UDC: 004.93 https://doi.org/10.15407/jai2024.04.207

V. Hura[1], L. Monastyrskii[2]
[1,2]Ivan Franko National University of Lviv, Ukraine
1, University St., Lviv, 79000
[1]volodymyr.gura@lnu.edu.ua
[2]lyubomyr.monastyrskyy@lnu.edu.ua
[1]https://orcid.org/0009-0007-8781-8970
[2]https://orcid.org/0000-0003-4782-9978

INFLUENCE OF AIR QUALITY MODEL PARAMETERS ON POLLUTION CONCENTRATION

Abstract. Air pollution poses a significant threat to public health, ecosystems, and the global climate. Accurate prediction and effective management of air quality are of paramount importance, which, in turn, rely on sophisticated air quality models. These models integrate a variety of atmospheric parameters to simulate the dispersion of pollutants in the complex urban atmosphere, yet the influence of specific input parameters on predicted pollution concentrations has not been fully elucidated. This comprehensive study assesses how variations in model input parameters can lead to divergent pollution concentration outputs, with the goal of identifying those that are most critical to model accuracy.

Using observational data from air quality monitoring stations in conjunction with meteorological records, the study explores the sensitivity of forecasted pollutant concentrations to fluctuations in model inputs such as emission source strength, atmospheric stability, wind speed and direction, diurnal heating patterns, chemical reaction rates, and boundary layer dynamics. Dispersion models are evaluated across different spatial and temporal scales to gauge their response to environmental variables and topographic features. The performance of these models is also assessed against satellite-derived pollutant measurements to encompass a broader geographical context.

Through the application of numerical simulations and statistical analyses, the study quantifies the relative impact of each parameter. Cross-validation techniques, along with uncertainty quantification methods, are applied to ensure the reliability of the conclusions drawn. The research also incorporates the use of machine learning tools to identify complex patterns in the environmental data that may be missed by traditional modeling approaches.

The abstract concludes that a detailed understanding of influential model parameters is essential for refining air quality predictions. Improvements in the accuracy of dispersion models will enable policymakers and urban planners to make better-informed decisions regarding air pollution control and mitigation strategies. This work forms the foundation for future advancements in the field of atmospheric sciences and encourages continued exploration into the interaction between anthropogenic activities, meteorological phenomena, and air quality outcomes.

Keywords: air quality modelling, pollutant dispersion, environmental parameters, model sensitivity, observational data, statistical analysis, meteorological influences, machine learning.

Introduction

In recent years, the global community has become increasingly aware of the adverse impacts of air pollution on human health, ecosystems, and the climate. Airborne pollutants such as particulate matter (PM), nitrogen oxides (NOx), sulfur dioxide (SO2), ozone (O3), and volatile organic compounds (VOCs) have been linked to respiratory diseases, cardiovascular problems, and various environmental issues including acid rain and global warming. Urbanization, industrial activities, and vehicular emissions intensify these pollution challenges, necessitating refined air quality modeling to better predict pollutant concentrations and inform mitigation strategies.

Air quality models (AQMs) are the primary tools used by environmental scientists to simulate the atmosphere's physical and chemical processes that influence pollutant transport and transformation. These models are essential for forecasting air pollution and for evaluating the potential impact of different pollution control options. The fidelity of AQMs to real-world conditions is highly dependent on their input parameters, such as emission rates, atmospheric composition, and meteorological data. Therefore, understanding the sensitivity of models to these inputs is crucial for enhancing their predictive accuracy.

A key aspect of this exploration is the analysis of meteorological parameters such as temperature, wind speed and direction, atmospheric stability, and humidity. These factors significantly affect the dispersion of pollutants by altering the rate of mixing and dilution in the atmosphere.

In addition to physical factors, this study also recognizes the importance of chemical

reactions occurring in the atmosphere, which can transform primary pollutants into secondary products with different dispersion characteristics and health implications.

Implementation

The implementation provides robust insights into the behavior and distribution of atmospheric pollutants, ultimately enabling more accurate air quality forecasting and more effective pollution management strategies.

Model Selection and Setup:

Selecting appropriate air quality models is a critical step in conducting an air quality assessment study. This process entails a careful examination of several factors, including the objectives of the study, the spatial and temporal scale of interest, the availability of data, computational resources, and the specific pollutants of concern. Below is a detailed approach to model selection and setup for an air quality modeling study:

1. Define Study Objectives:

Clarify the goals of the modeling exercise. This may involve understanding the dispersion of a specific pollutant, assessing exposure levels for public health studies, evaluating the effectiveness of emission control strategies, or forecasting air quality for regulatory compliance.

2. Identify the Scale of Interest:

Determine whether the study is focused on a local (point sources), urban (city scale), regional (across multiple cities or states), or global scale. Each scale may necessitate a different type of model with varying levels of detail and complexity.

