IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 12 September 2023, accepted 22 September 2023, date of publication 28 September 2023,
date of current version 4 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3320563

RESEARCH ARTICLE

Short Path Wind-Field Distance-Based Lagrangian Trajectory Model for Enhancing Atmospheric Dispersion Prediction Accuracy

SOUKAINA R'BIGUI[1], HIND R'BIGUI[2], AND CHIWOON CHO[1]
[1]School of Industrial Engineering, University of Ulsan, Ulsan 680749, Republic of Korea
[2]Digital Enterprise Department, NSOFT Company Ltd., Ulsan 44776, Republic of Korea

Corresponding author: Chiwoon Cho (chiwoon6@ulsan.ac.kr)

This work was supported by the 2023 Research Fund of University of Ulsan.

ABSTRACT Air pollution is a major global issue that not only threatens the safety of our planet but also poses risks to global health. Weather plays a crucial role in the rapid dispersion of air pollution. Various models have been used to predict air pollution; however, atmospheric pollution dispersion remains unpredictable, especially in relation to meteorological conditions. Our research scope focuses on developing an Air Diffusion Model using Future Wind and Pollutant sensing data to enhance prediction accuracy. In this paper, we present a new approach based on a mathematical model named the Short Path Distance based Lagrangian Trajectory Model (SPD-LTM). This model utilizes a trajectory approach and short path wind-field distance optimization to predict future air dispersion using pollutant sensing data. The framework developed in this work aims to model changes in Particulate Matter (PM2.5) and predict its concentration based on short path distance and time dependencies. The Lagrangian trajectory and concentration calculations are performed using the Hybrid Single-Particulate Lagrangian Integrated Trajectory algorithm (HYSPLIT). Then, we apply the short path distance algorithm using the Dijkstra algorithm. The obtained results demonstrate that the SPD-LTM outperforms the usual LTM and provides better accuracy to our predictive model.

INDEX TERMS PM2.5, air pollution, predictive model, short-path distance, trajectory model, particle trajectory, interpolation.

I. INTRODUCTION

The evolutional wave of the fast industrialization has been leading us to face a serious climate changes and health issues. This global issue refers mainly to the toxic elements existing in the air we consume [1], [2]. One of the most harmful contaminants in the air is PM2.5, which is a type of a small particulate matter that has a diameter of less than 2.5 micrometers [3], [4]. PM2.5 particles are incredibly lightweight and small, which allows them to remain suspended in the atmosphere for a considerable amount of time [5]. They are produced by a variety of sources, including motor vehicles, power plants, factories, and wildfires, and can be dispersed over enormous area by wind and atmospheric conditions [6]. When breathed, these small particles may spread deep into the lungs and even in the bloodstream, causing many health diseases [7]. The concentrations of PM2.5 in the atmosphere are measured in micrograms per cubic meter ($\mu g/m^3$) and are considered as a critical indicator of air quality. In many cities throughout the world, PM2.5 concentrations often surpass the Worth Health Organization's recommended limit of 10g/m3, reaching levels that are toxic to human health [8]. Atmospheric dispersion is a fundamental phenomenon that impacts the movements and distribution of air pollutants, especially PM2.5, in the atmosphere. It refers to the transport of pollutants through the air, driven by wind field, temperature gradients, and other meteorological conditions. Atmospheric dispersion models are used to forecast the movement and behavior of pollutants in the atmosphere, to help develop effective strategies for

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

lowering air pollution and improving air quality, especially in industrial areas [9], [10], [11], [12], [13], [14], [15]. One way to examine the movement of air pollutants, such as PM2.5, is through Lagrangian trajectory models. These models follow the movement of the air particles and can determine as well the source and path of air pollution. Indeed, lagrangian trajectory models use mathematical algorithms to simulate the transport of air pollutants over time and space. These models can take into consideration the impacts of wind, turbulence, and other meteorological conditions that affect the movement of air particles. By inputting data on the location and timing of air pollution emissions, as well as meteorological data such as wind speed and direction, these models may simulate the movement of air pollutants and provide insights into their dispersion patterns [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]. One of the most extensively used lagrangian trajectory models is the HYSPLIT. It employs both Lagrangian and Eulerian techniques to mimic the movement of air contaminants over huge distances and timescales. It has been used in a wide range of applications, including tracking the movement of pollutants from industrial and transportation sources, studying the dispersion of pollutants from wildfires and volcanic eruptions, and assessing the impact of long-range transport of air pollutants on human health and the environment [29], [30], [31], [32]. Additionally, air pollution dispersion can be enhanced through the use of Short Path Distance (SPD) models. Considering that SPD models take into account the distance that air particles move during a short period of time, often several hours, in order to anticipate more correctly their dispersion patterns [33], [34], [35]. The goal of this study is to implement the lagrangian trajectory approach and the short path distance approach, by presenting a new method. Comparisons are conducted between real time concentration data of the PM2.5, the LTM concentration results and the SPD-LTM concentration to observe the prediction improvement achieved.

II. RELATED WORK

Models of atmospheric dispersion are crucial tools for determining how air pollutants affect the environment and human health. These models are used to mimic the behavior of air pollutants in the atmosphere, including their transport, diffusion, and chemical changes, according to [36]. The use of these models to forecast the concentration and dispersion of fine PM2.5 in the atmosphere has attracted increasing interest in recent years. For instance, [37] predicted South Korean PM2.5 concentrations using the Community Multiscale Air Quality (CMAQ) model. Operational Multiscale Environment Model with Grid Adaptivity (OMEGA) is one sort of atmospheric dispersion model frequently used for PM2.5. OMEGA is a cutting-edge air quality model that combines a variety of physical and chemical processes that affect the transport and fate of air pollutants in the atmosphere, according to [38] and [39]. It simulates air quality at numerous scales, from regional to local. The CMAQ model, a complete air quality modeling system created by the EPA and other research institutes, is another frequently used atmospheric dispersion model for PM2.5. The CMAQ model, which has been widely used for regulatory and research reasons, comprises accurate models of atmospheric chemistry, emissions, and meteorology according to [40] and [41]. The Weather Research and Forecasting model with Chemistry (WRF-Chem) [42], the Comprehensive Air Quality Model with Extensions (CAMx) [43], and the HYSPLIT model [44] are a few additional atmospheric dispersion models that have been used for PM2.5. A class of atmospheric transport models called lagrangian dispersion models simulates the flow of air contaminants by following the path of individual particles or air parcels. Lagrangian models, which can take into account the impacts of complicated meteorological circumstances such atmospheric turbulence and convection, are particularly well-suited for modelling long-range transport of pollutants, according to [45]. The HYSPLIT model, created by NOAA [46], is a typical Lagrangian dispersion model. HYSPLIT is a popular model for atmospheric transport and dispersion modeling, according to [47], and has been used for a range of environmental applications, such as air pollution, radioactive discharges, and volcanic ash dispersion. The Flexible Particle (FLEXPART) model, created by the Norwegian Institute for Air Research, is another Lagrangian model that has gained prominence recently [45]. FLEXPART has reportedly been used extensively in air quality and climate research, and has been applied to a wide range of atmospheric transport and dispersion investigations, according to [48]. Other Lagrangian dispersion models, such as the Particle and Dispersion Model (PDM) [49] and the Lagrangian Particle Dispersion Model (LPDM) [50], have been used in atmospheric transport and dispersion modeling in addition to HYSPLIT and FLEXPART. Overall, it has been found that Lagrangian dispersion models are useful tools for modelling the transport and dispersion of air contaminants in the atmosphere. On the other hand, short path distance models, a subset of atmospheric dispersion models, mimic the movement and dispersion of air pollutants over short distances, often of the order of a few kilometers or less. They are helpful for forecasting the amount of air pollution in the vicinity of emission sources, including factories or traffic. The Atmospheric Dispersion Modeling System (AERMOD) [51] and the California Puff (CALPUFF) model [52] are examples of frequently used short path distance models. These models have been used in a variety of regulatory and research situations, including environmental impact assessment [53] and air pollution modeling [54], [55], [56]. In general, atmospheric dispersion models are now a crucial tool for determining how PM2.5 affects both air quality and human health. These models are being continuously improved and updated to increase their precision and application, and it is anticipated that they will play a crucial part in determining future air quality laws and regulations [57].

