T. AIJ Guideline for Practical Applications of CFD to Pedestrian Wind Environment around Buildings. *J. Wind Eng. Ind. Aerodyn.* **2008**, *96*, 1749–1761. [CrossRef]
38. Salazar, J.; Albani, R. Atmospheric Boundary Layer Flow Simulations with OpenFOAM Using a Modified k-epsilon Model Consistent with Prescribed Inlet Conditions. *ABCM Eng. Proc.* **2022**. [CrossRef]
39. Wieringa, J. Updating the Davenport roughness classification. *J. Wind. Eng. Ind. Aerodyn.* **1992**, *41*, 357–368. Available online: https://www.sciencedirect.com/science/article/pii/016761059290434C (accessed on 23 December 2025). [CrossRef]
40. Petroff, A.; Mailliat, A.; Amielh, M.; Anselmet, F. Aerosol dry deposition on vegetative canopies. Part I: Review of present knowledge. *Atmos. Environ.* **2001**, *42*, 3625–3653. [CrossRef]
41. Zhang, L.; Gong, S.; Padro, J.; Barrie, L. A size-segregated particle dry deposition scheme for an atmospheric aerosol module. *Atmos. Environ.* **2001**, *35*, 549–560. [CrossRef]

42. OpenCFD Ltd. OpenFOAM User Guide, Section 4.4: Mesh Generation with the snappyHexMesh Utility. Available online: https://www.openfoam.com/documentation/user-guide/4-mesh-generation-and-conversion/4.4-mesh-generation-with-the-snappyhexmesh-utility (accessed on 23 December 2025).
43. Blocken, B.; Stathopoulos, T.; Carmeliet, J. CFD Simulation of the Atmospheric Boundary Layer: Wall Function Problems. *Atmos. Environ.* **2007**, *41*, 238–252. [CrossRef]
44. Hargreaves, D.M.; Wrightet, N.G. On the use of the k–ϵ model in commercial CFD software to model the neutral atmospheric boundary layer. *J. Wind. Eng. Ind. Aerodyn.* **2007**, *95*, 355–369. [CrossRef]
45. Zoljalali, M.; Mohsenpour, A.; Omidbakhsh Amiri, E. Developing MLP-ICA and MLP Algorithms for Investigating Flow Distribution and Pressure Drop Changes in Manifold Microchannels. *Arab. J. Sci. Eng.* **2022**, *47*, 6477–6488. [CrossRef]
46. Ghazvini, M.; Varedi-Koulaei, S.M.; Ahmadi, M.H. Optimization of MLP Neural Network for Modeling Effects of Electric Fields on Bubble Growth in Pool Boiling. *Heat Mass Transf.* **2023**, *60*, 329–336. [CrossRef]
47. Hora, G.S.; Giometto, M.G. Surrogate Modeling of Urban Boundary-Layer Flows. *arXiv* **2023**, arXiv:2306.17807. [CrossRef]
48. Lumet, E.; Rochoux, M.C.; Jaravel, T.; Lacroix, S. Uncertainty-aware surrogate modeling for urban air pollutant dispersion prediction. *Build. Environ.* **2025**, *267*, 112287. [CrossRef]
49. Chen, T.; Li, R.; Hu, X.; Zhang, B.; Liu, Y.; Wang, L.; Gao, N. Machine learning as CFD surrogate models for rapid prediction of building-related physical fields: A review of methods and state-of-the-art. *Build. Environ.* **2025**, *285*, 113667. [CrossRef]
50. Caron, C.; Lauret, P.; Bastide, A. Machine Learning to speed up Computational Fluid Dynamics engineering simulations for built environments: A review. *Build. Environ.* **2025**, *206*, 108315. [CrossRef]
51. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-Generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019.
52. Yang, S.; Vinuesa, R.; Kang, N. Enhancing Graph U-Nets for Mesh-Agnostic Spatio-Temporal Flow Prediction. *arXiv* **2024**, arXiv:2406.03789. [CrossRef]
53. Lupo Pasini, M.; Reeve, S.T.; Zhang, P.; Choi, J.Y. *HydraGNN: Distributed PyTorch Implementation of Multi-Headed Graph Convolutional Neural Networks*; Technical Report; Oak Ridge National Laboratory: Oak Ridge, TN, USA, 2021. [CrossRef]
54. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.