3. Review Pollutants of Interest:

Decide which pollutants will be the focus of the study. This could be primary pollutants directly emitted from sources (e.g., NOx, SO2, PM) or secondary pollutants formed in the atmosphere through chemical reactions (e.g., ozone).

4. Evaluate Available Data:

Assess the data available for the study area, including emissions inventories, meteorological data, topographical information, and ambient air quality measurements. The availability and quality of these datasets can considerably influence model choice.

5. Select Model Type:

Consider the most suitable model type for the study's needs:

a. Gaussian plume models are relatively simple models suitable for simulating pollutant dispersion from point sources over short distances and in simple meteorological conditions.

b. Eulerian models divide the study area into a grid and simulate the pollutant dispersion and chemical transformations within each grid cell, making them ideal for regional or urban air quality assessments.

c. Lagrangian models track pollutant "parcels" as they move with the wind, providing detailed simulation of pollutant trajectories and are beneficial when assessing specific emission events or for complex terrains.

d. Chemical transport models (CTMs) incorporate both physical dispersion and chemical reactions, offering a comprehensive approach for simulating the formation of secondary pollutants on various scales.

6. Model Setup:

Once the model type is selected, set it up with appropriate input data:

a. Emissions data: Compile or obtain an emissions inventory that includes the rates, locations, and types of pollutants being released.

b. Meteorological data: Input detailed meteorological information such as wind speed and direction, temperature, radiation, humidity, and boundary layer properties.

c. Initial and boundary conditions: Set up the initial pollutant concentrations and boundary conditions of the model domain as per available observations or predictions.

d. Grid resolution and domain: Choose an appropriate grid resolution and model domain that appropriately balances computational demand with the desired level of detail.

e. Chemical mechanism: If applicable, select the chemical mechanism for the model that accurately simulates the chemical reactions relevant to the study.

7. Calibration Parameters:

Identify the calibration parameters, which may include factors such as deposition velocities, reactivity rates, and background

concentration levels, that can be adjusted to improve model performance against observed data.

8. Perform Test Runs:

Conduct initial test runs to ensure that the model operates as expected, and to identify potential issues with input data or model configuration that need to be addressed before proceeding to full-scale simulations.

Using this targeted approach, researchers can select an air quality model that is well-suited to their study objectives, available data, and computational resources. Proper model selection and meticulous setup are essential to produce credible and actionable results from air quality modeling efforts.

The proposed air quality monitoring system leverages the power of IoT and real-time data to provide users with accurate and up-to-date information about their immediate environment. Fig-1 shows the block of the Pm2.5 data taken each 50 meters from thermal station.

The Gaussian plume model is a simplification of the advection-diffusion equation used to estimate the distribution of a pollutant over space and time from a point source under steady-state conditions. The formula for the concentration C at a downwind location (x, y, z) is expressed as:

$$[C(x,y,z) = \frac{Q}{(2\pi\sigma_y\sigma_z u)}\exp\left(-\frac{y^2}{2\sigma_y^2}\right)\left[\exp\left(-\frac{(z-H)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+H)^2}{2\sigma_z^2}\right)\right]]$$

Where:

- (C) is the concentration at the point of interest,
- (Q) is the pollutant emission rate (mass/time),
- (u) is the mean wind speed along the x-axis (wind direction),
- (σ_y) and (σ_z) are the standard deviations of the pollutant concentration distribution in the crosswind (y) and vertical (z) directions, respectively,
- (H) is the effective height of the emission release (stack height plus plume rise), and
- (x, y, z) are the downwind, crosswind, and vertical distances from the emission source point.

Eulerian model solves the advection-diffusion or continuity equation on a fixed grid over the area of interest. The generic equation in three dimensions is:

$$[\frac{\partial C}{\partial t} + u \cdot \nabla C = \nabla \cdot (K\nabla C) + S - L]$$

Where:

- (C) is the concentration of a pollutant,
- (t) is time,
- (u) is the three-dimensional wind field vector,
- (K) is the turbulent diffusion coefficient matrix,
- (S) is the source term describing the rate of addition of the pollutant,
- (L) is the loss term accounting for chemical reactions and deposition,
- (∇) represents the spatial gradient operator.

It is important to note that Eulerian models often involve coupling atmospheric chemistry and using complex numerical methods to solve the partial differential equations.

Lagrangian model tracks pollution "parcels" as they move through space under the influence of the wind field and turbulent mixing. The change in concentration in a Lagrangian framework is often described by a stochastic differential equation (SDE):

$$[dX_t = u(X_t, t)dt + \sqrt{2K}dW_t]$$

Where:

- (X_t) represents the position of the pollutant parcel at time (t),
- ($u(X_t, t)$) is the velocity field (wind) at position (X_t)and time (t),
- (K) is the turbulent diffusion coefficient,
- (dW_t) is the Weiner process (stochastic term) representing the turbulent fluctuations.