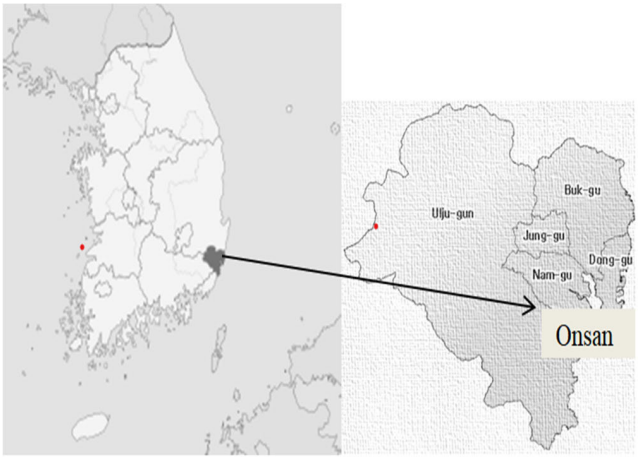

FIGURE 1. The study area ulsan city.

TABLE 1. Station Specification.

Air Quality Station Area	Onsan (Ulsan- South Korea)
Station Name	Deoksin-ro
Data Capturing Date (Time span)	01/01/2022-12/31/2022 (Hourly Data)
Latitude	35.5388° N
Longitude	129.1269° E

III. MATERIALS AND METHODS

A. PM2.5 STUDY AREA

The Onsan industrial zone in Ulsan City is one potential study location that could concentrate on PM2.5 and weather factors. Ulsan city is renowned for having high levels of air pollution, notably PM2.5, which has been associated to harmful health impacts, according to prior studies [58]. About 21 monitoring stations were set up by researchers in Ulsan for the collection of both PM2.5 and meteorological data [59]. The Korea National Institute of Environmental Research (KNIER) station near the Onsan industrial complex is one of the areas where they placed air quality monitoring equipment for PM2.5 [60]. Basically ulsan city has 5 districts: ulju-gun, buk-gu, jung-gu, namgu, and dong-gu. Onsan is located in the neighboring area of ulju-gun district as shown in Figure 1. The station is able to measure 6 metrics: SO2, CO, O3, NO2, PM10, and PM2.5. The measuring station has been installed in 1993 and operated by the Ulsan Institute of Health & Environment agency [61]. The monitoring station location is shown in Figure 2 [62]. As indicated in Table 1, The Onsan station had a Latitude of 35.5388°N and a longitude of 129.1269°E.

B. DATA COLLECTION

The observed hourly meteorological data of Ulsan city used in this study was obtained from the Korea Meteorological Administration (KMA) which is the responsible agency for the meteorological observations in South Korea. The detailed observed weather data was provided by the Automated Weather System (AWS), which is an automatic weather station network operated by the KMA [63]. PM2.5 data was retrieved from AirKorea which is a real-time air quality information disclosure system (www.airkorea.or.kr). The collected hourly data was from 1 January 2022 to 31 December 2022. As per the meteorological parameters, we retrieved wind speed, wind direction, temperature and humidity. Table 2 shows the related variables of the retrieved data at the monitored location of our study area. However, those units are commonly used in environmental monitoring, meteorology, and various scientific and data collection contexts to describe and quantify different parameters and variables. Year units represents time in years, typically used to indicate the calendar year when the data was collected. Month unit represents time in months, often used to specify the month within a year when the data was recorded. The Time is measured in hours, indicatiog the specific hours within a day when the data was taken(this is typically on a 24-hour clock (0∼23). The longitude is measured in degrees and represents the east-west positon of a location on the Earth's surface (The ranges from −180°(West) ∼ +180°(East) with pri. Latitude is also measured in degrees and represents the north-south position of a location on the Earth's surface. It ranges from −90° (South) to +90° (North), with 0° at the Equator. Temperature is measured in degrees Celsius (°C), which is a common unit for indicating temperature in the metric system. It measures the warmth or coldness of the air or a substance. Humidity is expressed as a percentage (%), representing the relative humidity of the air. It measures the amount of moisture present in the atmosphere relative to the maximum amount it can hold at a given temperature. Wind speed is measured in meters per second (m/s) and represents the rate at which air is moving horizontally past a point. It's commonly used to describe the intensity of wind. Wind direction is measured in degrees and indicates the direction from which the wind is blowing. It's typically measured in degrees clockwise from north (0°). PM2.5 concentration is measured in micrograms per cubic meter ($\mu g/m^3$). It quantifies the concentration of fine particulate matter in the air

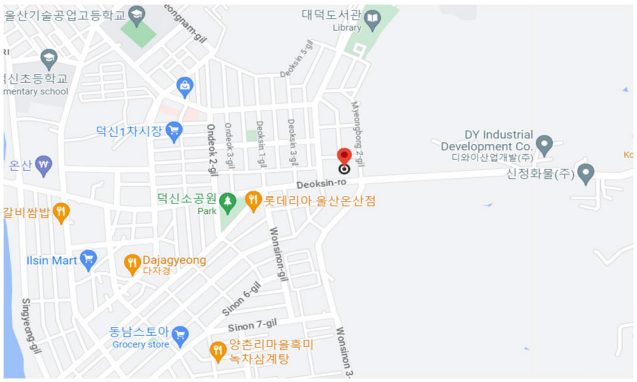

FIGURE 2. Monitoring station location of PM2.5.

TABLE 2. Data Related variable.

Related Variable	Unit
Year	year
Month	month
Hour	hour
Longitude (°)	degree
Latitude(°)	degree
Temperature	°C
Humidity	%
Windspeed	m/s
Winddirection	degree
Pm2.5	µg/m³

with a diameter of 2.5 micrometers or smaller, which can have adverse effects on health when inhaled.

C. DATA PREPROCESSING

In order to analyze the PM2.5 data and meteorological data, a data preprocessing approach is necessary. Firstly, the two datasets need to be merged and mapped together to establish a comprehensive dataset for analysis. This can be achieved by aligning the timestamps or common identifiers present in both datasets. We used common timestamps during data integration to merge the PM2.5 data and meteorological data into a single dataset. Z-score standardization method has been applied to the related variables of the meteorological data and the pm2.5 data. However the geological data variables such as latitude and longitude didnt't require any standardization Then, we ensured that both datasets have consistent and compatible formats. Interpolation techniques were used to fill in any missing values in the dataset. By estimating values based on the nearby accessible data points, interpolation aids in filling in the gaps in the data. We used spine interpolation, which uses smooth curves to approximatively fill in missing data, for handling missing values.