Chemical Transport Models (CTMs):

CTMs are comprehensive models that include atmospheric chemistry and physical transport processes to predict the formation, transport, and fate of chemical species in the

atmosphere. The general equation for a chemical species concentration $(C_i)$is:

$$[\frac{\partial C_i}{\partial t} + \nabla \cdot (uC_i) = \nabla \cdot (K\nabla C_i) + \sum_j R_j C_j + E_i - D_i]$$

Where:

- (C_i) is the concentration of chemical species (i),
- (t) is time,
- $(u)is$ the wind field vector,
- (K) is the diffusion coefficient matrix,
- (R_j) is the reaction rate for a reaction (j) in which (C_i) is involved,
- (E_i) is the emission rate of (i),
- (D_i) is the deposition rate of (i).

CTMs require highly sophisticated numerical techniques to solve the coupled non-linear equations related to chemistry, transport, and other atmospheric processes.

Accessfully implement the IoT-based air quality monitoring system, the following steps are essential:

1. Assemble the hardware components: Connect the pollutant sensors to the ESP32 microcontroller using appropriate analog or digital pins. Ensure that the connections are secure and follow best practices for managing wiring and layout.

2. Configure and program the ESP32: Set up the required settings for wireless connectivity, cloud services API keys, and other relevant configurations. Develop appropriate firmware to acquire data from the sensors, preprocess the data if necessary, and transmit it to the cloud platform in a format compatible with the chosen cloud service.

3. Cloud server setup: Create an account and configure the chosen cloud platform to receive, store, and analyze air quality data from the ESP32 microcontroller, enabling data visualization, tracking, and alerting.

4. Develop a user interface: Create a mobile application or web-based platform that allows users to interact with the system, access real-time air quality data, set up threshold-based alerts, and visualize trends in air quality parameters.

5. Calibration and maintenance: Regularly calibrate the sensors, update firmware, and perform other necessary maintenance tasks to ensure the system's continued accurate functioning and reliability [1].

Fig. 1. Empirical data taken from the location

Model Implementation:

The task of air quality monitoring is to use sensor data to obtain intelligible results about the condition of the air in real-time. Applying a recursive difference approach, we can build a model to estimate the level of

pollutants based on current and previous measurements.

Consider how the speed and direction of the wind affect the level of atmospheric air pollution using formulas from the field of dispersion modelling. The direction and speed of the wind are important for the level of atmospheric air pollution. The wind affects the transport, dispersion, and settlement of pollutants in the atmosphere. The level of pollution in a city can be influenced by aspects such as:

– Transport of pollutants: The wind can carry pollutants from industrial zones, agricultural lands (where pesticides or fertilizers may be used), power plants, roadways, and other emission sources to areas where people live or spend a lot of time.

– Dispersion of pollutants: Strong winds can disperse pollutants, thereby reducing their concentration in a particular zone. At the same time, a weak wind may cause the accumulation of pollutants over urban areas, causing a higher level of pollution.

– Direction of the wind: The direction from which the wind blows can be decisive in determining where pollutants are carried. If the wind blows from pollution zones (such as industrial areas) towards urban districts, an increased level of air pollution may be recorded.

– Wind speed: Wind speed is also important. At high wind speeds, pollutants disperse more quickly and thus, the level of pollution can decrease. On the other hand, at low speeds, especially in combination with a temperature inversion, pollutants can accumulate, increasing the level of pollution.

– Vertical processes: Winds blowing in the vertical plane can help lift pollutants up to higher layers of the atmosphere, where they can disperse over long distances. Also, winds can facilitate the scavenging (deposition) of pollutants, reducing their concentration in the lower air layers.

Assessment of the impact of winds on the level of air pollution requires the application of atmospheric dispersion models that incorporate various meteorological data, pollutant measurements, and information about emission sources. Such models can be used for forecasting AQI and developing strategies to reduce pollution's impact on the population.

To include the effect of wind direction in the model, one needs to introduce the wind velocity vector and its components into the convection-diffusion equation. The simple one-dimensional equation we considered does not consider wind direction, as it is one-dimensional and assumes that the source of pollution, the direction of pollutant spread, and the influence of wind direction have already been considered in establishing the boundary conditions.

In realistic models, the direction and speed of the wind are usually considered as vector fields (u(x,y,z,t)), which change in space and time. In three-dimensional space, the convection-diffusion equation may look like:

$$[\frac{\partial c}{\partial t} + \nabla \cdot (uc) = \nabla \cdot (D\nabla c) + S] \ ,$$

where (u) is the wind velocity vector, which has components in the x, y, and z directions (u, v, and w, respectively).

When we assume a steady-state (stable over time) and one-dimensional case, the equation can be generalized to include the direction of the wind as:

$$[\nabla \cdot (uc) = u\frac{\partial c}{\partial x}] \ ,$$

where (u) is the component of the wind speed in the direction of x, and $(c(x))$is the concentration of pollutants along the direction of the wind.

To integrate the influence of direction into the AQI calculations, one must make assumptions or measurements of the wind and adapt the model accordingly. However, including the advective (wind-caused) term in the model significantly complicates the solution and calculations [4].

From main formula it is evident that the wind speed (u) is significant - it is included in the denominator, thus as the wind speed increases, the concentration of the pollutant (\($C(x,y,z)$ \)) decreases. This means that with fast winds, pollutants disperse more quickly and efficiently, reducing the local airload.

As for the wind direction, it is not explicitly present in this formula, but it

significantly influences the parameters (y) and (z) and the choice of the emission source, which will be decisive for the pollution level at a specific point.

$$\left[\frac{\partial c}{\partial t} + u \cdot \nabla c = \nabla \cdot (\rho D \nabla c + \rho K \nabla c) + S\right]$$

The speed and direction of the wind play a key role in the distribution and dispersion of pollutants in the atmosphere, and therefore, on the AQI level in different places. Calculations based on such formulas can help determine high-risk areas of air pollution and form strategies to address them.

The impact of humidity on PM2.5 can be included in main equation, considering that the PM2.5 concentration changes with humidity. Considering our formula for PM2.5 concentration depending on humidity, the formula can be updated as follows:

$$[\frac{\partial(\rho' C')}{\partial t} + \nabla \cdot (\rho' u C') = \nabla \cdot (\rho' D \nabla C' + \rho' K \nabla C') + S]$$

Here (C') is the PM2.5 concentration, modified to account for the influence of relative humidity $\left((C' = C(1 + \kappa RH)^{\frac{1}{\nu}})\right)$, and (ρ') is the density of PM2.5, which is also determined by the influence of humidity on particles. Fig-2 shows have the Pm2.5 concentration determined by the influence of humidity.

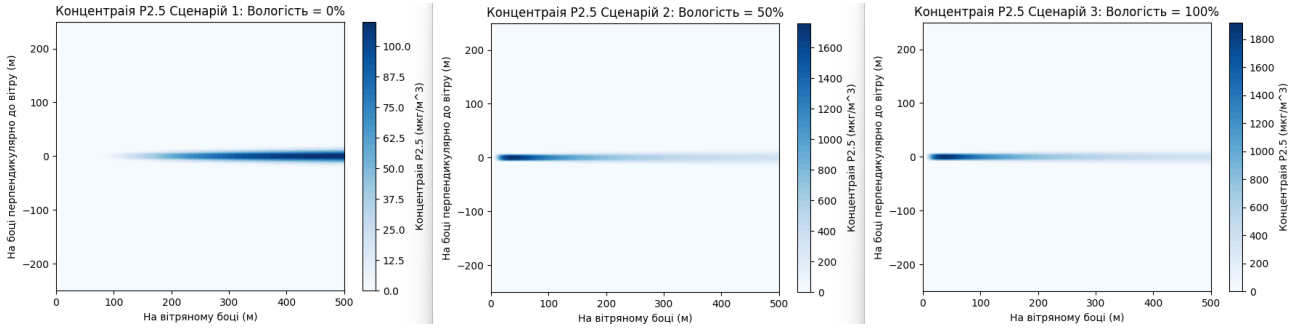

Fig. 2. Pm2.5 determined by the influence of humidity

Temperature is also a critical variable that affects PM2.5. During high temperatures, the rate of chemical reactions in the atmosphere increases, which can lead to the creation of additional PM2.5 through secondary atmospheric processes. Temperature can also affect the stability of the atmospheric layer and thus influence the vertical migration of PM2.5. Incorporating the influence of temperature into main equation is a somewhat complicated task as the temperature can affect various parameters of the equation.

Ways in which temperature can affect the parameters in the equation:

– Density (ρ): In gases, density generally decreases with increasing temperature (assuming pressure remains constant), according to the ideal gas law. At the same time, the density of PM2.5 particles may change depending on their chemical composition and properties.

– Molecular diffusivity (D) and turbulent diffusivity (K): Usually increase with temperature since they are related to the kinetic energy of molecules.

– Concentration (C): Can change depending on the temperature due to the properties of the chemical substances that make up PM2.5, as well as due to various chemical processes in the atmosphere.