The equation for each cubic polynomial segment between two neighboring data points for the cubic spline interpolation is written as follows:

Taking into account two neighboring data sets (x_i, z_i) and (x_{i+1}, z_{i+1}), where x_i and x_{i+1} are x-coordinates, while z_i and z_{i+1} are y-coordinates. The cubic spline equation for the segment between these two points is as follows:

$$S(x) = a \times (x_{i+1} - x)^3 + b \times (x - x_i)^3 + c \times (x_{i+1} - x)^3 + d \times (x - x_i)^3 \quad (1)$$

where:

- The cubic spline function is denoted by $S(x)$.
- The input value x is what we want to use to estimate the z-value.

The cubic spline must satisfy the following condition in order to compute the coefficients $a, b, c,$ and d. These conditions include:

1) Interpolation Condition: The spline passes through each data point, meaning $S(x_i) = z_i$ and $S(x_{i+1}) = z_{i+1}$.

2) Continuity Condition: The first and second derivatives of adjacent polynomials are equal at each data point to ensure a smooth transition.

In order to meet the requirements of the integration of HYSPLIT algorithm, we made sure that our data comprised regular time intervals with constant temporal resolution, geographical coordinates (latitude, longitude), and relevant meteorological variables (wind speed, wind direction, and temperature). These aspects need to be considered in order to properly comprehend air transport and dispersion. To visually assess the data on wind direction, we used a wind rose in our preprocessing part. The wind rose shows the frequency and distribution of measured wind directions, providing information about the dominant winds and wind resources. By collecting and aggregating wind direction data into directional sectors, we create the wind rose in Figure 3, which offers a thorough picture of the wind condition in our study region. The observation of pm2.5 in correlation with the other features is shown in Figure 5. After the observation of each feature in Figures 4, 5, 6 and 7 in correlation with the measured pm2.5 retrieved from the measuring station of the study area, we conclude that temperature and humidity had no significant impact on the pm2.5 concentration as observed in the analyzed on Figure 8. In consequence, we are taking in consideration wind speed and wind direction features. The observation and analysis of these features are depicted in Figures 9 and 10.

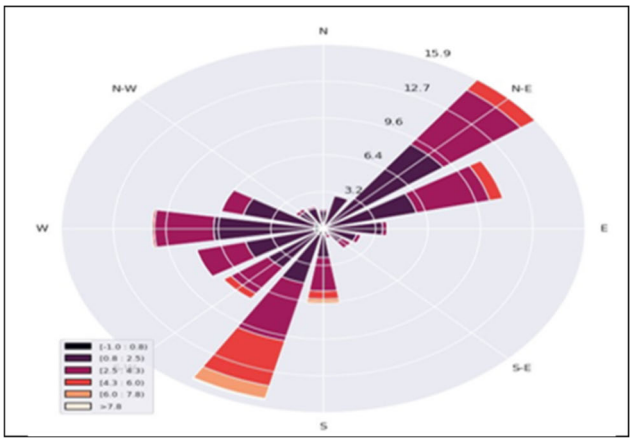

FIGURE 3. Wind rose of the wind speed and direction.

Table 3 present the detailed data information statistics for five different variables: Temperature, Humidity, Windspeed, Winddirection, and Pm2.5 measured. Each row in the table represents a statistical measure for these variables, and the columns provide specific information for each measure. For instance, count shows the number of data points available for each variable (There are 8760 data points for each variable). Mean, represents the mean (average) value of each variable. Std represents the standard deviation of each variable (Bigger number means more spread). Min shows the minimum value observed for each variable. First Quartile (25%) shows the

TABLE 3. Data information statistics.

Index	Temperature	Humidity	Windspeed	Winddirection	Pm2.5 measured
Count	8760.0	8760.0	8760.0	8760.0	8760.0
Mean	12.17133	0.70149	11.30131	193.82842	13.10273
Std	9.18279	0.18450	7.01055	103.90505	12.28054
Min	-11.12778	0.0	0.0	0.0	0.0
(First Quartile) 25%	5.58333	0.57	6.2307	129.0	2.0
(Second Quartile) 50%	11.57778	0.74	10.6421	190.0	11.0
(Third Quartile)75%	18.68472	0.85	14.4417	291.0	20.0
Max	37.75556	1.0	55.9314	359.0	75.0

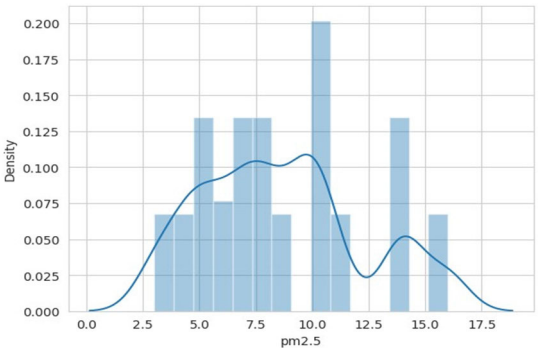

FIGURE 4. Measured pm2.5 distribution in the dataset.

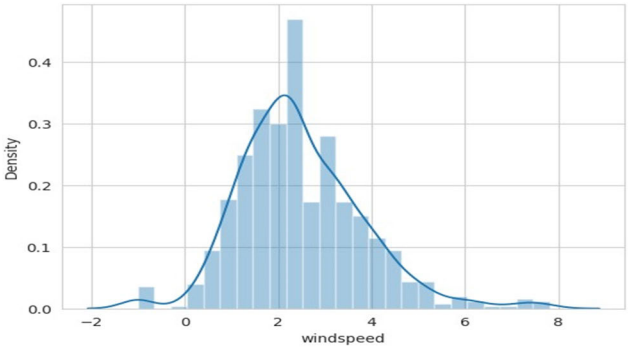

FIGURE 6. Windspeed distribution in the dataset.

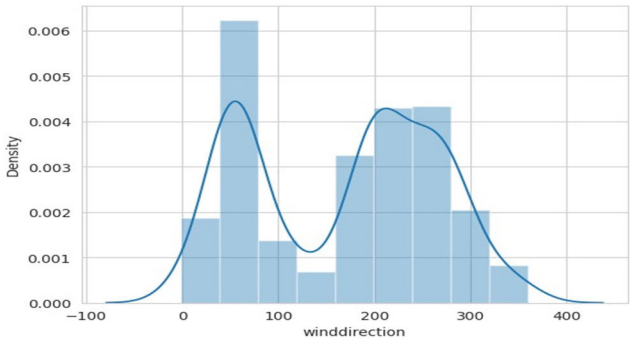

FIGURE 5. Winddirection distribution in the dataset.

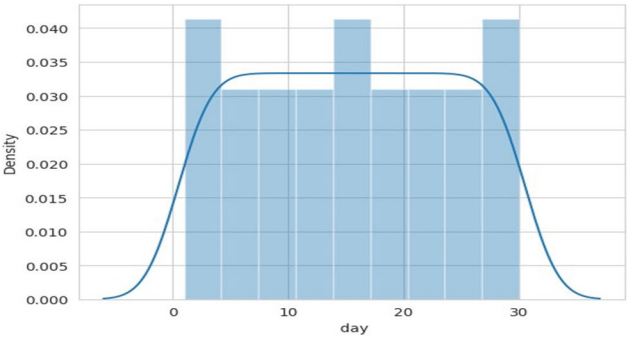

FIGURE 7. Time distribution in the dataset.

value at the 25th percentile of the data. Max gives the maximum (largest) value observed for each variable.

D. METHODOLOGY

In this part, we are discussing the attempted methodology used in this study for the pm2.5 concentration prediction. The process of this study is shown in Figure 11. The trajectories and concentrations of the pm2.5 were calculated using Hysplit algorithm. Equation (2) and equation (4) are the basic particle trajectory and particle concentration equations respectively [65]:

1. Particle trajectory:

$$p'(t+\Delta t) = p(t) + V(p,t).\Delta t \tag{2}$$

The advection of the particle is calculated as the average of the three-dimensional velocity vectors $V(p,t)$ at the initial

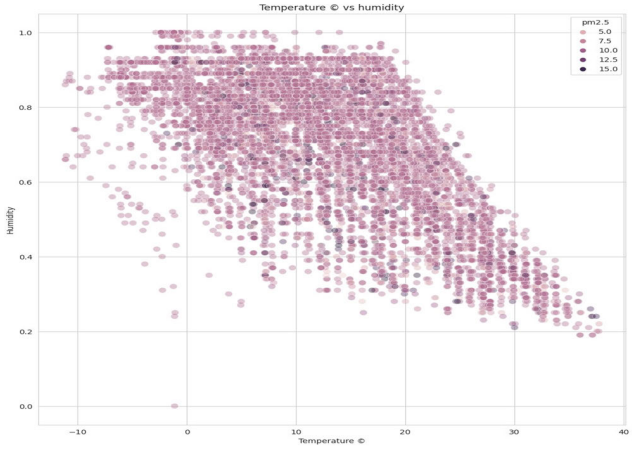

FIGURE 8. Temperature and humidity by measured pm2.5.

point $p(t)$ and the first-estimated position $p'(t+\Delta t)$ where t represents time. The final position is:

$$p(t+\Delta t) = p(t) + \frac{1}{2}\left[V(p,t) + V\left(p', t+\Delta t\right)\right]\Delta t \quad (3)$$

To find the final position of the particle at time $t+\Delta t(p(t+\Delta t))$, we update the initial position $p(t)$ by adding an average of the velocities at both time points ($tandt+\Delta t$). The velocity vector at the first-estimated position $p'(t+\Delta t)$ is considered alongside the current velocity vector $V(p,t)$ at time t. The averaging of velocities helps to account for the potential changes in the particle's speed or direction over the time interval Δt.

The term $1/2[V(p,t) + V(p', t+\Delta t)]$ represents the average velocity vector over the time intervalΔt, and when multiplied by Δt, it gives the displacement of the particle over that time step. Adding this displacement to the initial position $p(t)$ gives the final position of the particle at time $t+\Delta t$. In both space and time, the velocity vectors are linearly interpolated.

FIGURE 9. Windspeed and winddirection by measured pm2.5.

2. Particle concentration:

$$\Delta C = m\left(2\pi\sigma_h^2\Delta z\right)^{-1}\exp\left(\frac{-x^2}{2\sigma^2 h}\right) \quad (4)$$

where ΔC represents the change in particle concentration at a specific point in the three-dimensional space and m representing the mass of the particles. It scales the overall concentration level. σ_h is the standard deviation of the Gaussian distribution in the horizontal plane. It controls the spread or dispersion of the particle concentration in the horizontal direction, Δz is the variation or difference in height or altitude in the vertical (z) direction, x is the distance from the center point (where the concentration is at its peak). It is usually measured in the horizontal plane. σ is the standard deviation of the Gaussian distribution in the x-direction. It controls the spread of the concentration along the x-axis, and h is a constant parameter, which can influence the vertical spread of the concentration. Therefore, the spread of the distribution in both the horizontal and vertical directions is controlled by σ_h and Δz, respectively.

The cost function is used to quantify how well the predicted PM2.5 concentration aligns with the observed values while also considering the influence of the short path distance. By minimizing the cost function, the model aims to improve its predictive accuracy and effectively account for the relationship between PM2.5 concentration and the distance traveled along the short path. The cost function equation is as follows:

$$cost = W_1 \times (PM_{predicted} - PM_{observed})^2 + W_2 \times SD \quad (5)$$

The squared difference between the predicted and observed pm2.5 concentrations is combined in the cost function. By squaring the difference, the necessity of an accurate forecast is highlighted and greater errors are amplified. SD variable refers to the short path distance component, is a measurement of the length of the trajectory connected to the concentration of pm2.5. The distance of the trajectory has been calculated in this phase using the Dijkstra method. Our model carefully examines the effect of the trajectory distance on the prediction of pm2.5 concentration while implementing the short path distance in this cost function. W_1 and W_2 are weighting factors that determine the relative importance of the prediction error component and the short path distance component in the cost function. W_1 controls the influence of the prediction error component, while W_2 controls the influence of the short path distance component. The notations and explanations used in this cost function as well as the notations used in this study, are summarized in Table 4.

E. EVALUATION METRICS

We compared the observed concentrations of pm2.5, the predicted concentrations without the short path distance, and the predicted concentrations with the short path distance to assess the performance of our model and determine how much efficiency the short path distance would add to our prediction.

The statistical evaluation metrics listed in Equations 6, 7 and 8 were employed for this evaluation. y_i is the observed pm2.5 concentration, $\hat{y}_i$ is the predicted pm2.5 concentrations, $\bar{y}_i$ is the mean of observed values and n is the length of the test set. Measures of the difference between the observed value and the predicted value include Root Mean Square Eroor (RMSE), Mean Absolute Error (MAE) and Means squared error (MSE). The RMSE and MAE metrics show how sensitive and resistant the model is to greater errors, respectively. The prediction impact is better when the two values are lower. MSE evaluates how well the accuracy predicted results match the observed data. The lower MSE value, the better the effect can be predicted.

We calculated the MAE by taking the absolute difference between each observed y_i and predicted $\hat{y}_i$ value and then averaging those absolute differences over all data points n. The formula is as follows:

$$MAE = \frac{1}{n}\sum_{i-1}^{n}\left|y_i - \hat{y}_i\right| \tag{6}$$

This equation (6) calculates the absolute difference between each observed value y_i and its corresponding predicted value $\hat{y}_i$ for all data points ($i = 1$ to n). Then, it takes the sum $\sum$ of all these absolute differences. Finally, it divides the sum by the total number of data points n to obtain the mean, which gives the average absolute difference between the observed and predicted values.

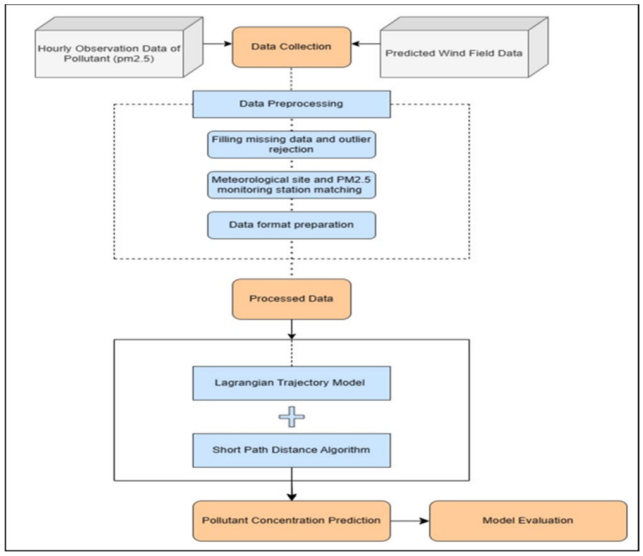

FIGURE 10. Overview of the study design.