The simplest way to use empirical correlation or phenomenological model to indicate how temperature changes one or more parameters in the equation. For example, that molecular diffusivity D and turbulent diffusivity K depend on temperature T using certain correlations or physical principles and express them as functions D(T) and K(T). Similarly, density ρ can be

expressed as ρ(T).

$$[\frac{\partial \rho(r,t)}{\partial t} + \nabla \cdot (\rho u) = \nabla \cdot \big(\rho D(r,t)\nabla c + \rho K\big(T(r,t), RH(r,t)\big)\nabla c\big) + S\big(T(r,t), RH(r,t)\big)]$$

As can be seen, understanding the relationship between such meteorological factors as wind, humidity, and temperature, and PM2.5, is crucial for predicting air quality and developing effective strategies for reducing air pollution.

Transitioning to the PM2.5 concentration graph from distance requires solving this equation for given initial and boundary conditions. In simple cases, this can be done analytically, but often numerical methods are necessary.

Let's assume that we model the diffusion of a pollutant in the atmosphere without considering advection and reaction. In this case, we can simplify the equation to

$$[D\nabla^2 c = \frac{\partial c}{\partial t}],$$

where D is the diffusion coefficient. Knowing the initial concentration, one can solve the diffusion equation for any time t and any distance from the source. We base our initial parameters as Q = 0.1, U = 0.5, H = 10.0, K = 1.5. Where Q is the emission rate (g/s), U is the wind speed (m/s), H is the height of the chimney (m), K is the coefficient of turbulent diffusion (m^2/s). Based on the parameters, we construct a graph of the dependence of PM2.5 concentration on distance (Fig 3).

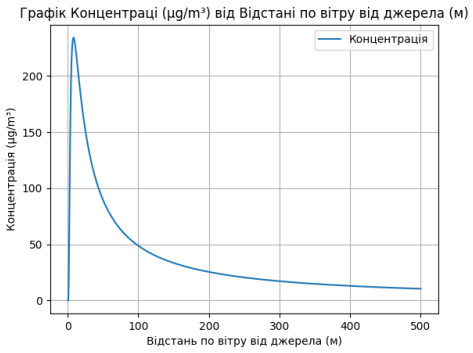

Fig. 3. Graph of PM2.5 (µg/m³) dependence on distance from the source (m)

Thus, we discussed how weather parameters such as humidity, pressure, and temperature affect the distribution and concentration of pollutants in the air, responsible for the Air Quality Index (AQI). Now let's move on to analyzing another important aspect – quantitative modeling of pollutant dispersion.

Applying indicators to the model, we calculate the concentration of pollutants at various distances from the emission source. Equation considers not only the height and wind speed but also air turbulence, so it is possible to predict the distribution of concentrations in both vertical and horizontal directions.

These calculations allow ecologists and engineers to determine the "impact area" of industrial emissions, predict areas of potentially higher pollution levels, and develop measures to reduce the negative impact on the environment. They are also useful in planning the placement of new production facilities and studying the impact of existing plants on the environment and public health.

Analyzing data from Fig 4, several key conclusions can be made about the behavior of the PM2.5 particle distribution model:

– Increasing the emission rate (Q): When Q increases, the maximum concentration also increases proportionally. This corresponds with intuition since a larger amount of pollutants released into the atmosphere increases the concentration near the source. This shows that controlling the emission rate is important for reducing the maximum concentration of pollutants [5,6].

– Changes in wind speed (U): With the increase in wind speed, the maximum concentration decreases. This is because stronger winds spread the pollutants more broadly, thereby reducing their concentration at any given location. This underscores the importance of wind speed in assessing and controlling pollution levels.

– Changes in chimney height (H): A higher emission height ensures better distribution of pollutants in the upper atmospheric layers, resulting in lower concentrations near the ground surface. Increasing H from 5 to 20 meters reduces the

maximum concentration of pollutants.

– Turbulent diffusion coefficient (K): In this analysis, the maximum concentration did not change with the variation of the diffusion coefficient, which may indicate that in the given scenario, with specified emission rate, wind speed, and source height, the impact of the diffusion coefficient K on pollutant concentration is minimal or that it may have a more significant effect on the spread of pollutants beyond the considered distance.