We have calculated the MSE by taking the squared difference between each observed y_i and predicted $\hat{y}_i$ value and then averaging those squared differences over all data points n. The formula is as follows:

$$MSE = \frac{1}{n}\sum_{i-1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{7}$$

where the equation (7) calculates the squared difference between each observed value y_i and its corresponding predicted value $\hat{y}_i$ for all data points. Then, it takes the sum $\sum$ of

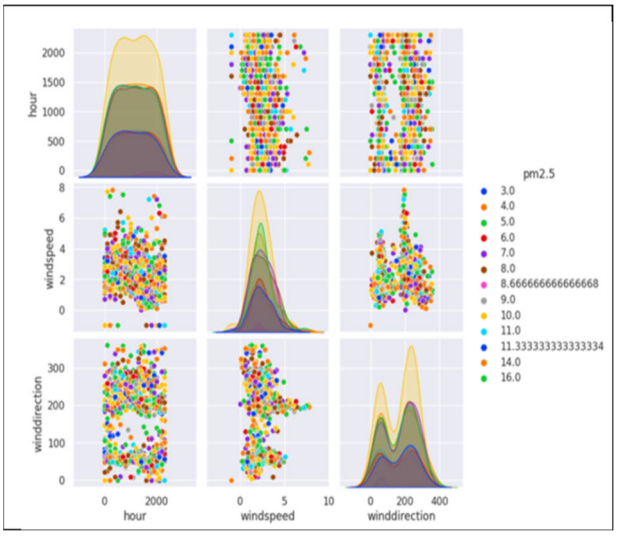

FIGURE 11. Correlation between pm2.5 and other features.

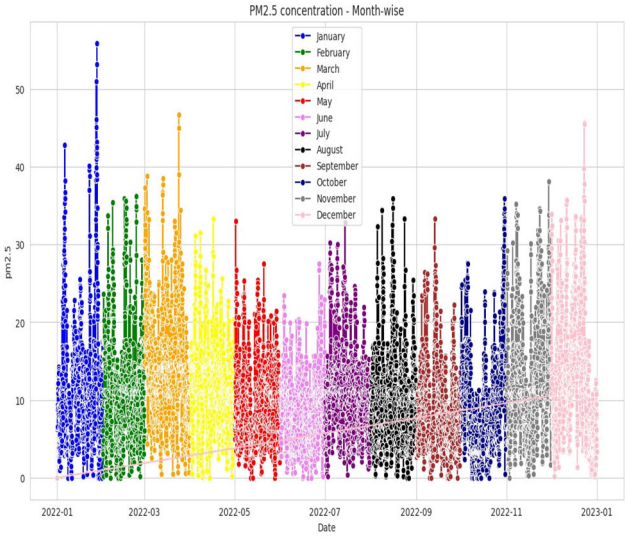

FIGURE 12. Pm2.5 concentration monthly results for the SPD-LTM prediction.

all these squared differences. Finally, it calculates the average squared difference between observed and predicted values by dividing the sum by the total number of data points n.

We obtained the RMSE by taking the square root of the MSE. It provides a measure of the typical error magnitude between the observed and predicted values. The formula is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i-1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{8}$$

The equation (8) calculates the average difference between the observed and predicted values, squares it, takes the mean of all the squared differences, and then takes the square root of that mean. The resulting value represents the typical

TABLE 4. Notations and explanations.

Notation	Explanation
$PM_{predicted}$	The predicted pm2.5 concentration
$PM_{observed}$	The observed pm2.5 concentration
W_1	Weight factor related to the prediction error component
W_2	Weight factor related to the short path distance component
SD	Variable refers to the short path distance, a measurement of the length of the trajectory connected to pm2.5 concentration
$p(t)$	Initial point of the particle at time t
$p'(t+\Delta t)$	first-estimated position of the particle $(t+\Delta t)$ time
$p(t+\Delta t)$	Final position of the particle at $(t+\Delta t)$ time of the
$V(p,t)$	Velocity vector of the point at time t of the particle
Δt, t	Time
ΔC	Particle concentration
Δz	A certain height about the source of emission
σ	Vertical dispersion coefficient
σ_h	Horizontal dispersion coefficient
h	Reference level at which the concentration is being computed
m	Emission rate
X	Horizontal distance from the emission source
$S(x)$	The cubic spline function for the interpolation
(x_i, z_i)	two neighboring data sets of the cubic spline function
a, b, c, d	Interpolation coefficients of cubic spline function
MAE	Mean Absolute Error
MSE	Mean Squared Error
$RMSE$	Root Mean Squared Error
yi	observed pm2.5 concentration
$\hat{y}_i$	predicted pm2.5 concentration
$\bar{y}_i$	the mean of observed values
n	the length of the test set

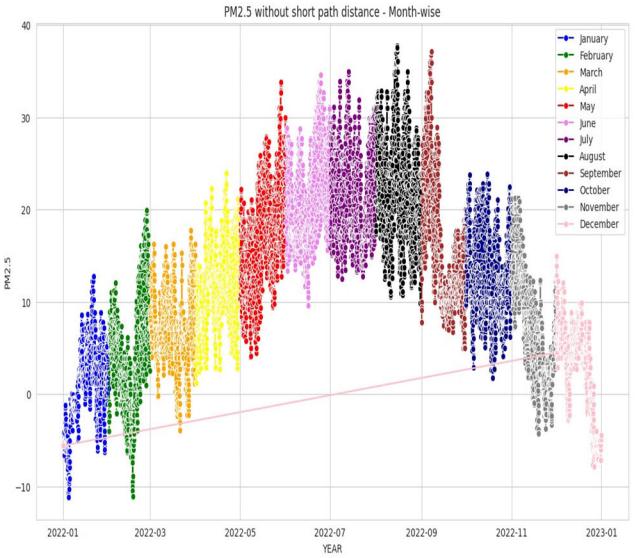

FIGURE 13. Pm2.5 concentration monthly results for the LTM prediction.

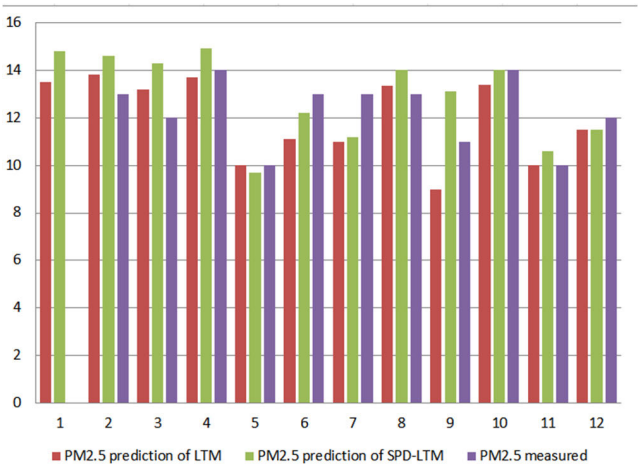

FIGURE 14. Comparison between the predicted pm2.5 of SPD-LTM and LTM and the measured pm2.5.

difference between the observed and predicted values, with lower values indicating better model performance.