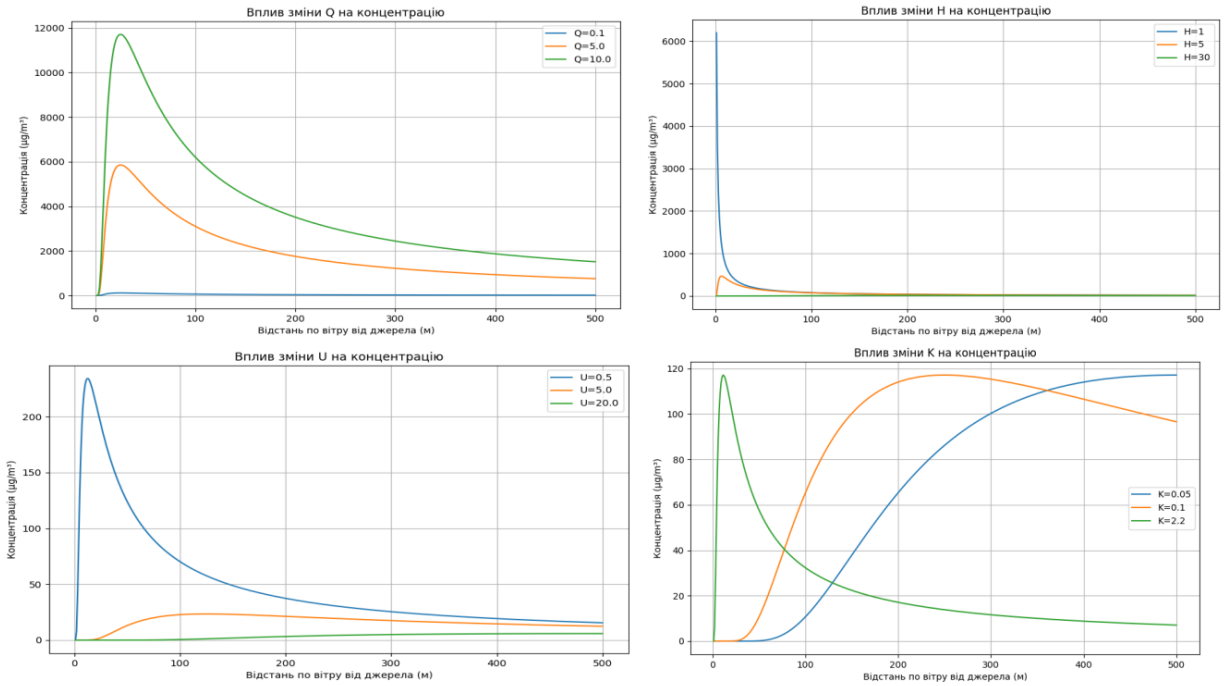

Fig. 4. Graph of PM2.5 (µg/m³) dependence on parameters

Solving and studying the inverse problem

The solution to the inverse problem involves finding effective parameters for air quality determination. The process of searching for coefficients for the model's dispersion parameters is carried out using two optimization methods: the brute force method (`brute`) and the Broyden-Fletcher-Goldfarb-Shanno optimization method (BFGS).

Before starting the optimization process, it is necessary to first define the model that will be used to describe the phenomenon and the objective function that will be minimized or maximized during optimization. In our case, the model is the distribution of the pollutant, describing how lateral and vertical dispersions affect the concentration of emissions at a certain point from the source.

The model parameters that need to be determined and optimized may include:

– Q - mass emission rate of the pollutant in g/s

– U - wind speed in m/s

– H - height of the smokestack in m

– K - diffusion coefficient that reflects the dispersion process of the pollutant in the environment.

Preliminary values for these parameters should be used that are typical for the problem or available from previous research, expert estimates, or technical standards.

The objective function that we optimize in this problem is the sum of the squares of the differences between the measured and predicted concentrations of the pollutant:

$$[\sum_i (C_{\text{pred},i} - C_{\text{meas},i})^2] \ ,$$

where $(C_{\text{pred},i})$ is the predicted concentration at point \(i\), calculated using the model, and $(C_{\text{meas},i})$ is the measured concentration at the same point.

The objective function calculates the deviation of predicted values from measured ones, and the task of optimization is to minimize this deviation by selecting optimal

values for the model's parameters.

Prior to commencing the optimization process, it is critical to gather and prepare the input data that the model will utilize. The following data are required for calculating the dispersion model parameters:

– Measured concentration levels of the pollutant at various points (distance from the emission source).

– Distances from the emission source to the points where measurements were taken.

This data was collected from experimental measurements in CSV format. To upload and process this data, we utilize the pandas library, which allows easy reading of data from various file formats and convenient operation with tables in Python:

```
import pandas as pd
df_loaded = pd.read_csv('pm25_data.csv')
measured_distances = df_loaded['L (m)'].values
measured_concentrations = df_loaded['Pm25 (μg/m³)'].values
```

As a result, we have two arrays: one with distances to measurement points from the emission source (`measured_distances`) and another with the measured concentrations of the pollutant (`measured_concentrations`). These data will serve as the basis for further analysis and optimization of the model's parameters.

The brute force method, also known as the exhaustive search method, is used to find an approximate optimal solution to the optimization problem by systematically reviewing each possible value within a set range.

When using the brute force method, the process has the following stages:

1. Selection of search range: We determine the range of values for each parameter being optimized. The range may be based on prior experience, expert assessments, or the range of observations. Each parameter receives its range, which is then divided into equal intervals (possible parameter values).

2. Enumeration: We systematically calculate the objective function for all combinations of parameter values within the established intervals, i.e., "rummaging through" all possible parameter values and finding the corresponding value of the objective function.

3. Evaluation: We assess the results for each combination of parameters by comparing the measured data with the predicted.

4. Finding the Minimum: We determine the set of parameters that provides the smallest (optimum) value of the objective function. Such a combination will approximate the best optimization problem solution

Although this method is simple and easily implementable, it can be very resource intensive as the number of calculations increases exponentially with each additional parameter or reduction in the exhaustive search grid step. Also, it provides only an approximation of the optimal solution and does not guarantee finding the absolute minimum of the objective function.

Example of using the `brute` method for optimizing one parameter using the `SciPy` library:

```
# Result of the brute force method
result = brute(
   objective_function,
   ranges=[(min_value, max_value)],  # Range of values for the parameter
   Ns=number_of_steps,               # Number of steps in the range
   full_output=True,
   finish=None                       # No additional optimizers after brute force
)
```

In the code, `min_value` and `max_value` are the search range limits for the parameter, and `number_of_steps` is the number of steps in this range, determining the discretization bound for the enumeration process. The `brute` function returns the optimization result, including the approximated optimum parameter value and the minimum value of the objective function.

Optimization with the BFGS Method.

The BFGS method (named after Broyden, Fletcher, Goldfarb, and Shanno) is an iterative method for finding the minimum of an objective function. It belongs to the class of quasi-Newton methods and is one of the most popular methods for unconstrained optimization functions.

The main steps of the BFGS optimization process are as follows:

- Initial Point Selection: We determine a starting point for the parameters from which the optimization process will begin. Often,

initial values are taken from the results of prior optimization, for example, using the brute force method.

– Gradient Calculation (derivative): The objective function is calculated along with its gradient—a vector of its first derivatives with respect to all parameters. The gradient indicates the direction of the steepest increase of the function.

– Direction of Search Determination: The direction in which the function is most likely to decrease is determined based on the gradient and previous calculations.

– Step Size Calculation: The size of the step is determined using a line search, which "probes" the objective function in the chosen direction to find the optimal step length.

– Parameter Update: Parameters are updated by the amount of the step in the chosen direction.

– Stop Conditions Check: If the parameters have changed only slightly or the objective function has ceased to decrease, the optimization process concludes, and a local minimum is found.

– Iteration: If the stopping conditions are not met, the process from steps 2 to 6 is repeated.

The BFGS is a more calculation-intensive method than brute force but is much more efficient for exploring multi-parametric spaces and typically converges to the optimum quickly [2,3].

To perform optimization with the BFGS method, the `minimize` function from the `SciPy` library can be used, specifying 'BFGS' as the optimization method:

```
initial_params = [initial_Q, initial_U, initial_H, initial_K]
# Perform the optimization
result = minimize(
   objective_function,
   initial_params,
   method='BFGS'
)
```

Here, `initial_params` is an array with the initial values of the parameters to be optimized. `result` is an object containing information about the optimization results, including the optimal parameters (`result.x`) and the value of the objective function at the optimum point (`result.fun`).

As a result of optimizing the parameters for the dispersion model, we obtained sets of parameter estimates that allow us to predict PM2.5 concentration at various distances from the emission source with a high degree of accuracy. This model reflects the real behavior of pollutants in the atmosphere and can be successfully applied to assess air quality.

Based on the optimization results, we observe that:

- The values of MSE and MAE are sufficiently low, indicating minor discrepancies between observed and modeled concentrations.

- Further analysis of the residuals shows they do not exhibit noticeable trends or patterns, which is evidence of the model's adequacy.

- A high R^2 value, close to 1, demonstrates that the model accurately predicts changes in concentration.

- Tests for homoscedasticity did not reveal any statistically significant indications of heterogeneity in the residuals, which speaks to the model's robustness under different conditions.

Considering the obtained results (Q: 10.0, U: 0.1, H: 5.0, K: 0.01), we can be confident that the optimized model is a reliable tool for determining air pollution levels.

Conclusion

In conclusion, the application of optimization techniques such as the brute force method and the BFGS method has proven to be effective in solving the inverse problem of determining effective parameters for air quality models. The process involved careful preparation of the necessary input data, selection of an appropriate dispersion model, and definition of a suitable objective function to minimize the discrepancy between modeled and measured pollutant concentrations.