IV. RESULTS & DISCUSSION

Our model's evaluation showed useful predictions of PM2.5 values. The MAE between the predicted and observed values for our SPD-LTM model was determined as 1.1400. Additionally, our model's MSE, which measures the typical squared difference between predicted and observed values, was 1.2400. Additionally, the RMSE, which measures the residuals' standard deviation, was discovered to be 1.4800. Compared to the short path distance model, the Lagrangian trajectory model had somewhat larger errors, as seen by the MAE of 1.2600, MSE of 1.3000, and RMSE of 1.6700. When short path distance is not taken into account, these measurements show a bigger average absolute difference, squared difference, and standard deviation of the residuals. The better model performance, as shown by the reduced MAE, MSE, and RMSE values, emphasizes the important role of the short path distance parameter in improving the precision of PM2.5 concentration forecasts. We were able to consider the particular transport pathways and fully comprehend the spatiotemporal distribution of PM2.5 concentrations by including short path distance. Figure 12 shows the pm2.5 concentration monthly results. We conclude that the average concentration varies from 0 to 20(μg/m^3) which is considered as good quality average. Contrary to the usual LTM results, the concentration of pm2.5 is not accurate as depicted in Figure 13. As a summary of our model performance, Figure 14 presents a comparison between the observed pm2.5 concentration, SPD-LTM predicted concentration and the LTM predicted concentration. The short path distance

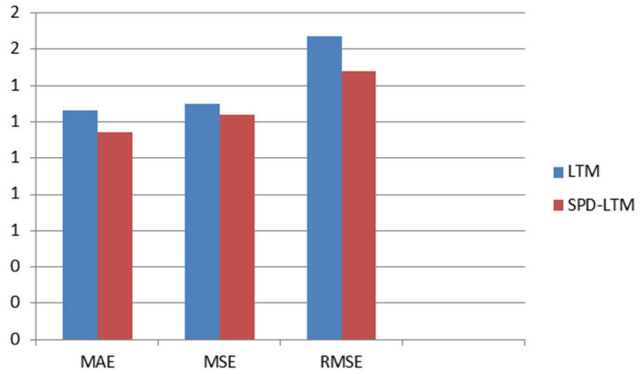

FIGURE 15. Model prediction performance comparison.

TABLE 5. Model prediction performance at the study location.

Method	LTM	SPD-LTM
MAE	1.2600	1.1400
MSE	1.3000	1.2400
RMSE	1.6700	1.4800

parameter's extra data enables more accurate calculations and depicts the intricate dynamics of pollution movement more clearly. These findings highlight the significance of coupling Lagrangian trajectory model with the short path distance, as well as the crucial impact of our proposed cost function in the short path distance has in improving PM2.5 concentration prediction. The importance of this parameter is shown by the enhanced performance of our model over the usual Lagrangian trajectory model. The results of this study have significant impact for managing air quality and making decisions on public health, and they offer insightful information for comprehending and combating PM2.5 pollution. Table 5 and Figure 15 show performance of SPD-LTM and a comparison with LTM.

V. CONCLUSION AND FUTURE WORK

In this study, a new framework for pm2.5 propagation was presented. This study shows how a Lagrangian trajectory model with short path distance can be used to estimate PM2.5 concentrations by introducing a new mathematical method. The reduced values of MAE, MSE, and RMSE show that the model outperforms both the real-time data and the traditional Lagrangian trajectory model. The inclusion of short path lengths greatly increases the accuracy and precision of the predictions and provides a more accurate understanding of the dynamics of pollutant transport. The results highlight the importance of considering specific transport pathways and their effects on PM2.5 concentrations. This method holds great promise for use in treatments related to public health, environmental planning, and air quality management. Exploring the full potential of this technology and incorporating it into decision support systems should be the focus of future studies. Future work should focus on several areas to further enhance the application of this methodology. Firstly, the model's performance can be assessed in different geographical regions and under varying meteorological conditions to evaluate its generalizability. Additionally, the integration of other relevant parameters such as land use patterns, emission sources, and meteorological factors could contribute to a more comprehensive predictive model. Furthermore, the incorporation of artificial intelligence, machine learning techniques and advanced data analytics methods may enhance the model's predictive capabilities.

AUTHOR CONTRIBUTION

Soukaina R'bigui: main idea and the methodology of the research; Soukaina R'bigui: designed the experiment, performed the simulations, and wrote the original manuscript; Hind R'bigui: improving the technical and grammatical contents of the manuscript; and Hind R'bigui and Chiwoon Cho: reviewed the manuscript and provided valuable suggestions to further refine the manuscript.

REFERENCES

[1] J. Park and S. Chang, "A particulate matter concentration prediction model based on long short-term memory and an artificial neural network," *Int. J. Environ. Res. Public Health*, vol. 18, no. 13, p. 6801, Jun. 2021, doi: 10.3390/ijerph18136801.

[2] T. Chuluunsaikhan, M. Heak, A. Nasridinov, and S. Choi, "Comparative analysis of predictive models for fine particulate matter in Daejeon, South Korea," *Atmosphere*, vol. 12, no. 10, p. 1295, Oct. 2021.

[3] F. Zhang, H.-R. Cheng, Z.-W. Wang, X.-P. Lv, Z.-M. Zhu, G. Zhang, and X.-M. Wang, "Fine particles ($PM_{2.5}$) at a CAWNET background site in central China: Chemical compositions, seasonal variations and regional pollution events," *Atmos. Environ.*, vol. 86, pp. 193–202, Apr. 2014, doi: 10.1016/j.atmosenv.2013.12.008.

[4] Q. Yang, Q. Yuan, T. Li, H. Shen, and L. Zhang, "The relationships between $PM_{2.5}$ and meteorological factors in China: Seasonal and regional variations," *Int. J. Environ. Res. Public Health*, vol. 14, no. 12, p. 1510, Dec. 2017, doi: 10.3390/ijerph14121510.

[5] W. Bai and F. Li, "$PM_{2.5}$ concentration prediction using deep learning in Internet of Things air monitoring system," *Environ. Eng. Res.*, vol. 28, no. 1, pp. 210450–210456, 2023, doi: 10.4491/eer.2021.456.

[6] World Health Organization. *Global Health Observatory (GHO) Data for Ambient Air Pollution*. [Online]. Available: http://www.who.int/gho/phe/outdoor_air_pollution/en/

[7] J. Kasznia-Kocot, M. Kowalska, R. L. Górny, A. Niesler, and A. Wypych-Slusarska, "Environmental risk factors for respiratory symptoms and childhood asthma," *Ann. Agric. Environ. Med.*, vol. 17, no. 2, pp. 221–229, 2010.

[8] K. Qadeer, W. U. Rehman, A. M. Sheri, I. Park, H. K. Kim, and M. Jeon, "A long short-term memory (LSTM) network for hourly estimation of $PM_{2.5}$ concentration in two cities of South Korea," *Appl. Sci.*, vol. 10, no. 11, p. 3984, Jun. 2020, doi: 10.3390/app10113984.

[9] J. Li, X. Li, K. Wang, and G. Cui, "Atmospheric $PM_{2.5}$ concentration prediction and noise estimation based on adaptive unscented Kalman filtering," *Meas. Control*, vol. 54, nos. 3–4, pp. 292–302, Mar. 2021, doi: 10.1177/0020294021997491.

[10] A. Kavousi-Fard, H. Samet, and F. Marzbani, "A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting," *Exp. Syst. Appl.*, vol. 41, no. 13, pp. 6047–6056, Oct. 2014.

[11] C. Wen, S. Liu, X. Yao, L. Peng, X. Li, Y. Hu, and T. Chi, "A novel spatiotemporal convolutional long short-term neural network for air pollution prediction," *Sci. Total Environ.*, vol. 654, pp. 1091–1099, Mar. 2019.

[12] J. Wang, P. Du, Y. Hao, X. Ma, T. Niu, and W. Yang, "An innovative hybrid model based on outlier detection and correction algorithm and heuristic intelligent optimization algorithm for daily air quality index forecasting," *J. Environ. Manag.*, vol. 255, Feb. 2020, Art. no. 109855.

[13] J. Cha and J. Kim, "Development of data mining algorithm for implementation of fine dust numerical prediction model," *J. Korea Inst. Inf. Commun. Eng.*, vol. 22, no. 4, pp. 595–601, 2018.