By utilizing the brute method, we were able to obtain a rough estimate of the optimal parameters for the model, which provided a valuable starting point for further refinement. While the brute force method is straightforward and easy to implement, its exhaustive nature makes it computationally expensive, particularly for complex models with multiple parameters.

Subsequent fine-tuning using the more advanced BFGS optimization algorithm allowed for a more detailed and efficient search for the optimal values of the model parameters. The BFGS method not only offered a faster convergence to the solution but also produced a set of parameters that significantly improved the model's predictive accuracy, as evidenced by the low values of mean squared error (MSE) and mean absolute error (MAE), as well as a high coefficient of determination (R^2).

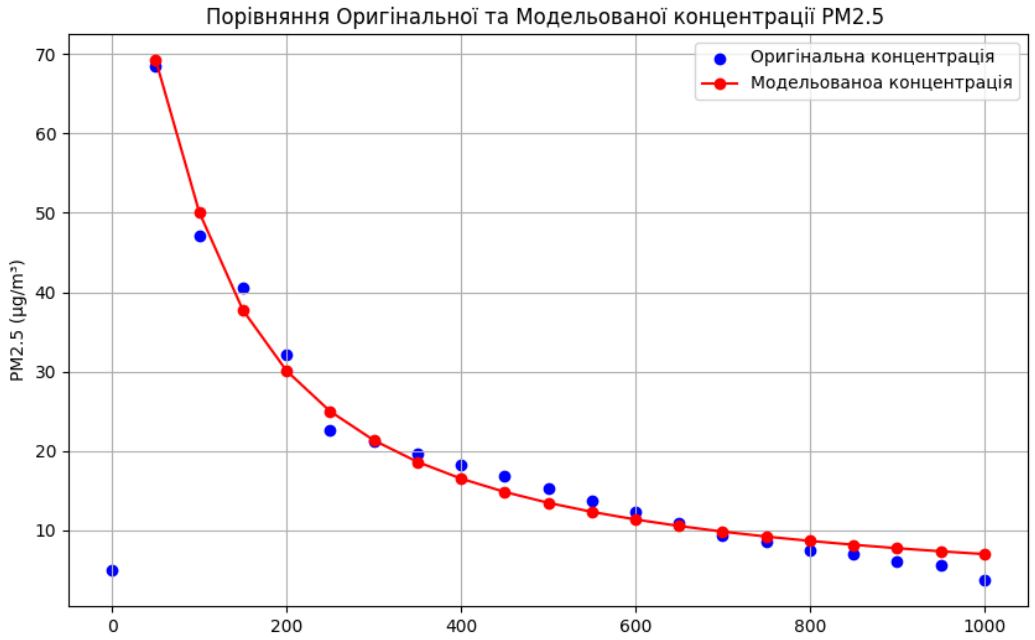

Fig. 5. Historical of air quality data by the time

The graphical analysis provided further confirmation that the optimized model could effectively characterize the concentration of PM2.5 pollutants as a function of distance from the emission source. The lack of systematic patterns in the residuals underscored the model's reliability, while tests for homoscedasticity supported its robustness across various conditions.

Ultimately, the combination of optimization methods and careful data analysis enabled the development of an air quality model capable of accurately estimating pollutant concentrations. Such models are crucial for environmental management and public health, as they allow for the assessment of pollution levels and the identification of necessary mitigating actions. They also serve as essential tools for policymakers and stakeholders to address air quality issues effectively.

References

1. Huang, R. (2014) High secondary aerosol contribution to particulate pollution during haze events in China. Nature, 514(7521), 218-222. Otry`mano http://dx.doi.org/10.1038/nature13774

2. Numerical Optimization, Nocedal, J. and Wright, S., 2006, Springer-Verlag New York, eBook ISBN: 978-0-387-40065-5

3. Thuillard M. Adaptive multiresolution search: How to beat brute force?. International Journal of Approximate Reasoning. 2004. Vol. 35, no. 3. P. 223–238. URL: https://doi.org/10.1016/j.ijar.2003.08.003 (date of access: 08.04.2024).

4. United States. Environmental Protection Agency. EPA organization. [Washington, D.C : U.S. Gov't. Print. Office, 1972.

5. Дифузія атмосферних домішок у ближній зоні від точкового висотного джерела в конвективному граничному шарі / О.Я. Скриник // Наукові праці Українського науково-дослідного гідрометеорологічного інституту: Зб. наук. пр. — 2009. — Вип. 258. — С. 43-56. — Бібліогр.: 12 назв. — укр.

6. Vaishali, Verma G., Das R. M. Influence of Temperature and Relative Humidity on PM2.5 Concentration over Delhi. MAPAN. 2023. URL: https://doi.org/10.1007/s12647-023-00656-8

The article has been sent to the editors 08.04.24.
After processing 15.10.24.
Submitted for printing 30.12.24.