[14] H.-J. Song, B.-D. Oh, J.-D. Kim, C.-Y. Park, and Y.-S. Kima, "Predicting concentration of PM10 using optimal parameters of deep neural network," *Intell. Autom. Soft Comput.*, vol. 25, no. 2, 2019.

[15] Y. Zhao, "Machine learning algorithms for predicting roadside fine particulate matter concentration level in Hong Kong Central," *Comp. Ecol. Softw.*, vol. 3, no. 3, p. 61, 2013.

[16] S. Ulfah and S. A. Awalludin, "Advection-diffusion model for the simulation of air pollution distribution from a point source emission," *J. Phys., Conf.*, vol. 948, Jan. 2018, Art. no. 012067, doi: 10.1088/1742-6596/948/1/012067.

[17] K. Lakshminarayanachari, K. S. Pai, M. S. Prasad, and C. Pandurangappa, "Advection-diffusion numerical model of air pollutants emitted from an urban area source with removal mechanisms by considering point source on the boundary," *Int. J. Appl. Innov. Eng. Manag.*, vol. 2, pp. 251–268, Jan. 2013.

[18] N. K. Arystanbekova, "Application of Gaussian plume models for air pollution simulation at instantaneous emissions," *Math. Comput. Simul.*, vol. 67, nos. 4–5, pp. 451–458, Dec. 2004.

[19] C. Zhao and G. Song, "Application of data mining to the analysis of meteorological data for air quality prediction: A case study in Shenyang," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 81, Aug. 2017, Art. no. 012097, doi: 10.1088/1755-1315/81/1/012097.

[20] J. I. Jeong, R. J. Park, J.-H. Woo, Y.-J. Han, and S.-M. Yi, "Source contributions to carbonaceous aerosol concentrations in Korea," *Atmos. Environ.*, vol. 45, no. 5, pp. 1116–1125, Feb. 2011.

[21] D. G. F. Huilier, "An overview of the Lagrangian dispersion modeling of heavy particles in homogeneous isotropic turbulence and considerations on related LES simulations," *Fluids*, vol. 6, no. 4, p. 145, Apr. 2021, doi: 10.3390/fluids6040145.

[22] J. D. Wilson and B. L. Sawford, "Review of Lagrangian stochastic models for trajectories in the turbulent atmosphere," *Boundary-Layer Meteorol.*, vol. 78, nos. 1–2, pp. 191–210, Feb. 1996.

[23] L.-P. Wang and D. E. Stock, "Stochastic trajectory models for turbulent diffusion: Monte Carlo process versus Markov chains," *Atmos. Environ. A, Gen. Topics*, vol. 26, no. 9, pp. 1599–1607, Jun. 1992.

[24] J.-S. Shuen, L.-D. Chen, and G. M. Faeth, "Evaluation of a stochastic model of particle dispersion in a turbulent round jet," *AIChE J.*, vol. 29, no. 1, pp. 167–170, Jan. 1983.

[25] A. M. Reynolds, "On the formulation of Lagrangian stochastic models for heavy-particle trajectories," *J. Colloid Interface Sci.*, vol. 232, no. 2, pp. 260–268, Dec. 2000.

[26] K. Launay, D. Huilier, and H. Burnage, "An improved Lagrangian method for predicting the long-time turbulent dispersion in gas-particle flows," in *Proc. ASME Summer Fluids Eng. Meeting*, Washington, DC, USA, 1998.

[27] F. Toschi and E. Bodenschatz, "Lagrangian properties of particles in turbulence," *Annu. Rev. Fluid Mech.*, vol. 41, no. 1, pp. 375–404, Jan. 2009.

[28] J. M. Stockie, "The mathematics of atmospheric dispersion modeling," *SIAM Rev.*, vol. 53, no. 2, pp. 349–372, Jan. 2011.

[29] Y. Li, D. Q. Tong, F. Ngan, M. D. Cohen, A. F. Stein, S. Kondragunta, X. Zhang, C. Ichoku, E. J. Hyer, and R. A. Kahn, "Ensemble $PM_{2.5}$ forecasting during the 2018 camp fire event using the HYSPLIT transport and dispersion model," *J. Geophys. Res., Atmos.*, vol. 125, no. 15, Aug. 2020, Art. no. JD032768.

[30] N. Fedorova, V. Levit, A. O. da Silva, and D. M. B. dos Santos, "Low visibility formation and forecasting on the northern coast of Brazil," *Pure Appl. Geophys.*, vol. 170, no. 4, pp. 689–709, Apr. 2013.

[31] J. Guang, Y. Xue, L. Mei, Y. Li, Y. Wang, H. Xu, and J. Guo, "Forecasting air quality by integration of satellite data and hysplit trajectory model," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2010, pp. 2773–2776.

[32] A. Donnelly, O. Naughton, B. Broderick, and B. Misstear, "Short-term forecasting of nitrogen dioxide (NO_2)," *Environ. Model. Assess.*, vol. 22, no. 3, pp. 231–241, 2017.

[33] L. Li, J. Gong, and J. Zhou, "Spatial interpolation of fine particulate matter concentrations using the shortest wind-field path distance," *PLoS ONE*, vol. 9, no. 5, May 2014, Art. no. e96111.

[34] A. Consortini, C. Innocenti, and G. Paoli, "Estimate method for outer scale of atmospheric turbulence," *Opt. Commun.*, vol. 214, nos. 1–6, pp. 9–14, Dec. 2002.

[35] E.-B. Moon, T.-I. Jeon, and D. R. Grischkowsky, "Long-path THz-TDS atmospheric measurements between buildings," *IEEE Trans. Terahertz Sci. Technol.*, vol. 5, no. 5, pp. 742–750, Sep. 2015.

[36] Q. Ye, M. A. Upshur, E. S. Robinson, F. M. Geiger, R. C. Sullivan, R. J. Thomson, and N. M. Donahue, "Following particle-particle mixing in atmospheric secondary organic aerosols by using isotopically labeled terpenes," *Chem*, vol. 4, no. 2, pp. 318–333, Feb. 2018.

[37] S. Park, M. Shin, J. Im, C.-K. Song, M. Choi, J. Kim, S. Lee, R. Park, J. Kim, D.-W. Lee, and S.-K. Kim, "Estimation of ground-level particulate matter concentrations through the synergistic use of satellite observations and process-based models over South Korea," *Atmos. Chem. Phys.*, vol. 19, no. 2, pp. 1097–1113, Jan. 2019.

[38] T. Dunn, D. P. Bacon, and P. Boris, "Development of the operational multiscale environment model with grid adaptivity (OMEGA) and its aerosol transport and diffusion model (ATDM)," Air & Waste Management Association, Pittsburgh, PA, USA, Tech. Rep., 1994.

[39] T. J. Dunn, D. P. Bacon, P. C. Lee, M. S. Hall, and A. Sarma, "Parallelization development and testing of the operational multiscale environment model with grid adaptivity (OMEGA)," *Appl. Briefs*, vol. 808, p. 14, Jan. 2002.

[40] K. W. Appel, S. L. Napelenok, K. M. Foley, H. O. Pye, C. Hogrefe, D. J. Luecken, J. O. Bash, S. J. Roselle, J. E. Pleim, H. Foroutan, and W. T. Hutzell, "Description and evaluation of the community multiscale air quality (CMAQ) modeling system version 5.1," *Geosci. Model Dev.*, vol. 10, no. 4, pp. 1703–1732, 2017.

[41] K. W. Appel, S. Napelenok, C. Hogrefe, G. Pouliot, K. M. Foley, S. J. Roselle, J. E. Pleim, J. Bash, H. O. Pye, N. Heath, and B. Murphy, "Overview and evaluation of the community multiscale air quality (CMAQ) modeling system version 5.2," in *Air Pollution Modeling and its Application XXV*, vol. 35. Berlin, Germany: Springer, 2018, pp. 69–73.

[42] L. K. Berg, M. Shrivastava, R. C. Easter, J. D. Fast, E. G. Chapman, Y. Liu, and R. A. Ferrare, "A new WRF-chem treatment for studying regional-scale impacts of cloud processes on aerosol and trace gases in parameterized cumuli," *Geoscientific Model Develop.*, vol. 8, no. 2, pp. 409–429, Feb. 2015.

[43] X. Chang, S. Wang, B. Zhao, J. Xing, X. Liu, L. Wei, Y. Song, W. Wu, S. Cai, H. Zheng, D. Ding, and M. Zheng, "Contributions of inter-city and regional transport to $PM_{2.5}$ concentrations in the Beijing-Tianjin-Hebei region and its implications on regional joint air pollution control," *Sci. Total Environ.*, vol. 660, pp. 1191–1200, Apr. 2019.

[44] B. Chen, A. F. Stein, N. Castell, J. D. de la Rosa, A. M. S. de la Campa, Y. Gonzalez-Castanedo, and R. R. Draxler, "Modeling and surface observations of arsenic dispersion from a large cu-smelter in southwestern Europe," *Atmos. Environ.*, vol. 49, pp. 114–122, Mar. 2012.

[45] A. Stohl, C. Forster, A. Frank, P. Seibert, and G. Wotawa, "Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2," *Atmos. Chem. Phys.*, vol. 5, no. 9, pp. 2461–2474, Sep. 2005.

[46] R. R. Draxler and G. D. Hess, "An overview of the HYSPLIT_4 modelling system for trajectories," *Aust. Meteorol. Mag.*, vol. 47, no. 4, pp. 295–308, 1998.

[47] A. F. Stein, R. R. Draxler, G. D. Rolph, B. J. B. Stunder, M. D. Cohen, and F. Ngan, "NOAA's HYSPLIT atmospheric transport and dispersion modeling system," *Bull. Amer. Meteorological Soc.*, vol. 96, no. 12, pp. 2059–2077, Dec. 2015.

[48] I. Pisso, E. Sollum, H. Grythe, N. I. Kristiansen, M. Cassiani, S. Eckhardt, D. Arnold, D. Morton, R. L. Thompson, C. D. Groot Zwaaftink, and N. Evangeliou, "The Lagrangian particle dispersion model FLEXPART version 10.4," *Geosci. Model Dev.*, vol. 12, no. 12, pp. 4955–4997, 2019.

[49] I. Park and I. W. Seo, "Modeling non-fickian pollutant mixing in open channel flows using two-dimensional particle dispersion model," *Adv. Water Resour.*, vol. 111, pp. 105–120, Jan. 2018.

[50] J. D. Fast and R. C. Easter, "A Lagrangian particle dispersion model compatible with WRF," in *Proc. 7th Annu. WRF User's Workshop*, Jun. 2006, pp. 19–22.

[51] S. Arunachalam, B. H. Baek, A. Holland, Z. Adelman, F. S. Binkowski, A. Hanna, T. Thrasher, and P. Soucacos, "An improved method to represent aviation emissions in air quality modeling systems and their impacts on air quality," in *Proc. 13 Conf. Aviation, Range Aerosp. Meteorol.*, New Orleans, LA, USA, 2008, pp. 1–15.

[52] J. S. Scire, D. G. Strimaitis, and R. J. Yamartino, "A user's guide for the CALPUFF dispersion model," *Earth Tech.*, vol. 521, pp. 1–521, Jan. 2000.

[53] S. Gulia, A. Kumar, and M. Khare, "Performance evaluation of CALPUFF and AERMOD dispersion models for air quality assessment of an industrial complex," Tech. Rep., 2015.

[54] U. Im, R. Bianconi, E. Solazzo, I. Kioutsioukis, A. Badia, A. Balzarini, R. Baró, R. Bellasio, D. Brunner, C. Chemel, and G. Curci, "Evaluation of operational on-line-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone," *Atmos. Environ.*, vol. 115, pp. 404–420, Aug. 2015.

[55] V. Sinha, V. Kumar, and C. Sarkar, "Chemical composition of pre-monsoon air in the indo-gangetic plain measured using a new air quality facility and PTR-MS: High surface ozone and strong influence of biomass burning," *Atmos. Chem. Phys.*, vol. 14, no. 12, pp. 5921–5941, Jun. 2014, doi: 10.5194/acp-14-5921-2014.

[56] A. Gowardhan, H. Spoon, D. A. Riechers, E. González-Alfonso, D. Farrah, J. Fischer, J. Darling, C. Fergulio, J. Afonso, and L. Bizzocchi, "The dual role of starbursts and active galactic nuclei in driving extreme molecular outflows," *Astrophysical J.*, vol. 859, no. 1, p. 35, May 2018.

[57] J. H. Seinfeld and S. N. Pandis, *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. Hoboken, NJ, USA: Wiley, 2016.

[58] S.-J. Lee, H.-Y. Lee, S.-J. Kim, H.-J. Kang, H. Kim, Y.-K. Seo, H.-J. Shin, Y. S. Ghim, C.-K. Song, and S.-D. Choi, "Pollution characteristics of $PM_{2.5}$ during high concentration periods in summer and winter in Ulsan, the largest industrial city in South Korea," *Atmos. Environ.*, vol. 292, Jan. 2023, Art. no. 119418.

[59] Ulsan: Metropolitan. *Real-Time Air Quality Monitoring Data*. [Online]. Available: http://www.ulsan.go.kr/s/uihe/realtime/area/index.ulsan?area=238121&mId=0010030050010020000&datatype=data&group=0&term=30

[60] W. J. Choi, B. Jung, D. Lee, H. Kang, H. Kim, and H. Hong, "An investigation into the effect of emissions from industrial complexes on air quality in the ulsan metropolitan city utilizing trace components in PM2.5," *Appl. Sci.*, vol. 11, no. 21, p. 10003, Oct. 2021.

[61] *Air, Korea*. [Online]. Available: https://www.airkorea.or.kr/eng/stationInformation?pMENU_NO=158

[62] [Online]. Available: https://www.airkorea.or.kr/web/images/station_map/238120.gif

[63] Korea Meteorological Administration. *Automated Weather System (AWS)*. [Online]. Available: https://web.kma.go.kr/eng/weather/aws.jsp

SOUKAINA R'BIGUI received the B.S. degree in mathematical education from the Higher School of Teachers, Mohammed V University, Morocco, in 2020. Currently, she is pursuing the M.S. degree with the School of Industrial Engineering, University of Ulsan, South Korea. Her research interests include artificial intelligence, machine learning, and mathematical modeling.

HIND R'BIGUI received the Ph.D. degree in industrial engineering from the University of Ulsan, South Korea, and the State Engineering degree in industrial engineering from the National School of Applied Sciences, Fes, Morocco. She is currently a Consultant and a Solution Architect of SIEMENS Smart Factory Solutions, NSOFT Company Ltd., South Korea. Her research interests include process mining, process modeling, BPM, robotic process automation, smart factory, and digital transformation.

CHIWOON CHO is currently a Full Professor in industrial engineering with the University of Ulsan, South Korea. He has a lot of industry experiences, including Hyundai Heavy Industries, LG CNS, and Samsung SDS.